ASSESSING THE PREVALENCE OF SUSPICIOUS ACTIVITIES IN ASPHALT PAVEMENT CONSTRUCTION USING ALGORITHMIC LOGICS AND MACHINE LEARNING

by

Mostofa Najmus Sakib

A thesis

submitted in partial fulfillment

of the requirements for the degree of

Master of Science in Civil Engineering

Boise State University

August 2020

BOISE STATE UNIVERSITY GRADUATE COLLEGE

DEFENSE COMMITTEE AND FINAL READING APPROVALS

of the thesis submitted by

Mostofa Najmus Sakib

Thesis Title: Assessing the Prevalence of Suspicious Activities in Asphalt Pavement Construction Using Algorithmic Logics and Machine Learning

Date of Final Oral Examination: 09 July 2020

The following individuals read and discussed the thesis submitted by student Mostofa Najmus Sakib, and they evaluated their presentation and response to questions during the final oral examination. They found that the student passed the final oral examination.

Mojtaba Sadegh, Ph.D.	Chair, Supervisory Committee
Debakanta Mishra, Ph.D.	Co-Chair, Supervisory Committee
Yang Lu, Ph.D.	Member, Supervisory Committee

The final reading approval of the thesis was granted by Mojtaba Sadegh, Ph.D., Chair of the Supervisory Committee. The thesis was approved by the Graduate College.

DEDICATION

Dedicated to my loving parents and all the researchers who devoted their life in the path of science

ACKNOWLEDGMENTS

First and foremost, I would like to convey my sincere gratitude towards my advisor Dr. Mojtaba Sadegh and co-advisor Dr. Debakanta Mishra for their constant support towards this challenging yet memorable journey of learning. They have constantly supported and motivated me to reach the goal of completing this thesis. Their mentorship has helped me evolve as an independent researcher. I will always cherish the memory of this journey and the freedom to research. The amount of learning would be helpful for me towards a better future.

I would also like to express my appreciation to Dr. Yang Lu for generously serving on my thesis committee. Alongside my advisors and committee members the other faculty members of the Civil Engineering Department at Boise State University have always been helpful and never turned me away. I have learned a lot about programming and Machine Learning from courses taught by Dr. Dylan Mikesell and Dr. Michael Ekstrand. Their courses prepared me to complete the tasks of this thesis.

This research is supported by the Idaho Transportation Department (ITD) through the project no RP 275. The author is grateful for the assistantship granted through the project. Project manager of this project, Mike Copeland extensively helped me to understand the complex datasets and problems associated with this research. It would have been impossible for me to finalize the tasks without his explanations.

Finally, my gratitude and thanks to my family and all the graduate students of the Civil Engineering Department at Boise State University.

ABSTRACT

Quality Control (QC) and Quality Assurance (QA) is a planned systematic approach to secure the satisfactory performance of Hot mix asphalt (HMA) construction projects. Millions of dollars are invested by government and state highway agencies to construct large-scale HMA construction projects. QC/QA is statistical approach for checking the desired construction properties through independent testing. The practice of QC/QA has been encouraged by the Federal Highway Administration (FHWA) since the mid 60's. However, the standard QC/QA practice is often criticized on how effective such statistical tests and how representative the reported material tests are. Material testing data alteration in the HMA construction sector can render the QC/QA practice ineffective and shadow the performance of asphalt pavements.

The American Society of Civil Engineers estimates that $340 billion is lost globally each year due to corruption in the construction industry. Asphalt pavement construction consists of several sectors, including construction and transportation, which are prone to potential suspicious activities. There is approximately 18 billion tons of asphalt pavement on American roads, which makes the costs of potential suspicious activities unacceptably large.

The Idaho Transportation Department (ITD) relies on contractor-produced QC test results for the payment of the HMA pavement projects. In 2017, a case study by FHWA found some unnatural trends where 74% of the ITD test results didn't match with the contractor results. ITD's approach to track down the accuracy of mix design and volumetric

test data set the off-stage of this research to mark out instances of suspicious activities in asphalt pavement projects.

The first objective of this research was to develop algorithmic logics to recognize the patterns of discrepancies in agency- and contractor-produced QC/QA test results. This was possible with a unique dataset that ITD collected from several dozen HMA projects, in which all instances of data entry into the material testing report file was recorded in the background, without the operators' knowledge. My solution was bifurcated into development of an algorithm combining the logics to automatically detect and categorize suspicious instances when multiple data entries were observed. Modern data mining approaches were also used to explore the latent insights and screen out suspicious incidences to identify the chances of suboptimal materials used for paving and extra payment in HMA pavement projects. I have also successfully prompted supervised machine learning techniques to detect suspicious cases of data alterations.

The second step of this research was to calculate the monetary losses due to data alteration. I replicated ITD's procedure for HMA payment calculation, and quantified payment-related parameters and associated payment for each project for two cases: 1. when the first parameter value categorized as Suspicious Alteration (S.A.) was used for payment calculation, and 2. when the last S.A. parameter value was used for payment. It was evident from my findings that there has been overpayment on construction projects across Idaho due to material testing data alterations. Overall, based on the available audit data, I found that overpayments have ranged from $14,000 to $360,000. Further analysis showed that alteration of each major material testing parameter's value can cause roughly $1,000 to $5,000 overpayment. I also note that data alteration did not always cause monetary gains.

Other possible motives may include passing Percent Within Limit (PWL) criteria and precision criteria. Throughout the research, I strive to automate a suspicious activity detection system and calculate the associated excessive payment.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF ABBREVIATIONS

AI	Artificial Intelligence
DOT	Department of Transportation
FHWA	Federal Highway Administration
HMA	Hot Mix Asphalt
ITD	Idaho Transportation Department
ML	Machine Learning
P.C.	Plausible Correction
QA	Quality Assurance
QC	Quality Control
S.A.	Suspicious Alteration

CHAPTER 1: INTRODUCTION AND BACKGROUND

Introduction and Research Problems

The Federal Highway Administration (FHWA) has been encouraging contractors and Departments of Transportation (DOTs) from the mid 1960's to use statistics-based quality control and quality assurance (QC/QA) to ensure pavement products fulfill the design specifications provided by the highway agency (Akkinepally & Attoh-okine, 2006). QC/QA specifications are a combination of end result specifications and materials and methods specifications (Akkinepally & Attoh-okine, 2006; Transportation Research Board Glossary, 2018). Contractors and highway agencies typically collect material testing data and later statistically compare them through F & T tests to ensure the required quality of the product used in transportation infrastructures is achieved. DOTs, usually being short of physical and financial resources, have limited capability of sampling and testing all projects. Hence, often the material testing data for the department is also collected by third party contractors. With limited control of the DOT over material testing and reporting, the collected datasets are vulnerable to probable data alteration. Such a case of data alteration evidence was seen in a recent study by the Idaho Transportation Department (ITD). Under the scope of this study, a research project was initiated to classify the data alteration instances to plausible corrections and suspicious alterations. While I refrain from using the term "fraud" while referring to detected suspicious data alterations in this study, given that a pure data mining approach is not able to detect/classify fraud, I provide a brief literature review on fraudulent activities in various sectors in the following paragraphs. This helps in putting a worst-case scenario – which might or might not have materialized – into broader context.

Fraud is a willful act that can be associated with the intention of gaining financial benefit, which is obviously against the law (Wang et al., 2006; Ngai et al., 2011). The World Bank estimates that fraudulent activities cost the global economy around $2.6 trillion annually, which is equal to 5% of the global gross domestic product (GDP) in 2016 (United Nations, 2018). Loss of public assets through corruption can significantly affect the limited resources of a country. Any damage in the public fund through corruption can increase the revolutionary feelings among people as they see their tax dollars being used in wrongful ways (Power and Taylor, 2011). This wicked problem may lead down the path to more corruption and can disrupt economic progress.

Corruption and fraud have been a critical global issue in the construction and transportation sectors. Despite the existence of corruption in public construction projects, it is one of the less attended sectors in efforts against corruption. Corruption acts as a barrier against the growth of developing countries and the continuation of growth in developed countries (Treisman, 2007; Tabish and Jha, 2011; Loosemore and Lim, 2015; Locatelli et al., 2017). Transparency International identifies the construction sector as the largest corrupted sector compared to other sectors such as banking, insurance, securities, etc. The construction sector is prone to corruption because of the complex and convoluted involvement of different parties (Krishnan, 2009).

While traditionally various investigation methods have been used for detection of fraud, the application of Machine Learning (ML) and Artificial Intelligence (AI) in unearthing fraud and corruption is receiving a lot of attention in the literature in recent years (Stockemer, 2018; Sun and Medaglia, 2019; Tang et al., 2019). According to Forbes magazine, there are 2.5 quintillion bytes of data produced each day at the current pace. It is not humanly possible to monitor or foresee fraudulent attempts within the large volume of data, although even small data alteration might lead to significant losses. With the advancement of modern data analytics capacity, the application of AI in the public sector has received a growing interest (de Sousa et al., 2019). AI can significantly contribute to untangling fraud related evidences by working closely with large scale datasets (Lima and Delen, 2020).

Research Objectives and Tasks

The initial objective of this research was to develop a logic-based algorithm to distinguish between instances of Plausible Correction (P.C.) and Suspicious Alteration (S.A.) in audit data from material testing reports of several Hot Mix Asphalt (HMA) projects. Also, applicability of well-established ML algorithms in classifying large-scale audit data from construction sites was tested in this research. Audit data were acquired from ITD, which included all instances of value entry for mix design parameters in HMA projects. A VBA macro was encoded by ITD into the Excel reporting files that registered all data entries for each parameter in each test, and hence provided extensive and invaluable information about data alteration in material testing reports. All material testing data were reported to ITD through these excel files. These VBA encoded files included all the audit data, whether a value for any parameter was entered once or multiple times. This presented the opportunity for taking a close look at all the modifications/alterations for any reported parameter.

A 2017 forensic case study by ITD first highlighted the inconsistent mix design parameter data from QC/QA test results. It was suspected from their study that the collected data might have been reported inappropriately several times. While various reasons can explain data alterations, the worst-case scenario corresponds to the situation where data alteration is directed to achieve increased pay factors or to pass substandard projects. The objective of this research was to classify construction audit data into either green (Plausible Correction / P.C.) or red (Suspicious Alteration / S.A.) zone. This probable data alteration can often lead to not only financial losses but also poorly paved roadways. Alongside the classification task, the monetary loss associated with suspicious alteration of material testing report data in multiple HMA projects across Idaho was calculated.

Research goals include:

i. Development of a logic-based algorithm to classify repetitive data entries in construction projects' audit data to P.C. and S.A.

ii. Application of ML algorithms to evaluate whether or not patterns recognized by logic-based algorithms are evident to machine as well.

iii. Monetary analysis to quantify the amount of economic loss due to S.A. cases for the analyzed projects.

Research tasks carried out to accomplish the above described goals were:

i. *Review of existing literature*: Existing literature on data alteration and fraudulent attempts were studied to understand the underlying reasons for such acts on a global scale. Unfortunately, there is not much research available on data alteration/manipulation in the construction sector. A majority of the fraudulent cases have been registered in banking, insurance, securities, commodities, and the corporate sectors. Other aspects of corruption in the literature, including bribery, embezzlement, kickbacks in construction sector were also checked.

ii. *Data organization and cleaning:* I received audit data files from ITD, which had the recorded, altered data from material testing results. The dataset was large in volume and needed proper "cleaning" before the application of logic-based algorithms. Additionally, more data, i.e., test summary, lot information, volume of material, etc. was organized/cleaned for the later part of the analysis.

iii. *Development of algorithmic logics:* At the initial part of the research, one project data was examined manually to untangle the general trend of data alteration. This resulted in several cases of probable data alteration as well as typing errors. Subsequently, more projects were manually analyzed to see if such patterns exist in different projects, and if there are other patterns in the altered data. Later these findings were converted to if/else cases and assembled to an algorithm to detect similar cases for all projects.

iv. *Application of supervised ML algorithms:* Alongside the development of customized algorithm I also focused on the application of ML algorithms to assess the effectiveness of strategies of the logic-based work. Several renowned ML techniques, including K-nearest neighbor, logistic regression, decision tree/random forest, neural network, support vector machine, and discriminant analysis, were used on the audit data. Due to the unavailability of categorized data, I used my previously classified data as the training/validation source of ML classifications. None of the projects had a large enough dataset to fit a machine learning model; therefore, I merged data from all projects to train/test the models. This task further confirmed the logics that were developed in the earlier step to be consistent.

v. *Monetary analysis:* In this step, I quantified the amount of money that should have been paid if there was no data alteration. The idea was to check if there has been any overpayment in the asphalt pavement projects.

vi. *Comparison of overpayment and pass/fail of payment parameters:* At the final stage of this thesis the amount of money that has been overpaid for each project was reported. There were also some pass/fail tests for the payment parameters prior to the payment calculation. An overall comparison of those pass/fail tests is also shown for each project.

Findings from the tasks carried out under the scope of this master's thesis research have been documented in two manuscripts prepared to be submitted to peer-reviewed journals. Table 1-1 lists the different tasks and how they were divided between the two manuscripts.

Table 1-1 Tasks carried out under the scope of the current master's thesis research, and corresponding manuscripts

Tasks	Name	Manuscript
1	Developing algorithmic logics to classify construction projects' Audit data to Plausible Correction (P.C) and Suspicious Alteration (S.A.)	Manuscript #1
2	Application of well-established supervised machine learning algorithms to detect probable P.C. & S.A.	
3	Monetary analysis to quantify the amount of economic loss due to S.A. cases	Manuscript #2

This thesis consists of four chapters. The first chapter presents a detailed background of the research problem, and outlines the research questions and hypotheses that were addressed through this master's thesis research. Brief descriptions of different tasks carried out to accomplish the overall research goal have been provided. Descriptions of the tasks and the corresponding findings have been divided into two technical manuscripts, which constitute Chapters Two and Three of the current thesis. The first manuscript (Chapter Two of the thesis) details the development of the logic-based algorithm to distinguish between different categories of data alteration during quality control and acceptance testing. Audit data provided by ITD has been used to identify different data alteration patterns and for the development of the logic-based algorithm. Supervised ML techniques played a supporting role during this task. Different ML approaches were implemented, and their accuracies were compared against the previously developed algorithmic logics.

The second manuscript (Chapter Three of the thesis) focuses entirely on quantifying the financial impact of data inconsistencies encountered in HMA quality control and acceptance testing. The primary objective was to highlight the extent of impact that data inconsistencies can have on the overall costs to state highway agencies. Chapter Four summarizes major findings from the current study and provides recommendations for future research that can lead to the implementation of improved quality control and acceptance testing practices by state and local highway agencies.

References

Akkinepally, R., and Attoh-okine, N. (2006). Quality control and quality assurance of hot mix asphalt construction in Delaware. Delaware Center for Transportation, University of Delaware, Report DCT 173.

Krishnan, C. (2009). Combating corruption in the construction and engineering sector: The role of Transparency International. Leadership and Management in Engineering, 9(3), 112-114.

Lima, M. S. M., and Delen, D. (2020). Predicting and explaining corruption across countries: A machine learning approach. Government Information Quarterly, 37(1), 101407.

Locatelli, G., Mariani, G., Sainati, T., and Greco, M. (2017). Corruption in public projects and megaprojects: There is an elephant in the room! International Journal of Project Management, 35(3), 252–268.

Loosemore, M., and Lim, B. (2015). Inter-organizational unfairness in the construction industry. Construction Management and Economics, 33(4), 310–326.

Ngai, E. W. T., Hu, Y., Wong, Y. H., Chen, Y., and Sun, X. (2011). The application of data mining techniques in financial fraud detection: A classification framework and an academic review of literature. Decision Support Systems, 50(3), 559–569.

Power, T. J., and Taylor, M. M. (2011). Corruption and Democracy in Brazil: The Struggle for Accountability. University of Notre Dame Press, Notre Dame, Indiana.

de Sousa, W. G.,, de Melo, E. R. P., Bermejo, P. H. D. S., Farias, R. A. S., and Gomes, A. O. (2019). How and where is artificial intelligence in the public sector going? A literature review and research agenda. Government Information Quarterly, 36(4), 101392.

Stockemer, D. (2018). The internet: An important tool to strengthening electoral integrity. Government Information Quarterly, 35(1), 43–49.

Sun, T. Q., and Medaglia, R. (2019). Mapping the challenges of Artificial Intelligence in the public sector: Evidence from public healthcare. Government Information Quarterly, 36(2), 368–383.

Tabish, S. Z. S., and Jha, K. N. (2011). Analyses and evaluation of irregularities in public procurement in India. Construction Management and Economics, 29(3), 261–274.

Tang, Z., Chen, L., Zhou, Z., Warkentin, M., and Gillenson, M. L. (2019). The effects of social media use on control of corruption and moderating role of cultural tightness-looseness. Government Information Quarterly, 36(4), 101384.

Transportation Research Board (2018). Glossary of Transportation Construction Quality Assurance Terms, Seventh Edition. Transportation Research Circular E-C235. The National Academies of Sciences, Engineering, and Medicine.

Treisman, D. (2007). What Have We Learned About the Causes of Corruption from Ten Years of Cross-National Empirical Research? Annual Review of Political Science, 10(1), 211–244.

United Nations. (2018). Meetings Coverage & Press Releases: Global Cost of Corruption at Least 5 Per Cent of World Gross Domestic Product, Secretary-General Tells Security Council, Citing World Economic Forum Data. Accessed at: https://www.un.org/press/en/2018/sc13493.doc.htm, retrieved on August 1, 2020.

Wang, J. H., Liao, Y. L., Tsai, T. M., and Hung, G. (2006). Technology-based financial frauds in Taiwan: Issues and approaches. Conference Proceedings - IEEE International Conference on Systems, Man and Cybernetics, 2, 1120–1124.

CHAPTER 2: DEVELOPING ALGORITHMIC LOGICS AND APPLICATION OF MACHINE LEARNING TO CLASSIFY CONSTRUCTION PROJECTS' AUDIT DATA TO PLAUSIBLE CORRECTION AND SUSPICIOUS ALTERATION

Introduction

Construction of a cost-effective, well-performing pavement section is largely dependent on sound construction practices and material quality control. The process of Quality Control (QC)/ Quality Assurance (QA) involves QC testing by the contractor, and acceptance testing by the state Department of Transportation (DOT) to ensure all required standards are met. QC/QA is a combined procedure consisting of materials/methods or end-result specification. Depending on a particular DOT's policies, payments to contractors are made upon comparison of the quality control and acceptance testing data. As these tests are conducted on random samples collected from the same population, it is expected that the test results would 'agree' with each other. For cases where the results do not 'agree', further investigation is required to identify the source of the discrepancy. Inconsistencies in quality control and acceptance testing data can ultimately lead to poor-performing pavements.

A 2017 forensic investigation into Hot Mix Asphalt (HMA) projects' material testing reports in Idaho revealed unnatural trends and inconsistencies between the data that contractors reported and the data that ITD collected. In fact, only 26% of the contractor results were in good agreement with the ITD-produced test results. This motivated my study to investigate the prevalence and sources of data inconsistencies in HMA quality control and acceptance testing data. To further investigate the reasons behind this data discrepancy, ITD engineers inserted a VBA Macro code into ITD's material testing report Excel files, which recorded all instances of data entry for each parameter in the background (not visible to the operator). This audit data was then made available to the Boise State research team. All recorded instances of data entry were investigated for 15 available projects from the year 2018, and modern data mining and logic development approaches were implemented to classify repetitive data entries into two categories: (1) Plausible Correction (PC); and (2) Suspicious Alteration (S.A.). Through extensive manual analysis of audit data, I found patterns of P.C. and S.A. instances, which were then coded into logic-based computer programs that automatically classified all audit data. Note that the audit files comprised data from both QC as well as acceptance testing. Therefore, the data files analyzed in this study may have been generated by the contractor (during QC testing) or the agency (during acceptance testing). Also, it is important to note that both contractor and agency hired third-party testing laboratories to run the tests on multiple occasions. Therefore, the test data could also have been generated by a third-party laboratory. Nevertheless, for the purpose of this study, if a data file was finally signed by the agency (DOT), it was treated as agency (DOT) data irrespective of whether the tests were physically run in the DOT lab or a third-party lab. Similarly, if a data file was submitted by the contractor, it was treated as contractor' data even if the tests might have been physically performed at a third-party testing laboratory. To avoid inherent bias during the data analysis and interpretation, this thesis uses the names "Entity 1" and "Entity 2" to refer to agency and contractor data, not necessarily in the same order. In other words, it has not been disclosed to the reader whether Entity 1 represents data from the agency or contractor. The same is the case for Entity 2.

Overall, I found that there were 2,268 instances where the alteration of 595 unique parameters by Entity 1 could be classified as S.A. Similarly, considering the data for Entity 2, 387 unique parameters were altered a total of 1,266 times, with the alterations classified as S.A. Similarly, considering P.C. occurrences, Entity 1's data accounted for 316 unique parameters being altered 660 times; from Entity 2's audit files, the alteration of 280 unique parameters for a total of 587 times can be categorized as P.C. Further, I evaluated the potential of supervised Machine Learning (ML) algorithms to detect the patterns that were captured in the logic-based analysis. Trained over combined data from all projects, Support Vector Machine and Discriminant Analysis models exceeded accuracy rates of 70%, pointing to their ability to observe similar patterns in the data as those manually set. Further, I pose that if large homogenous data (e.g. from one large project rather than from multiple projects) were used to train the models, the model performances could have improved significantly.

Background and Problem Statement

QC acts as a checklist of procedures to confirm the quality of a paving work based on certain specifications set by highway agencies in the contract documents. QC processes are required to be followed by the contractors to ensure the longevity of a newly paved work is secured. Before formally accepting a project, typically, a product is verified by the state/contracting agencies through sampling/testing or inspecting to identify products compliance with the product requirement. QC and acceptance can be jointly defined as QC/QA, which includes evaluation of design, development of plans and specifications, awarding of contracts, and maintenance, among others, to ensure satisfactory performance.

Federal agencies instruct the state departments of transportation (DOTs) to maintain a QA program to carefully inspect the materials used for highway/transportation infrastructure (Coenen et al., 2019). DOTs generally follow different standard specification/test results to evaluate the pavement and quantify the payment (Newcomb et at., 2016; Al-Khayat et al., 2020).

QC/QA is important to maintain the quality and meet the specified quality thresholds. Any deviation from the design specifications can result in sub-standard work and reduce the life span of HMA pavements. QA typically follows a statistics-based approach, i.e. F & T test, to test whether or not contractor-reported QC material testing data and those of the state DOT come from the same population (Coenen et al., 2019). Although passing the agreement tests should ensure a good quality product, it is important that the reported test data be representative of the actual material used for pavement. Examples of data alteration in material testing reports have been recently detected in ITD's investigations. While I refrain from using fraud for the detected suspicious data alterations in this study, given that a pure data mining approach is not able to detect/classify fraud, I provide a brief literature review on fraudulent activities in various sectors in the following paragraph. This helps in putting a worst case from this research scenario – which might or might not have materialized – into broader context.

According to the formal definition of the Oxford dictionary, fraud is an act of deception, an intentional concealment, omission or perversion of truth, to (1) gain unlawful or unfair advantage, (2) induce another to part with some valuable item or surrender a legal right, or (3) inflict injury in some manner. Willful fraud is a criminal offense that calls for severe penalties. The Association of Certified Fraud Examiners (ACFE, 2010) classifies fraud cases to asset misappropriation, corruption, and financial statement fraud. The occurrence of fraud is widespread in sectors like banking, insurance, securities, health, commodities mass market, and the corporate sector (Atwood et al., 2006; Srivastava et al., 2008; Perols, 2011; Perols and Lougee, 2011; Markelevich and Rosner, 2013; West et al., 2015; Perols et al., 2017; Jain and Shinde, 2019).

Recently, media and the public have shown a surge of interest in revealing and preventing corruption in the construction industry. A report in the New York Times (Bagli, 2018) stated investigators eye a possible $100 million in construction fraud. An executive of a large construction company anonymously claimed such a big amount of overpayment in New York as part of bribery, bid-rigging and kickbacks. Another article published in Oregon Public Broadcasting (Manning, 2019) reported that construction fraud was filed against a contractor working on school construction in Portland. This fraudulent case was responsible for nearly $3 million in construction overpayments. A similar case was seen in a billion-dollar school modernization project where three contractors were accused of fraud (Craig, 2019). All of them were accused of “pass-through” contracts where they allowed a minority owned company to be receiving illicit payment without completing any sort of works. Similar cases were also reported globally. A forensic investigation on a construction company in Toronto revealed an $80 million trail of phony invoices by allegedly mimicking the names of legitimate sub-contractors on several key projects (Harvey, 2019). Such works resulted in contractors stopping their work, suppliers shutting down for no payment and finally walking to the path of bankruptcy. The investigation of bankruptcy revealed the alleged fraud payments running from 2011. Further, China demolished three high-rise buildings as part of anti-corruption where it was stated as a “serious breach of planning regulations" that posed a major safety risk (Hewitt, 2015). Based on the original plan two of the three buildings were supposed to be of 31 and the other one 35 floors, but after finishing they were found to be 41, 58 and 65 floors high, respectively. One of them was a total of 88 meters taller than it should have been.

The American Society of Civil Engineers estimated that corruption consumes $340 billion (U.S. dollars) each year in the global construction industry (Sohail and Cavill, 2008; Kyriacou et al., 2015). The construction industry indeed has a reputation for corruption, asset misappropriation, and bribery across the globe (Zarkada-Faser and Skitmore, 2000; Sohail and Cavill, 2008). Corruption in construction takes several forms, including bribery, embezzlement, kickbacks, and fraud. The Organization for Economic Co-operation and Development (OECD) reports extraction, construction, and transportation sectors to be the leading corrupt sectors in the world based on a study of over 400 cases worldwide (Robertson, 2014). There are several causes of fraudulent activities, including conflict of interests, tight margins, monopolistic service delivery, political interference, fragmented nature, low transparency in project selection, involvement of multiple stakeholders in a complex structure, variety of human psychological behavior preferences, large flow of public money, and competitive tendering process (Rodriguez et al., 2005; Sohail and Cavill, 2008; De Jong et al., 2009; Gunduz and Onder, 2012; Nordin et al., 2013).

Asphalt pavement construction projects involve extraction, construction and transportation sectors, making them vulnerable to fraudulent activities. The National Asphalt Pavement Association (NAPA, n.d.) reports that in 2013 state and local governments spent more than $110 billion and the federal government spent $46 billion on the nation's highways asphalt pavement, pointing to massive public tax dollars invested in this sector and highlighting the importance of scientific investigation of potentially suspicious activities in this sector.

Many of the state transportation agencies, including the Idaho Transportation Department (ITD), rely on contractor-produced Quality Control (Q.C.) test results for calculating payments for HMA pavement projects (Hand et al., 2020). Note that starting from the year 2020, ITD has stopped the practice of considering contractor-reported test data for pay factor calculations. Nevertheless, the current research study was undertaken in 2018 and focused on ITD's QC/QA approach in effect through the end of 2019. A 2017 forensic investigation by ITD looked into 13 preselected pavement projects and found that out of 77 material testing reports, only 26% of the tests showed agreement between the ITD-generated results and the contractor-reported test values. This alarming mismatch not only can impact pavement projects' pay-factors, but also can have significant repercussions concerning the pavement service life and maintenance costs. Further inspection revealed that 40% of the investigated projects showed moderate distress two to five years after construction, whereas the design life of the pavements was 20 years.

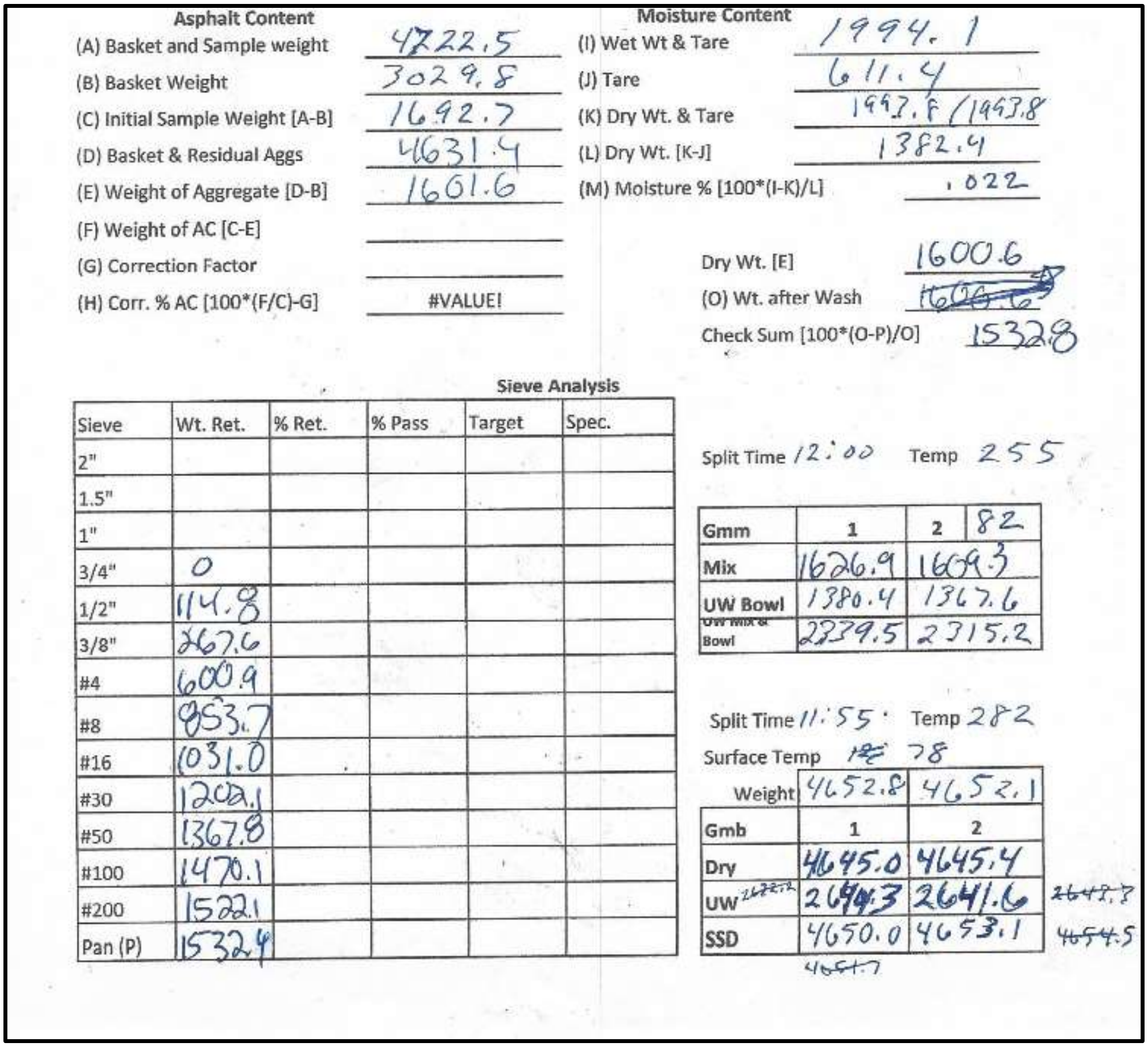
Objective and Scope

The objective of this research was to develop a framework to learn the patterns in the audit material test results for several HMA projects and classify the observed data alterations into Plausible Correction (P.C.) and Suspicious Alteration (S.A.). I use this terminology as detection of fraudulent activity requires a forensic analysis that cannot entirely be captured in a data mining approach. I first developed a logic-based algorithm based on an expert categorization of audit data into P.C./S.A. instances. Subsequently, I developed several supervised ML models to evaluate their capability to recognize the patterns in the labeled audit data.

Data Alteration

Both forensic analysis and anecdotal interviews with ITD and consulting engineers point to the possible existence of data alterations in HMA project reports (Dutton, 2020). This is concerning given substandard materials that might have been used for construction of some pavement projects that may result in lower than expected service life, higher maintenance costs, and in extreme cases even lower safety. ITD is investing $535 million (both federal and state funds) in Idaho highways in 2021 and a similar amount each year afterward by 2027; and suspicious activities and altered material testing values have the potential to cost taxpayers millions of dollars (ITD, 2019).

Figure 2-1 shows an image of a laboratory datasheet submitted to ITD during one of the HMA projects being looked into. As seen from the datasheet, the values in several fields were altered and over-written several times during the course of testing. This is particularly evident from the Under-Water (UW) and Saturated Surface Dry (SSD) weights. Some of this can be attributed to the possibility that scale readings were affected by the testing environment (such as excessive wind draft in the laboratory). However, repeated occurrence of such trends raises serious concerns about the quality of the test results. Moreover, such instances of alteration were also observed in cases where the test data were directly entered into the Excel-based data form (instances of data alteration in the Excel-based form were obtained through the embedded macro code). This emphasizes the importance of studying the extent of such data inconsistencies in the reported values, and developing approaches to prevent future occurrences of such poor testing practices.



Asphalt Content

(A) Basket and Sample weight	4722.5
(B) Basket Weight	3029.8
(C) Initial Sample Weight [A-B]	1692.7
(D) Basket & Residual Aggs	4631.4
(E) Weight of Aggregate [D-B]	1601.6
(F) Weight of AC [C-E]	
(G) Correction Factor	
(H) Corr. % AC [100*(F/C)-G]	#VALUE!

Moisture Content

(I) Wet Wt & Tare	1994.1
(J) Tare	611.4
(K) Dry Wt. & Tare	1993.8 / 1993.8
(L) Dry Wt. [K-J]	1382.4
(M) Moisture % [100*(I-K)/L]	.022

Dry Wt. [E] 1600.6
(O) Wt. after Wash ~~1600.6~~
Check Sum [100*(O-P)/O] 15328

Sieve Analysis

Sieve	Wt. Ret.	% Ret.	% Pass	Target	Spec.
2"					
1.5"					
1"					
3/4"	0				
1/2"	114.8				
3/8"	267.6				
#4	600.9				
#8	953.7				
#16	1031.0				
#30	1202.1				
#50	1367.8				
#100	1470.1				
#200	1522.1				
Pan (P)	1532.4				

Split Time 12:00 Temp 255

Gmm	1	2 82
Mix	1626.9	1609.3
UW Bowl	1380.4	1367.6
UW Mix & Bowl	2339.5	2315.2

Split Time 11:55 Temp 282
Surface Temp ~~12~~ 78

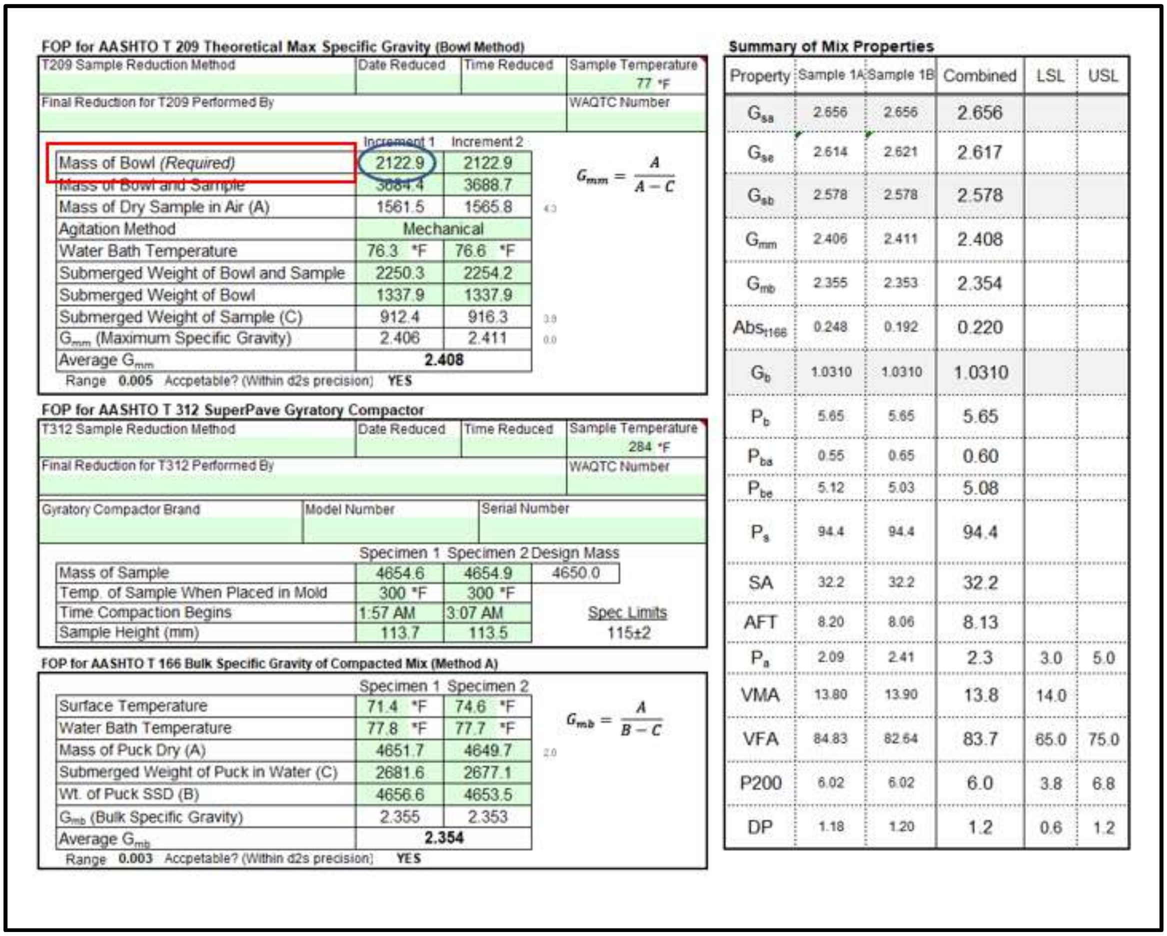
Gmb	1	2	
Weight	4652.8	4652.1	
Dry	4645.0	4645.4	
UW	2644.3	2641.6	~~2641.7~~
SSD	4650.0	4653.1	~~4654.5~~
	~~4654.7~~		

Figure 2-1 Data alteration on a paper data reporting sheet

Description of Available Audit Data

I acquired the material testing reported audit files from ITD for several HMA projects completed in Idaho during the year 2018. These Excel files comprised an internal audit algorithm (embedded by ITD) to record the sequences of changing parameter values in the background (not visible to the operator). Figure 2-2 presents a screenshot of a typical data input file to record material testing data. For example, if an operator inputs the value (2122.9 in this case) for Mass of bowl (red box) for increment 1 (blue circle), that value is recorded under U32 (corresponding cell number for mass of bowl (increment 1) in the

Excel file). If the operator deletes the value (2122.9) and registers a new value (2500 for example) both values are registered under U32 in the audit file with the corresponding time stamp.



FOP for AASHTO T 209 Theoretical Max Specific Gravity (Bowl Method)

T209 Sample Reduction Method	Date Reduced	Time Reduced	Sample Temperature
			77 °F
Final Reduction for T209 Performed By			WAQTC Number

	Increment 1	Increment 2
Mass of Bowl *(Required)*	2122.9	2122.9
~~Mass of Bowl and Sample~~	3684.4	3688.7
Mass of Dry Sample in Air (A)	1561.5	1565.8
Agitation Method	Mechanical	
Water Bath Temperature	76.3 °F	76.6 °F
Submerged Weight of Bowl and Sample	2250.3	2254.2
Submerged Weight of Bowl	1337.9	1337.9
Submerged Weight of Sample (C)	912.4	916.3
G_{mm} (Maximum Specific Gravity)	2.406	2.411
Average G_{mm}	**2.408**	

Range **0.005** Accpetable? (Within d2s precision) **YES**

$$G_{mm} = \frac{A}{A-C}$$

FOP for AASHTO T 312 SuperPave Gyratory Compactor

T312 Sample Reduction Method	Date Reduced	Time Reduced	Sample Temperature
			284 °F
Final Reduction for T312 Performed By			WAQTC Number
Gyratory Compactor Brand	Model Number	Serial Number	

	Specimen 1	Specimen 2	Design Mass
Mass of Sample	4654.6	4654.9	4650.0
Temp. of Sample When Placed in Mold	300 °F	300 °F	
Time Compaction Begins	1:57 AM	3:07 AM	Spec Limits
Sample Height (mm)	113.7	113.5	115±2

FOP for AASHTO T 166 Bulk Specific Gravity of Compacted Mix (Method A)

	Specimen 1	Specimen 2
Surface Temperature	71.4 °F	74.6 °F
Water Bath Temperature	77.8 °F	77.7 °F
Mass of Puck Dry (A)	4651.7	4649.7
Submerged Weight of Puck in Water (C)	2681.6	2677.1
Wt. of Puck SSD (B)	4656.6	4653.5
G_{mb} (Bulk Specific Gravity)	2.355	2.353
Average G_{mb}	**2.354**	

Range **0.003** Accpetable? (Within d2s precision) **YES**

$$G_{mb} = \frac{A}{B-C}$$

Summary of Mix Properties

Property	Sample 1A	Sample 1B	Combined	LSL	USL
G_{sa}	2.656	2.656	2.656		
G_{se}	2.614	2.621	2.617		
G_{sb}	2.578	2.578	2.578		
G_{mm}	2.406	2.411	2.408		
G_{mb}	2.355	2.353	2.354		
Abs_{t166}	0.248	0.192	0.220		
G_b	1.0310	1.0310	1.0310		
P_b	5.65	5.65	5.65		
P_{ba}	0.55	0.65	0.60		
P_{be}	5.12	5.03	5.08		
P_s	94.4	94.4	94.4		
SA	32.2	32.2	32.2		
AFT	8.20	8.06	8.13		
P_a	2.09	2.41	2.3	3.0	5.0
VMA	13.80	13.90	13.8	14.0	
VFA	84.83	82.64	83.7	65.0	75.0
P200	6.02	6.02	6.0	3.8	6.8
DP	1.18	1.20	1.2	0.6	1.2

Figure 2-2 Audit file to record material testing data

The dataset has several interesting characteristics:

i. Material test reporting Excel files had a VBA script embedded, which had a unique ability to record each data entry typed in the excel sheet. This develops a chronological record of all values entered into the spreadsheet in the form of an audit log. Inspection of this audit log can give a clear picture of how the test results were recorded. Figure 2-3 presents a screenshot of the audit log file for one of the projects. Note that all identifying information, such as project name, test date, test time, testing lab, among others, have been removed from the figures in this manuscript to ensure the anonymity of the testing/reporting entities.

ii. Audit data was available for both quality control as well as acceptance tests. In other words, records of data entries were available for certain projects irrespective of whether the tests were performed by the contractor (or a third-party testing laboratory hired by the contractor) or the state DOT (or a third-party testing laboratory hired by the state). As already mentioned, the primary objective of the current research was to study the data alteration patterns during HMA quality control and acceptance testing. The discussions in this manuscript do not focus on whether the data alterations were carried out by representatives of the contractor or the state DOT.

iii. All parameters that would affect the payments of each project were also provided, which are listed in Table 2-1. There is a total of 27 different parameters that affect the payment. They are categorized into three different categories (major/moderate/minor).

Sample	Cell	Value	Time
Test(25)	U37	2255.1	4:30:41 AM
Test(25)	U37	2256.1	4:30:58 AM
Test(25)	U37	2256.1	4:31:07 AM
Test(25)	U63	4554.6	5:31:38 AM
Test(25)	U63	4654.6	5:32:07 AM
Test(25)	Z37	2271.6	4:30:45 AM
Test(25)	Z37	2273.6	4:30:52 AM
Test(25)	Z131	1782.2	4:31:47 AM
Test(25)	Z131	1782.3	4:33:17 AM
Test(25)	S112	4603.1	4:31:25 AM
Test(25)	S112	4586.1	4:32:55 AM
Test(25)	S114	4514.5	4:31:31 AM
Test(25)	S114	4513.5	4:32:02 AM
Test(25)	S114	4515.5	4:32:13 AM
Test(25)	S114	4516.5	4:32:19 AM
Test(25)	S114	4516.1	4:32:23 AM
Test(25)	S114	4500.9	4:33:02 AM

Figure 2-3 Screenshot of the Audit Log File showing data alteration in excel file

Table 2-1 Material testing parameters and their impacts on pay-factor related parameters

Cell Description	Voids in the Mineral Aggregate (VMA)	Air Voids	Density	Major/ Moderate/ Minor Effect
Mass of Bowl (Increment 1) (U32)	Yes	Yes	Yes	Major
Mass of Bowl and Sample Dry (Increment 1) (U33)	Yes	Yes	Yes	Major
Submerged Weight of Bowl and Sample (Increment 1) (U37)	Yes	Yes	Yes	Major
Submerged Weight of Bowl (Increment 1) (U38)	Yes	Yes	Yes	Major
Mass of Bowl (Increment 2) (Z32)	Yes	Yes	Yes	Major
Mass of Bowl and Sample Dry (Increment 2) (Z33)	Yes	Yes	Yes	Major
Submerged Weight of Bowl and Sample (Increment 2) (Z37)	Yes	Yes	Yes	Major
Submerged Weight of Bowl (Increment 2) (Z38)	Yes	Yes	Yes	Major
Mass of Puck Dry (Specimen 1) (U61)	Yes	Yes	No	Major
Submerged Weight of Puck in Water (Specimen 1) (U62)	Yes	Yes	No	Major
Weight of Puck SSD (Specimen 1) (U63)	Yes	Yes	No	Major
Mass of Puck Dry (Specimen 2) (Z61)	Yes	Yes	No	Major
Submerged Weight of Puck in Water (Specimen 2) (Z62)	Yes	Yes	No	Major
Weight of Puck SSD (Specimen 2) (Z63)	Yes	Yes	No	Major
Mass Basket Assembly (S111)	Yes	No	No	Moderate
Mass Basket Assembly & Initial Sample (S112)	Yes	No	No	Moderate
Mass Basket Assembly & Final Aggregate (S114)	Yes	No	No	Moderate
Ignition Furnace Correction Factor (S116)	Yes	No	No	Moderate

Calibration Factor (AP114)	Yes	No	No	Moderate
Uncorrected Binder Content (AP115)	Yes	No	No	Moderate
Pan Mass (N128)	Yes	No	No	Minor
Mass Pan and Initial Sample (N129)	Yes	No	No	Minor
Drying Cycle 1 Mass Pan and Sample (Z129)	Yes	No	No	Minor
Drying Cycle 2 Mass Pan and Sample (Z130)	Yes	No	No	Minor
Drying Cycle 3 Mass Pan and Sample (Z131)	Yes	No	No	Minor
Drying Cycle 4 Mass Pan and Sample (Z132)	Yes	No	No	Minor
Drying Cycle 5 Mass Pan and Sample (Z133)	Yes	No	No	Minor

iv. Payment affecting parameters are similar for both department and contractor-reported data (Table 2-1). However, parameters that affect Density are only reported by the state DOT data. Those parameters are enlisted in Table 2-2. These parameters are monitored by ITD to decide on whether a particular asphalt mix meets specifications or not (VMA and Air Voids), and also whether a constructed pavement section has been adequately compacted or not (main line density). Reading 1 and 2 and Device Used are reported more than one time for each lot. So, if there are 2 tests in lot 1, then for reading 1, test 1 and 2 values would be registered in cell AC37 and AC38, respectively. Basically, there are only three parameters (Reading 1 and 2, Device used) in the density-related data.

Table 2-2 Material testing parameters (density) and their impacts on pay-factor related parameters

Cell Description	Voids in the Mineral Aggregate (VMA)	Air Voids	Density	Major/Minor Effect
Reading 1 (AC37-AC61)	No	No	Yes	Major
Reading 2 (AG37-AG61)	No	No	Yes	Major
Device Used (X37-X61)	No	No	Yes	Major

v. Total number of material testing parameters (department/contractor/density) is summarized in Table 2-3.

Table 2-3 Total number of material testing parameters

Parameter type	Total number (Department and Contractor)	Total number (Density)
Parameters with **major** impact	14	3-75
Parameters with **moderate** impact	6	0
Parameters with **minor** impact	7	0

Classification of Parameter Changes to Plausible Correction and Suspicious Alteration

The following section describes the approach adopted to categorize the data alterations into two groups: (1) Plausible Correction (P.C.) or (2) Suspicious Alteration (S.A.). The whole process was accomplished in several steps.

i. The first step was to separate the repeated data from the non-repeated incidents. Non-repeated data represent cases where no change in values was recorded for certain parameters in the input form.

ii. The second step involved manual inspection of all the repeated (altered) data to identify any existing patterns. Data alterations identified through this approach were categorized into P.C. and S.A.

Plausible Correction (P.C.): The incidents where values were likely not changed deliberately. The most likely cause of such changes was mistyping while entering the data from paper reports in the excel files.

Suspicious Alteration (S.A.): The incidents of altered values that I could not attribute to typographical and other cases of mistakes, after exhaustive consultation with advisors and engineers. Such alterations may have been done intentionally to reach the desired value, potentially change the payment, and/or modify a certain test outcome.

iii. Third step was to find general patterns in P.C. and S.A. cases.

iv. A total of 7 and 4 general patterns were found for the P.C. and S.A. categories, respectively.

v. Algorithmic logics were devised for each case, and computer codes were developed to automatically detect and categorize data value changes

<u>Development of Algorithms and Code:</u>

This was accomplished in several steps:

i. Initially, all cells with repeated values (more than one entry) associated with the pay effecting parameters in each project were identified.

Sample	Cell	Value	Time
Test(17)	U32	2123.2	11:53:46 PM
Test(17)	U32	2123.3	12:53:08 AM
Test(22)	U32	4655.3	9:44:05 PM
Test(22)	U32	2123	9:44:16 PM
Test(17)	U33	3658.3	12:50:59 AM
Test(17)	U33	3687.5	2:36:39 AM
Test(44)	U33	3653.2	12:20:25 AM
Test(44)	U33		12:54:03 AM
Test(44)	U33	3670.5	2:31:53 AM
Test(45)	U33	3670.5	1:59:31 AM
Test(45)	U33		2:31:46 AM
Test(45)	U33	3651.1	4:27:12 AM
Test(53)	U33	3645.1	1:35:29 AM
Test(53)	U33	3649	1:35:34 AM
Test(53)	U33	3654.3	1:35:41 AM
Test(8)	U33	3627.9	3:13:05 AM
Test(8)	U33	3672.9	3:13:20 AM
Test(9)	U33	3690	3:28:25 AM
Test(9)	U33	3696.6	3:28:40 AM
Test(9)	U33	3690.6	3:29:46 AM
Test(9)	U33	3699.6	3:29:54 AM

Figure 2-4 Repeated data entry (third column) of pay affecting parameters (second column; e.g. U32) for tests in a project (first column; e.g. Test(17)). Time of data entry is presented in column 4.

ii. Repeated cells are then separated per parameter name. In Figure 2-5, for example, parameter U32 (mass of bowl) is separated.

Sample	Cell	Value	Time
Test(17)	U32	2123.2	11:53:46 PM
Test(17)	U32	2123.3	12:53:08 AM
Test(22)	U32	4655.3	9:44:05 PM
Test(22)	U32	2123	9:44:16 PM

Figure 2-5 Separation of cells based on parameter name

iii. For each parameter, one set of samples (i.e. tests) is then considered at a time. Figure 2-5 had both sample Test(17) and Test(22), but in this step, we only consider one set of samples, i.e. Test(17) (Figure 2-6).

Sample	Cell	Value	Time
Test(17)	U32	2123.2	11:53:46 PM
Test(17)	U32	2123.3	12:53:08 AM

Figure 2-6 Separation of cells based on test/sample

iv. Cells are then run through a series of algorithms to determine cases of P.C. and S.A.

v. When there are multiple repetitions for a single parameter, each two consecutive entries (for instance, 1^{st} and 2^{nd} entry of a series of alterations) are considered a pair, and these pairs are run through the algorithms to be labelled P.C. or S.A. This is repeated for all pairs (e.g. 2^{nd} and 3^{rd}, 3^{rd} and 4^{th}, and so on). Once the serial comparison is completed, the first and last entries are considered a pair, and a similar analysis is done. I noted that there were some cases where the values were changed by a very small amount in every repetition, but this was done multiple times. In this case, each pair was labeled as P.C., but the comparison of first and last entries showed S.A. If the result is P.C. for all the pairs, the entire group is labeled as P.C. Upon detection of S.A., the entire group is labeled S.A. This procedure is visually represented in Figure 2-7.

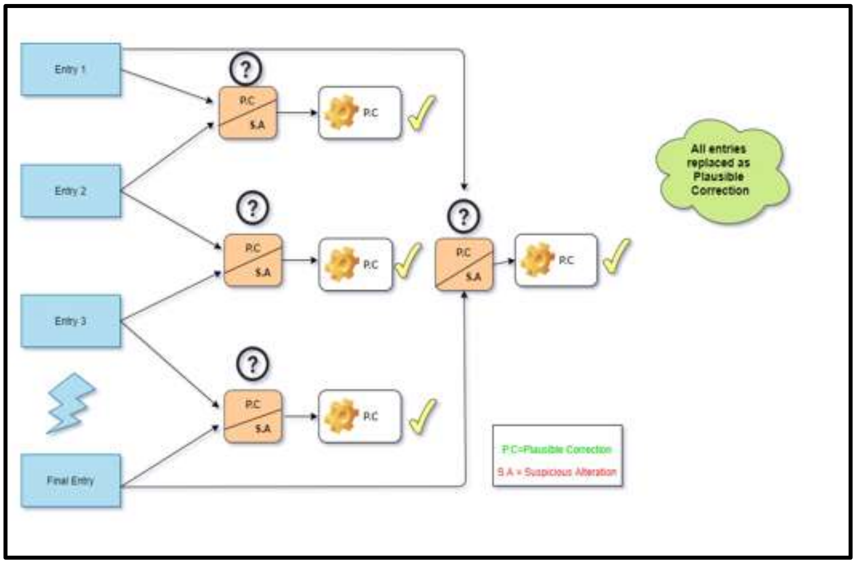


Figure 2-7 Methodology for Suspicious Alteration/Plausible Correction detection

<u>Scenarios to Categorize Data Alteration as Plausible Correction</u>

Case 1: One digit may be pressed instead of a neighboring key

While typing a digit, there is always a chance that another digit is mistakenly pressed instead of the desired number. For my analysis, I have considered a keypad like that of Figure 2-8, because in most of the desktop computers the keypad has this format. Here, I have considered all the possible cases that can happen when typing a number.

Figure 2-8 Plausible correction (case 1)

Usually, the neighboring keys surrounding a particular number key have the highest probability of being mistakenly pressed. As in Figure 2-8 (left), if we consider number 5, the closest buttons to 5 are 2, 4, 6, 8. I assume the probability of mistakenly pressing any of these digits instead of 5 is the same. Similarly, for the number 8 (refer to Figure 2-8, right), the closest keys are for numbers 5, 7, and 9. An algorithm was developed to label the repetition as P.C. if the number of repetitions is only one (there has been a change only from the 1st case to the 2nd case) and only one digit (at any position) is changed. This method is considered for all numbers from 0 to 9, and a series of neighboring possibilities are considered in each individual possible case. The algorithm first separates each digit of a number. In the next step, the algorithm does an element by element comparison and tries to identify if the changed digit fits in the closest neighboring category.

Sample	Cell	Value	Time
Test(26)	U37	2250.7	4:35:19 AM
Test(26)	U37	2251.7	4:35:49 AM

Figure 2-9 Plausible correction (case 1)—example

In Figure 2-9, for example, the number of changes/repetitions is only one and it is for one digit only (2250.7 versus 2251.7). My algorithm eliminates all the similar digits between the two entries except for the 4th digit. Then, a comparison is made for the unmatched digit, which is 0 versus 1 in this case. Since 1 fits in the adjacent neighboring rule of 0, this is considered a P.C.

Case 2: One or two digits were missed while typing

A very common scenario of plausible correction is 1 or 2 digits were missed while trying to type quickly or simply because the desired digit was not pressed properly. An individual might want to press 123, but instead, he/she presses 13 and misses 2. This is a clear case of an honest mistake or P.C. The logic that was used here is that if the 2nd entry is smaller than 80% or larger than 120% of the first input, then it is a P.C. I pose that S.A.s are generally around the vicinity of the actual value but are altered to return a better result. When the two values are too different, it is most probably a P.C. case.

An element-wise comparison is simply not possible in this case because the missing number can be any digit at any place. Generally, if a number is missed, the first entry becomes much smaller than the final or corrected entry. Hence a percentage difference can help determine this case. However, there is no fixed percentage threshold that I can specify to accurately determine the missed number case, but through the manual analysis of data, the appropriate threshold was found to be 20% above or below the final entry. In this case, the change would be considered as a P.C. only if the number of repetitions is only one.

Sample	Cell	Value	Time
Test(53)	U37	236.2	1:27:52 AM
Test(53)	U37	2236.2	1:27:54 AM

Figure 2-10 Plausible correction (case 2)—example

Figure 2-10 shows an example of a missed digit case of a P.C. The typist tried to insert 2236.2, but instead, he/she initially typed 236.2 missing the digit 2 and later corrected it.

Case 3: Order of digits were reversed while typing

A very often case of P.C. is typing digits in the wrong order, for example, 34 instead of 43.

Sample	Cell	Value	Time
Test(52)	N129	1234.6	4:47:43 AM
Test(52)	N129	1243.6	4:47:48 AM

Figure 2-11 Plausible correction (case 3)—example

Figure 2-11 depicts a case of order of digits being reversed while tying. The user wanted to type 1243.6, but instead, he/she typed 1234.6.

Case 4: Exact same value was typed twice

The initial inspection of the dataset showed that, in some cases, the exact same value was entered twice for a single parameter. This happened quite often. A logic was added in my algorithm to identify this type of P.C., as in Figure 2-12.

Sample	Cell	Value	Time
Test(51)	N129	1204.4	2:37:20 AM
Test(51)	N129	1204.4	2:54:07 AM

Figure 2-12 Plausible correction (case 4)—example

Case 5: Cell was empty at first and was filled in the second entry

Manual inspection revealed some cases where the cell was empty at first, but it was filled later. A possible reason might be that the VBA script records everything, even a single click, as an input while nothing was actually entered. The user then inputted the actual entry, for example as in Figure 2-13. This is a possible situation I considered P.C.

Sample	Cell	Value	Time
Test(5)	Z129		3:23:50 AM
Test(5)	Z129	1838.3	3:53:07 AM

Figure 2-13 Plausible correction (case 5)—example

Case 6: Digits that are hand-written similarly, if only repeated once, are considered a P.C.

Another case of P.C. is the numbers that look alike in handwriting can be entered instead of one another. Test results are usually logged in a paper sheet and are later digitized into the ITD provided Excel file. It is evident that handwriting would not be similar for all people, and there is a possibility of typing a digit instead of the actual digit due to their similarity in handwriting. For instance, 1 might look like 7 or 9 in the handwriting of various people (Figure 2-14). Another combination can be 6/8/0. In any of these combinations, it is essential that the number of repetitions must be only one. If the number of repetitions is more than one, it is more likely to be an S.A. case.

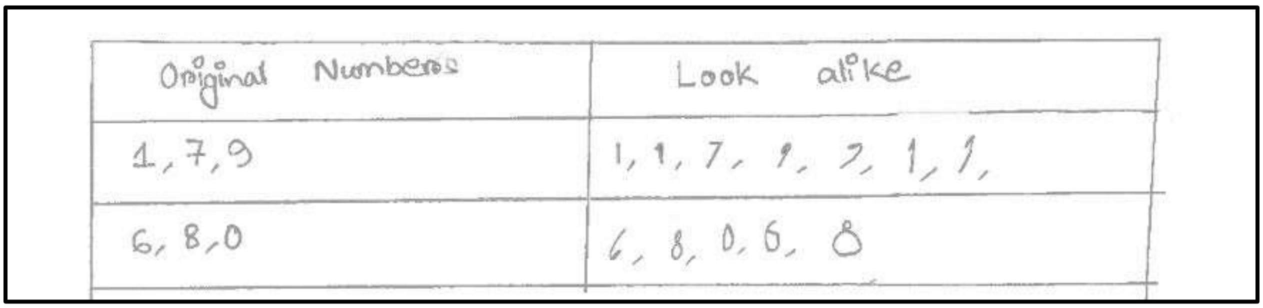


Figure 2-14 Look-wise case of Plausible Correction

Sample	Cell	Value	Time
Test(35)	Z37	2266	3:04:11 AM
Test(35)	Z37	2268	3:04:28 AM

Figure 2-15 Plausible correction (case 6)—example

Figure 2-15 shows a change of digit from 6 to 8, which is most probably a P.C. There is a point of argument here that this can fit in both cases, that the number was changed deliberately, or a simple look wise mistake has occurred. It is not possible to state with certainty that this is a P.C. or a S.A. case, since this is a subjective issue. I have concluded that if the number of changes is more than 1 (more than 1 repetition) the likelihood is higher toward S.A., whereas if the number of changes is only one, it aligns well with the P.C. case. Figure 2-16 shows a case where the changes could have been categorized as look wise error, but since the number of changes was more than one, this is no longer considered a P.C. case and it rather falls into a S.A. case.

Sample	Cell	Value	Time
Test(51)	U38	1337.9	4:47:39 AM
Test(51)	U38	1337.8	4:47:45 AM
Test(51)	U38	1337.6	4:47:55 AM

Figure 2-16 Plausible correction (case 6)—example

Case 7: Difference between two entries is too high

There have been some cases where the difference between two successive entries is too high. These incidents can also be differentiated through the percentage calculation. If the first entry is less than 80% or greater than 120% of the 2^{nd} entry, then the change is likely a P.C. There might be several reasons for this P.C. case, including reporting a parameter value for another parameter or reporting the parameter value from one test/sample to another test/sample.

Sample	Cell	Value	Time
Test(22)	Z32	4655.4	9:44:08 PM
Test(22)	Z32	2150.6	9:44:19 PM

Figure 2-17 Plausible correction (case 7)—example

Figure 2-17 is a clear example of a large difference between successive entries, which can be considered as a P.C. Here the 2^{nd} entry was less than 50 percent of the first case (4655.4 versus 2150.6), so this is most probably a P.C. case.

Scenarios to Categorize Data Alteration as Suspicious Alteration

Case 1: Changing values in a pattern or following a combination

S.A. cases mostly followed a pattern of change. In most cases, the number of changes is more than one, and the values are changing by a value of 1/2/10 in the positive or negative direction.

Sample	Cell	Value	Time
Test(34)	U37	2335.2	2:42:08 AM
Test(34)	U37	2338.2	2:43:04 AM
Test(34)	U37	2340.2	2:43:12 AM
Test(34)	U37	2339.2	2:43:39 AM
Test(34)	U37	2334.2	2:54:16 AM
Test(34)	U37	2336.2	2:54:23 AM
Test(34)	U37	2337.2	2:54:37 AM

Figure 2-18 Suspicious alteration (case 1)—example

Figure 2-18 presents a clear indication of a S.A. case. Here, the total number of changes is 6 times. The value was increased in the first two cases, reduced on the 3rd and 4th alterations, but then in the final two incidents, it increased again.

Case 2: Decimal values are eliminated.

In some S.A. cases, the digits after the decimal point are eliminated (e.g. Figure 2-19). In general, this might be a very small change, but even small changes in the sample of Hot Mix Asphalt (HMA) can have high impacts. Therefore, these cases are also considered as S.A. in my algorithm.

Sample	Cell	Value	Time
Test(38)	Z38	1355.4	6:07:51 AM
Test(38)	Z38	1355	6:14:13 AM

Figure 2-19 Suspicious alteration (case 2)—example

Case 3: Parameter values were changed but returned to the initial value

A clear case of altering data is presented in Figure 2-20, where the values were changed but later returned to the original value. Here, initially, the value was entered as 1945.4, which was changed to 1943, but later brought back to 1945.4. Although the value didn't change, I considered this as exploring values potentially for the wrong reasons and labeled it as S.A.

Sample	Cell	Value	Time
Test(19)	Z129	1945.4	1:17:53 AM
Test(19)	Z129	1943	10:57:30 AM
Test(19)	Z129	1945.4	10:57:39 AM

Figure 2-20 Suspicious alteration (case 3)—example

Case 4: Parameters were first assigned a value but finally changed to zero or removed entirely

There have been times, especially for parameters with small values, that the values were completely deleted or replaced with a value of zero. For example, in Figure 2-21, for sample Test(1) the value was set to 0.26 and replaced with zero. I considered this change as S.A.

Sample	Cell	Value	Time
Test(1)	AP114	0.26	11:31:55 PM
Test(1)	AP114	0	1:16:11 AM
Test(3)	AP114	1	1:25:46 AM
Test(3)	AP114	0.26	1:25:51 AM
Test(3)	AP114	0	1:25:59 AM

Figure 2-21 Suspicious alteration (case 4)—example

Scenarios to Uncertain Cases: Plausible Correction or Suspicious Alteration

A very interesting finding in my analysis indicated that there were incidents where the repetitions might fall in either S.A. or P.C. cases, an example of which is shown in Figure 2-22 (values changing from 4531.5 to 4532.5 and then to 4530.5). The first change was from 1 to 2, which might be considered P.C. In the second change, the digit 2 was replaced with 0, which is likely to be a S.A. However, there is enough room for argument to fit these cases in other categories. But the number of changes can be informative here. It is unlikely that both cases were a typo, hence this case is considered as S.A.

Sample	Cell	Value	Time
Test(8)	S114	4531.5	3:12:26 AM
Test(8)	S114	4532.5	3:12:34 AM
Test(8)	S114	4530.5	3:12:38 AM

Figure 2-22 Plausible Correction /Suspicious Alteration (case 1)—example

Impact of time stamp

Although S.A. cases generally occur in a relatively short period of time, I could not determine a definite relationship between P.C./S.A. cases with time that can be explored in a computer code (Figures 2-23 ,2-24, 2-25 and 2-26). Both categories have examples where a change occurred instantly or after some time.

Sample	Cell	Value	Time
Test(17)	U32	2123.2	11:53:46 PM
Test(17)	U32	2123.3	12:53:08 AM

Figure 2-23 Plausible Correction relationship with time—example 1

Sample	Cell	Value	Time
Test(16)	U37	2238.2	2:35:58 AM
Test(16)	U37	2239.2	2:36:09 AM

Figure 2-24 Plausible Correction relationship with time—example 2

Sample	Cell	Value	Time
Test(44)	U33	3653.2	12:20:25 AM
Test(44)	U33		12:54:03 AM
Test(44)	U33	3670.5	2:31:53 AM

Figure 2-25 Suspicious alteration relationship with time—example 1

Sample	Cell	Value	Time
Test(9)	U37		3:28:36 AM
Test(9)	U37	2269.8	3:28:51 AM
Test(9)	U37	2279.8	3:30:05 AM
Test(9)	U37	2269.8	3:30:13 AM
Test(9)	U37	2259.8	3:31:11 AM
Test(9)	U37	2262.8	3:31:20 AM

Figure 2-26 Suspicious alteration relationship with time—example 2

Results of P.C./S.A. Classification Algorithm

I applied the algorithms explained earlier to all audit data from the available project's datasets (separately for entities 1 and 2) to determine P.C. and S.A. cases. For each project, I determined the number of unique parameters that were altered, and the total number of times those parameters were altered. I also separated parameters with major/moderate/minor impacts on pay factor to analyze whether or not one category might be more susceptible to alteration than others.

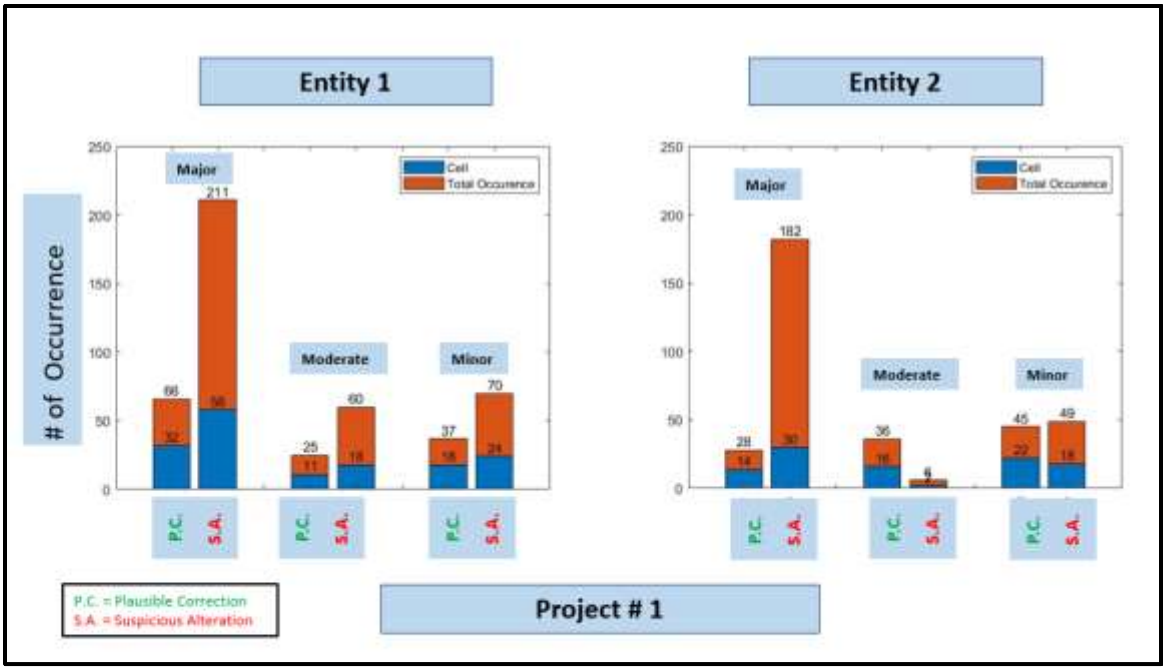


Figure 2-27 Number of occurrences of P.C./S.A. for project #1

Figure 2-27 shows the total number of altered parameters and frequency of alterations for project #1, as an example. Figure 2-27 shows the entity 1-reported statistics on the left side and the entity 2-reported statistics on the right side. In this project and for major parameters in entity 1-reported data, there were a total of 32 unique parameters that fell within the P.C. cases, and these parameters were changed a total of 66 times (an average of roughly one change per parameter). I observed a greater number of S.A. cases for the entity 1-reported major parameters, with a total of 58 parameters being changed 211 times (an average of roughly 2.5 changes per parameter). The higher average number of changes for major parameters in the case of S.A. compared to P.C. (2.5 versus 1) implies that there are some suspicious activities potentially to tune the parameter values to obtain certain outcomes. For moderate parameters in the P.C. category, 11 unique parameters were changed 25 times (an average of roughly 1 change per parameter), and in the S.A. category, 18 unique parameters were changed 60 times (an average of roughly 2 changes per parameter). Finally, 18 unique minor parameters were changed 37 times for the P.C. category, and 24 parameters were changed 70 times for the S.A. category. I observed an interesting scenario in this analysis as the number of changes for S.A. is roughly 2 times per unique parameter, whereas P.C. cases show roughly 1 change per unique parameter. While this can be partly an artifact of the devised algorithms, my careful manual investigation of P.C./S.A. categorized audit data confirm that algorithms are performing accurately. I attribute this observation to the P.C. cases being unintentional, and if an error/mistake occurred, it is usually corrected in the second entry. This is, however, quite different in the S.A. cases due to the potentially intentional nature of the alterations as the operator seeks a certain outcome and tries to fine tune the reported value to reach the intended result. The parameters are indeed altered multiple (≥ 2) times, which resulted in a high number of changes for major/moderate/minor S.A. cases.

A similar trend is observed in the entity 2-reported data for this project. A total of 14 major parameters in the P.C. category was changed 28 times, and 30 major parameters in the S.A. category were changed 182 times. In the case of moderate parameters in the P.C. category, 16 parameters were changed 36 times, whereas in the S.A. category 2 parameters were altered 6 times. For minor parameters in the P.C. category, 22 parameters were altered 45 times, and in the S.A. category 18 parameters were altered 49 times. Surprisingly, data alteration seems to be less pronounced in the entity 2 data compared to the entity 1-reported data. My further investigation showed that entity 1-reported values can be altered to either confirm the entity 2-reported data or to ensure entity 2 data is used for payment calculation, among other reasons.

I conducted this analysis on all available projects and reported their results in Table 2-4. Figures 2-28, 2-29 and 2-30 visually depict three example project results (projects #4, #7, #9).

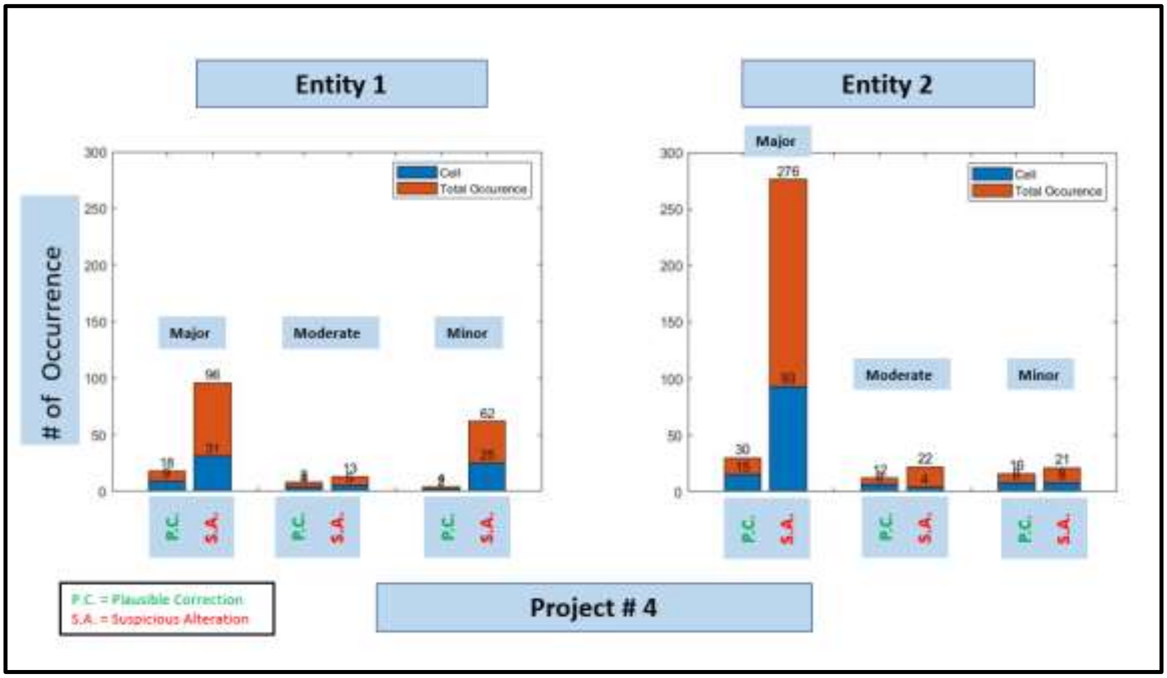


Figure 2-28 Number of occurrences of P.C./S.A. for project #4

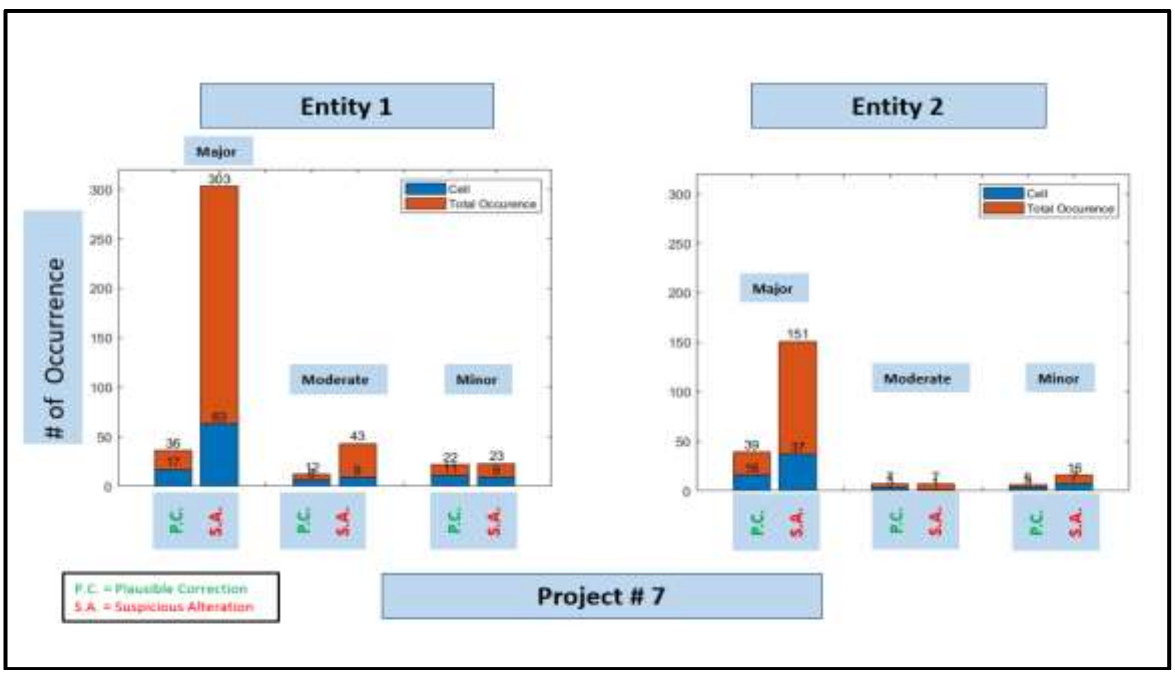


Figure 2-29 Number of occurrences of P.C./S.A. for project #7

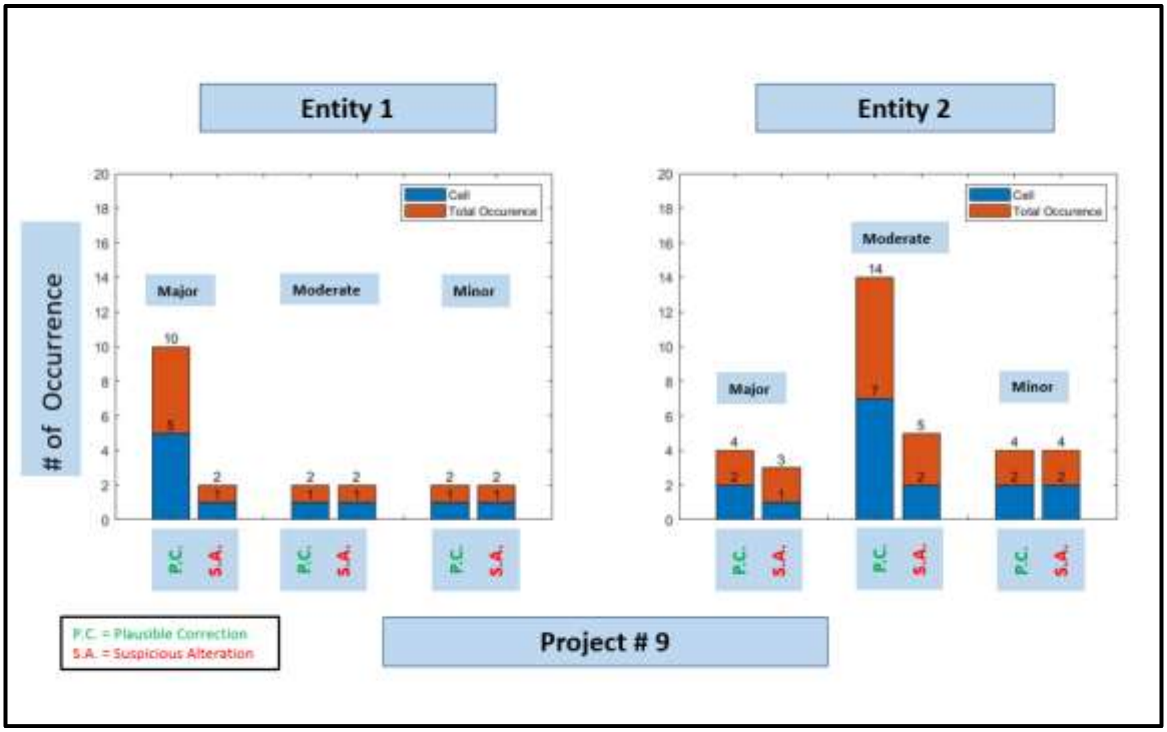


Figure 2-30 Number of occurrences of P.C./S.A. for project #9

Table 2-4 Unique and total number of material testing parameter changes

Project Number	Classification Category	Entity 1						Entity 2					
		Parameters with **major** impact		Parameters with **moderate** impact		Parameters with **minor** impact		Parameters with **major** impact		Parameters with **moderate** impact		Parameters with **minor** impact	
		Unique changes	All changes	Unique changes	All changes	Unique changes	All changes	Unique changes	All changes	Unique changes	All changes	Unique changes	All changes
Project 1	S.A.	58	211	18	60	24	70	30	182	2	6	18	49
	P.C.	32	66	11	25	18	37	14	28	16	36	22	45
Project 2	S.A.	94	404	18	53	26	81	26	66	0	0	0	0
	P.C.	29	64	14	30	17	41	11	22	0	0	0	0
Project 3	S.A.	2	6	2	8	0	0	0	0	0	0	0	0
	P.C.	10	20	3	6	0	0	0	0	0	0	0	0
Project 4	S.A.	31	96	5	13	25	62	93	276	4	22	8	21
	P.C.	9	18	4	8	2	4	15	30	6	12	8	16
Project 5	S.A.	19	52	2	5	2	6	39	87	6	12	9	22
	P.C.	10	20	1	2	5	10	48	98	18	37	30	63
Project 6	S.A.	25	73	3	9	0	0	0	0	0	0	0	0
	P.C.	11	23	1	2	1	2	0	0	0	0	0	0
Project 7	S.A.	63	303	9	43	9	23	37	151	1	7	7	16
	P.C.	17	36	6	12	11	22	16	39	3	7	3	6
Project 8	S.A.	33	138	1	7	6	19	19	77	2	5	5	17
	P.C.	13	26	7	14	7	14	13	26	3	6	6	15
Project 9	S.A.	1	2	1	2	1	2	1	3	2	5	2	4
	P.C.	5	10	1	2	1	2	2	4	7	14	2	4
Project 10	S.A.	7	28	2	17	0	0	0	0	0	0	0	0
	P.C.	0	0	0	0	0	0	0	0	0	0	0	0
Project 11	S.A.	8	17	1	3	2	4	17	60	10	37	5	19

	P.C.	8	16	11	22	4	9	13	28	5	10	4	9
Project 12	S.A.	7	17	2	6	1	3	26	56	5	10	3	7
	P.C.	4	8	3	7	0	0	3	6	2	4	5	10
Project 13	S.A.	0	0	1	2	1	2	0	0	0	0	0	0
	P.C.	1	2	0	0	0	0	0	0	0	0	0	0
Project 14	S.A.	66	334	7	27	5	14	0	0	0	0	0	0
	P.C.	20	41	11	22	5	11	0	0	0	0	0	0
Project 15	S.A.	3	21	3	23	1	2	10	49	0	0	0	0
	P.C.	0	0	2	4	1	2	3	8	2	4	0	0

Application of Machine Learning Algorithms for P.C./S.A. Classification

Several supervised machine learning (ML) algorithms have been used for classification of Plausible Correction (P.C.) and Suspicious Alteration (S.A.) cases. The main purpose of this exploratory analysis is to determine whether or not the human-detected patterns in the audit data are verified by the machine. Upon successful implementation, this adds a level of confidence to my analysis. Statistical techniques and ML algorithms are widely used for fraud detection in various sectors (Bell and Carcello, 2000; Lin et al., 2003; Caudill et al., 2005; Kotsiantis et al., 2006; Kirkos et al., 2007; Perols, 2011; Ngai et al., 2011). A supervised machine learning algorithm learns a function through labeled input data and produces output for new unlabeled data.

In the absence of independent training data for P.C./S.A. classification, I used the classified data from the previous section to evaluate various ML algorithms. A non-exhaustive list of well renowned and frequently used classification algorithms includes K-Nearest neighbor, Logistic Regression, Decision Tree, Neural Network, Support Vector Machine and Discriminant analysis. I successfully applied these algorithms to my datasets for P.C./S.A. classification purpose. These algorithms provide valuable insights to my analysis by assessing their suitability for the detection of suspicious activities in material testing reports. Here, I briefly introduce the employed ML algorithms, and refer the interested reader to "the elements of statistical learning" (Hastie et al., 2008) for detailed information.

Description of the ML algorithms

K-Nearest Neighbor: K-Nearest Neighbor (KNN) algorithm performs on the assumption that similar things occur in close proximity and in groups. A KNN algorithm generally stores the available scenarios and classifies them based on the similarity measure. This algorithm is widely used in real-life cases, such as recommender systems for recommending products on Amazon, movies on Netflix, or videos on Youtube because of its non-parametric nature that relaxes the need for assumption about the distribution of data.

Logistic Regression: Logistic regression is ideal for categorical variables. It is widely used in categorization of spam versus non-spam email, and fraud versus non-fraud credit card activity, among others. Logistic regression is a classification algorithm used to assign observations to a discrete set of classes, for example binary cases. This algorithm is named after the core method of the function, which is a logistic sigmoid function. This function is basically an S-shaped curve that can project any real number between the range of 0 to 1 but cannot reach the limits.

Decision Tree/Random Forest: Decision tree is a proper machine learning model for both classification and regression. The model performs "*If this than that*" with a certain condition for the final result. A decision tree model iterates through the dataset for partitioning data into categories. Random forest is basically a combination of several decision trees. This binary splitting method is very efficient as it can narrow down the probable options very quickly from a large number of classes.

Neural Network: Neural networks are multi-layer networks of neurons designed to recognize patterns. The neural network algorithm is modeled loosely after the human brain. This algorithm mimics the operation procedure of a human brain to identify the relationships in a set of available data. One excellent aspect of the neural network is its ability to adapt to changing inputs. Neurons in a neural network represent a mathematical function, which is responsible for collecting and classifying information based on the requirement of the user. The neural network consists of multiple layers of interconnected nodes.

Support Vector Machine: Support vector machine (SVM) algorithms typically find a hyperplane that can efficiently distinguish between data points. This hyperplane is termed as the decision boundary, and anything falling on one side of this line is considered one group. SVM models can solve both classification and regression problems. They use a technique called “kernel trick” for transforming data, from which the hyperplane is detected.

Discriminant Analysis: Discriminant analysis is a supervised machine learning technique used for dimensionality reduction. Ideally, this algorithm is used to classify between two or three classes and separate project features from a higher dimension to a lower order dimension. The generic concept of a discriminant analysis model is very similar to a principal component analysis, but through the discriminant analysis axes that maximizes the separation between multiple classes is found.

Data Preprocessing for ML Algorithms

For the ML algorithms to perform successfully, a large data set is generally required. The bigger the training dataset is, the superior the model learning, and thereby, the higher the model performance would be. None of the project datasets was big enough and typically had 300/400 rows of instances. I, therefore, enlarged the dataset by combining all the entity 1-reported project data. Similarly, the entity 2 dataset was also combined for all the projects. Thereby I created two datasets that were sufficient to train the ML models.

Data were preprocessed before being fed into the ML algorithms. Figure 2-31 represents the original format of the dataset after removing the non-repeated cases. For ML purposes, all repeated parameter values are required to be presented row-wise. For this purpose, all the repeated values of a certain parameter in a certain test are presented in adjacent columns, as shown in Figure 2-32.

Sample	Cell	Value	Time
Test(17)	U32	2123.2	11:53:46 PM
Test(17)	U32	2123.3	12:53:08 AM
Test(22)	U32	4655.3	9:44:05 PM
Test(22)	U32	2123	9:44:16 PM
Test(8)	U33	3627.9	3:13:05 AM
Test(8)	U33	3672.9	3:13:20 AM

Figure 2-31 Repeated data points to be used for training ML algorithms (original format)

Sample	Cell	Value1	Value2	Value3	Time	Var_time_gap	Var_effect_type	Var_error
Test(17)	U32	2123.2	2123.3		12:53:08 AM	3562	Major	0
Test(22)	U32	4655.3	2123		9:44:16 PM	11	Major	0
Test(44)	U33	3653.2	NA	3670.5	2:31:53 AM	5870	Major	1

Figure 2-32 Row-wise rearranged data (first step preprocessing for ML)

We also need to create a consistent dataset, meaning each row (one parameter for one test) should have the same number of columns. However, a parameter value might be repeated once and another might be repeated five times. To create a consistent matrix format, we need to consider the maximum number of repetitions for all parameters. In case a parameter has fewer repetitions that the maximum number of repetitions, the last value was copied in the remaining columns (Figure 2-33).

Sample	Cell	value1	value2	value3	value4	value5	value6	value7	value8	value9	Number_of_change	Date_time	Time_gap	Effect	S.A.	Timestamp
Test(17)	U 32	2123.2	2123.3	2123.3	2123.3	2123.3	2123.3	2123.3	2123.3	2123.3	1	12:53:00 AM	3562	Major	0	1.394424
Test(22)	U 32	4655.3	2123.0	2123.0	2123.0	2123.0	2123.0	2123.0	2123.0	2123.0	1	9:44:00 PM	11	Major	0	-1.749008
Test(17)	U 33	3658.3	3687.5	3687.5	3687.5	3687.5	3687.5	3687.5	3687.5	3687.5	1	2:36:00 AM	6340	Major	1	0.936485
Test(44)	U 33	3653.2	NaN	3670.5	3670.5	3670.5	3670.5	3670.5	3670.5	3670.5	2	2:31:00 AM	5870	Major	1	1.584196
Test(45)	U 33	3670.5	NaN	3651.1	3651.1	3651.1	3651.1	3651.1	3651.1	3651.1	2	4:27:00 AM	6926	Major	1	-0.557791
Test(53)	U 33	3645.1	3649.0	3654.3	3654.3	3654.3	3654.3	3654.3	3654.3	3654.3	2	1:35:00 AM	7	Major	1	0.248621
Test(8)	U 33	3627.9	3672.9	3672.9	3672.9	3672.9	3672.9	3672.9	3672.9	3672.9	1	3:13:00 AM	15	Major	0	0.438393

Figure 2-33 Matrix formatted data

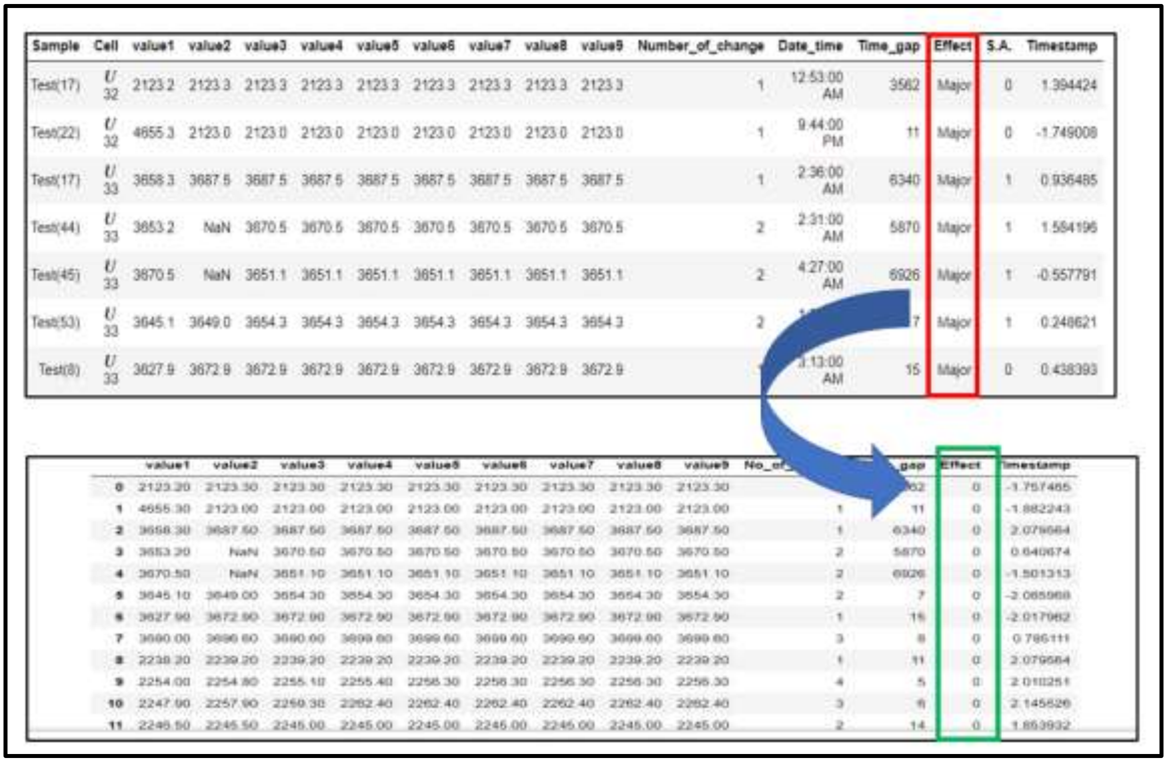
For example, in Figure 2-33, I noticed that the maximum number of repetitions for a certain parameter in a test project is 8 (simply an example project), hence, we need a total of nine values (columns) for each row (parameter). When I put all projects together for the ML application, the number of columns increases to 23 for entity 2-reported data and 29 for entity 1-reported data. Alongside the maximum number of changes, I also counted the actual number of changes for each cell. This was also provided in an extra column. The very first row in Figure 2-33, for example, shows mass of bowl (U32) for Test(17) has only one repetition, so there are only two values for this parameter. We need seven more values to fill up the matrix. Hence, the last value was copied to the remaining seven columns (titled "value") for mass of bowl (U32). This was applied to all parameters.

Although my human-based effort to include time and date of data entry failed to detect a conclusive pattern, I examined whether or not such pattern in evident to the machine. To this end, date that was originally in the mm/dd/yyyy format and time, were converted to timestamp format in order to obtain a unique value. Now we can use the date and time information as a feature in ML application.

Sample	Cell	value1	value2	value3	value4	value5	value6	value7	value8	value9	Number_of_change	Date_time	Time_gap	Effect	S.A.	Timestamp
Test(17)	U_{32}	2123.2	2123.3	2123.3	2123.3	2123.3	2123.3	2123.3	2123.3	2123.3	1	12:53:00 AM	3562	Major	0	1.394424
Test(22)	U_{32}	4655.3	2123.0	2123.0	2123.0	2123.0	2123.0	2123.0	2123.0	2123.0	1	9:44:00 PM	11	Major	0	-1.749008
Test(17)	U_{33}	3658.3	3687.5	3687.5	3687.5	3687.5	3687.5	3687.5	3687.5	3687.5	1	2:36:00 AM	6340	Major	1	0.936485
Test(44)	U_{33}	3653.2	NaN	3670.5	3670.5	3670.5	3670.5	3670.5	3670.5	3670.5	2	2:31:00 AM	5870	Major	1	1.584196
Test(45)	U_{33}	3670.5	NaN	3651.1	3651.1	3651.1	3651.1	3651.1	3651.1	3651.1	2	4:27:00 AM	6926	Major	1	-0.557791
Test(53)	U_{33}	3645.1	3649.0	3654.3	3654.3	3654.3	3654.3	3654.3	3654.3	3654.3	2	1:35:00 AM	7	Major	1	0.248621
Test(8)	U_{33}	3627.9	3672.9	3672.9	3672.9	3672.9	3672.9	3672.9	3672.9	3672.9	1	3:13:00 AM	15	Major	0	0.438393

Figure 2-34 Conversion of Date &Time to the timestamp format

Similarly, the effect of parameter on the pay factor (major/minor/moderate) was an important feature to help the ML algorithms classify repetitions to P.C./S.A. The categorical data type was required to be converted to numerical format. Label encoding, which is a powerful tool that can convert the categorical/text data into numerical data, was used for this purpose.

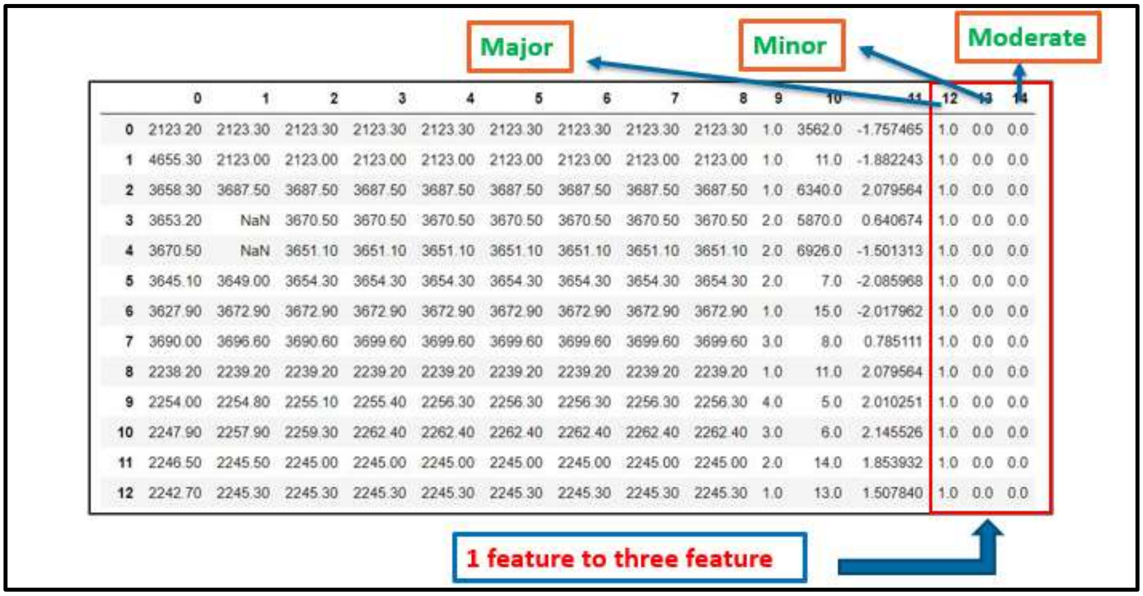


Sample	Cell	value1	value2	value3	value4	value5	value6	value7	value8	value9	Number_of_change	Date_time	Time_gap	Effect	S.A.	Timestamp
Test(17)	U_{32}	2123.2	2123.3	2123.3	2123.3	2123.3	2123.3	2123.3	2123.3	2123.3	1	12:53:00 AM	3562	Major	0	1.394424
Test(22)	U_{32}	4655.3	2123.0	2123.0	2123.0	2123.0	2123.0	2123.0	2123.0	2123.0	1	9:44:00 PM	11	Major	0	-1.749008
Test(17)	U_{33}	3658.3	3687.5	3687.5	3687.5	3687.5	3687.5	3687.5	3687.5	3687.5	1	2:36:00 AM	6340	Major	1	0.936485
Test(44)	U_{33}	3653.2	NaN	3670.5	3670.5	3670.5	3670.5	3670.5	3670.5	3670.5	2	2:31:00 AM	5870	Major	1	1.584196
Test(45)	U_{33}	3870.5	NaN	3651.1	3651.1	3651.1	3651.1	3651.1	3651.1	3651.1	2	4:27:00 AM	6926	Major	1	-0.557791
Test(53)	U_{33}	3645.1	3649.0	3654.3	3654.3	3654.3	3654.3	3654.3	3654.3	3654.3	2	[illegible]	7	Major	1	0.248621
Test(8)	U_{33}	3627.9	3672.9	3672.9	3672.9	3672.9	3672.9	3672.9	3672.9	3672.9	[illegible]	3:13:00 AM	15	Major	0	0.438393

	value1	value2	value3	value4	value5	value6	value7	value8	value9	No_of_[illegible]	[illegible]_gap	Effect	Timestamp
0	2123.20	2123.30	2123.30	2123.30	2123.30	2123.30	2123.30	2123.30	2123.30	[illegible]	[illegible]62	0	-1.757465
1	4655.30	2123.00	2123.00	2123.00	2123.00	2123.00	2123.00	2123.00	2123.00	1	11	0	-1.882243
2	3658.30	3687.50	3687.50	3687.50	3687.50	3687.50	3687.50	3687.50	3687.50	1	6340	0	2.079664
3	3653.20	NaN	3670.50	3670.50	3670.50	3670.50	3670.50	3670.50	3670.50	2	5870	0	0.640674
4	3670.50	NaN	3651.10	3651.10	3651.10	3651.10	3651.10	3651.10	3651.10	2	6926	0	-1.501313
5	3645.10	3649.00	3654.30	3654.30	3654.30	3654.30	3654.30	3654.30	3654.30	2	7	0	-2.065968
6	3627.90	3672.90	3672.90	3672.90	3672.90	3672.90	3672.90	3672.90	3672.90	1	15	0	-2.017962
7	3690.00	3696.60	3690.00	3699.60	3699.60	3699.60	3699.60	3699.00	3699.60	3	8	0	0.786111
8	2238.20	2239.20	2239.20	2239.20	2239.20	2239.20	2239.20	2239.20	2239.20	1	11	0	2.079664
9	2254.00	2254.80	2255.10	2255.40	2256.30	2256.30	2256.30	2256.30	2256.30	4	5	0	2.010251
10	2247.90	2257.90	2259.30	2262.40	2262.40	2262.40	2262.40	2262.40	2262.40	3	6	0	2.145626
11	2246.50	2245.50	2245.00	2245.00	2245.00	2245.00	2245.00	2245.00	2245.00	2	14	0	1.853932

Figure 2-35 Label encoded parameter effects

This approach assigned values of 0, 1, 2 to major/moderate/minor categories, respectively. Although the label encoder successfully converted the categorical data to numerical data, the ML algorithm would assume the categorical data with higher integer value is greater/more important than others. So, another method, “one hot encoding” was used to solve this issue by turning the categorical numbers to binary vectors. One hot encoder creates a vector with three binary digits. Value of 1 in the first column represents major parameters, whereas value of 1 in the second and third columns represent minor and moderate parameters, respectively, as shown in Figure 2-36.



	0	1	2	3	4	5	6	7	8	9	10	11	12	13	14
0	2123.20	2123.30	2123.30	2123.30	2123.30	2123.30	2123.30	2123.30	2123.30	1.0	3562.0	-1.757465	1.0	0.0	0.0
1	4655.30	2123.00	2123.00	2123.00	2123.00	2123.00	2123.00	2123.00	2123.00	1.0	11.0	-1.882243	1.0	0.0	0.0
2	3658.30	3687.50	3687.50	3687.50	3687.50	3687.50	3687.50	3687.50	3687.50	1.0	6340.0	2.079564	1.0	0.0	0.0
3	3653.20	NaN	3670.50	3670.50	3670.50	3670.50	3670.50	3670.50	3670.50	2.0	5870.0	0.640674	1.0	0.0	0.0
4	3670.50	NaN	3651.10	3651.10	3651.10	3651.10	3651.10	3651.10	3651.10	2.0	6926.0	-1.501313	1.0	0.0	0.0
5	3645.10	3649.00	3654.30	3654.30	3654.30	3654.30	3654.30	3654.30	3654.30	2.0	7.0	-2.085968	1.0	0.0	0.0
6	3627.90	3672.90	3672.90	3672.90	3672.90	3672.90	3672.90	3672.90	3672.90	1.0	15.0	-2.017962	1.0	0.0	0.0
7	3690.00	3696.60	3690.60	3699.60	3699.60	3699.60	3699.60	3699.60	3699.60	3.0	8.0	0.785111	1.0	0.0	0.0
8	2238.20	2239.20	2239.20	2239.20	2239.20	2239.20	2239.20	2239.20	2239.20	1.0	11.0	2.079564	1.0	0.0	0.0
9	2254.00	2254.80	2255.10	2255.40	2256.30	2256.30	2256.30	2256.30	2256.30	4.0	5.0	2.010251	1.0	0.0	0.0
10	2247.90	2257.90	2259.30	2262.40	2262.40	2262.40	2262.40	2262.40	2262.40	3.0	6.0	2.145526	1.0	0.0	0.0
11	2246.50	2245.50	2245.00	2245.00	2245.00	2245.00	2245.00	2245.00	2245.00	2.0	14.0	1.853932	1.0	0.0	0.0
12	2242.70	2245.30	2245.30	2245.30	2245.30	2245.30	2245.30	2245.30	2245.30	1.0	13.0	1.507840	1.0	0.0	0.0

Figure 2-36 One hot encoded vectors for effect types

Finally, since parameters are on different scales, we need to normalize the data to ensure certain parameter values do not spuriously impact the outcome. So, the standard scalar function of python was used to normalize all the data. The standard scalar function assumes the data to be normally distributed and scale them such that the data is now centered around 0 and with a standard deviation of 1. The final dataset is similar to that of Figure 2-37. I then divided the total dataset into two parts. One for the training purpose and the other part for the evaluation. I used 2/3 of the data for training purposes and 1/3 for testing purposes.

```
array([[-0.44517741, -0.5185548 , -0.4943244 , ...,  0.88819417,
        -0.59408853, -0.46871843],
       [ 1.33323244, -0.51878526, -0.49454862, ...,  0.88819417,
        -0.59408853, -0.46871843],
       [ 0.63299366,  0.68305791,  0.67475555, ...,  0.88819417,
        -0.59408853, -0.46871843],
       ...,
       [ 1.24628203,  1.33218382,  1.30481206, ..., -1.12587994,
        -0.59408853,  2.13347701],
       [-1.93621544, -2.14966609, -2.08108025, ..., -1.12587994,
        -0.59408853,  2.13347701],
       [-1.93621544, -2.14946636, -2.08108025, ..., -1.12587994,
        -0.59408853,  2.13347701]])
```

Figure 2-37 Normalization of the data for training ML algorithms

Evaluation of ML algorithms

Once the training and test datasets are prepared, I trained various ML models and evaluated their performance using the accuracy score function of "scikit learn" toolbox in python. I have used the sigmoid activation function and adam optimizer to train the neural network model. For the loss function, I used the "binary crossentropy" which is compatible with sigmoid. This type of loss function is ideal for binary classification tasks. For the logistic regression, I have used the L2 or ridge regression as a penalty. Ridge regression adds "squared magnitude" of coefficient and penalty term to the loss function. This technique helps to avoid overfitting. For the k-Nearest Neighbor algorithm, I used a neighbor value of 5, which yielded the best result in my prediction. These functions compare predicted P.C./S.A. with the training and test data. The performance of selected models for entity 1 and entity 2 data are listed in Tables 2-5 and 2-6, respectively. I also conducted this analysis on each project separately, which showed that every single project does not provide enough information to train the ML models (not shown here). It is noteworthy that the entity 2 dataset had a total of 737 sample data points (rows or in other words unique parameters). Parameters were changed up to a maximum of 22 times in the entity 2-reported data, meaning the 23rd value was the final reported value. Similarly, on the entity 1 side, there were a total of 892 sample points (unique parameters) changing up to a maximum of 28 times.

Table 2-5 Performance of supervised ML algorithms on combined entity 1 datasets

Supervised ML model	Accuracy Score
K-Nearest Neighbor	69%
Logistic Regression	69%
Decision Tree/Random Forest	66%
SVM (Linear)	73%
Discriminant Analysis	72%
Neural Network	39%

Table 2-6 Performance of supervised ML algorithms on combined entity 2 datasets

Supervised ML model	Accuracy Score
K-Nearest Neighbor	69%
Logistic Regression	69%
Decision Tree/Random Forest	66%
SVM (Linear)	72%
Discriminant Analysis	72%
Neural Network	39%

All models, except for Neural Network, generally perform at an acceptable level, with the best model (SVM) resulting in an accuracy of 73% and 72% for entity 1- and entity 2-reported data. In both cases, the neural network had the lowest accuracy score of 39%. The performance of SVM and Discriminant Analysis models is at an acceptable level given the complexity of the data, and in the presence of potential outlier information in the reported data. Moreover, different projects had to be merged to generate a large enough dataset for ML applications, which resulted in merging non-homogeneous data from various projects. All in all, I pose that ML algorithms performed successfully, confirming that the human detected logics are also differentiable with machine.

Conclusion

The construction industry is exigent for national opulence and growth. It boosts the economy and augments the country's Gross Domestic Product (GDP). In any developed or developing country, both public and private owned construction sectors play a pivotal role in the growth of the country. The devastating impression of suspicious activities and alteration in reported data can impose turmoil between the government agencies and contractors. This also refutes the perception of the construction industry in front of the general public.

QC/QA is an integral step to ensure the quality of the HMA construction works. This statistics-based approach has been followed by state highway agencies for quite a period now. However, there are some concerns about representativeness of the reported material testing data. My study focused on potential data alteration during the QC/QA processes. This has a significant impact; as potential alterations on the reported data can jeopardize the quality of asphalt pavements and cause overpayment on HMA projects.

Through this research, I analyzed an audit dataset of material testing reports that registered all value entries in the Excel reporting files. The series of changes in parameter values can shed important insights on the potential sources of discrepancies that are observed between contractor test results and those of the transportation departments and the mix design. I first manually analyzed all the provided instances of changes in the parameter values, and determined the general patterns in data reporting. I categorized these instances to two general categories of Plausible Correction (P.C.) and Suspicious Alteration (S.A.). I then developed logic-based computer algorithms to automatically classify all instances of parameter value changes to P.C. and S.A. I then rigorously evaluated the automatic classification results to evaluate computer algorithms' performance. My results show that a total of 595 and 316 unique parameters were changed 2,268 and 660 times that can be categorized as S.A. and P.C., respectively, in entity 1-reported data. For entity 2-reported data, a total of 387 and 280 unique parameters were changed 1,266 and 587 times that can be categorized as S.A. and P.C., respectively. My results indicated that major parameters were altered four to five times on average per parameter. Parameter values for plausible correction cases were mostly changed one time.

I also successfully prompted supervised machine learning technique to detect S.A. instances from P.C. cases. Given the unavailability of independent labeled data, I utilized the categorized data from my logic-based analysis to train the ML algorithms. Supervised ML algorithms like Support Vector Machine and Discriminant Analysis, achieving accuracy levels of more than 70%, parades well harmony with the logic-based categorized results.

My findings emphasize the necessity of an advanced cumulative approach to improve QC/QA process. A better approach is needed to remove probable unethical course of actions and bring more rigor to QC/QA analysis.

References

Al-Khayat, H., Gurganus, C., Newcomb, D. E., and Sakhaeifar, M. S. (2020). Developing Specification Limits for Hot Mix Asphalt Properties and Impact on Pay Factors. Journal of Transportation Engineering Part B: Pavements, 146(3), 1–8.

Association of Certified Fraud Examiners (ACFE). (2010). Report to the nations on occupational fraud and abuse. Accessed at: https://www.acfe.com/uploadedFiles/ACFE_Website/Content/documents/rttn-2010.pdf, retrieved on August 1, 2020.

Atwood, J. A., Robinson-Cox, J. F., and Shaik, S. (2006). Estimating the Prevalence and Cost of Yield-Switching Fraud in the Federal Crop Insurance Program. American Journal of Agricultural Economics, 88(2), 365–381.

Bagli, C. V. (2018). Investigators Eye Possible $100 Million Construction Fraud. The New York Times. Accessed at: https://www.nytimes.com/2018/02/26/nyregion/bloomberg-interior-construction-fraud.html, retrieved on August 1, 2020.

Bell, T. B., and Carcello, J. V. (2000). A decision aid for assessing the likelihood of fraudulent financial reporting. Auditing, 19(1), 168–184.

Caudill, S. B., Ayuso, M., and Guillén , M. (2005). Fraud detection using a multinominal logit model with missing information. The Journal of Risk and Insurance, 72(4), 539-550.

Coenen, A. R., Pforr, J. E., Hefel, S. A., and Paye, B. C. (2019). State DOT Implementation of Statistical Analysis and Percent Within Limits. Transportation Research Record, 2673(2), 583–592.

Craig, G. (2019). Three contractors accused of fraud in billion-dollar school modernization project. Democrat and Chronicle. Accessed at: https://www.democratandchronicle.com/story/news/2019/08/27/three-contractors-accused-fraud-construction-plan-reach-deal-feds/2131956001/, retrieved on August 1, 2020.

De Jong, M., Henry, W. P., and Stansbury, N. (2009). Eliminating corruption in our engineering/construction industry. Leadership and Management in Engineering, 9(3), 105–111.

Dutton, A. (2020). Idaho highway contractors changed records hundreds of times. Then, they got bonuses. Idaho Statesman. Accessed at: https://www.idahostatesman.com/news/local/investigations/watchdog/article239690033.html, retrieved on August 1, 2020.

Gunduz, M., and Önder, O. (2013). Corruption and Internal Fraud in the Turkish Construction Industry. Science and Engineering Ethics, 19(2), 505–528.

Hand, A. J. T., Nimeri, M. A., Hajj, E. Y., Sebaaly, P. E., West, R. C., Heitzman, M. A., Yin, F., Hughes, C. S., and Tayabji, S. (2020). Procedures and Guidelines for Validating Contractor Test Data. NCHRP Research Report 946, National Academies of Sciences, Engineering, and Medicine 2020, Washington, DC: The National Academies Press. https://doi.org/10.17226/25823.

Harvey, I. (2019). Forensic investigation into $80M Bondfield fraud digs deeper. Daily Commercial News by ConstructConnect. Accessed at: https://canada.constructconnect.com/dcn/news/government/2019/12/forensic-investigation-into-80m-bondfield-fraud-digs-deeper, retrieved on August 1, 2020.

Hastie, T., Tibshirani, R., and Friedman, J. (2008). The Elements of Statistical Learning. Springer Series in Statistics. https://web.stanford.edu/~hastie/Papers/ESLII.pdf

Hewitt, D. (2015). China To Demolish New Skyscraper As Part Of Anti-Corruption Campaign. International Business Times. Accessed at: https://www.ibtimes.com/china-demolish-new-skyscraper-part-anti-corruption-campaign-2225814, retrieved on August 1, 2020.

ITD. (2019). FY20 to FY26: Idaho Transportation Investment Program (ITIP). Accessed at: https://apps.itd.idaho.gov/Apps/Fund/itip2020/draft/FY20_Draft_itip.pdf, retrieved on August 1, 2020.

Jain, A., and Shinde, S. (2019). A Comprehensive Study of Data Mining-based Financial Fraud Detection Research. IEEE 5th International Conference for Convergence in Technology (I2CT), 2–5.

Kirkos, E., Spathis, C., and Manolopoulos, Y. (2007). Data Mining techniques for the detection of fraudulent financial statements. Expert Systems with Applications, 32(4), 995–1003.

Kotsiantis, S., Koumanakos, E., Tzelepis, D., and Tampakas, V. (2006). Forecasting Fraudulent Financial Statements using Data Mining. International Journal Of Computational Intelligence, 3(2), 104–110.

Kyriacou, A. P., Muinelo-Gallo, L., and Roca-Sagalés, O. (2015). Construction corrupts: empirical evidence from a panel of 42 countries. Public Choice, 165(1–2), 123–145.

Lin, J. W., Hwang, M. I., and Becker, J. D. (2003). A fuzzy neural network for assessing the risk of fraudulent financial reporting. Managerial Auditing Journal, 18(8), 657–665.

Markelevich, A., and Rosner, R. L. (2013). Auditor fees and fraud firms. Contemporary Accounting Research, 30(4), 1590–1625.

Manning, R. (2019). Portland Public Schools Respond To Fraudulent $3 Million Construction Payment. Oregon Public Broadcasting. Accessed at: https://www.opb.org/news/article/portland-public-schools-fraud-3-million-construction-payment/, retrieved on August 1, 2020.

National Asphalt Pavement Association (NAPA). (n.d.). Market Facts. Accessed at: https://www.asphaltpavement.org/, retrieved on August 1, 2020.

Newcomb, D. E., Gurganus, C., Al-Khayat, H., Sakhaeifar, M., and Epps, J. A. (2016). Review Of Oregon Department Of Transportation Asphalt Mix Sprcification, Phase II, 1–26.

Ngai, E. W. T., Hu, Y., Wong, Y. H., Chen, Y., and Sun, X. (2011). The application of data mining techniques in financial fraud detection: A classification framework and an academic review of literature. Decision Support Systems, 50(3), 559–569.

Nordin, R. M., Takim, R., and Nawawi, A. H. (2013). Behavioural Factors of Corruption in the Construction Industry. Procedia - Social and Behavioral Sciences, 105, 64–74.

Perols, J. (2011). Financial statement fraud detection: An analysis of statistical and machine learning algorithms. Auditing: A Journal of Practice & Theory, 30(2), 19–50.

Perols, J. L., and Lougee, B. A. (2011). The relation between earnings management and financial statement fraud. Advances in Accounting, 27(1), 39–53.

Perols, J. L., Bowen, R. M., Zimmermann, C., and Samba, B. (2017). Finding needles in a haystack: Using data analytics to improve fraud prediction. The Accounting Review, 92(2), 221–245.

Robertson, L. (2014). Scale of international bribery laid bare by new OECD report. Organisation for Economic Co-operation and Development. Accessed at: https://www.oecd.org/daf/anti-bribery/scale-of-international-bribery-laid-bare-by-new-oecd-report.htm, retrieved on August 1, 2020.

Rodriguez, D., Waite, G., & Wolfe, T. (2005). Global corruption report 2005. Transportation International. https://images.transparencycdn.org/images/2005_GCR_Construction_EN.pdf

Sohail, M., and Cavill, S. (2008). Accountability to prevent corruption in construction projects. Journal of Construction Engineering and Management, 134(9), 729–738.

Srivastava, A., Kundu, A., Sural, S., and Majumdar, A. K. (2008). Credit card fraud detection using Hidden Markov Model. IEEE Transactions on Dependable and Secure Computing, 5(1), 37–48.

West, J., Bhattacharya, M., and Islam, R. (2015). Intelligent Financial Fraud Detection Practices: An Investigation. In: International Conference on Security and Privacy in Communication Systems, 186–203.

Zarkada-Fraser, A., and Skitmore, M. (2000). Decisions with moral content: Collusion. Construction Management and Economics, 18(1), 101–111.

CHAPTER 3: MONETARY ANALYSIS TO QUANTIFY THE AMOUNT OF ECONOMIC LOSS DUE TO DATA ALTERATION

Abstract

The Idaho Transportation Department (ITD) is responsible for collecting material testing data from Hot Mix Asphalt (HMA) projects across Idaho to evaluate their quality through statistical quality control/quality assurance (QC/QA) assessment. ITD conducted a forensic investigation on contractor reported QC data, where only 26% of the contractor results were in good agreement with the ITD-produced test results. This pointed out the question of suspicious alterations in material testing data. ITD incorporated a VBA macro into the material testing report Excel files, which recorded every instance of parameter value entry. These files provided a sequence of value changes for many parameter values. This change in material testing data can originate from operator/equipment error as well as intentional/unintentional data alteration in an HMA project. In any form, those error/mistakes risk the quality of the end product and can cause monetary loss. In this chapter, I analyzed the monetary impact of such data alteration and calculated the payments with and without data alterations. A majority of the analyzed projects showed a significant over-payment due to data alterations. My analysis also showed that in the absence of data alteration, only one third of the lots, for which audit data was available, would pass the percent within limit thresholds – i.e. were at an acceptable level.

Introduction

State highway agencies have adopted the use of QC/QA specifications programs for the construction of asphalt pavement in recent decades (Butts and Ksaibati, 2002). This specification is adopted to ascertain better performing and long-lasting roadways through decreasing the deviation in asphalt production materials from the design level. The Transportation Research Board (TRB) has defined QC/QA as the combination of end result specifications, and materials and methods specifications. QA specifications in general represent the quality level in statistical terms i.e. mean, standard deviation, percent within limits, among others (Akkinepally & Attoh-okine, 2006). Departments of transportation (DOTs) usually cover numerous projects, and usually lack the needed resources to conduct QA analyses in house, and hence hire third-party contractors to conduct QA testing (Coenen et al., 2019). The statistical specifications of QC/QA are prone to multiple errors which can occur both intentionally and unintentionally. Individual personnel or equipment can potentially lead to unexpected errors, and intentional statistical parameter/material data alteration can pursue certain goals. The target of this work was to quantify the financial impacts of data alterations in HMA projects. While I refrain from using fraud for the detected suspicious data alterations in this study, given that a pure data mining approach is not able to detect/classify fraud, I provide a brief literature review on fraudulent activities in various sectors in the following paragraph. This helps in putting a worst case from this research scenario – which might or might not have materialized – into broader context.

Fraud and financial crime negatively impact a variety of sectors and people, ranging from the public to investors (Perols, 2011), and hence have attracted a great deal of attention in recent times (Ngai et al., 2011). With the advancement of technology and digitization of the paper-based financial works, contrary to expectations, fraudulent activities associated with financial statements have significantly increased (Coutorie, 1995). Financial fraud detection is cardinal for preventing grave consequences of fraudulent activities. The complex nature of financial fraud, however, makes it challenging to prevent such incidents (Salem, 2012). Data mining and artificial intelligence have been widely used as anomaly detection methods, finding interesting patterns and hidden truth in an ever-increasing amount of available data, to detect, deter and prevent fraudulent activities (Frawley et al., 1992; Turban et al., 2007; Bose and Mohapatro, 2011).

The construction industry is widely known for its association with corruption and fraud, due to its complex and heterogenous nature, as well as complicated involvement of third-party contractors (Gunduz and Onder, 2013). The global construction market is worth around 3,200 billion USD per year (Sohail and Cavill, 2008), and this huge flow of money makes this sector vulnerable and prone to fraudulent activities. Corruption in construction is remarkably active at various stages, ranging from selection of contractors, ordering construction materials, and bribing officials to pass substandard works and manipulating construction data to increase payment, among others (Sohail and Cavill, 2008). Corruption in public construction projects are believed to be more prevalent, and also detrimental, in developing countries because of resource limitation and deficiency in institutional capacity to detect and prevent fraud (Hardoon and Heinrich, 2011). Several factors like the uniqueness of the project, intense competition between contractors, several and often inconsistent levels of bureaucracy for obtaining official approvals, and flexibility in project delays and overruns contribute to prevalence of fraudulent activities in the construction sector and thereby cause suboptimal project deliverables (Rodriguez et al., 2005; Bowen et al., 2007).

Corruption and fraud in the construction industry can negatively impact the desired objectives in various ways, including but not limited to, cost overruns, poor quality, less efficient project selection, and increasing maintenance costs (Kenny, 2006, 2009; Kyriacou et al., 2015). Financial fraud analysis is still a new and underexplored aspect of the construction sector. Most of the literature has focused on blackmail, bribery, embezzlement, increased project costs, and tendering uncertainty (Sohail and Cavill, 2008; Le et al., 2014; Locatelli et al., 2017). Other forms of financial fraud, like credit card fraud, corporate fraud, telecommunications fraud, and money laundering, however, have the focus of much research and analysis in recent times (Ngai et al., 2011).

I received a unique dataset of material testing reports for Hot Mix Asphalt (HMA) construction projects in Idaho, that recorded every instance of data entry in the Excel reporting file. Data recording was conducted in the background with a VBA code, and was not apparent to the data reporting personnel. This provided a series of data entry for some material testing parameters, which show data alteration in many parameters. It is expected that each parameter be reported as observed, and hence being reported only once, although typographical errors may result in multiple entries for some parameters. The patterns observed in some parameters in the audit data, however, cannot be simply explained as typographical errors. As described in Chapter 2, I applied a series of logic-based algorithms to categorize all instances of multiple (more than 1) data entry as either Plausible Correction (P.C.) or Suspicious Alteration (S.A.). I refrain from using a blanket statement of “fraudulent activities”, as a mere data mining approach may not justify categorizing all suspicious changes in the reported data as fraud. However, I pose that S.A. instances cannot simply and readily be explained as typographical errors or other forms of mistakes.

The aim of this study is to analyze the financial repercussions and impacts of S.A. instances. It is plausible that data alterations can occur for monetary benefit or personal/institutional advantage. Suspicious alterations may also have been done to obtain bonus payments, avoid repetition of faulty tests and works, and pass substandard work.

Scope of Work

The scope of the current work was to calculate the monetary loss that occurred in HMA pavement projects due to alterations in material testing reports. In the previous chapter, I differentiated the Suspicious Alteration (S.A.) instances from the Plausible Correction (P.C.) cases for multiple data entry values in volumetric testing reports. This chapter will demonstrate the economic impact of S.A. cases. I calculated the required financial payment to contractors if only the first acceptable instance of S.A. data entry was used, and compared it to the project payment based on the reported values (final S.A. instances). I considered the last entry for all P.C. instances and adopted the final reported values for the missing parameters. The basic procedure was to go through the exact same calculation procedures followed by the Idaho Transportation Department (ITD) for monetary calculation, quantify the payment-related parameters and associated payment for each lot in each project. To avoid inherent bias during the monetary analysis, this thesis uses the names “Entity 1” and “Entity 2” to refer to agency and contractor data, not necessarily in the same order. In other words, it has not been disclosed to the reader whether Entity 1 represents data from the agency or contractor. The same is the case for Entity 2.

Monetary Calculation

ITD has a certain set of rules to determine how a contractor will be paid for a Hot Mix Asphalt (HMA) project. Several input parameters, like Mass of Bowl, Mass Pan and Initial Sample, and Calibration factor, are calculated while performing an HMA project.

Once a test is completed, test results are grouped as lots based on pre-specified lot calculation rules. Payment is finally calculated per lot. The required input parameters are translated into a group of asphalt mix design properties such as G_{mm} (Theoretical maximum specific gravity), G_{mb} (Bulk specific gravity), P_a (Air voids), VMA (Voids in the mineral aggregate), and VFA (Voids filled with asphalt), among others. These mix design properties are then used as acceptance criteria at the start of the production. Out of these calculated mix design properties, three variables, namely Air voids, VMA, and Mainline Density (Percent compaction), are used for final payment calculation. All the project data that we received were from before 2020, so the calculation procedure is from earlier ITD payment conventions.

Fig. 3-1 illustrates the overall representation of the generic input parameters tested in the lab/plant and later converted to mix design properties. These Excel sheets are identified as “ITD-0777” form. The input parameters are shown on the left-hand side, and the calculated mix design properties are located on the right-hand side. Generally, these calculations are done for two samples (Sample 1A and Sample 1B), which are then averaged, and the combined values of Air voids, VMA, and Mainline Density (Percent Compaction) are used for payment calculation.

FOP for AASHTO T 209 Theoretical Max Specific Gravity (Bowl Method)

T209 Sample Reduction Method	Date Reduced	Time Reduced	Sample Temperature
			77 °F
Final Reduction for T209 Performed By			WAQTC Number

$$G_{mm} = \frac{A}{A - C}$$

	Increment 1	Increment 2
Mass of Bowl *(Required)*	2122.9	2122.9
Mass of Bowl and Sample	3684.4	3688.7
Mass of Dry Sample in Air (A)	1561.5	1565.8
Agitation Method	Mechanical	
Water Bath Temperature	76.3 °F	76.6 °F
Submerged Weight of Bowl and Sample	2250.3	2254.2
Submerged Weight of Bowl	1337.9	1337.9
Submerged Weight of Sample (C)	912.4	916.3
G_{mm} (Maximum Specific Gravity)	2.406	2.411
Average G_{mm}	**2.408**	

Range **0.005** Accpetable? (Within d2s precision): **YES**

FOP for AASHTO T 312 SuperPave Gyratory Compactor

T312 Sample Reduction Method	Date Reduced	Time Reduced	Sample Temperature
			284 °F
Final Reduction for T312 Performed By			WAQTC Number
Gyratory Compactor Brand	Model Number	Serial Number	

	Specimen 1	Specimen 2	Design Mass
Mass of Sample	4654.6	4654.9	4650.0
Temp. of Sample When Placed in Mold	300 °F	300 °F	
Time Compaction Begins	1:57 AM	3:07 AM	Spec Limits
Sample Height (mm)	113.7	113.5	115±2

FOP for AASHTO T 166 Bulk Specific Gravity of Compacted Mix (Method A)

$$G_{mb} = \frac{A}{B - C}$$

	Specimen 1	Specimen 2
Surface Temperature	71.4 °F	74.6 °F
Water Bath Temperature	77.8 °F	77.7 °F
Mass of Puck Dry (A)	4651.7	4649.7
Submerged Weight of Puck in Water (C)	2681.6	2677.1
Wt. of Puck SSD (B)	4656.6	4653.5
G_{mb} (Bulk Specific Gravity)	2.355	2.353
Average G_{mb}	**2.354**	

Range **0.003** Accpetable? (Within d2s precision): **YES**

Summary of Mix Properties

Property	Sample 1A	Sample 1B	Combined	LSL	USL
G_{sa}	2.656	2.656	2.656		
G_{se}	2.614	2.621	2.617		
G_{sb}	2.578	2.578	2.578		
G_{mm}	2.406	2.411	2.408		
G_{mb}	2.355	2.353	2.354		
Abs_{t166}	0.248	0.192	0.220		
G_b	1.0310	1.0310	1.0310		
P_b	5.65	5.65	5.65		
P_{ba}	0.55	0.65	0.60		
P_{be}	5.12	5.03	5.08		
P_s	94.4	94.4	94.4		
SA	32.2	32.2	32.2		
AFT	8.20	8.06	8.13		
P_a	2.09	2.41	2.3	3.0	5.0
VMA	13.80	13.90	13.8	14.0	
VFA	84.83	82.64	83.7	65.0	75.0
P200	6.02	6.02	6.0	3.8	6.8
DP	1.18	1.20	1.2	0.6	1.2

Figure 3-1 Typical data input file for asphalt pavement projects

Once the payment-related parameters are calculated for each test, tests are grouped to form a lot, and payments are calculated based on some statistical tests on the lot data (details later).

Lot Grouping: Payment factors are calculated for each lot, but based on F and T tests from a group of tests that might include several lots. Grouping is done to enhance the diagnostic power of F and T tests. If the group consists of only one lot, then payment is calculated for that individual lot, whereas if the lot group has multiple lots then payment is calculated for all the lots together. ITD has set certain defining formulas to group the lots for payment. For each lot, a few parameters define payment related calculations including "Start of evaluation range" (lot number from where the evaluation would start) and "End of evaluation Range" (lot number for which the payment would be calculated). For example, in Fig. 3-2, for lot 2, the evaluation range started from lot 2 and also ended at 2. So, for this lot, no other lot is grouped for payment calculation. For lot 6, the evaluation started at lot 4 and ended at 6. So, all the tests from lots 4, 5, and 6 would be grouped together for payment calculation of lot 6.

LotStatus		
LotNumber	Start of Evaluation Range	End of Evaluation Range
1	1	1
2	2	2
3	2	3
4	4	4
5	4	5
6	4	6
7	7	7
8	8	8
9	9	9
10	10	10
11	11	11
12	12	12
13	12	13
14	14	14
15	15	15
16	16	16
17	17	17

Figure 3-2 Lot evaluation range for payment calculation

Test Statistics: Mean and standard deviation value for Air Voids, VMA, and Mainline Density of a lot group both from the entity 1 and the entity 2-reported data are calculated first. From those values, a pass/fail test check is done using F & T tests. If p-values for both Air Voids and VMA are below 0.05, then they pass the test. So, we have two p-values from the F test for Air Voids and VMA. Similarly, there is another p-value check for T test for both parameters. If data are passed based on both F and T tests for both Air Voids and VMA, then the project lot gets a green signal, and entity 2 data is selected for payment. If in any of these tests, p-value exceeds 0.05 (rejected), then the test fails, and the entire lot is rejected for payment based on the entity 2 data; instead, the entity 1-reported data is selected for calculating payment factor.

Determination of Percent Within Limits (PWL): The next step of the calculation of payment factor is the determination of PWL values. The lot average Air Voids, VMA, and Mainline Density values are considered, and through a series of calculations, PWL values are measured. The final payment factor for all three payment affecting parameters is computed through the following equation (3-1).

$$Pay\ factor = \frac{55+0.5\times PWL}{100} \tag{3-1}$$

The final payment value is then computed for the lot, using:

$$Payment = \ Pay\ factor \times Quantity\ represented\ by\ lot \times Contract\ unit\ price \tag{3-2}$$

Here, "Quantity represented by lot" is the total volume of asphalt pavement produced in the lot and "Contract unit price" is the unit price to be paid to the contractor.

Formation of Input Data for Monetary Calculation

I created two sets of data: first and last reported S.A. value, which will subsequently be used for monetary impact analysis. My hypothesis is that the first "acceptable" S.A. value is the original value that was measured for that parameter, whereas the last value is the final reported value after alterations. The difference in payment calculations for these two cases is assumed to be the monetary loss to suspicious activities in the material testing reports. As a reminder, we have three types of data: non-repeated data (one value is reported) and repeating data with P.C. and S.A. categorization (multiple data entry were recorded for each parameter). Since only the S.A. data can be held responsible for any sort of economic impact, I have selected first and last entry of the S.A. cases. The P.C. and non-repeated cases don’t have any influence on the monetary value, so they adopted their reported values. Also, any missing parameter value is assigned its reported value.

Sample	Cell	Value	Time	Var_effect_typ	Var_err
Test(16)	U62	2804.2	4:06:48 PM	Major	S.A.
Test(16)	U62	2805.2	4:07:07 PM	Major	S.A.
Test(16)	U62	2806.2	4:07:09 PM	Major	S.A.
Test(16)	U62	2808.2	4:07:11 PM	Major	S.A.
Test(10)	U63	2811.4	11:52:19 PM	Major	P.C.
Test(10)	U63	4823	11:52:27 PM	Major	P.C.

Figure 3-3 Classified Plausible Correction (P.C.) and Suspicious Alteration (S.A.) data

As an example, in Fig. 3-3, cell U62 (Submerged weight of puck in water (specimen 1)) from test Test(16) has three repetitions with a total of four values and falls in the S.A. category. Hence, the first value of 2804.2 was selected for my first dataset (that will be used for original payment calculation) and the last value of 2808.2 was selected for the second dataset (that will be used for payment calculation after alterations). Cell U63 (Weight of puck SSD (specimen 1)) from Test(10) falls in the P.C. category. So, I picked the final value of 4823 for both datasets. For non-repeated cells, the single corresponding value was kept for both datasets.

A Python code was generated to accomplish these steps. The code is designed to adopt the first and last values of S.A. and to take the last value of P.C. from the previously categorized audit data, and to take the final reported value for all non-repeating and missing variables. A sample of the newly generated dataset is presented in Fig. 3-4. I included tests on the rows and parameters/cells associated with each test in the columns.

	\$BE\$124	\$U\$25	\$O\$149	\$S\$112	\$Z\$33	\$U\$62	\$Z\$38	\$Z\$62	\$O\$18	\$S\$162	\$Z\$52	\$AG\$47	\$Z\$63
Test(1)	22444	22444	1062.8	4599.6	3688.7	2681.6	1337.9	2677.1		22444	4649.7	22444	4653.5
Test(2)	22444	22444	1074.8	4584.5	3691	2680.6	1337.9	2673.8	0.911805556	22444	4655.2		4650.7
Test(3)	22444	22444	1056.5	4608.8	3692.1	2677.6	1337.9	2678.3	0.927083333	22444	4655.3		4653.8
Test(4)	22444	22444	1099.1	4587.5	3697.5	2673.8	1356.6	2668.4	0.913194444	22444	4654.6	22444	4645.1
Test(5)	22444	22444	1111.2	4592.5	3715.1	2659.8	1356.6	2655.7	0.00625	22444	4654.6	22444	4645.2
Test(6)	22444	22444	116.9	4629	3720.7	2666	1356.6	2665.4		22444	4654.5	22444	4653
Test(7)	22444	22444	1106.4	4601	3707	2671	1356.2	2670.8	0.9375	22444	4658	22444	4658.3
Test(8)	22444	22444	1135.1	4612.1	3708.7	2666	1356.2	2654.8	0.026388889	22444	4656.9	22444	4657.4
Test(9)	22444	22444	1130.3	4598.8	3725.6	2667.7	1356.3	2670	0.954861111	22444	4656.8	22444	4663.7
Test(10)	22444	22444	1181.4	4605.1	3725.8	2667.2	1356.6	2657		22444	4655.6	22444	4646
Test(11)	22444	22444	1229	4774.9	3726.9	2655.3	1356.6	2650.7	0.472222222	22444	4656.3	22444	4664.5
Test(12)	22444	22444	1111.3	4601.6	3704.1	2657.1	1356.6	2656.9	0.092361111	22444	4656.1	22444	4656.1
Test(13)	22444	22444	1116.7	4599.6	3700.2	2663.5	1356.2	2664	0.860416667	22444	4656.6	2244	4659.6
Test(14)	22444	22444	1109.1	4602	3692.3	2676.7	1356.2	2675.5	0.980555556	22444	4656.2	22444	4659.4
Test(15)	22444	22444	1143.1	4636.4	3699.6	2670	1356.2	2668.9	0.0625	22444	4657	22444	4657.6
Test(16)	22444	22444	1100.9	4608.1	3686.3	2656.1	1356.3	2657.5	0.922222222	22444	4656	22444	4657
Test(17)	22444	22444	1098.6	4608.6	3686.3	2656.1	1356.3	2658.5	0.965277778	22444	4656.3	22444	4657
Test(18)	22444	22444	1142.5	4624.2	3709.3	2670	1356.5	2667.4	0.928472222	22444	4655	22444	4645.9
Test(19)	22444	22444	1112.9	4603	3685.5	2667.1	1354.6	2663.4	0.958333333	22444	4655.6	22444	4653.9
Test(20)	22444	22444	1172.8	4618	3704.3	2665.5	1354.6	2668.5	0.041666667	22444	4655.9	22444	4653
Test(21)	22444	22444	1149.8	4635.9	3695.6	2674.4	1354.6	2672.3	0.104166667	22444	4655.5	22444	4656.3

Figure 3-4 Input dataset for monetary calculation: rows show test number and columns represent parameter values associated with each test

This step was associated with some challenges. There were instances in which the first or last S.A. and last P.C. data had an empty cell, which precluded us from calculating monetary values. These empty cells created unreasonably large, negative or not-a-number (NaN) values for my target parameters (Air Voids/VMA/Mainline Density). Hence, I devised some strategies to fill empty values. For the first entry, if the value was empty, I selected the second cell value; if the second was empty, I looked for the next one and continued until I found a value. A similar process was done for obtaining the value of last cell but in a reverse order. I plugged the cell value before the last cell if the last one was empty. I continued these steps from the last cell backwards until I found a value. Fig. 3-5 demonstrates a missing first entry for cell \$U\$37 (Submerged weight of bowl and sample (increment 1)), for which the next value was adopted.

Sample	Cell	Value	Time	Var_effect_typ	Var_err	T
Test(9)	U37		3:28:36 AM	Major	S.A.	
Test(9)	U37	2269.8	3:28:51 AM	Major	S.A.	
Test(9)	U37	2279.8	3:30:05 AM	Major	S.A.	
Test(9)	U37	2269.8	3:30:13 AM	Major	S.A.	
Test(9)	U37	2259.8	3:31:11 AM	Major	S.A.	
Test(9)	U37	2262.8	3:31:20 AM	Major	S.A.	

Figure 3-5 Empty cell for some parameters

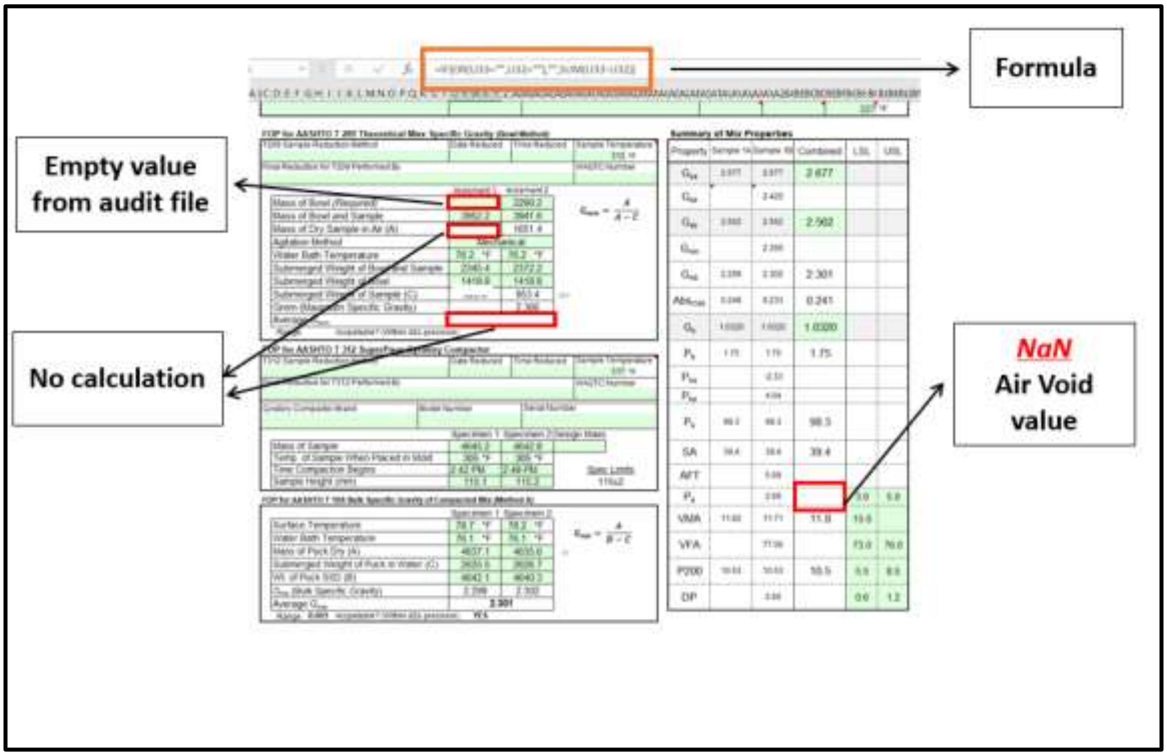


Figure 3-6 Effect of empty/NaN cells on calculated payment parameters

Fig. 3-6 demonstrates an example problem associated with having a NaN value for an input parameter. Because we had missing values for one of the cells, several calculations were not possible and resulted in NaN value for Air Voids. Since secondary parameter values (payment-related parameters) depend on various primary parameters, lack of primary parameter values will preclude calculation of secondary values.

This final dataset was used to test and apply the formulas from "ITD-0777" to evaluate the monetary values (payments based on first and last S.A. values). A code was prepared in Matlab in this step, which replicated the original calculation flow of the "ITD-0777" file and extracts the Air Voids, VMA, and Mainline Density values. To ensure the accuracy of the calculations, another code was prepared at this step to plug in the parameter values directly into the "ITD-0777" file. This enabled me to calculate the parameter values both from the coded program and from the "ITD-0777" file. I cross-checked several projects to ensure the calculated monetary values through my code and "ITD-0777" file were exactly matching. The direct monetary calculation through my code was much faster as it could automatically produce all the test parameters of a project. Figure 3-7 demonstrates the calculation of Air voids and VMA values for each test of a sample project through the direct calculation in my code. Similar values were obtained from the "ITD-0777" file.

	1	2	3
1	1	2.2530	13.8470
2	2	2.4360	13.7259
3	3	2.3511	13.9852
4	4	2.7840	13.7067
5	5	3.9509	14.1524
6	6	3.2369	14.0890
7	7	3.3439	13.9108
8	8	4.0792	14.2173
9	9	3.6062	14.0943
10	10	4.0250	13.6778
11	11	4.7919	15.0584
12	12	4.1021	14.5232
13	13	3.6756	14.2097
14	14	3.3518	13.8192
15	15	3.4700	13.9369
16	16	3.9073	14.5932
17	17	3.8401	14.6379
18	18	3.2519	13.9749
19	19	3.4856	14.2646
20	20	3.4537	14.3423
21	21	3.3110	13.7850
22	22	3.8958	14.5322
23	23	3.9206	14.5253
24	24	3.9042	14.6839
25	25	3.4800	14.2475
26	26	3.8978	14.5547
27	27	3.5027	14.0945
28	28	4.3486	14.2865
29	29	3.8418	13.6723

Figure 3-7 Calculated Air voids (column 2) and VMA (column 3) for an example project

FOP for AASHTO T 209 Theoretical Max Specific Gravity (Bowl Method)

T209 Sample Reduction Method	Date Reduced	Time Reduced	Sample Temperature 77 °F
Final Reduction for T209 Performed By			WAQTC Number

$$G_{mm} = \frac{A}{A - C}$$

	Increment 1	Increment 2
Mass of Bowl *(Required)*	2122.9	2122.9
Mass of Bowl and Sample	3684.4	3688.7
Mass of Dry Sample in Air (A)	1561.5	1565.8
Agitation Method		
Water Bath Temperature	76.3 °F	76.6 °F
Submerged Weight of Bowl and Sample	2250.3	2254.2
Submerged Weight of Bowl	1337.9	1337.9
Submerged Weight of Sample (C)	912.4	916.3
G_{mm} (Maximum Specific Gravity)	2.406	2.411
Average G_{mm}	2.408	

Range **0.005** Accpetable? (Within d2s precision) **YES**

T312 Sample Reduction Method	Date Reduced	Time Reduced	Sample Temperature 284 °F
Final Reduction for T312 Performed By			WAQTC Number
Gyratory Compactor Brand	Model Number	Serial Number	

	Specimen 1	Specimen 2	Design Mass
Mass of Sample		4649.7	
Temp. of Sample When Placed in Mold	300 °F	300 °F	
Time Compaction Begins	1:57 AM	3:07 AM	Spec Limits
Sample Height (mm)	113.7	113.5	115±2

FOP for AASHTO T 166 Bulk Specific Gravity of Compacted Mix (Method A)

$$G_{mb} = \frac{A}{B - C}$$

	Specimen 1	Specimen 2
Surface Temperature	80.4 °F	79.6 °F
Water Bath Temperature	77.8 °F	77.7 °F
Mass of Puck Dry (A)	4651.7	4649.7
Submerged Weight of Puck in Water (C)	2681.6	2677.1
Wt. of Puck SSD (B)	4656.6	4653.5
G_{mb} (Bulk Specific Gravity)	2.355	2.353
Average G_{mb}	2.354	

Range **0.003** Accpetable? (Within d2s precision) **YES**

Summary of Mix Properties

Property	Sample 1A	Sample 1B	Combined	LSL	USL
G_{sa}	2.656	2.656	2.656		
G_{se}	2.614	2.621	2.617		
G_{sb}	2.578	2.578	2.578		
G_{mm}	2.406	2.411	2.408		
G_{mb}	2.355	2.353	2.354		
Abs_{166}	0.248	0.192	0.220		
G_b	1.0310	1.0310	1.0310		
P_b	5.65	5.65	5.65		
P_{ba}	0.55	0.65	0.60		
P_{be}	5.12	5.03	5.08		
P_s	94.4	94.4	94.4		
SA	32.2	32.2	32.2		
AFT	8.20	8.06	8.13		
P_a	2.09	2.41	2.3	3.0	5.0
VMA	13.80	13.90	13.8	14.0	
VFA	84.83	82.54	83.7	65.0	75.0
P200	6.02	6.02	6.0	3.8	6.8
DP	1.18	1.20	1.2	0.6	1.2

Figure 3-8 Calculation of Air voids and VMA through ITD-0777 file

Fig. 3-8 shows an example "ITD-0777" file where all the input values have been inserted, and calculations were done by the internal formulas of this sheet. Since this procedure is lengthy and can only be done for one test at a time, the developed code that replicates "ITD-0777" file was used for the remainder of my analysis. However, I randomly selected 3 tests from each project to cross-check individual test results with the previously discussed code produced results.

Unavailability of Audit Files: Unfortunately, we didn't have the audit files for all projects. On many occasions, the audit files didn't have the recorded values for all the tests of a project. Sometimes there were no audit files for neither entity 1 nor entity 2-reported data. Since we need data both from both entities, I considered the reported values from the project files where audit values were missing.

Lot Number	Test Number	Sample Time	Air Voids	VMA
1	1	7:38:00 AM	4.12	15.61
1	2	1:10:00 PM	5.51	15.90
1	3	1:30:00 PM	5.27	15.79
2	4	9:58:00 AM	3.59	15.00
2	5	12:40:00 PM	3.55	15.10
2	6	3:30:00 PM	4.09	15.50
2	7	4:30:00 PM	3.21	14.90
3	8	8:00:00 AM	4.15	15.25
3	9	9:30:00 AM	4.38	15.86
3	10	12:15:00 PM	3.69	15.05
4	11	7:00:00 AM	2.77	14.29
4	12	8:45:00 AM	2.35	13.80
4	13	12:00:00 PM	2.88	14.28
4	14	2:45:00 PM	3.25	14.40
4	15	3:00:00 PM	3.39	14.76
5	16	9:45:00 AM	3.83	15.09
5	17	11:30:00 AM	4.37	15.18
5	18	4:20:00 AM	5.08	15.90
6	19	7:30:00 AM	3.43	14.63
6	20	7:45:00 AM	4.27	15.42
6	21	11:00:00 AM	4.94	15.76
7	22	9:45:00 AM	4.07	15.68
7	23	12:45:00 PM	4.02	15.01
7	24	1:50:00 PM	3.93	14.99
7	25	4:25:00 PM	3.98	15.21
8	26	9:20:00 AM	4.79	15.68
8	27	3:06:00 PM	4.44	15.34
8	28	3:41:00 PM	4.65	15.49
9	29	6:40:00 AM	5.33	16.21
9	30	8:25:00 AM	6.30	16.93
9	31	11:25:00 AM	5.49	16.12
9	32	3:21:00 PM	4.59	15.74
10	33	7:08:00 AM	3.80	14.70
10	34	8:32:00 AM	3.97	14.80
10	35	3:00:00 PM	3.06	14.14
10	36	3:41:00 PM	4.56	15.08
11	37	6:36:00 AM	3.48	14.53
11	38	9:38:00 AM	4.28	14.72
11	39	1:40:00 PM	3.84	14.32
12	40	6:50:00 AM	3.58	14.20
12	41	10:30:00 AM	4.55	15.16
12	42	12:20:00 PM	5.02	15.24
12	43	3:15:00 PM	4.88	15.23
12	44	4:00:00 PM	4.90	15.24
13	45	7:30:00 AM	4.51	15.27
13	46	11:50:00 AM	4.60	15.23
13	47	6:46:00 AM	5.26	15.68
13	48	11:00:00 AM	4.47	15.18
13	49	12:23:00 PM	4.20	14.59
	50	4:40:00 PM	4.45	14.93
14	51	7:20:00 AM	4.67	15.08
14	52	9:30:00 AM	4.16	14.77
14	53	10:08:00 AM	4.19	14.65
14	54	2:06:00 PM	4.94	15.28
14	55	4:00:00 PM	4.82	15.16
15	56	7:50:00 AM	4.23	14.61
15	57	9:35:00 AM	4.52	14.98
15	58	12:35:00 PM	4.28	15.27
15	59	3:15:00 PM	4.15	15.21
16	60	8:08:00 AM	4.81	15.22
16	61	10:35:00 AM	4.46	15.08
16	62	12:05:00 PM	4.74	15.71
16	63	1:51:00 PM	4.25	15.10
16	64	3:33:00 PM	3.91	14.71
17	65	7:10:00 AM	5.23	15.88
17	66	8:43:00 AM	4.76	15.58
17	67	11:15:00 AM	3.36	14.61
17	68	12:40:00 PM	3.86	15.02
17	69	2:47:00 PM	4.25	15.11
18	70	7:25:00 AM	3.59	14.33
18	71	8:42:00 AM	4.66	15.42
18	72	11:25:00 AM	5.21	15.64
19	73	7:00:00 AM	3.66	14.46
19	74	8:30:00 AM	4.23	15.05
19	75	10:45:00 AM	3.38	14.18
19	76	2:05:00 PM	3.77	14.58
20	77	3:30:00 PM	3.08	13.98
20	78	8:00:00 AM	3.19	14.17
20	79	12:20:00 PM	3.98	14.68
20	80	1:55:00 PM	3.56	14.30
21	81	6:00:00 AM	4.67	15.02
21	82	9:45:00 AM	4.25	14.61
21	83	2:20:00 PM	3.61	14.49
22	84	6:45:00 AM	5.24	15.74
22	85	8:35:00 AM	4.50	14.99
22	86	9:55:00 AM	4.19	14.71
22	87	11:35:00 AM	4.06	14.64
22	88	3:22:00 PM	3.68	14.34
23	89	7:10:00 AM	3.85	14.78
23	90	9:05:00 AM	3.81	14.55
23	91	12:20:00 PM	5.07	15.38
23	92	3:25:00 PM	4.58	15.31
23	93	4:14:00 PM	4.53	15.13
24	94	7:30:00 AM	4.49	14.93
24	95	11:30:00 AM	4.55	15.02
24	96	3:15:00 PM	4.52	15.27
24	97	2:07:00 PM	3.94	14.68
24	98	4:00:00 PM	4.40	15.12
25	99	8:45:00 AM	3.44	14.58
25	100	9:31:00 AM	3.84	14.57
25	101	11:33:00 AM	3.70	14.51

Figure 3-9 Total number of tests done for an example project (Project 1)

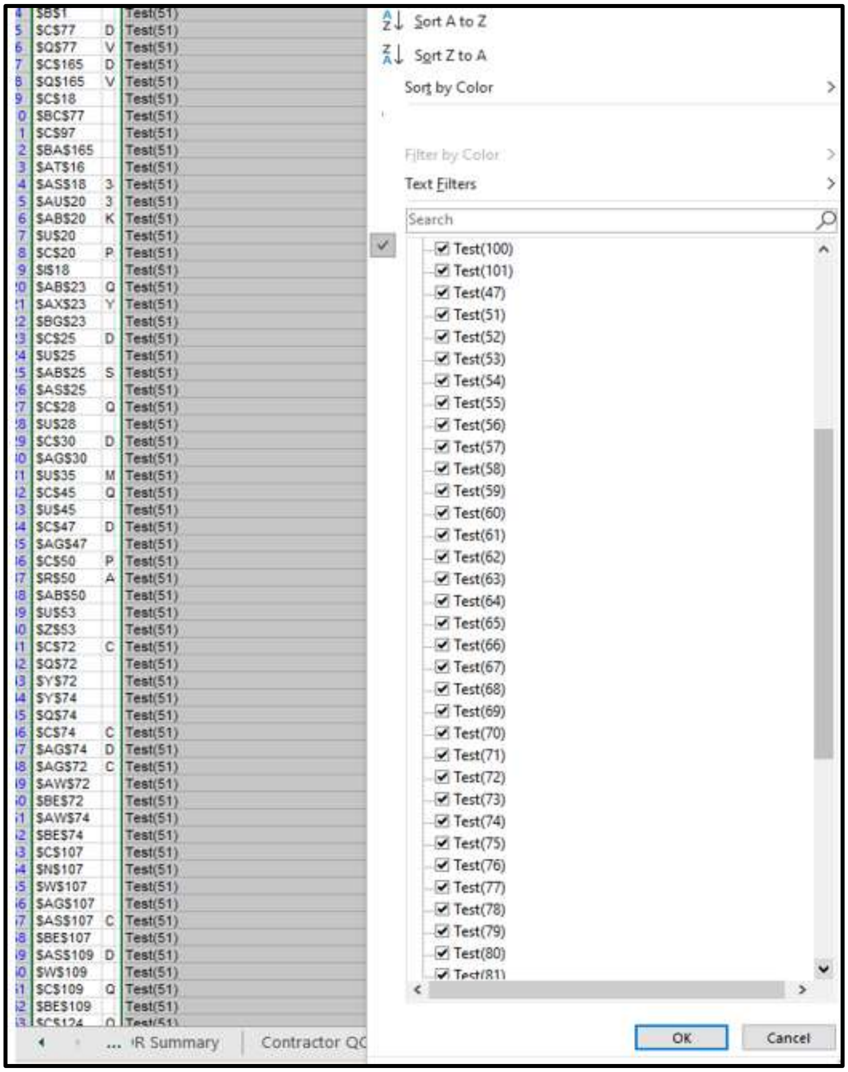


Figure 3-10 Available tests in the audit file for an example project

Project #1 shown in Fig. 3-9 has a total of 101 reported tests from the entity 1-reported data, while in the audit file we only have data for 52 tests (Fig. 3-10). All tests in audit file from Test(1) to Test (50) were missing except for Test (47). For the monetary calculations, I used the reported values for the missing tests. The reported values were exactly the same in both input datasets, so they did not induce any monetary difference. But the available tests from the audit file showed a significant difference in the monetary values (shown later).

Although I successfully filled the empty cells and missing audit values with the reported ones, I faced some issues while trying to calculate the pay factor parameters. I found negative and unreasonably large secondary parameter values based on the first S.A. primary parameters. Fig. 3-11 shows an example attempted monetary parameter calculation, where I observed large negative Air Voids values even after removing all the empty cells from the input parameter set.

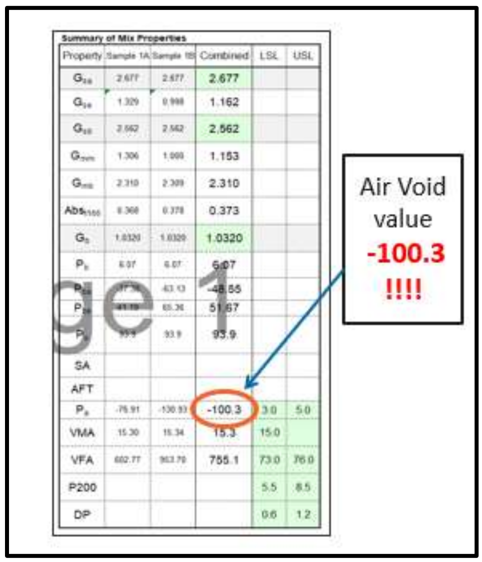


Figure 3-11 Calculated negative Air voids value

I investigated the sources of those negative and unreasonably big values by referring back to the ITD-0777 source file. It is noteworthy that it takes around 10-15 minutes to write the input values to the Excel file (done automatically with a Python code on a laptop) and generate Air Voids/VMA values for a single test. Through trial and error, I was able to discover the reasons for those unusual values, which are presented under different cases as shown below.

Case1: The first case that was borne out of my investigation was an input that was unreasonably smaller than an ideal value for a parameter (Figs. 3-12 and 3-13). Fig. 3-12 shows that mass of bowl for increment 1 has a value that was far lower than its ideal value, whereas the value for increment 2 was much closer to its ideal value. The smaller input resulted in a large negative Air voids value. Similarly, on other occasions, with lower inputs, I observed positive Air voids values, but the value was unreasonably large.

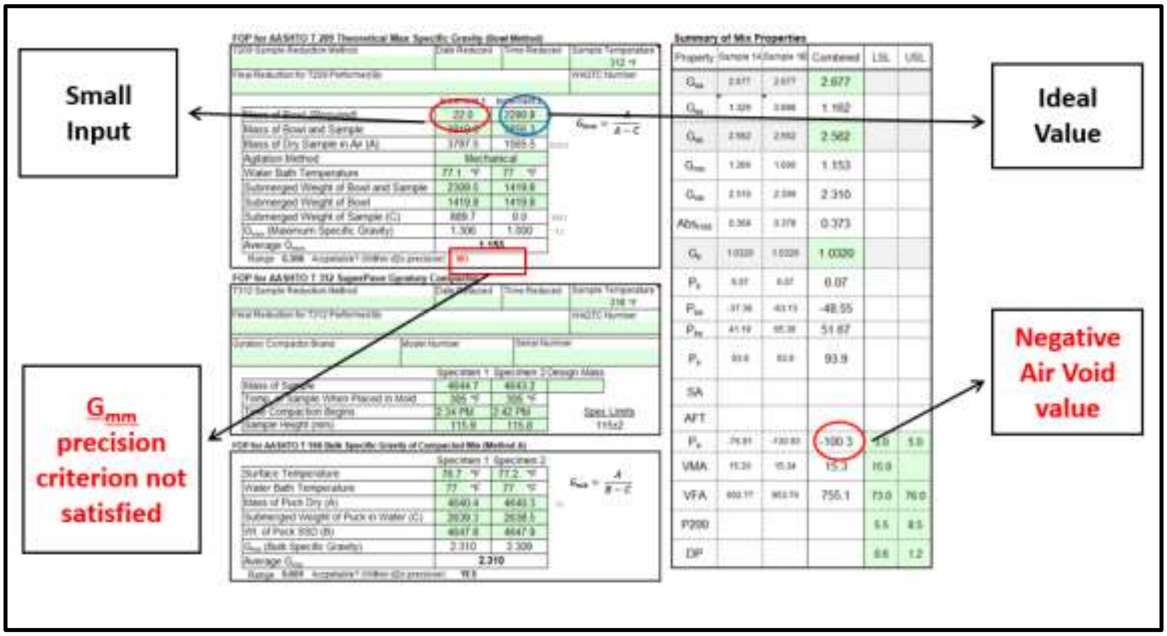


Figure 3-12 Calculated negative Air voids value due to unreasonably small primary parameter

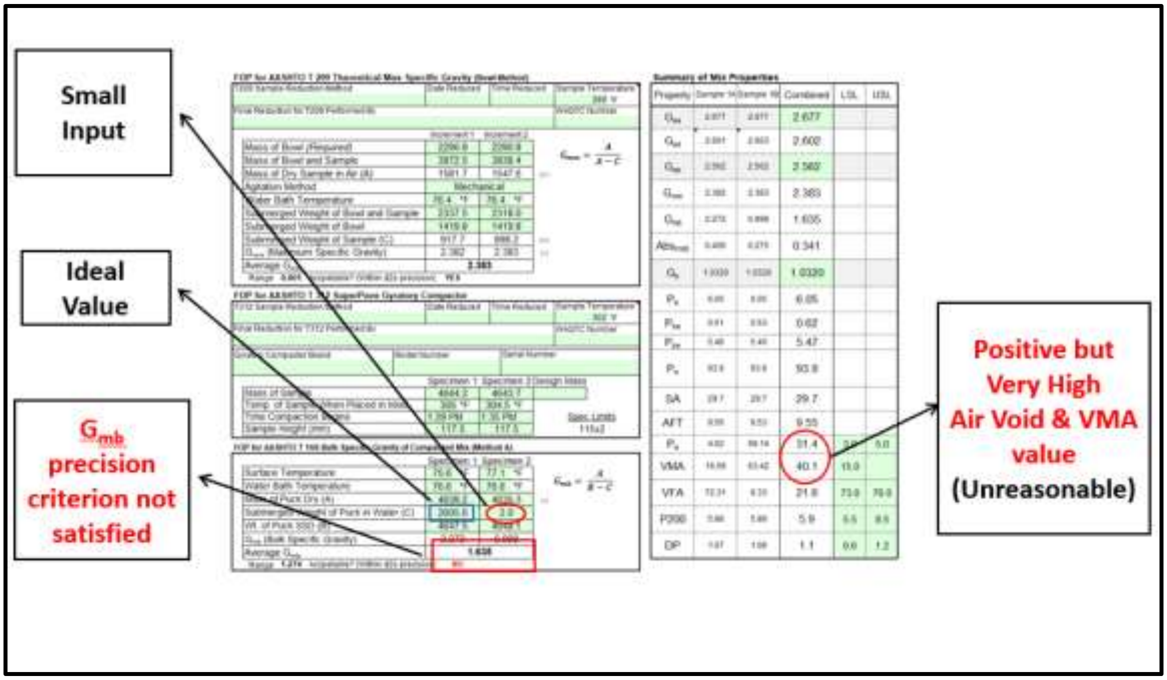


Figure 3-13 Calculated unreasonable Air voids and VMA values from unreasonably small input

Case2: In some tests, I observed unreasonably large primary parameter values producing unreasonable secondary parameters (Fig. 3-14). For example, the mass of bowl for increment 2 was 22,290, which was much higher than the ideal value (2,290). This directly affected the Air Voids calculation, which took a value that was much higher than expected. The value of 22,290 was a typing error value, which in this case, was the last typing error value. The audit file recorded this value as the final reported value, which obviously cannot be used for monetary calculation. In this case, I either adopted the previous/succeeding reasonable parameter value from the audit file, or if this was not possible (e.g. for plausible corrections), I took the final reported value for this parameter.

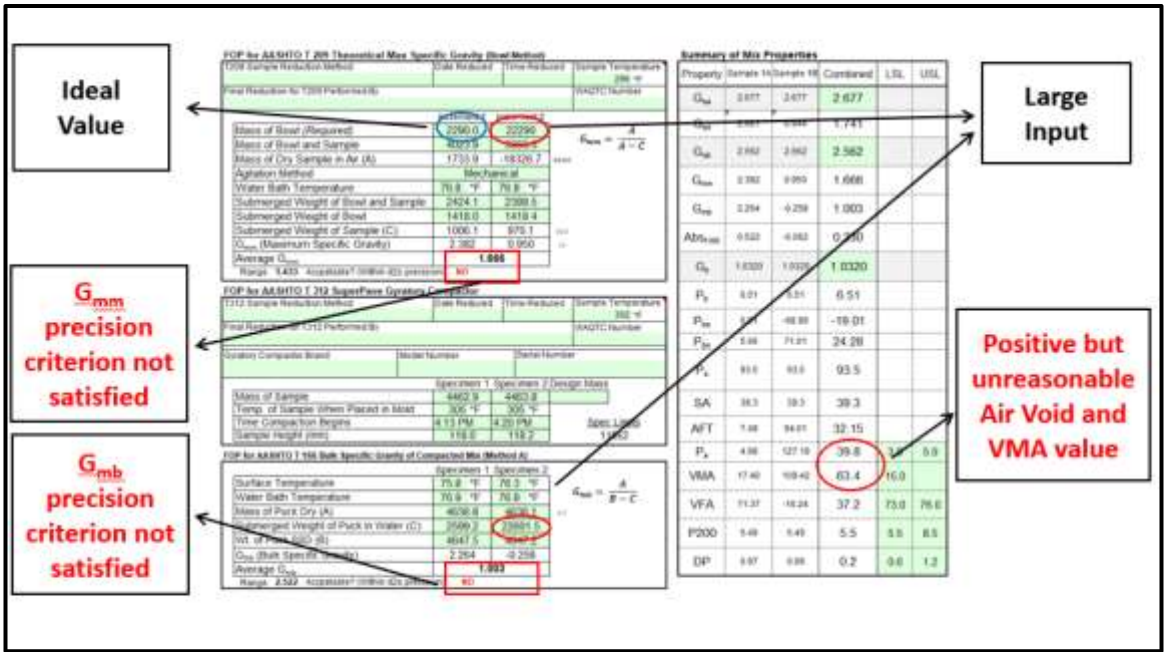


Figure 3-14 Calculated higher unreasonable Air voids and VMA value from a large input value

Case 3: Some audit values were exactly the same for multiple cells (Fig. 3-15). This was probably due to the wrong input by a data entry person. A possible explanation can be that while the operator was trying to insert the values for a cell, they probably put the value in an adjacent cell. For example, the submerged weight of bowl and sample and the submerged weight of bowl both were set as 1,367.6, which resulted in a value of 0 for the weight of sample.

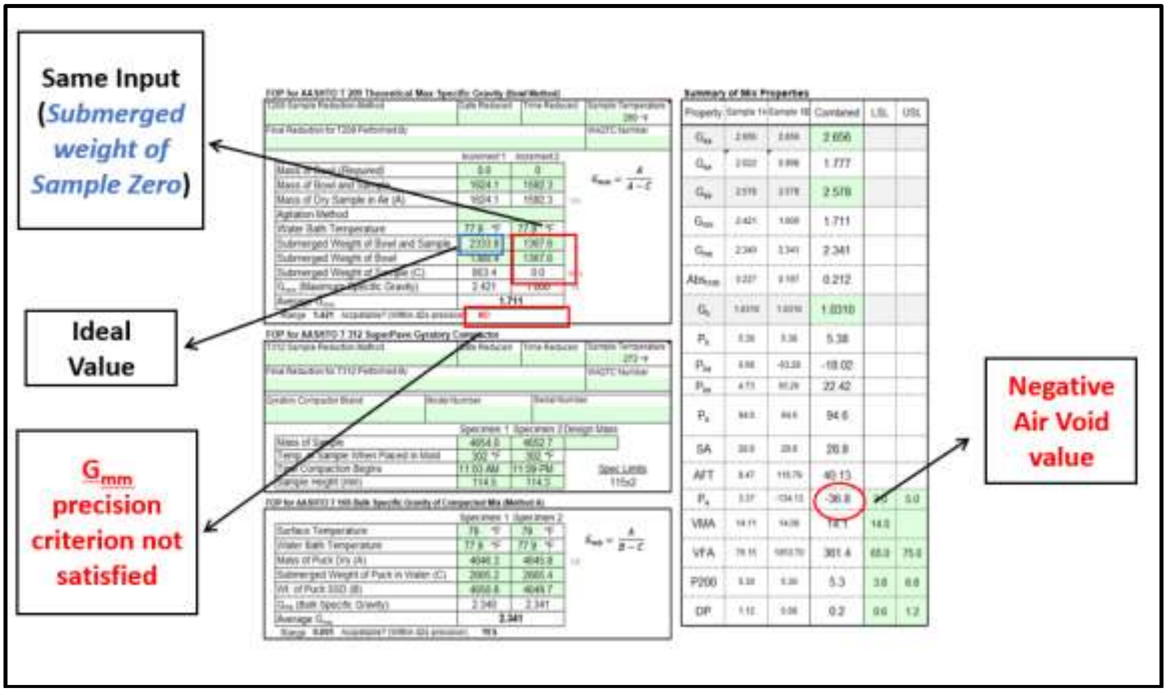


Figure 3-15 Calculated negative Air voids value due to similar values inserted for adjacent cells

Case 4: In some occasions, the later value (e.g. mass of bowl and sample) was smaller than the first value (e.g. mass of bowl), which is obviously not reasonable. Fig. 3-16 shows such an example for which a test had a mass of bowl value higher than the mass of bowl and sample, which resulted in a large negative Air Voids value.

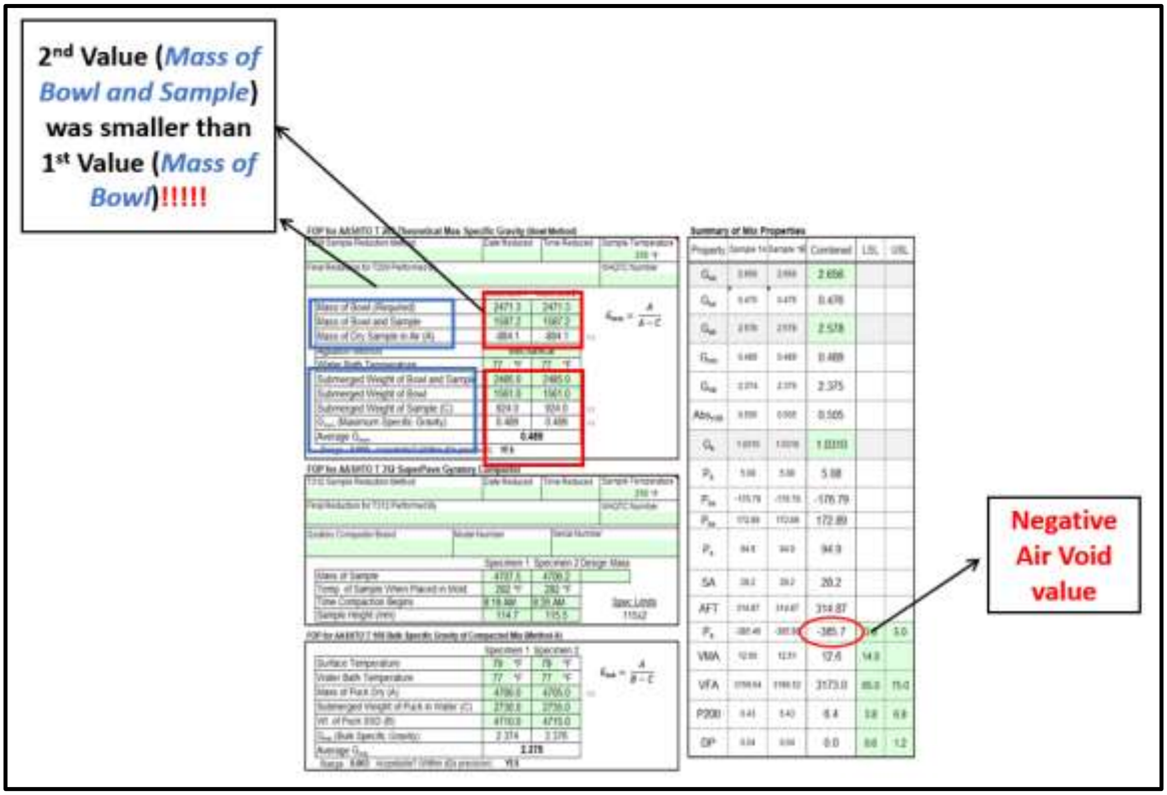


Figure 3-16 Calculated negative Air voids value due to mass of bowl and sample being less than mass of bowl

Test and Lot Information: For the purpose of calculating the monetary value as well as removing unreasonable values, we need the Test and Lot information. From the "Testing Summary" sheet of ITD-0777 file (reported material testing data), I retrieved all the Tests and Lot information about each project (Fig. 3-17).

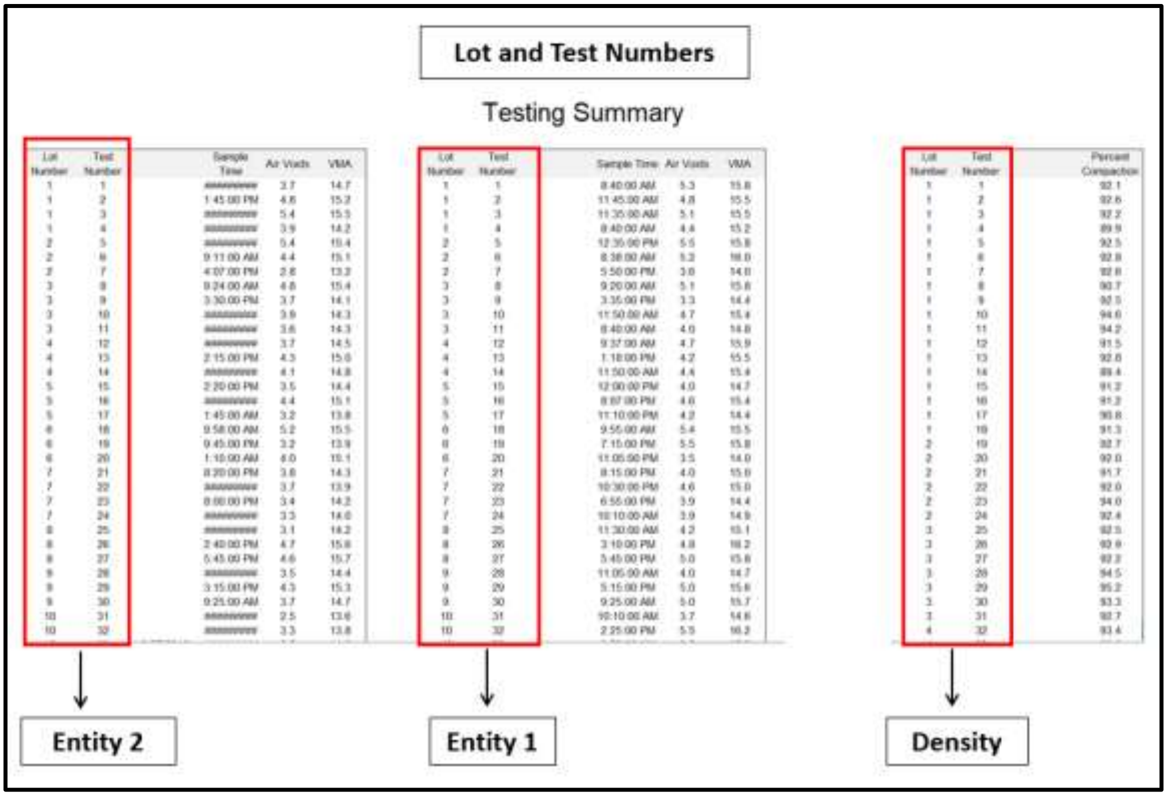


Figure 3-17 Lot and test information for a test project

Parameter Values for Missing Tests: There were several tests where we didn't have any value from the audit file. For the sake of the monetary analysis, we need values for all tests of a project. Hence, I replaced all the missing values with recorded values prior or after the missing value in the audit file, or if not available, with the final reported values. It is more often that final reported values (those that were formally used for payment calculation) were used to replace missing values.

Removing Unreasonable Parameter Values: The first and last entry for S.A. and the last entry of P.C. from audit files were unreasonable on some occasions. In order to remove them and only select reasonable values, I enforced multiple conditions through the following steps:

i. All the reported and audit values were taken for a parameter. For example, all values U32 (mass of bowl) for a project was considered as a list.

ii. Missing values from the list were removed at the first step.

iii. There were a couple of outlier values in some lists. For example, from the aforementioned case 1, a value of 22 was unreasonable for U32 (mass of bowl). This outlier value was removed using the Matlab's "rmoutlier" function. This removed any value that was outside three standard deviations range from the median.

iv. I noticed that "rmoutlier" did not remove all the unreasonable values, hence, I put a second criterion in place. If a value was greater than 1.2×mean or lower than 0.8×mean then it was removed. This threshold is set by expert opinion, and was manually checked for all tests in all projects to ensure its validity.

v. Some reasonable values, however, were removed through the process of step iv. In order to reintroduce the reasonable values to the list, the range of final reported values for each parameter was checked (Fig. 3-18). If a removed parameter value fell within this range, it was reintroduced in the final list.

Parameter	Lower Limit	Upper Limit
U32	2421.9	2422.4
U33	3928.3	4633.4
U37	2352	2980.6
U38	1263.3	1717.5
Z32	2474	2475
Z33	3978.7	4457
Z37	2471.9	3084.3
Z38	1367	1717.5
U61	4684.9	4707
U62	2730.2	2852.8
U63	4692.6	4717.1
Z61	4663.1	4712.5
Z62	2723.4	2759.1
Z63	4672.9	4720.8
S111	3006.6	3317.3
S112	4775.5	5311
S114	4549.3	5203.6
S116	0.28	0.28

Figure 3-18 Lower and upper limit value for parameters

After completing all these steps, the desired dataset (two sets of parameter values, i.e. first and last S.A. values, for all tests) was finally ready for calculating the secondary parameters (Air voids/VMA/Mainline Density) that are used for monetary analysis.

Although I removed the unreasonable values there is still the possibility of getting smaller/larger/negative secondary parameter values for first S.A. entry. This is probably another reason why the data was altered to match with the ideal ranges for Air Voids (2-4) and VMA (12-16). Figs. 3-19, 3-20, 3-21 show cases in which even seemingly reasonable values of primary parameters resulted in secondary parameter values that do not fall in the acceptable range.

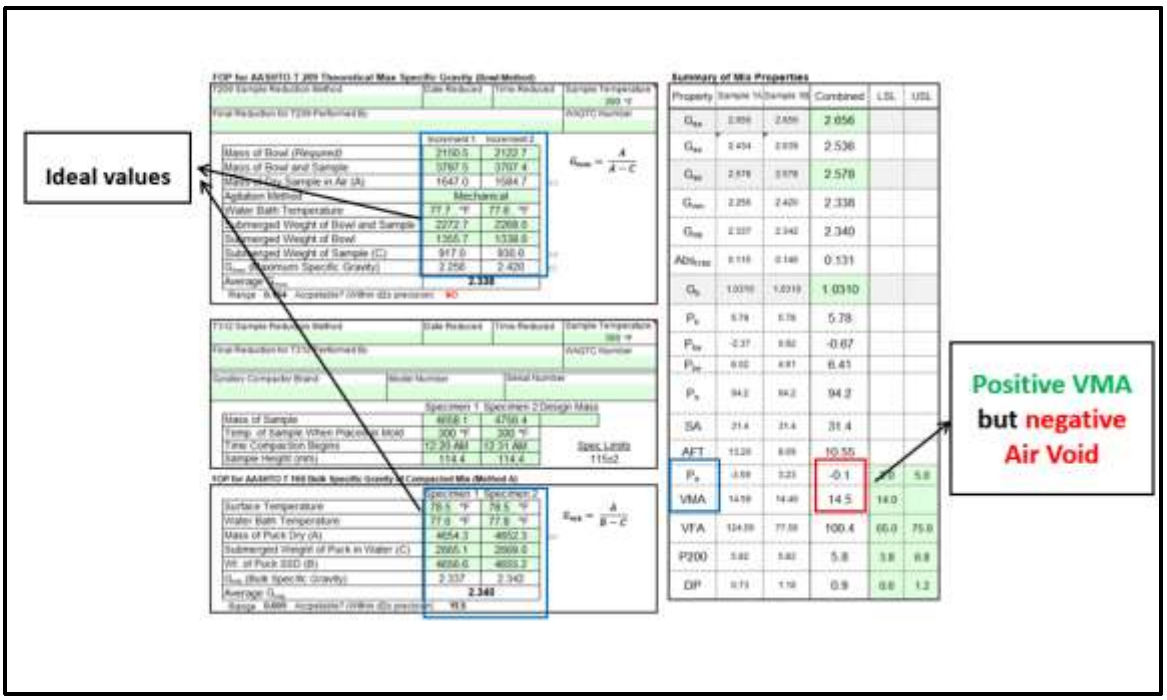


Figure 3-19 Unreasonable calculated Air voids and VMA with reasonable primary parameter values

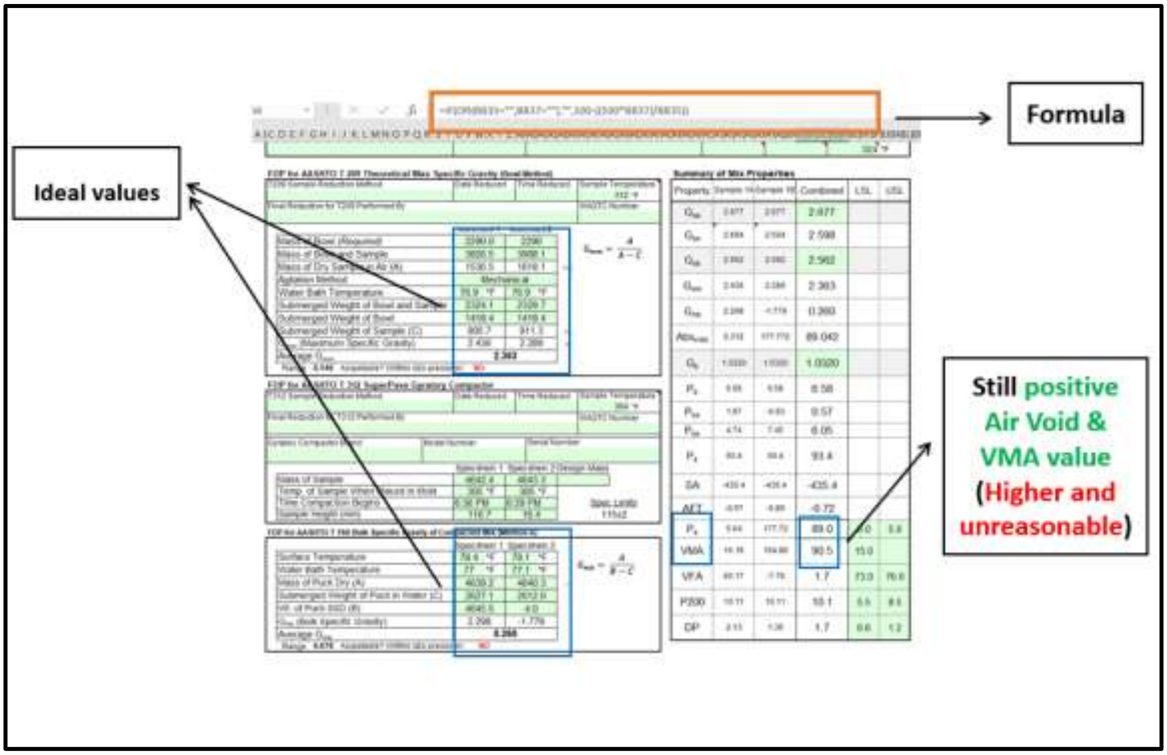


Figure 3-20 Unreasonable calculated Air voids and VMA with reasonable primary parameter values

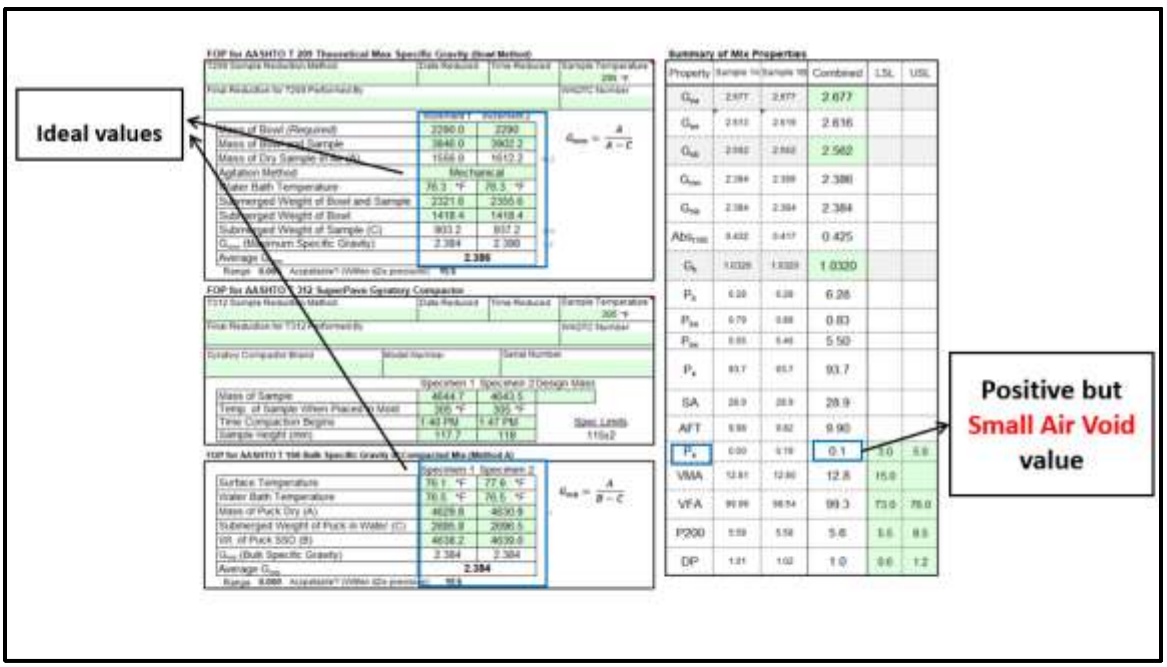


Figure 3-21 Unreasonable calculated Air voids and VMA with reasonable primary parameter values

Lot Grouping: Based on the calculated Air Voids/VMA/Mainline Density parameter values, several lot groupings, that were originally used for monetary calculations, should have been changed, and many tests should have been rejected in the first place (Fig. 3-22). However, it's not possible during my analysis steps to ask for a redo of the tests in the field and recalculate the secondary parameters, so I considered the lot grouping as reported.

				AirVoids	VMA
Lot Range	2 - 3				
		X_c Avg		4.222799966	14.76476385
		X_o Avg		3.551989801	14.1973697
		S_c^2		0.225765469	0.168285178
		S_o^2		0.352827358	0.080193902
		F-Statistic		1.562804795	2.09847849
		F-Critical		5.285236852	6.853075629
		P-Value		0.56988216	0.431684528
		Alpha		0.05	0.05
		Pass/Fail		Pass	Pass
				0	0
		t-Statistic		2.352781401	2.89631451
		t-Critical		2.17881283	2.17881283
		P-Value		0.036527557	0.013417847
		Alpha		0.05	0.05
		Pass/Fail		Fail	Fail
Lot	3	Verified?	No	1	1
3 Not Verified					
Entity 1					

Figure 3-22 An example case of lot calculated parameters failing the statistical tests

Results of Monetary Analysis

The final payment-related parameter values were calculated for all tests of each project and all projects, which are presented here. Detailed results and plots for project #1 are described in this section, and summary results for all projects are presented in a Table format.

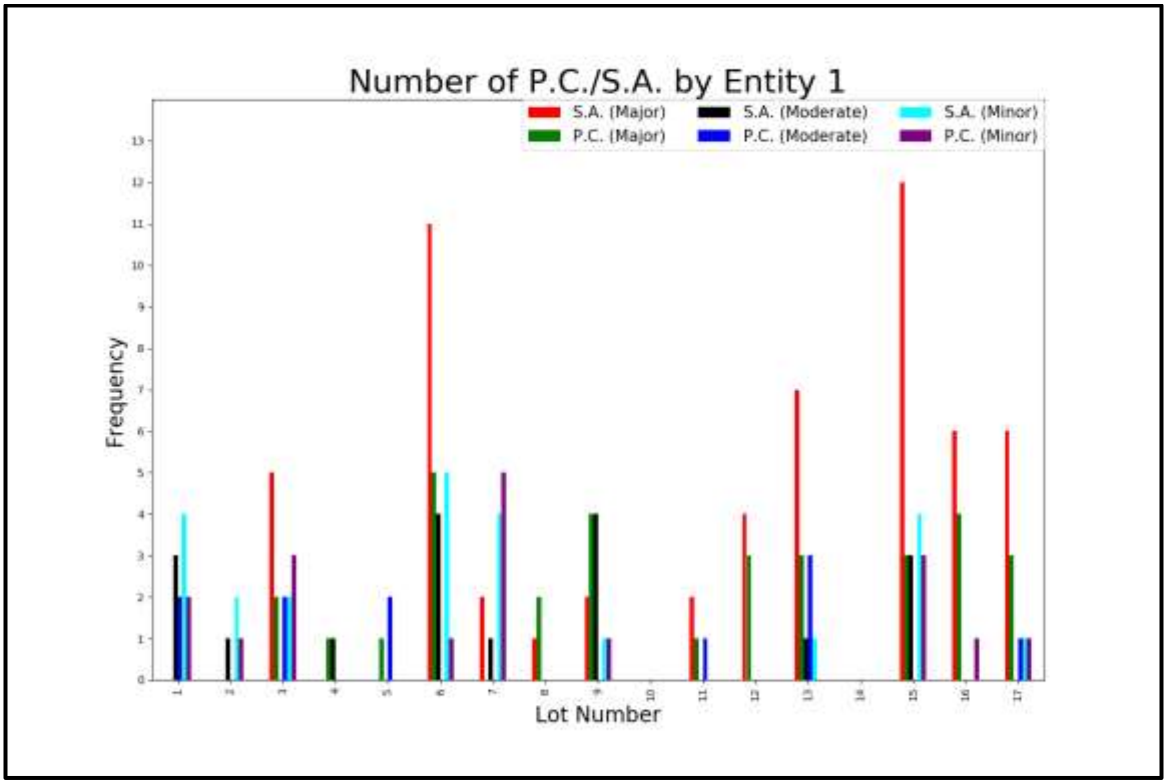


Figure 3-23 Number of unique P.C./S.A. parameter changes for each lot and each parameter type for the entity 1-reported data for project #1

Fig. 3-23 presents the number of unique cells that were changed in each lot for project #1. The graph shows data for three separate categories of major/moderate/minor parameters for both P.C./S.A. instances. Lot 3, for example, has 5 instances of S.A. and 2 instances of P.C. for major parameters. This graph presents the unique number of cells/parameters that were affected, not the number of times these cells were changed. The total number of times these cells were changed was much higher because each cell was changed multiple times.

I observed the maximum number of S.A. for major parameters in lot 15 (Fig. 3-23). It will be shown later that frequency of S.A. parameters does not necessarily have a monotonic relationship with payment, rather changes might be due to a variety of reasons including passing Percent Within Limits (PWL) or precision criteria.

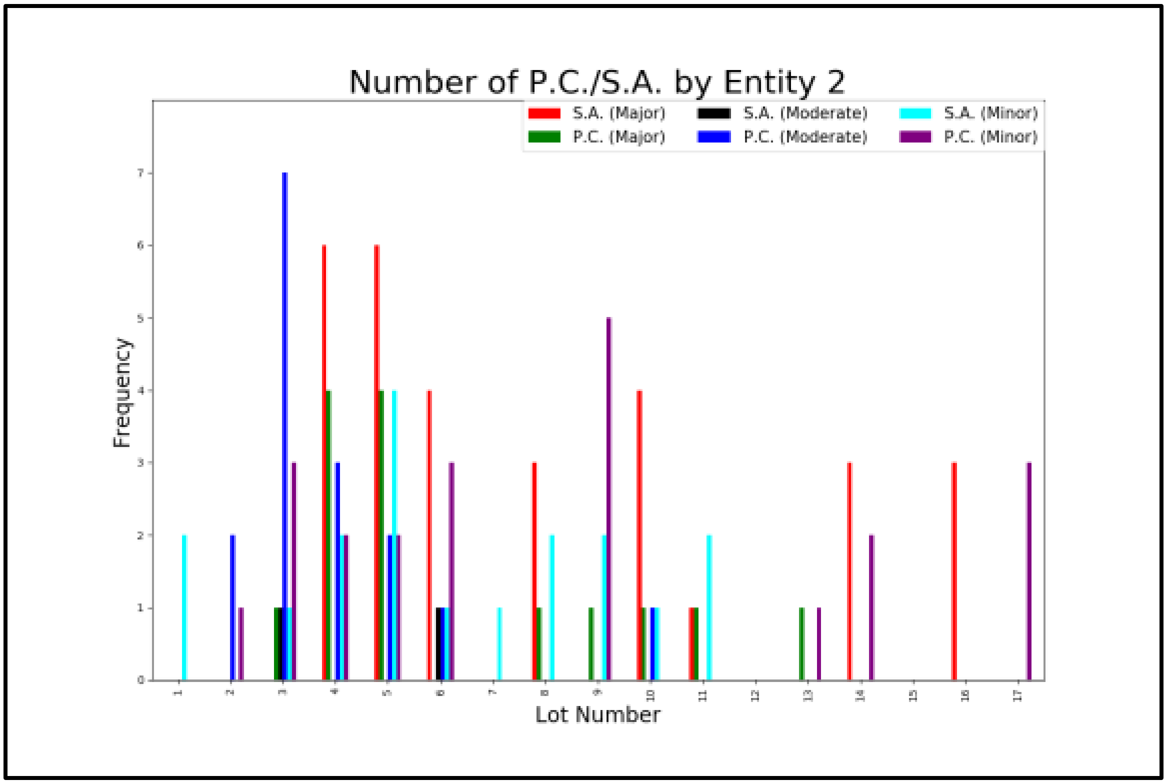


Figure 3-24 Number of unique P.C/S.A. parameters for entity 2 tests for project #1

I did not observe any direct relationship between the number of P.C. or S.A. changes in entity 1- reported versus the entity 2-reported data. Both datasets are prone to having multiple parameter value changes.

Before performing the monetary analysis, these primary parameters are checked for precision level in G_{mm} (Theoretical maximum specific gravity), G_{mb} (Bulk specific gravity), and P_b (Asphalt binder content, percent by total mass of mixture) parameters. One of the precision checks is shown in the Fig. 3-25, where G_{mm} precision didn't pass (results as No) for this example test. For project #1, I presented the precision results for each test both on the entity 1 and entity 2 data in Fig. 3-26 (green: pas – red: fail). Multiple tests didn't pass the precision test.

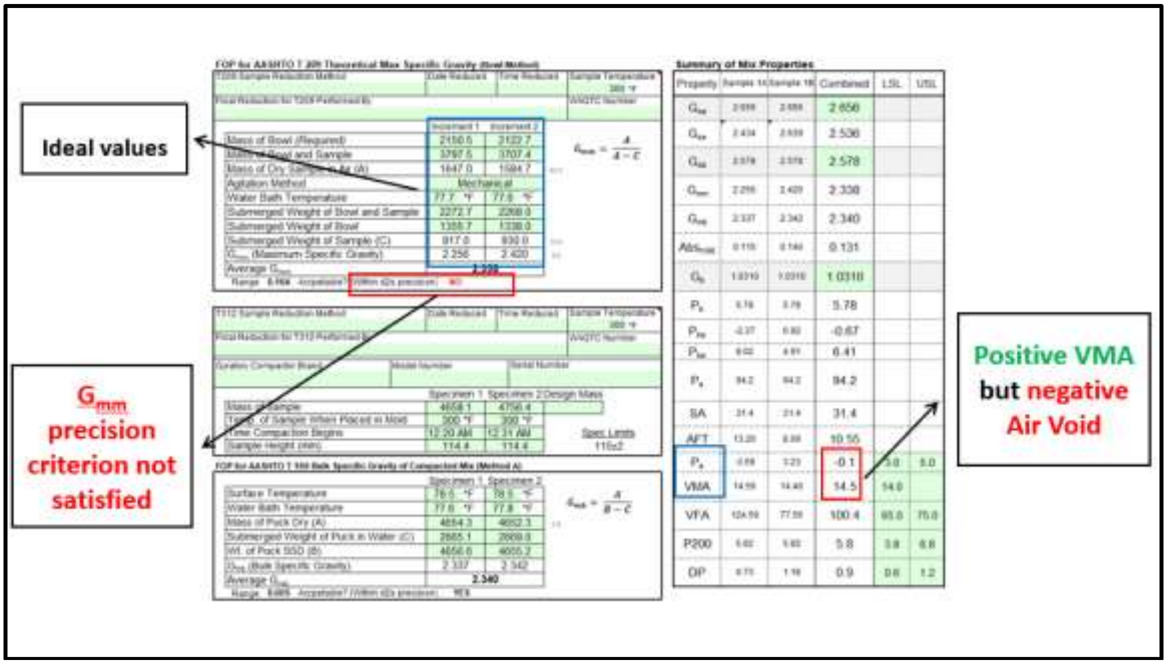


Figure 3-25 Precision criterion not satisfied for an example project

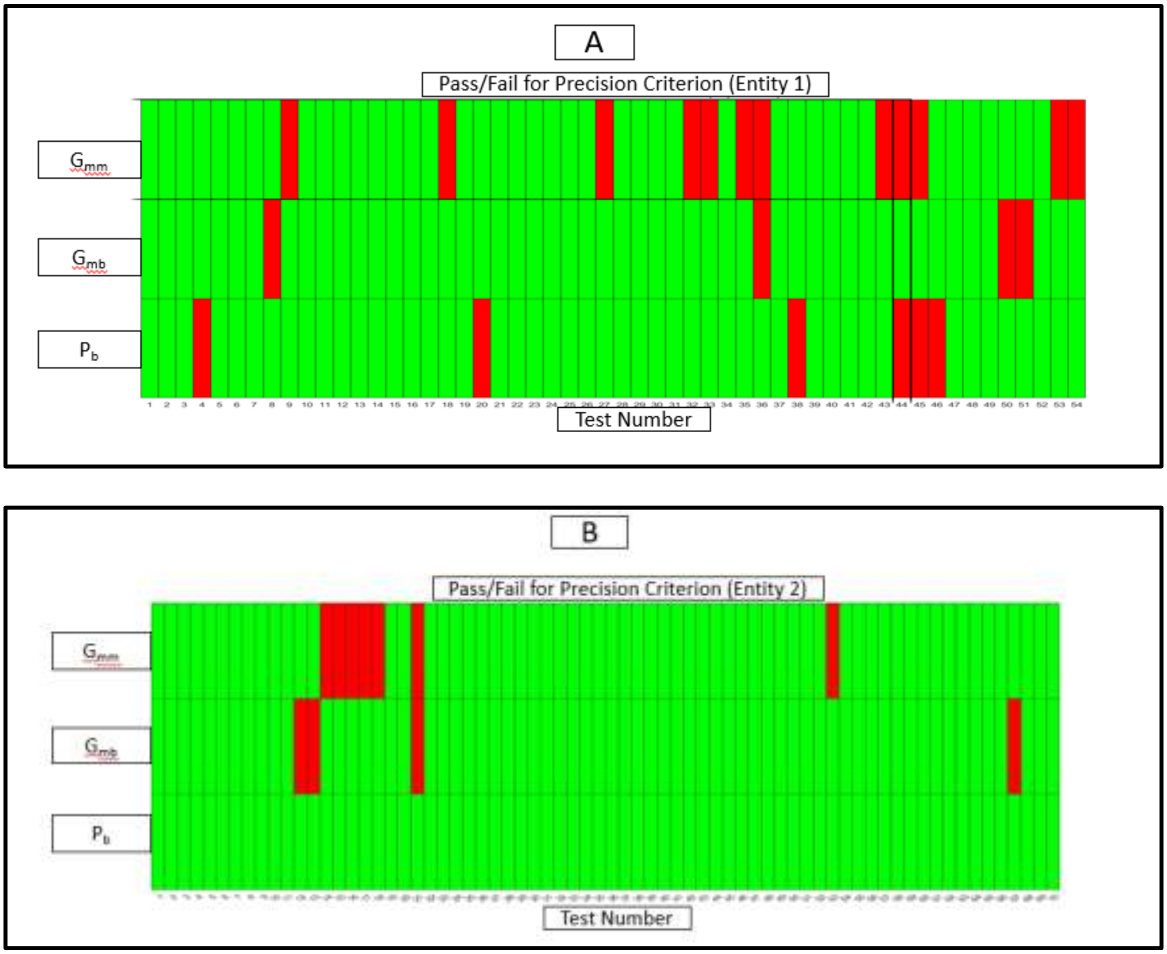


Figure 3-26 Precision criterion for each test of project #1 (upper entity 1, lower entity 2). Green shows pass and red represents fail.

<u>Acceptance Check</u>

Monetary analysis starts with two statistical tests (F and T tests) to determine whether entity 1-reported data should be used, or the entity 2-reported data is to be used. Then the selected data goes through the "quality level analysis" for Air Voids/VMA/Mainline Density which subsequently determines whether or not the lot is at an acceptable level. Fig 3-27 shows an example graph with Accept (green)/Reject (red)/Stop Production (black) levels for Percent Within Limits (PWL) for Air Voids, VMA and Mainline Density for project #1. These checks were done for the first S.A. entry cases to see if the first value was considered for payment, how many lots should have been rejected. This analysis indicates that even before considering payment, several lots might have been rejected straight away. Usually, for the three payment factor related parameters, this acceptability check is done with the following generic value check.

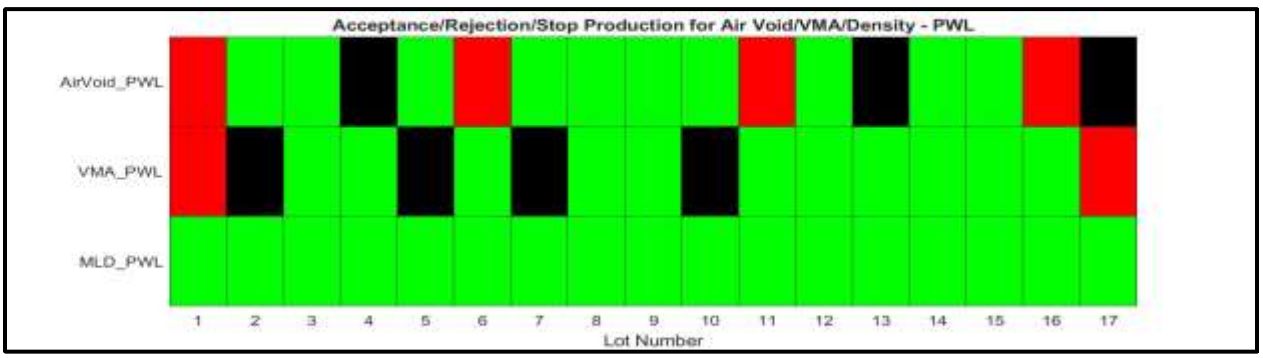
$$PWL_{Air\ Void/\ VMA/\ Density} > 60 = \ Acceptable$$

$$PWL_{Air\ Void/\ VMA/\ Density} > 40 = \ Stop\ Production, Action\ Needed$$

$$PWL_{Air\ Void/\ VMA/\ Density} < 40 = \ Reject\ Level$$

Lot Acceptance Status		
AirVoids		Acceptable
VMA		Acceptable
	0	Reject Level
	0	
	0	
MLD		Acceptable

Figure 3-27 Acceptance check for payment related parameters



-- Acceptable Level, -- Reject Level, -- Stop Production,

Figure 3-28 Lot-wise Acceptance/Rejection/Stop production according to Percent Within Limit (PWL) for project #1

Fig. 3-28 shows that multiple lots might have been rejected based on the PWL check. The first row presents results for Air Voids, the second row is for VMA, and the last row is for Mainline Density. Five lots out of the total 17 got rejected in the parameter's quality level analysis check. Further, only 6 lots out of the 17 were at an acceptable level.

I now focus on the monetary analysis of data alterations, based on the first and last acceptable entry for S.A. cases. In Fig. 3-29, the Green bars show calculated monetary value for the first acceptable S.A. parameter values. As discussed earlier, for the unchanged parameter values (no alteration) and for P.C. cases, the reported value and last P.C. value were selected for monetary analysis, respectively. The red bar shows calculated payment based on the last entry for S.A. parameters. Yellow bars present the original reported payment. These payment levels are calculated for each lot separately.

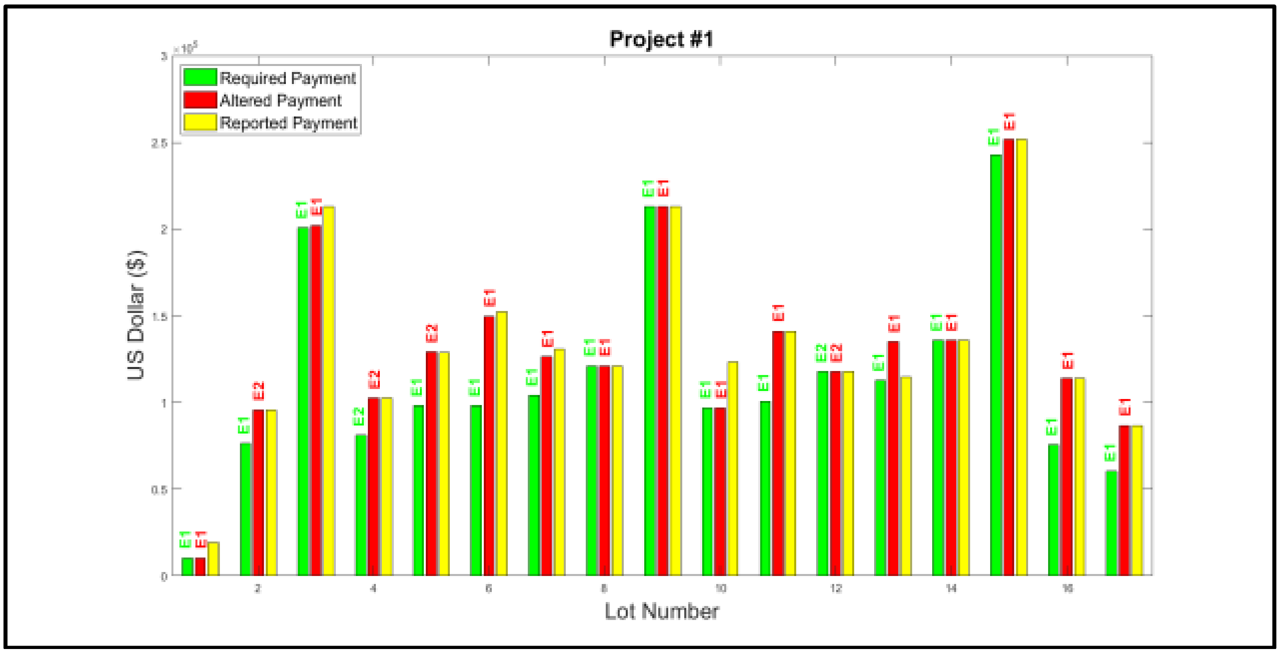


Figure 3-29 Lot-wise payment for project #1. Green bars show payment based on the first S.A. parameter values, red bars present payment based on last S.A. parameter values, and yellow bars show the actual payment formally made.

There were some lots for which my final calculated value didn't match the reported formal value from the projects. There are two reasons for this observation:

i. Some of the lots had "dispute resolution" status, which was resolved by collecting data by a third party. However, we didn't have any audit data from the third party. So, my calculated value was different from the originally reported payments.

ii. As discussed earlier, audit files did not necessarily record all data entry, meaning that final reported parameter value might not be included in the audit file. I observed some instances that the last value recorded in the audit file was not equal to the reported one. Because of the irregularity of the data in the audit file for some lots, my calculations did not match the exact reported value in a few cases.

Bars in Fig. 3-29 are labeled as E1 and E2, which represent Entity 1 and Entity 2, respectively. This shows which reported data was chosen for payment analysis based on the F and T tests. For lot 2, for example, if the initially reported values were considered, Entity 1-reported data should have been used for payment, whereas due to alteration, entity 2 data were used for payment. This resulted in an overpayment of around 20,000 dollars (+20%) for this lot. It is evident in Fig. 3-29 that for several lots payment should have been less if the initial entry value for parameters was chosen for payment analysis.

There were originally about 30 projects obtained from ITD that had some sort of audit file included. Out of the 30 projects, however, 18 either were missing audit files or reported values were unavailable. I hence focused on the 12 projects for which I could calculate payments. In the rest of this chapter, I will present all results for these projects.

Table 3-1 shows cumulative monetary value based on the first and last S.A. parameter values and also the final/formal reported payment. This table includes all the available number of audit tests from entity 1 and entity 2 as well as the cumulative monetary values for the projects. In most projects, there was a significant amount of overpayment.

Table 3-1 Calculated payments for first and last S.A. parameter values, and the formally paid amount for each project. Table also enlists statistics of total number of lots and available audit files from the entity 1 and entity 2 reports

Project Number	Total Lot	Total Test (E2)	Available Audit (E2)	Total Test (E1)	Available Audit (E1)	Total Sheet (Dens)	Available Audit (Dens)	Payment (First S.A.)	Payment (Last S.A.)	Formal Payment (Reported with Dispute Resolution)
Project 1	17	70	70	54	54	21	21	$1,945,217	$2,228,807	$2,260,795
Project 2	14	67	67	67	67	15	14	$2,492,391	$2,853,563	$3,215,331
Project 3	5	16	No Data	16	16	12	5	$568,890	$583,246	$579,831
Project 4	25	101	51	101	52	27	13	$3,962,182	$4,082,441	$4,217,759
Project 5	50	241	84	150	12	57	5	$9,860,811	$9,906,251	$9,897,883
Project 6	21	79	No Data	74	50	50	33	$1,976,327	$2,030,917	$2,040,929
Project 7	5	16	16	16	16	10	10	$762,583	$989,563	$989,797
Project 8	4	14	14	13	11	8	6	$586,866	$709,034	$709,243
Project 9	3	9	9	9	9	3	3	$195,573	$192,578	$212,967
Project 10	11	51	No Data	33	4	13	3	$1,756,489	$1,878,476	$1,952,210
Project 11	13	25	25	42	42	13	13	$1,117,583	$1,142,740	$1,525,770
Project 12	17	54	14	51	11	19	6	$1,907,322	$1,906,912	$2,306,717

Table 3-2 summarizes all Percent Within Limit (PWL) results for all projects. This table provides details about the number of lots in each project, number of lots for which audit files were available, and number of lots for which audit files are available for both entity 1 and entity 2. These tables further enlists the number of lots that might have been rejected (at least based on one parameter, i.e. Air Voids, VMA, or Mainline Density), accepted or was at stop production level. Projects #8 and #9 did not have even a single lot that was accepted (Table 3-2), whereas project #5 had the highest fraction of accepted lot (90% of all lots). On average, 8-50% of the lots should have been stopped and reformed the lot/redid the test, which indicates a considerable proportion of the lots would have been rejected.

Table 3-2 Summary of acceptance/rejection and stop production for PWL analysis for each project

Project	Total Lot	Available No of Lot (E1)	Missed Lot (E1)	Available No of Lot (E2)	Missed Lot (E2)	Total Accepted	Total Rejected	Total Stop Production, Action Needed	Available No of Lots (Mutual between E1 & E2)	Accepted (Mutual between E1 & E2)	Rejected (Mutual between E1 & E2	Stop Production, (Mutual between E1 & E2)
Project 1	17	15	10, 14	15	12, 15	6 (35%)	5 (29%)	7 (41%)	13	3 (23%)	5 (38%)	6 (46%)
Project 2	14	14		14		6 (43%)	7 (50%)	2 (14%)	14	6 (43%)	7 (50%)	2 (14%)
Project 3	5	4	1	No data	No data	4 (80%)	0 (0%)	1 (20%)	0	0 (0%)	0 (0%)	0 (0%)
Project 4	25	12	2 to 13, 21	14 to25	1 to 13	20 (80%)	2(8%)	3 (12%)	11	9 (82%)	1 (9%)	1 (9%)
Project 5	50	5	2 to 46	34 to 50	1 to 33	45 (90%)	1 (2%)	4 (8%)	4	3 (75%)	1 (25%)	0 (0%)

Project 6	21	12	1 to 7, 11, 19	No data	No data	16 (76%)	2 (10%)	3 (14%)	0	0 (0%)	0 (0%)	0 (0%)
Project 7	5	5	0	5	0	1 (20%)	2 (40%)	2 (40%)	5	1 (20%)	2 (40%)	2 (40%)
Project 8	4	4	0	4	0	0 (0%)	3 (75%)	2 (50%)	4	0 (0%)	3 (75%)	2 (50%)
Project 9	3	3	0	2	1	0 (0%)	3 (100%)	1 (33%)	2	0 (0%)	2 (100%)	1 (50%)
Project 10	11	1	1 to 10	No data	No data	2 (18%)	2 (18%)	1 (9%)	0	0 (0%)	0 (0%)	0 (0%)
Project 11	13	10	4, 5, 9	8	9 to 13	3 (23%)	9 (69%)	3 (23%)	6	1 (17%)	4 (67%)	2 (33%)
Project 12	17	4	5 to 17	5	6 to 17	7 (41%)	6 (35%)	6 (35%)	4	0 (0%)	4 (100%)	2 (50%)

Table 3-3 Summary of payment change, and number of unique S.A. parameters involved for each project

Project Number	Total Lot	Total payment change ($) (first and last S.A.)	Total major S.A. unique parameters	Total payment change per unique major S.A. parameter ($/parameter)	Total S.A. unique parameters	Total payment change per unique S.A. parameter ($/parameter)
Project 1	17	$283,590	60	4,727	103	2,753
Project 2	14	$361,172	94	3,842	138	2,617
Project 3	5	$14,356	0		0	
Project 4	25	$120,258	38	3,165	64	1,879
Project 5	50	$45,440	33	1,377	45	1,010
Project 6	21	$54,590	20	2,729	22	2,481
Project 7	5	$226,980	47	4,829	66	3,439
Project 8	4	$122,168	36	3,394	45	2,715
Project 9	3	$-2,995	1	-2,995	5	-599
Project 10	11	$121,987	7	17,427	9	13,554
Project 11	13	$25,158	14	1,797	23	1,094
Project 12	17	$-409	7	-58	10	-41

Table 3-3 summarizes the calculated overpayment for each project, as well as the average extra payment per unique parameter changed. In this table, the total major S.A. unique parameters and total S.A. unique parameters represent either entity 1 or entity 2 based on which of them were selected. For example, on lot 1 of a project, either the entity 1 or 2 is selected for payment based on the statistical test results. If entity 1 is selected, then major S.A. unique and total S.A. unique for entity 1 is considered. Similarly, for lot 2 based on statistical test results if entity 2 is selected, then the major S.A. unique and total S.A. is considered for entity 2. So, the final value of major S.A. unique and total S.A. represented in this table is a summation of all the lot of a project from either entity 1 or entity 2 based on which of them were selected on each individual lot. The maximum amount of extra payment was seen on project #2, where more than $361,000 were overpaid. In this project, 94 major and a total of 138 parameters were altered. The high number of alterations resulted in a massive monetary change in this project. A majority of the analyzed projects had a significant amount of overpayment. For some projects (9 and 12) I saw a reduction in payment, although the sheer value of reduction is minimal. It is also noteworthy that there were also some lots in different projects for which detected S.A. values resulted in minor decrease in payment, but for the entire project, the summation of all lots resulted in over-payment. It is also interesting to observe in this table that each S.A. parameter change resulted in roughly $1,000-$5,000 extra payment in each project. The audit files did not necessarily capture all changes in reported parameter values, and I expect if those are factored in, the change in payment can be even higher.

Relationship between S.A. Instances and Payment

An essential question is whether or not data alteration always translated into financial impacts. The answer is "No". Although my main objective was to capture the economic repercussions of data alterations on the projects, I observed that they did not necessarily translate into monetary changes all the time. Through in-depth analysis, I investigated the potential reasons for this observation. An overall comparison of the monetary-related parameter (Air Voids/VMA/Mainline Density) values from the primary parameters for first S.A. parameter entry and final reported parameter is shown in Fig. 3-30. The upper part (green) and lower part (red) of Fig. 3-30 present all the test values for first S.A. entry and final reported entry for a particular lot in project #7. Looking closely, most of the test values are different between the two cases (green versus red).

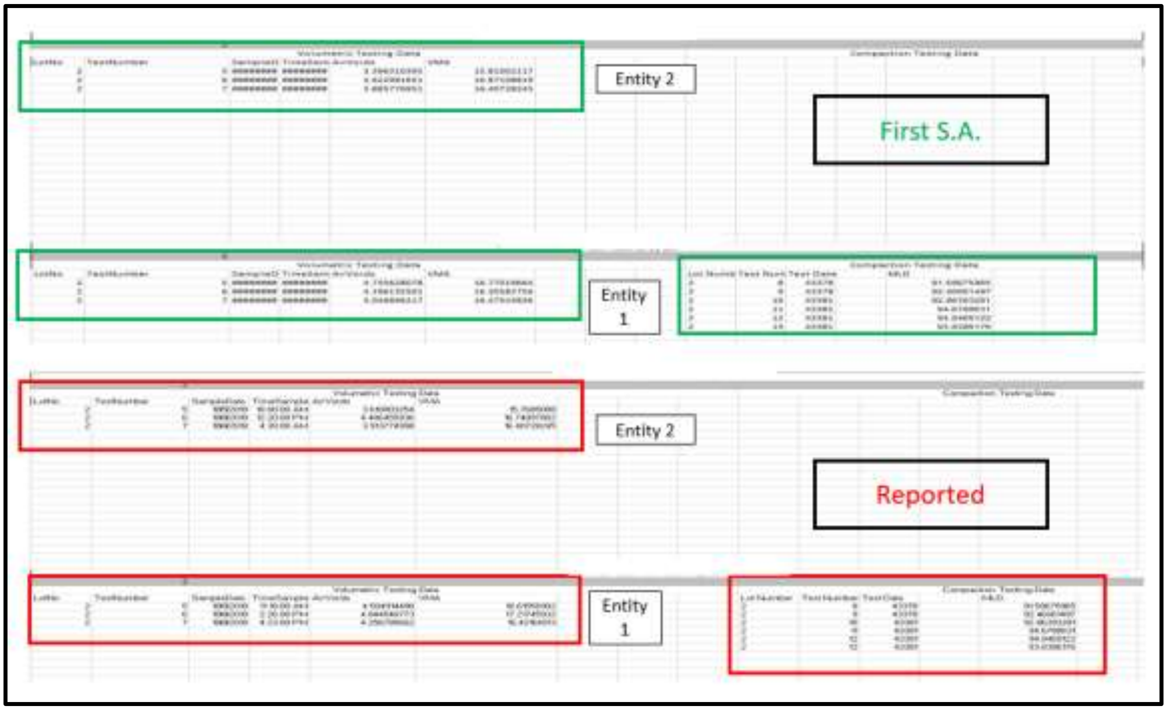


Figure 3-30 Lot-wise calculated Air voids/VMA/Mainline Density parameters based on first S.A. and final reported parameter values (project #7)

But this is not all that we need for monetary calculation. The next step is to form the lot groups. This particular lot was grouped with its previous lot (Fig. 3-31). Like the individual group, this lot group also had evidence of changed value for most of the tests.

Figure 3-31 Formation of lot group (project #7)

As described earlier, the second step of the financial analysis is to check the acceptability of the entity 1/entity 2 data through the F and T tests (Fig. 3-32).

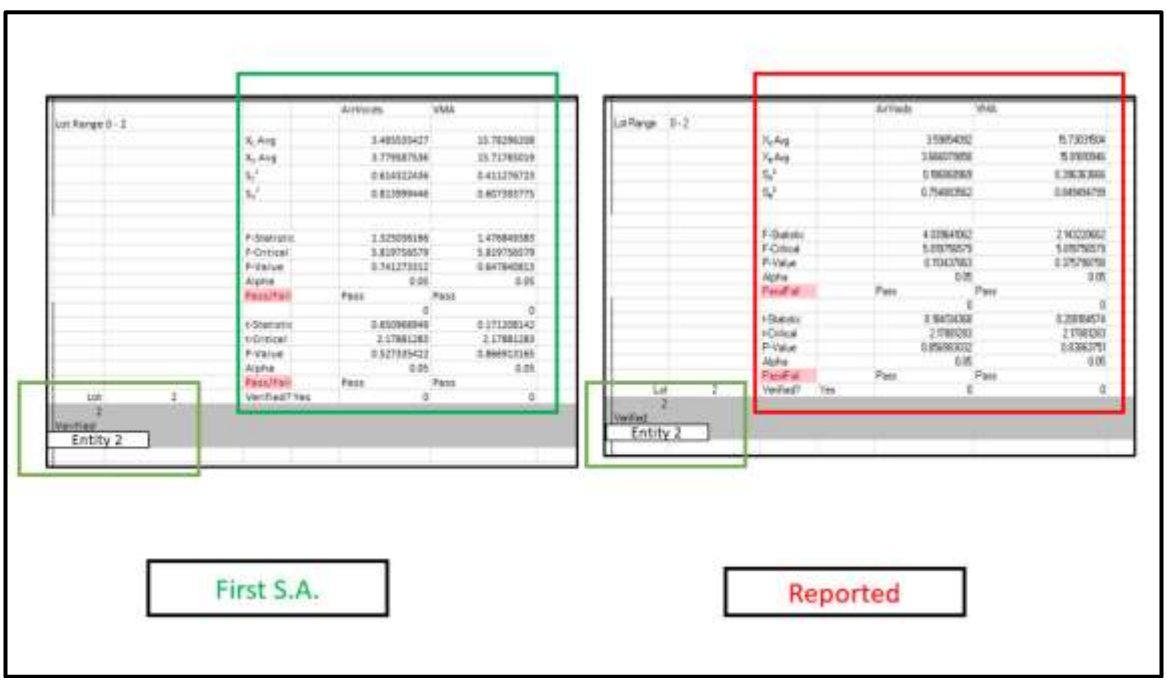


Figure 3-32 Selection of entity 1/entity 2 test result based on F and T tests (project #7)

It is evident that although there were clear data alteration instances, based on F and T tests, entity 2 data were used for payment calculations. No matter how many times the entity 1 data was changed, it doesn't go into the payment calculation steps.

The next target was to calculate the Unweighted Pay Factor. In this step, the average value of Air Voids, VMA, and Mainline Density is used. This average value, often, can compensate for the test value change, hence not resulting in payment change. Some reported test values were lower than the first S.A. instances, and some were higher. Since a mean value is taken, we often had a very close overall value from both calculations. For instance, the average Air Voids value was 3.96 from my first S.A. calculation, whereas it was 3.97 in the reported section. Similarly, the average VMA value came up as 16.36 from my first S.A. calculation, and it was reported as 16.30. Despite all the clear alterations done on the earlier steps, averaged monetary-related parameters can take values very close to the original values (Fig. 3-33).

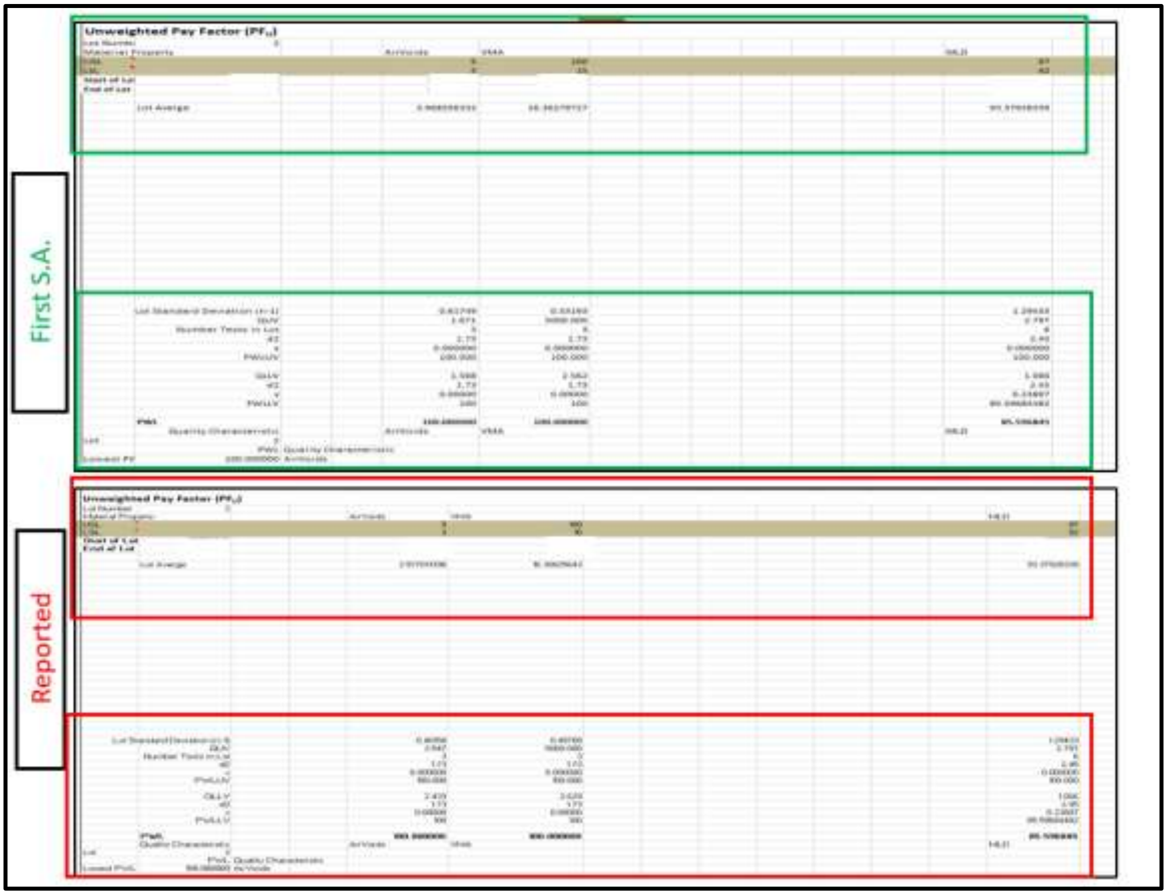


Figure 3-33 Calculation of unweighted pay factor (project #7)

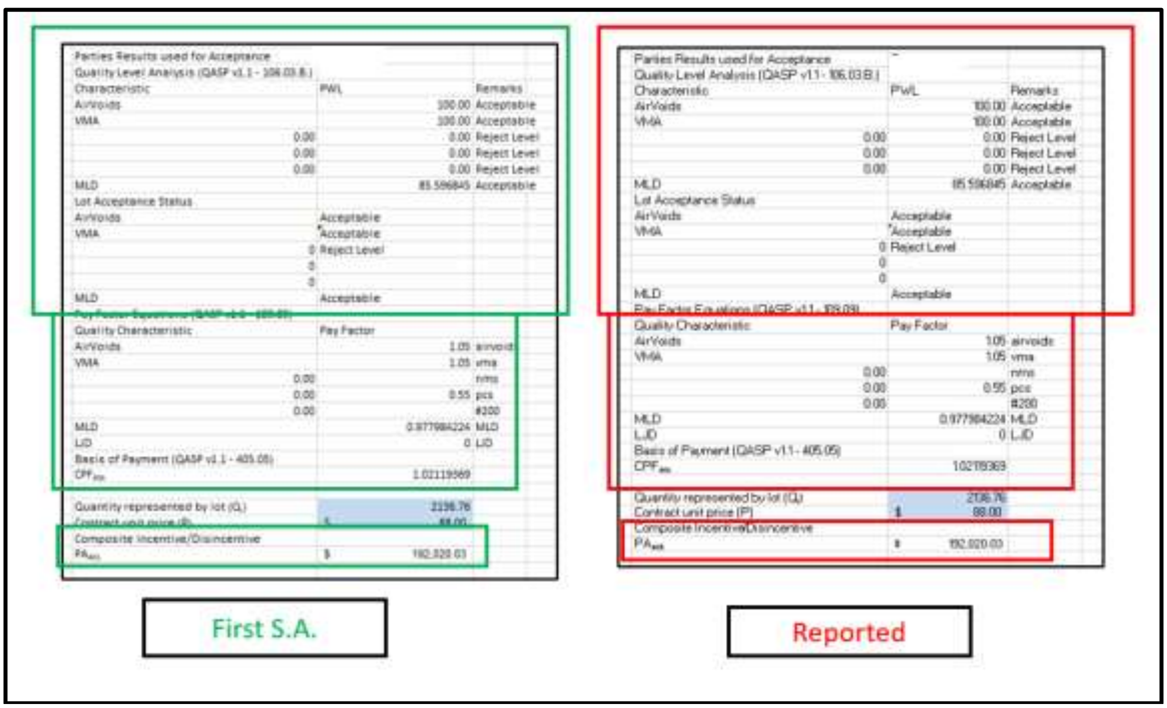


Figure 3-34 Calculation of PWL and monetary value (project #7)

The last step is to determine the Percent within Limit (PWL) value and calculate the monetary values (Fig. 3-34). We can see in Figs. 3-32 & 3-33 that because the average value of the secondary parameters was almost equal; the PWL value came precisely the same for these specific tests. The end result was, hence, an identical payment value for both scenarios. I argue that for some cases no matter how many times data alteration has been done, there might still be zero payment impact. Obviously, this does not apply to all projects and tests. As shown in Tables 3-1 and 3-2, data alteration has often resulted in overpayment to entity 2.

Conclusion

Construction projects are generally performed in a complex dynamic environment and are highly sensitive to data alteration and suspicious activities. Failure to take adequate measures to protect these sensitive tasks against corruption results in higher costs and time overruns in construction projects. This research leverages the availability of a unique audit dataset (recording sequence of all entered parameter values in a material testing form) to calculate monetary impacts of potential suspicious alteration of material testing reports. Such claim of data alteration upholds the necessity for reformation of traditional QC/QA practice which seems to be vulnerable to suspicious intentional or unintentional digitalized data error and can cause loss in monetary values. I have successfully replicated the monetary payment calculation procedures followed by Idaho Transportation Department and calculated lot-wise payments for various lots of 12 Hot Mix Asphalt projects prior to and after data alterations. Majority of the projects prompted overpayment, even with the conservative approach that was taken for monetary calculations. Further, a great majority of the analyzed lots did not pass the Percent Within Limit thresholds.

References

Akkinepally, R., and Attoh-okine, N. (2006). Quality control and quality assurance of hot mix asphalt construction in Delaware. University of Delaware.

Bose, I., and Mahapatra, R. K. (2001). Business data mining - A machine learning perspective. Information and Management, 39(3), 211–225.

Bowen, P., Akintoye, A., Pearl, R., and Edwards, P. J. (2007). Ethical behaviour in the South African construction industry. Construction Management and Economics, 25(6), 631–648.

Butts, N. E., and Ksaibati, K. (2002). Asphalt Pavement Quality Control/Quality Assurance Programs in the United States. Paper No. 0123, presented at the 82nd Transportation Research Board Annual Meeting,Washington, DC.

Coenen, A. R., Pforr, J. E., Hefel, S. A., and Paye, B. C. (2019). State DOT Implementation of Statistical Analysis and Percent Within Limits. Transportation Research Record, 2673(2), 583–592.

Coutorie, L.E. (1995). The future of high-technology crime: A parallel Delphi study. Journal of Criminal Justice, 23(1), pp.13-27.

Frawley, W. J., Piatetsky-Shapiro, G., and Matheus, C. J. (1992). Knowledge discovery in databases: An overview. AI Magazine, 13(3), 57–70.

Gunduz, M., and Önder, O. (2013). Corruption and Internal Fraud in the Turkish Construction Industry. Science and Engineering Ethics, 19(2), 505–528.

Hardoon, D., and Heinrich, F. (2011). Bribe payers index. Transparency International. https://www.transparency.org/en/publications/bribe-payers-index-2011

Kenny, C. (2006). Measuring and Reducing the Impact of Corruption in Infrastructure. World Bank Policy Research Working Paper 4099.

Kenny, C. (2009). Transport construction, corruption and developing countries. Transport Reviews, 29(1), 21–41.

Kyriacou, A. P., Muinelo-Gallo, L., and Roca-Sagalés, O. (2015). Construction corrupts: empirical evidence from a panel of 42 countries. Public Choice, 165(1–2), 123–145.

Le, Y., Shan, M., Chan, A. P. C., and Hu, Y. (2014). Investigating the causal relationships between causes of and vulnerabilities to corruption in the Chinese public construction sector. Journal of Construction Engineering and Management, 140(9), 1–12.

Locatelli, G., Mariani, G., Sainati, T., and Greco, M. (2017). Corruption in public projects and megaprojects: There is an elephant in the room! International Journal of Project Management, 35(3), 252–268.

Ngai, E. W. T., Hu, Y., Wong, Y. H., Chen, Y., and Sun, X. (2011). The application of data mining techniques in financial fraud detection: A classification framework and an academic review of literature. Decision Support Systems, 50(3), 559–569.

Perols, J. (2011). Financial statement fraud detection: An analysis of statistical and machine learning algorithms. Auditing: A Journal of Practice & Theory, 30(2), 19–50.

Rodriguez, D., Waite, G., & Wolfe, T. (2005). Global corruption report 2005. Transportation International. https://images.transparencycdn.org/images/2005_GCR_Construction_EN.pdf

Salem, M. S. M. (2012). An overview of research on auditor's responsibility to detect fraud on financial statements. Journal of Global Business Management, 8(2), 218-229.

Sohail, M., and Cavill, S. (2008). Accountability to prevent corruption in construction projects. Journal of Construction Engineering and Management, 134(9), 729–738.

Turban, E., Aronson, J. E., Liang, T. P., and Sharda, R. (2007). Decision Support and Business Intelligence Systems. Eighth Edition, Pearson Education.

CHAPTER 4: SUMMARY, COMCLUSIONS AND RECOMMENDATIONS FOR FUTURE RESEARCH

Summary and Conclusion

This thesis emphasized two specific applied science problems. The first problem discussed in Chapter 2 featured the significance of suspicious activities in Hot Mix Asphalt (HMA) construction projects. The objective was to devise a model to display suspicious activity detection strategies to government agencies and prove the necessity of reforming the traditional QC/QA practice. Such data alteration can occur from simple human mistake or intentional instances. In the age of data science and big data, corruption is considered encyclopedic and it actively challenges modern society in every aspect. A modern data-centric optimized solution is required for such problems, which encouraged us to take the machine learning route in my research.

Chapter three of this thesis was focused on quantifying the monetary losses due to Suspicious Alteration attempts summarized from Chapter 2. In this section, I show that in almost all of the analyzed projects, altered data resulted in an overpayment.

Major findings from this research include:

i. There was evidence of data alteration both in the digital format (Excel sheets) and manual entries (paper-based data reporting).

ii. A total of 7 Plausible Correction and 4 Suspicions Alteration cases were identified from the audit datasets.

iii. Out of the three payment affecting categories (major/minor/moderate) defined by the Idaho Transportation Department for Hot Mix Asphalt (HMA) pavement parameters; major parameters observed most data alterations, and the number of alterations was significantly higher compared to the other two categories.

iv. Supervised machine learning algorithms, like K-nearest neighbor, logistic regression, support vector machine, and discriminant analysis, exhibited good performances in categorizing Plausible Correction (P.C.) and Suspicious Alteration (S.A.) cases. The high accuracy score of these models supports my logic-based categorization of P.C. and S.A. cases.

v. HMA testing parameters are run through a series of equations to calculate lot-wise payment for each project. If the first suspicious alteration was considered almost half of the lots couldn't pass the precision check. Further, only about 1/3 of the lots – with available audit data – would have passed percent-within-limit thresholds.

vi. Majority of the projects had a significant amount of overpayment ranging from \$14,000 to more than \$360,000. Major unique parameter changes were also higher on projects where the overpayment was higher.

vii. On some projects (2 out of 12) there was a minor reduction (-\$400 to -\$2,500) in payment if the first S.A. parameter values were considered.

viii. Data alterations didn't always result in a change in monetary value. There were multiple occasions where data was altered but no monetary change was observed, but did result in a change in pass/fail of percent-within-limit thresholds.

ix. I considered the same lot formation values available from the reported files. However, if the first S.A. cases were considered, a lot of tests would fail which would have required a new lot formation. Since, all these projects were already completed, and test redone and lot reformation is not possible, I considered the reported lots. The lot reformation could have resulted in more overpayment than shown in my results.

Recommendations for Future Research

i. I considered multiple cases of P.C. and S.A. from the digitized files. However, paper-based data alteration cases were not considered in my analysis. If there are enough paper-based data alteration attempts available, such cases should be included in the algorithm. This would ensure a more robust approach in detecting data alteration attempts in HMA construction projects.

ii. Rigorous training of field engineers and technicians (from both contractor and agency side) involved in HMA production, quality control, and acceptance testing. Emphasis should be on the importance of test accuracy and repeatability, and how they affect the end product

iii. Extensive review of agency-adopted specifications related to HMA mix design and construction. Special care should be taken to ensure the specifications and tolerances are developed based on materials commonly used in the region. Setting “unreasonable” targets for material quality will ultimately lead to undesirable practices and inferior pavement performance.

iv. I have applied supervised machine learning technique in my analysis. Due to limited size of the available dataset, I did not try unsupervised machine learning techniques. If a similar larger dataset is available from ITD or other transportation/government agencies, unsupervised Machine Learning (ML) techniques can also be applied.

v. I couldn't find a significant relationship between time of data entry and S.A. cases. A research path can be to implement ML techniques to discover the relationship between time stamp and probable S.A. attempts.

vi. The lot reformation was not possible in my analysis. If there is another way of lot reformation after the project has been completed another approach of payment calculation can be done.