ACKNOWLEDGEMENT

We truly acknowledge the cooperation and help made by Ms. Tripti Rathee , Assistant Professor Department of Information Technology, Maharaja Surajmal Institute of Technology, C-4 Janakpuri , Delhi , Guru Gobind Singh Indraprastha University. She has been a constant source of guidance throughout the course of this project. We are also thankful to our friends and family whose silent support led us to complete our project.

CERTIFICATE

This is to certify that the project entitled " Analysing Factors Affecting Delays in Road Construction" is a bonafide work carried out by Shrishti Soni , Ayush Aggarwal , Palak Singhal and Komal Singh under my guidance and supervision and submitted in partial fulfilment of the requirement for the award of the degree of Bachelor of Technology in Information Technology of Guru Gobind Singh Indraprastha University, Delhi. The work embodied in this project has not been submitted for any other degree or diploma.

Ms. Tripti Rathee

Assistant Professor

Department of Information Technology

MSIT, Delhi

Mr Manoj Malik

Head , Department of Information Technology

MSIT, Delhi

ABSTRACT

Construction projects in India are experiencing widespread delays. Due to a dramatic shift in the capacity and volume of the Indian construction sector over the last decade, the need of a systematic analysis of the reasons of delays and developing a clear understanding among the industry professionals are highly crucial. Using a selected set of 45 attributes, this research first identified the key factors impacting delay in Indian construction industry and then established the relationship between the critical attributes for developing prediction models for assessing the impacts of these factors on delay. A questionnaire and personal interviews have formed the basis of this research. Factor analysis and regression modelling were used to examine the significance of the delay factors.

Regression model indicates slow decision from owner, poor labour productivity, architects reluctance for change and rework due to mistakes in construction are the reasons that affect the overall delay of the project significantly.

These findings are expected to be significant contributions to Indian construction industry in controlling the time overruns in construction contracts.

Table of Contents

CHAPTER 1 : INTRODUCTION

1 Introduction

The Indian construction sector has acted as an engine of growth for the Indian economy for over the past five-decades and becoming a basic input for the socio-economic development of the country. Construction is the second largest economic activity after agriculture, and has contributed around 6 to 9% of India's GDP over the past five years while registering 8 to 10% growth per annum. The investments made in construction were reported to be close to USD 50 billion in 2008 with persistent growth pattern expected for much of the next decade. Contribution of the industry in terms of employment is also significant providing 31.46 million jobs; with about 1.25 million engineering jobs in 2008–2009. As per government data, the demand for construction manpower is projected to grow at a consistent pace of 8%–9%, thereby resulting in an annual addition of around 2.5 million jobs to the existing stock with approximately 125,000 new engineering jobs being added annually. Regardless of the economic importance and employment generation of the sector, issues such as low productivity, limited mechanisation and lack of professionally qualified employees plague the industry. While the importance of Indian construction sector over the past five years has grown significantly, lack of sophistication across the construction supply chain is one of the key issues in the industry. There is strong evidence of inconsistent performance of Indian construction projects and the trend is growing rapidly. Projects are reportedly failing across all the key performance measures including cost, time and quality performances. While understanding of the intrinsic factors affecting all these key performance measures is still an area of investigation at least in the Indian context, this research focuses on the analysis of performance in terms of timely delivery of construction projects. By earlier estimate, over 40% projects have reportedly been suffering from poor performance across the country [1]. In a separate study comparing the performance of international development projects [2] reported that construction projects in India showed the worst schedule performance [2]. The study found that in India average schedule overrun is the highest (55% of actual schedule) compared to the other nations. Construction projects, especially infrastructure projects, in India have come under tremendous international scrutiny in the wake of the recent 2010 Commonwealth Games Hindustan Times.The current status report published by the Ministry of Statistics and Programme

Implementation (MOSPI) highlighted that out of the 951 projects being monitored 309 projects have cost overruns and 474 projects are behind schedule. MOSPI has reported that "Of the total reported cost increase of USD 12.4 billion, USD 8.4 billion is on 466 delayed projects" (www.mospi.nic.in)[4]. Reasons for these problems range from land acquisition, improper planning and budgeting, to poor coordination and monitoring of the projects.

With this growing volume, schedule performance of the Indian construction sector is a certainly significant topic for investigation. While many studies have published on causes and factors affecting schedule and cost performance, most of the studies are area specific Kim et al.. Applicability of such research in Indian construction context still remains unexplored. There is a strong need to understand the attributes that cause delays, understand the impact of these attributes, combine them into factors, and decipher the interdependencies between these factors. Thus, the primary objectives of this research are to identify the various attributes for construction delay, to identify the relationship between these attributes by statistical methods and to predict the impact of these identified attributes on construction delay using a regression model in the Indian construction sector.

CHAPTER 2 : LITERATURE REVIEW

2 Literature review

Delay in construction projects has been a research topic for decades. Research conducted in this area is broadly divided into two streams—one stream relating to attributes and factors that cause project delays and second stream relating to delay analysis. Some location specific work related to delay analysis reported [5], [6] and [1] highlighted the complexity on this issue across many countries. The first stream of literature focusing upon delay attributes and factors which is more relevant to this research is reviewed below.

A Kazaz[7] reported the causes of delay on Turkish public-sector construction projects in the 1970s and 1980s by surveying public agencies and contractors involved in public sector projects. This study divided the identified factors into those that are influenced by national economic policies and those that can be controlled by the public agencies and contractors . Investigating the factors causing delay in construction projects in United Arab Emirates, [8] reported that over 50% of construction projects experience delay due to factors such as delay in approval of construction drawings, poor pre-planning and slow decision making process. Comparing the key factors of construction delay across UAE, the Kingdom of Saudi Arabia (KSA) and Lebanon, the research asserted that delay in approval, owner's slow decision making and material shortages are common causes of construction delay across the region. However, the findings that other high ranked factors in UAE had no significant impact in KSA construction projects clearly highlight the fact that factors causing construction delay cannot be considered common across the countries. There is a clear need for critical analysis and validation of the factors in Indian context as well.

A survey to identify project delays in Saudi Arabia was conducted [9] reporting lack of agreement between project stakeholders in such identification. [10] repeated this study in Saudi Arabia to highlight the chronic nature of the problem and disparity in the views of the project stakeholders. [11] provided time performance guideline for Singaporean contractors working in India.

[5] identified main causes of delays in Egyptian construction projects and concluded that different parties of construction don't agree on the relative importance of various factors of delay, mostly blaming each other of delays using importance index and spearman rank correlation similar to [12]. He also identified the importance of team effort in the success of a project.

In Nigeria,[13] identified 43 factors of delay under 9 categories based on the works of [14] and [15]. Based on covariance analysis and Pareto analysis it was found out that 88% of the factors contribute to 90% of delays and thereby concluded that there is no discernable difference among the different delay factors and none of them really stands out as a largest contributor to the problem.

In a separate study, 16] identified that the most common cause out of the listed 73 causes of delay identified by all parties of construction is change of orders using frequency index (FI) and severity index (SI).

In India, 1] identified the project success and failure attributes and their latent property failure attributes being: conflict among project participants, ignorance and lack of knowledge, presence of poor project specific attributes and non-existence of cooperation, hostile socio economic and climatic condition, reluctance in timely decision, aggressive competition at tender stage, short bid preparation time.

[15] reviewed the causes of delays in housing projects and identified main categories as: client-, consultant-, and contractor-caused delays, and extraneous factors in Nigeria. The research asserted that client-caused delays predominately arise from design variation in projects.[17] reviewed the causes of delays and cost overruns and found that there was a very good agreement between the respondents on those factors that could cause delays and cost overrun. The four most important items agreed on by the contractor, consultants, and public clients surveyed were the financing of and payment for completed works, poor contract management, change in site conditions, and shortages of materials.

From the above selected literature review, it has been apparent that in most studies, priority has been given to identifying the critical causes based on perceptions of different parties in construction. However, quantification of the dependencies of one factor over others has not found widespread coverage. For instance steps taken to control a critical reason might trigger a situation where other factor becomes critical and cause even more delay than earlier anticipated. Hence it is important to identify the relationship between various factors of delay. Work is yet to be done in identifying the relationship between the various reasons of delay and also prediction of impact on that delay.

2.1 Research and Analysis Flow

The research flow is mentioned in below flow chart for sake of simplicity which ends with the final conclusion of the entire research.

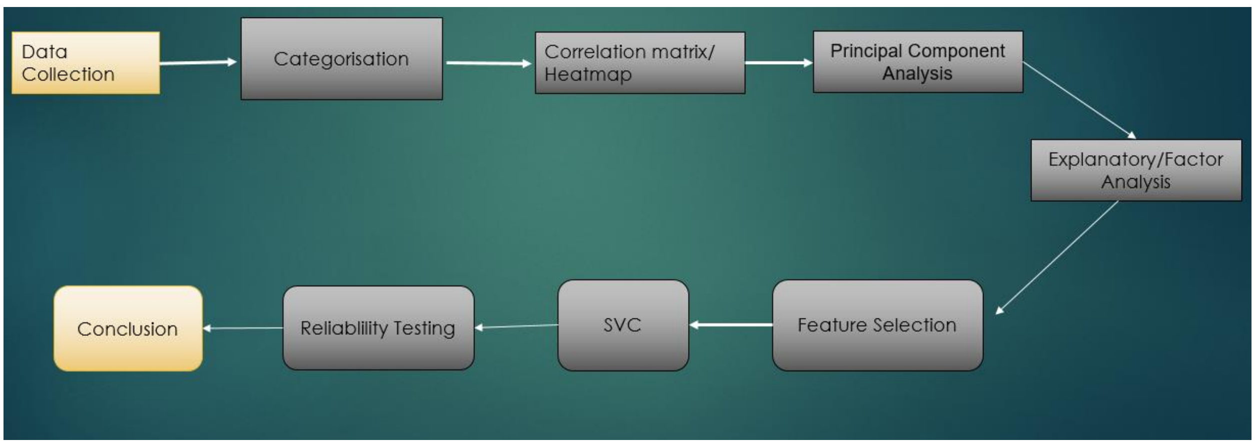

Figure 2.1: Flow Chart

CHAPTER 3 : RESEARCH METHODOLOGY

3 Research Methodology

For this research, a questionnaire survey approach has been adopted to find the impact of various attributes on delay in the Indian construction sector drawing from various international researchers mentioned above. A survey of construction professionals representing various stakeholders involved in construction projects in India was conducted. Heterogeneity of respondents is an important criteria in capturing the impact of various attributes on construction delay [6]. In this study, the heterogeneity in the survey sample was maintained by approaching to the selected group of respondents representing the key industry roles across the construction sector.

3.1 Preparation of Questionnaire

Identification of critical attributes for the study and preparation of questionnaire is a crucial step for the success of the re-search. Significant amount of work has already been done on causes of construction delay and there is a well documented and peer-reviewed set of delay attributes available in the literature. For this research, the questionnaire has been prepared by incorporating the key delay attributes reported in the literature. A total of 45 delay attributes were identified under six broad categories namely project related, site related, process related, human related, authority related and technical issues. To reflect the cross-section of the already available delay attributes inf the Indian context, personal interviews with Indian construction experts were also conducted. The final questionnaire survey was on design based on these two inputs. The attributes are listed in Table 1. A five point Likert scale (1 very low, 2 low, 3 average, 4 high, 5 very high) was adopted where respondents were asked to rank the importance and impact of a particular attribute on delay in one of their selected projects. The research was designed to be used with two statistical techniques namely factor analysis and regression modeling [3]. In addition, descriptive analysis was also performed on the attributes using the raw data collected in the survey. Descriptive analysis is an important measure for ranking the attributes in terms of their criticality as perceived by the respondents. This is similar to the analysis of the basic statistics on collected samples to investigate the trends of perceptions of certain industry practices based on first hand experiences of the practitioners. As such analysis does not provide any meaningful outcomes in terms of understanding the clustering effects of the similar attributes and the predictive capacity, further analysis is required using advanced statistical methods. Factor analysis was used to reduce the attributes for investigating the clustering effects while regression analysis was performed for deriving a predictive model based on the best fit attributes for forecasting time performance in the project [3];[17].

Descriptive statistics namely Relative Importance Index (RII) has been used to highlight the relative importance of attributes as perceived by the respondents [12]; [12]; [1]. Factor analysis is primarily used to get greater in-sight among numerous correlated but seemingly unrelated attributes into a much fewer underlying factors [3]; [1]. The results form a firm basis for identifying the criticality of attributes on construction impact. However the analysis is unable to depict the underlying relationship.

An attempt to achieve this multiple regression analysis is considered a most suitable method to derive the relationship between the attributes [3]. With these research design issues in mind a survey of Indian construction professionals was conducted. Various methods such as email, online, mail, and telephone discussions were used to collect the information from experts.

3.2 Respondent's profile

All the respondents identified had experience in relatively large engineering construction projects in the Indian context. The sample consisted of owners, architects, structural engineers, service engineers, project managers, contract administrators, design managers and construction managers. Table 2 shows a brief description of respondents' profile in terms of professional role and experience who participated in the study. As seen, the mix of disciplines was well proportioned in the sample. In order to get the best possible response commensurate by the experience and expertise, introductory conversations and email contacts were made with each respondent to explain and make the objectives of the research clear. A total of 110 questionnaires were mailed both by hard copy and via email, out of which 77 valid responses were obtained with a response rate of 70%. Though the sample size is relatively small, the quality of the responses was considered to be highly reliable for the analysis due to relevant industry experiences, personal level interactions and clear understanding of the questionnaires among the respondents (Vaus, 2001). Amongst the respondents, the highest proportion (66%) was from the contractors involved in construction activities followed by the clients (21%). Respondents from the roles of architects and design managers were 13%. The average experience of the respondents was about 15 years with 21% over 20 years.

3.3 Tertiary source

Data gathering was done through tertiary sources as well with help Beautiful Soup web scraping python libraries. Mostly research paper and open source and libraries data are collected.

Beautiful Soup is a Python package for parsing HTML and XML documents (including having malformed markup, i.e. non-closed tags, so named after tag soup). It creates a parse tree for parsed pages that can be used to extract data from HTML, which is useful for web scraping.

```
from bs4 import BeautifulSoup
import urllib.request

with       urllib.request.urlopen(https://data.gov.in/keywords/civil-
works') as response:
    soup = BeautifulSoup(response, 'html.parser')
    for anchor in soup.find_all('a'):
        print(anchor.get('href', '/'))
```

CHAPTER 4 : ATTRIBUTES AND MATHEMATICAL VALIDITY OF FACTOR ANALYSIS

4.1 Ranking of Attributes

Many researchers , and [8]; [1] are of the opinion that mean and standard deviation of each individual attribute is not a suitable measure to assess overall rankings as they do not reflect any relationship between them and hence used RII which can be calculated using the following equation:

RII(Relative Importance Index) = $\sum \mathrm{W} / (A * N)$

W Weight given to each attribute by respondent

A Highest weight

N Total number of respondents

The attributes are arranged in ascending order of ranks, attribute with highest RII or rank 1 indicates that it has the maximum impact on the delay while the attribute with lowest rank indicates that it has the least impact on delay duration. However RII doesn't talk about the relationship between the various attributes.

To identify if there is a relationship between the selected attributes, Spearman rank correlation is used (Mansfield, 1994). It assesses how well the relationship between two variables can be described using a monotonic function. The sign of the Spearman correlation indicates the direction of association between X and Y. A Spear-man correlation of zero indicates that there is no tendency for Y to either increase or decrease when X increases.

CATEGORY	SUB SET
Category 1: Materials	
A	Delay in delivering material to construction site
B	Monopoly of material by some suppliers
C	Prices fluctuation
D	Types of material availability at local market
E	Lack of consultant's knowledge of available materials
Category 2: Labor and Equipment	
F	Low productivity
G	Poor work execution
H	shortage of laborers
I	Mistakes happens during construction phase
J	Unavailability of equipment
K	Lack of engineering experience
L	Construction cost underestimation
M	Poor site management
Category 3: Financing	
N	Delayed payment for completed work
O	Cash flow problem during construction
P	High rate of interest
Q	Contractor's financial difficulties
R	Increase in salaries of skilled laborers
Category 4: Design and Documentation	
S	Change design
T	Errors and omission in design
U	Incomplete design drawings and specifications at tendering stage
V	Delay in preparation and approval of drawing
W	Inaccurate quantity take-off
X	Long period from planning of project to construction
Category 5: Management and Organization	
Y	Delays in issuing information to the contractor during construction stage
Z	No practical use of earned value management system
AA	Lack of cost planning/monitoring during pre and post contract stage

AB	Poor communication between client, consultant and contractor
AC	Insufficient coordination between local authority who is responsible about road projects and service authorities
AD	Delays in decisions making by the approval authorities
Category 6: Schedule	
AE	Change schedule
AF	Stoppages because of work being rejected by consultant
Category 7: Contractual issues	
AG	Lowest bid price
AH	Insufficient time for preparation of contract documents
AI	Discrepancies in contract documents
AJ	Underestimation of time for completion of project
AK	Extension of time with cost claims
Category 8: Scope of work	
AL	Scope and specification changes
AM	Changes in client requirements
AN	Frequency of variation orders and additional works
AO	Rework due to error in execution
Category 9: External Issues	
AP	Weather conditions
AQ	Accidents during construction
AR	Delay in getting NOCs from different government authorities

Table 1: Data Set

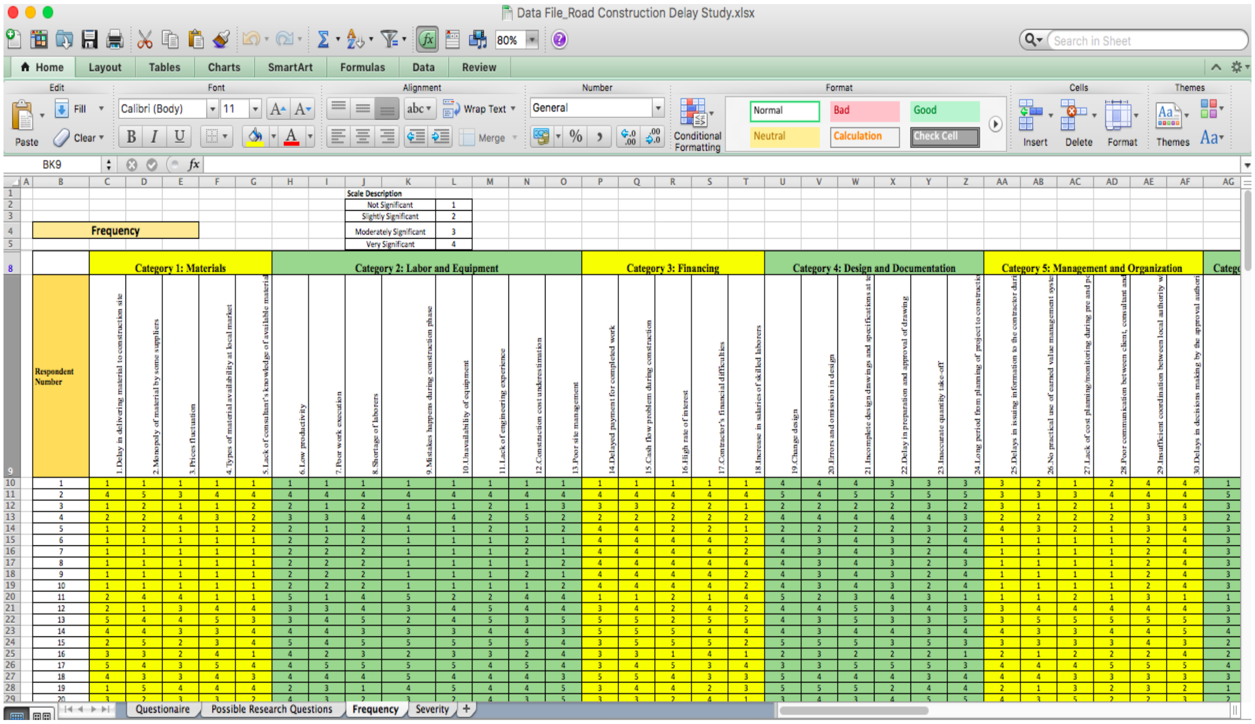

Figure 4.1: Data Set

4.2 Factor analysis for Data Set

Factor analysis is a powerful statistical technique that aims at providing greater insight among numerous correlated but seemingly unrelated variables into a much fewer underlying factors [3]. In order to evaluate the adequacy of the survey data for factor analysis, Kaiser– Meyer–Olkin (KMO) test and Bartlett's test of sphericity [17] were conducted. The value of KMO represents the ratio of the squared correlation between variables to the squared partial correlation between variables. It varies from 0 to 1. A value close to 1 indicates that the pattern of correlations is relatively compact and hence factor analysis should give distinct and reliable results [17]. A minimum value of 0.5 has been suggested. In this research of the 45 at-tributes, a total of 27 attributes are selected based on correlation.

The other 18 were found to have no significant correlation with one another and hence were excluded from further analysis. The KMO value for selected 27 variables is 0.78 which is deemed good for this study.

Components extracted: Principal component analysis is used to reduce numerous correlated attributes into much fewer underlying factors. As discussed, 27 out 45 attributes are used in this analysis. A total of seven principal components (factors) were evolved. Results are tabulated in Table 4. These seven factors explained 70.64% of total variance. Varimax rotation is used for better interpretation of results on the orthogonal factors.

Factor named lack of commitment explains 11.61% of total variance of the linear component (factor) and contains four attributes. Commitment from all the parties involved is essential for successful completion of any project [1]. First attribute site accidents due to lack of safety measures is due to lack of commitment from both client and contractor towards the project. Site accidents not only harm individuals and consume time, but also it is observed that productivity of labour reduces significantly after an accident. Time is also wasted in attending to accidents and replacing the injured person by a person with lesser or irrelevant skills. This then relates to the efforts required on training and development. These can be avoided if client and contractor are committed to appropriate safety measures adopted on the site. Second attribute lack of motivation for contractor for early finish (i.e. no incentive for early finish etc.) clearly links lack of commitment from the client and other stakeholders. Third attribute use of improper or obsolete construction methods is a result of unprofessional engagement and perhaps without an appropriate commitment to project from the contractor. Improper construction method compromises the safety and quality standards and affects the productivity, which potentially increases the duration of the project. Fourth attribute delay in material delivery by vendors shows the lack of commitment in terms of contractor's procurement planning prior to construction phase of project. Ignorance of the lead time for material delivery by the vendors potentially result material shortage, which has reportedly been one of the significant causes of schedule delay across construction projects.

4.2.1 Materials

Only very few construction materials are locally available in Brunei, so it relies on imported materials in general. The only available materials locally are: ordinary Portland cement, sand, aggregates, timber, bricks and glass. However, their local supplies cannot wholly meet the demand of the local construction industry. Therefore, these materials, along with other construction materials, are imported to meet the demand, mainly from China, Indonesia and Malaysia.This takes longer time for sourcing, procuring and transporting of the materials. Moreover, land transport is the only delivery method available within Brunei that results in to a long lead time for materials.

With higher demand of materials, for example during the construction of national housing projects, which involve the construction of hundreds or thousands of houses, local suppliers experience shortages of their stock.Saudi Arabia also experienced project delays due to similar causes during their booming construction industry in 1990s . On the other hand, locally available materials may also encounter shortage in production. For example, one of the Brunei industrial areas had suffered power failure for several days a few years back, resulting in to temporary stoppage of cement production and the factory ran out of stock. During the times of higher demand (i.e.increased number of ongoing projects), Brunei also suffers from shortage of bricks at times and contractors need to import bricks from neighbouring provinces of Malaysia. Moreover, local authority limits timber logging to only 17% by the year 2045, although 75% of Brunei is covered with forest. This may cause shortage in timber supply and Brunei may need to import it from neighbouring countries to cater for current needs.

4.2.2 Financing

Construction delay, which is often encountered in engineering constructions, is defined by scholars as the completion date of the project implementation being later than the planning date which leads to the schedule delays of the whole contract, or the owner and the contractor being not able to meet the completion date on the contract that has been signed. Meanwhile engineering construction delay not only causes the time delay to opportunity cost, resulting in the owner being not able to deliver the project in time so as to reduce investment effect, but also increases cost and reduces the enterprise prestige of the contractor. This problem affects not only the construction industry, but also the whole economy of the country. This issue is especially obvious in developing countries. Scholars have done a lot of research on the cause of this problem, pointing out that financial factor is one of the main reasons affecting the construction progress. Research has shown that financial factor is one of the main reasons for delays of engineering construction progress and cost overrun in Ghana. In Thailand, survey and design, the contractor and the relevant factors of supervision are the important factors influencing the large engineering project delays, of which the lack of resources, such as the contractor's financial difficulties matters a lot. Some scholars summarized 31 specific factors influencing large-scale project delays in Malaysia, including contractors, owner, supervision, and the external environment, among which the financial problems are the main factors influencing the project delay. In different countries and places, under different stages of practice, the risk factors influencing the construction progress are not the same. Scholars have discussed the situation in Vietnam, some of which think that the owner's ability of management, participants' ability and the external environment are the main reasons behind the success of the project. The analysis from the perspectives of different relevant persons shows that fiscal capacity shortage is one of the seven factors influencing Vietnamese engineering construction progress. Some scholars that have carried out further empirical research on large engineering projects in Vietnam summarizes 7 types and 59 specific factors, in which fiscal capacity is one of the main factors influencing the construction schedule delays in Vietnam , besides which policy factors and natural condition also affects the project construction schedule delays.

4.2.3 Labor and Equipment

Available literature contend that proper project planning, availability of materials, equipment and adequate labour are key critical success factors for the successful implementation of building construction projects. A number of studies have been carried out in those key critical dimensions in order to assess their relative contributions to schedule delays in the construction industry. Aibinu and Odeyinka identified financial difficulties, equipment breakdown and maintenance problems, planning and scheduling problems, material and equipment shortages, slow mobilization and shortage of manpower as main contributors to this category of delay factors. Al-Khalil and Al-Ghafly observed that financing and cash flow challenges, poor project management and inadequate manpower were key considerations. Al-Kharashi and Skitmore contend that poor qualification of contractor's technical staff, poor site management and supervision and difficulty in financing the project were critical. Arditi et al., observed that inadequate supply of materials, and contractor's financial difficulties were the main causes of delay. Assaf and Al-Hejji identified the contractor related delay factors as; difficulties in financing project by contractor, conflicts in sub-contractors schedule in execution of project, rework due to errors during construction, conflicts between contractor and other parties (consultant and owner), poor site management and supervision by contractor, poor communication and coordination by contractor with other parties, ineffective planning and scheduling of project by contractor, improper construction methods implemented by contractor, delays in sub-contractors work, inadequate contractor's work, frequent change of sub-contractors .

4.2.4 Design and Documentation

Another factor here is incomplete drawings and discrepancies of contract documents are major factors cause the delay. The majority of delays can be traced to inconsistent or to incomplete detailing of drawings, incorrect dimensions of walls and openings, inadequate detailing of difficult locations and inconsistent detailing. Many contracts were bid on the basis of incomplete information, require extensive changes during construction. Edwin H. 2005 stated that the design responsibility is transferred from the owner organization to the Consultant who is responsible for the design management in the Project and to be delivered by the design procedure system. "Reference" stated that lack of information about the type of contracts, the conditions of contracts, major design issues, standard specifications and major design criteria to managers and engineers who works in consultant offices are one of the major problems that construction industry sector in gulf region is still suffering and generates many problems during the execution. Many problems occur in schematic and detailed design where conflicts between structural and services drawings becomes the norm. This creates difficulties in getting the approval of the final detailed design and other tender documents from state authorities.Another major factor of delays caused by the Consultant is the inability of effectively managing and preparing the contract document including bill of quantities and the approved drawings which creates a large margin of errors and omission in quantities. Poor appreciation of the design management process is another factor that causes delay by the Consultant. Inspection process during the work operations, the approval process duration of submittals and approvals of project materials and the technical site experience of the inspector who gets the instruction from the Consultant resident engineer are also factors attributes the project delay.

4.2.5 Management and Organization

Factor named Management and Organization explains that 10.96% of total variance of the linear component comprises four key attributes. Inefficient site management is mentioned as one of the top five reasons for construction delay . First attribute ambiguity in specifications and potential for conflicting interpretation by parties poses an enormous challenge to the contractors for rolling out an appropriate management plan onsite. Second attribute poor labour productivity is caused either by employing unskilled labour or due to lack of proper supervision over them which comes under inefficient management skills of the supervisor onsite. In case there is unavailability of workforce with the required skill set and hiring of unskilled labour is inevitable, they must be trained properly before putting them at work. Third attribute lack of control over sub contractor reflects the inefficient management skills of main contractor. This perhaps links to lack of clear contractual framework and objective criteria for engaging subcontractors in Indian projects. Lack of control over subcontractor may lead to unwanted conflicts, low productivity and development of negative attitudes on the site. Fourth attribute inadequate experience of contractor is due to lack of site management skills of the client. Inexperienced contractor may not be able to cope up with the progress of work or may not understand the complexity of project leading to misinterpretation and confusion. Inadequate experience of contractor in turn leads to improper management of site and thus cause time overruns.

4.2.6 Scope of work

Factor, named as Scope of work in project scope explains that about 10.57% of total variance of the linear component contains five attributes. First attribute rework due to change in design or deviation order (with a factor loading of 0.861) is an effect caused by improper design brief and poor coordination between the owner, designer and engineer. Second attribute rework due to errors in execution (with a factor loading of 0.657) contributes to delay as the rework itself consumes time and resources. Rework due to errors in execution implies project manager's lack of understanding of scope or design of the project.. Increase in scope of work at a later stage delays the project completion due to change in quantities and change in project schedule. Increase in scope of work may further delay project due to unavailability of appropriate spare resources with the contractors. In fact, increase in scope of work results into a complete drain out of the contractor's resources and reduce his capability to follow the time plan. Fifth attribute improper storage of materials leading to damage when necessary (with a factor loading of 0.524) occurs due to lack of awareness and negligence by the project manager on proper inventory planning and storage of materials onsite. Improper storage resulting to the non availability of material when needed is one of the key causes across most projects and should be considered seriously in relation to required lead time for procuring materials when needed.

4.3 Mathematical validity of analysis

Once factors have been extracted, it is necessary to cross check if factor analysis measured what was intended to be measured i.e. the attributes in each factor formed collectively explain the same measure within target dimensions [3].If attributes truly form the factor identified, it is understood that they should reasonably correlate with one another but not the perfect correlation though. By calculating Pearson correlation using SPSS we can estimate the extent of correlation among various variables. The values of Pearson correlation are tabulated in Table 5. We find that Pearson bivariate correlations are greater than 0.4 in most of the cases among different attributes in all the factors. From these results, we can ensure that factors formed in factor analysis contain attributes which are related. For reliability analysis, which is required to ensure the construct of the model over time (i.e. consistency of measured attributes and scale), Cronbach's alpha test was performed on entire data as well as attributes in each factor which are shown in Table 6. The value of Cα could be anywhere in the range of 0 to 1, where a higher value denotes the greater internal consistency and vice versa.

The value of Cα is inflated by a large number of variables, so there is no set interpretation as to what is an acceptable limit. However, a rule of thumb applies to most situations with the following ranges: Cα> 0.9 denotes excellent, 0.9 > Cα> 0.8 as good, 0.8 > Cα> 0.7 as acceptable, 0.7 > Cα> 0.6 as questionable, 0.6 > Cα> 0.5 as poor and
0.5 > Cα denotes unacceptable . The value of Cα for all attributes calculated is 0.944 which is considered to be excellent.

4.4 Spearman's rank correlation

In statistics, Spearman's rank correlation coefficient or Spearman's rho, named after Charles Spearman and often denoted by the Greek letter (rho) or as , is a nonparametric measure of rank correlation (statistical dependence between the rankings of two variables). It assesses how well the relationship between two variables can be described using a monotonic function. The Spearman correlation between two variables is equal to the Pearson correlation between the rank values of those two variables; while Pearson's correlation assesses linear relationships, Spearman's correlation assesses monotonic relationships (whether linear or not). If there are no repeated data values, a perfect Spearman correlation of +1 or −1 occurs when each of the variables is a perfect monotone function of the other.

Intuitively, the Spearman correlation between two variables will be high when observations have a similar (or identical for a correlation of 1) rank (i.e. relative position label of the observations within the variable: 1st, 2nd, 3rd, etc.) between the two variables, and low when observations have a dissimilar (or fully opposed for a correlation of −1) rank between the two variables. Spearman's coefficient is appropriate for both continuous and discrete ordinal variables.

Both Spearman's and Kendall's can be formulated as special cases of a more
general correlation coefficient.

Spearman rank correlation is a non-parametric test that is used to measure the degree of association between two variables. The Spearman rank correlation test does not carry any assumptions about the distribution of the data and is the appropriate correlation analysis when the variables are measured on a scale that is at least ordinal.

The following formula is used to calculate the Spearman rank correlation:

$$\rho = 1 - \frac{6\sum di^2}{n(n^2-1)}) \qquad \text{...... (i)}$$

ρ= Spearman rank correlation

di= the difference between the ranks of corresponding variables

n= number of observations

Another option is simply to use the full version of Spearman's formula (actually a slightly modified Pearson's r), which will deal with tied ranks:

$$\rho = \frac{Sxy}{SxSy} = \frac{\frac{1}{n}\sum_{i=1}^{n}(R(xi)-\overline{R(x)}).(R(yi)-\overline{R(y)})}{\sqrt{\frac{1}{n}\sum_{i=1}^{n}(R(xi-\overline{R(x)})^2.(\frac{1}{n}\sum_{i=1}^{n}(R(yi)-\overline{R(y)}))^2)}} \qquad \text{...... (ii)}$$

Eq 2: Spearman's formula

Effect size: Cohen's standard may be used to evaluate the correlation coefficient to determine the strength of the relationship, or the effect size. Correlation coefficients between .10 and .29 represent a small association, coefficients between .30 and .49 represent a medium association, and coefficients of .50 and above represent a large association or relationship.

Ordinal data: In an ordinal scale, the levels of a variable are ordered such that one level can be considered higher/lower than another. However, the magnitude of the difference between levels is not necessarily known. An example would be rank ordering levels of education. A graduate degree is higher than a bachelor's degree, and a bachelor's degree is higher than a high school diploma. However, we cannot quantify how much higher a graduate degree is compared to a bachelor's degree. We also cannot say that the difference in education between a graduate degree and a bachelor's degree is the same as the difference between a bachelor's degree and a high school diploma.

4.5 Spearman over Pearson coefficients

The Pearson and Spearman correlation coefficients can range in value from −1 to +1. For the Pearson correlation coefficient to be +1, when one variable increases then the other variable increases by a consistent amount. This relationship forms a perfect line. The Spearman correlation coefficient is also +1 in this case.

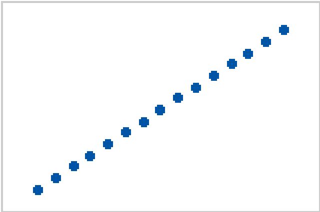

Figure 4.2: Pearson = +1, Spearman = +1

If the relationship is that one variable increases when the other increases, but the amount is not consistent, the Pearson correlation coefficient is positive but less than +1. The Spearman coefficient still equals +1 in this case.

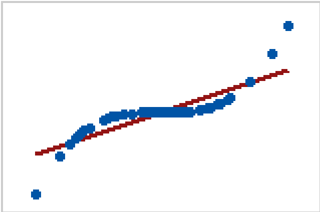

Figure 4.3: Pearson = +0.851, Spearman = +1

When a relationship is random or non-existent, then both correlation coefficients are nearly zero.

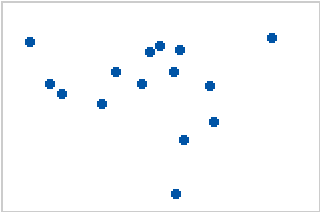

Figure 4.4: Pearson = −0.093, Spearman = −0.093

If the relationship is a perfect line for a decreasing relationship, then both correlation coefficients are −1.

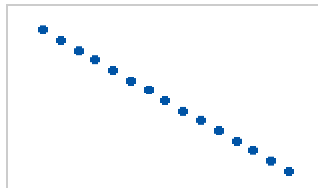

Figure 4.5: Pearson = −1, Spearman = −1

If the relationship is that one variable decreases when the other increases, but the amount is not consistent, then the Pearson correlation coefficient is negative but greater than −1. The Spearman coefficient still equals −1 in this case

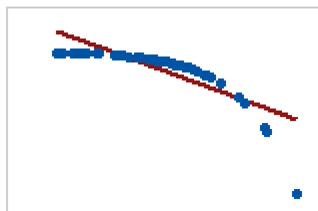

Figure 4.6: Pearson = −0.799, Spearman = −1

Correlation values of −1 or 1 imply an exact linear relationship, like that between a circle's radius and circumference. However, the real value of correlation values is in quantifying less than perfect relationships. Finding that two variables are correlated often informs a regression analysis which tries to describe this type of relationship more.

4.5.1 Other nonlinear relationships

Pearson correlation coefficients measure only linear relationships. Spearman correlation coefficients measure only monotonic relationships. So a meaningful relationship can exist even if the correlation coefficients are 0. Examine a scatter plot to determine the form of the relationship.

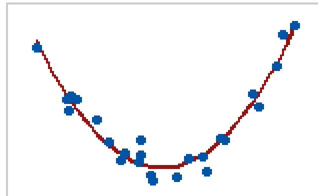

Figure 4.7: Coefficient of 0

This graph shows a very strong relationship. The Pearson coefficient and Spearman coefficient are both approximately 0.

4.5.2 Heat Map

A heat map (or heatmap) is a graphical representation of data where the individual values contained in a matrix are represented as colors. It is a bit like looking a data table from above. It is really useful to display a general view of numerical data, not to extract specific data point. It is quite straightforward to make a heat map, as shown on the examples below. However be careful to understand the underlying mechanisms. We can probably normalise matrix, choose a relevant colour palette, use cluster analysis and thus permute the rows (Responses) and the columns(Factors) of the matrix to place similar values near each other according to the clustering.

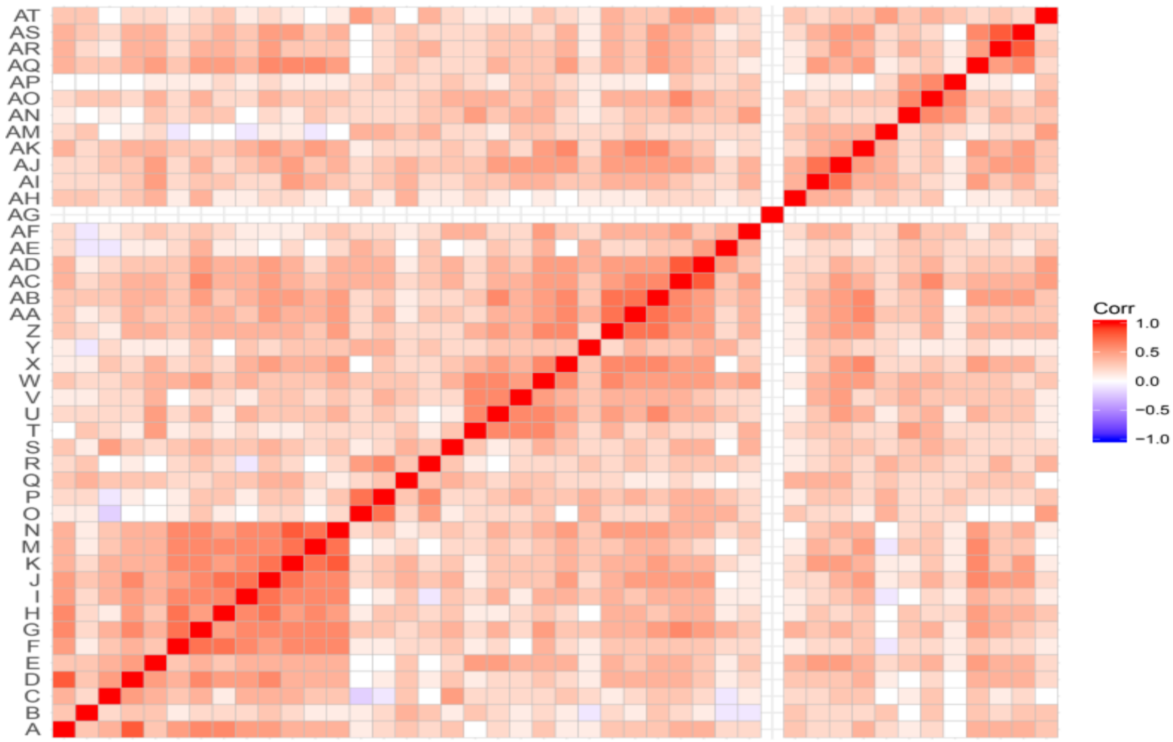

Figure 4.8: Heat Map

4.6 Factor analysis in Detail

Factor analysis is a technique that is used to reduce a large number of variables into fewer numbers of factors. This technique extracts maximum common variance from all variables and puts them into a common score. As an index of all variables, we can use this score for further analysis. Factor analysis is part of general linear model (GLM) and this method also assumes several assumptions: there is linear relationship, there is no multicollinearity, it includes relevant variables into analysis, and there is true correlation between variables and factors. Several methods are available, but principal component analysis is used most commonly.

4.6.1 Principal component analysis

This is the most common method used by researchers. PCA starts extracting the maximum variance and puts them into the first factor. After that, it removes that variance explained by the first factors and then starts extracting maximum variance for the second factor. This process goes to the last factor.

Principal Component Analysis, or PCA, is a dimensionality-reduction method that is often used to reduce the dimensionality of large data sets, by transforming a large set of variables into a smaller one that still contains most of the information in the large set.

Reducing the number of variables of a data set naturally comes at the expense of accuracy, but the trick in dimensionality reduction is to trade a little accuracy for simplicity. Because smaller data sets are easier to explore and visualize and make analyzing data much easier and faster for machine learning algorithms without extraneous variables to process.

So to sum up, the idea of PCA is simple—reduce the number of variables of a data set, while preserving as much information as possible.

Below is given Scree plot for the analysis

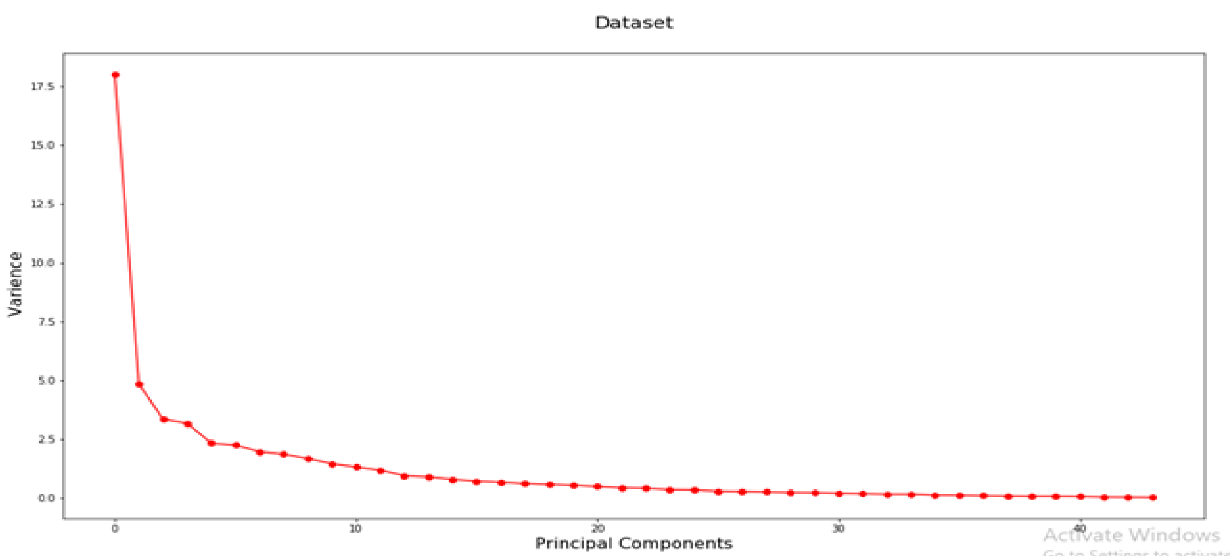

Figure 4.9: PCA

Step by step explanation

Step 1: Standardization

The aim of this step is to standardize the range of the continuous initial variables so that each one of them contributes equally to the analysis.

More specifically, the reason why it is critical to perform standardization prior to PCA, is that the latter is quite sensitive regarding the variances of the initial variables. That is, if there are large differences between the ranges of initial variables, those variables with larger ranges will dominate over those with small ranges (For example, a variable that ranges between 0 and 100 will dominate over a variable that ranges between 0 and 1), which will lead to biased results. So, transforming the data to comparable scales can prevent this problem.

Mathematically, this can be done by subtracting the mean and dividing by the standard deviation for each value of each variable.

$$z = \frac{value - mean}{standard\ deviation} \qquad \text{…… (iii)}$$

Once the standardization is done, all the variables will be transformed to the same range.

Step 2: Covariance Matrix computation

The aim of this step is to understand how the variables of the input data set are varying from the mean with respect to each other, or in other words, to see if there is any relationship between them. Because sometimes, variables are highly correlated in such a way that they contain redundant information. So, in order to identify these correlations, we compute the covariance matrix.

The covariance matrix is a p × p symmetric matrix (where p is the number of dimensions) that has as entries the covariances associated with all possible pairs of the initial variables. For example, for a 3-dimensional data set with 3 variables x, y, and z, the covariance matrix is a 3×3 matrix of this from:

$$\begin{bmatrix} Cov(x,x) & Cov(x,y) & Cov(x,z) \\ Cov(y,x) & Cov(y,y) & Cov(y,z) \\ Cov(z,x) & Cov(z,y) & Cov(z,z) \end{bmatrix} \qquad \text{…… (iv)}$$

Since the covariance of a variable with itself is its variance (Cov(a,a)=Var(a)), in the main diagonal (Top left to bottom right) we actually have the variances of each initial variable. And since the covariance is commutative (Cov(a,b)=Cov(b,a)), the entries of the covariance matrix are symmetric with respect to the main diagonal, which means that the upper and the lower triangular portions are equal.

Step 3: Compute the eigenvectors and eigenvalues of the covariance matrix to identify the principal components

Eigenvectors and eigenvalues are the linear algebra concepts that we need to compute from the covariance matrix in order to determine the principal components of the data. Before getting to the explanation of these concepts, let's first understand what do we mean by principal components.

Principal components are new variables that are constructed as linear combinations or mixtures of the initial variables. These combinations are done in such a way that the new variables (i.e., principal components) are uncorrelated and most of the information within the initial variables is squeezed or compressed into the first components. So, the idea is 10-dimensional data gives you 10 principal components, but PCA tries to put maximum possible information in the first component, then maximum remaining information in the second and so on, until having something like shown in the scree plot below.

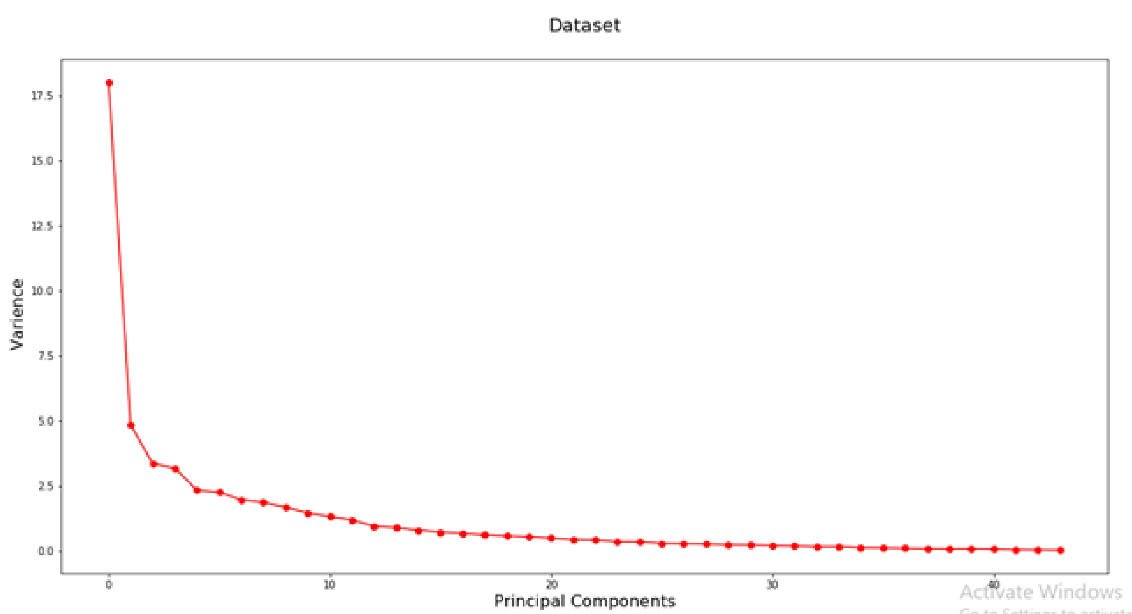

Figure 4.10: PCA Output

4.7 Feature Selection

Feature selection, also known as variable selection, attribute selection or variable subset selection, is the process of selecting a subset of relevant features (variables, predictors) for use in model construction. Feature selection techniques are used for several reasons like simplification of models to make them easier to interpret by users, enhanced generalization by reducing overfitting (formally reduction of variance), shorter training times and to avoid any shortcomings of dimensionality. The biggest advantage of using Feature Selection is that the data contains some features that are either redundant or irrelevant, and can thus be removed without incurring much loss of information. Feature selection techniques are often used in domains where there are many features and comparatively few samples.

This algorithm can be seen as the combination of a search technique for proposing new feature subsets, along with an evaluation measure which scores the different feature subsets. The simplest algorithm is to test each possible subset of features finding the one which minimizes the error rate. This is an exhaustive search of the space, and is computationally intractable for all but the smallest of feature sets.

4.8 Bartlett's Test

Bartlett's test tests the hypothesis that the correlation matrix is equal to the identity matrix. It is a test of sphericity checks whether or not the observed variables inter-correlate at all using the observed correlation matrix against the identity matrix. If the test found statistically insignificant, you should not employ a factor analysis.

In this Bartlett 's test, the p-value should be less than significant level or near to 0 so that test is statistically significant, indicating the observed correlation matrix is not an identity matrix.

Test the hypothesis that the correlation matrix is equal to the identity matrix.

H0: The matrix of population correlations is equal to 1.

H1: The matrix of population correlations is not equal to 1.

The formula for Bartlett's Sphericity test is:

$$-1*(n-1-((2p+5)/6))*\ln(\det(R))-1*(n-1-((2p+5)/6))*\ln(\det(R))$$

where R det(R) is the determinant of the correlation matrix, and p is the number of variables.

Parameters: x (array-like) – The array from which to calculate sphericity. It returns:-

1) statistic (float) – The chi-square value.

2) p_value (float) – The associated p-value for the test.

4.9 Cronbach's Alpha Reliability Testing

Reliability comes to the forefront when variables developed from summated scales are used as predictor components in objective models. Since summated scales are an assembly of interrelated items designed to measure underlying constructs, it is very important to know whether the same set of items would elicit the same responses if the same questions are recast and re-administered to the same respondents. Variables derived from test instruments are declared to be reliable only when they provide stable and reliable responses over a repeated administration of the test.

The ALPHA option in PROC CORR provides an effective tool for measuring Cronbach's alpha, which is a numerical coefficient of reliability. Computation of alpha is based on the reliability of a test relative to other tests with same number of items, and measuring the same construct of interest . This paper will illustrate the use of the ALPHA option of the PROC CORR procedure from SAS(R) to assess and improve upon the reliability of variables derived from summated scales.

Sixteen questions using Likert-type scales (1 = strongly agree; 6 = strongly disagree) from a national agricultural and food preference policy survey were administered nationwide. Usable survey forms, totaling 1,111, were received and processed using the PROC FACTOR and PROC CORR procedures of SAS. Three common factors were extracted during factor analysis and were interpreted to represent "subsidy policy" factors, "regulatory policy" factors, and "food safety policy" factors (Santos, Lippke, & Pope, 1998).

To make the demonstration on Cronbach's alpha possible, SB8, which was a variable previously deleted during factor analysis, was restored in the data set. SB8 was used to demonstrate how a poorly selected item on a summated scale can affect the resulting value of alpha. It should be noted here that factor analysis is not required in the determination of Cronbach's alpha.

After factor analysis, it is a common practice to attach a descriptive name to each common factor once it is extracted and identified. The assigned name is indicative of the predominant concern that each factor addresses. In SAS, a RENAME FACTOR(i)='descriptive name' statement would do the job. In this example, this can be accomplished by

RENAME FACTOR1=SUBSIDY
FACTOR2=REGULATE
FACTOR3=FSAFETY

While labeling is critical, it definitely makes for an easy identification of which construct is running on what particular procedure. At this point, the named common factors can now be used as independent or predictor variables. However, most experienced researchers would insist on running a reliability test for all the factors before using them in subsequent analyses.

```
Console ~/Downloads/
> library(psych)
> alpha(da)
Some items ( AQ AR ) were negatively correlated with the total scale and
probably should be reversed.
To do this, run the function again with the 'check.keys=TRUE' option
Reliability analysis
Call: alpha(x = da)

  raw_alpha std.alpha G6(smc) average_r S/N    ase mean   sd
      0.94      0.94    0.99      0.23  15 0.0069  3.3 0.55

 lower alpha upper     95% confidence boundaries
0.93 0.94 0.95

 Reliability if an item is dropped:
    raw_alpha std.alpha G6(smc) average_r S/N alpha se
A        0.94      0.94    0.98      0.23  15   0.0070
B        0.94      0.94    0.98      0.23  15   0.0068
C        0.94      0.94    0.98      0.23  15   0.0069
D        0.94      0.94    0.98      0.23  15   0.0071
E        0.94      0.94    0.98      0.23  15   0.0071
F        0.94      0.94    0.98      0.23  15   0.0070
G        0.94      0.93    0.98      0.23  14   0.0072
H        0.94      0.94    0.98      0.23  15   0.0071
I        0.94      0.94    0.98      0.23  15   0.0071
J        0.94      0.93    0.98      0.23  14   0.0072
K        0.94      0.94    0.98      0.23  14   0.0072
L        0.94      0.94    0.98      0.23  15   0.0071
M        0.94      0.94    0.98      0.23  14   0.0072
N        0.94      0.94    0.98      0.23  15   0.0069
O        0.94      0.94    0.98      0.23  15   0.0070
P        0.94      0.94    0.98      0.23  15   0.0070
Q        0.94      0.94    0.98      0.23  15   0.0069
R        0.94      0.94    0.98      0.23  15   0.0070
S        0.94      0.94    0.98      0.23  15   0.0070
T        0.94      0.94    0.98      0.23  15   0.0071
U        0.94      0.94    0.98      0.23  15   0.0070
```

Figure 4.11: Cronbach Table

CHAPTER 5 : RESULTS AND OUTPUT

5.1 Results and Output

The results and output of all the 9 categories are generalized here and demonstrated below category wise.

Category 1: Materials

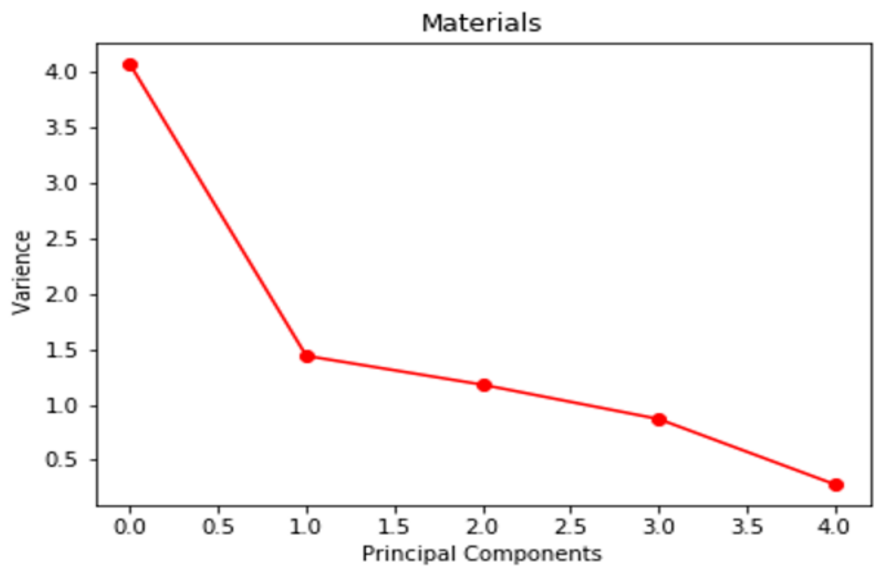

Figure 5.1: Material Scree Plot

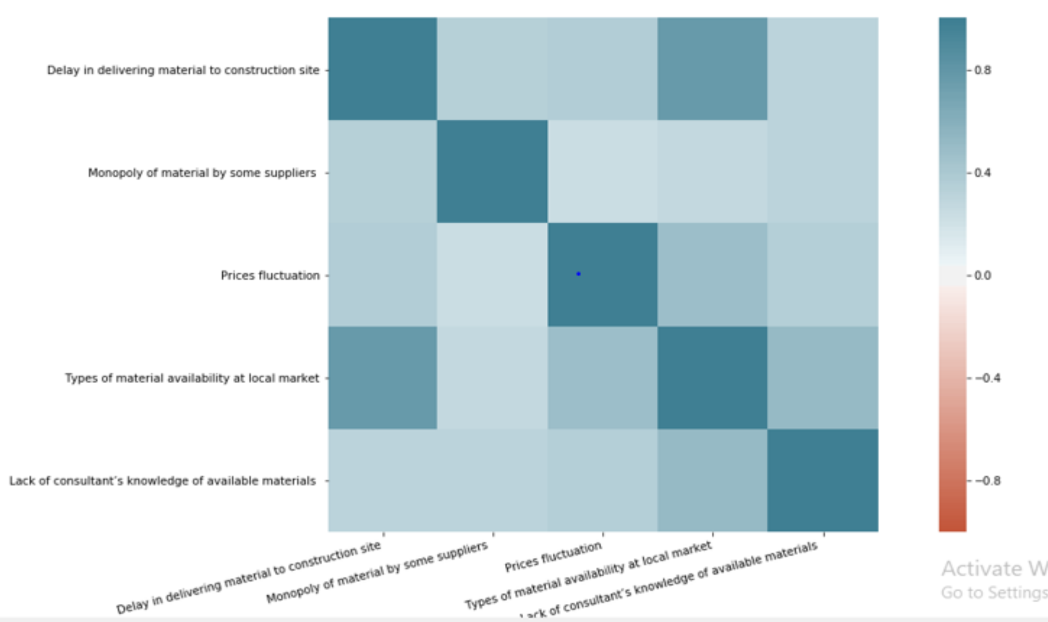

Figure 5.2: Material Heatmap

```
IPython console
Console 1/A

In [3]: from factor_analyzer.factor_analyzer import
calculate_bartlett_sphericity
   ...: chi_square1,p_value1=calculate_bartlett_sphericity(x1)
   ...: print('\nChi Square Value (Category 1): ',chi_square1)
   ...: print('\np Value (Category 1): ',p_value1)

Chi Square Value (Category 1):  235.03091103703633

p Value (Category 1):  3.654473744029261e-45
```

Figure 5.3: Labor and Equipment Chi_square & p_value

```
Console 1/A

In [7]: uniq1 = fa.get_uniquenesses()
   ...: print('\nUniqueness (Category 1):\n',uniq1)
   ...: ev1, v1 = fa.get_eigenvalues()
   ...: print('\nEigenValues (Category 1):\n',ev1)

Uniqueness (Category 1):
 [0.05812532 0.70768769 0.81224008 0.21872354 0.4670037 ]

EigenValues (Category 1):
 [2.63442753 0.81696845 0.72714019 0.63474533 0.18671851]
```

Figure 5.4: Labor and Equipment Uniqueness & EigenValues

```
Console 1/A

In [12]: from sklearn.feature_selection import SelectKBest
    ...: from sklearn.feature_selection import chi2
    ...:
    ...: # Feature extraction
    ...: test = SelectKBest(score_func=chi2, k=1)
    ...: fit = test.fit(x1, y)
    ...:
    ...: # Summarize scores
    ...: np.set_printoptions(precision=2)
    ...: print(fit.scores_)
[0.14 1.14 0.85 2.19 2.91]
```

Activate Windows
Go to Settings to activate Windows.

Figure 5.5: Labor and Equipment Feature Selection

Category 2: Labor and Equipment

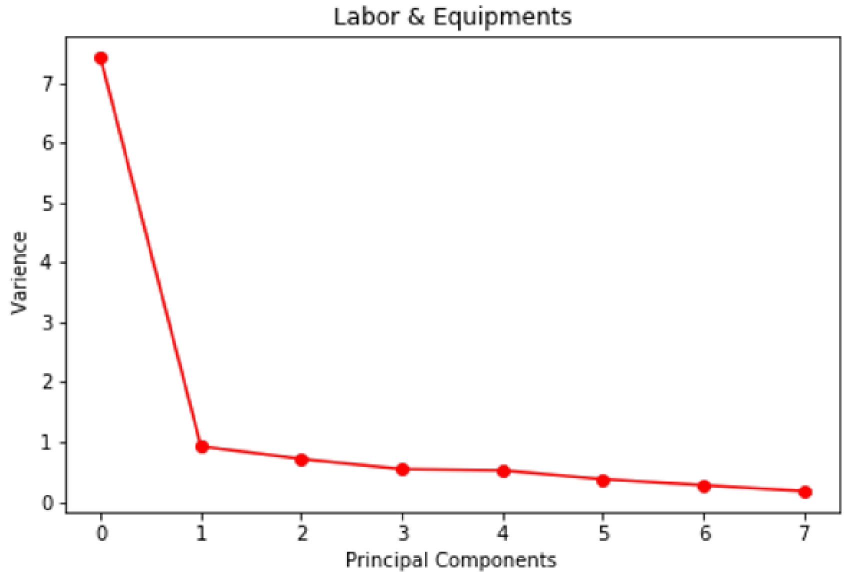

Figure 5.6: Labor and Equipment Scree Plot

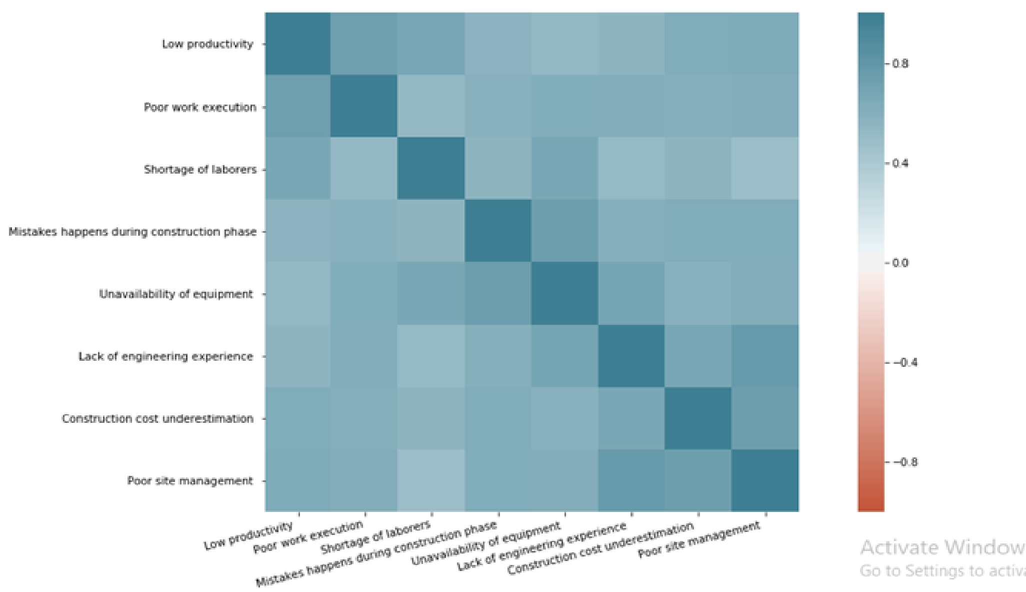

Figure 5.7: Labor and Equipment Heatmap

```
Console 1/A
In [41]: from factor_analyzer.factor_analyzer import
calculate_bartlett_sphericity
    ...: chi_square2,p_value2 = calculate_bartlett_sphericity(x2)
    ...: print('\nChi Square Value (Category 2): ',chi_square2)
    ...: print('\np Value (Category 2): ',p_value2)

Chi Square Value (Category 2):  832.069878170697

p Value (Category 2):  1.8693853088766914e-157
```

Figure 5.8: Labor and Equipment Chi_square & p_value

```
Console 1/A

Uniqueness (Category 2):
 [-0.25574942  0.71884025  0.56992181  0.66896488 -0.2874541   0.41171781
  0.53865559 -0.07211556]

EigenValues (Category 2):
 [5.35510013 0.66076843 0.58914814 0.42056533 0.36980169 0.26356376
 0.18771224 0.15334028]
```

Figure 5.9: Labor and Equipment Uniqueness & EigenValues

```
Console 1/A
In [14]: from sklearn.feature_selection import SelectKBest
    ...: from sklearn.feature_selection import chi2
    ...:
    ...: # Feature extraction
    ...: test = SelectKBest(score_func=chi2, k=1)
    ...: fit = test.fit(x2, y)
    ...:
    ...: # Summarize scores
    ...: np.set_printoptions(precision=2)
    ...: print(fit.scores_)
[3.45 3.99 0.83 2.23 0.06 3.5   8.8   3.89]
```

Activate Windows

Figure 5.10: Labor and Equipment Feature Selection

Category 3: Financing

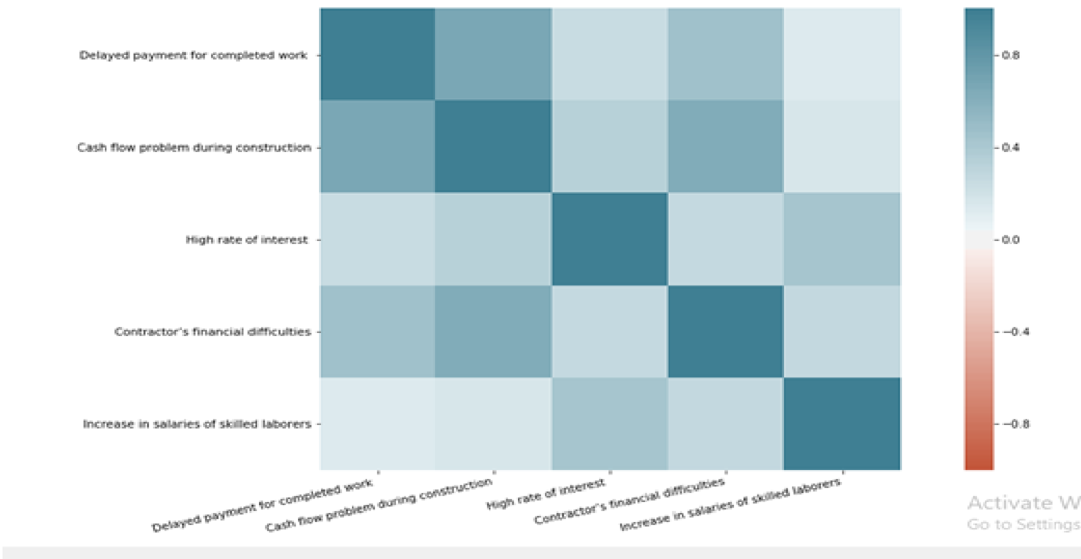

Figure 5.11: Financing Heatmap

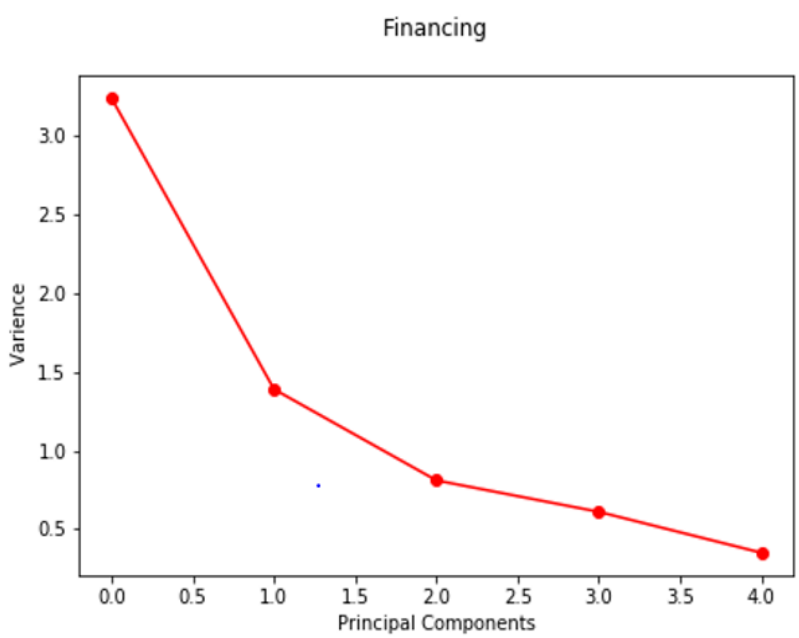

Figure 5.12: Scree Plot

```
In [14]: from sklearn.feature_selection import SelectKBest
    ...: from sklearn.feature_selection import chi2
    ...:
    ...: # Feature extraction
    ...: test = SelectKBest(score_func=chi2, k=1)
    ...: fit = test.fit(x2, y)
    ...:
    ...: # Summarize scores
    ...: np.set_printoptions(precision=2)
    ...: print(fit.scores_)
[3.45 3.99 0.83 2.23 0.06 3.5  8.8  3.89]
```

Figure 5.13: FinancingChi_square & p_value

```
In [17]: from factor_analyzer import FactorAnalyzer
    ...: fa3 = FactorAnalyzer()
    ...: fa3.fit_transform(x3)
    ...:
    ...: # Check Eigenvalues
    ...: ev3, v3 = fa3.get_eigenvalues()
    ...: r_ev3 = np.round(ev3, decimals=2)
    ...: print('\nEigenValues (Category 3):\n',ev3)

EigenValues (Category 3):
 [2.49 1.14 0.61 0.48 0.28]
```

Figure 5.14: Financing Uniqueness & EigenValues

```
In [18]: from sklearn.feature_selection import SelectKBest
    ...: from sklearn.feature_selection import chi2
    ...:
    ...: # Feature extraction
    ...: test = SelectKBest(score_func=chi2, k=2)
    ...: fit = test.fit(x3, y)
    ...:
    ...: # Summarize scores
    ...: np.set_printoptions(precision=2)
    ...: print(fit.scores_)
[0.5  0.23 1.58 0.11 3.01]
```

Figure 5.15: Financing Feature Selection

Category 4: Design and Documentation

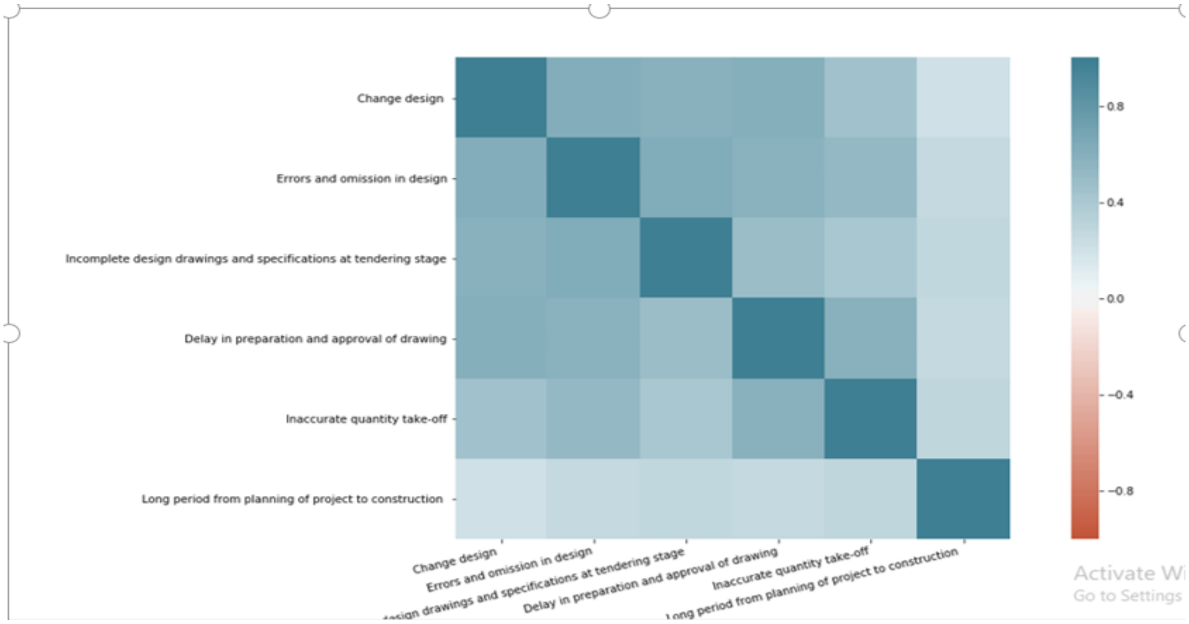

Figure 5.16: Design and Documentation Heatmap

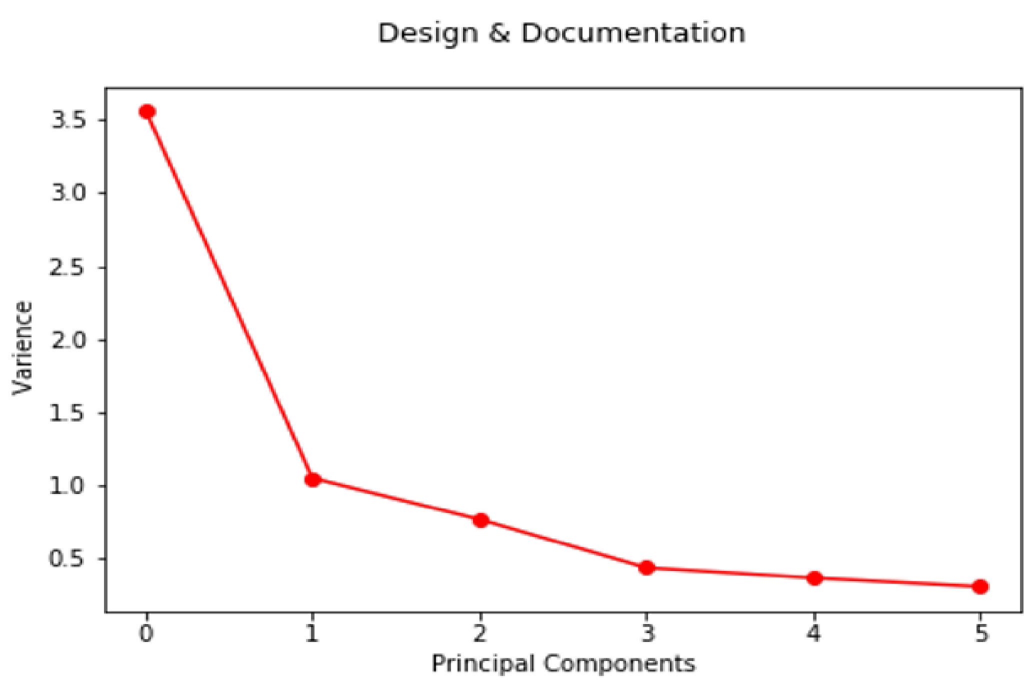

Figure 5.17: Design and Documentation Scree Plot

```
IPython console
Console 1/A
In [20]: from factor_analyzer.factor_analyzer import
calculate_bartlett_sphericity
    ...: chi_square4, p_value4 = calculate_bartlett_sphericity(x4)
    ...: print('\nChi Square Value (Category 4): ',chi_square4)
    ...: print('\np Value (Category 4): ',p_value4)

Chi Square Value (Category 4):  323.87466883003947

p Value (Category 4):  2.8780733079822416e-60
```

Figure 5.18: Design and DocumentationChi_square & p_value

```
IPython console
Console 1/A
In [21]: from factor_analyzer import FactorAnalyzer
    ...: fa = FactorAnalyzer()
    ...: fa.fit_transform(x4)
    ...:
    ...: # Check Eigenvalues
    ...: ev4, v4 = fa.get_eigenvalues()
    ...: r_ev4 = np.round(ev4, decimals=2)
    ...: print('\nEigenValues (Category 4):\n',ev4)

EigenValues (Category 4):
 [3.35 0.87 0.65 0.44 0.34 0.34]
```

Activate Windows

Figure 5.19: Design and Documentation Uniqueness & EigenValues

```
IPython console
Console 1/A
In [22]: from sklearn.feature_selection import SelectKBest
    ...: from sklearn.feature_selection import chi2
    ...:
    ...: # Feature extraction
    ...: test = SelectKBest(score_func=chi2, k=1)
    ...: fit = test.fit(x4, y)
    ...:
    ...: # Summarize scores
    ...: np.set_printoptions(precision=2)
    ...: print(fit.scores_)
[1.89 0.8  0.5  1.83 1.45 1.97]
```

Activate Windows

Figure 5.20: Design and Documentation Feature Selection

Category 5: Management and Organization

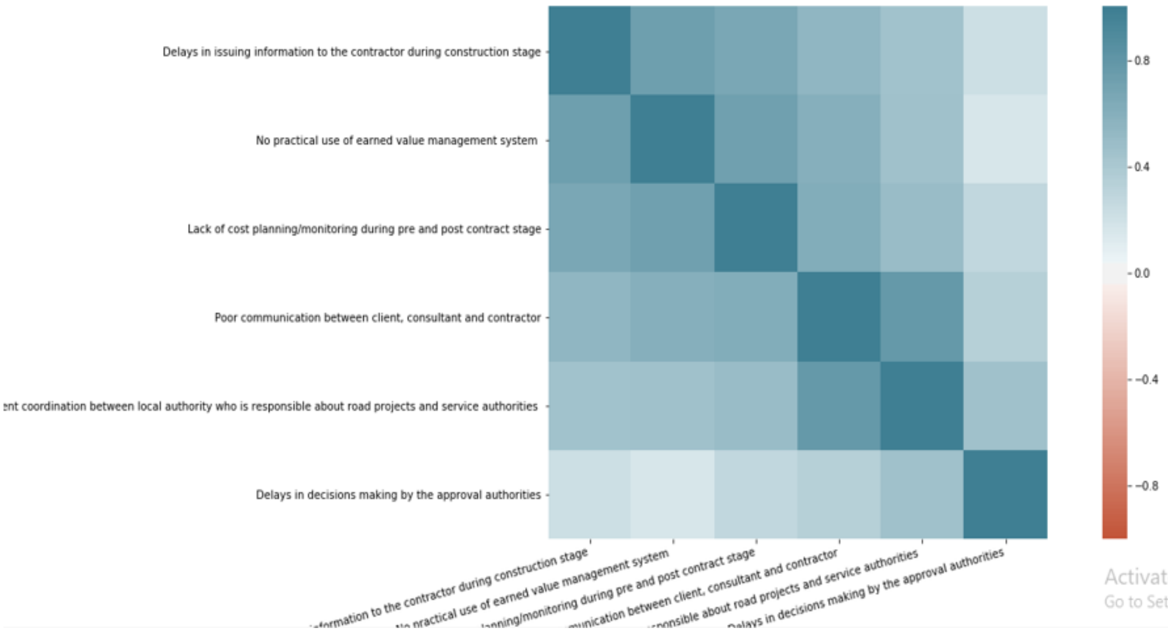

Figure 5.21: Management and Organization Heatmap

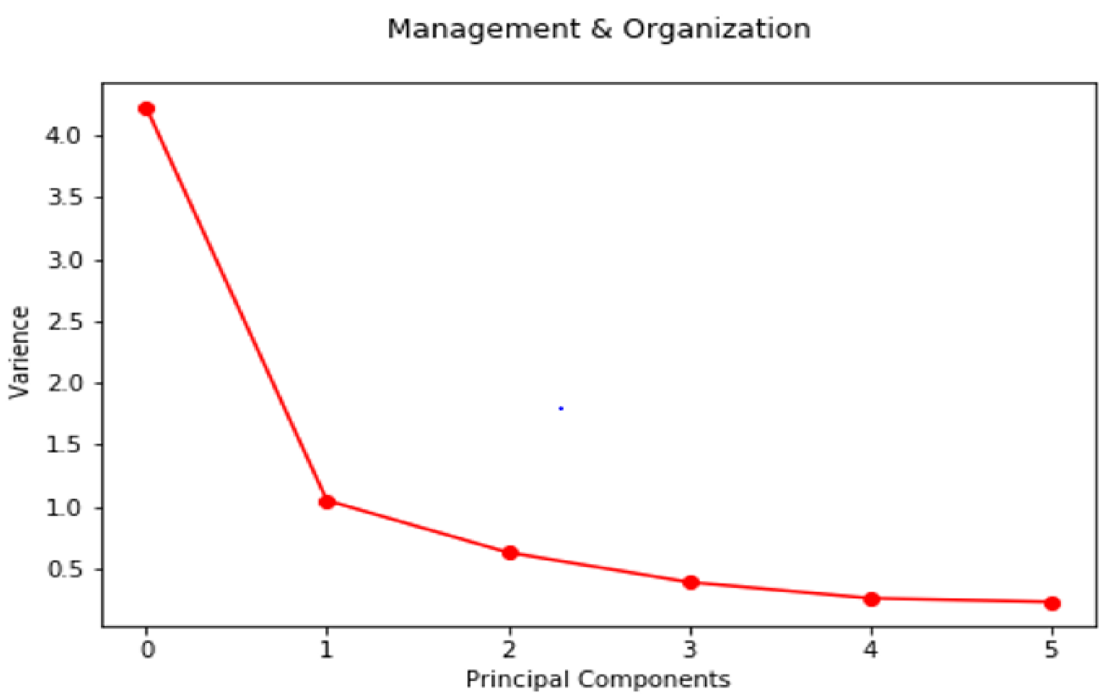

Figure 5.22: Management and Organization Scree Plot

```
In [24]: from factor_analyzer.factor_analyzer import
calculate_bartlett_sphericity
    ...: chi_square5, p_value5 = calculate_bartlett_sphericity(x5)
    ...: print('\nChi Square Value (Category 5): ',chi_square5)
    ...: print('\np Value (Category 5): ',p_value5)

Chi Square Value (Category 5):  471.10071079473954

p Value (Category 5):  3.5250241538566654e-91
```

Figure 5.23: Management and Organization Chi_square & p_value

```
In [25]: from factor_analyzer import FactorAnalyzer
    ...: fa = FactorAnalyzer()
    ...: fa.fit_transform(x5)
    ...:
    ...: # Check Eigenvalues
    ...: ev5, v5 = fa.get_eigenvalues()
    ...: r_ev5 = np.round(ev5, decimals=2)
    ...: print('\nEigenValues (Category 5):\n',ev5)

EigenValues (Category 5):
 [3.61 1.04 0.59 0.33 0.25 0.19]
```

Figure 5.24: Management and Organization Uniqueness & EigenValues

```
In [26]: from sklearn.feature_selection import SelectKBest
    ...: from sklearn.feature_selection import chi2
    ...:
    ...: # Feature extraction
    ...: test = SelectKBest(score_func=chi2, k=2)
    ...: fit = test.fit(x5, y)
    ...:
    ...: # Summarize scores
    ...: np.set_printoptions(precision=2)
    ...: print(fit.scores_)
[7.51 4.14 4.96 6.7  1.2  1.07]
```

Figure 5.25: Management and Organization Feature Selection

Category 6: Schedule

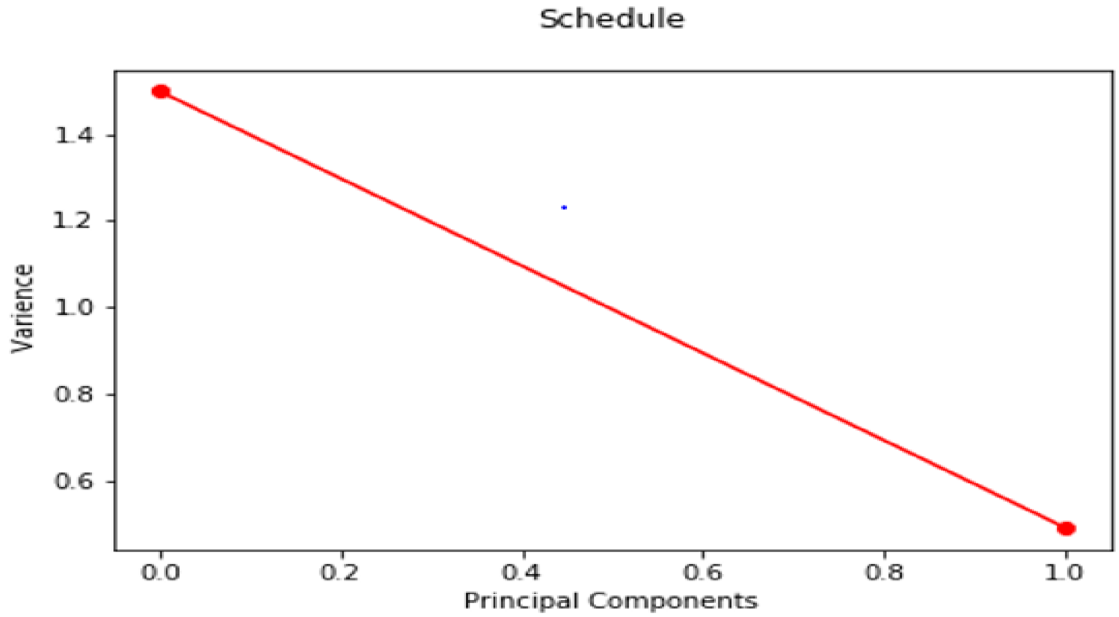

Figure 5.26: Schedule Code Scree plot

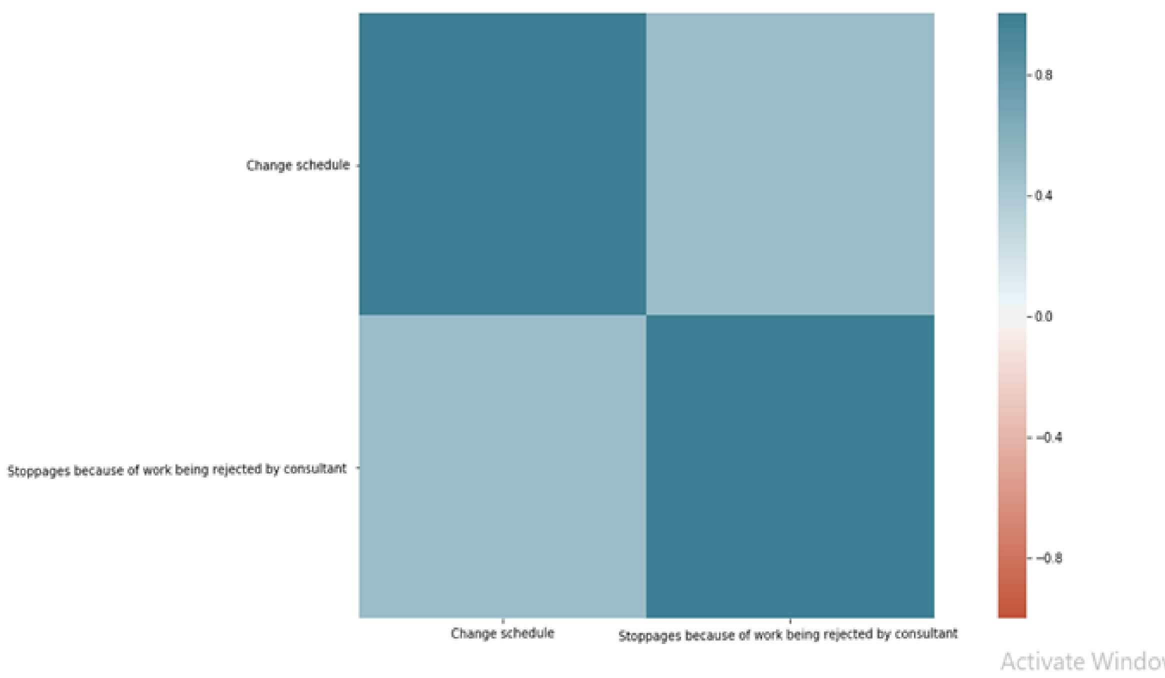

Figure 5.27: Schedule Heatmap

```
Console 1/A

In [28]: from factor_analyzer.factor_analyzer import
calculate_bartlett_sphericity
    ...: chi_square6, p_value6 = calculate_bartlett_sphericity(x6)
    ...: print('\nChi Square Value (Category 6): ',chi_square6)
    ...: print('\np Value (Category 6): ',p_value6)

Chi Square Value (Category 6):  40.22666548656567

p Value (Category 6):  1.15756030538956e-10
```

Figure:5.28 Schedule Chi_square & p_value


```
Console 1/A

In [28]: from factor_analyzer.factor_analyzer import
calculate_bartlett_sphericity
    ...: chi_square6, p_value6 = calculate_bartlett_sphericity(x6)
    ...: print('\nChi Square Value (Category 6): ',chi_square6)
    ...: print('\np Value (Category 6): ',p_value6)

Chi Square Value (Category 6):  40.22666548656567

p Value (Category 6):  1.15756030538956e-10
```


Figure 5.29: Schedule Uniqueness & EigenValues

```
IPython console
Console 1/A

In [30]: from sklearn.feature_selection import SelectKBest
    ...: from sklearn.feature_selection import chi2
    ...:
    ...: # Feature extraction
    ...: test = SelectKBest(score_func=chi2, k=1)
    ...: fit = test.fit(x6, y)
    ...:
    ...: # Summarize scores
    ...: np.set_printoptions(precision=2)
    ...: print(fit.scores_)
[3.84 0.59]
```

Figure 5.30: Schedule Feature Selection

Category 7: Contractual issues

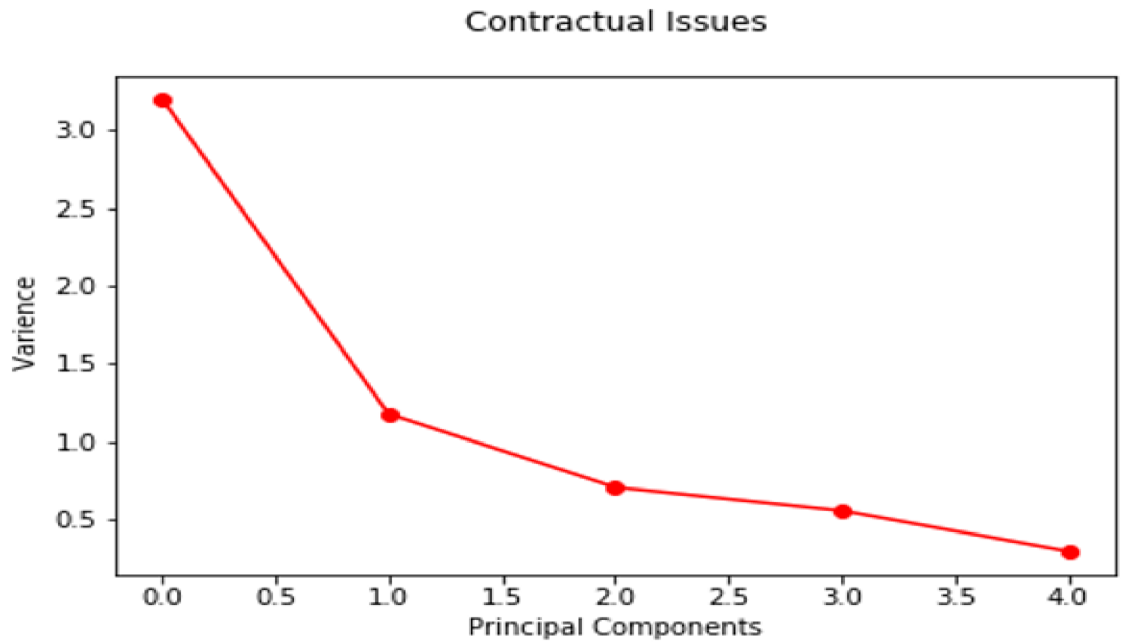

Figure 5.31: Contractual issues Scree Plot

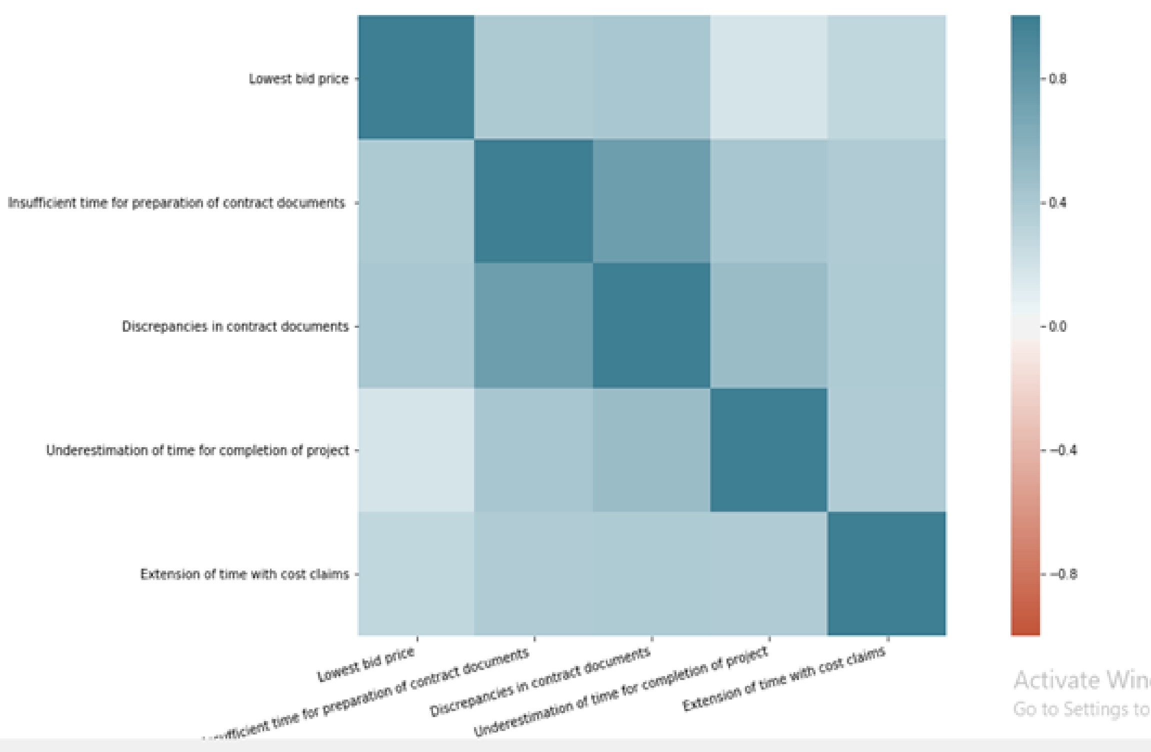

Figure 5.32: Contractual issues Heatmap

```
Console 1/A

In [32]: from factor_analyzer.factor_analyzer import
calculate_bartlett_sphericity
    ...: chi_square7 ,p_value7 = calculate_bartlett_sphericity(x7)
    ...: print('\nChi Square Value (Category 7): ',chi_square7)
    ...: print('\np Value (Category 7): ',p_value7)

Chi Square Value (Category 7):  214.31800331265228

p Value (Category 7):  7.948959650407999e-41
```

Figure 5.33: Contractual issues Chi_square & p_value

```
Console 1/A

In [33]: from factor_analyzer import FactorAnalyzer
    ...: fa7 = FactorAnalyzer()
    ...: fa7.fit_transform(x7)
    ...:
    ...: # Check Eigenvalues
    ...: ev7, v7 = fa7.get_eigenvalues()
    ...: r_ev7 = np.round(ev7, decimals=2)
    ...: print('\nEigenValues (Category 7):\n',ev7)

EigenValues (Category 7):
 [2.67 0.83 0.71 0.53 0.26]
```

Activate Windows

Figure 5.34: Contractual issues Uniqueness & EigenValues

```
IPython console
Console 1/A

In [34]: from sklearn.feature_selection import SelectKBest
    ...: from sklearn.feature_selection import chi2
    ...:
    ...: # Feature extraction
    ...: test = SelectKBest(score_func=chi2, k=1)
    ...: fit = test.fit(x7, y)
    ...:
    ...: # Summarize scores
    ...: np.set_printoptions(precision=2)
    ...: print(fit.scores_)
[4.16 6.61 4.42 1.93 0.25]
```

Activate Windows

Figure 5.35: Contractual issues Feature Selection

Category 8: Scope of work

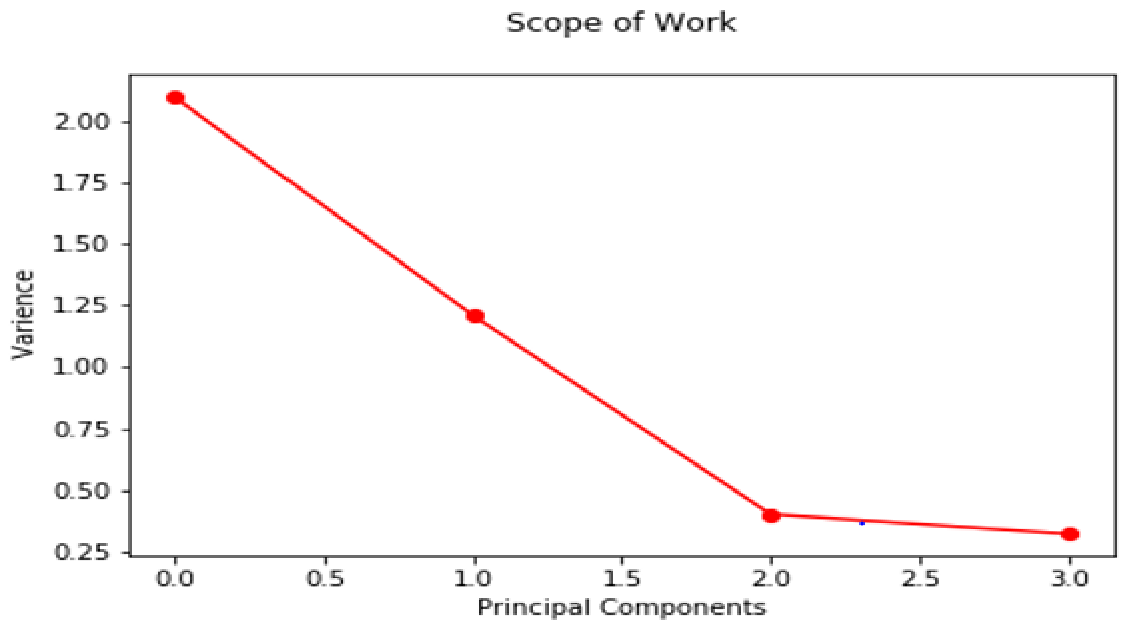

Figure 5.36: Scope of work Scree Plot

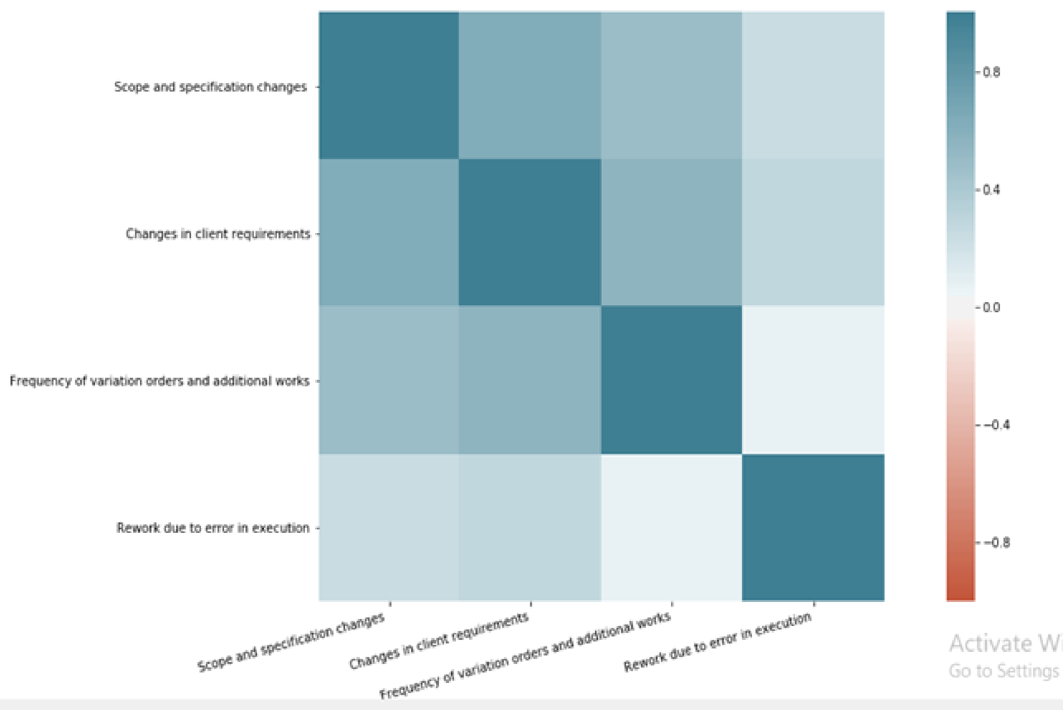

Figure 5.37: Scope of work Heatmap

```
In [36]: from factor_analyzer.factor_analyzer import
calculate_bartlett_sphericity
    ...: chi_square8, p_value8 = calculate_bartlett_sphericity(x8)
    ...: print('\nChi Square Value (Category 8): ',chi_square8)
    ...: print('\np Value (Category 8): ',p_value8)

Chi Square Value (Category 8):  145.37344506538028

p Value (Category 8):  3.5760837298003455e-29
```

Figure 5.38: Scope of work Chi_square & p_value

```
In [37]: from factor_analyzer import FactorAnalyzer
    ...: fa8 = FactorAnalyzer()
    ...: fa8.fit_transform(x8)
    ...:
    ...: # Check Eigenvalues
    ...: ev8, v8 = fa8.get_eigenvalues()
    ...: r_ev8 = np.round(ev8, decimals=2)
    ...: print('\nEigenValues (Category 8):\n',ev8)

EigenValues (Category 8):
 [2.23 0.94 0.48 0.35]
```

Figure 5.39: Scope of work Uniqueness & EigenValues

```
In [38]: from sklearn.feature_selection import SelectKBest
    ...: from sklearn.feature_selection import chi2
    ...:
    ...: # Feature extraction
    ...: test = SelectKBest(score_func=chi2, k=1)
    ...: fit = test.fit(x8, y)
    ...:
    ...: # Summarize scores
    ...: np.set_printoptions(precision=2)
    ...: print(fit.scores_)
[1.07 3.8  1.32 1.46]
```

Figure 5.40: Scope of work Feature Selection

Category 9: External issues

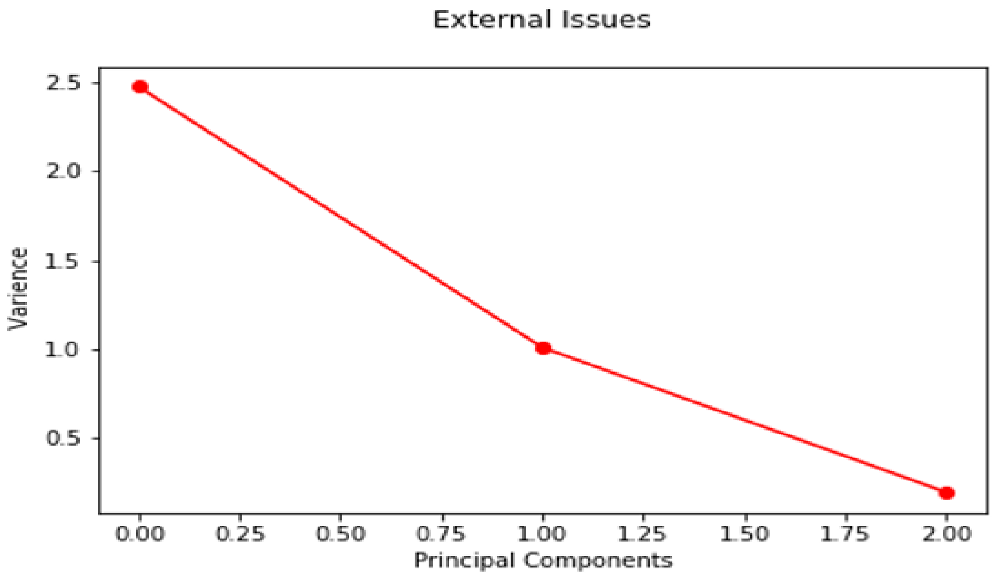

Figure 5.41: External issues Scree Plot

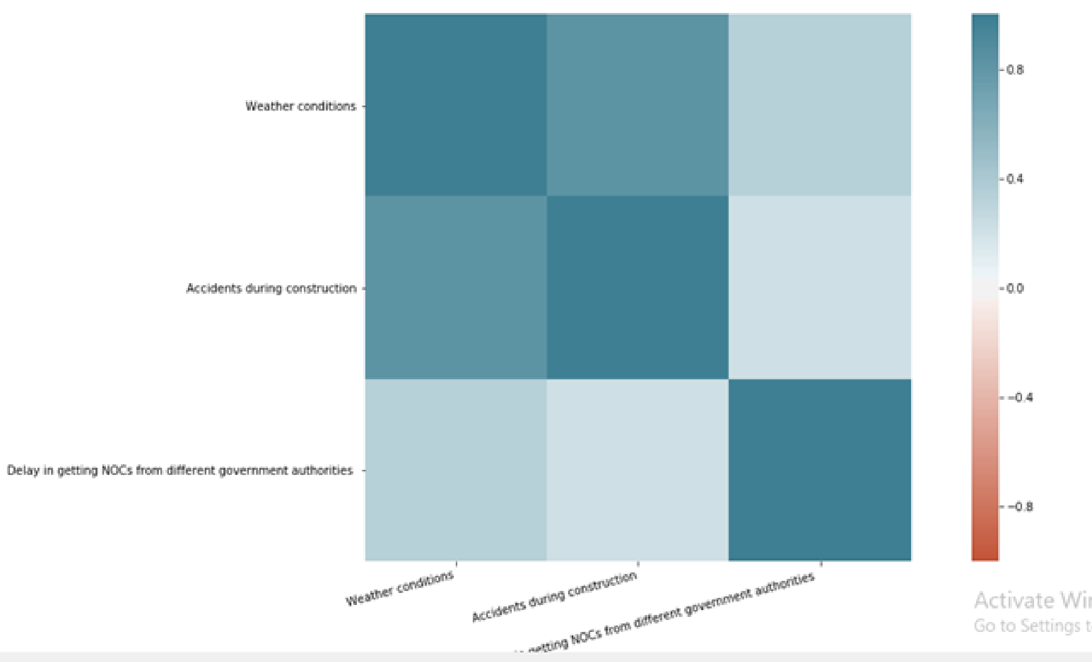

Figure 5.42: External issues Heatmap

CHAPTER 6 : CONCLUSION

This research reveals that one of the most critical factors of construction delay is the lack of commitment. This finding is indeed a clear contrast to the findings [5] that financial problem of a contractor is the most important cause. Inefficient site management is certainly another key factors affecting time performance of most construction projects in India. This is perhaps due to lack of formal training among the site professionals who usually develop their supervisory skills by experience. Most notably, the importance of this factor, however in different orders, has been identified in the previous research on cost performance context in Indian construction projects [1]. Based on RII we can infer that material shortage is the most significant factor in construction delays. This finding supports the findings by [6] where material shortage was one of the key factors affecting time delay in Malaysian construction projects. The result of the regression model asserted that slow decision from owner, poor labour productivity, architects' reluctance for change and rework due to mistakes in construction are the reasons affecting the overall delay of Indian construction project significantly. While lack of commitment with four key attributes was found to be the most influencing factor, none of these attributes found to have any significant predictive power in the regression analysis. However, slow decision from owner and rework due to errors in executive (regression coefficients 0.368 and 0.326 respectively) are found to have greater predictive power in the regression model. Poor site management (with regression coefficient 0.299) is one of the other key attributes clearly affecting delay in Indian construction projects. Similar poor labour productivity and consultant's reluctance for change (with regression coefficients 0.165 and 0.177 respectively) also need attention in achieving time success in Indian projects.

Despite a clear understanding of these key factors among the research communities, a sincere attempt to address this chronic issue of time overrun is yet to materialise amongst practitioners in the Indian construction industry. Traditionally, the approach to managing construction is quite ad-hoc on Indian projects and need for adopting a systematic approach has not been realised across the board. This became evident on the world arena during the execution stage of programmes and projects during the recently concluded Commonwealth Games 2010. In the advent of the rapid urbanisation and fast growth in construction industry, these factors must be considered and well-integrated in the mainstream construction processes for improving industry practices across the construction projects. Benchmarking data on performance of Indian construction projects is not available.

Recently efforts to measure, especially on large infrastructure projects, time and cost performance have begun. In 2009, the Bureau of Indian Standards released the first guideline on construction project management. The path forward requires setting of standards, benchmarking performance, skills development, and research undertakings in project management. Via this research the authors have attempted to highlight some of these issues. Consequently these findings are envisioned to be significant contributions to the Indian construction industry in controlling the time overruns on construction projects.

6.1. Limitations of the research

Though best efforts were put in this research and findings do make a significant contribution for industry, this research has some limitations. First the sample size of 77 is considered to be on the smaller side for statistical analysis. Secondly the respondents are not evenly distributed among the professional roles which may have induced some bias in responses. Thus the model formed may be further honed based on detailed discussions and suggestions from industry experts. The relationship between various reasons of delay and its impact on overall project delay has to be detailed further which is the author's intended future work.

CHAPTER 7 : REFERENCES

7 References

[1] Iyer, K.C. and Jha, K.N., 2005. Factors affecting cost performance: evidence from Indian construction projects. *International journal of project management*, *23*(4), pp.283-295.

[2] Ahsan, K., & Gunawan, I. (2010). "Analysis of cost and schedule performance of international development projects". *International Journal of Project Management*, *28*(1), 68-78.

[3] Doloi, H., Sawhney, A., Iyer, K.C. and Rentala, S., 2012. Analysing factors affecting delays in Indian construction projects. *International journal of project management*, *30*(4), pp.479-489.

[4] www.mospi.nic.in

[5] Abd El-Razek, M.E., Bassioni, H.A. and Mobarak, A.M., 2008. Causes of delay in building construction projects in Egypt. *Journal of construction engineering and management*, *134*(11), pp.831-841.

[6] Sambasivan, M. and Soon, Y.W., 2007. Causes and effects of delays in Malaysian construction industry. *International Journal of project management*, *25*(5), pp.517-526.

[7] Kazaz, A., Ulubeyli, S. and Tuncbilekli, N.A., 2012. Causes of delays in construction projects in Turkey. *Journal of Civil Engineering and Management*, *18*(3), pp.426-435.

[8] Faridi, A.S. and El-Sayegh, S.M., 2006. Significant factors causing delay in the UAE construction industry. *Construction Management and Economics*, *24*(11), pp.1167-1176.

[9] Al-Khalil, M.I. and Al-Ghafly, M.A., 1999. Delay in public utility projects in Saudi Arabia. *International Journal of Project Management*, *17*(2), pp.101-106.

[10] Al-Kharashi, A. and Skitmore, M., 2009. Causes of delays in Saudi Arabian public sector construction projects. *Construction Management and Economics*, *27*(1), pp.3-23.

[11] Ling, F.Y.Y. and Hoi, L., 2006. Risks faced by Singapore firms when undertaking construction projects in India. *International Journal of Project Management*, *24*(3), pp.261-270.

[12] Fugar, F.D. and Agyakwah-Baah, A.B., 2010. Delays in building construction projects in Ghana. Construction Economics and Building, 10(1-2), pp.103-116.

[13] Aibinu, A.A. and Odeyinka, H.A., 2006. Construction delays and their causative factors in Nigeria. *Journal of construction engineering and management*, *132*(7), pp.667-677.

[14] Bramble, B.B. and Callahan, M.T., 1992. Disruption and Lost Productivity.

[15] Odeyinka, H.A. and Yusif, A., 1997. The causes and effects of construction delays on completion cost of housing projects in Nigeria. *Journal of Financial Management of Property and Construction*, *2*, pp.31-44.

[16] Assaf, S.A. and Al-Hejji, S., 2006. Causes of delay in large construction projects. *International journal of project management*, *24*(4), pp.349-357.

[17] Mansfield, N.R., Ugwu, O.O. and Doran, T., 1994. Causes of delay and cost overruns in Nigerian construction projects. *International journal of project Management*, *12*(4), pp.254-260.