## Executive Summary

In today’s world business problems are easily fixed if we record the preferences of customers in the form of data and run analytics on them to understand the trend, the likes and dislikes and thus propose solutions that can bridge the gap between the end customers and company. In this research, we analyzed the requirement of the students and their needs which would drive them to use the college cafeteria extensively by increasing their frequency of visit. The data regarding the need, the likes and dislikes about the college cafeteria was collected in the form of a survey. The data was then recorded in an excel sheet. The variables were identified and linear regression analysis was performed on the dataset and the influence of each variable on the dependent variable was verified against the error acceptance level which 5%.The inference from the model is that one unit change in the dependent variable will be explained 19% by the independent variables as the adjusted r square value is 0.19.The projected started around 15th of March 2017 and analysis was submitted by 5th of May 2017. The project is very important for the cafeteria management in order to understand the drawback they have which when resolved can fix increase the student visiting the cafeteria and in turn the revenue. The main objective of the project is to understand problems or the reasons which doesn’t please students in visiting the cafeteria and suggest solutions which can be proposed to the management to work on to improve the service and get rid of the presumptions that they might have stopped them from improvement. Thus, this project suggests solution which is feasible which will improve the revenue of cafeteria.

# INTRODUCTION

Texas A&M University Commerce runs its own cafeteria where the students of the university can have unlimited buffet for 8.84$ during any meal of the day. The number of students enrolled in college during a semester is 13000 approximately and of which 30% of the students stay on campus. It is estimated that the international students enrolled with the university is as close as 900 and 70% of which stay near campus. Despite of these facts the college canteen is never the first choice for most of these students. These draw the attention of the management and consider the issue behind this set back.

## Background Problem

The purpose of this project is mainly to understand and evaluate various variables that could possibly be act as the reason behind low turn around and the college cafeteria not being the first choice for any international student or the on campus students. In this project we can analyze the pitfalls and come up with solution to increase the students visit to the cafeteria for their meal. The goal is to run conduct a rigorous survey and to apply the statistical model on the data and identify and evaluate the variables. The objective of the project is to identify and understand the influence of these variables on the frequency of the students visit to the cafeteria. The expected result from the project is once after identifying the influencing factors suggest the best possible solution to the management of the cafe in order to increase the student frequency.

## Overall goal

The goal of the project is to collect the student feedback carefully with no discrepancy and then identify the variables required and run the regression on the dataset, identify the relationship between the variable and propose solution which would increase the frequency of students visit to the cafeteria.

**Project values to the Stakeholder**

The project will give insight about the pitfalls to the cafeteria management and the solution to improve their revenue by increasing the frequency of visit by students. In the case of students they are benefited by this project indirectly as they will find a better service at cafeteria. In the case of the analysts who conduct this project they will get a feel of the real time data the scent, will help them learn and understand a new skillset, help them learn how to overcome the pits fall or the unseen issues that any new business analyst would face and finally the analysts will get to use and learn R and Tableau.

## Scope

Scope of the project is to identify and analyze the independent variable which influences the frequency of visit of the students to the cafeteria and to validate the statistical model accuracy which is used in the analysis for the prediction. In this analysis from the list of variables that we collect from the survey we pick few variables which and enforce the relation of these independent variables with that of the frequency of visit of the students to the cafeteria. Acceptance of the error in the project is just 5% anything beyond this percentage is not considered to be a reliable value to consider the analysis further. Project deliverables will have a report on the variables involved and their positive or negative impact on the frequency of the student visiting cafeteria, it will also suggest the possible measures that could obtain the desired result. While the survey is conducted and the data is collected there is a good chance that few data doesn’t make sense or they are incomplete or they are just are inappropriate; such data are excluded from the analysis. Additionally few variables which don’t impact the end result are excluded from analysis. Constraint of the project is that we can record the preference of sample number of students who visit cafeteria and try to apply our result to the entire population of student in the university. And another constrain is that there is a good chance of the data being biased and the model doesn’t fit best to explain the desired result. We accept that not all the variables actually impact the frequency to visits and also there is some accepted level of the R square value in the statistical analysis which doesn’t defines the model has some accepted percentage of error variance and the model is never hundred percent errors free. And also that if the model proves the data is biased then we take another set of sample and conduct the survey.

## Stake Holder Registered

In this project we have two set of stake holders. Internal stake holder and external stake holders are:

Internal Stake holders: Project team members, they are Divyaja Gundrathi, Muthulakshmi Venkatakrishnan, Emmanuel Issac Akpo and Shivani Dharmapuri

External Stake holder: Professor Son Bui, to whom the analysis and the end result, will be presented and the students who take the survey, the students who would befit from the change once the management implements the changes suggested from the derived analysis. Cafeteria management whose gets mainly benefitted from the research analysis. Measurable Objectives : Measureable objectives are the determination of the measure of the variable influence on student preference by designing survey questions and distributing them across students electronically or by pamphlets. The variables involved in the Project are as below.

|  |  |
| --- | --- |
| **Table 1:** Explains the variables measured in the project | |
| Serial Number | Variables Measured |
| **1** | Job status |
| **2** | Spice of food |
| **3** | Cafeteria Prime Hours |
| **4** | Cost of the meal |
| **5** | Service Provided |
| **6** | Ambience |
| **7** | Variety in food availability |
| **8** | Type of Meat |
| **9** | Quality of Food |
| **10** | Quality of Food |
| **11** | Locality |

## Resource Requirements

|  |  |
| --- | --- |
| **Table 2:** Explains the resource information | |
| Number of Resources Needed | Skillset Required |
| **4** | Decide on the variables involved |
| Design Survey Questions |
| Design Survey Template |
| Print and Distribute Survey |
| Excellent knowledge on Regression Analysis |
| Expertise in R |
| Expertise in Tableau |
| Good at Documentation  Leadership  Team player |

This Project needs four resources that have excellent leadership quality and are team players. The resources must have a skillset to identify the variables, generate survey question and design the survey that can be shared electronically with students. Which can then be distributed and the survey response can be recorded and maintained for the analysis. Additionally the team needs to have extensive knowledge in R analytical tool which will be used for data analysis and Tableau, a tool which will be used in presenting the result during the project submission. Finally, the solution must be proposed and documented for further reference.

## Plans to obtain resources:

* Resources are selected through three rounds of interview.
* Candidates who are perusing or have completed their Masters in Business Analytics with a GPA 3.0 and above are considered for interview process.
* Candidates who demonstrate excellent communication skills and presentation skill in the management round are selected for second round of interview.
* Candidates who answer the technical questions about R and Tableau get through for the third HR round.
* Candidates who answer the situational questions with compelling leadership quality and team spirit are selected into the team.

**Cost of the resources**

Below gives the break-down of the cost involved in the project:

Cost involved in Resources:

Total number of resource required for the project : 4

Total number of hours required from resource (in hours): 832

Cost per hour involved in this project (in USD) $ : 20

Total number of cost per resource (in USD) $ : = 16640

Cost involved in tools:

R $ : 0

Tableau trail version (in USD) $ : 0

Total cost (in USD) $ : 0

Cost involved in Survey:

Cost involved in Printing 1 pamphlet (in USD) $ :0.10

Cost involved in printing 30 pamphlet (in USD) $ : 30\* 0.10

Total cost (in USD) $ : 3

Total Cost Involved: $16643

Thus the total cost involving all the expenditure to execute this project successfully is $16643.

## Risk Registered:

1. The Risk involved in this project is that the survey result from sample might be biased as the survey might have circulated more amongst certain group of people whose geographical background might be same and hence the result might not be as expected.

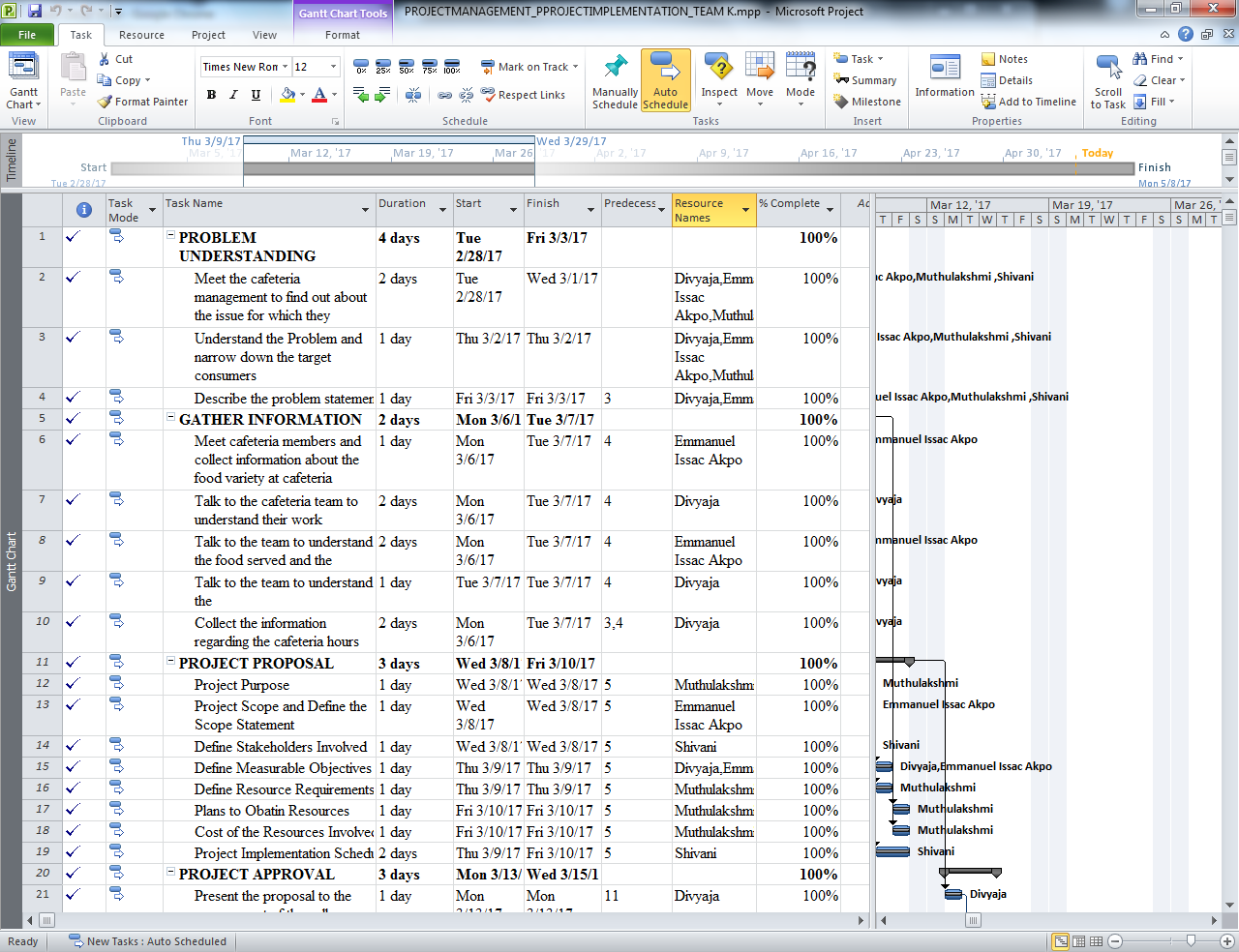
2. Survey takers might not take the survey with interest and might end up giving incorrect dataset.

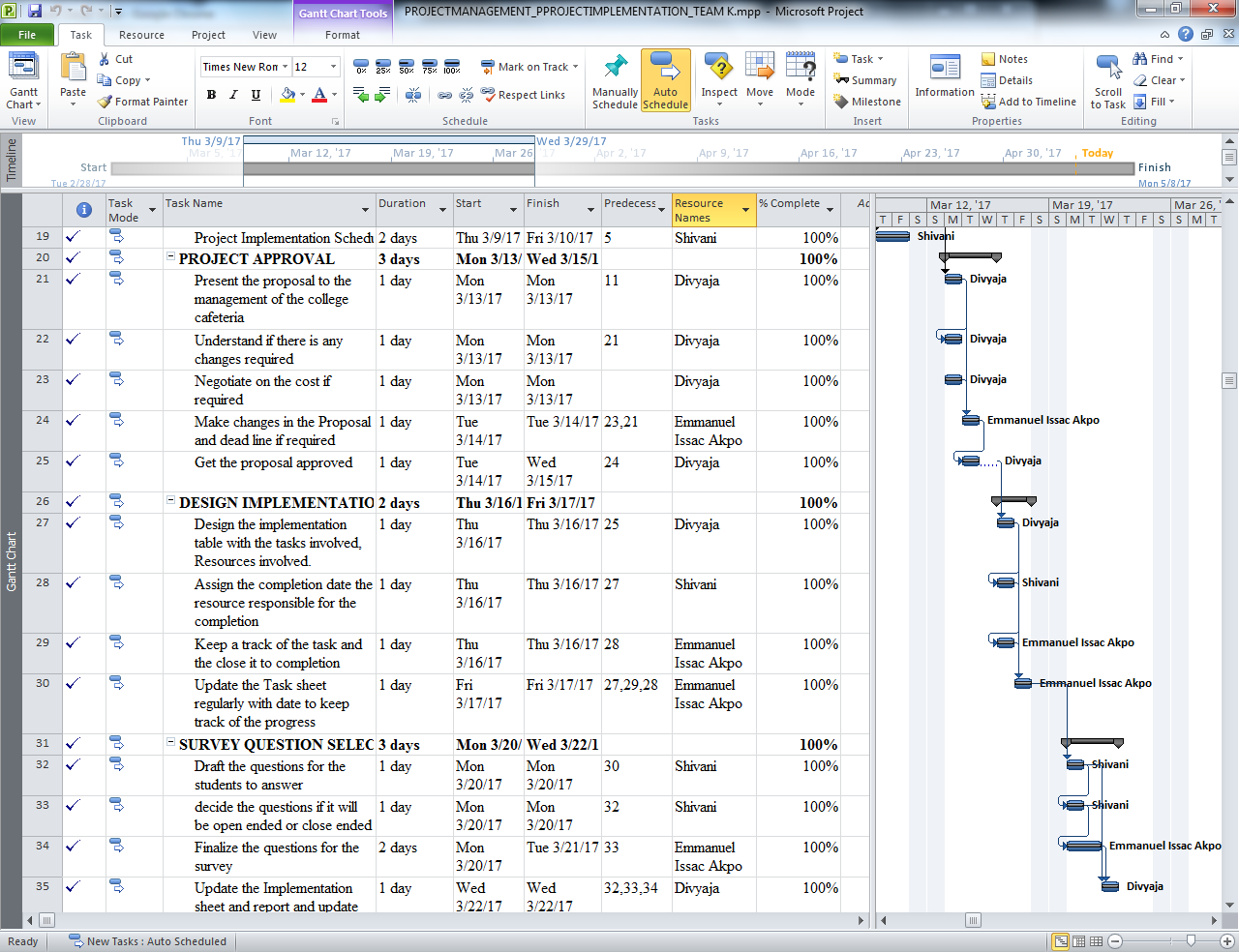
3. Inefficiency or running short of resources required to conduct the survey; which might result to the over burden of the other resources in the team.

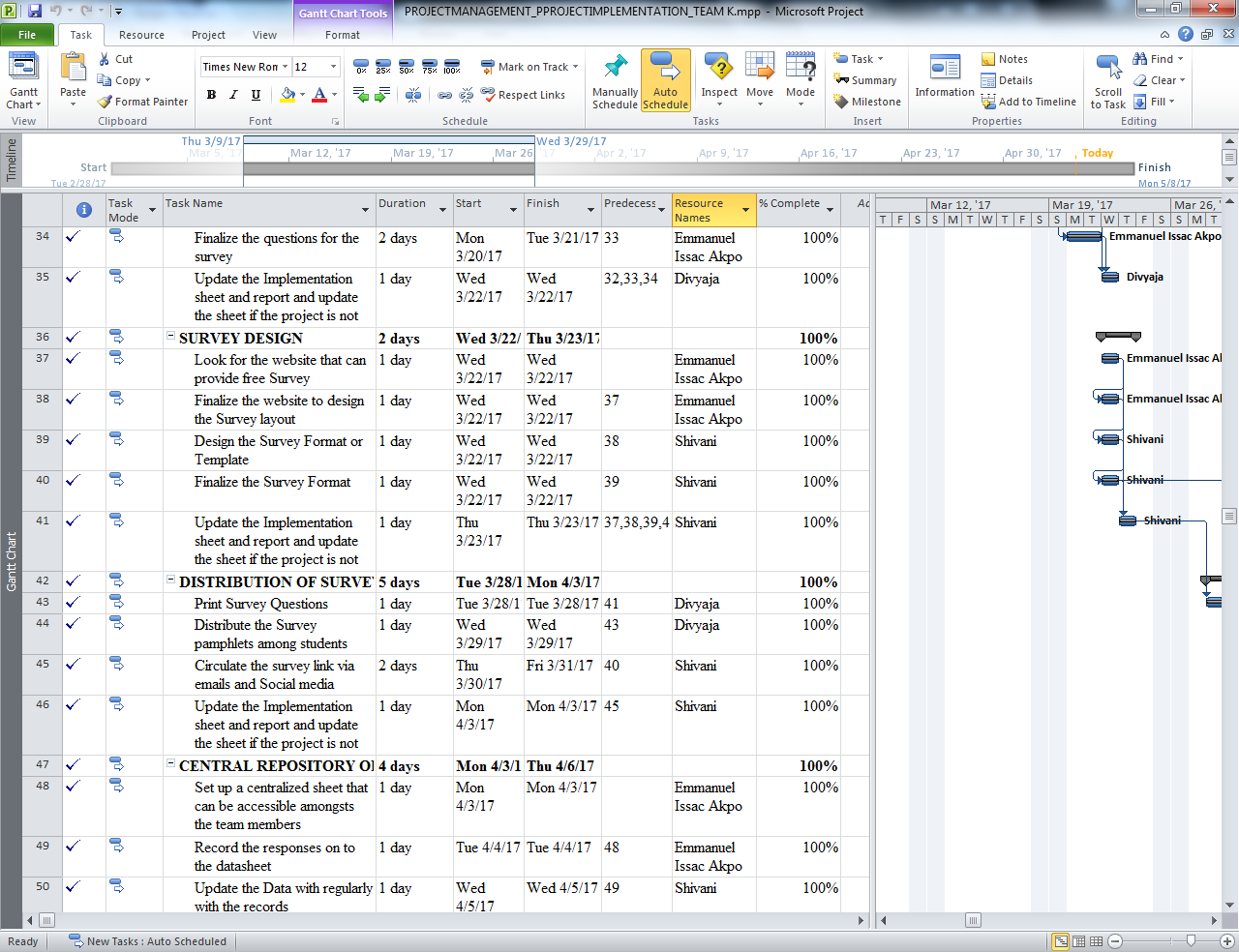
4. Issues with the installation of the application and management of the application; which might lead to find out a work around to perform the analysis.

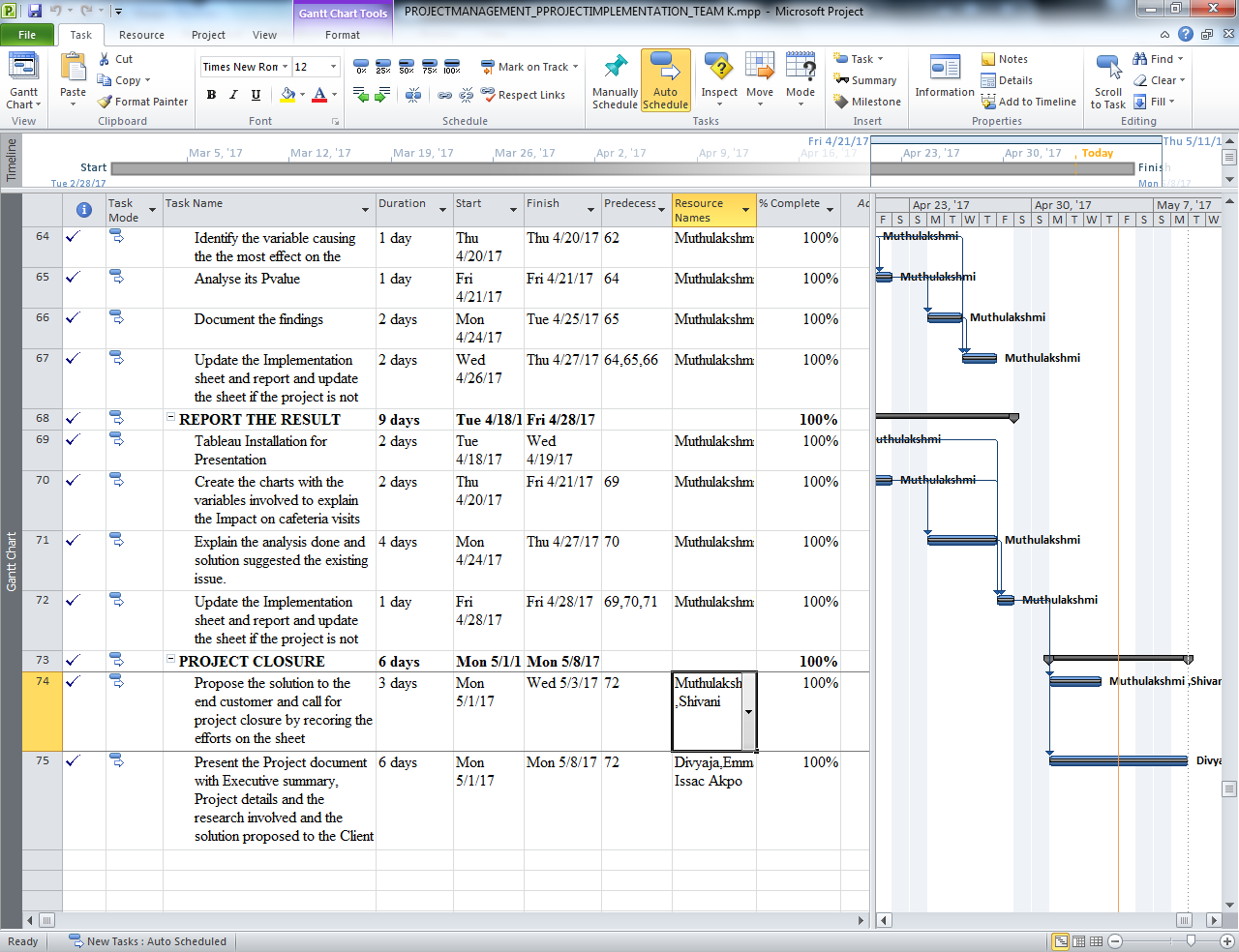
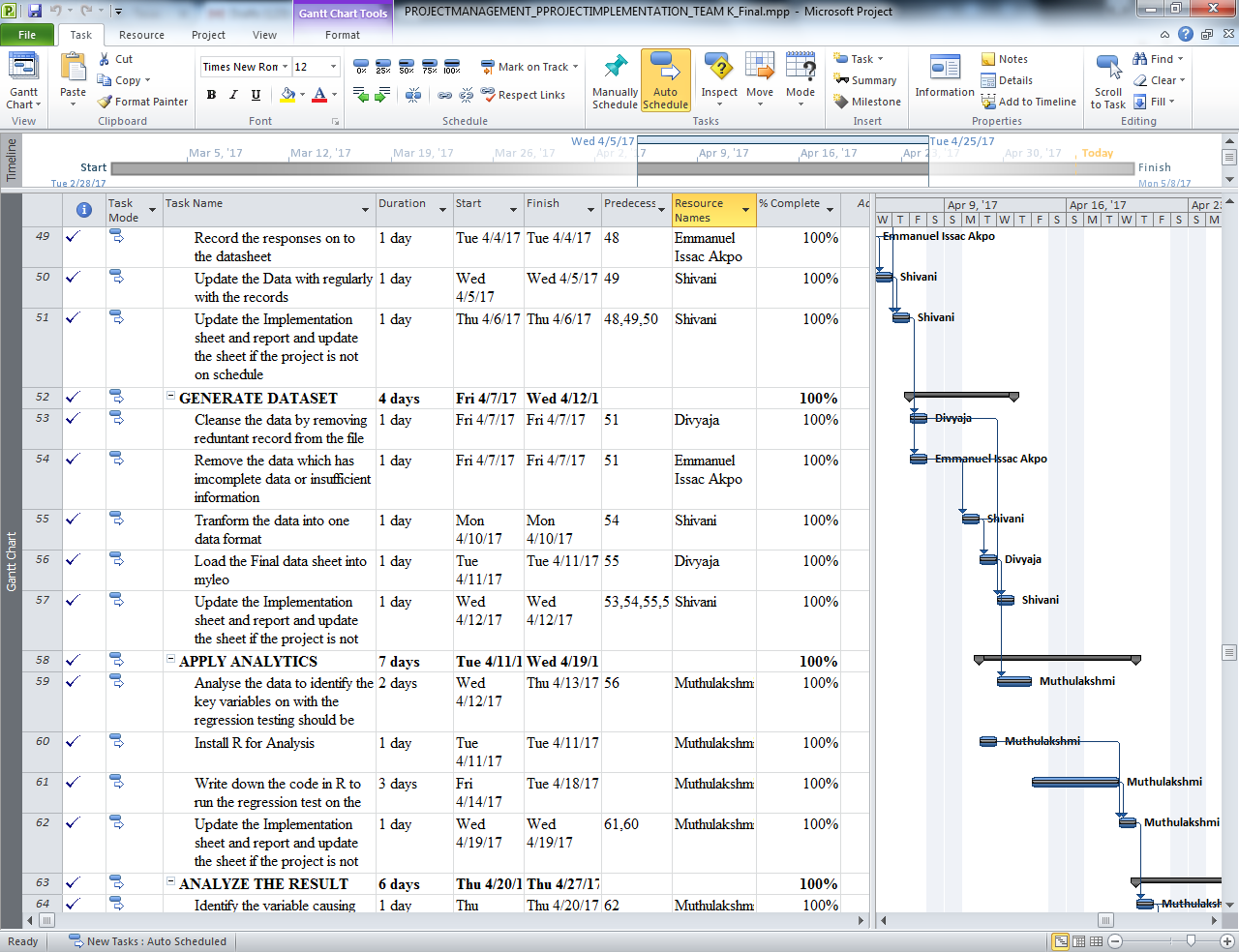
5. Shortage of Project cast due to incorrect budget calculation and allocation.

# WBS

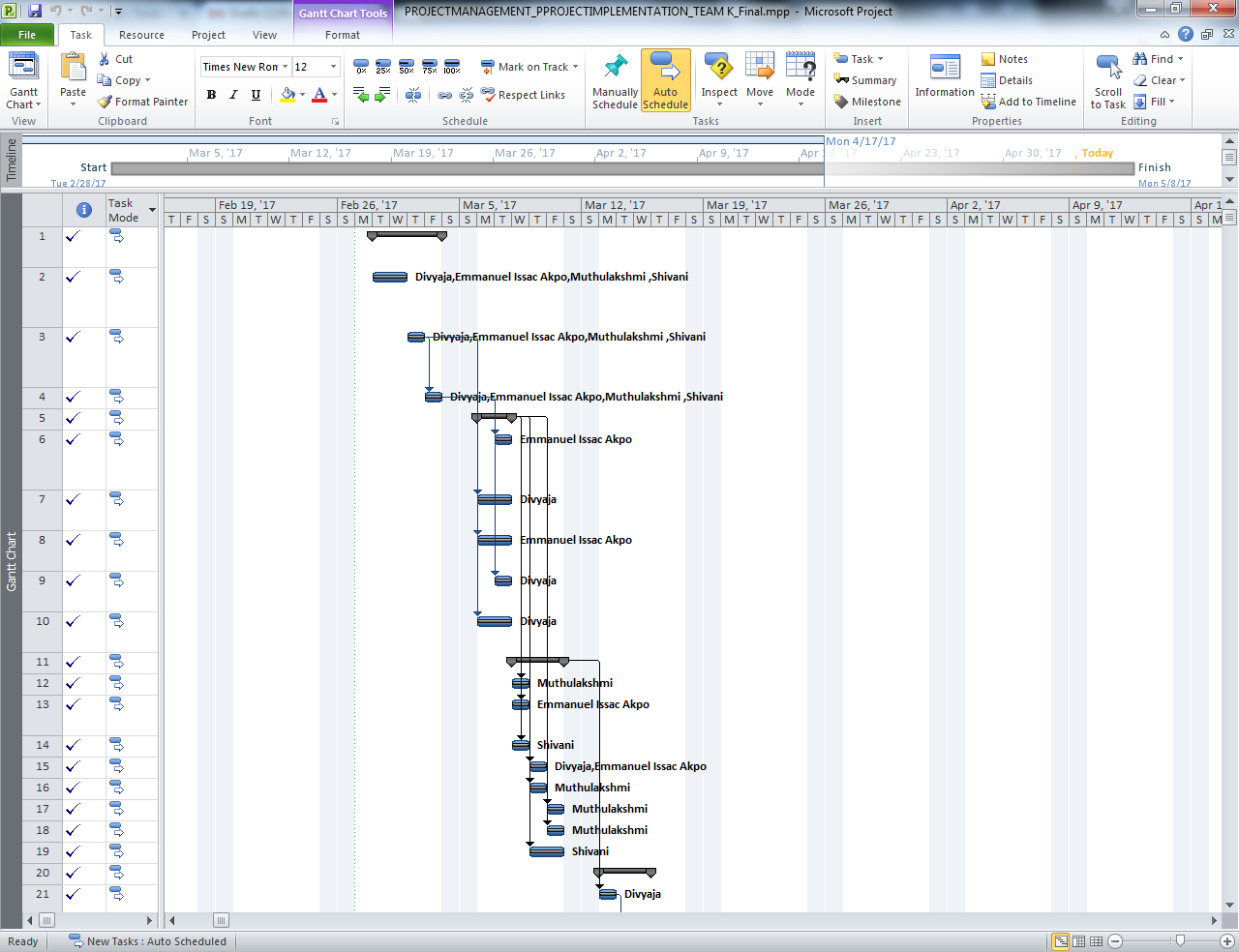


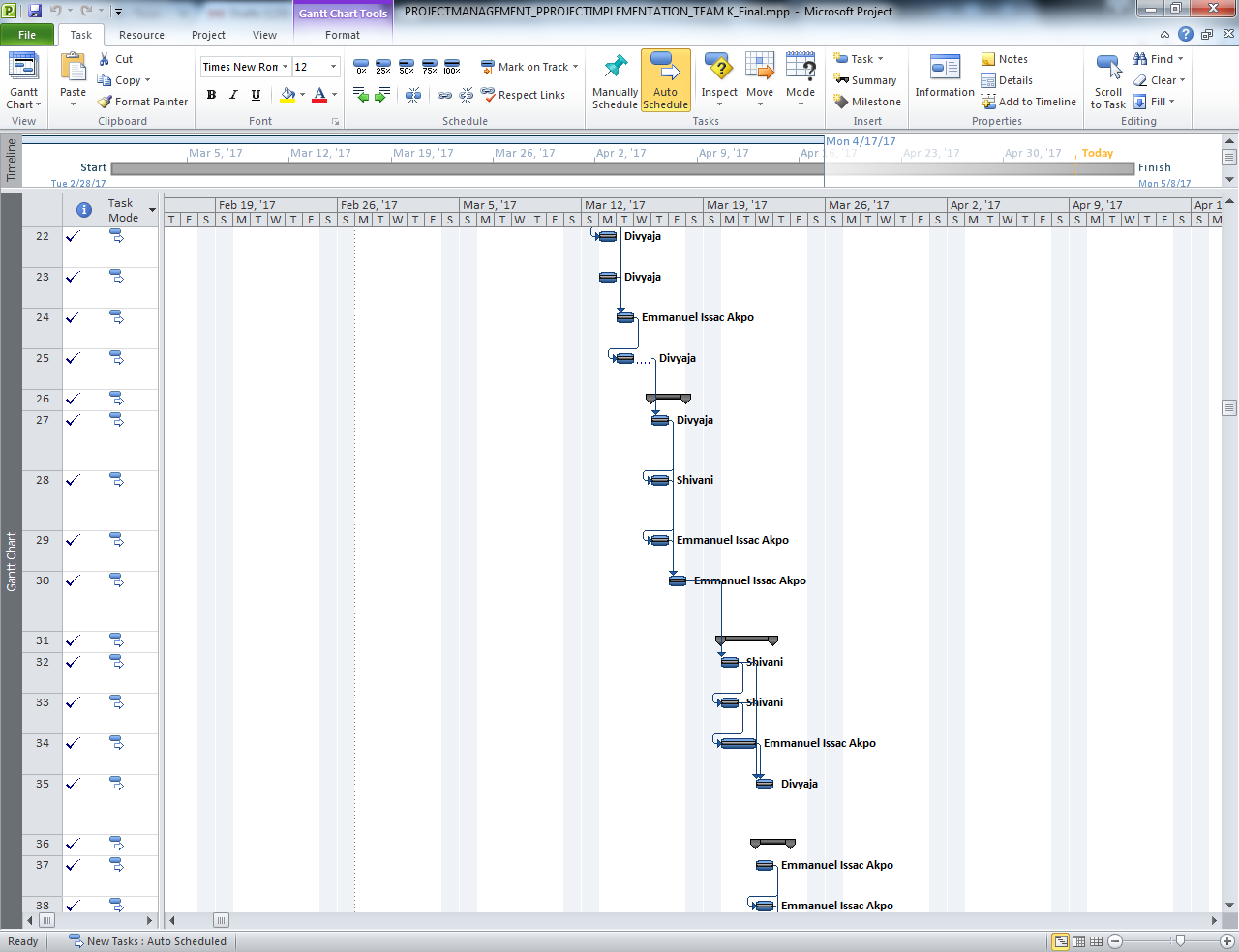


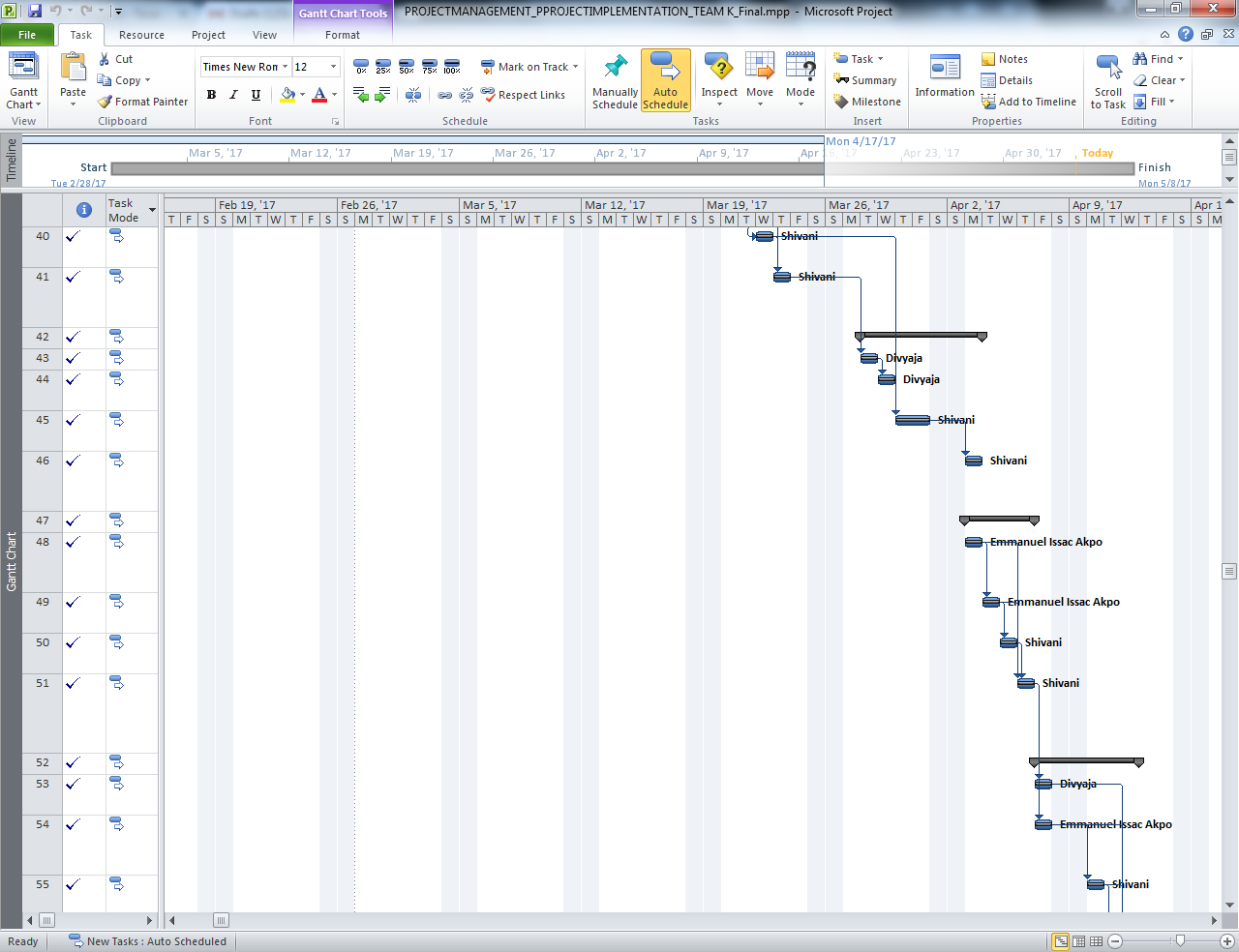


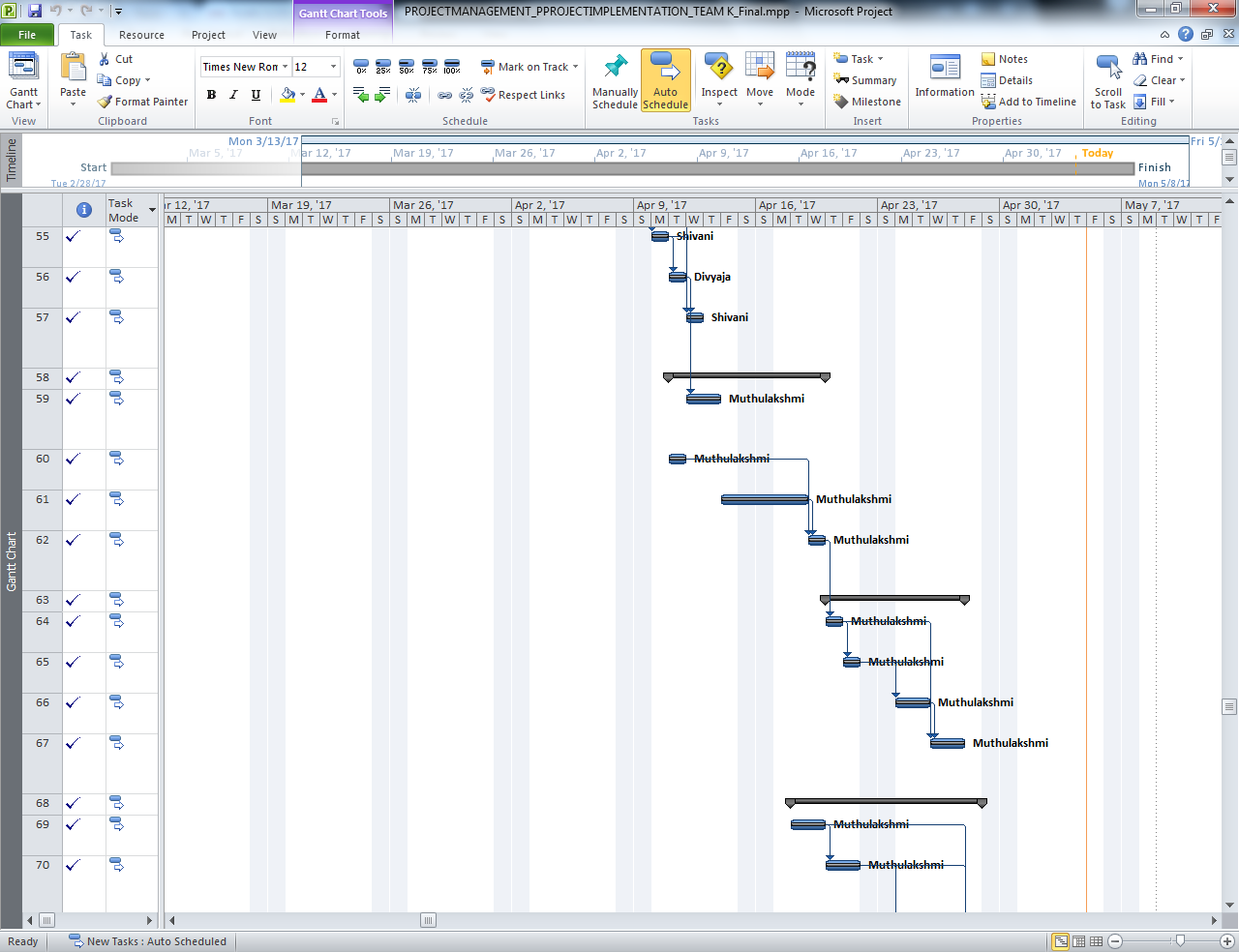


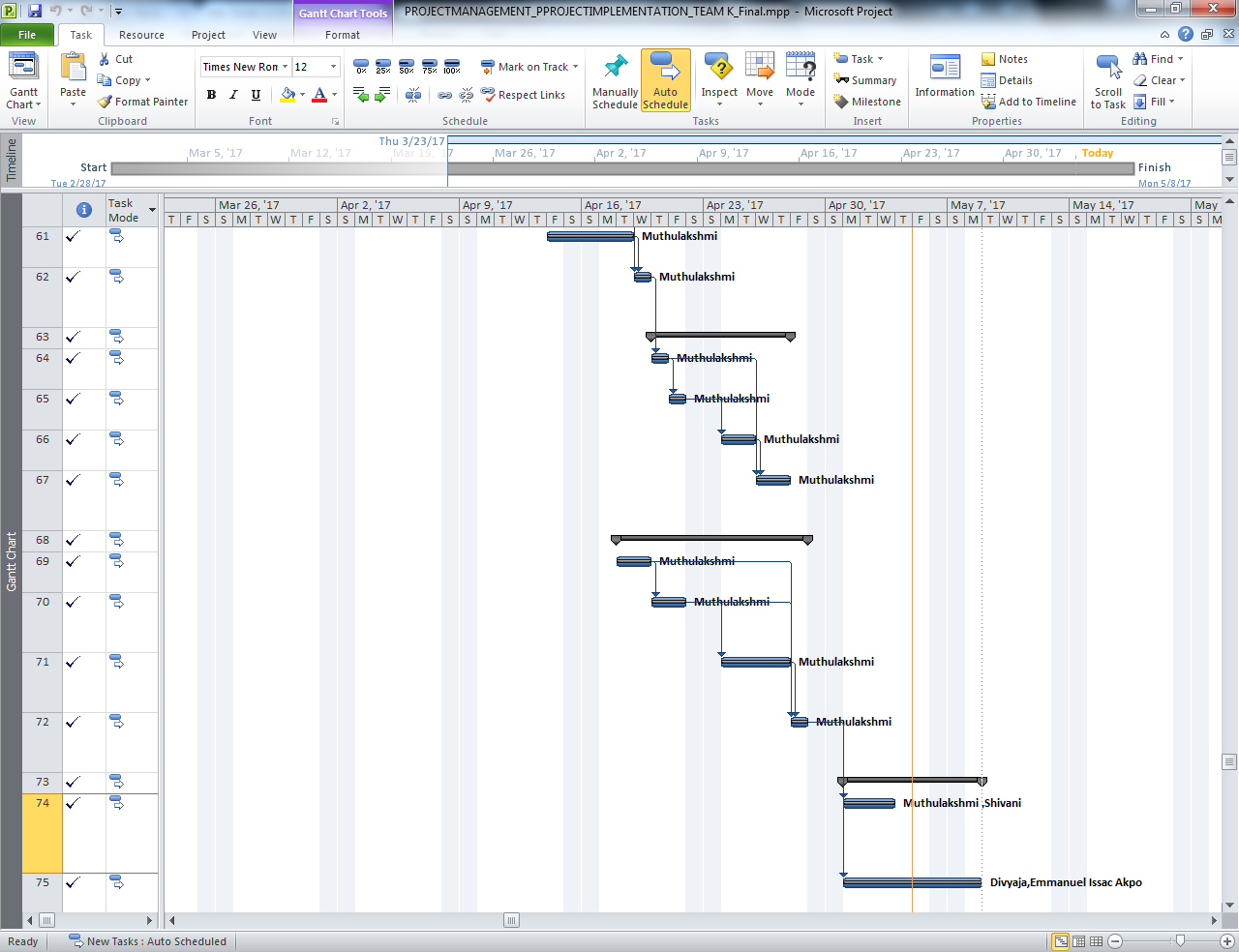
# Gantt Chart











# Methodology

This section explains about the approach considered to conduct the analysis to achieve the end result. Before working on the actual case a set of analysis was done to understand how other analyst have conducted the survey and reported the results and steps they took to overcome the obstacles. A literature review was done before taking up the project on a similar issue. Understanding the learning from similar project a careful study on the case was performed.

## Literature Review

## Satisfaction with food surveys:

A similar study was conducted at a college in northwestern Pennsylvania about university cafeteria to evaluate the overall customer service. It is imperative to conduct research about college students’ food service because offering quality education is paramount, many colleges who wish to retain their students, attention to food services on campus could be important. The research model was built from the available information and tested using data from survey conducted in college. The result showcased the three most important variables that explain student satisfaction include staff behavior, food quality and price. Also, there were some other significant variables like atmosphere, responsiveness and cleanliness. Attention to these variables could enable those in charge of food services to deliver great value and satisfaction. This would help the committee who is in charge of food services to deliver greater value and satisfaction to enhance students overall experience.

**A study on the preferences of College Cafeteria**

Thus as considering the purpose of the study we need to find out the dependent variable that we are planning to analyze. In our case the dependent variable is “Frequency visit” of the students to cafeteria as this defines the problem. Now we narrow down further in understanding the dependent variable that could influence the “Frequency visit”, upon brainstorming we could come up with key variables such as the Price, Taste, Quantity, Ambiance, Timing of the cafeteria, Services provided at the cafeteria, Quality of the food provided in the cafeteria, Job status of the individual, Preference of food these are some factors that can influence the visits of the students to the college cafeteria.

Now keeping these variables in kind a set of questions is drafted and it is ensured that the questions don’t affect any one in belonging to a particular background or to offend the survey taker. Once they are finalized the survey is distributed amongst students over emails and some surveys are given out to fill when students gets out of cafeteria. Thus the response of sample size of 100 were recorded through surveys then the results were documented in an excel sheet. Once the data was collected as a part of cleansing the incomplete records were removed. Then the data is standardized to a single format. The answers of the questions are reframed into 0s and 1s.And questions were grouped into categorical data as required.

As the research is more on the analysis of the value the research questions are below

Research Question 1: H0 : If a student owns a job his/her frequency to the café is more

Ha : The job of a student is not related with the frequency of his/her visit to the café;

Research Question2: H0: If a student is stays on campus his or her frequency is more

Ha: The place of stays doesn’t affect the frequency of visit to cafeteria.

Research question 3: H0: Cuisine in cafeteria affects the frequency of visit of students

Ha: Cuisine in cafeteria does not affect the frequency of visit of students

Research question 4: H0: Quality in cafeteria affects the frequency of visit of students

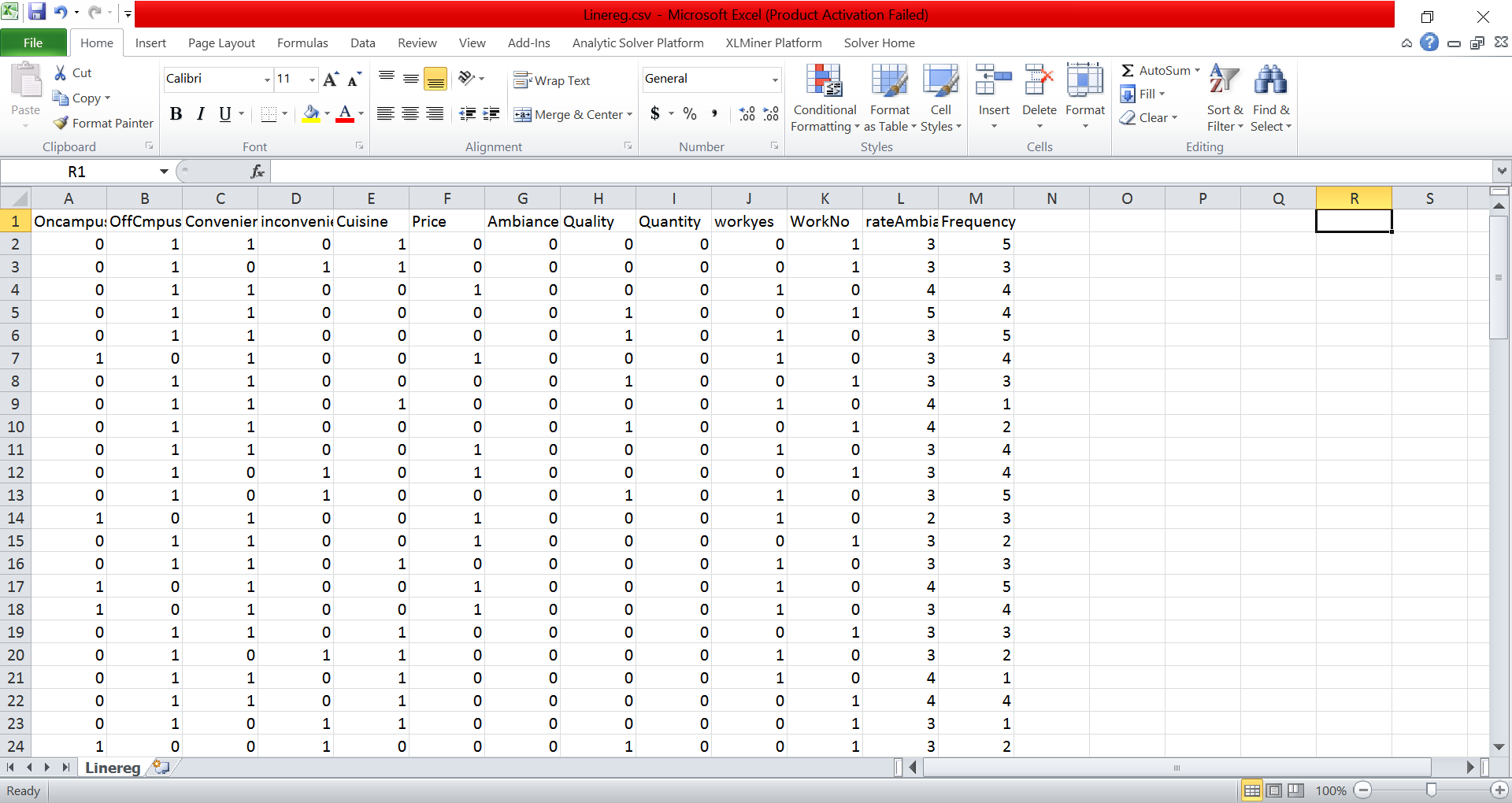
Ha: Quality in cafeteria does not affect the frequency of visit of students

These research questions are analyzed using linear regression analysis. As we know Y is the output variable the dependent variable and X is the independent variable. The independent variables considered are

Oncampus ;OffCmpus; Convenient; inconvenient Cuisine Price; Ambiance Quality Quantity; Frequency work; rateAmbiance; workyes;WorkNo

Using these dependent and independent variables regression analysis is done. To perform this we need the programming language R. Once the installation of the programming language is done we use R to analyze the data set.

The dataset used is as below is the set of the selected variables involved.

 Please find explanation of the variables as below:

stay on campus(yes)?: as variable **Oncampus**

stay on campus(No): as variable **OffCmpus**

Time convenient?(yes): as variable **Convenient**

Time convenient?(No): as variable in **Convenient**

Why do you prefer other places than cafeteria? as variable in **Cuisine**

Why do you prefer other places than cafeteria? as variable in **Ambiance**

Why do you prefer other places than cafeteria? as variable in **Quantity**

Why do you prefer other places than cafeteria? as variable in **Quality**

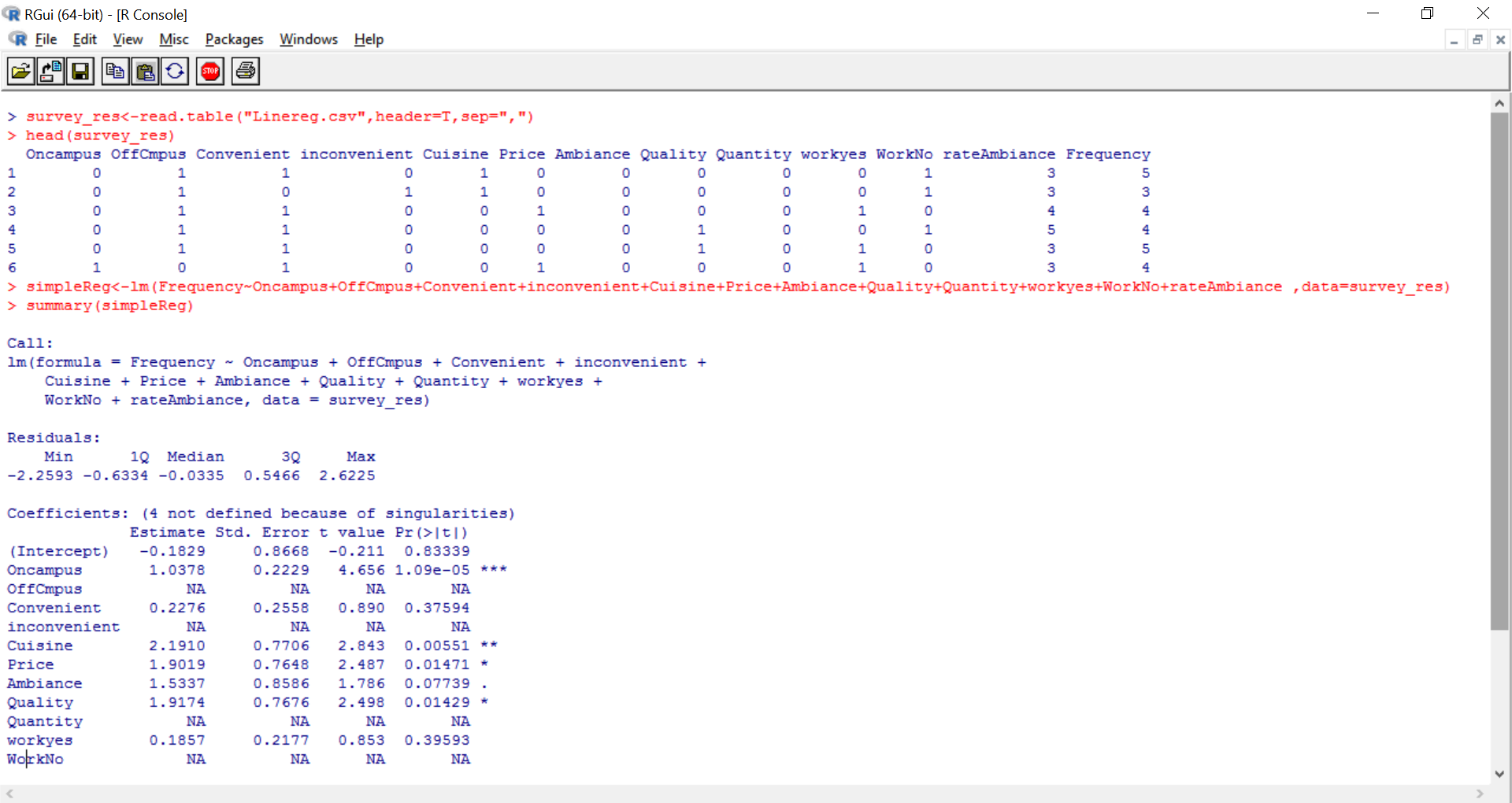
Why do you prefer other places than cafeteria? as variable in **Price**

Do you work ?(yes) as variable **workyes**

Do you work ?(No) as variable **WorkNo**

Do you like the ambiance? as variable **rateAmbiance**

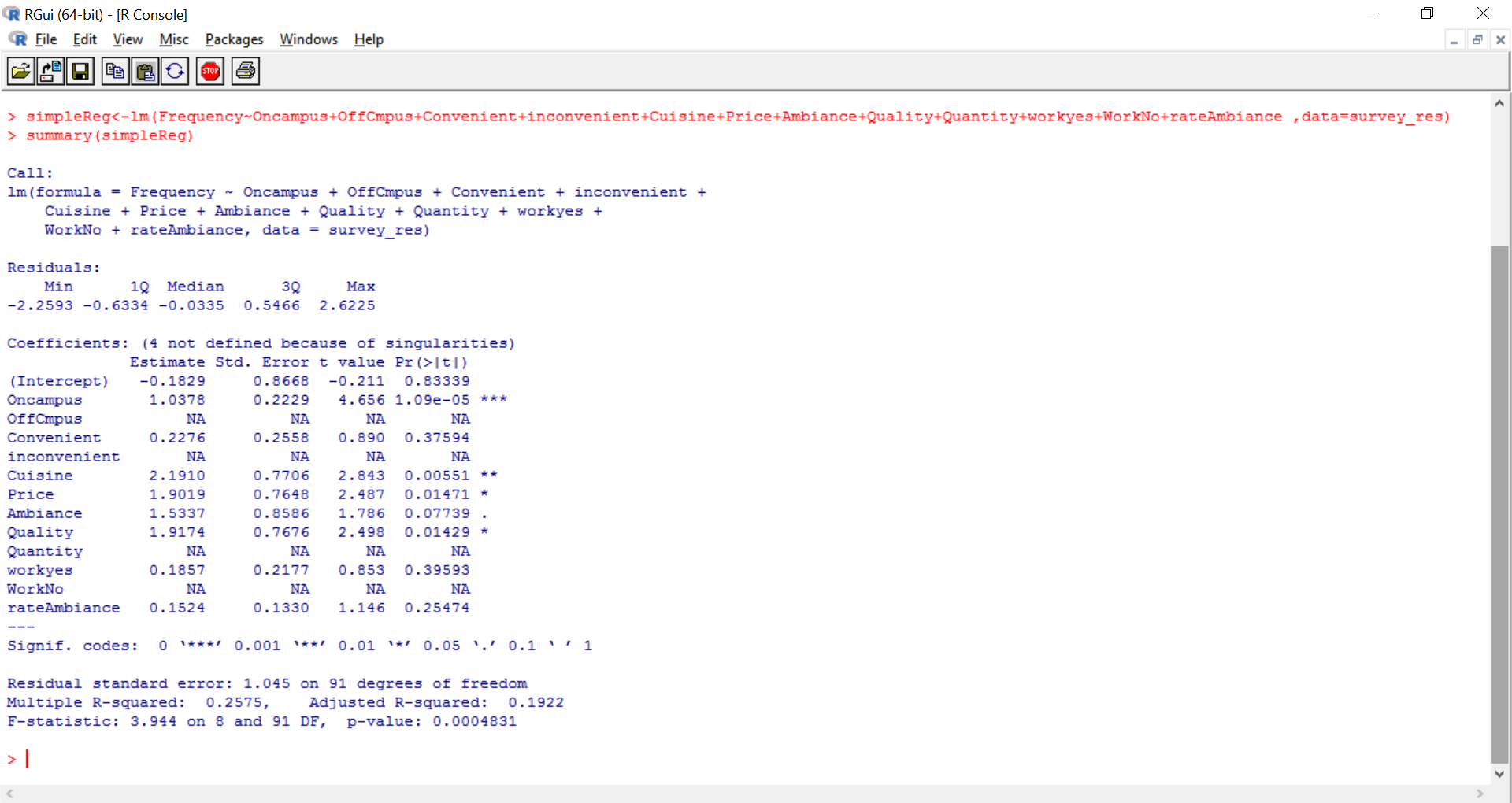
Once these variables added to the .csv file, we run the regression as below on R.



Once the regression test is run successfully the result is save for analysis. The survey data is also used for analyzing ranking the preference of cuisine using tableau.

# **Discussion**

For any data analytics model the interpretation of result is more important to understand so that one can come up with the resolution to fix the problem.

Below example shows the result of regression.

As we know frequency is the dependent and the other list of variables is independent variables.

The result shows the relationship between these variables the relationship is as below

Frequency = -0.1829 + 1.0378\*Oncampus + 0.2276\*Convenient + 2.1910\*Cuisine + 1.9019\*Price + 1.5337\*Ambiance + 1.9174\*Quality + 0.1857\*workyes + 0.1524\*rateAmbiance

Call:

lm(formula = Frequency ~ Oncampus + OffCmpus + Convenient + inconvenient +

Cuisine + Price + Ambiance + Quality + Quantity + workyes +

WorkNo + rateAmbiance, data = survey\_res)

Residuals:

Min 1Q Median 3Q Max

-2.2593 -0.6334 -0.0335 0.5466 2.6225

Coefficients: (4 not defined because of singularities)

Estimate Std. Error t value Pr(>|t|)

(Intercept) -0.1829 0.8668 -0.211 0.83339

Oncampus 1.0378 0.2229 4.656 1.09e-05 \*\*\*

OffCmpus NA NA NA NA

Convenient 0.2276 0.2558 0.890 0.37594

In convenient NA NA NA NA

Cuisine 2.1910 0.7706 2.843 0.00551 \*\*

Price 1.9019 0.7648 2.487 0.01471 \*

Ambiance 1.5337 0.8586 1.786 0.07739 .

Quality 1.9174 0.7676 2.498 0.01429 \*

Quantity NA NA NA NA

workyes 0.1857 0.2177 0.853 0.39593

WorkNo NA NA NA NA

rateAmbiance 0.1524 0.1330 1.146 0.25474

---

Signif. codes: 0 ‘\*\*\*’ 0.001 ‘\*\*’ 0.01 ‘\*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 1.045 on 91 degrees of freedom

Multiple R-squared: 0.2575, Adjusted R-squared: 0.1922

F-statistic: 3.944 on 8 and 91 DF, p-value: 0.0004831

The adjusted R- squared value is 0.1922 a unit change in the frequency of the visits of the students to the cafeteria is estimated by 19.2% percent of change in these independent variables. Note that the confidence level that we have considered for this project is as high as 99%. Thus the pvalue must be as low as possible for the variable to be significant. As shown in the above equation as we can see the intercept is negative. You can see that the oncampus has a positive relation with the frequency, which means any increase in the students stays in campus increases their frequency of visit to the Cafeteria by 1.0378; but on the other hand if the student stays off campus then their frequency of visit if impacted by **-0.1829.The p-value of this variable is very low 1.09e-05, which makes this variable “oncampus” very significant.**

Convenient timing is positively correlated by 0.2276 with frequency of the students visit who can make a visit to the cafeteria during the cafeteria hours while the hours impacts on the frequency is by -.1829 for the ones who couldn’t get to the their meal from cafeteria post the cafeteria hours. Further the P-value of this variable is quite high which is 0.375, leaving the variable insignificant.

The relationship above shows clearly that if the cafeteria improves the cuisine then it can increase the students’ frequency of visit by 2.1910, further analyzing at the P-value 0.00551 of this variable says that the value is significant.

Similarly, can see from the result that the price of the influences the visit of the frequency by 1.9019, now to figure out if the students are satisfied with the current price of the meal we add another variable called acceptable price and run the regression to understand the correlation between these two variables. The P-value of Price variable is 0.01471 which explains that the variable is significant and plays a very important role in determining frequency.

Further analyzing we can see that ambiance also has a positive impact on the frequency, it affects by 1.53; now if we further relate the variables we can understand that the increase in the quality of ambiance is essential to rate the ambiance which impacts the frequency by 0.152.But as the p value of both these variable is quite high 0.07739 and 0.25474 which makes them insignificant.

Similarly quality of the food has a positive influence on the frequency which is by 1.92.This is a very important variable as the p-value is 0.01429 which is acceptable. Now when we analyze the third variable we can see that the value of ‘workyes‘which shows that the frequency of the visits to the cafeteria increases if the students has a job by 0.1857. But as this variable has a P value of 0.39593, it makes the variable insignificant.

Now as we consider our research questions,

Research Question 1: H0 : If a student owns a job his/her frequency to the café is more

Ha : The job of a student is not related with the frequency of his/her visit to the café;

Now since the p-value of this variable is quite high which is 0.39593> a

We know alpha = 0.05, so, 0.39593> 0.05 we fail to reject the Null Hypothesis, which indicates that the null hypothesis may be true.

Research Question2: H0: If a student is stays on campus his or her frequency is more

Ha: The place of stays doesn’t affect the frequency of visit to cafeteria.

Here the p-value of on campus is very low 1.09e-05, which is greater than a =0.05, which means we reject the null as 1.09e-05<0.05. Hence, Ha : The place of stays doesn’t affect the frequency of visit to cafeteria is true. Below are the graphs which depict the relationship between the variables and frequency.

Research question 3: H0: Cuisine in cafeteria affects the frequency of visit of students

Ha: Cuisine in cafeteria does not affect the frequency of visit of students

Here the p-value of on campus is very low 0.00551, which is greater than a =0.05, which means we reject the null as 0.00551 <0.05. Hence, Ha : Cuisine in cafeteria affects the frequency of visit of students.

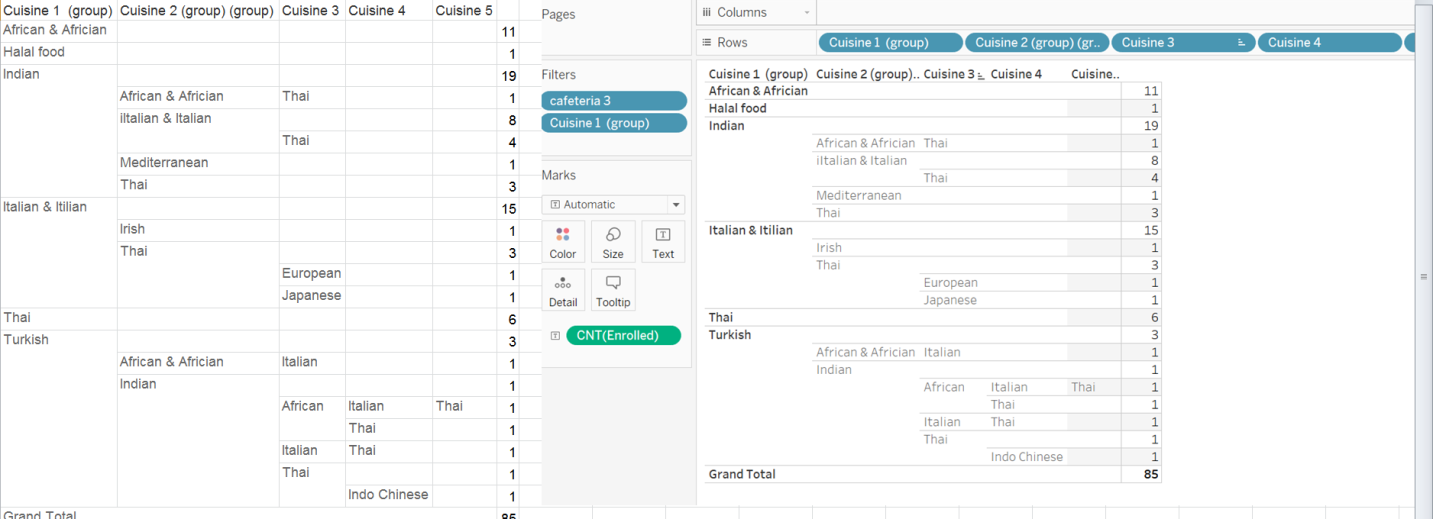
Research question 4: H0: Quality in cafeteria affects the frequency of visit of students

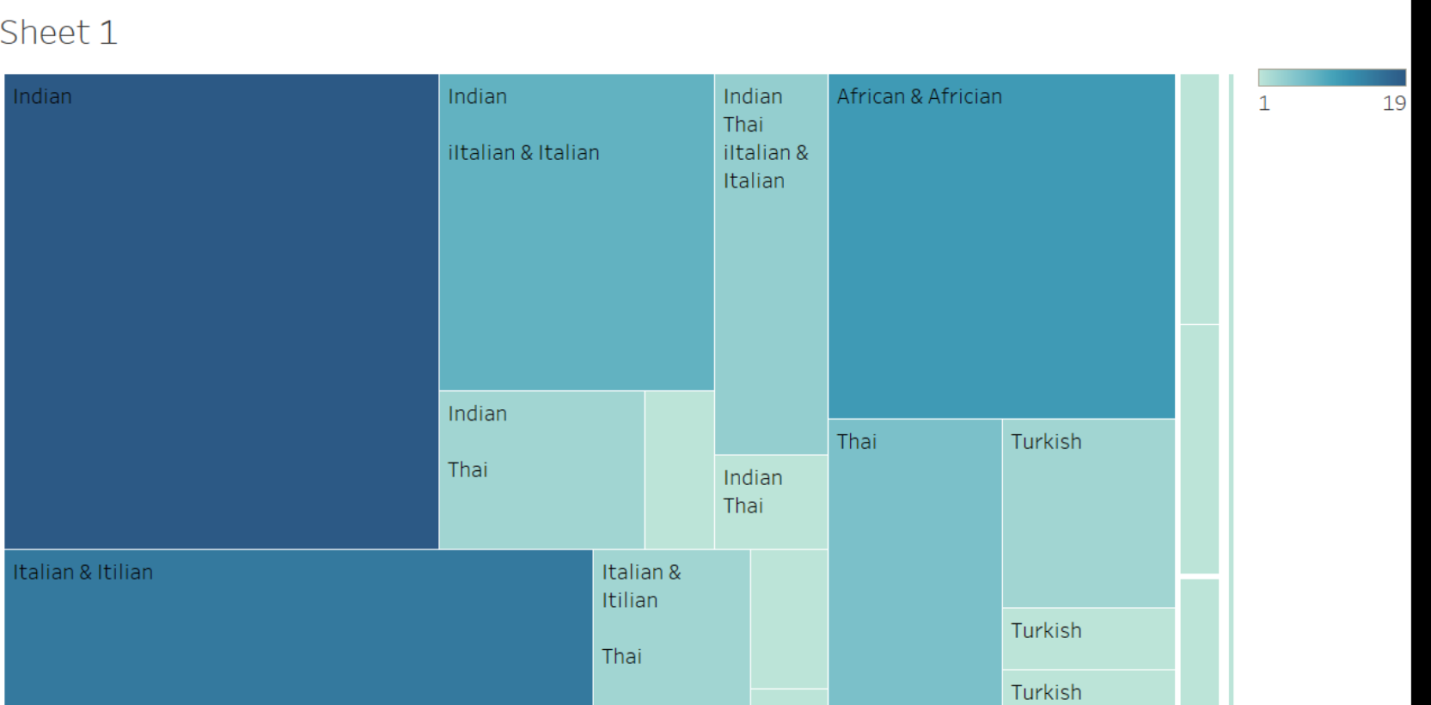
Ha: Quality in cafeteria does not affect the frequency of visit of students

Here the p-value of on campus is very low 0.01429, which is greater than a =0.05, which means we reject the null as 0.01429 <0.05. Hence, Ha : Quality in cafeteria affects the frequency of visit of students.

Below are the graphs which depict the relationship between the variables and frequency.

Using tableau we find group the below cuisine to rate the preferred cuisines. As shown in the below screen shot Indian food is preferred the most, followed by Thai and African American.





## Solution Proposed

Thus as per our analysis it is made clear that the main variables impacting the frequency of the students visits to the cafeteria are cuisine, price and quality. So it is suggested to give importance to these factors. By offering various cuisines we can increase the student visit and bring in more new students to have food at the café. By introducing new cuisines like Indian, Thai and African American we might attract more students to cafeteria. Secondly the price meal is handled without reducing the price by introducing a-la carte, so that in case you give choice to the student to pay for what they eat. Thus, not affecting the ones who prefer Quantity over Quality can still prefer Buffet. Another suggestion would be to introduce schemes where students who go as a group gets some food coupons. Thus the cafeteria will become the best place to have food with friends.

Thirdly, enhance the ambiance by having a karaoke stage so that students can also find the place refreshing and also enjoy their meal.

## Total Cost &Duration

Total cost involved in the project execution was a little more than the cost allocated as the survey was collected more through pamphlets than through the online survey links as the link was unavailable after at times. As it took more time in collecting the survey and recording them the number of hours spent by the resources in entering data , creating a datasheet and in the collection of the survey responses demanded for increase in the working hours of the employees.

The duration of the project increased secondly because of the issues faced during installation of the software with R. In the R software we had to add packages such as “ggplot“ but during this the system couldn’t reach the mirror servers for the download of the packages which caused a delay and several errors. Additionally in Tableau the licensing of this application for the students had to go through some approval process so it took a couple of days from the application support team to approve the license for the complete setup of the application. Due to these issues the Total Project hours overshadowed the estimated project hours by 24 ie. 3 days of work hours of a resource.

Cost involved in Resources:

Total number of resource required for the project: varies with individual:

(208+200+184+264)

Total number of hours required each resource (in hours): 832+3days ie.832+24 =856

Cost per hour involved in this project (in USD) $ : 20

Total number of cost per resource (in USD) $ : 20\*856 = 17120

Cost involved in tools:

R $ : 0

Tableau trail version (in USD) $ : 0

Total cost (in USD) $ : 0

Cost involved in Survey:

Cost involved in Printing 1 pamphlet (in USD) $ :0.10

Cost involved in printing 30 pamphlet (in USD) $ : 60\* 0.10

Total cost (in USD) $ : 6

Cost of the project in total :

Cost involved in Survey+ Cost involved in the tool +Cost involved in the Survey=17120+0+6 =$17126

Estimated Budget: $16643

Cost of the project: $17126

Difference in Cost: Estimated Budget - Cost of the project =$17126-$16643=$480

The project cost increased by 480$ and the delay in the project completion by 24 hours.

Thus total cost invested in the successful execution of the project is $17126.

**Learning from the Project**

In this project we had many learning that must be documented so that we can use them as reference and not commit the same mistakes while executing other projects in the future.

1. Firstly we must always add buffer time when we allocate time to every task. The estimated time of completion of the project must have more time than the actual completion.

2. While collecting data through surveys we must ensure we don’t distribute the survey to just one category of people. It is not just the number count of the survey that is considered. It is the sample size along with the divergence in the sample responses.

3. When the task is not equally distributed amongst the resources there is a good chance that we drain out the energy and the enthusiasm of the resource and the spirit in the team.

4. Any such setbacks in a team it is mandatory to inform and get it fixed.

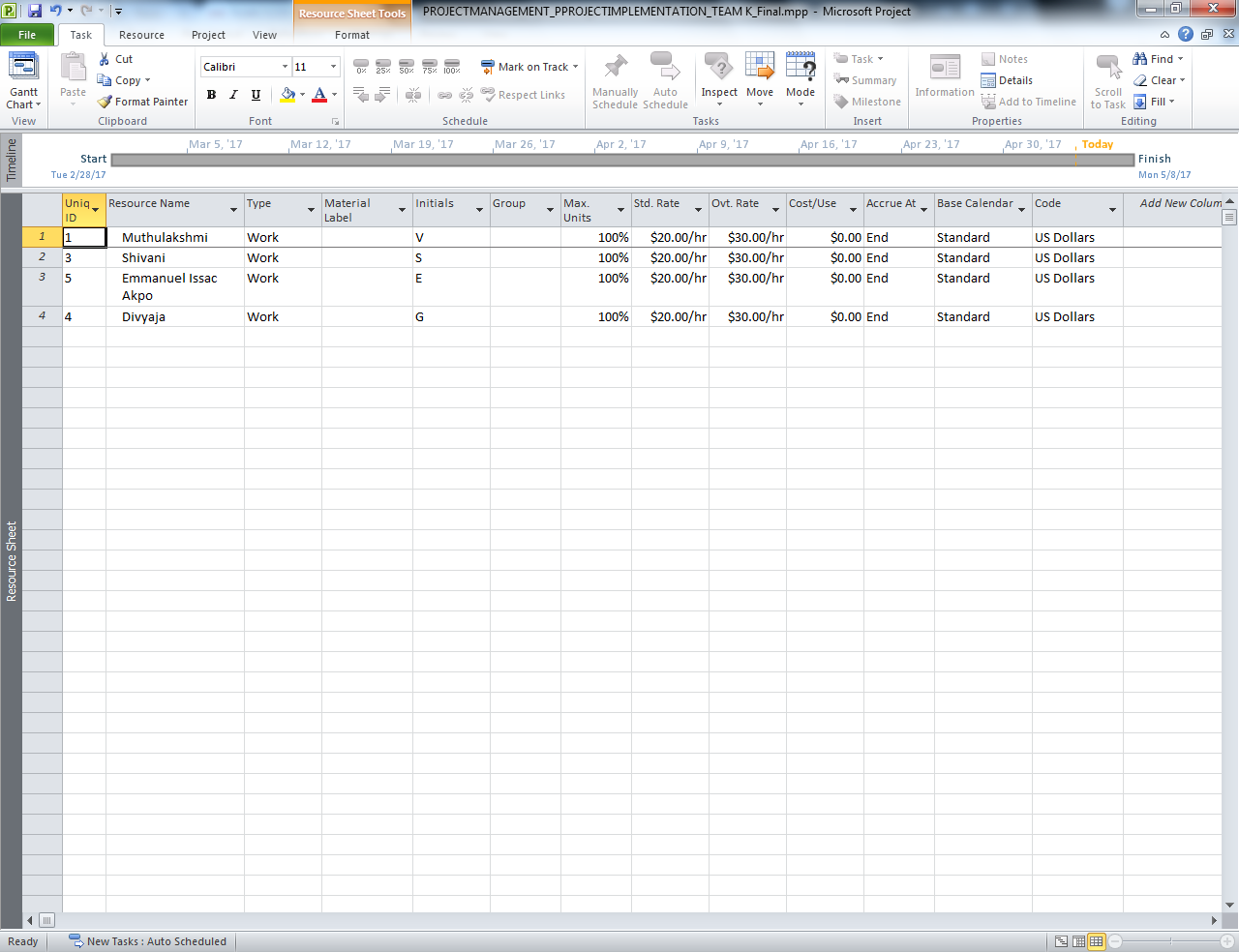
5. It is also good to have funds on serve that could be used in the project when the project runs shot of resource.

6. It is always good to do POC on the software in order to understand their setbacks, and have a clear idea of what is required for the software installation and get prepared.

**Conclusion**

This project provided a good insight about how the data can be used to record the feed of your consumers .It showed how Hypothesis testing and multiple regression testing can be used to understand the relation between the variables which are of concern. This project helped us to come up with solutions that when implemented will help overall business of the college cafeteria.

# Appendix

**Resource Table**

## Individual Resource hours