Analysis of facilities provided on Austin-Bergstrom International Airport

Name: Ranu Kishor Parate
Student ID: X17161452
Course: MSc in Data Analytics (Cohort A)
Subject: Analytical Customer Relationship Management
Lecturer: Prof. Vikas Sahni
Data of Submission: 03/04/2019

Analysis of facilities provided on Austin-Bergstrom International Airport

Ranu Parate
x17161452
MSc in Data Analytics (Cohort A)
National College of Ireland

Abstract - The analysis of the passenger's reviews and rating on the facilities provided on the airport plays an important part for contributing to the airport's economy. Thus, by analyzing this with the help of data mining techniques can help the airport management to take care of the facilities because of which the overall satisfaction rate is decreasing. In this report of analytical customer relationship management, it provides the data mining techniques that can tell the factors affecting the overall satisfaction rate of Austin-Bergstrom International Airport. The linear regression model is performed for the analysis. This analysis is done on the various facilities that the passengers can use by making use of analytical tool.

Keywords—linear regression, analysis, regression, spss, data mining, efficiency, customer relationship management, multiple linear regression

I. INTRODUCTION

In 21st century, individuals barely envision their life without travelling. Nowadays, travelling has become the lifestyle for people. Travelling around the world by using airways as a mode of transport is becoming common these days as the people prefer saving time by travelling with the airways, due to this there is the rapid change in the airport industry. Passengers generally spend around two to three hours on the airport before the departure. Also due to delays of the flights the passengers need to wait on airport.

On the airport the passenger makes use of facilities. There are many facilities which passengers use within the airport. The facilities such as parking facilities, washroom facilities, ground transportation, check-in wait time and many more facilities on the airport play an important role. The ratings and the views of the passengers regarding the facilities helps in increasing the economy of the airport.

In this project the analysis is done on the overall satisfaction rate provided by the passenger. The survey is carried out quarterly which provides the rating of all the facilities provided on the Austin-Bergstrom International Airport. For analyzing the overall satisfaction of the passenger, the multiple linear regressions are used, and this analysis is done on IBM SPSS tool.

A. Proposal

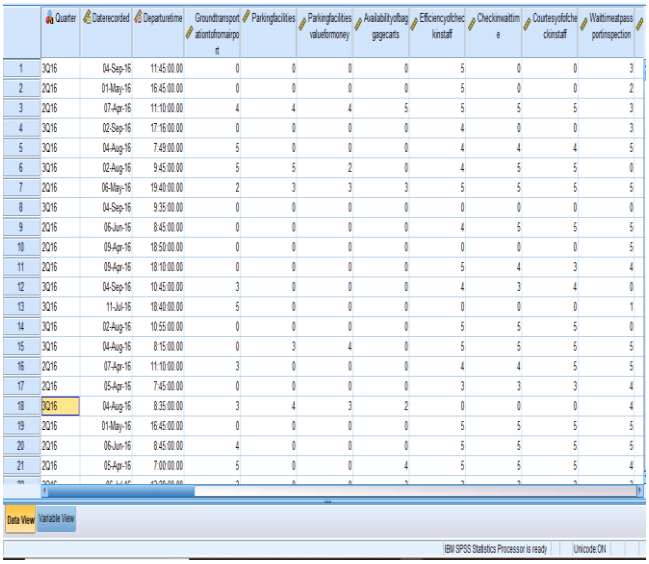
The main objective of this project is to analyze the facilities according to the ratings provided by the passenger. These are the facilities that are affecting the overall satisfaction rate quarterly for the year 2015 to 2017. Keeping in mind the customer satisfaction, this project will help the management of the airport to improve the facilities by which the overall satisfaction rate is decreasing.

B. Dataset

Dataset for analyzing the facilities provided on airport is taken from data.world website. This dataset provides the rating of passengers on different facilities. This data is the passenger survey conducted quarterly on facilities of Austin-Bergstrom International Airport's. The rating provided by the passengers is from 0 to 5 here 0 signifies the worst facility whereas 5 signifies the excellent facility. There are 3502 records and 37 attributes in the dataset.

The attributes consist of the facilities, departure date, departure time and the quarter of particular year.

The dataset for this analysis is downloaded from: https://catalog.data.gov/dataset/airport-quarterly-passenger-survey



	Quarter	Daterecorded	Departuretime	Groundtransportationtofromairport	Parkingfacilities	Parkingfacilitiesvalueformoney	Availabilityofbaggagecarts	Efficiencyofcheckinstaff	Checkinwaittime	Courtesyofcheckinstaff	Waittimeatpassportinspection
1	3Q16	04-Sep-16	11:45:00.00	0	0	0	0	5	0	0	3
2	2Q16	01-May-16	16:45:00.00	0	0	0	0	5	0	0	2
3	2Q16	07-Apr-16	11:10:00.00	4	4	4	5	5	5	5	3
4	3Q16	02-Sep-16	17:15:00.00	0	0	0	0	4	0	0	3
5	3Q16	04-Aug-16	7:49:00.00	5	0	0	0	4	4	4	5
6	3Q16	02-Aug-16	9:45:00.00	5	5	2	0	4	5	5	0
7	2Q16	06-May-16	19:40:00.00	2	3	3	3	5	5	5	5
8	3Q16	04-Sep-16	9:35:00.00	0	0	0	0	0	0	0	0
9	2Q16	06-Jun-16	8:45:00.00	0	0	0	0	4	5	5	5
10	2Q16	09-Apr-16	18:50:00.00	0	0	0	0	0	0	0	5
11	2Q16	09-Apr-16	18:10:00.00	0	0	0	0	5	4	3	4
12	3Q16	04-Sep-16	10:45:00.00	3	0	0	0	4	3	4	0
13	3Q16	11-Jul-16	18:40:00.00	5	0	0	0	0	0	0	1
14	3Q16	02-Aug-16	10:55:00.00	0	0	0	0	5	5	5	0
15	3Q16	04-Aug-16	8:15:00.00	0	3	4	0	5	5	5	5
16	2Q16	07-Apr-16	11:10:00.00	3	0	0	0	4	4	5	5
17	2Q16	05-Apr-16	7:45:00.00	0	0	0	0	3	3	3	4
18	3Q16	04-Aug-16	8:35:00.00	3	4	3	2	0	0	0	4
19	2Q16	01-May-16	16:45:00.00	0	0	0	0	5	5	5	5
20	2Q16	06-Jun-16	8:45:00.00	4	0	0	0	5	5	5	5
21	2Q16	05-Apr-16	7:00:00.00	5	0	0	4	5	5	5	4

Data View | Variable View

IBM SPSS Statistics Processor is ready | Unicode:ON

Figure 1: Dataset

Figure 1 shows the screenshot the dataset having quarter, date received, departure tome and various facilities. Some of the facilities used for analysis are:

- Ground transportation to/from airport
- Parking facilities
- Availability of baggage carts
- Check-in wait time
- Wait time at passport inspection
- Wait time of security inspection
- Restaurants
- Flight information screens
- Walking distance inside terminal
- Cleanliness of airport terminal
- Availability of washrooms
- Internet access
- Speed of baggage delivery

II. Business Strategies to be Implemented

The goal of this project is to find the correlation between the overall satisfactions and other facilities according to the ratings provided by the passengers. This will help to understand collaboration between the passengers and the various facilities. The multiple linear regression models are used to make the analysis and to generate the correlation among the facilities.

Based on the analysis done there are many facilities which need to be improved for increasing the overall satisfaction rate.

III. Related Work

Main aim of this paper [2] is for the progress and development of service quality provided on the airport. For this paper the passenger's expectation was investigated. In this paper the researchers have done the qualitative and quantitative research on approximately 1,000 airports. According to the researcher's the service quality provided on the airport is hierarchically constructed that have three important dimensions namely: interaction, diversion and function. The main objective of the quantitative study was to develop, design, implement and verify the test result. The data for this research was collected from the purchased list of frequent flyers. Here the data was analysis was done using confirmatory factor analysis and exploratory analysis.

The objective of the paper [3] is to provide the tool that will measure the satisfaction rate of passenger travelling with airways and, recognizing the service aspects that are provided on the terminal so that the airport can provide the proper facilities to the passenger. The dataset regarding passenger satisfaction was collected from the survey. This survey is based on the facilities provided on Lamezia Terme Airport. For this research the regression model is used. To perform this logistic regression on the facilities the independent variable considered was judgements on the services and the dependent variable was the overall satisfaction rate. According to the researcher's the logistic model built in the research can provide a better way to identify the required actions for improving the facilities and also to improve the customer relationship.

This paper [4] aims to add the information by enhancing the understanding and knowledge of service quality and the fare or price that affects the satisfaction rate of the passengers of the airlines providing all the facilities and airlines with low cost. The data was collected from both the airlines; full facilities provider and low-cost facilities. This collection of data was done with cross-sectional approach. 200 records were gathered from each type of airlines using sampling method. The conclusion and the result of the research is that the

airline providing full services are more grateful and satisfied compared to the low-cost airline. Moreover, the result of regression analysis on the low-cost airline illustrate that the service accuracy, punctuality and promptness, attitude and nature of employee, and the fare has significantly influenced the customer satisfaction. Where-as, in full service providing airlines the attitude and boldness of the employees, evidence and the fare or prices are important and significant predictors of customer satisfaction. This research concludes that in both type of airlines the quality of services particularly the attitude and nature of service staff and employees and the pricing of the airline requires more attention for building the customer satisfaction rate and the customer relationship.

The study examined in this paper [5] is about the quality of services provided on the aircraft transport that have effect on passenger's satisfaction. The research shows the factors having positive effect on the quality of service and the factors that have the maximum and minimum effect on the service quality present on international airport. The data used for this research is the sample of 270 records. The data that is analyzed is from the passengers of three different classes; economy, business and premium. The results got from the research propose that there are various attributes in-flight quality of service are most important according to the different classes. The main services which are discussed in this report are in-flight service, in-flight digital service and back-office operations. The research shows that the three services are significantly linked to the quality of international air travel. Additionally, the research shows that the services delivered by the airline companies plays an important on the passenger's satisfaction.

IV. Methodology

Methodology is the important part in the analytics to show several factors that get affected by the factor that is predictable. The data consists of the ratings provided by the passengers on the facilities that are present on the Austin-Bergstrom International airport. In this analysis the multiple linear regression analysis is used to identify the correlation of other facilities on the overall satisfaction of the passenger. In this dataset for applying multiple linear regression one dependent variable and various independent variables are considered. By applying multiple linear regression, the facilities by which the overall rating is getting low can be identify. By doing this the facility which is contributing more for the low overall rating can be improved and in this way the economy can be increased.

A. *Business Question(s)*

1. By which facilities the passenger's overall satisfaction rate is decreasing?

2. How the facilities that are having low rating can be improved to increase the economy?

3. Which facilities are responsible for low satisfaction rate of passengers?

B. *Hypothesis*

H_0: There is significant relation between the overall satisfaction rate and other facilities provided on the airport.
H_1: There is no significant relation between the overall satisfaction rate and other facilities provided on the airport.

V. Implementation

Before applying the multiple linear regression models to the dataset some assumption test need to be done. This test includes the checking of null values, outliers in spss software. After this test, the null values were replaced by the average of the respective columns in Microsoft Excel. Once the data is cleaned the dependent variable and independent variables are chosen. Here the dependent variable is the overall satisfaction rate and the independent variables are all the facilities for which the passengers have provided ratings. Once the multiple line regression is applied the correlation between the attributes can be analyze and accordingly the statistically significant attributes can be checked and the conclusion can be made whether to reject the null hypothesis or to accept it. After cleaning the dataset, the data is than imported in the spss tool for implementing the data mining techniques.

Multiple Linear Regression

The Multiple Linear Regression (MLR) is another form of linear regression analysis. It is one of the statistical techniques that make use of various dependent and independent variables or attributes. The aim of MLR is to build the linear

model relationship between the dependent and independent variables [1].

The formula of MLR is:

$$y_i = \beta_0 + \beta_{1x1} + \beta_{2x2} + \cdots + \beta_{pxip} + \varepsilon \text{ for } i = 1,2 \ldots n.$$

Before applying the MLR to the dataset some assumptions need to check. The assumptions are linearity, multicollinearity, homoscedasticity and normal distribution.

For performing the analysis, the data is imported in spss tool. After importing the data, in the variable view the necessary changes were made like changing the datatype, measures and labelling the columns. Finally, the multiple linear regressions are selected for applying it on the dataset. Following are the output received after applying the MLR in spss.

A descriptive statistics shown in figure 2 is the description of all the dependent and independent attributes. The general overview of the dataset is provided in the descriptive statistics. N here shows the total number of attributes or variable. So, 3501 is the total number of attributes. Mean represents the average of the attributes according to the dependent and independent variables. The standard deviation is also calculated according to the mean in the descriptive statistics.

Descriptive Statistics

	Mean	Std. Deviation	N
Overall satisfaction	1.83	2.158	3501
Ground transportation to/from airport	2.19	2.167	3501
Parking facilities	1.13	1.848	3501
Parking facilities (value for money)	1.02	1.695	3501
Availability of baggage carts	1.04	1.773	3501
Efficiency of check-in staff	3.78	1.704	3501
Check-in wait time	3.79	1.716	3501
Courtesy of of check-in staff	3.78	1.768	3501
Wait time at passport inspection	3.34	1.945	3501
Courtesy of inspection staff	3.44	1.890	3501
Courtesy of security staff	3.96	1.417	3501
Thoroughness of security inspection	4.08	1.263	3501
Wait time of security inspection	4.02	1.282	3501
Feeling of safety and security	4.19	1.194	3501
Ease of finding your way through the airport	4.50	.864	3501
Flight information screens	4.23	1.334	3501
Walking distance inside terminal	4.39	.907	3501
Ease of making connections	.35	1.225	3501
Courtesy of airport staff	3.59	1.853	3501
Restaurants	2.97	1.944	3501
Restaurants (value for money)	2.56	1.802	3501
Availability of banks/ATM/money changing	.89	1.690	3501
Shopping facilities	1.89	1.997	3501
Shopping facilities (value for money)	1.55	1.775	3501
Internet access	1.90	1.936	3501
Business/executive lounges	.47	1.302	3501
Availability of washrooms	3.91	1.423	3501
Cleanliness of washrooms	3.80	1.513	3501
Comfort of waiting/gate areas	4.00	1.018	3501
Cleanliness of airport terminal	4.37	.831	3501
Ambience of airport	4.23	.876	3501
Arrivals passport and visa inspection	2.66	2.151	3501
Speed of baggage delivery	.92	1.689	3501
Customs inspection	1.32	1.878	3501

Figure 2: Descriptive Statistics

Correlations

		Overall satisfaction	Ground transportation to/from airport	Parking facilities	Parking facilities (value for money)	Availability of baggage carts	Efficiency of check-in staff	Check-in wait time	Courtesy of of check-in staff
Pearson Correlation	Overall satisfaction	1.000	-.011	.008	.019	.017	.071	.061	.052
	Ground transportation to/from airport	-.011	1.000	.251	.257	.241	.078	.099	.106
	Parking facilities	.008	.251	1.000	.905	.245	.071	.077	.077
	Parking facilities (value for money)	.019	.257	.905	1.000	.249	.069	.070	.065
	Availability of baggage carts	.017	.241	.245	.249	1.000	.127	.153	.154
	Efficiency of check-in staff	.071	.078	.071	.069	.127	1.000	.819	.823
	Check-in wait time	.061	.099	.077	.070	.153	.819	1.000	.864
	Courtesy of of check-in staff	.052	.106	.077	.065	.154	.823	.864	1.000
	Wait time at passport inspection	.087	.073	.067	.078	.109	.236	.238	.228
	Courtesy of inspection staff	.071	.094	.055	.061	.119	.239	.247	.258
	Courtesy of security staff	.070	.083	.055	.052	.075	.294	.293	.303
	Thoroughness of security inspection	.063	.085	.056	.059	.076	.339	.341	.334
	Wait time of security inspection	.095	.081	.051	.067	.059	.378	.370	.334
	Feeling of safety and security	.067	.078	.044	.050	.073	.323	.328	.318
	Ease of finding your way through the airport	.042	.025	.035	.033	-.019	.145	.142	.140

Figure 3: Correlations

Figure 3 gives the information about the correlations or the relationship among the dependent and independent variables. Here the dependent variable is Overall Satisfaction and all other are independent variables. The significant correlations are shown in the correlation table.

Variables Entered/Removed[a]

Model	Variables Entered	Variables Removed	Method
1	Arrivals passport and visa inspection	.	Stepwise (Criteria: Probability-of-F-to-enter <= .050, Probability-of-F-to-remove >= .100).
2	Ambience of airport	.	Stepwise (Criteria: Probability-of-F-to-enter <= .050, Probability-of-F-to-remove >= .100).
3	Speed of baggage delivery	.	Stepwise (Criteria: Probability-of-F-to-enter <= .050, Probability-of-F-to-remove >= .100).
4	Cleanliness of airport terminal	.	Stepwise (Criteria: Probability-of-F-to-enter <= .050, Probability-of-F-to-remove >= .100).

Figure 4: Variables Entered/Removed

Figure 4 shows the information on the variables which are entered or removed for performing the analysis. This output is generated when stepwise method is used for performing the regression. Variables entered show the name of variables which are included in the regression. Variables removed show the names of variables which are not considered for the regression. And last column Method it gives the information about the method used in the analysis.

Model Summary[l]

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate	Change Statistics				Durbin-Watson
					R Square Change	F Change	df1	df2	Sig. F Change	
1	.876[a]	.767	.767	1.042	.767	11517.428	1	3499	.000	
2	.898[b]	.807	.807	.948	.040	730.248	1	3498	.000	
3	.907[c]	.823	.823	.908	.016	314.593	1	3497	.000	
4	.909[d]	.826	.826	.901	.003	51.706	1	3496	.000	
5	.909[e]	.827	.827	.898	.001	25.168	1	3495	.000	
6	.910[f]	.828	.828	.895	.001	26.931	1	3494	.000	
7	.911[g]	.829	.829	.893	.001	18.359	1	3493	.000	
8	.911[h]	.830	.830	.891	.001	15.115	1	3492	.000	
9	.911[i]	.830	.830	.890	.000	9.938	1	3491	.002	
10	.911[j]	.831	.830	.889	.000	6.103	1	3490	.014	
11	.912[k]	.831	.830	.889	.000	4.207	1	3489	.040	1.987

Figure 5: Model Summary

Model summary shown in figure 5 gives the information about the model that is how suitable the model is. This suitability of the model is shown by the calculated values of R, RSquare, Standard deviation error, etc.

R value in the model summary indicates the simple correlations. This correlation is the relationship between the predicted value by the model and the value which is observed by the outcome.

RSquare is the amount of variance that is calculated by the model. It shows the total variance of the dependent attribute that can be explained by the independent attributes.

Adjusted R Square is the estimated value of the R Square. It provides the overview of how well the model is build. As R Square value is very high due to a greater number of independent variables then the adjusted R Square value is calculated.

In the model summary the Durbin Watson Test value is also calculated. By this test it gives the information whether the hypothesis of independent error is acceptable. Here the calculated value of Durbin Watson Test is 1.987.

Figure 6 gives the information about the coefficients. The information provided in the model column is the predictor variables.

The b-values in the coefficient table give information about that degree to which it affects the outcome of predictor if all other predictors are held constant.

The coefficient table also tells the significance value. If the significance value is less than 0.05, it means it is statistically significant.

Coefficients[a]

Model		Unstandardized Coefficients		Standardized Coefficients			Correlations			Collinearity Statistics	
		B	Std. Error	Beta	t	Sig.	Zero-order	Partial	Part	Tolerance	VIF
1	(Constant)	4.170	.028		148.953	.000					
	Arrivals passport and visa inspection	-.879	.008	-.876	-107.319	.000	-.876	-.876	-.876	1.000	1.000
2	(Constant)	2.128	.080		26.696	.000					
	Arrivals passport and visa inspection	-.902	.007	-.899	-120.267	.000	-.876	-.897	-.893	.987	1.013
	Ambience of airport	.497	.018	.202	27.023	.000	.099	.416	.201	.987	1.013
3	(Constant)	1.944	.077		25.223	.000					
	Arrivals passport and visa inspection	-.818	.009	-.815	-94.941	.000	-.876	-.849	-.675	.687	1.456
	Ambience of airport	.446	.018	.181	24.940	.000	.099	.389	.177	.961	1.041
	Speed of baggage delivery	.194	.011	.152	17.737	.000	.604	.287	.126	.692	1.446
4	(Constant)	1.641	.087		18.792	.000					
	Arrivals passport and visa inspection	-.821	.009	-.818	-95.877	.000	-.876	-.851	-.677	.685	1.461
	Ambience of airport	.342	.023	.139	14.920	.000	.099	.245	.105	.577	1.733
	Speed of baggage delivery	.190	.011	.149	17.542	.000	.604	.284	.124	.690	1.448
	Cleanliness of airport terminal	.173	.024	.066	7.191	.000	.081	.121	.051	.583	1.715
5	(Constant)	1.643	.087		18.871	.000					
	Arrivals passport and visa inspection	-.799	.010	-.796	-83.077	.000	-.876	-.815	-.585	.539	1.855
	Ambience of airport	.339	.023	.138	14.869	.000	.099	.244	.105	.577	1.734

Figure 6: Coefficients

Figure 7 shows the ANOVA table. The information in the ANOVA table is divided between Regression, Residual and total. Here Sum of Square is related with the three sources of variance. Degree of Freedom is denoted by df that is related with the three sources of variance. df for total variance is calculated as N-1, N is the total number of variances.

ANOVA[a]

Model		Sum of Squares	df	Mean Square	F	Sig.
1	Regression	12503.380	1	12503.380	11517.428	.000[b]
	Residual	3798.533	3499	1.086		
	Total	16301.913	3500			
2	Regression	13159.413	2	6579.707	7324.047	.000[c]
	Residual	3142.500	3498	.898		
	Total	16301.913	3500			
3	Regression	13418.782	3	4472.927	5425.292	.000[d]
	Residual	2883.131	3497	.824		
	Total	16301.913	3500			
4	Regression	13460.802	4	3365.201	4140.895	.000[e]
	Residual	2841.111	3496	.813		
	Total	16301.913	3500			
5	Regression	13481.116	5	2696.223	3340.651	.000[f]
	Residual	2820.797	3495	.807		
	Total	16301.913	3500			
6	Regression	13502.692	6	2250.449	2809.019	.000[g]
	Residual	2799.221	3494	.801		
	Total	16301.913	3500			
7	Regression	13517.327	7	1931.047	2422.316	.000[h]
	Residual	2784.586	3493	.797		
	Total	16301.913	3500			
8	Regression	13529.328	8	1691.166	2129.980	.000[i]
	Residual	2772.585	3492	.794		

Figure 7: ANOVA

Mean Square in ANOVA is the sum of squares that is divided by its respective df (degree of freedom). Mean Square (Regression) when divided by Mean Square (Residuals) the F value is generated.

VI. Visualizations

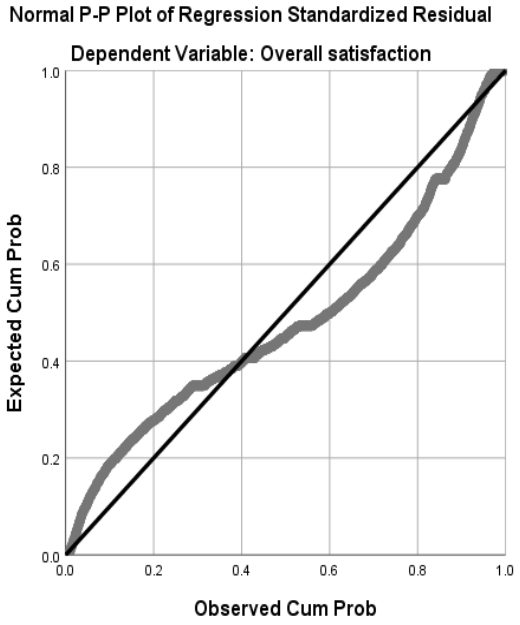


Figure 8: p-p plot

Figure 8 shows the p-p plot. The p-p plot is not linear there are many outliners which are lying outside the linear line. As most of the points are outside the linearity this means that the assumptions are not met.

VII. Result and Conclusion

From the implementation of the multiple linear regressions using spss it can be concluded that the null hypothesis is not rejected. This means the significant relation exists between the overall satisfaction rate and other facilities provided on the airport.

There are various many facilities that are statistically significant and there is a correlation between the facilities and overall satisfaction rate.

The facilities which are statistically significant can be improved so that the overall satisfaction rate will get increase and the passenger will enjoy the facilities provided on the airport.

VIII. References

[1] Investopedia (2019) *Multiple Linear Regression – MLR Definition* [Online] Available at: https://www.investopedia.com/terms/m/mlr.asp [Accessed 23 March 2019].

[2]. Fodness, D. and Murray, B. (2007) 'Passengers' expectations of airport service quality'. Journal of Services Marketing, 21(7): pp. 492-506, Emerald [Online] Available at: https://doi.org/10.1108/08876040710824852 [Accessed 27 March 2019].

[3]. Eboli, L., and Mazzulla, G. (2009) 'An ordinal logistic regression model for analyzing airport passenger satisfaction'. EuroMed Journal of Business, 4(1): pp. 40–57, Available at: https://doi:10.1108/14502190910956684 [Accessed 27 March 2019]

[4]. Suhartanto, D. and Noor, A. (2012) 'Customer satisfaction in the airline industry: The role of service quality and price', Asia Tourism Forum. Indonesia, May 2012, pp. 1-9.

[5]. Archana, R. and Subha, M (2012) 'A Study on service quality and passenger satisfaction on Indian Airlines'. International Journal of Multidisciplinary Research, 2(2): pp. 50-63, Available at: http://www.prosperfinancialmanagement.com/assets/client_files/PDFs/jd.pdf [Accessed 28 March 2019].

IX. Work Log

Time	Work done
Week 1	Topic research and finalization
Week 4	Finding the datasets
Week 6	Implementation of the method
Week 7	Evaluation of the result and conclusion
Week 8	Report writing