UNIVERSITY OF TEXAS AT ARLINGTON

2228-INSY-5339-002

PRINCIPLES OF BUSINESS DATA MINING

DATA MINING PROJECT FINAL REPORT

PROFESSOR: Dr. JAYARAJAN SAMUEL

GROUP NUMBER – 10

Submitted by:

Safoora Naureen (1002050531)
Dumpa Akash Reddy (1002025351)
Spandana Sanapureddy (1002102072)
Rakesh Maram (1002024578)

AIRLINE PASSENGER SATISFACTION

SOURCE OF DATA SET:

The data set is available in Kaggle.com:
https://www.kaggle.com/datasets/teejmahal20/airline-passenger-satisfaction

DESCRIPTION:

The statistics on airline passenger happiness include information on the customer satisfaction assessment index based on several factors. Each row represents a client, and each column lists the attributes of the customers as defined in the Metadata columns. The dataset includes statistics on

- Airline passengers, where satisfaction is the prediction column (TARGET VARIABLE).
- The services that each passenger has signed up for: Type of Travel, Class, Flight Distance, Inflight Wi-Fi service, Departure, Arrivals, Baggage Claim, Online-boarding, gate location, type of food, Ease of Online booking, Seat comfort, Inflight entertainment, on-board service, leg room service, check-in service, inflight service and cleanliness
- Passenger demographic data includes Id, Gender, Age Range, and Customer Type.

The airlines would greatly benefit from learning how client happiness affects business success and growth through comment analysis to better understand what the client demands. The best way to determine customer satisfaction and look at the many areas where a business may improve is through customer feedback.

DESCRIPTION OF VARIABLES IN THE DATA SET:

DATASET: This dataset contains an airline passenger satisfaction survey.

No. of Rows	129880
No. of columns	24

OBJECTIVE:

As the COVID cases seem to decrease in the past few months, many people are turning towards travel, more than in pre-COVID times, as there were lockdowns and quarantines for most of the time. There are over 5000 airline choices and it was an 872-billion-dollar revenue before the pandemic. This way, airlines have the best time to improve their businesses and make up for the loss in the past few years. For that, companies would turn towards achieving higher passenger satisfaction and look for ways to evaluate their service.

The inclusion of the expectations accomplishes two objectives:

- It can first be used to examine the factors that affect satisfaction.
- The dataset may be split into independent components and dependent factors in order to develop models that forecast consumer satisfaction (the satisfaction column).

In order to develop a remedy or proposal to forecast the behaviors of passenger satisfaction levels, the Airline Passenger Satisfaction Survey requested assistance from its data analytics team. We selected the dataset below, which shows consumer happiness over time, as part of our investigation.

BUSINESS PROBLEM:

Airline passengers are generally divided into Business class and Economy class. Business class tickets often cost 4 to 5 times as much as economy class tickets. It indicates that compared to economy class, business class generates 3 to 5 times more income for airlines. The data analysis team will thus focus on a business class of travellers. Airlines may learn what regions have room for service development after reviewing the data, and they can also investigate the crucial elements that influence consumer happiness. The airline firm asks the data analysis team to look into and forecast which aircraft services may be enhanced without damaging the economy class to boost the number of customers who select business class travel. Following data research, changes must be made to encourage more people to choose the business class. Meaning that airlines may comprehend which regions have room for service enhancement and also study the crucial elements that influence consumer contentment.

The analysis of customer feedback data is currently being used to pinpoint the amenities that might significantly increase the number of consumers and enhance the airline company. For every facility offered by the airline, we have ratings from over 100k consumers, along with a final happy, neutral, or unsatisfied comment.

VARIABLES:

- Gender: Gender of the passengers (Female, Male).
- Customer Type: The customer type (Loyal customer, Dis-loyal customer).
- Age: The actual age of the passengers.
- Type of Travel: Purpose of the flight of the passengers (Personal Travel, Business Travel).
- Class: Travel class in the plane of the passengers (Business, Eco, and Eco Plus).
- Flight distance: The flight distance of this journey.
- Inflight Wi-Fi service: Satisfaction level of the inflight Wi-Fi service (0: Not Applicable; 1-5).
- Departure/Arrival time convenient: Satisfaction level of Departure/Arrival time convenient.
- Ease of Online booking: Satisfaction level of online booking.
- Gate location: Satisfaction level of Gate location.
- Food and drink: Satisfaction level of Food and drink.
- Online boarding: Satisfaction level of online boarding.
- Seat comfort: Satisfaction level of Seat comfort.
- Inflight entertainment: Satisfaction level of inflight entertainment.

- On-board service: Satisfaction level of On-board service.
- Leg room service: Satisfaction level of Leg room service.
- Baggage handling: Satisfaction level of baggage handling.
- Check-in service: Satisfaction level of Check-in service.
- Inflight service: Satisfaction level of inflight service.
- Cleanliness: Satisfaction level of Cleanliness.
- Departure Delay in Minutes: Minutes delayed when departure.
- Arrival Delay in Minutes: Minutes delayed when Arrival.
- Satisfaction: Airline satisfaction level (Satisfied, Neutral / Dissatisfied).

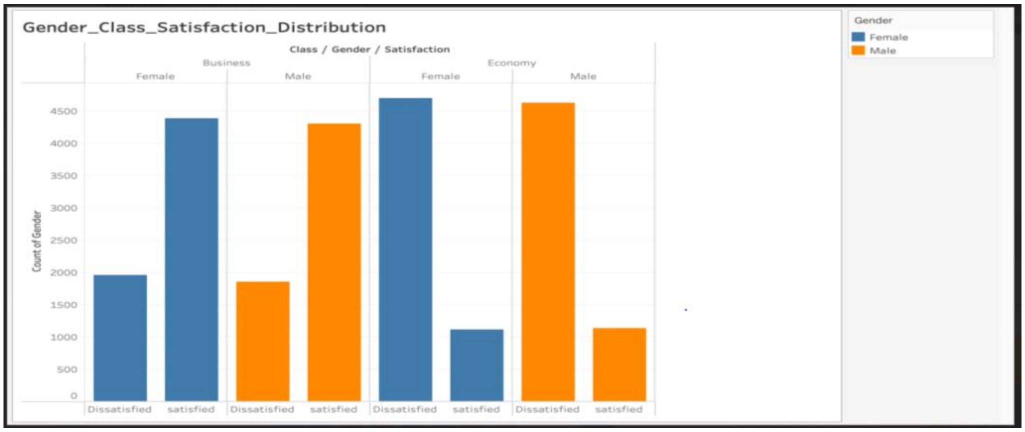
STEPS TO BE FOLLOWED:

- Exploring the Data.
- Data visualization and Stats summary.
- Filling the missing values.
- Splitting the data to training and test sets.
  - a) Training data (80%) – For training model.
  - b) Testing data (20%) – For scoring model against new data.
- Models are built using Stepwise Logistic Regression, Decision tree with Entropy Criterion, Gradient Boosting and Deep Learning Model using Neural Network.
- Evaluating the model using the training set and then with test set to predict accuracy.
- Make conclusions on the insights derived.

How many observations in the dataset?	129880
How many binary/categorical variables?	24 (Categorical)
How many continuous variables?	4 (Continuous)
What is the outcome / target variable?	Satisfaction
If binary or categorical: What percentage of the variables belong to each class.	56% are neutral/dissatisfied 44% satisfied
If continuous: What is the mean value of the target variable?	Not continuous

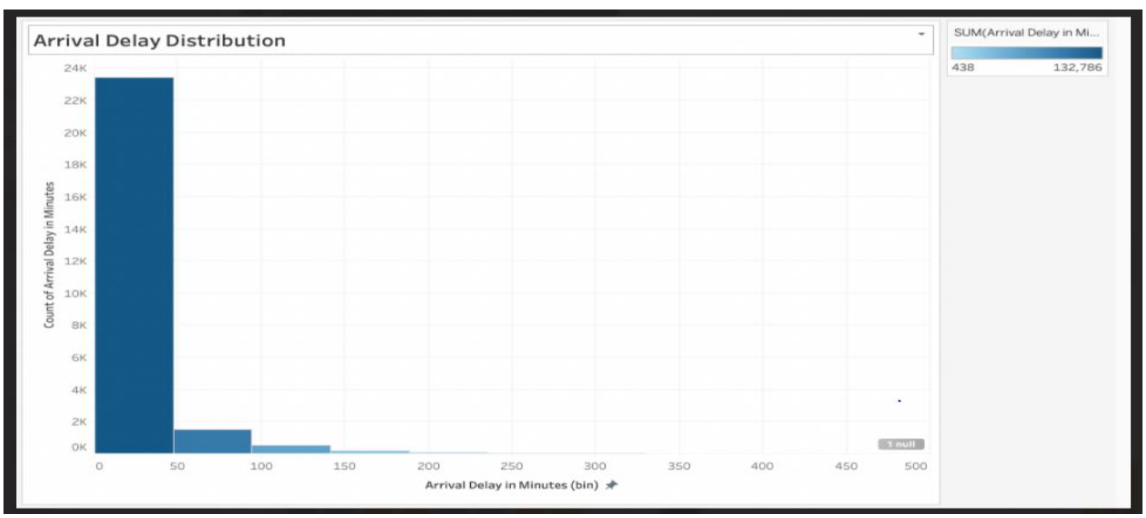
VISUALIZATION TECHNIQUES USING TABLEAU:

- As per the below plot, we have checked the satisfaction rate among male and female customers in both business and economy classes.



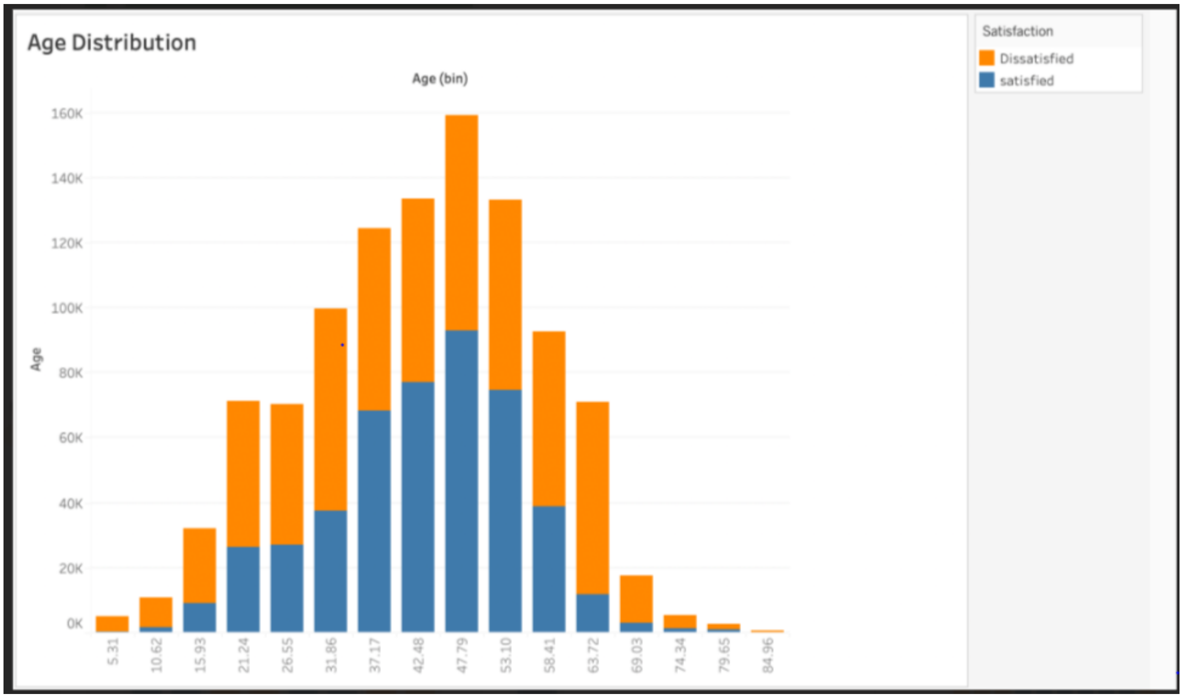
The incidence of service dissatisfaction appears to be approximately comparable between the genders given that more than 50% of each gender report concerns. The airline has two different price tiers for Business and Economy tickets. Given that over 80% of travellers in various classes are unsatisfied, there could be particulars for the Economy class. Business class appears to be performing the best, with a 31% grade for issues that affect satisfaction.

- Below histogram shows the Arrival delay in Minutes and the Count of the Arrival Delay in Minutes.



For every airline, delays are a major issue. The magnitude of the delays may compel airlines to offer refunds and overnight accommodations in addition to having an effect on customers' satisfaction, according to the governing regulations. The improper placement of personnel and planes might potentially have an impact on schedule. Flight delays should then be analysed independently after identifying probable causes and deciding whether comparable issues may be eliminated. The lengthier the wait, the more likely it is that consumers would be disappointed, according to an analysis of the total delays, arrival delays, and exit delays.

- As per the below plot, we see the satisfaction and dissatisfaction rate among different age groups.



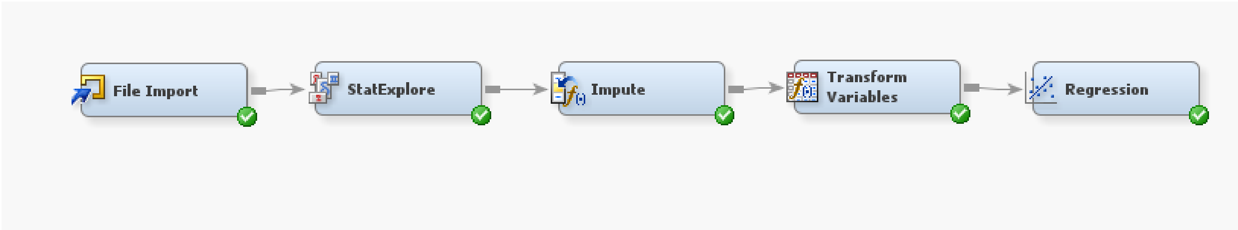
The average passenger age on the airline is around 40 years old, yet it transports passengers as young as 6 and as elderly as 85. Age groups between 40 and 60 were found to have the highest likelihood of finding the service sufficient, whilst those above 60 had the highest likelihood of finding the service inadequate.

<u>How are we using data to answer the business problem?</u>

• We need to run regression models for business class and economy class separately.
• Get which variables are statistically significant for business class but have very less to none significance to economy class.
• Run the model for all observations with shortlisted statistically significant variables of business class.

<u>MODEL BUILDING THROUGH SAS ENTERPRISE MINER:</u>

1. Creating regression model for business class passengers and economy class passengers separately.



- Firstly, we imported the excel file containing business class passengers into the SAS enterprise miner using the file import node.
- We then marked the type of variables namely Class as rejected because we have only business class values in the dataset. We also marked ID as ID for the dataset and ran a stat-explore node.
- The statistics revealed that Arrival delay in minutes contains missing values. As seen from visualizations in the tableau, the arrival delay in the minutes' column was skewed to the right. Hence, the column was imputed with its median value using the impute node.
- Further, depending on the distributions of variables seen through exploring option in SAS, the columns Departure_Delay_in_Minutes, Flight_Distance, IMP_Arrival_delay_in_Minutes were given a log 10 transformation using the transform variables node.
- We ran regression for imputed and transformed variables to get the estimates as shown below

Analysis of Maximum Likelihood Estimates

Parameter		satisfaction	DF	Estimate	Standard Error	Wald Chi-Square	Pr > ChiSq	Standardized Estimate	Exp(Est)
Intercept		satisfied	1	-11.6039	0.1297	8005.09	<.0001		0.000
Age		satisfied	1	-0.0117	0.00121	93.35	<.0001	-0.0825	0.988
Baggage_handling		satisfied	1	0.3123	0.0210	220.36	<.0001	0.1928	1.367
Checkin_service		satisfied	1	0.4702	0.0136	1193.30	<.0001	0.3072	1.600
Cleanliness		satisfied	1	0.3765	0.0163	533.96	<.0001	0.2569	1.457
Customer_Type	Loyal Customer	satisfied	1	1.2152	0.0224	2940.52	<.0001		3.371
Departure_Arrival_time_convenien		satisfied	1	-0.0831	0.0140	35.26	<.0001	-0.0690	0.920
Ease_of_Online_booking		satisfied	1	-0.0721	0.0148	23.84	<.0001	-0.0586	0.930
Gate_location		satisfied	1	0.0954	0.0148	41.82	<.0001	0.0719	1.100
Inflight_entertainment		satisfied	1	-0.0839	0.0206	16.65	<.0001	-0.0574	0.920
Inflight_service		satisfied	1	0.2750	0.0227	146.50	<.0001	0.1695	1.316
LG10_Departure_Delay_in_Minutes		satisfied	1	0.0848	0.0359	5.59	0.0181	0.0326	1.089
LG10_IMP_Arrival_Delay_in_Minute		satisfied	1	-0.3908	0.0358	118.98	<.0001	-0.1509	0.677
Leg_room_service		satisfied	1	0.3824	0.0151	640.22	<.0001	0.2575	1.466
On_board_service		satisfied	1	0.4415	0.0185	571.02	<.0001	0.2900	1.555
Online_boarding		satisfied	1	0.7953	0.0149	2861.31	<.0001	0.5344	2.215
Seat_comfort		satisfied	1	0.1605	0.0166	93.76	<.0001	0.1057	1.174
Type_of_Travel	Business travel	satisfied	1	1.9132	0.0414	2136.14	<.0001		6.775

- Similar process was followed for economy class to get the following estimates.

Analysis of Maximum Likelihood Estimates

Parameter		satisfaction	DF	Estimate	Standard Error	Wald Chi-Square	Pr > ChiSq	Standardized Estimate	Exp(Est)
Intercept		satisfied	1	-5.5815	0.1516	1355.35	<.0001		0.004
Age		satisfied	1	-0.00752	0.000925	65.99	<.0001	-0.0694	0.993
Checkin_service		satisfied	1	0.1765	0.0113	245.57	<.0001	0.1269	1.193
Cleanliness		satisfied	1	0.0708	0.0154	21.10	<.0001	0.0528	1.073
Customer_Type	Loyal Customer	satisfied	1	0.8070	0.0208	1504.56	<.0001		2.241
Ease_of_Online_booking		satisfied	1	-0.1696	0.0155	119.19	<.0001	-0.1225	0.844
Gate_location		satisfied	1	0.0338	0.0125	7.29	0.0069	0.0222	1.034
Inflight_service		satisfied	1	-0.0385	0.0129	8.91	0.0028	-0.0254	0.962
Inflight_wifi_service		satisfied	1	1.1566	0.0197	3435.57	<.0001	0.7840	3.179
LG10_Flight_Distance		satisfied	1	-0.2035	0.0424	23.01	<.0001	-0.0369	0.816
LG10_IMP_Arrival_Delay_in_Minute		satisfied	1	-0.5215	0.0213	597.07	<.0001	-0.2055	0.594
Leg_room_service		satisfied	1	0.0215	0.0109	3.88	0.0489	0.0159	1.022
On_board_service		satisfied	1	0.1543	0.0117	173.91	<.0001	0.1117	1.167
Online_boarding		satisfied	1	0.0506	0.0176	8.23	0.0041	0.0369	1.052
Seat_comfort		satisfied	1	0.0431	0.0155	7.75	0.0054	0.0323	1.044
Type_of_Travel	Business travel	satisfied	1	1.0137	0.0171	3523.92	<.0001		2.756

2. Comparing the statistically significant variable estimates from both business and economy class.

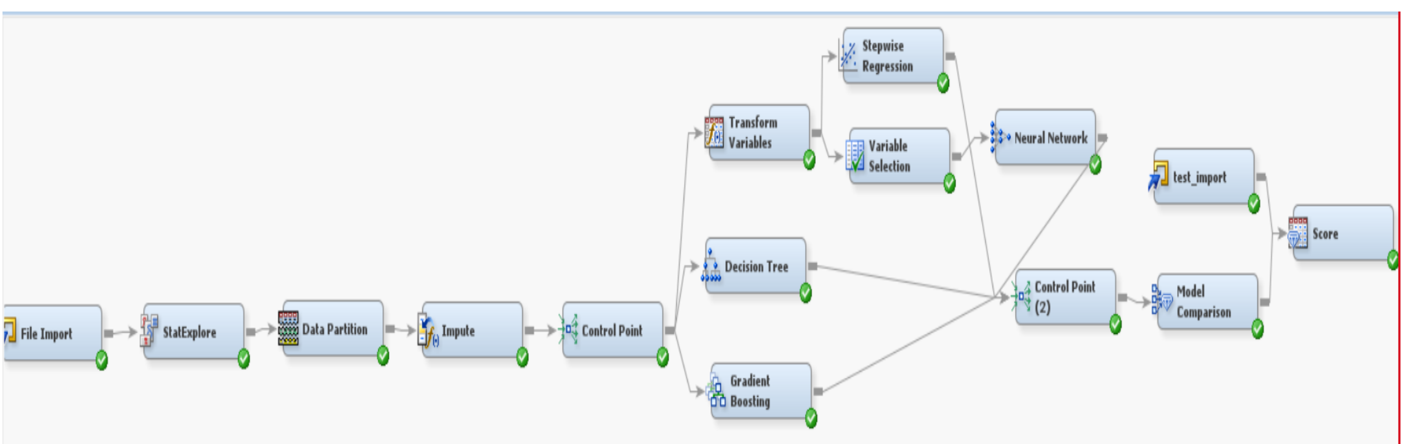
The variables which have p-values <0.0001 were taken for each class and the estimates of both classes were compared. **The absolute estimates with a higher effect on the business class than that of the economy class were taken after sorting them in descending order**.
For example, the type of travel = business affects the business class by 1.932 times which is higher than 1.0137 times that of economy class. Similarly, if the customer type is a loyal customer, the target variable for the business class is affected 1.2 times when compared to 0.8 times for that economy class. Online boarding is of high importance to business class as the beta value of the variable is 0.7953 which is very high compared to 0.056 of the economy class. Similarly, Type_of_Travel, Customer_Type, Online_boarding, Checkin_service, On_board_service, LG10_IMP_Arrival_Delay_in_Minute, Leg_room_service, Cleanliness, Baggage_handling, Inflight_service, Seat_comfort, Gate_location, Inflight_entertainment, Departure_Arrival_time_convenience, and Age were statistically significant and have higher beta values on business class when compared to economy class. The variables namely Ease_of_Online_booking, Inflight_wifi_service, and LG10_Flight_Distance effect economy class with a higher magnitude than the business class. Such variables are not taken into consideration. After that, **an exogenous business decision was made to take the variables with an estimate value effect of greater than 0.3 times the target variable** as shown below.

	A	B	C	D	E	F	G	H	I
1		Business	Economy				Business	Economy	
2	Type_of_Travel	1.9132	1.0137			**Intercept**	11.6039	5.5815	
3	Intercept	-11.6039	-5.5815			**Type_of_Travel**	1.9132	1.0137	
4	Leg_room_service	0.3824	0.0215			**Customer_Type**	1.2152	0.807	
5	Inflight_entertainment	-0.0839				**Online_boarding**	0.7953	0.0506	
6	Checkin_service	0.4702	0.1765			**Checkin_service**	0.4702	0.1765	
7	Departure_Arrival_time_convenien	-0.0831				**On_board_service**	0.4415	0.1543	
8	Age	-0.0117	-0.00752			**LG10_IMP_Arrival_Delay_in_Minute**	0.3908	0.5215	
9	Ease_of_Online_booking	-0.0721	-0.1696			**Leg_room_service**	0.3824	0.0215	
10	On_board_service	0.4415	0.1543			**Cleanliness**	0.3765	0.0708	
11	Online_boarding	0.7953	0.0506			**Baggage_handling**	0.3123		
12	Gate_location	0.0954	0.0338			**Inflight_service**	0.275	0.0385	0.3 cutoff
13	LG10_IMP_Arrival_Delay_in_Minute	-0.3908	-0.5215			**Seat_comfort**	0.1605	0.0431	
14	Inflight_service	0.275	-0.0385			**Gate_location**	0.0954	0.0338	
15	Seat_comfort	0.1605	0.0431			**Inflight_entertainment**	0.0839		
16	Cleanliness	0.3765	0.0708			**Departure_Arrival_time_convenien**	0.0831		
17	Baggage_handling	0.3123				**Age**	0.0117	0.00752	
18	Customer_Type	1.2152	0.807			Ease_of_Online_booking	0.0721	0.1696	
19						Inflight_wifi_service		0.0197	
20						LG10_Flight_Distance		0.0424	

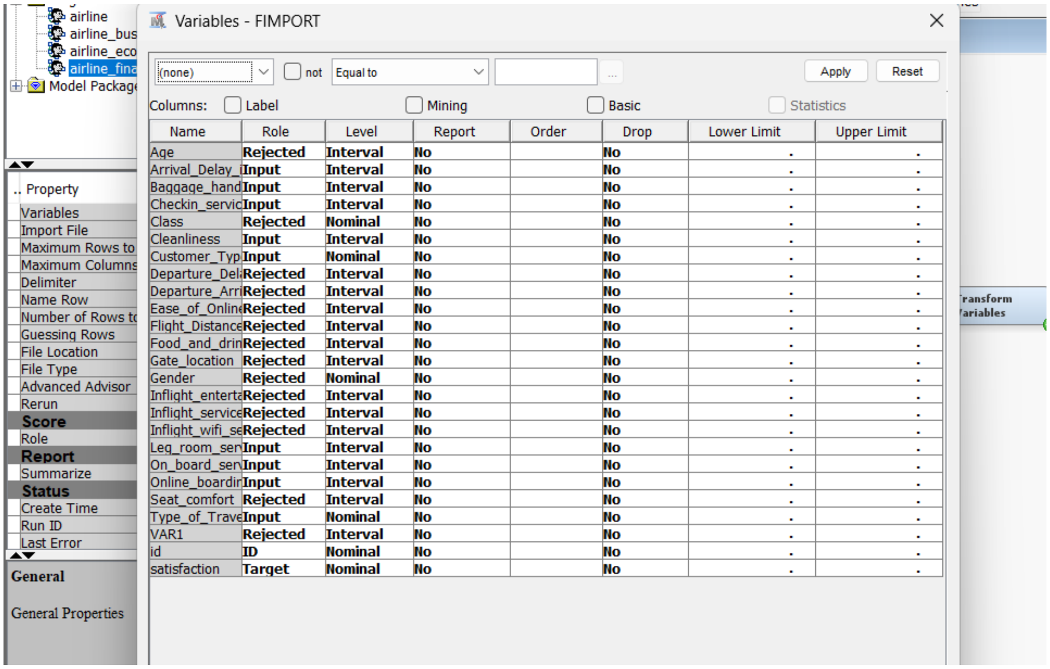
The final regression equation for business class would be as follows

*Y (**Business Class Satisfaction)*** = -11.6039 +1.9132(Type_of_travel = Business) + 1.2152 (Customer_Type = Loyal customer) + 0.7953(Online_Boarding) +0.4702(Check_in_Service) + 0.4415(On_Board_Service) - 0.3908(LG10_IMP_Arrival_Delay_in_Minutes) + 0.3824(Leg_room_service) + 0.3765(Cleanliness) + 0.3123(Baggage_Handling)

3. Using variables with high statistical significance to business class for building various algorithms for the entire dataset.



- Now that the factors which affect the business class are known, it is time to define the accurate magnitudes with which they affect the business class.
- For that, the dataset containing all the observations is imported into SAS Enterprise Miner, and variables except the ones in the above regression equation are rejected keeping Satisfaction as the target variable.



Name	Role	Level	Report	Order	Drop	Lower Limit	Upper Limit
Age	Rejected	Interval	No		No	.	.
Arrival_Delay_	Input	Interval	No		No	.	.
Baggage_hand	Input	Interval	No		No	.	.
Checkin_servic	Input	Interval	No		No	.	.
Class	Rejected	Nominal	No		No	.	.
Cleanliness	Input	Interval	No		No	.	.
Customer_Typ	Input	Nominal	No		No	.	.
Departure_Del	Rejected	Interval	No		No	.	.
Departure_Arr	Rejected	Interval	No		No	.	.
Ease_of_Onlin	Rejected	Interval	No		No	.	.
Flight_Distance	Rejected	Interval	No		No	.	.
Food_and_drin	Rejected	Interval	No		No	.	.
Gate_location	Rejected	Interval	No		No	.	.
Gender	Rejected	Nominal	No		No	.	.
Inflight_enterta	Rejected	Interval	No		No	.	.
Inflight_service	Rejected	Interval	No		No	.	.
Inflight_wifi_se	Rejected	Interval	No		No	.	.
Leg_room_ser	Input	Interval	No		No	.	.
On_board_ser	Input	Interval	No		No	.	.
Online_boardir	Input	Interval	No		No	.	.
Seat_comfort	Rejected	Interval	No		No	.	.
Type_of_Trave	Input	Nominal	No		No	.	.
VAR1	Rejected	Interval	No		No	.	.
id	ID	Nominal	No		No	.	.
satisfaction	Target	Nominal	No		No	.	.

- The variables' distribution and statistical moments are understood using the stat-explore node.
- After that, the data is divided into 80% for testing and 20% for validation so that the model can train itself on 80 percent of the data and further validate the results with 20 percent of it.
- The missing values in the Arrival delay in the minute column are filled with median value because the tableau histogram of the variable showed the data is highly skewed to the right. Mean is not a proper method of imputation here. This process is done using the impute node.
- As our target variable is binary, logistic regression is the go-to method for predicting the random variable function. To gain the advantage of both forward and backward selection, the stepwise logistic regression model is used. The model resulted in an accuracy of 86.3831 percent.
- The final regression equation through stepwise logistic regression becomes as follows:

Y ***(Satisfaction)*** = -9.2923 + 0.2625(Baggage_Handling) +0.3221(Check_in_Service) + 0.3038(Cleanliness) + 1.0025 (Customer_Type = Loyal customer) - 0.4439(LG10_IMP_Arrival_Delay_in_Minutes) + 0.3072(Leg_room_service)) + 0.3856(On_Board_Service) + 0.7395(Online_Boarding +1.6805(Type_of_travel = Business)

Analysis of Maximum Likelihood Estimates

Parameter		satisfaction	DF	Estimate	Standard Error	Wald Chi-Square	Pr > ChiSq	Standardized Estimate	Exp(Est)
Intercept		satisfied	1	-9.2923	0.0696	17824.30	<.0001		0.000
Baggage_handling		satisfied	1	0.2625	0.0112	548.12	<.0001	0.1709	1.300
Checkin_service		satisfied	1	0.3221	0.00911	1250.24	<.0001	0.2242	1.380
Cleanliness		satisfied	1	0.3038	0.00880	1190.70	<.0001	0.2196	1.355
Customer_Type	Loyal Customer	satisfied	1	1.0025	0.0141	5023.46	<.0001		2.725
LG10_IMP_Arrival_Delay_in_Minute		satisfied	1	-0.4439	0.0153	840.27	<.0001	-0.1735	0.642
Leg_room_service		satisfied	1	0.3072	0.00914	1130.97	<.0001	0.2227	1.360
On_board_service		satisfied	1	0.3856	0.0102	1420.21	<.0001	0.2739	1.471
Online_boarding		satisfied	1	0.7395	0.00964	5879.10	<.0001	0.5500	2.095
Type_of_Travel	Business travel	satisfied	1	1.6805	0.0147	13146.93	<.0001		5.368

- After using the stepwise logistic regression, a decision tree was used as it is not much affected by outliers or the distribution of the independent variables. The decision tree split the dataset taking entropy as the criterion for the split with 91.1325 accuracy which was far better than the stepwise logistic regression model.
- This motivated the team to use gradient boosting as it trains itself from various weaker models to provide an algorithm with higher accuracy. The gradient boosting model provided the dataset with 91.623 percent accuracy.
- After using the simple regression techniques, an effort was made to use deep learning for predicting the magnitude of the statistically significant variables. But, as the addition of a single variable to the neural networks increases the activation function by n-fold, the variable selection process was done to take the variables with maximum effect on the satisfaction.
- However, the neural network returned a model with an accuracy of 78.0253 percent which was the least of all the models used.
- The distribution of False Negatives, True Negatives, False Positives, and True Positives along with the model accuracies for the validation dataset is shown below.

Validation Dataset					
	False Negative	**True Negative**	**False Positive**	**True Positive**	**Accuracy %**
Stepwise Logistic Regression	1655	10602	1175	7351	86.3831
Decision Tree with entropy criterion	1031	10963	814	7975	91.1325
Gradient Boosting	1105	11141	636	7901	91.623
Neural Network	2681	9891	1886	6325	78.0253

- We have compared the models using the train data for Mean Square Error and Misclassification Rate, Model comparison node was used. Resulted in the statistics as follows:

Fit Statistics
Model Selection based on Valid: Mean Square Error (_VMSE_)

Selected Model	Model Node	Model Description	Valid: Mean Square Error	Train: Average Squared Error	Train: Misclassification Rate	Valid: Average Squared Error	Valid: Misclassification Rate
	Boost	Gradient Boosting	.	0.05688	0.07897	0.06059	0.08377
	Tree	Decision Tree	.	0.06771	0.08898	0.06827	0.08877
Y	Reg	Stepwise Regression	0.10153	0.09909	0.13141	0.10153	0.13617
	Neural	Neural Network	0.15237	0.15052	0.21721	0.15237	0.21975

- The Gradient Boosting Model has the lowest Mean Square Error and Misclassification Rate (0.08377), as can be seen from the above chart.
- This means that gradient boosting has the maximum accuracy in predicting the magnitudes by which the independent variables influence the target variable. Therefore, the model's outputs are observed.

Variable Importance

Obs	NAME	LABEL	NRULES	NSURROGATES	IMPORTANCE	VIMPORTANCE	RATIO
1	Online_boarding	Online boarding	977	456	1.00000	1.00000	1.00000
2	Type_of_Travel	Type of Travel	209	121	0.93380	0.92057	0.98584
3	Cleanliness		1315	1080	0.85020	0.83967	0.98761
4	Leg_room_service	Leg room service	1203	1401	0.70296	0.68062	0.96822
5	Baggage_handling	Baggage handling	1287	1516	0.47663	0.45629	0.95731
6	IMP_Arrival_Delay_in_Minutes	Imputed: Arrival Delay in Minutes	1685	2030	0.45713	0.43170	0.94437
7	On_board_service	On-board service	1176	1502	0.45378	0.42993	0.94744
8	Customer_Type	Customer Type	323	370	0.42764	0.42894	1.00303
9	Checkin_service	Checkin service	1353	1071	0.37908	0.35570	0.93832

- This demonstrates that the three key criteria that affect the business class are online boarding, the kind of travel, and cleanliness.

Data Role=VALIDATE Output Type=CLASSIFICATION

Variable	Numeric Value	Formatted Value	Frequency Count	Percent
I_satisfaction	.	NEUTRAL OR DISSATISFIED	12257	58.9761
I_satisfaction	.	SATISFIED	8526	41.0239

- After importing the test data, the score node in SAS Enterprise Miner was used to estimate the happiness of more customers using the model with the best accuracy. On the unexposed test data, the gradient boosting model was applied to forecast consumer happiness.
- The gradient boosting model projected that out of all 20783 consumers in the dataset, 58.9761 percent would be indifferent or unhappy and 41.0239 percent would be satisfied.

CONCLUSION AND MANAGERIAL INSIGHTS

- According to our model, the very relevant factors for business class are Online_Boarding, Type of travel, Cleanliness, Leg_room_service, Baggage Handling, IMP_Arrival_Delay_In_Minutes, Onboard_Service, Customer_Type, Checkin_Service

- From the resultant variable, we can observe that the three most important variables for the enhancement of the business class sale are online boarding, the kind of trip, and cleanliness.

- Given that the aforementioned amenities provide a sizable profit for business class travelers while causing little to no change for those traveling in economy class, we advise the managerial team to make more efforts to improve them. To see a noticeable difference in the number of passengers, we advise the managerial team to alter these factors.

- As business class passengers invest more money, they expect the onboard process to be easy and not much time-consuming. We suggest the team make changes according to that and make it easier for the business class passengers.

- We also suggest to the team to understand the passenger's purpose of traveling at the time of booking whether the passenger is traveling for business or for personal so that we can display ticketing options accordingly.

- We suggest the team make sure to take extra effort to keep the area clean by keeping individual sanitizers to individual seats and keeping the fixed intervals cleaning service to make sure the area looks clean always.

- We hope our data analysis and suggestions from it will help in improving business class sales without affecting economy class sales.