SYRACUSE UNIVERSITY

AIRLINES ANALYSIS

FINAL PROJECT REPORT

IST 707 – DATA ANALYTICS (SPRING 2020)

Aditya Tornekar

Paridhi Rajyaguru

Rishabh Upadhye

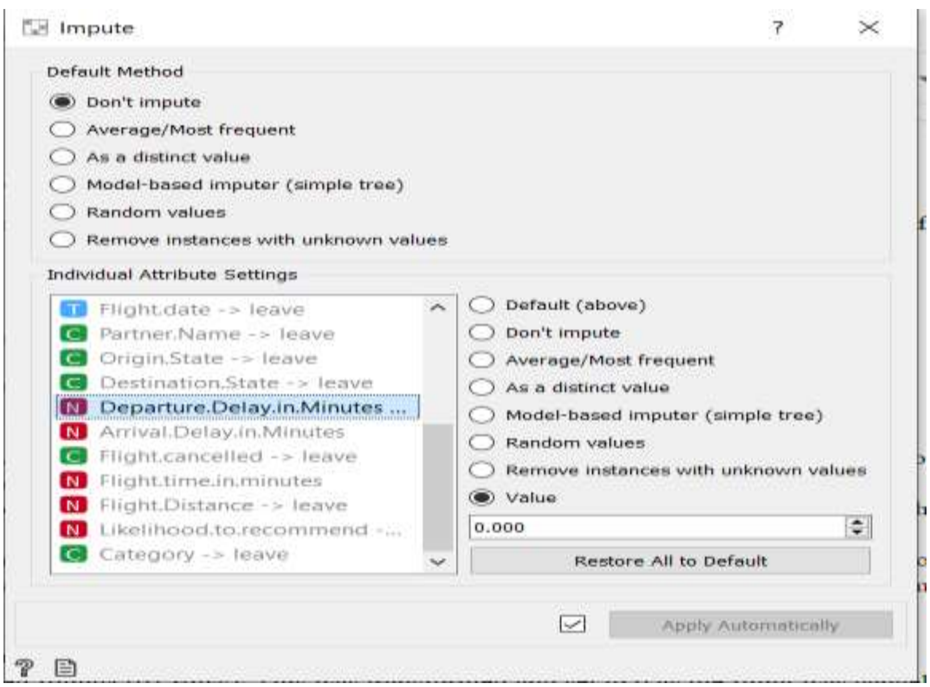
05/03/2020

Data Pre-processing / Data Preparation Phase:

No duplicates to be removed as each record is associated with a unique customer. There is a total of 4985 observations.

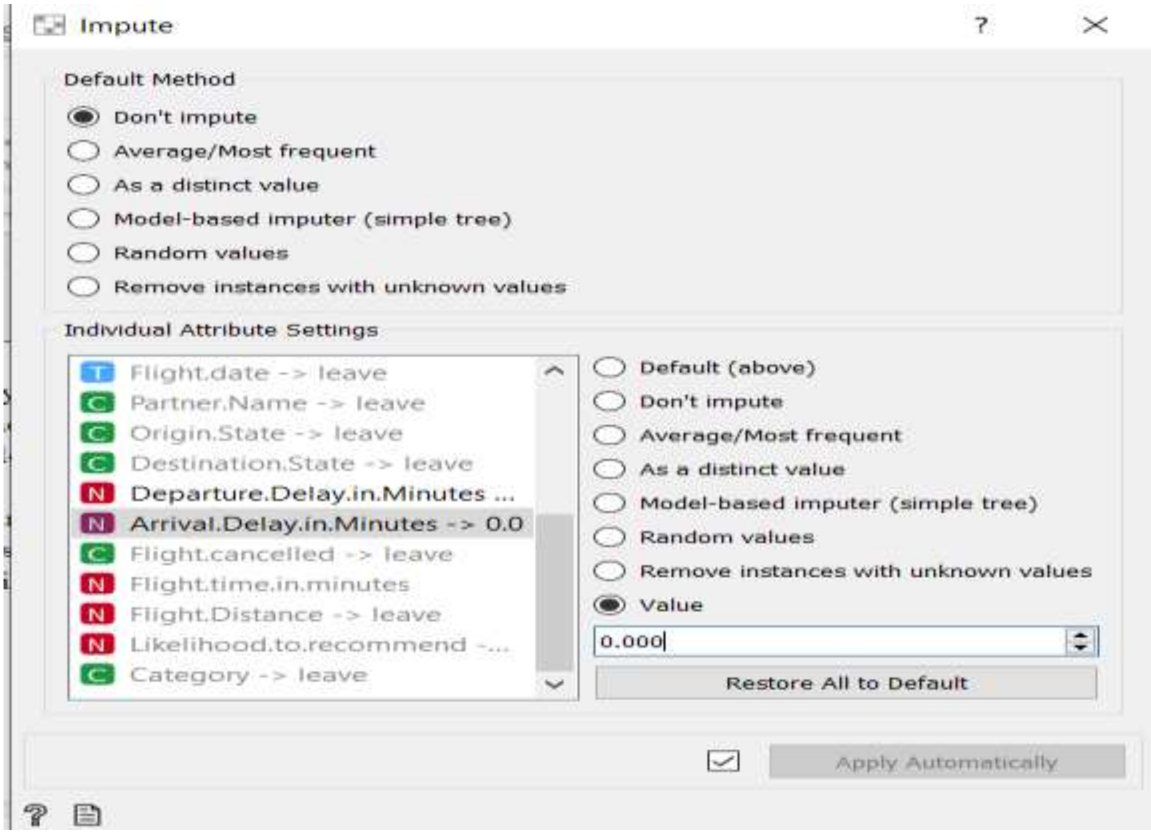
Phase1: Mitigate Missing Data

1) Departure.Delay.in.Minutes: Departure delay in minutes was "NA", only when a flight was cancelled (99 observations), this was changed to 0 as the flight was canceled.



2) Arrival.Delay.in.Minutes: This variable had “NA” values for 112 observations, regardless of flight being cancelled or not.

a) For canceled flights: A total of 101 observations were transformed and set to 0 as the flight was canceled.

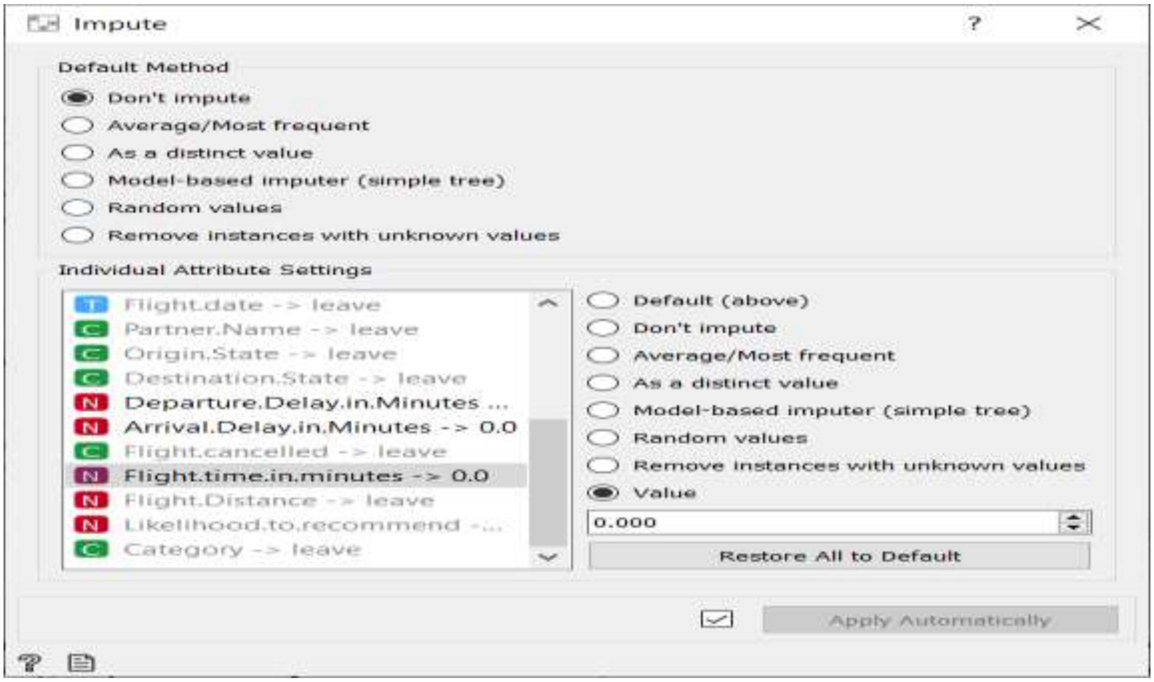


b) For non-canceled flights: Remaining 11 observations having “NA” values in this scenario were handled using Departure.Delay.in.Minute’s variable, as that could have been the minimum amount of possible delay and the same values were used to replace “NA”. This part was handled using MS Excel.

Destination.State	Flight.cancelled	Flight.Distance	Likelihood.to.recommend	Category	Departure.Delay.in.Minutes	Arrival.Delay.in.Minutes	Flight.time.in.minutes
Maryland	No	1246	8	Passive	106	106	NA
Alabama	No	682	9	Promoter	29	29	NA
Missouri	No	328	4	Detractor	5	5	NA
Arkansas	No	589	8	Passive	0	0	NA
Missouri	No	393	8	Passive	0	0	NA
Missouri	No	436	4	Detractor	0	0	NA
New York	No	762	7	Passive	51	51	NA
Utah	No	368	9	Promoter	0	0	NA
Texas	No	1214	8	Passive	110	110	NA
California	No	480	6	Detractor	0	0	NA
Missouri	No	448	9	Promoter	151	151	NA

3) Flight.time.in.minutes: This variable had “NA” values for 112 observations.

a) For canceled flights (101 rows): This was transformed and set to 0 as the flight was canceled.

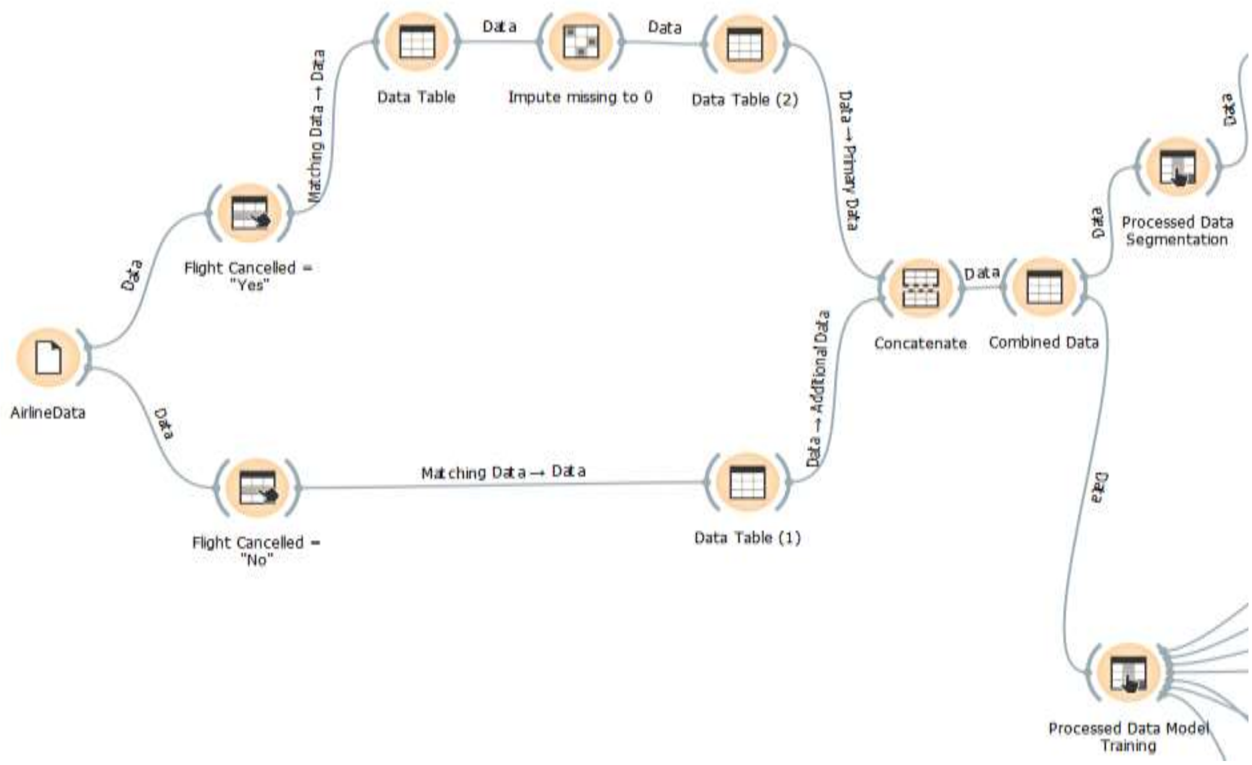


b) For non-canceled flights (11 rows): “NA” values in this scenario were handled by taking the average flight time of other flights that had the same flight distance. Interpolation technique or using an average of all observations will be wrong as flight time is dependent on flight distance. This part was handled using MS Excel.

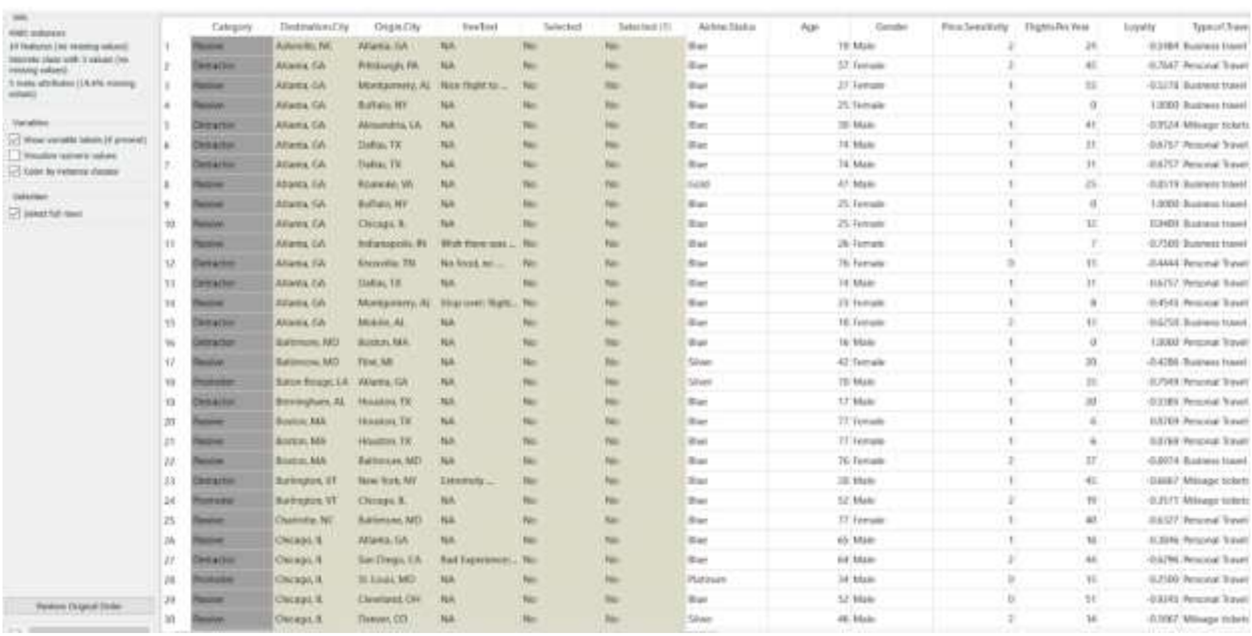
Destination.State	Flight.cancelled	Flight.Distance	Likelihood.to.recommend	Category	Departure.Delay.in.Minutes	Arrival.Delay.in.Minutes	Flight.time.in.minutes
Maryland	No	1246	8	Passive	106	106	NA
Alabama	No	682	9	Promoter	29	29	NA
Missouri	No	328	4	Detractor	5	5	NA
Arkansas	No	589	8	Passive	0	0	NA
Missouri	No	393	8	Passive	0	0	NA
Missouri	No	436	4	Detractor	0	0	NA
New York	No	762	7	Passive	51	51	NA
Utah	No	368	9	Promoter	0	0	NA
Texas	No	1214	8	Passive	110	110	NA
California	No	480	6	Detractor	0	0	NA
Missouri	No	448	9	Promoter	151	151	NA

Destination.State	Departure.Delay.in.Minutes	Arrival.Delay.in.Minutes	Flight.cancelled	Flight.time.in.minutes	Flight.Distance
Maryland	106	106	No	157	1246
Alabama	29	29	No	104	682
Missouri	5	5	No	51	328
Arkansas	0	0	No	135	589
Missouri	0	0	No	62	393
Missouri	0	0	No	67	436
New York	51	51	No	112	762
Utah	0	0	No	58	368
Texas	110	110	No	159	1214
California	0	0	No	72	480
Missouri	151	151	No	70	448

Snapshot of Data Processing workflow:

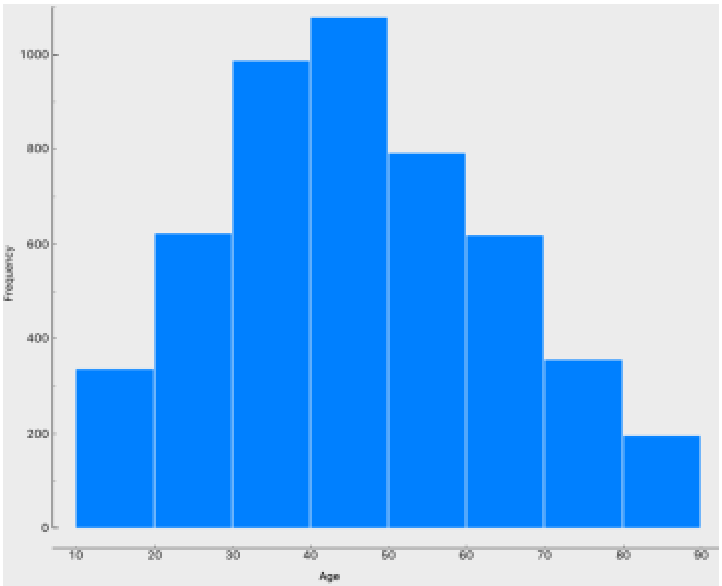


Snapshot of data after data processing:



Phase2: Summarize Variables

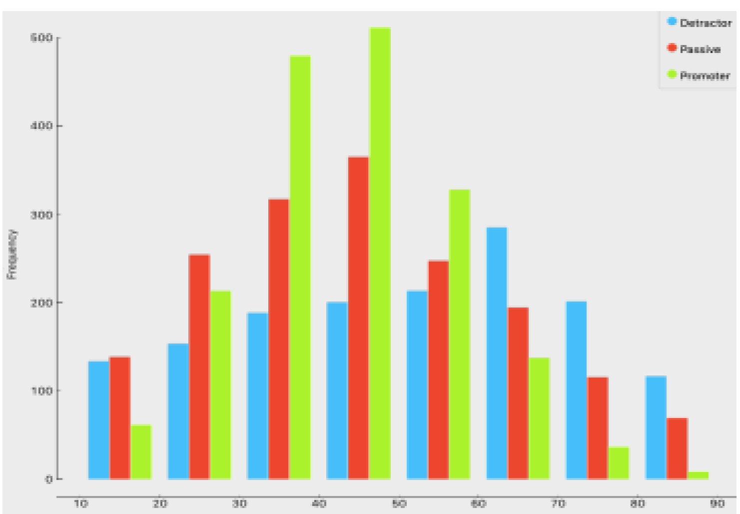
- **Age:** Age has an almost normal distribution, as seen below. We can split the Age into three categories: low (10-30), moderate (40-60), and high (>60).



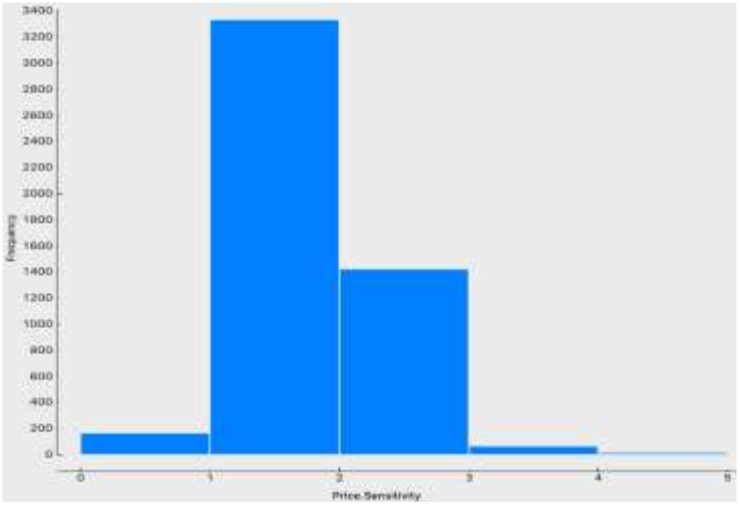
On examining age, with category, we get a better idea of which age range tends to be a promoter/detractor.

Most of the moderate age range is high in promoters, while the higher age range has significantly higher detractors (and a negligible portion of promoters). Another trend observed is that the proportion of detractors keeps increasing gradually with an increase in age, till about the 50-60 year bracket.

Snapshot of Age variable based on NPS categories:



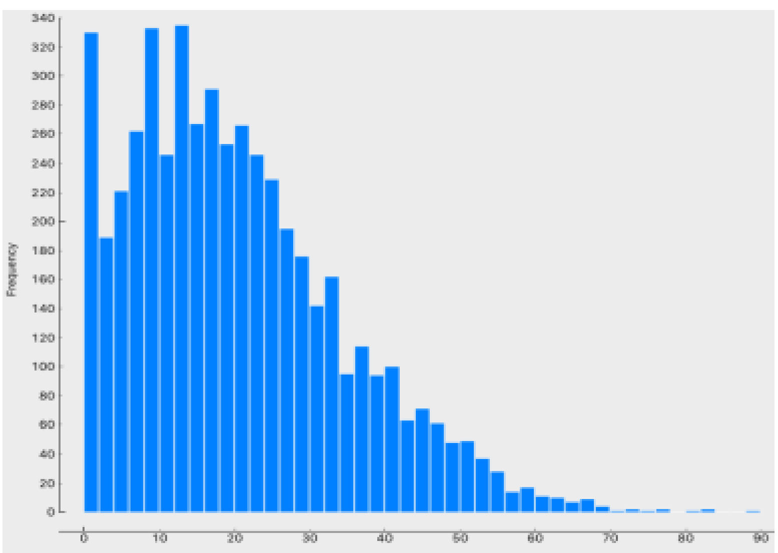
➢ **Price Sensitivity:** This attribute has a narrow normal distribution, with a high peak and small tails. It lies mostly in the mid 1-3 range.



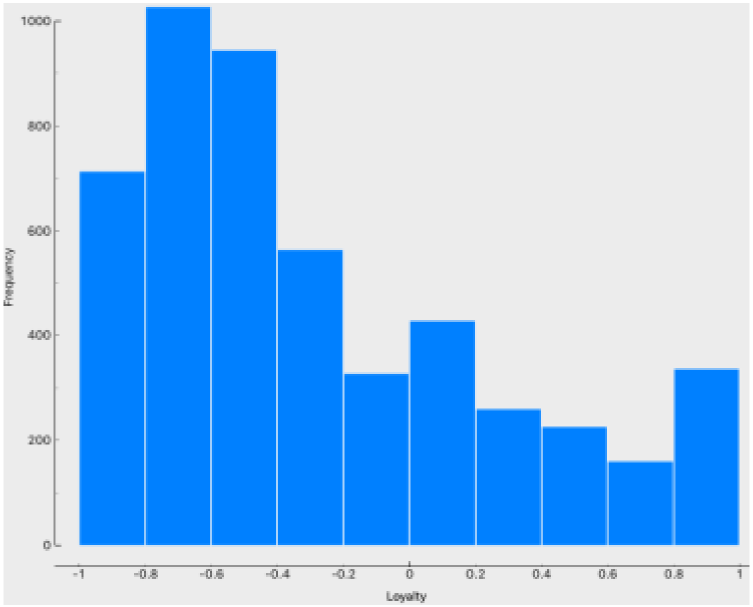
On further examination with other variables:

- Class, it is observed that "Economy" Class dominates passengers dominate the graph, which is to be expected. "Business" and "Economy Plus" have negligible price-sensitive customers
- Gender, the overall data suggests females are more price-sensitive than males.
- Partner Airlines, "Cheap seats" passengers are more sensitive than others, which is to be expected.

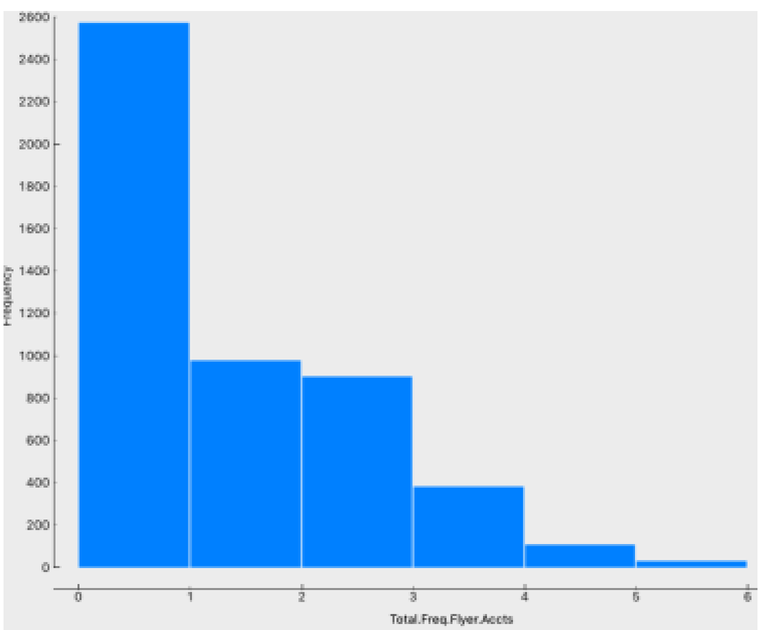
➢ **Flights per year:** The distribution is heavily left-skewed, with a small tail. It is concentrated mostly to 0-30 flights per year, regardless of airline status, partner airline or class of travel.



- **Loyalty:** The distribution is also left-skewed, with a gradual tail on the right. Over 70% of the customers fall in the negative loyalty range.

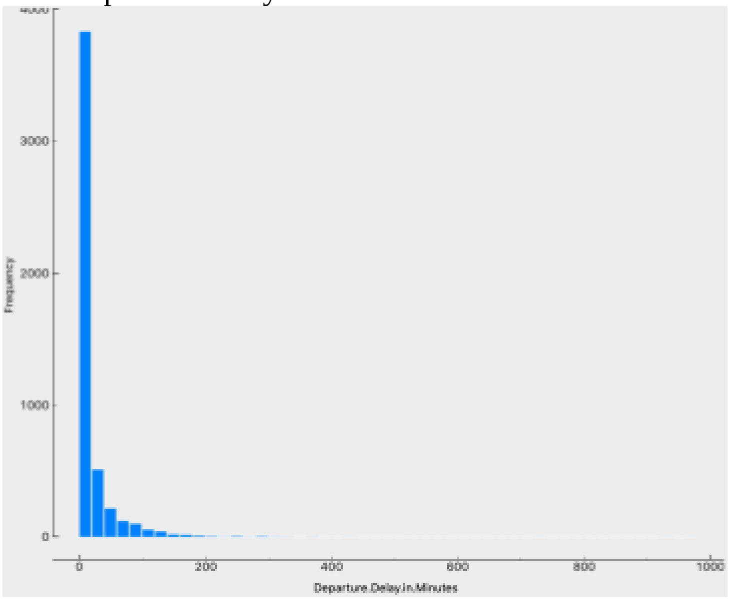


- **Total Frequent Flyer Accounts (FFA):** This distribution is also starkly skewed on the left, with a small thin tail. Over 50% of the customers have either none or 1 FFA, while almost 40% have 2-3 FFA.
  On examining further with Category, it is observed that customers with 0-1 FFA have a higher number of detractors, whereas those with more FFA's tend to be more populated by promoters.

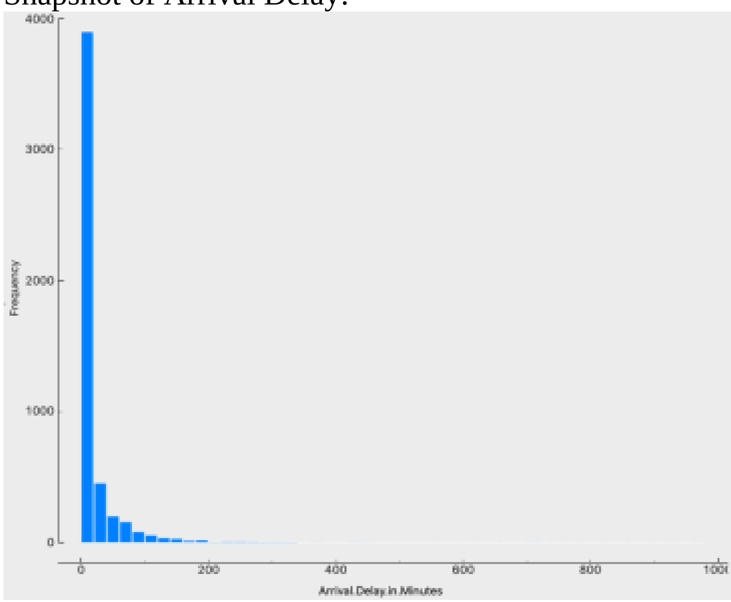


- **Departure Delay and Arrival Delay in Minutes:** The distributions for both these attributes are nearly identical. They are left-skewed with a sharp peak and long tail. Majority of the delays (>75%), whether during departure or arrival, are confined to 20 minutes or less.

Snapshot of Departure Delay:

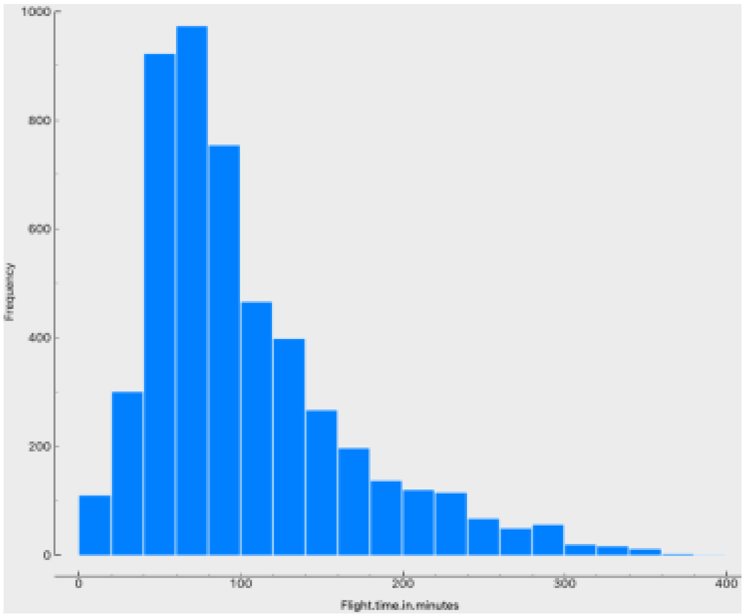


Snapshot of Arrival Delay:

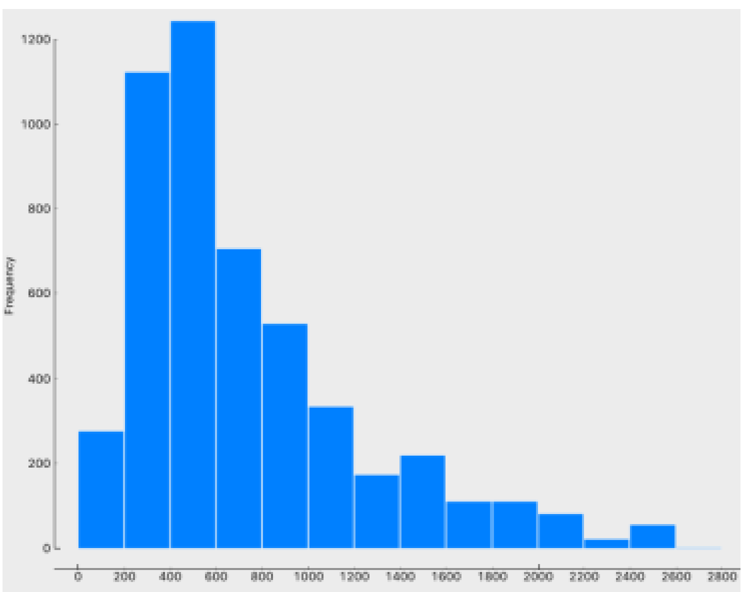


➢ **Flight time in minutes, Flight Distance:** Both these distributions are similar in nature - Somewhat left-skewed with trailing edges on the right. This makes sense as Flight Time is proportional to Flight Distance, and we observe that in the graphs below. Almost half of the data is concentrated in 500-700 miles of flight distance, which corresponds to around 40-100 minutes of flying time.

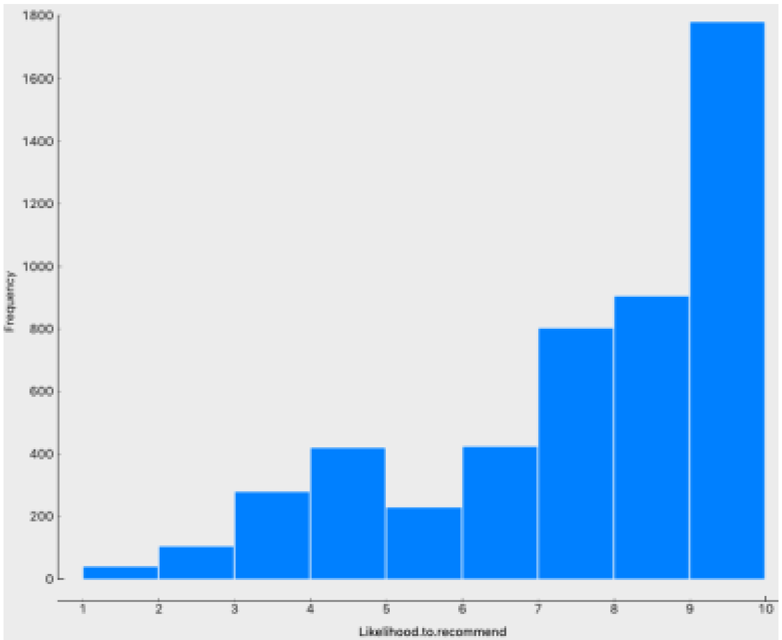
Snapshot of Flight time in minutes:



Snapshot of Flight Distance:



- **Likelihood to Recommend:** This distribution is right-skewed with a gradual left tail. In our data, we have around 30% detractors, with the remaining 70% passengers almost evenly split between passives and promoters.



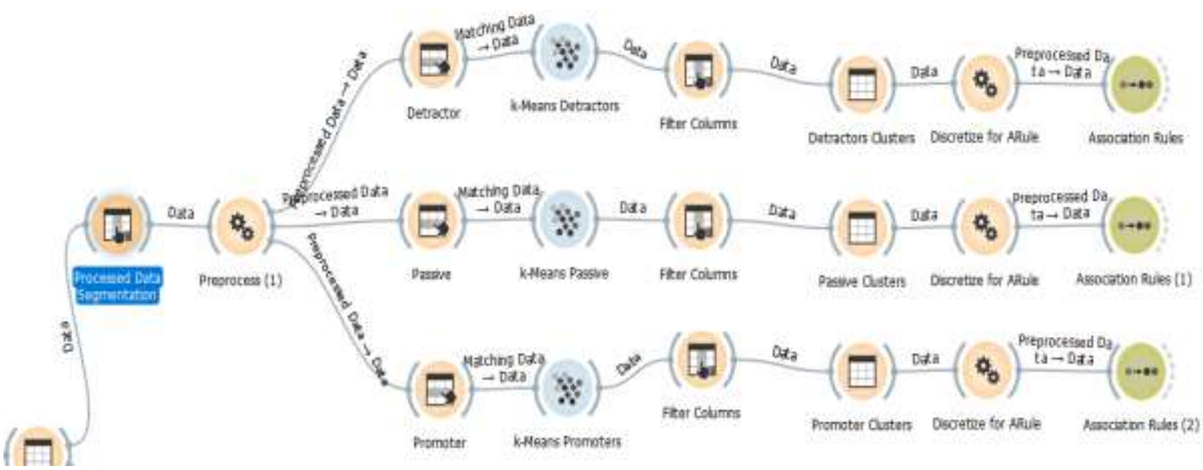
On further examination with other attributes, some interesting things we noticed:

- A high number of promoters are those that fly with airlines having "Silver" airline status, whereas detractors fly with "Blue" airline status
- A sizable portion of promoters also flies for "Business" type of travel, whereas detractors tend to be flying for "Personal" travel.

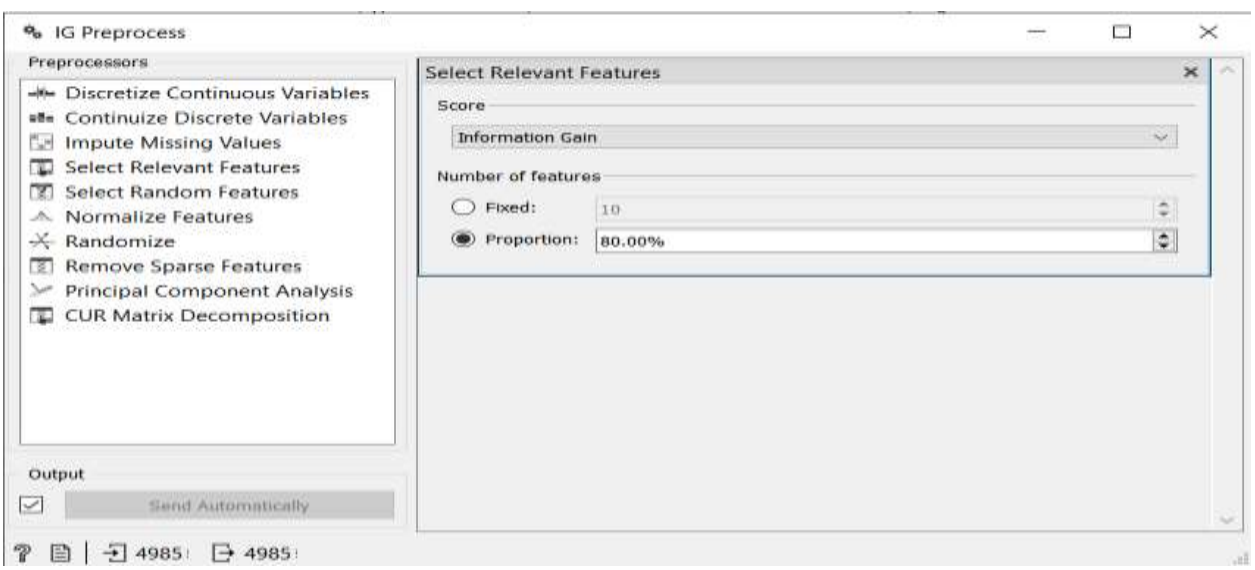
Exploratory Analysis Phase:

Phase3: Segmentation of Population

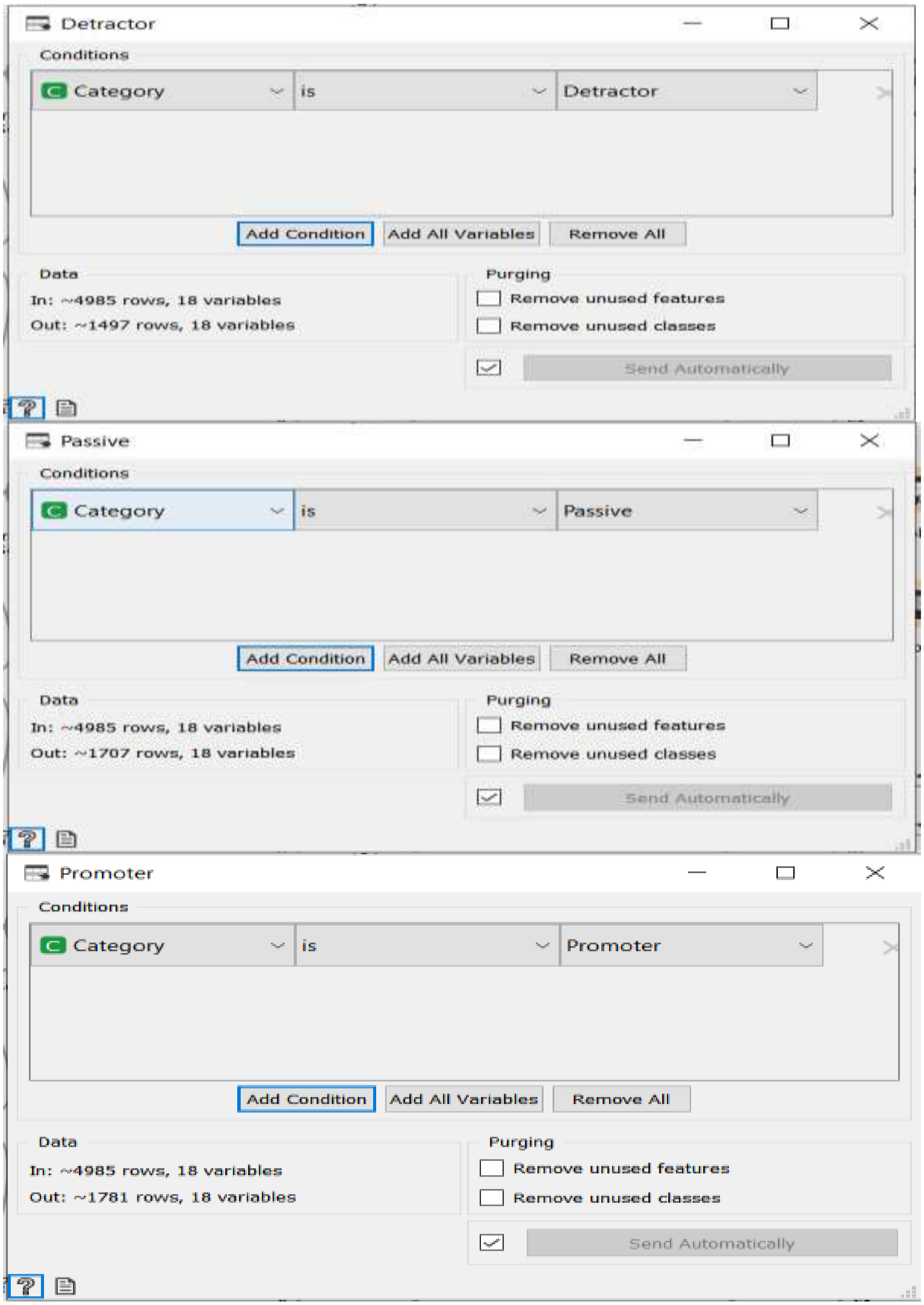
Here the 4985 observations are analyzed for patterns and associations using unsupervised machine learning techniques.



The data is first pre-processed to get relevant features contributing towards analyzing each category. 80% of the relevant features are considered for further cluster analysis using information gain filter.

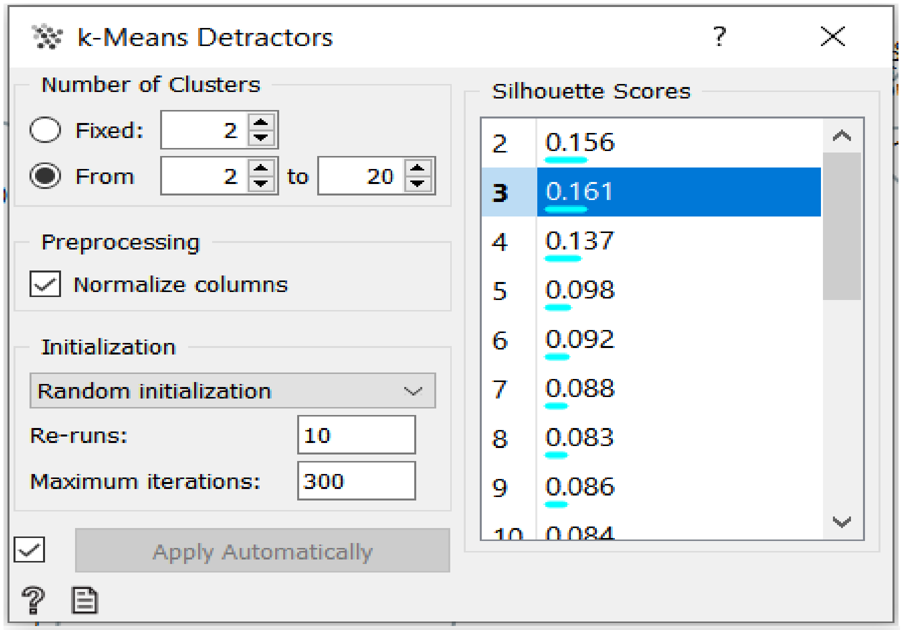


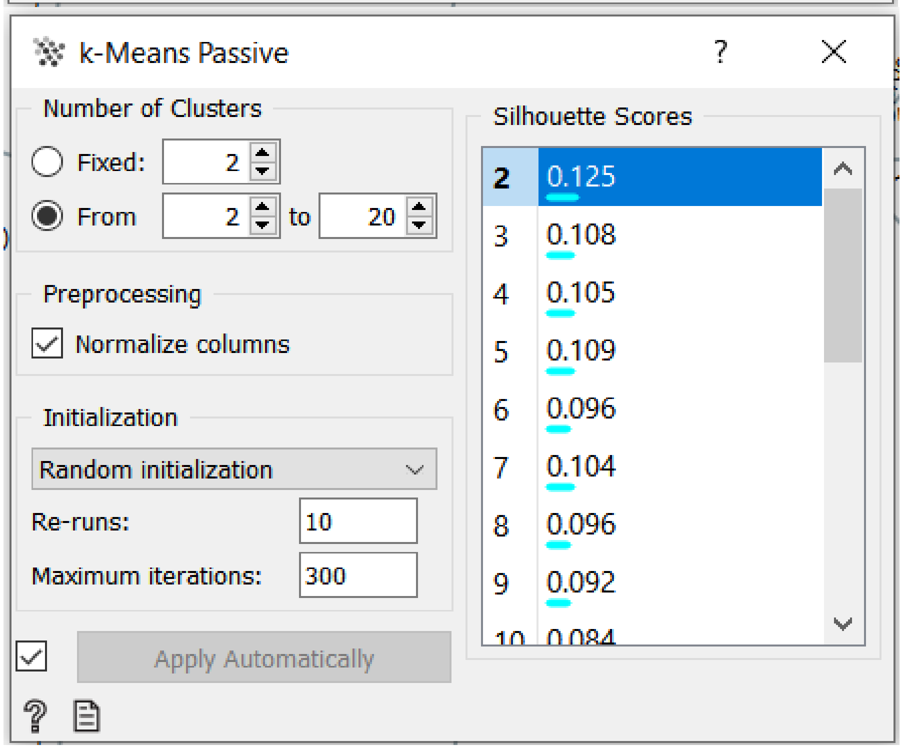
The observations are further segregated into 3 categories Detractors, Passive & Promoter using Select Columns filter Orange feature so that each of the customer categories is used for cluster creation and individual analysis. This type of analysis is useful as the clusters and associations for one category might not be the same for the other.

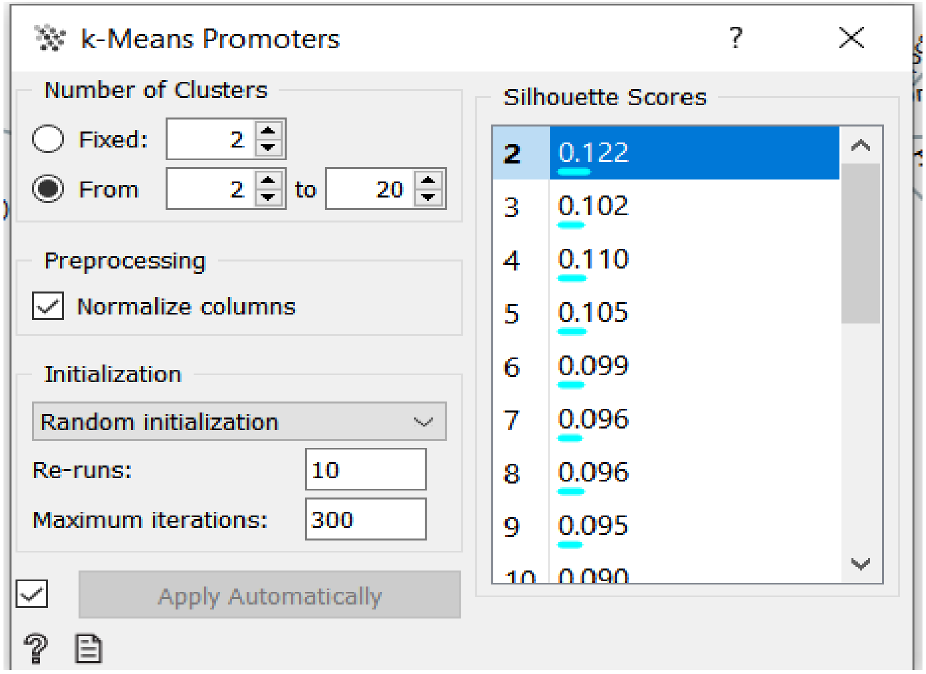


Cluster Formation:

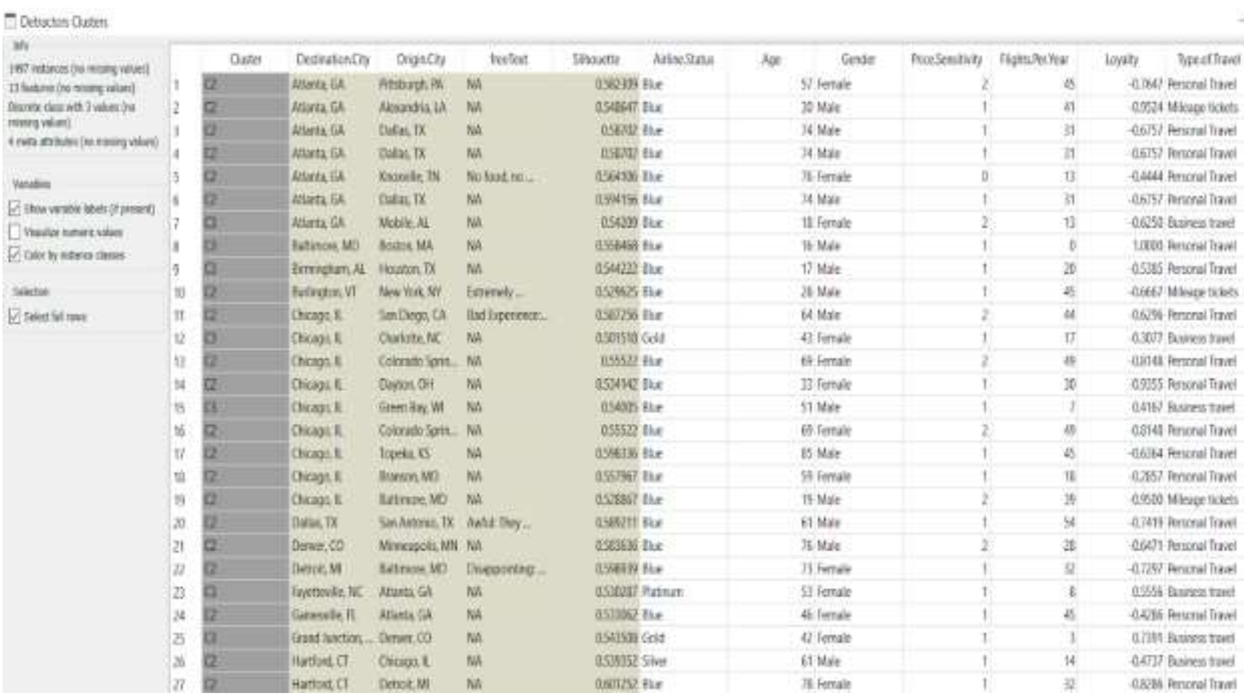
K-Means algorithm is used to create clusters on each category of Detractor, Passive, and Promoter. The number of clusters is decided using the Silhouette score analysis which tells how far the sample cluster is from its neighboring clusters. A range of 2-20 clusters was tested and forming 2 clusters in each category was optimal except the Detractors where 3 clusters were to be formed, this optimal clustering was done using the K-Means algorithm with a Silhouette score over 0.12 in all three cases.







Below the data, table viewer shows the cluster assigned to each observation mentioned in the 1st column. The Detractors are divided into three clusters C1, C2 & C3.



	Cluster	DestinationCity	OriginCity	freeText	Silhouette	AirlineStatus	Age	Gender	PriceSensitivity	FlightsPerYear	Loyalty	TypeofTravel
1	C2	Atlanta, GA	Pittsburgh, PA	NA	0.582309	Blue	57	Female	2	45	-0.7647	Personal Travel
2	C2	Atlanta, GA	Alexandria, LA	NA	0.548647	Blue	30	Male	1	41	0.9524	Mileage tickets
3	C2	Atlanta, GA	Dallas, TX	NA	0.56702	Blue	34	Male	1	31	-0.6757	Personal Travel
4	C2	Atlanta, GA	Dallas, TX	NA	0.56702	Blue	34	Male	1	31	0.6757	Personal Travel
5	C2	Atlanta, GA	Knoxville, TN	No food, no ...	0.564306	Blue	76	Female	0	13	-0.4444	Personal Travel
6	C2	Atlanta, GA	Dallas, TX	NA	0.594156	Blue	34	Male	1	31	-0.6757	Personal Travel
7	C1	Atlanta, GA	Mobile, AL	NA	0.54209	Blue	18	Female	2	13	-0.6250	Business travel
8	C3	Baltimore, MD	Boston, MA	NA	0.558468	Blue	16	Male	1	0	1.0000	Personal Travel
9	C1	Birmingham, AL	Houston, TX	NA	0.544222	Blue	17	Male	1	20	-0.5385	Personal Travel
10	C2	Burlington, VT	New York, NY	Extremely ...	0.529625	Blue	28	Male	1	45	-0.6667	Mileage tickets
11	C2	Chicago, IL	San Diego, CA	Bad Experience:...	0.587256	Blue	64	Male	2	44	0.6296	Personal Travel
12	C3	Chicago, IL	Charlotte, NC	NA	0.501510	Gold	43	Female	1	17	0.3077	Business travel
13	C2	Chicago, IL	Colorado Sprin...	NA	0.55522	Blue	69	Female	2	49	-0.8148	Personal Travel
14	C2	Chicago, IL	Dayton, OH	NA	0.524142	Blue	33	Female	1	30	0.9355	Personal Travel
15	C3	Chicago, IL	Green Bay, WI	NA	0.54005	Blue	51	Male	1	7	0.4167	Business travel
16	C2	Chicago, IL	Colorado Sprin...	NA	0.55522	Blue	69	Female	2	49	0.8148	Personal Travel
17	C2	Chicago, IL	Topeka, KS	NA	0.596136	Blue	85	Male	1	45	-0.6364	Personal Travel
18	C2	Chicago, IL	Branson, MO	NA	0.557967	Blue	59	Female	1	18	0.2857	Personal Travel
19	C2	Chicago, IL	Baltimore, MD	NA	0.528867	Blue	19	Male	2	39	0.9500	Mileage tickets
20	C2	Dallas, TX	San Antonio, TX	Awful. They ...	0.589711	Blue	61	Male	1	54	-0.7419	Personal Travel
21	C2	Denver, CO	Minneapolis, MN	NA	0.583636	Blue	76	Male	2	28	0.6471	Personal Travel
22	C2	Detroit, MI	Baltimore, MD	Disappointing: ...	0.598839	Blue	73	Female	1	32	-0.7297	Personal Travel
23	C1	Fayetteville, NC	Atlanta, GA	NA	0.530287	Platinum	53	Female	1	8	0.5556	Business travel
24	C2	Gainesville, FL	Atlanta, GA	NA	0.573062	Blue	46	Female	1	45	-0.4286	Personal Travel
25	C1	Grand Junction, ...	Denver, CO	NA	0.541508	Gold	42	Female	1	3	0.7391	Business travel
26	C2	Hartford, CT	Chicago, IL	NA	0.538352	Silver	61	Male	1	14	-0.4737	Business travel
27	C2	Hartford, CT	Detroit, MI	NA	0.601252	Blue	78	Female	1	32	0.8286	Personal Travel

The Passive and Promoters are divided into two clusters each named C1 & C2 for each category.

Snapshot of Passive clusters:

Passive Clusters

	Cluster	Destination.City	Origin.City	freeText	Silhouette	Airline.Status	Age	Gender	Price.Sensitivity	Flights.Per.Year	Loyalty	Type.of.Travel
1	C2	Asheville, NC	Atlanta, GA	NA	0.511581	Blue	19	Male	2	24	-0.5484	Business travel
2	C2	Atlanta, GA	Montgomery, AL	Nice flight to ...	0.524466	Blue	27	Female	1	55	-0.5278	Business travel
3	C1	Atlanta, GA	Buffalo, NY	NA	0.574741	Blue	25	Female	1	0	1.0000	Business travel
4	C2	Atlanta, GA	Roanoke, VA	NA	0.554445	Gold	47	Male	1	25	-0.8519	Business travel
5	C1	Atlanta, GA	Buffalo, NY	NA	0.574741	Blue	25	Female	1	0	1.0000	Business travel
6	C1	Atlanta, GA	Chicago, IL	NA	0.541456	Blue	25	Female	1	12	0.0400	Business travel
7	C2	Atlanta, GA	Indianapolis, IN	With there was...	0.515269	Blue	26	Female	1	7	-0.7500	Business travel
8	C2	Atlanta, GA	Montgomery, AL	Stop over flight...	0.50096	Blue	23	Female	1	8	0.4545	Personal Travel
9	C1	Baltimore, MD	Flint, MI	NA	0.510186	Silver	42	Female	1	20	-0.4286	Business travel
10	C2	Boston, MA	Houston, TX	NA	0.518733	Blue	77	Female	1	6	0.0769	Personal Travel
11	C2	Boston, MA	Houston, TX	NA	0.516099	Blue	77	Female	1	6	0.0769	Personal Travel
12	C2	Boston, MA	Baltimore, MD	NA	0.57229	Blue	76	Female	2	37	-0.8974	Business travel
13	C2	Charlotte, NC	Baltimore, MD	NA	0.572584	Blue	77	Female	1	40	-0.6327	Personal Travel
14	C2	Chicago, IL	Atlanta, GA	NA	0.550481	Blue	65	Male	1	18	-0.3846	Personal Travel
15	C2	Chicago, IL	Cleveland, OH	NA	0.555325	Blue	52	Male	0	51	-0.9245	Personal Travel
16	C1	Chicago, IL	Denver, CO	NA	0.526095	Silver	46	Male	2	14	-0.5667	Mileage tickets
17	C2	Chicago, IL	Omaha, NE	NA	0.56797	Blue	56	Female	2	38	-0.9487	Business travel
18	C2	Chicago, IL	San Antonio, TX	NA	0.5664	Blue	60	Female	1	33	-0.7368	Mileage tickets
19	C2	Cleveland, OH	Albany, NY	NA	0.550012	Blue	62	Female	1	55	-0.7647	Mileage tickets
20	C2	Cleveland, OH	Detroit, MI	NA	0.575678	Blue	70	Male	1	36	-0.8462	Personal Travel
21	C2	Dallas, TX	New Orleans, LA	NA	0.523729	Silver	39	Female	1	36	-0.7561	Business travel
22	C2	Houston, TX	Grand Rapids, MI	NA	0.553505	Blue	52	Female	1	37	-0.7339	Personal Travel
23	C2	Houston, TX	Lafayette, LA	NA	0.498196	Blue	47	Female	2	22	-0.6296	Personal Travel
24	C2	Houston, TX	Las Vegas, NV	NA	0.568635	Blue	71	Female	1	26	-0.6250	Personal Travel
25	C2	Huntsville, AL	Houston, TX	NA	0.527302	Blue	56	Male	1	41	-0.6400	Business travel
26	C2	Lansing, MI	Chicago, IL	NA	0.562807	Blue	37	Male	1	46	-0.7337	Personal Travel
27	C2	Las Vegas, NV	San Francisco, CA	NA	0.537203	Silver	41	Male	1	56	-0.8000	Personal Travel
28	C2	Milwaukee, WI	Atlanta, GA	NA	0.537911	Blue	22	Female	2	22	-0.9130	Personal Travel
29	C1	Montgomery, AL	Atlanta, GA	NA	0.575231	Blue	37	Female	1	0	1.0000	Personal Travel

Snapshot of Promoters clusters:

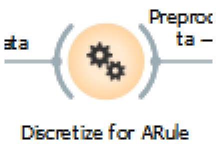
Promoter Clusters

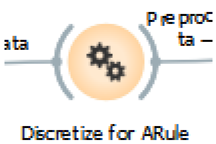
	Cluster	Destination.City	Origin.City	freeText	Silhouette	Airline.Status	Age	Gender	Price.Sensitivity	Flights.Per.Year	Loyalty	Type.of.Travel
1	C2	Baton Rouge, LA	Atlanta, GA	NA	0.573339	Silver	70	Male	1	35	-0.7949	Personal Travel
2	C2	Burlington, VT	Chicago, IL	NA	0.541073	Blue	52	Male	2	19	-0.3571	Mileage tickets
3	C1	Chicago, IL	St. Louis, MO	NA	0.497163	Platinum	34	Male	0	15	-0.2500	Personal Travel
4	C1	Chicago, IL	Indianapolis, IN	NA	0.512131	Platinum	37	Male	2	19	-0.4615	Business travel
5	C1	Chicago, IL	Detroit, MI	NA	0.553182	Silver	35	Female	0	0	1.0000	Personal Travel
6	C2	Cleveland, OH	Denver, CO	NA	0.567088	Silver	39	Female	2	26	-0.7931	Business travel
7	C2	Fayetteville, AR	Atlanta, GA	NA	0.56791	Silver	54	Female	1	26	-0.6774	Personal Travel
8	C1	Green Bay, WI	Chicago, IL	NA	0.499518	Blue	17	Female	1	11	-0.3750	Business travel
9	C2	Hartford, CT	Chicago, IL	NA	0.519739	Silver	65	Male	2	3	0.1429	Business travel
10	C2	Memphis, TN	Chicago, IL	NA	0.543616	Blue	67	Male	1	14	-0.3333	Mileage tickets
11	C1	Minneapolis, MN	Milwaukee, WI	NA	0.520276	Blue	48	Female	1	7	0.0000	Business travel
12	C2	New York, NY	Detroit, MI	NA	0.56161	Blue	58	Female	1	26	-0.4857	Mileage tickets
13	C2	New York, NY	St. Louis, MO	NA	0.555405	Silver	30	Male	1	47	-0.7407	Mileage tickets
14	C2	Orlando, FL	Providence, RI	NA	0.541654	Blue	32	Female	1	13	-0.6250	Business travel
15	C1	Salt Lake City, UT	Chicago, IL	NA	0.554849	Silver	39	Male	1	4	0.5789	Business travel
16	C1	Salt Lake City, UT	Chicago, IL	NA	0.554849	Silver	39	Male	1	4	0.5789	Business travel
17	C2	Springfield, MO	Denver, CO	NA	0.532566	Blue	50	Male	2	24	-0.5000	Business travel
18	C1	Washington, DC	Albany, NY	NA	0.492382	Blue	19	Female	1	16	-0.3333	Business travel
19	C1	Akron, OH	Denver, CO	NA	0.504719	Blue	28	Female	1	13	-0.4444	Business travel
20	C2	Akron, OH	Denver, CO	NA	0.572385	Blue	53	Female	1	26	-0.9259	Business travel
21	C1	Akron, OH	Denver, CO	NA	0.495671	Silver	59	Female	1	3	0.3333	Business travel
22	C1	Albany, NY	Orlando, FL	NA	0.522155	Gold	55	Female	2	0	1.0000	Business travel
23	C2	Albany, NY	Chicago, IL	NA	0.580843	Blue	55	Female	2	34	-0.9429	Business travel
24	C2	Albany, NY	Baltimore, MD	NA	0.544983	Gold	32	Male	1	22	-0.6923	Business travel
25	C2	Albany, NY	Chicago, IL	NA	0.530785	Silver	49	Female	1	9	-0.2000	Business travel
26	C2	Albany, NY	Baltimore, MD	NA	0.559779	Silver	30	Female	2	22	-0.9130	Business travel
27	C2	Albuquerque, NM	Las Vegas, NV	NA	0.5361	Silver	29	Female	2	10	-0.5000	Business travel
28	C2	Albuquerque, NM	Las Vegas, NV	NA	0.565026	Blue	62	Male	2	36	-0.8000	Personal Travel
29	C2	Albuquerque, NM	Dallas, TX	NA	0.550982	Blue	25	Male	2	29	-0.5263	Business travel

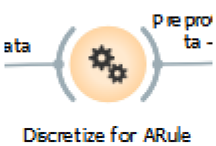
NPS Category Segmentation:

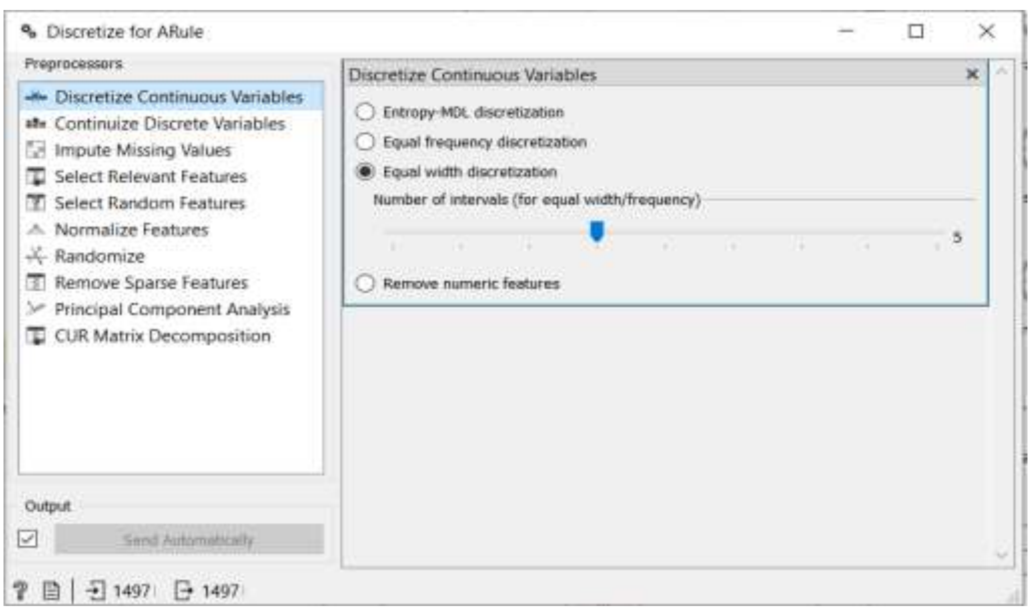
➢ **Detractor Clusters:**

Each observation in clusters C1, C2 & C3 is checked for creating a frequent itemset list using Association Rules. The data is discretized before generating Association rules.

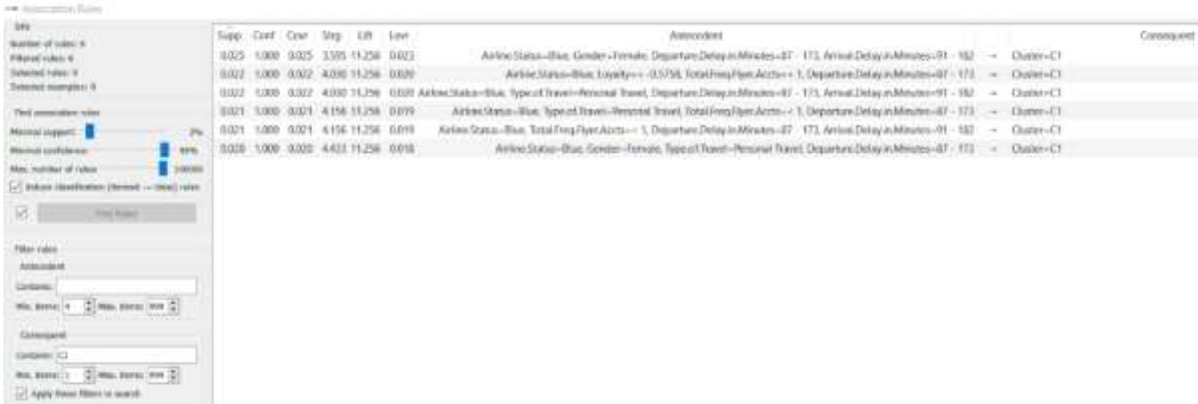








- Association Rules for Cluster C1:



- Association Rules for Cluster C2:

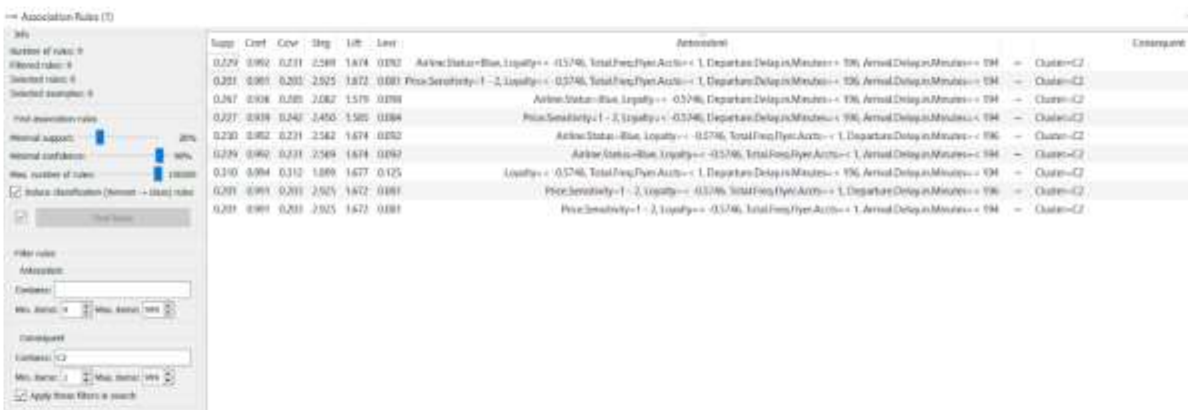
- Association Rules for Cluster C3:

➢ Passive Clusters:

Each observation in clusters C1 & C2 is checked for creating a frequent itemset list using Association Rules. The data is discretized before generating rules.

- Association Rules for Cluster C1:

- Association Rules for Cluster C2:



- **Promoter Clusters:**

  Each observation in clusters C1 & C2 is checked for creating a frequent itemset list using Association Rules. The data is discretized before generating rules.

  - Association Rules for Cluster C1:

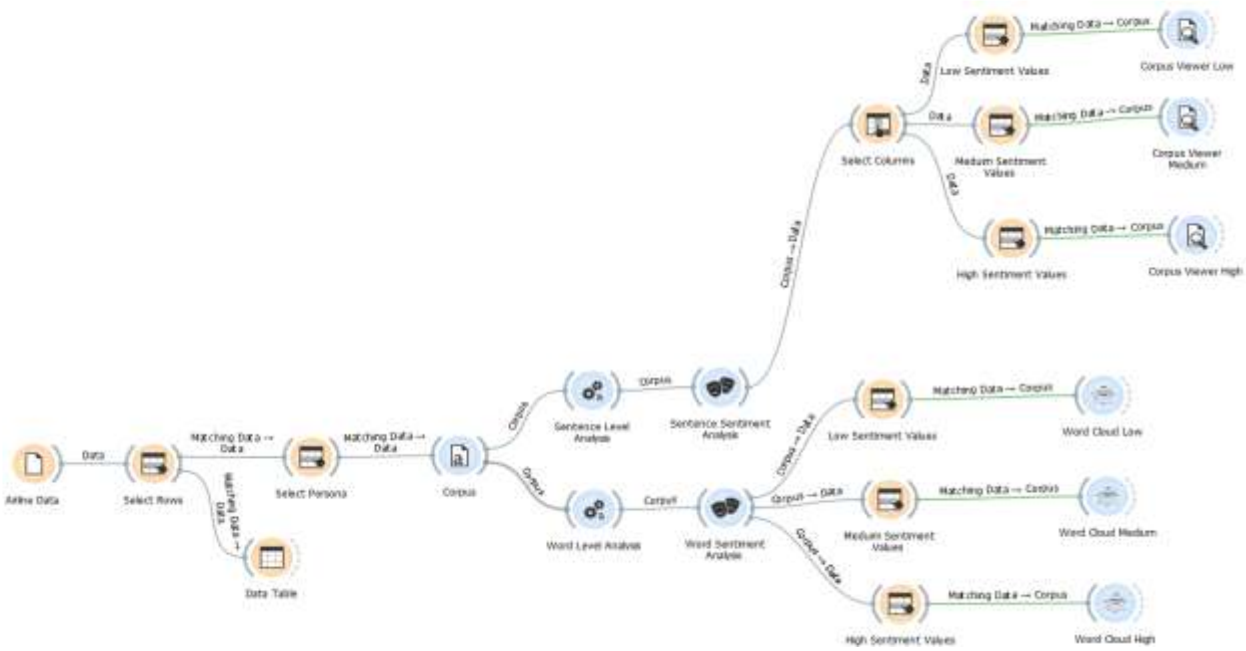
  - Association Rules for Cluster C2:

Segments created from Clustering Analysis and Association Rules:

	DETRACTORS		PASSIVE		PROMOTERS	
PERSONA/SEGMENT						
C1	Airline.Status	Blue	Price.Sensitivity	1 to 2	Age	29 to 43 years
	Type.of.Travel	Personal Travel	Flights.Per.Year	Less than 13	Flights.Per.Year	Less than 16
	Total.Freq.Flyer.Accts	Less than 1	Type.of.Travel	BusinessTravel	Type.of.Travel	BusinessTravel
	Departure.Delay.in.Minutes	87-173 mins	Departure.Delay.in.Minutes	Less than 196 mins	Departure.Delay.in.Minutes	Less than 74 mins
			Arrival.Delay.in.Minutes	Less than 194 mins	Arrival.Delay.in.Minutes	Less than 68 mins
C2	Airline.Status	Blue	Airline.Status	Blue	Price.Sensitivity	1 to 1
	Total.Freq.Flyer.Accts	Less than 1	Loyalty	Less than -0.5746	Type.of.Travel	BusinessTravel
	Departure.Delay.in.Minutes	Less than 87 mins	Total.Freq.Flyer.Accts	Less than 1	Total.Freq.Flyer.Accts	Less than 1
	Gender	Female	Departure.Delay.in.Minutes	Less than 196 mins	Departure.Delay.in.Minutes	Less than 74 mins
	Arrival.Delay.in.Minutes	Less than 91 mins				
C3	Airline.Status	Blue				
	Type.of.Travel	BusinessTravel				
	Departure.Delay.in.Minutes	Less than 87 mins				
	Arrival.Delay.in.Minutes	Less than 91 mins				

Phase4: Sentiment Analysis

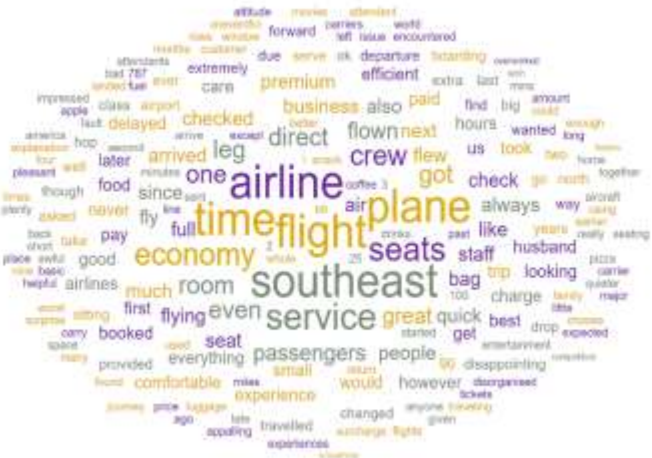
Here we have about 101 Free texts (comments) responses that were received by the airlines. Based on the personas obtained through the segmentation we observed the data and obtained some insightful information about the different airline customers.



- High Sentiment Cloud:

Here as per the 21 records in high sentiment, we can observe from the word cloud that southeast airline gives a comfortable experience to its customers, most of the customers have rated their experience as good and they find the service provided by the airline good.

● Medium Sentiment Cloud:



Here as per the 42 records in medium sentiment, we can find out that the economy is more focused on, so the customers are trying to focus more on the economy class of the airline. There might or might not be a problem with the seats of the airline. Again, the service and crew are focused on which tells us that they play a key role for our medium sentiment customers.

● Low Sentiment Cloud:

Here from the 38 instances, we can observe that again Service is more focused on, so this category of customers is not happy with the service provided by the airline. There is also a time issue which shows us that there is a delay in the arrival time of the flights.

- **FlyFast Airways Sentiment Analysis:**

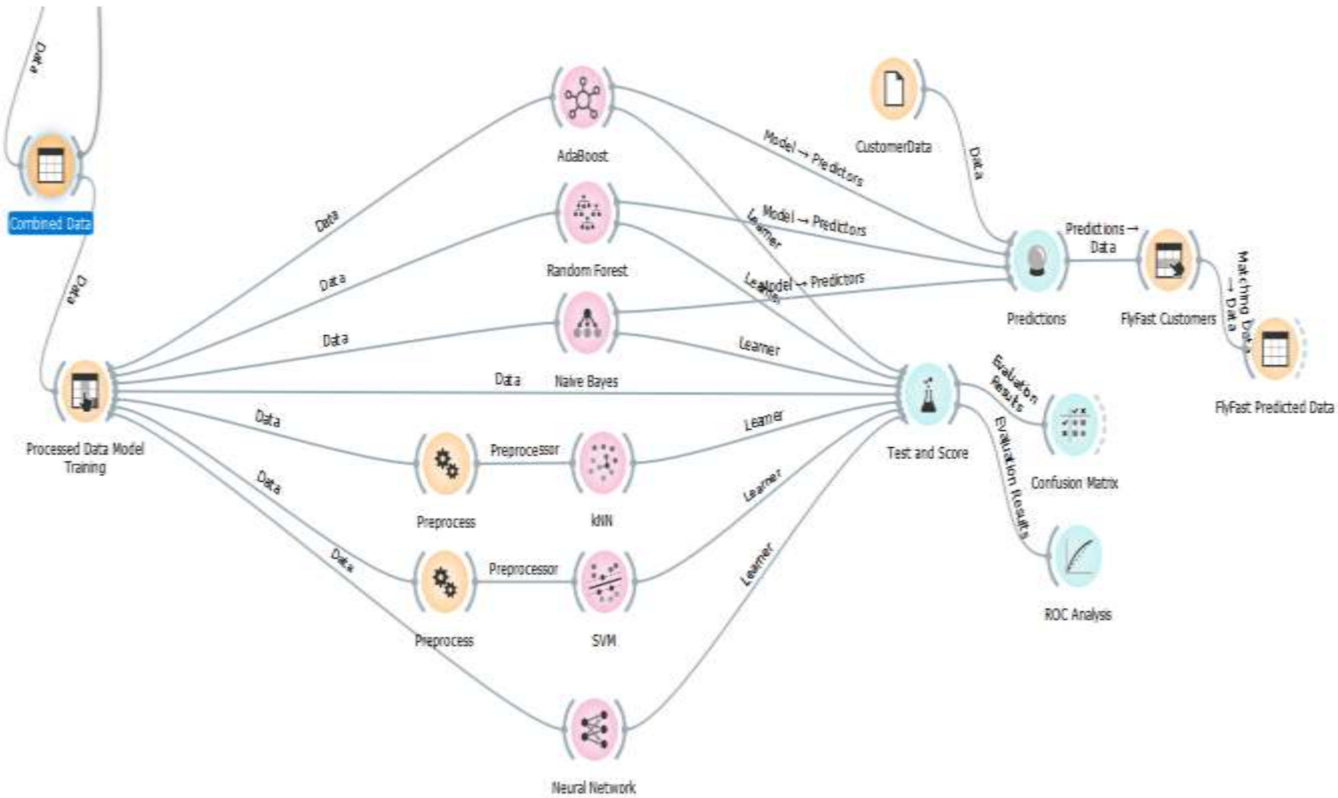
    - Low Sentiment Analysis:

Here the customers are more concerned about the flight time of the flight this might be the reason behind more detractors for this partner airline. Again, the service of this flight is more focused on, So the service of this flight should be more focused on.

    - High Sentiment Analysis:

We didn't get a better analysis of high sentiments in fly fast which shows us that the number of detractors is large, and we should focus on the areas mentioned in the low sentiment analysis.

__Phase5: Predictive Modelling:__



The above snapshot shows all the predictive models used for predicting the NPS category based on the training dataset of 4985 survey data instances.

Starting from the models which gave the least accuracy, we have Neural Networks and kNN models which gave similar results in terms of accuracy and precision/recall values.

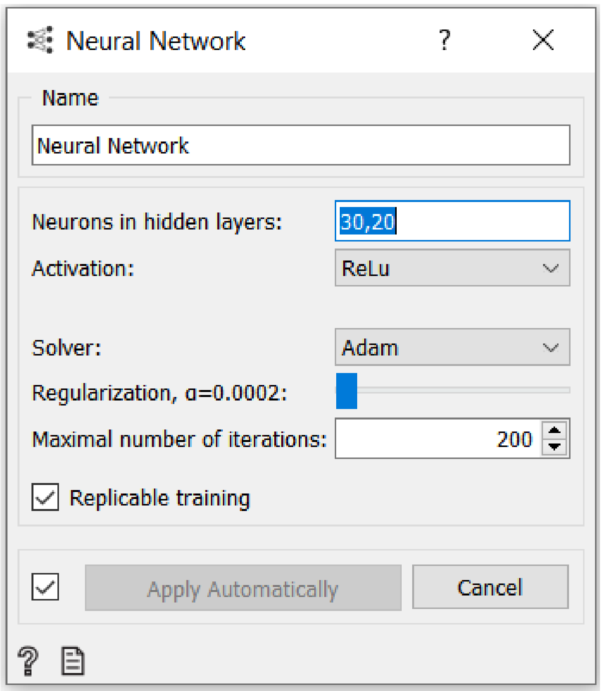
1) Neural Networks:

Neural Networks are categorized by a number of "hidden layers" of neurons between the input and output and are used to make classifications and predictions. The hidden layers are in place to allow the study of several complex relationships that might exist in our data and assist in better classification.

Pre-processing:

No pre-processing step is required for Neural Networks as it can handle dirty data and recognizes sophisticated patterns in the data.

Model Optimization:



- **Activation Function**

  Keeping other parameters constant, we tried different activation functions: "ReLu", "Identity" and "Logistic". In the case of logistic and identity functions, not only did it take longer to build the model, but accuracy dropped significantly (<0.4). ReLu was the only one that showed an improved performance.

- **Solver**

  This indicates the optimization function used. The "L-BFGS-B" adversely affects the model. The "SGD" (stochastic gradient descent) algorithm offers a slight improvement in accuracy but generates a biased model. The default "Adam", solver technique gives the best accuracy for all three NPS categories.

- **Neurons in hidden layers**

  We have added two hidden layers with 30 Neurons in the first layer and 20 neurons in the second layer. High numbers of neurons can cause overfitting, so the default 100 neurons were brought down to these numbers.

- **Iterations**

  For the number of iterations as well, changing it did not produce any significant changes in model performance so it has defaulted to 200.

- **Regularization**

  We use regularization to add penalties to the model and to prevent the network from overfitting. Changing the value of α did not affect the network in any way either.

Model Evaluation:

Classification Accuracy:

- Promoter: 0.557
- Passive: 0.531
- Detractor: 0.580

Confusion Matrix:

Show: Proportion of actual

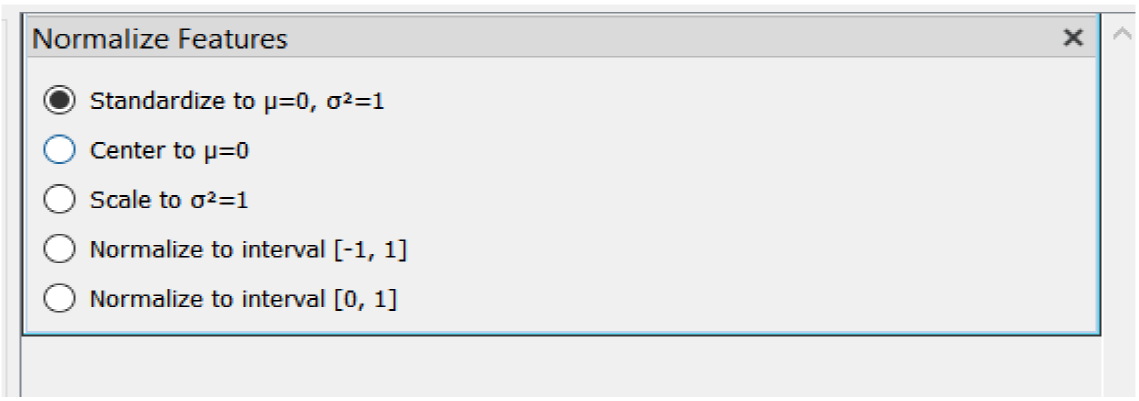
Actual \ Predicted	Detractor	Passive	Promoter	Σ
Detractor	29.9 %	40.1 %	30.0 %	1497
Passive	30.0 %	40.0 %	30.1 %	1707
Promoter	30.0 %	40.0 %	30.0 %	1781
Σ	1494	1994	1497	4985

2) K – Nearest Neighbor (KNN):

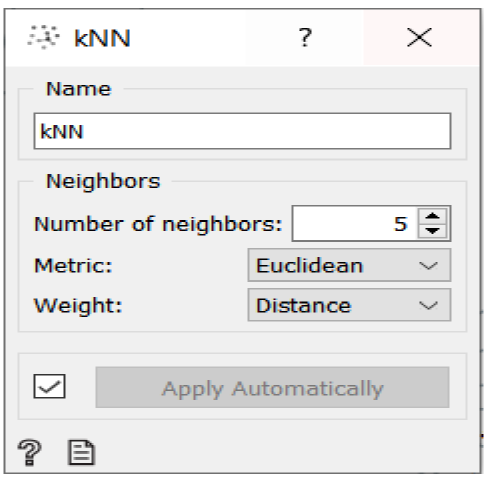
The K – Nearest Neighbor algorithm works on the simple technique of identifying what values its nearest data points have and when a new data point comes in, it is assigned a value based on a vote from the nearest neighbor. This algorithm checks for its neighbors based on distance measure hence it is important to have all the features to be on the same scale to avoid any dominance of non-important variables just because of distance.

Pre-Processing:

The data needs to be pre-processed and the features are standardized, which means the features now will be used based on Z-score values thus bringing everything on the same scale.



Model Optimization:



- Number of Neighbors:
  When the data was tested on different values, 5 was the optimal number of neighbors for getting the highest accuracy. For our dataset, Euclidean distance was used for optimizing the algorithm.

Model Evaluation:

The KNN model has below-listed classification accuracy for different NPS categories of target variable:

- Promoters: 0.636
- Passive: 0.590
- Detractors: 0.703

Confusion Matrix:

Show: Proportion of actual

Predicted

Actual	Detractor	Passive	Promoter	Σ
Detractor	44.4 %	30.3 %	25.3 %	1497
Passive	23.6 %	37.8 %	38.5 %	1707
Promoter	13.9 %	29.7 %	56.4 %	1781
Σ	1315	1628	2042	4985

The KNN model is among the lowest accuracy models with low recall value when compared to Random Forest and Naïve Bayes models.

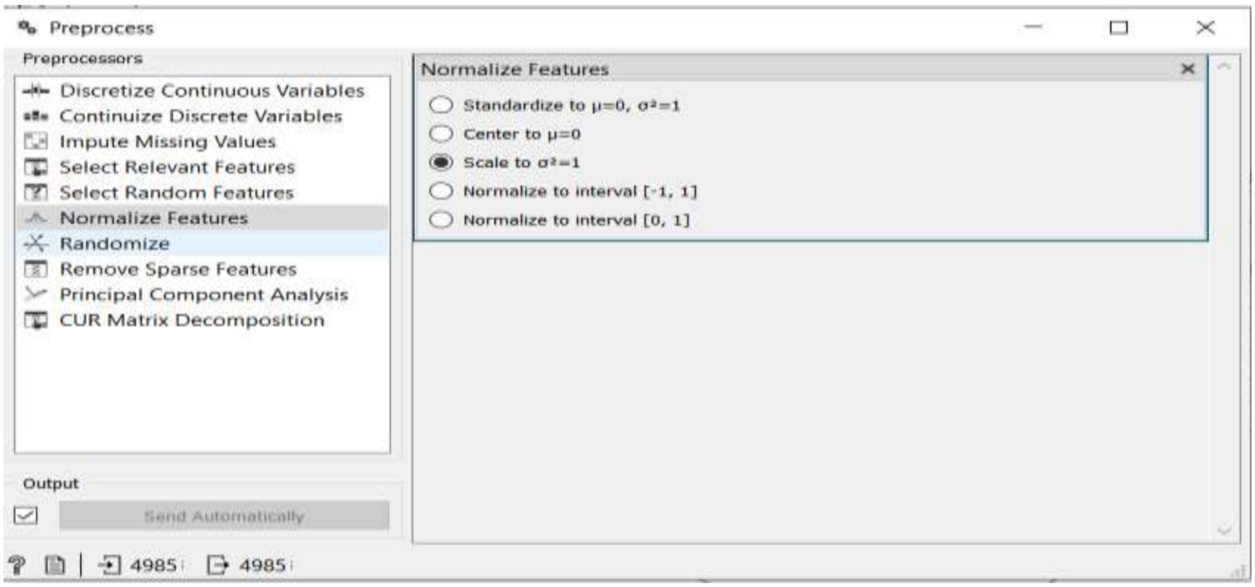
3) Support Vector Machines:

An SVM performs classification tasks by constructing hyperplanes in a multidimensional space that separates data points of different class labels. Here, we have three classes: Promoter, Passive, and Detractor, into which the customers are classified.

Pre-processing:

Half of the features in our data are numeric, with varying units of measurement. They also vary widely in the range of values. Since SVM finds an optimal hyperplane by evaluating the distance vectors among variables, it is important to scale these variables before employing SVM. If we do not scale the features, those with the widest range may unfairly dominate over others.

We used the "Scale to $\sigma^2=1$" option under "Normalize features"

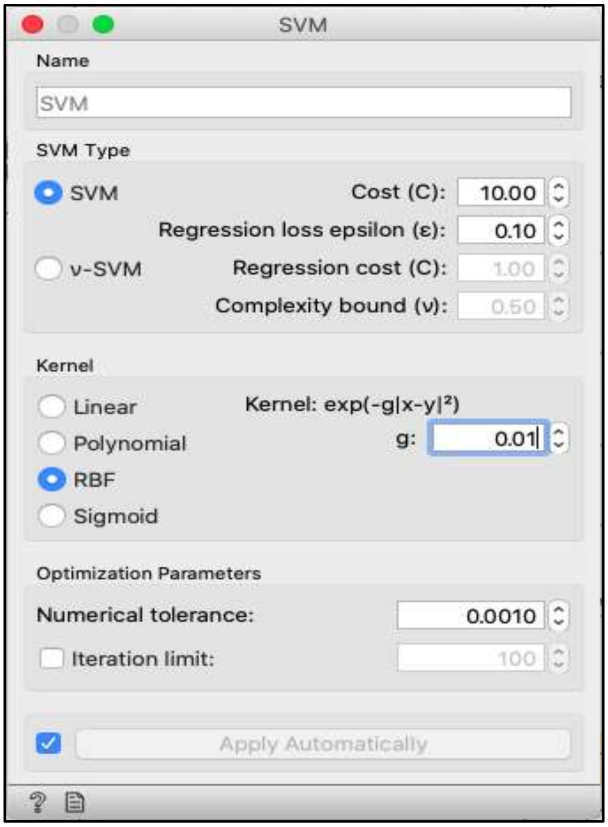


First, we employed the SVM model with default parameters and checked its performance.

It had a low classification accuracy in the range of 0.503 - 0.584 for each of the classes. To improve this, we tried tuning some parameters such as the kernel, cost (of error), and gamma.

Model Optimization:

Our final model has the following specifications:



- **Kernel: Radial Basis Function**

  The kernel is the function used to calculate similarity or closeness between data points. We used RBF as we have as many as 19 features in our data that we use for classification. RBF is suitable for such high dimensions, unlike the Linear kernel.

- **C parameter: C=10**

  This can be called the penalty parameter. It handles the tradeoff between a smooth boundary and classifying the training points correctly. A large C value will classify all training data most accurately and create a fine boundary but may lead to overfitting. Smaller C values will create a bigger boundary.

- **Gamma: g=0.01**

  Gamma is useful for multi-dimensional hyperplanes. The higher the gamma the more exactly it tries to fit the data. But too high a value can cause overfitting.

Model Evaluation:

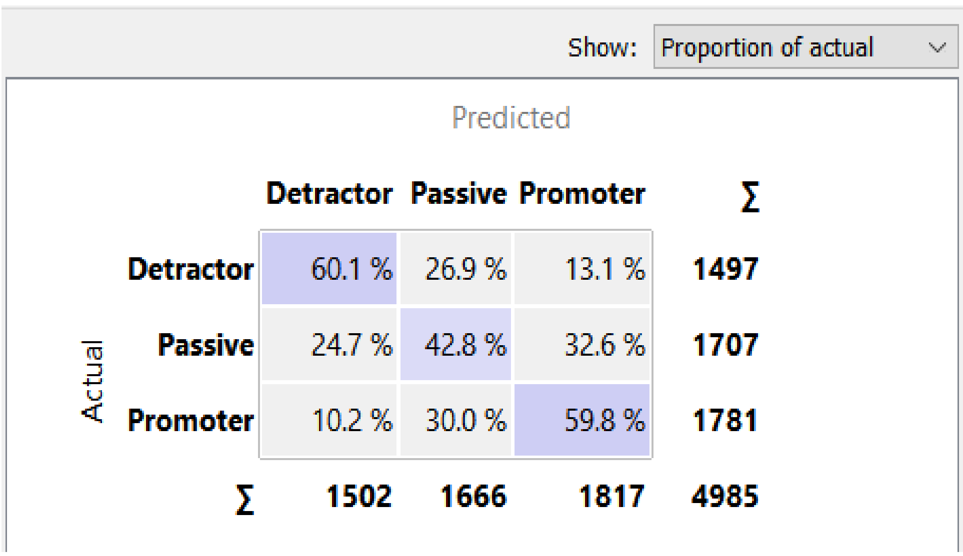
The above SVM model has classification accuracy as:

- Promoter: 0.706
- Passive: 0.616
- Detractor: 0.759

There is a slight improvement from the previous model with default parameters. However, it is still slightly less accurate in comparison to other predictive models that we have used.

Confusion Matrix:

From the confusion matrix below, we can observe the proportion of correct and incorrect classifications:


Show: Proportion of actual

Predicted

Actual	Detractor	Passive	Promoter	∑
Detractor	60.1 %	26.9 %	13.1 %	1497
Passive	24.7 %	42.8 %	32.6 %	1707
Promoter	10.2 %	30.0 %	59.8 %	1781
∑	1502	1666	1817	4985

We see that SVM can correctly predict close to 60% of the instances within any of the three categories. Also, among all the incorrect NPS category classifications, it has the highest amount of cases misclassified as "Passive".

4) Naïve Bayes:

We also employed the Naive Bayes algorithm over the data for classification. It is a fast and simple probabilistic classifier based on Bayes' theorem with the assumption of feature independence. It yields competitive classification accuracies as shown:

Model Evaluation:

The classification accuracy for the Naïve Bayes Model is as follows:

- Promoter: 0.714
- Passive: 0.620
- Detractor: 0.779

Confusion Matrix:

Show: Proportion of actual

Actual \ Predicted	Detractor	Passive	Promoter	Σ
Detractor	63.6 %	22.8 %	13.6 %	1497
Passive	26.0 %	35.6 %	38.5 %	1707
Promoter	6.3 %	25.3 %	68.3 %	1781
Σ	1508	1400	2077	4985

From the confusion matrix, we can see that it can correctly classify well over 60% of the promoters and detractors. It seems to have some trouble correctly classifying passive cases though. The model performs better than SVM, KNN, and Neural Network model and gives accuracy close to that of Random Forest and Adaptive Boost models.

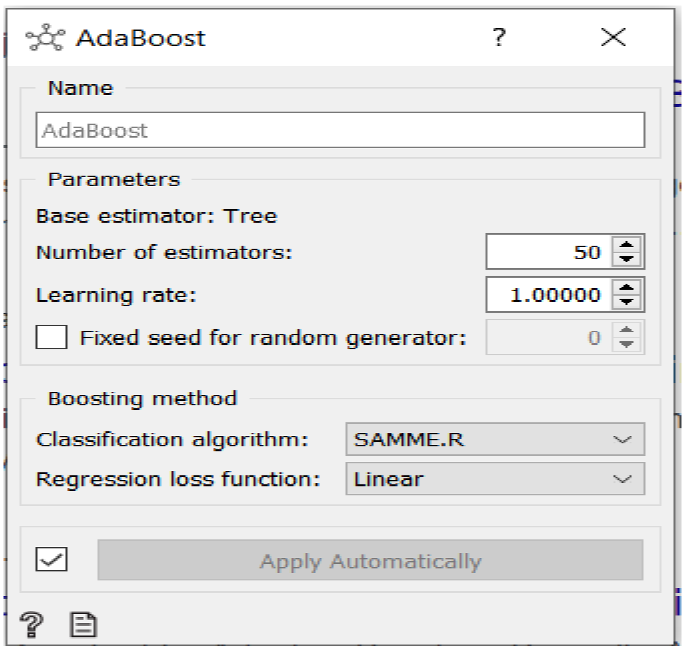
5) Adaptive Boosting (AdaBoost):

The AdaBoost is an ensemble learner algorithm that combines weak learners and adapts to the hardness of each training sample. It is similar to the Random Forest algorithm but AdaBoost uses a sequential ensemble technique as opposed to Random Forest which runs trees in parallel. The AdaBoost technique works on reducing errors from the previous round of predictors, which might bring variance issues leading to overfitting issues. Generally, this algorithm is used with other learning algorithms to boost performance.

Pre-Processing:

Adaptive Boosting is similar to Random Forest; therefore, no pre-processing is required as the algorithm is based on decision trees as well.

Model Optimization:



- **The number of estimators:**
  This is the maximum number of estimators at which boosting is terminated. It has defaulted to 50.

- **Learning Rate:**
  It is set to 1 so that it considers only the most recent information and errors. This is like a Long Short-Term Memory concept (LSTM).

- The fixed seed option was disabled as we did not want to reproduce the results. SAAME.R which considers probability for weight estimation gave better results than SAMME and Linear regression loss function was used.

Model Evaluation:

The AdaBoost model has below-listed classification accuracy for different NPS categories of target variable:

- Promoters: 0.716
- Passive: 0.641
- Detractors: 0.804

Confusion Matrix:

Show: Proportion of actual

Predicted

Actual	Detractor	Passive	Promoter	∑
Detractor	60.9 %	23.8 %	15.2 %	1497
Passive	18.6 %	45.2 %	36.2 %	1707
Promoter	4.2 %	27.8 %	67.9 %	1781
∑	1304	1625	2056	4985

The AdaBoost model is among the highest accuracy models with good recall value when compared to Random Forest and Naïve Bayes models. The model does better in predicting Detractors and Promoters but lacks in Passive NPS category prediction.

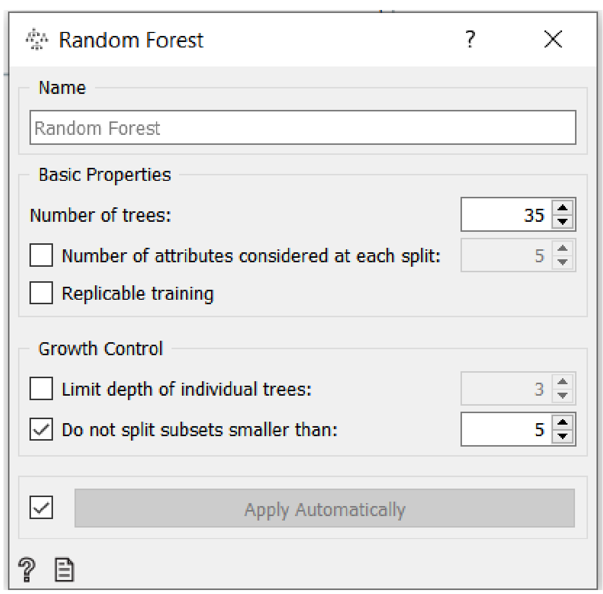
6) Random Forest:

Random forest is based on decision trees, where the algorithm creates different decision trees based on a different sample of data and different independent sets of explanatory variables. Then based on bagged random decision trees, a voting mechanism is carried out, and based on multiple such bags we arrive at a decision. This method reduces variance due to different samples of data and solves bias problems due to avoiding any over/under prediction of the NPS category.

Pre-Processing:

No pre-processing is required for Random Forest as the algorithm is based on decision trees and decision trees are not sensitive to outliers. Prediction is the average/majority class, where outlier data handling is not required. We can do better to create better decision trees in general by removing variables that do not contribute to making any decisions that have been done in the common pre-processing part used for all other algorithms.

Model Optimization:



- Number of Trees:
  After checking for different values of a total number of trees to be considered on this data for Random Forest, 35 was the best number bringing the highest accuracy and recall value.

- Do not split subsets smaller than:
  This option is a pre-pruning option to avoid individual decision trees having less depth, in our case trees less than 5 levels will not be split for further consideration.

Model Evaluation:

The Random Forest model has below-listed classification accuracy for different NPS categories of target variable:

- Promoters: 0.724
- Passive: 0.660
- Detractors: 0.822

Confusion Matrix:

Show: Proportion of actual

Predicted

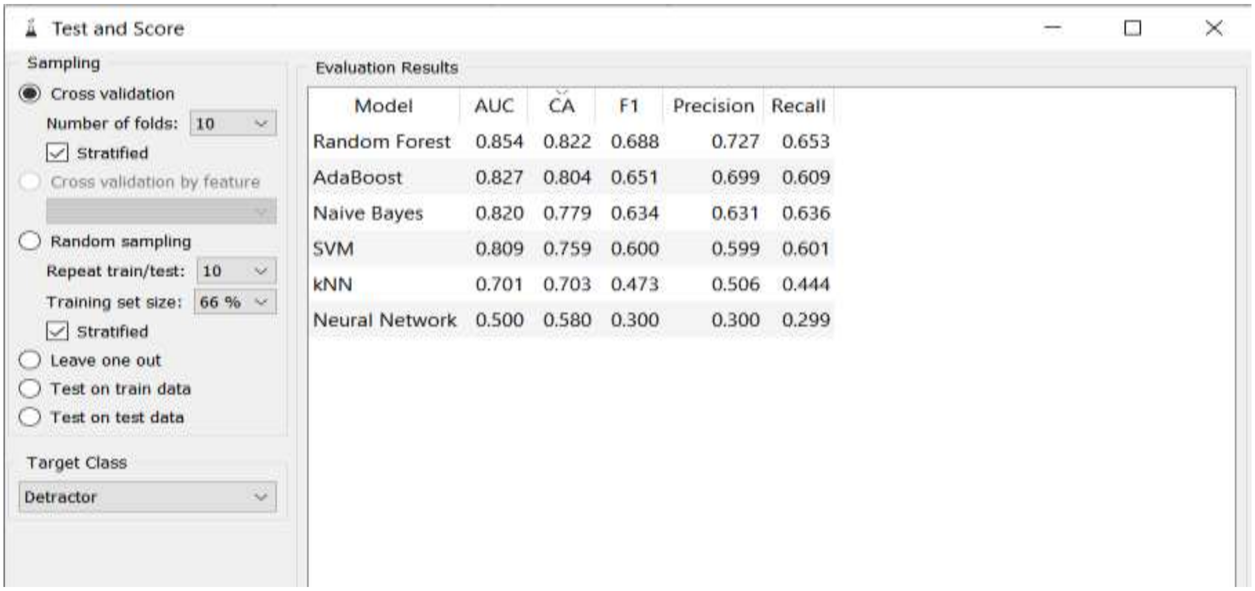
Actual	Detractor	Passive	Promoter	∑
Detractor	65.3 %	20.1 %	14.6 %	1497
Passive	17.6 %	43.5 %	38.9 %	1707
Promoter	3.6 %	24.1 %	72.2 %	1781
∑	1343	1473	2169	4985

The model not only shows the highest accuracy among all other models but also gives better recall value among others which are important in this scenario.

Overall Model Evaluation and Comparison:

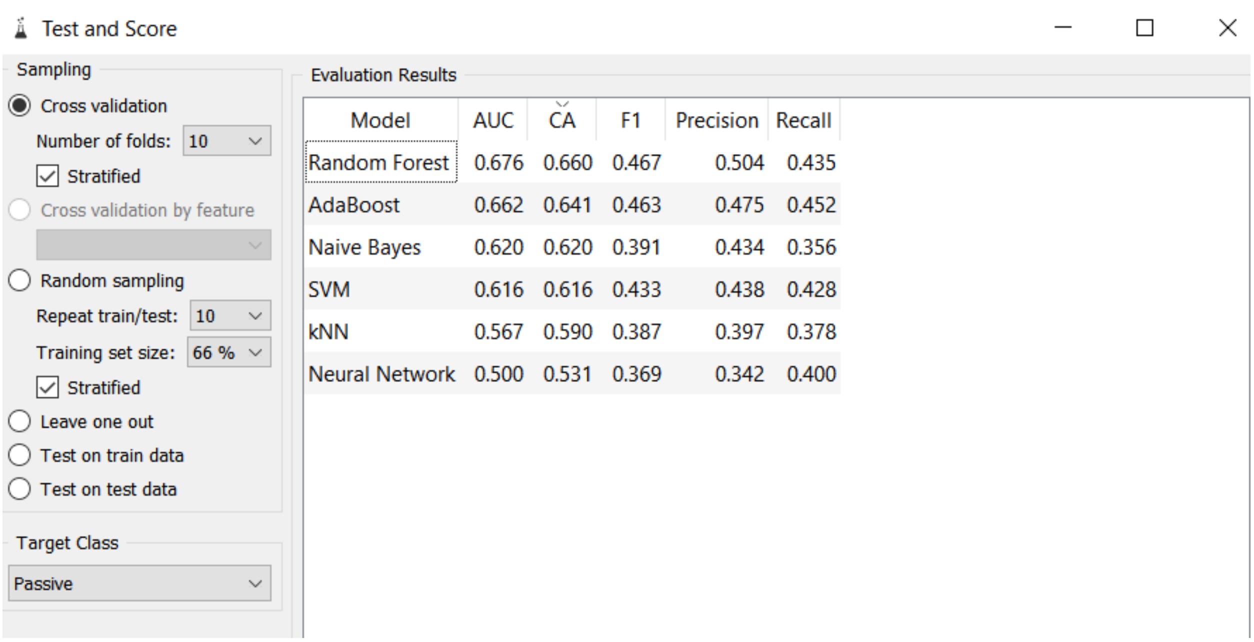
➢ **Test and Score**

- **Detractors**: Clearly, Random Forest performs better considering all the performance metrics like classification accuracy, F1 score as well as precision and recall. Highest accuracy of 82.2% in prediction of detractors



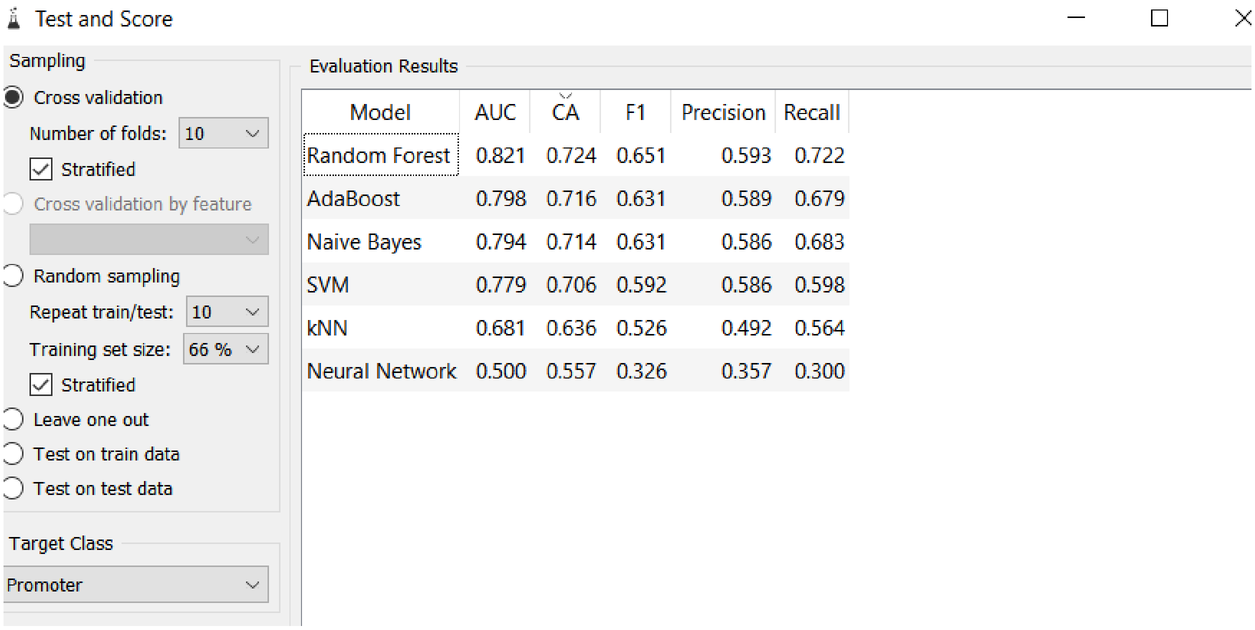
Model	AUC	CA	F1	Precision	Recall
Random Forest	0.854	0.822	0.688	0.727	0.653
AdaBoost	0.827	0.804	0.651	0.699	0.609
Naive Bayes	0.820	0.779	0.634	0.631	0.636
SVM	0.809	0.759	0.600	0.599	0.601
kNN	0.701	0.703	0.473	0.506	0.444
Neural Network	0.500	0.580	0.300	0.300	0.299

- **Passive:** In this NPS category prediction as well, the Random Forest performs better than all the other models with the highest accuracy of 66% in passive customers prediction



Model	AUC	CA	F1	Precision	Recall
Random Forest	0.676	0.660	0.467	0.504	0.435
AdaBoost	0.662	0.641	0.463	0.475	0.452
Naive Bayes	0.620	0.620	0.391	0.434	0.356
SVM	0.616	0.616	0.433	0.438	0.428
kNN	0.567	0.590	0.387	0.397	0.378
Neural Network	0.500	0.531	0.369	0.342	0.400

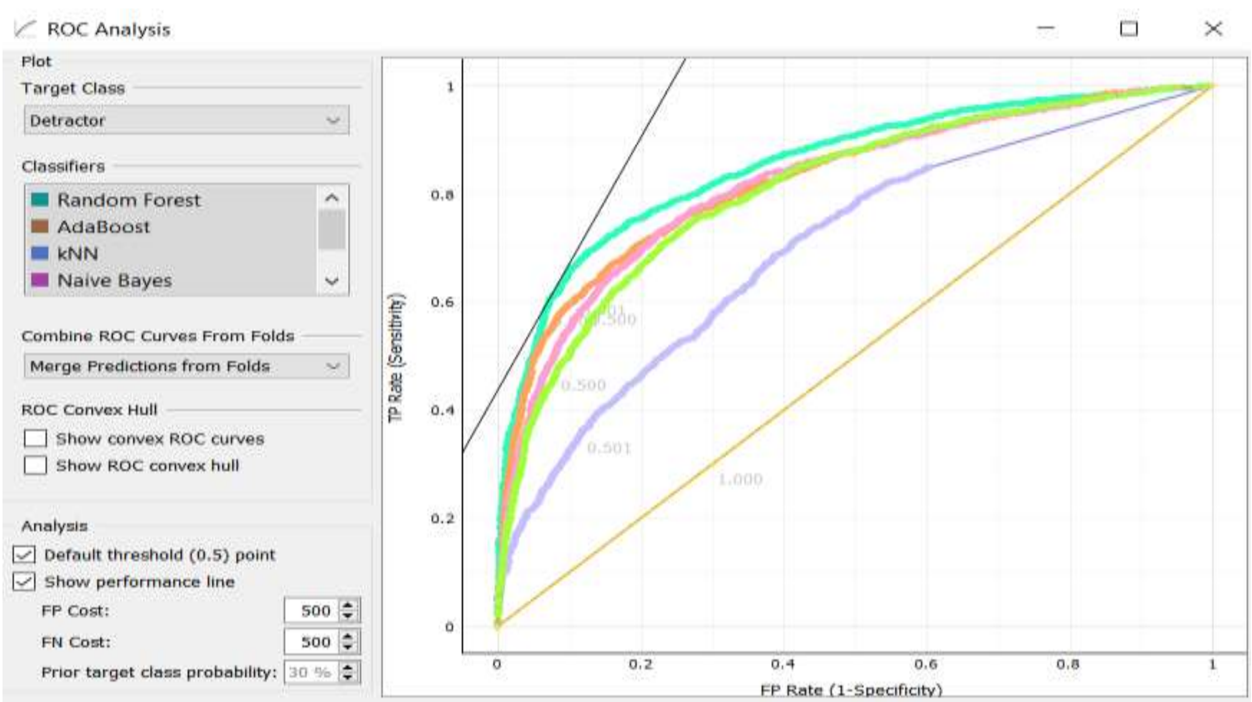
- **Promoters:** Finally, even for Promoters NPS category prediction, the Random Forest model performs the best with the highest accuracy of 72.4 %



Model	AUC	CA	F1	Precision	Recall
Random Forest	0.821	0.724	0.651	0.593	0.722
AdaBoost	0.798	0.716	0.631	0.589	0.679
Naive Bayes	0.794	0.714	0.631	0.586	0.683
SVM	0.779	0.706	0.592	0.586	0.598
kNN	0.681	0.636	0.526	0.492	0.564
Neural Network	0.500	0.557	0.326	0.357	0.300

➢ **ROC-AUC analysis:**

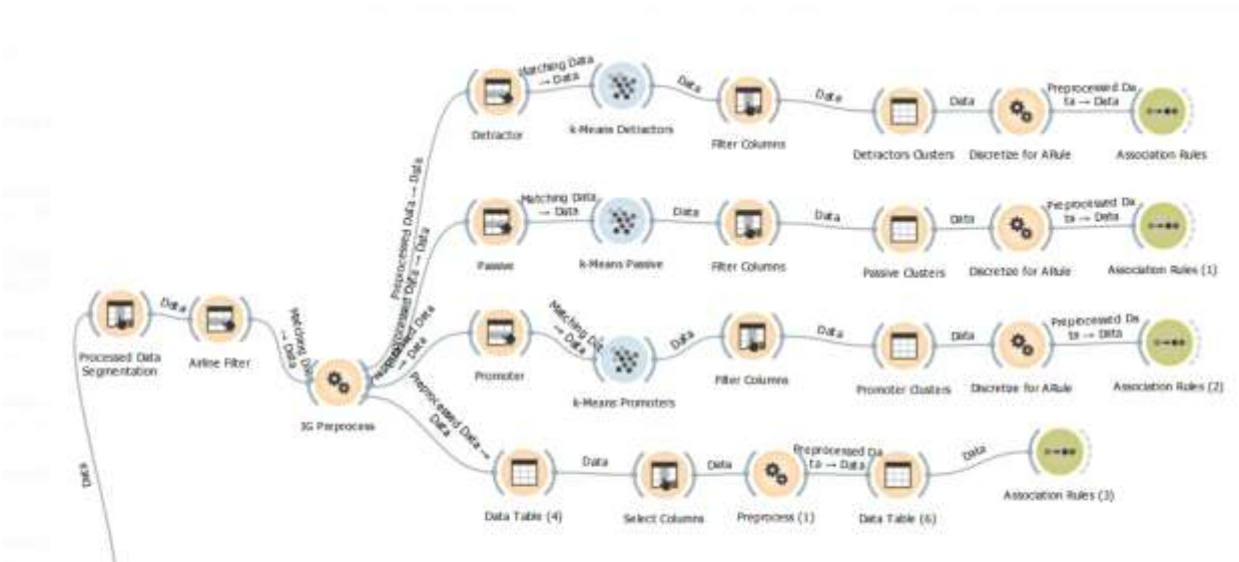
The below snapshot is the ROC analysis for Detractors and similar results are observed in other NPS categories as well. The Random Forest curve in the dark green which is closest to the Y-Axis TP Rate has the highest Area Under Curve.



Business Recommendations Development Phase:

Phase6: Business Analysis:

Using Association Rules, data were analyzed for all three airlines to get frequent items that were used for making business rules based on formed Persona in each NPS category.



- **FlyFast Airways Inc.**

NPS Category:

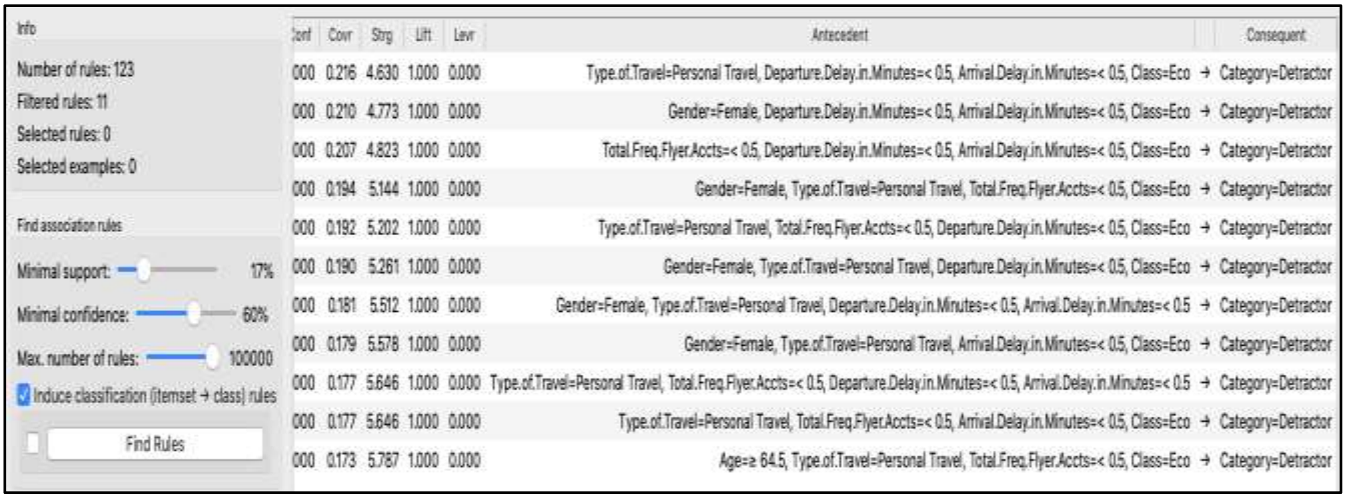
1) **Detractors**

What makes a FlyFast passenger a detractor?

- Females traveling for personal purposes.
- Flying in “Economy” class
- No frequent flyer accounts
- Delay, whether arrival or departure, <0.5 minute

Since the **delay in flight times is almost negligible, it is not the cause of disdain** among the passengers. Moreover, we see a trend of minimal delays, if any, among the passive and promoter fliers too. Similarly, each of the three classes has majority passengers flying “Economy”, so that need not be of concern either.

What stands out here are the **Female passengers, especially those on personal travel**. From the rules below, there is overwhelming evidence that this subset makes up the majority of our detractors. Also, detractors do not seem to have any Frequent Flyer Accounts and are probably unable to enjoy many benefits that they entail.



Conf	Covr	Strg	Lift	Levr	Antecedent		Consequent
000	0.216	4.630	1.000	0.000	Type.of.Travel=Personal Travel, Departure.Delay.in.Minutes=< 0.5, Arrival.Delay.in.Minutes=< 0.5, Class=Eco	→	Category=Detractor
000	0.210	4.773	1.000	0.000	Gender=Female, Departure.Delay.in.Minutes=< 0.5, Arrival.Delay.in.Minutes=< 0.5, Class=Eco	→	Category=Detractor
000	0.207	4.823	1.000	0.000	Total.Freq.Flyer.Accts=< 0.5, Departure.Delay.in.Minutes=< 0.5, Arrival.Delay.in.Minutes=< 0.5, Class=Eco	→	Category=Detractor
000	0.194	5.144	1.000	0.000	Gender=Female, Type.of.Travel=Personal Travel, Total.Freq.Flyer.Accts=< 0.5, Class=Eco	→	Category=Detractor
000	0.192	5.202	1.000	0.000	Type.of.Travel=Personal Travel, Total.Freq.Flyer.Accts=< 0.5, Departure.Delay.in.Minutes=< 0.5, Class=Eco	→	Category=Detractor
000	0.190	5.261	1.000	0.000	Gender=Female, Type.of.Travel=Personal Travel, Departure.Delay.in.Minutes=< 0.5, Class=Eco	→	Category=Detractor
000	0.181	5.512	1.000	0.000	Gender=Female, Type.of.Travel=Personal Travel, Departure.Delay.in.Minutes=< 0.5, Arrival.Delay.in.Minutes=< 0.5	→	Category=Detractor
000	0.179	5.578	1.000	0.000	Gender=Female, Type.of.Travel=Personal Travel, Arrival.Delay.in.Minutes=< 0.5, Class=Eco	→	Category=Detractor
000	0.177	5.646	1.000	0.000	Type.of.Travel=Personal Travel, Total.Freq.Flyer.Accts=< 0.5, Departure.Delay.in.Minutes=< 0.5, Arrival.Delay.in.Minutes=< 0.5	→	Category=Detractor
000	0.177	5.646	1.000	0.000	Type.of.Travel=Personal Travel, Total.Freq.Flyer.Accts=< 0.5, Arrival.Delay.in.Minutes=< 0.5, Class=Eco	→	Category=Detractor
000	0.173	5.787	1.000	0.000	Age=≥ 64.5, Type.of.Travel=Personal Travel, Total.Freq.Flyer.Accts=< 0.5, Class=Eco	→	Category=Detractor

2) Passive

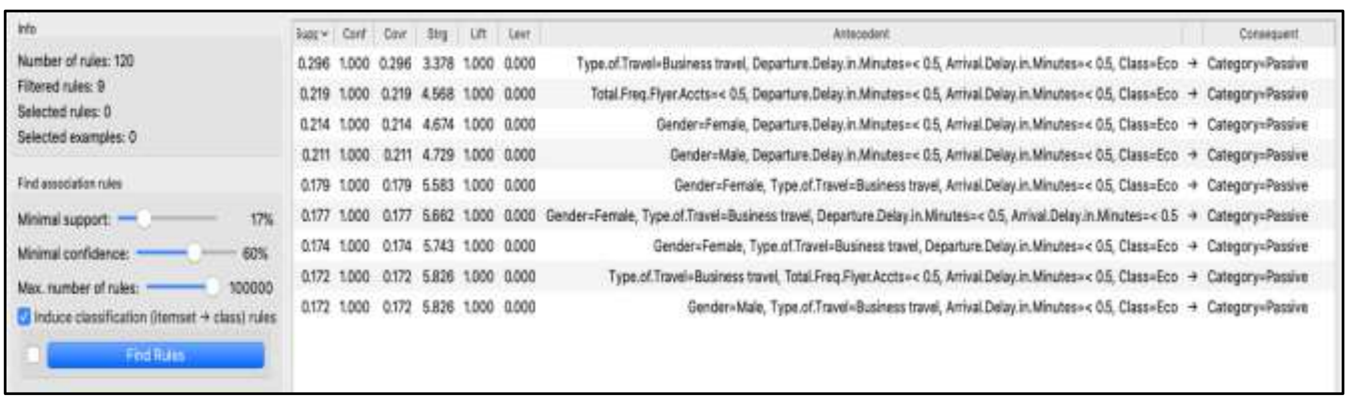
What makes a FlyFast passenger passive?

- Females traveling for Business purposes
- Flying in "Economy" class
- Delay in flight times is negligible

From the rules below, we can confirm our earlier hypothesis that the minor flight delays, if any, and "Economy" class are not exclusive among the detractors.

Here, we observe **Female passengers again, but those on Business travel.** So, females on business travel seem to be having a better experience than those on personal.

We also see a significant portion of males, regardless of the type of travel among passive customers.



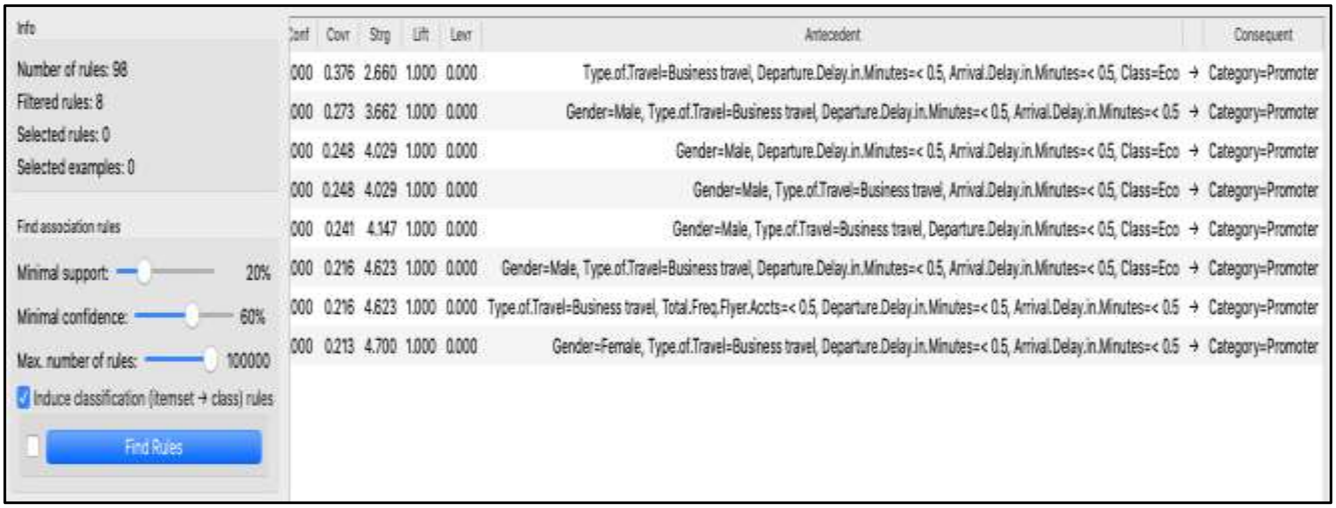
Supp	Conf	Covr	Strg	Lift	Levr	Antecedent		Consequent
0.296	1.000	0.296	3.378	1.000	0.000	Type.of.Travel=Business travel, Departure.Delay.in.Minutes=< 0.5, Arrival.Delay.in.Minutes=< 0.5, Class=Eco	→	Category=Passive
0.219	1.000	0.219	4.568	1.000	0.000	Total.Freq.Flyer.Accts=< 0.5, Departure.Delay.in.Minutes=< 0.5, Arrival.Delay.in.Minutes=< 0.5, Class=Eco	→	Category=Passive
0.214	1.000	0.214	4.674	1.000	0.000	Gender=Female, Departure.Delay.in.Minutes=< 0.5, Arrival.Delay.in.Minutes=< 0.5, Class=Eco	→	Category=Passive
0.211	1.000	0.211	4.729	1.000	0.000	Gender=Male, Departure.Delay.in.Minutes=< 0.5, Arrival.Delay.in.Minutes=< 0.5, Class=Eco	→	Category=Passive
0.179	1.000	0.179	5.583	1.000	0.000	Gender=Female, Type.of.Travel=Business travel, Arrival.Delay.in.Minutes=< 0.5, Class=Eco	→	Category=Passive
0.177	1.000	0.177	5.662	1.000	0.000	Gender=Female, Type.of.Travel=Business travel, Departure.Delay.in.Minutes=< 0.5, Arrival.Delay.in.Minutes=< 0.5	→	Category=Passive
0.174	1.000	0.174	5.743	1.000	0.000	Gender=Female, Type.of.Travel=Business travel, Departure.Delay.in.Minutes=< 0.5, Class=Eco	→	Category=Passive
0.172	1.000	0.172	5.826	1.000	0.000	Type.of.Travel=Business travel, Total.Freq.Flyer.Accts=< 0.5, Arrival.Delay.in.Minutes=< 0.5, Class=Eco	→	Category=Passive
0.172	1.000	0.172	5.826	1.000	0.000	Gender=Male, Type.of.Travel=Business travel, Arrival.Delay.in.Minutes=< 0.5, Class=Eco	→	Category=Passive

3) **Promoter**

What makes a FlyFast passenger a promoter?

- Travelers, BOTH male, and female on business
- “Economy” class
- Flight delays are negligible

Here, we see an abundance of **both male and female passengers on Business travel**. Thus, we can say that in general, those on business travel have a better flying experience than those on personal travel.



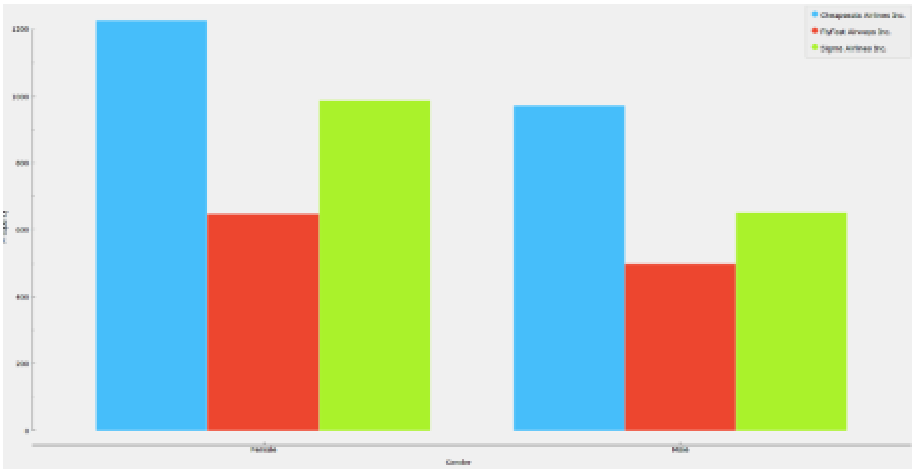
Understanding rival airlines Cheapseats Airways and Sigma Airways:

- **Cheapseats Airlines Inc.**

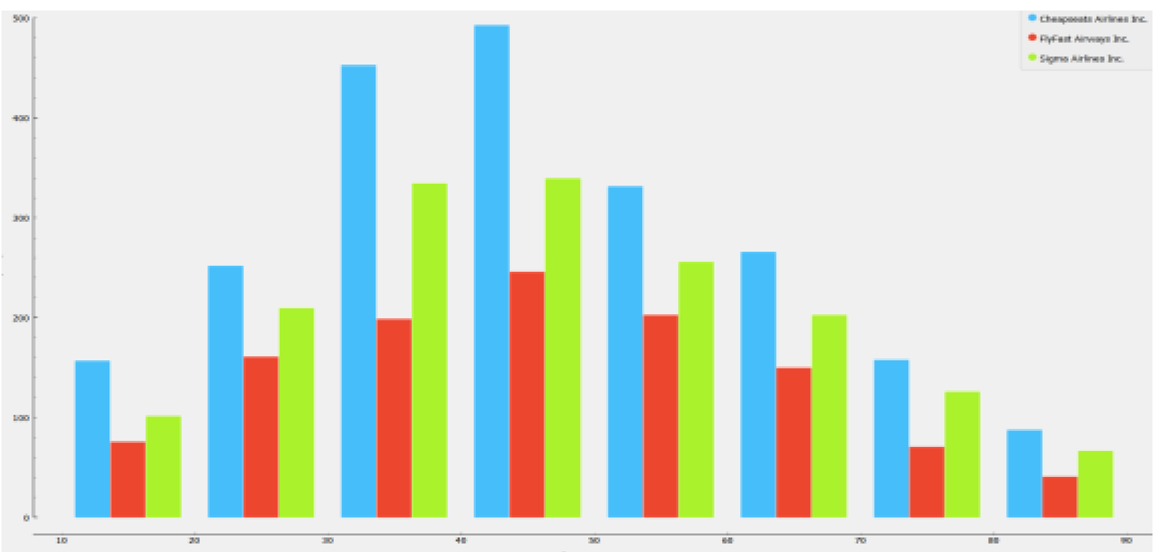
NPS Category:

1) **Promoter:** What makes a Cheapseats Airlines Customer Promoter?

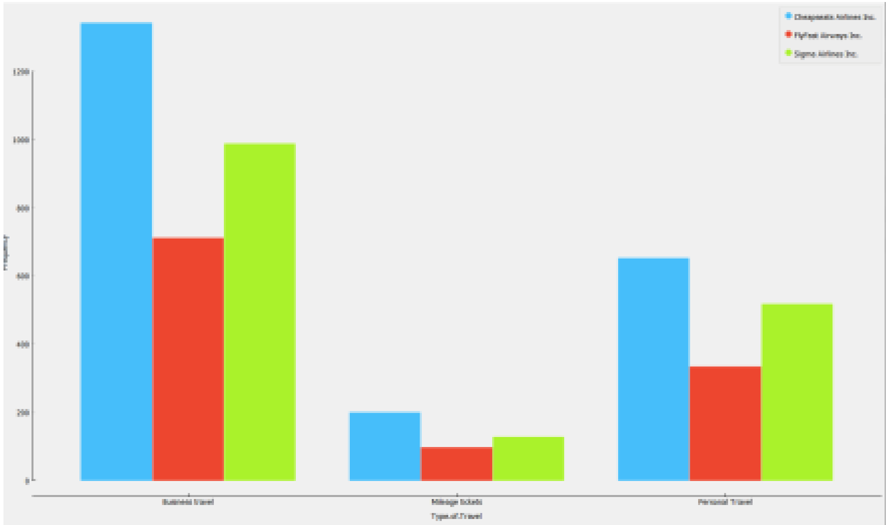
We observed that the airlines are more popular between the Men as they are the promoters of its airlines. Highlighted in Blue is Cheapseats followed by FlyFast in Red and Sigma in Green.



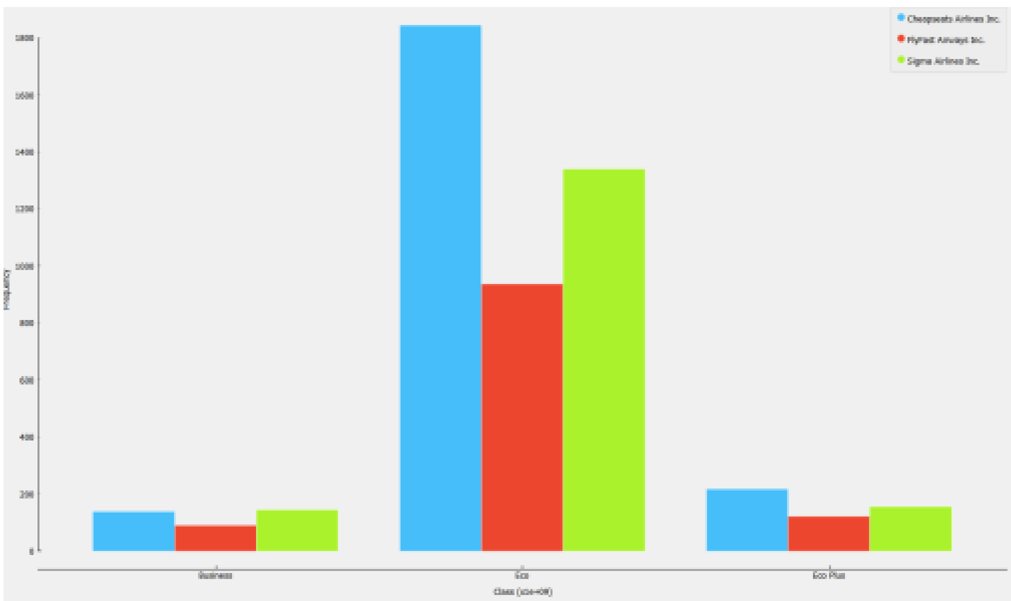
Another insight of promoters that we gained is the customers traveling in cheap seats are more price-sensitive and are people between the age of 29-43. Through the below graph, we can observe that most numbers of customers traveling fall in this category and they are Promoters.



As the number of Business Travelers is most in cheap seats and as per our analysis they tend to fall under the promoters. So, the airline was successful in retaining its business travelers.



The economy class is also popular among the promoters. And the greatest number of customers traveling in the airlines fly under the economy class.



So as per the analysis, we can conclude that by having the highest customer rate the airline was successful in retaining most flyers as promoters. The reason might be that the people falling under the age group 29-43 are mostly married and have a family and this category prefers economy class as they are more sensitive to price, turning more people to fly with the airlines. Business travelers also prefer airlines because of the cheap rates of economy class and better quality provided.

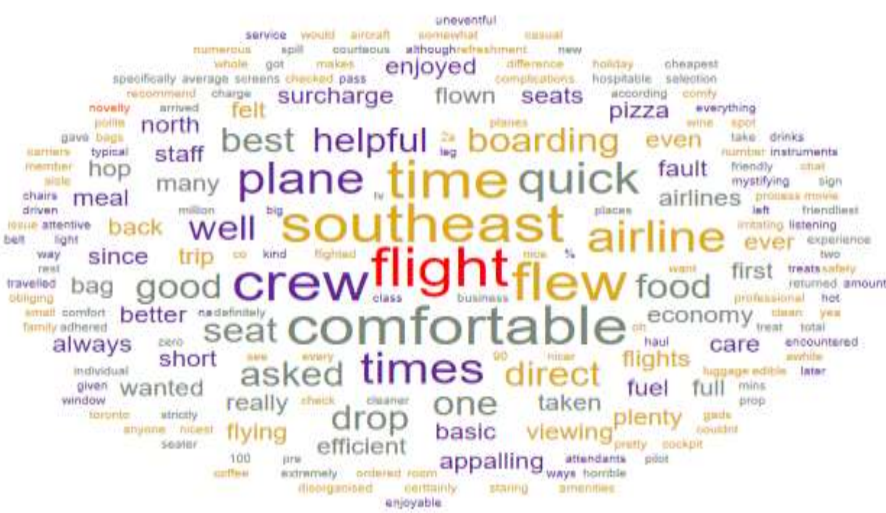
- Age group between 29-43
- Gender is Male
- Type of travel is Business
- Type of class is Eco

Supp	Conf	Covr	Strg	Lift	Levr	Antecedent		Consequent
0.119	0.631	0.189	2.007	1.667	0.048	Gender=Male, Type.of.Travel=Business travel, Price.Sensitivity=1 - 2, Departure.Delay.in.Minutes=< 60	→	Category=Promoter
0.119	0.631	0.189	2.007	1.667	0.048	Gender=Male, Type.of.Travel=Business travel, Price.Sensitivity=1 - 2, Departure.Delay.in.Minutes=< 60, Arrival.Delay.in.Minutes=< 84	→	Category=Promoter
0.103	0.631	0.163	2.327	1.667	0.041	Gender=Male, Class=Eco, Type.of.Travel=Business travel, Price.Sensitivity=1 - 2, Departure.Delay.in.Minutes=< 60	→	Category=Promoter
0.103	0.631	0.163	2.327	1.667	0.041	Gender=Male, Class=Eco, Type.of.Travel=Business travel, Price.Sensitivity=1 - 2, Departure.Delay.in.Minutes=< 60, Arrival.Delay.in.Minutes=< 84	→	Category=Promoter
0.127	0.628	0.203	1.868	1.657	0.051	Gender=Male, Type.of.Travel=Business travel, Price.Sensitivity=1 - 2	→	Category=Promoter
0.110	0.627	0.176	2.158	1.655	0.044	Gender=Male, Class=Eco, Type.of.Travel=Business travel, Price.Sensitivity=1 - 2	→	Category=Promoter
0.107	0.623	0.171	2.210	1.646	0.042	Gender=Male, Class=Eco, Type.of.Travel=Business travel, Price.Sensitivity=1 - 2, Arrival.Delay.in.Minutes=< 84	→	Category=Promoter
0.123	0.623	0.198	1.915	1.645	0.048	Gender=Male, Type.of.Travel=Business travel, Price.Sensitivity=1 - 2, Arrival.Delay.in.Minutes=< 84	→	Category=Promoter
0.104	0.619	0.168	2.251	1.634	0.040	Gender=Male, Type.of.Travel=Business travel, Flights.Per.Year=< 18	→	Category=Promoter
0.101	0.616	0.163	2.320	1.625	0.039	Gender=Male, Type.of.Travel=Business travel, Flights.Per.Year=< 18, Arrival.Delay.in.Minutes=< 84	→	Category=Promoter
0.121	0.613	0.196	1.928	1.619	0.046	Type.of.Travel=Business travel, Age=29 - 43, Departure.Delay.in.Minutes=< 60, Arrival.Delay.in.Minutes=< 84	→	Category=Promoter
0.121	0.612	0.197	1.924	1.616	0.046	Type.of.Travel=Business travel, Age=29 - 43, Departure.Delay.in.Minutes=< 60	→	Category=Promoter
0.101	0.609	0.165	2.295	1.607	0.038	Class=Eco, Type.of.Travel=Business travel, Age=29 - 43, Departure.Delay.in.Minutes=< 60, Arrival.Delay.in.Minutes=< 84	→	Category=Promoter
0.101	0.607	0.166	2.280	1.603	0.038	Class=Eco, Type.of.Travel=Business travel, Age=29 - 43, Departure.Delay.in.Minutes=< 60	→	Category=Promoter
0.125	0.606	0.206	1.835	1.599	0.047	Type.of.Travel=Business travel, Age=29 - 43, Arrival.Delay.in.Minutes=< 84	→	Category=Promoter
0.116	0.606	0.191	1.979	1.599	0.043	Airline.Status=Silver, Arrival.Delay.in.Minutes=< 84	→	Category=Promoter
0.109	0.602	0.181	2.088	1.588	0.040	Airline.Status=Silver, Departure.Delay.in.Minutes=< 60	→	Category=Promoter
0.109	0.602	0.181	2.088	1.588	0.040	Airline.Status=Silver, Departure.Delay.in.Minutes=< 60, Arrival.Delay.in.Minutes=< 84	→	Category=Promoter
0.104	0.601	0.173	2.106	1.587	0.039	Class=Eco, Type.of.Travel=Business travel, Age=29 - 43, Arrival.Delay.in.Minutes=< 84	→	Category=Promoter

Promoter Word Cloud:

Below are some pointers to be considered why Cheapseats has promoters. These pointers can also be considered by FlyFast Airways for better services.

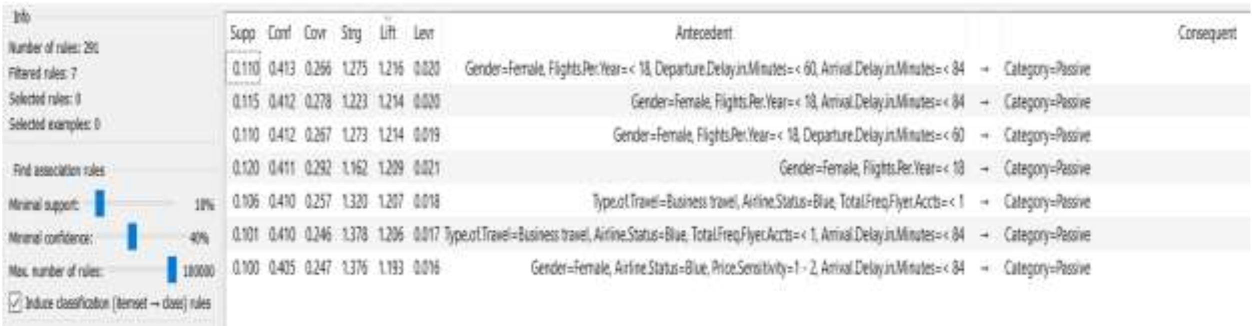
- Cheapseats crew and quick services
- Food provided by Cheapseats
- Comfortable seats and easy boarding



2) **Passive:** What makes a Cheapseats Airlines customer Passive?

The female flyers with less than 20 flights per year with blue airline status and are sensitive to price are the ones who tend to fall under the passive category.

- Female flyers
- Airline Status “Blue” customers
- Less than 20 flights per year
- Departure and Arrival Delay close to 3 hours
- Sensitive to price



Supp	Conf	Covr	Strg	Lift	Levr	Antecedent		Consequent
0.110	0.413	0.266	1.275	1.216	0.020	Gender=Female, Flights.Per.Year=< 18, Departure.Delay.in.Minutes=< 60, Arrival.Delay.in.Minutes=< 84	→	Category=Passive
0.115	0.412	0.278	1.223	1.214	0.020	Gender=Female, Flights.Per.Year=< 18, Arrival.Delay.in.Minutes=< 84	→	Category=Passive
0.110	0.412	0.267	1.273	1.214	0.019	Gender=Female, Flights.Per.Year=< 18, Departure.Delay.in.Minutes=< 60	→	Category=Passive
0.120	0.411	0.292	1.162	1.209	0.021	Gender=Female, Flights.Per.Year=< 18	→	Category=Passive
0.106	0.410	0.257	1.320	1.207	0.018	Type.of.Travel=Business travel, Airline.Status=Blue, Total.Freq.Flyer.Accts=< 1	→	Category=Passive
0.101	0.410	0.246	1.378	1.206	0.017	Type.of.Travel=Business travel, Airline.Status=Blue, Total.Freq.Flyer.Accts=< 1, Arrival.Delay.in.Minutes=< 84	→	Category=Passive
0.100	0.405	0.247	1.376	1.193	0.016	Gender=Female, Airline.Status=Blue, Price.Sensitivity=1 - 2, Arrival.Delay.in.Minutes=< 84	→	Category=Passive

3) **Detractor:** What makes a Cheapseats Airlines customer a Detractor?

In cheap seats, the customers who are personal types of travelers, those who have less took a flight less than once, have the airline status as blue and there is a delay in arrival are mostly the detractors of cheap seat airlines.

- Airline Status “Blue” customers
- **Personal Travel** purpose customers
- Departure and Arrival Delay close to 1 hour
- Personal type of travelers
- Took less than one flight

Supp	Conf	Covr	Strg	Lift	Levr	Antecedent		Consequent
0.145	0.620	0.233	1.207	2.202	0.079	Class=Eco, Type.of.Travel=Personal Travel, Departure.Delay.in.Minutes=< 60, Arrival.Delay.in.Minutes=< 84	→	Category=Detractor
0.137	0.637	0.216	1.306	2.263	0.077	Type.of.Travel=Personal Travel, Total.Freq.Flyer.Accts=< 1, Departure.Delay.in.Minutes=< 60, Arrival.Delay.in.Minutes=< 84	→	Category=Detractor
0.151	0.710	0.212	1.328	2.523	0.091	Type.of.Travel=Personal Travel, Airline.Status=Blue, Departure.Delay.in.Minutes=< 60, Arrival.Delay.in.Minutes=< 84	→	Category=Detractor
0.130	0.730	0.178	1.579	2.592	0.080	Type.of.Travel=Personal Travel, Airline.Status=Blue, Total.Freq.Flyer.Accts=< 1, Arrival.Delay.in.Minutes=< 84	→	Category=Detractor
0.138	0.733	0.189	1.492	2.602	0.085	Class=Eco, Type.of.Travel=Personal Travel, Airline.Status=Blue, Arrival.Delay.in.Minutes=< 84	→	Category=Detractor

Detractor Persona Word Cloud for Cheapseats:

Here, FlyFast Airways can consider the pointers to avoid repeating mistakes carried out by Cheapseats which resulted in their customers becoming Detractors.

- Economy Class customers are unhappy
- Flights seats are probably not comfortable in economy class
- Time is a parameter to be considered

- **Sigma Airlines Inc.**

NPS Category:

1) **Promoter:** What makes a Sigma Airlines Customer promoter?

- C1 Cluster:

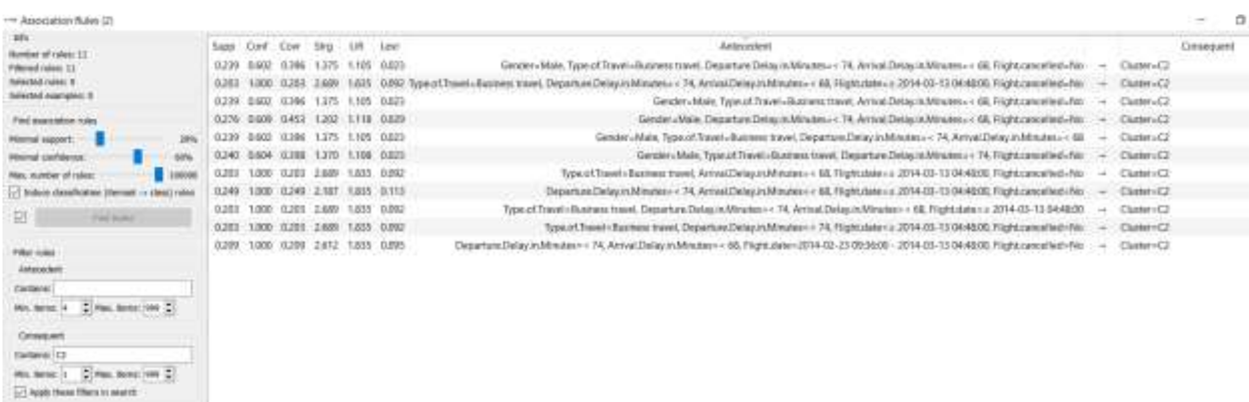
It is recommended that a customer traveling for business purposes should be targeted for assigning frequent flyer account and such customers should be given hotel benefits or promo offers/coupons for keeping the customer happy.

- Business travel purpose customers
- Delay in Departure and Arrival less than 1 hour
- No flight cancellation

- C2 Cluster:

It is recommended that Male customers with Business purposes should be given offers at the partnered airport restaurants for Food & Drinks expiring few hours before and after the flight so that customer is engaged and might have a good and enjoyable time during the flight and after the flight as well.

- Male Business purpose travelers
- Delay in Departure and Arrival less than 1 hour
- No flight cancellation



2) **Passive:** What makes a Sigma Airlines Customer Passive

- Cluster C1:
  It is recommended that customers traveling with Airline Status Blue be upgraded for better services if they have frequent flyer account and the delay time must be reduced at least close to 1 hour which might make them a promoter.
    - Airline Status “Blue” customers
    - Frequent Flyer Accounts <= 1
    - Departure and Arrival Delay close to 3 hours
    - Flights not canceled

- Cluster C2:

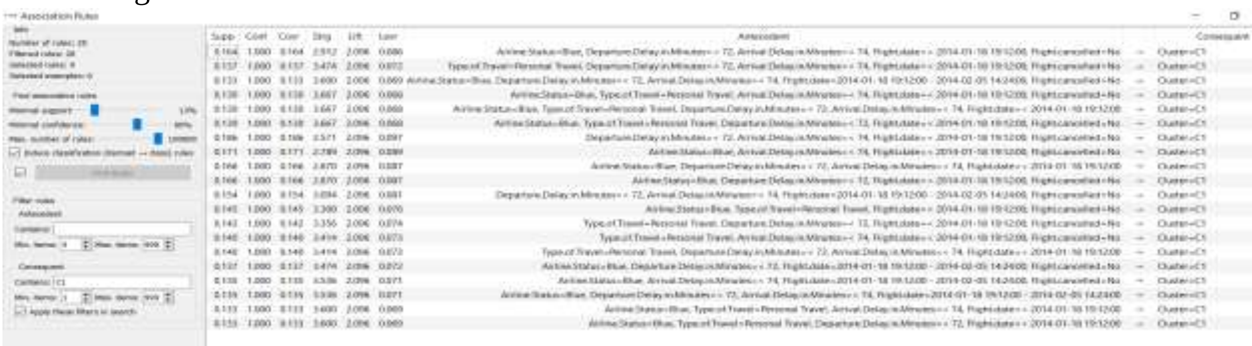
  It is recommended that customers traveling with Airline Status Blue be upgraded for better services if they have frequent flyer account and the delay time must be reduced at least close to 1 hour which might make them a promoter.
  Also, customers sensitive to price must be given better offers through partnered payment modes and should be given airport shopping coupons so that customer is engaged and spends in partnered shops for compensating the price sensitivity solution.
    - Airline Status “Blue” customers
    - Frequent Flyer Accounts <= 1
    - Departure and Arrival Delay close to 3 hours
    - Flights not canceled
    - Price Sensitivity between 1 to 2

3) **Detractor:** What makes a Sigma Airlines Customer Detractor

- Cluster C1:

It is recommended that as the customer is traveling for personal reasons, they should be offered shopping coupons at Airport partnered shops so that they can buy something if required for their family members. Blue status customers should be upgraded to better Airline status if they are possible frequent flyers.

- Airline Status "Blue" customers
- **Personal Travel** purpose customers
- Departure and Arrival Delay close to 1 hour
- Flights have not canceled still a Detractor



Cluster C2:

It is recommended that as the customer is traveling for personal reasons, they should be offered shopping coupons at Airport partnered shops so that they can buy something if required for their family members. Female customers sensitive to price and who may have longer layover time can be offered airport salon coupons and better-discounted ticket prices through partnered payment modes.

- Airline Status "Blue" customers
- Female customers sensitive to price
- **Personal Travel** purpose customers
- Departure and Arrival Delay close to 1 hour
- Flights have not canceled still a Detractor
- Price Sensitivity between 1 to 2

Phase7: Marketing Plan for FlyFast Airways Inc.

Detractors NPS Category: Overall, our Detractors have a majority of Females, on personal travel.

- People on personal travel are always likely to recommend our service if they have a good experience and should not be sidelined. We recommend introducing travel discounts, limited time offers (maybe in collaboration with tourism agencies, hotels) for this segment.
- On some further examination, we found that among these females, on personal travel, most of them had age>=64.5 years. Reserving preferred seats, providing baggage assistance during boarding to ensure a more comfortable flight would be a good option too.
- We should be sensitive to the specific needs of females to ensure they have a satisfactory experience. Introducing a separate lactation room, not accessible to the general public is an option.

Passive NPS Category: Among Passives, we see a large portion of Females again, but those traveling for business purposes.

- Knowing that we have a sizable portion of the same segment among Promoters too, it is plausible that with a few perks, we can nudge them to become a Promoter as well.
- Business travelers usually already have preferred boarding and seats. So we can enhance their experience in other ways - qualitative food and options for all types of diets.
- Improving onboard entertainment for longer flights.
- Encouraging them to get "Frequent Flyer" accounts, by promoting the benefits of acquiring "sky miles" and redeeming them for upgrades, dining or shopping coupons, etc.

Promoters NPS Category: Among promoters, we have a blend of male and female passengers with one common attribute - Business Travel.

- We recognize that this subset represents some of our happiest customers, and we should maintain our current standard of service to retain them.
- They too should be encouraged to enroll in and enjoy the benefits of "Frequent Flyer" programs.

- Airline-owned or partner lounges and bars at airports should be accessible, and occasional discounts/schemes if possible, can be offered.