

PROJECT REPORT

Flying Hospitality Improvement

Submitted towards the partial fulfillment of the criteria for award of Post Graduate Program In Analytics And Artificial Intelligence by Imarticus

Submitted By:

Goutham M

Course and Batch: PGAA-01 Jan-2021



Abstract

Customers are not satisfying good due to few uncomfortable situations in Flight. Major issue arises due to the usage of seat recline. As per the survey dataset we are trying to segment customers based on their past experience. In this project we are going to recommend a suitable seat for the customers and make them feel satisfied.

Acknowledgements

We are using this opportunity to express our gratitude to everyone who supported us throughout the course of this project. We are thankful for their aspiring guidance, invaluable constructive criticism and friendly advice during the project work. We are sincerely grateful to them for sharing their truthful and illuminating views on a number of issues related to the project.

Further, we were fortunate to have great teachers who readily shared their immense knowledge in data analytics and guided us in a manner that the outcome resulted in enhancing our data skills.

We wish to thank, all the faculties, as this project utilized knowledge gained from every course that formed the PGAA-01 program.

We certify that the work done by us for conceptualizing and completing this project is original and authentic.

Date: 12 Sep 2021

Place: Chennai

Goutham M

Certificate of Completion

I hereby certify that the project titled “**Flying Hospitality Improvement**” was undertaken and completed under my supervision by Goutham M from the batch of PGAA-01 (Jan 2021)

Date: 12 Sep 2021

Place: Chennai

Table of Contents

Abstract.....	2
Acknowledgements.....	3
Certificate of Completion	4
INTRODUCTION.....	6
Title & Objective of the study.....	6
Need of the Study.....	6
Business Model of Enterprise.....	6
Data Sources.....	6
Tools & Techniques.....	6
DATA UNDERSTANDING AND PREPARATION.....	7
Missing Value Analysis and Treatment.....	7
Encoding.....	10
Distance Metrics.....	10
SELECTING THE OPTIMUM CLUSTER.....	11
Dendrogram.....	11
Elbow Method.....	11
K-MODE CLUSTER SEGMENTATION.....	12
ANALYSIS.....	12
Data insert.....	12
Customer Segmentation analysis.....	12
CONCLUSION.....	14

INTRODUCTION

Title & Objective of the study

'Flying Hospitality Improvement' is the project I am working upon which falls under Tourism & Travel. The CSV files contain Flight survey data. The primary purpose of working on this project is to segment the customer by using the past survey data.

Need of the Study

In this study we are going to recommend a suitable seat for the customers and make them feel satisfied.

Business Model of Enterprise

Selecting the relevant variables from the dataset and arranging their values in order of importance to create a model to segment the customers by performing different types of algorithms on the data.

Data Sources

Flying Etiquette Survey Data contains the information about the satisfaction on their recent flight journey. The dataset contains the information like age, gender, Household Income, Education, Location (Census Region) and Survey questions of the customers.

Data Set Description:
Contains 1040 rows and 27 columns

Tools & Techniques

Tools: Jupyter Notebook.
Techniques: KModes.

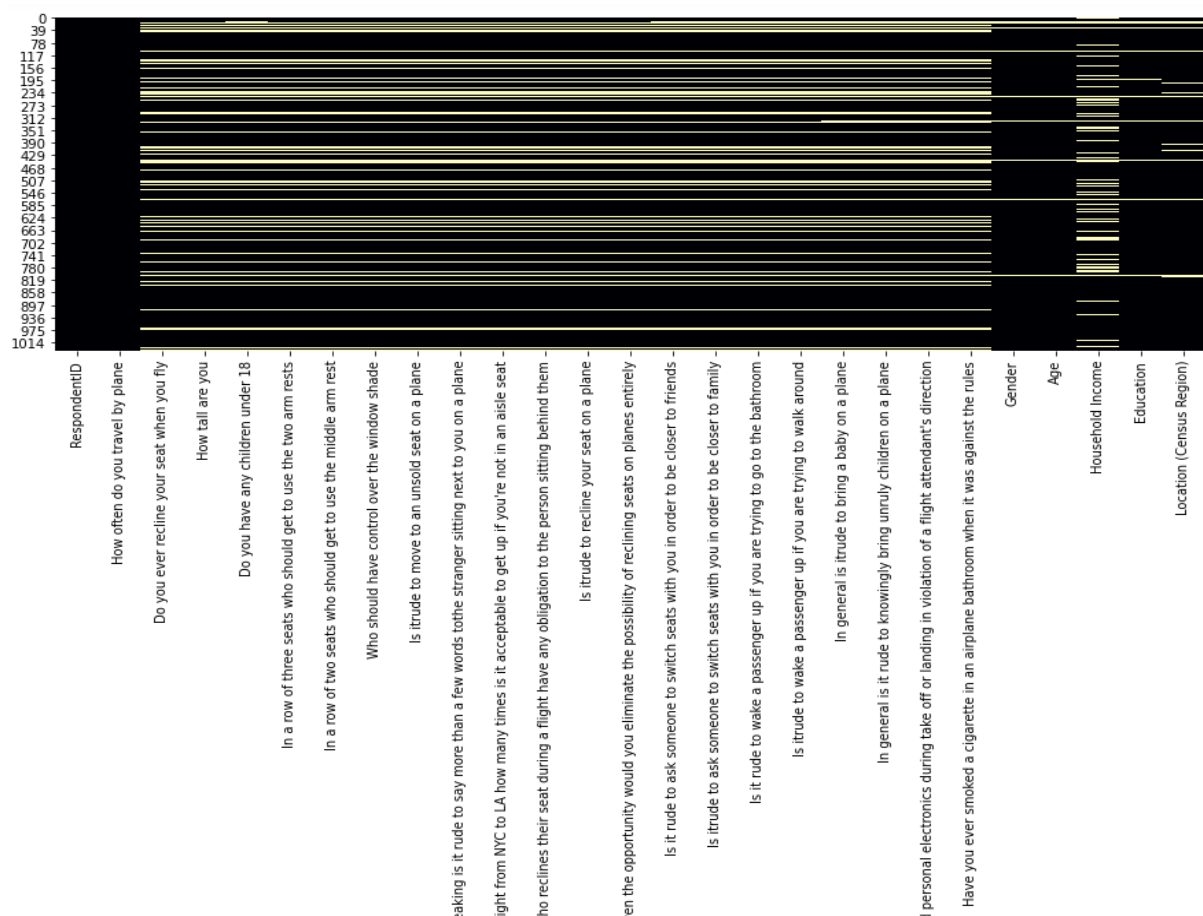
DATA UNDERSTANDING AND PREPARATION

One of the first steps we engaged in was to outline the sequence of steps that we will be following for our project. Each of these steps are elaborated below

After importing the required libraries, a sequence of steps was followed to perform data analysis.

Missing Value Analysis and Treatment

After printing the shape of the data, we gain that the dataset consists of 1040 observations and 27 variables. The initial step was to check the missing values in each variable for a better view, plot a heatmap of the dataset for visualizing the missing values as shown below:



It is evident from the above heatmap that our dataset contains a lot of missing values and we cannot use feature that has so many missing values.

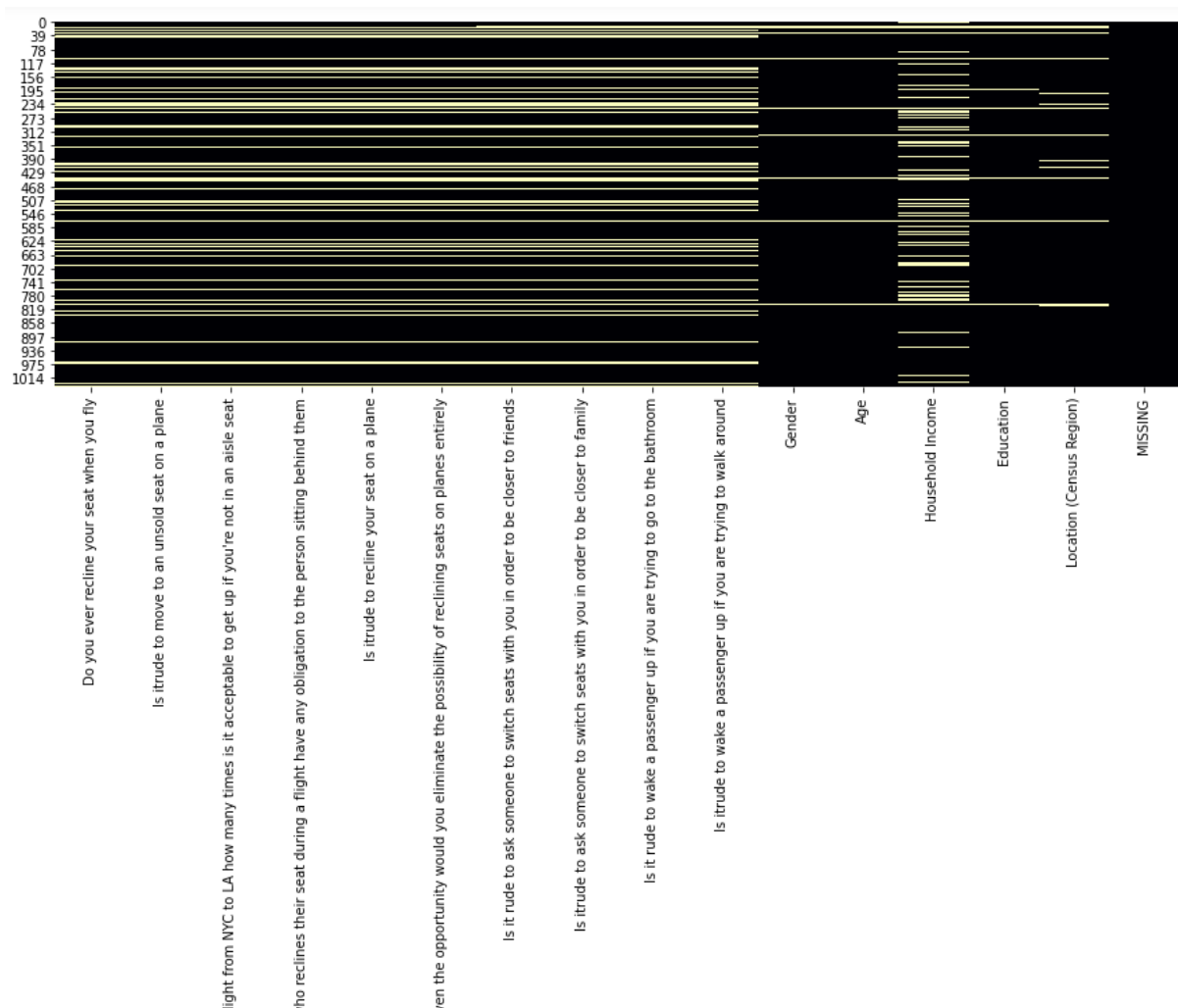
Above heatmap shows the intensity of values that are missing in every column. All the light-coloured columns represent the number of missing values present in that specific column.

Firstly, we are dropping the following because they are irrelevant to our problem statement:

- RespondentID - they all are unique numbers.
- Do you have any children under 18
- In a row of three seats who should get to use the two arm rests

- In a row of two seats who should get to use the middle arm rest
- Who should have control over the window shade
- Generally speaking, is it rude to say more than a few words to the stranger sitting next to you on a plane
- In general, is it rude to bring a baby on a plane
- In general, is it rude to knowingly bring unruly children on a plane
- Have you ever used personal electronics during take-off or landing in violation of a flight attendants direction
- Have you ever smoked a cigarette in an airplane bathroom when it was against the rules

We are then left with **17 variables**. Now visualizing the missing values in each column after dropping the variables, we get the following heatmap:



By comparing the above two heatmaps, it is clearly seen that the number of missing values has been reduced.

The remaining missing values present are treated by using **Mode** and a newly added column **"Missing"**.

In **Missing** column, we calculated the count of missing values for each record. In which ever record we find value is more than 0 are dropped.

```
1 df['MISSING'] = source[[
2     "Do you ever recline your seat when you fly",
3     "Is it rude to move to an unsold seat on a plane",
4     "On a 6 hour flight from NYC to LA how many times is it acceptable to get up if you're not in an aisle seat",
5     "Under normal circumstances does a person who reclines their seat during a flight have any obligation to the person sitt:
6     "Is it rude to recline your seat on a plane",
7     "Given the opportunity would you eliminate the possibility of reclining seats on planes entirely",
8     "Is it rude to ask someone to switch seats with you in order to be closer to friends",
9     "Is it rude to ask someone to switch seats with you in order to be closer to family",
10    "Is it rude to wake a passenger up if you are trying to go to the bathroom",
11    "Is it rude to wake a passenger up if you are trying to walk around"]].apply(lambda x: x.isnull().sum(), axis='columns')
12
```

The missing values of the following variables are treated with **Missing** Column:

- Do you ever recline your seat when you fly
- Is it rude to move to an unsold seat on a plane
- On a 6-hour flight from NYC to LA how many times is it acceptable to get up if you're not in an aisle seat
- Under normal circumstances does a person who reclines their seat during a flight have any obligation to the person sitting behind them
- Is it rude to recline your seat on a plane
- Given the opportunity would you eliminate the possibility of reclining seats on planes entirely
- Is it rude to ask someone to switch seats with you in order to be closer to friends
- Is it rude to ask someone to switch seats with you in order to be closer to family
- Is it rude to wake a passenger up if you are trying to go to the bathroom
- Is it rude to wake a passenger up if you are trying to walk around

While the missing values of the following variables are treated with **Mode**:

- Gender
- Age
- Household Income
- Education
- Location (Census Region)

```
1 df["Gender"] = df["Gender"].fillna(df["Gender"].mode()[0])
2 df["Age"] = df["Age"].fillna(df["Age"].mode()[0])
3 df["Household Income"] = df["Household Income"].fillna(df["Household Income"].mode()[0])
4 df["Education"] = df["Education"].fillna(df["Education"].mode()[0])
5 df["Location (Census Region)"] = df["Location (Census Region)"].fillna(df["Location (Census Region)"].mode()[0])
```

After the complete treatment of the missing values, it is evident from the below heatmap that the dataset is now clean and ready.



Encoding

To convert the Categorical data to numerical data, we are doing One Hot using Pandas – `get_dummies()` method.

```
1 cluster_data = pd.get_dummies(cluster_data)
2 cluster_data.columns
```

Distance Metrics

Distance Metrics to know the input data pattern in order to make any Data Based decision. Distance metric helps in improving the performance of Clustering. A distance function provides distance between the elements of a set.

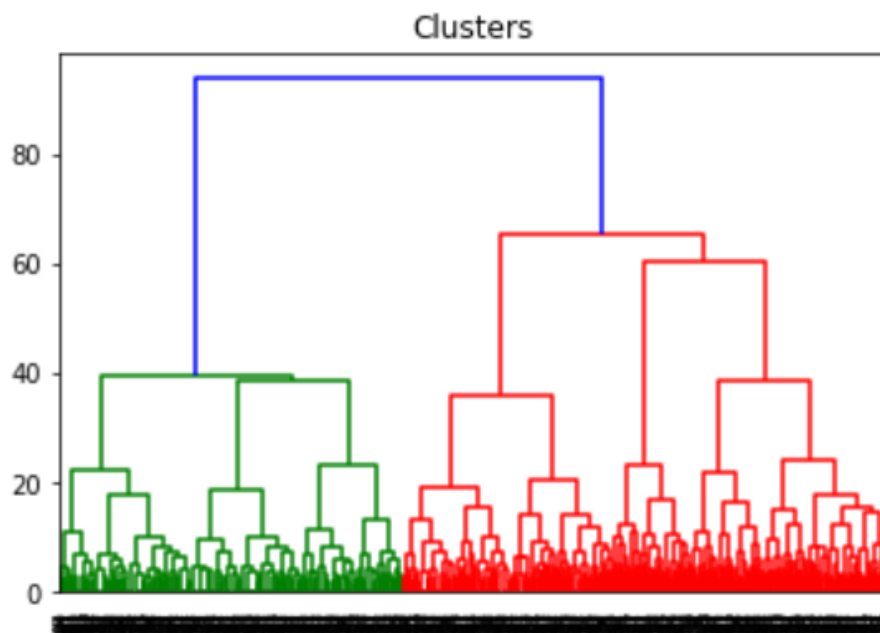
```
1 from sklearn.neighbors import DistanceMetric
2 cluster_data = DistanceMetric.get_metric('dice').pairwise(cluster_data)
3 cluster_data
```

SELECTING THE OPTIMUM CLUSTER

We are going to use the data after calculating the distance metric to find the optimum cluster using Elbow method and dendrogram.

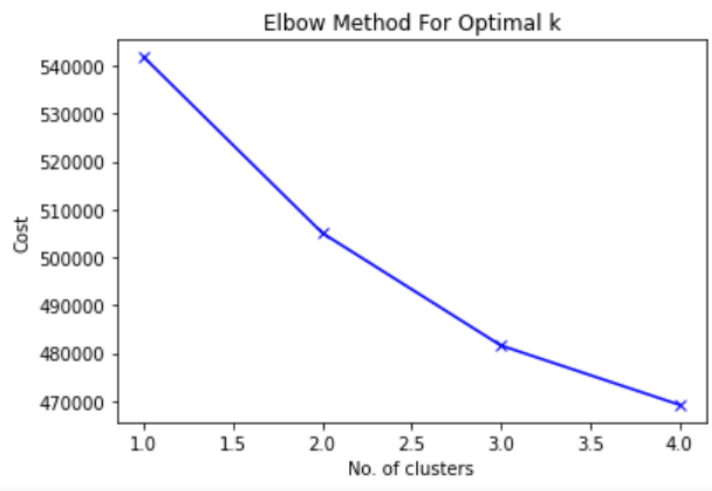
Dendrogram

Using linkage function we are finding the distance between two clusters, so here we have passed the matrix of features, and method "ward," the popular method of linkage in hierarchical clustering. The remaining lines of code are to describe the labels for the dendrogram plot. As per the below dendrogram we can easily see the clusters formed.



Elbow Method

In cluster analysis, the elbow method is used in determining the number of clusters in a data set. The method consists of plotting the explained variation as a function of the number of clusters, and picking the elbow of the curve as the number of clusters to use.



Based on Dendrogram & Elbow Method came to a conclusion that the optimum cluster is 3.

K-MODE CLUSTER SEGMENTATION

Using K-Mode we are going to segment the customers.

```
1 # Building the model with 3 clusters
2 kmode = KModes(n_clusters=3, init = "random", n_init = 5, verbose=1)
3 clusters = kmode.fit_predict(cluster_data)
4 clusters
```

ANALYSIS

Data insert

Inserting the clustering result with the df (required data fields for problem solving), to understand the customer segmentation.

```
1 df.insert(0, "Cluster", clusters, True)
2 df
```

Customer Segmentation analysis

Understand the customer Segmentation based on Crosstab analysis.

- Trying to under the cluster segment based on Gender in the below image. We can see that majority of the customers falls under **Cluster1**

```
1 pd.crosstab(df.Cluster, df.Gender)
```

Cluster	Gender	
	Female	Male
0	150	132
1	189	168
2	110	101

- As per our problem statement we need to find the clusters based on recline usage.

```
1 out=pd.crosstab(df.Cluster, df["Do you ever recline your seat when you fly"],normalize='index')
2 round(out*100,0)
```

Cluster	Do you ever recline your seat when you fly					
	About half the time	Always	Never	Once in a while	Usually	
0	15.0	30.0	7.0	17.0	30.0	
1	12.0	6.0	36.0	33.0	13.0	
2	14.0	12.0	9.0	43.0	21.0	

In the above image we evidently figure out that our customers are segmented properly

- Customers belong to Cluster 0 are **using the recline Very often**
 - Customers belong to Cluster 1 are **not using the recline**
 - Customers belong to Cluster 2 are **using the recline Once in a While**
- Trying to understand more on the clusters formed, will move to the next survey analysis.

```

1 out=pd.crosstab(df.Cluster, df["Under normal circumstances does a person who reclines their seat during a flight have any obligation to the person sitting behind them"],normalize='index')
2 round(out*100,0)

```

	Under normal circumstances does a person who reclines their seat during a flight have any obligation to the person sitting behind them	No, the person on the flight has no obligation to the person behind them	Yes, they should not recline their chair if the person behind them asks them not to
Cluster			
0	71.0	29.0	
1	23.0	77.0	
2	13.0	87.0	

As per the previous and above table we are getting an understanding that:

1. Customers belong to Cluster 0 are **using the recline very often and they are not compromising on their convenience**
 2. Customers belong to Cluster 1 are **not using the recline and they are accepting the fellow member's feedback**
 3. Customers belong to Cluster 2 are **using the recline Once in a While and they are accepting the fellow member's feedback**
- Will move a little deeper in to the next analysis

```

1 out=pd.crosstab(df.Cluster, df["Is itrude to recline your seat on a plane"],normalize='index')
2 round(out*100,0)

```

	Is itrude to recline your seat on a plane	No, not rude at all	Yes, somewhat rude	Yes, very rude
Cluster				
0	92.0	7.0	1.0	
1	23.0	59.0	18.0	
2	74.0	24.0	1.0	

As per the above tables we are getting an idea about the segmentation formed.

- a. Customers belong to **Cluster 0**
 - i. Using the recline very often
 - ii. They are not compromising on their convenience
 - iii. They are saying it's not rude at all to recline your seats on a plane
- b. Customers belong to **Cluster 1**
 - i. Not using the recline
 - ii. They are accepting the fellow member's feedback
 - iii. They are saying its somewhat rude to recline your seats on a plane
- c. Customers belong to **Cluster 2**
 - i. Using the recline Once in a While
 - ii. They are accepting the fellow member's feedback
 - iii. They are saying it's not rude at all to recline your seats on a plane

CONCLUSION

We don't need to focus much on Customers belong to Cluster 1 & 2 because they are using the recline Once in a While or not using and they are also adjusting with the fellow member's request.

But the Customers belong to Cluster 0 are Using the recline very often also they are not compromising on their convenience and they are saying it's not rude at all to recline your seats on a plane.

We need to focus on Customers who belong to Cluster 0 because it will create an unfriendliness and dissatisfaction between the Customers and Crew members. This dissatisfaction will lead to miss our customers.

Based on our model we are going to segment the customer and recommend a proper seat when they are booking the flight ticket. This will lead to Hospitality between the Customers and Crew members

FURTHER IMPROVEMENT AREAS

Now we got the dependent variable, based on the dependant variable we are going to create a supervised learning model. If an existing customer who had not taken the survey before and at the time of ticket booking, we can take a survey of their past experience. Based on the survey result our recommendation system will help them in selecting the best seat for their travel. Which will help us to increase customer Hospitality level and organization growth.

THANK YOU