# AI in Private Travel Agencies

ANAND K S

6th October, 2022.

# 1.0 Abstract:

Artificial Intelligence is a vast area of computer science that studies how to teach computers to think and act, like a human.

Machine Learning is a subset of AI, important, but not the only one. In a nutshell, Machine Learning is about building models that predict the result with the high accuracy on the basis of the input data. Using statistical methods, it enables machines to improve their accuracy as more data is fed in the system.

# 2.0 Problem statement:

Travel agents help clients schedule everything involved in planning a vacation, from booking flights and hotels to arranging for day trips and cruises. The communication and research travel agents perform for their clients can often be carried out in a home office just as easily an outside office.

A home-based business can seem like an easy way to make money compared to working a traditional job, but like any job, it requires hard work to be successful. Working from home can present distractions that can impede productivity and take the focus away from work. Self-motivation and the ability to stay focused are essential for home-based workers.

# 3.0 Market / Customer/ Business Need Assessment:

# The World of Anything You Want at Any Time.

Travelling is a highly emotional experience and has an endless choice of options that fit each budget and type of traveller. Solo travel, corporate trip or family vacation — travellers with different goals on their mind want the app to suggest the correct packages right from the start.

“When do airlines drop prices?”, “Are flights really cheaper on Tuesdays?” and a bunch of similar questions are typed into Google search field many times every day. Aren’t they the best evidence that people are eager to find the most optimal price on, sure thing, the option that appeared in their mind? It seems that everybody has had such a situation: you’ve just found a good flight deal, then returned back to the website in a while and whoa! , the ticket price ramped up (or if you’re lucky, fell down).

Flight price generation engine works according to certain rules and might take into account some of the following parameters:



Machine learning is known for finding hidden patterns that a human’s eye might not even spot. For example, ML algorithm can detect anomalies after a big fare increase, taking into account a combination of features that caused such an increase.

Travel providers can help travellers find the best time to book a hotel or to buy a cheap ticket by leveraging machine learning. When the deal becomes available, the app will likely send notification to user.

# 4.0 Target specifications and Characterization:

### 4.1) Sales optimisation

AI-powered models can empower sales managers on all the stages of sales funnel:

* Discovery phase helping customers find where to go
* Engagement phase by figuring out what option is best for each customer
* Conversion & Retention phases by working on crowdsourcing data

Travel history, previous behavioural patterns, purchase history and belonging to loyalty program — if such data is fed to AI based system, it can help sales teams generate personalized travel package.

### ****4.2) Price optimisation****

Dynamic pricing technology infused by AI can help pinpoint buying patterns so accurately that airlines can synchronize their pricing strategies in real-time and present the right price at the right time.

The data required for these targets can be gathered on numerous folds as shown in the

following figure.

# 

The above figure represents the various means that can be used as features to develop the required AI model through an appropriate algorithm which is discussed later in the document.

The different sources used to produce the data columns as features include Third party data, where we can gauge the consumer tastes and preferences through various options they choose in different scenarios which include the previous hotels they have booked with their ID, data from the government regarding their previous tourist advents among other things.

Data can also be collected through their social media interactions in the past, which describe their likes and dislikes.

# 5.0 External searches:

How machine learning and AI can improve Travel Services

This article is originally published on Django Stars blog.(Alexander Bulanov, Oct 3 2018)

A lot of the content in my document is based on the details provided in this article.

Data sets:

<https://www.kaggle.com/datasets/leomauro/argodatathon2019?select=flights.csv>

<https://archive.ics.uci.edu/ml/machine-learning-databases/00484/>

https://www.kaggle.com/datasets/djusdjus/hotel-rooms-prices-dataset?select=hotels.xlsx

The first of the two datasets provided above represent the flight pricing corresponding to a set of

influencing factors as features.

The numerous data fields or features in the flight fare pricing data set from kaggle are:

Travel code

User code

From (source station)

To (Destination airport)

Flight type

Time

Distance

Agency

Date, etc.

The distance determines the base price of the flight, on top of which the agency determines a fare or tax which depends on the quality of the air-lanes and the date or the period of booking also influences the fare. Festive periods for instance tend to have higher fares and so do weekends.

The second dataset is obtained from trip advisor.

This data set is populated by crawling TripAdvisor.com. Reviews on destinations in 10 categories mentioned across East Asia are considered. Each traveler rating is mapped as Excellent (4), Very Good (3), Average (2), Poor (1), and Terrible (0) and average rating is used against each category per user.

Attribute Information:

Attribute 1 : Unique user id  
Attribute 2 : Average user feedback on art galleries  
Attribute 3 : Average user feedback on dance clubs  
Attribute 4 : Average user feedback on juice bars  
Attribute 5 : Average user feedback on restaurants  
Attribute 6 : Average user feedback on museums  
Attribute 7 : Average user feedback on resorts  
Attribute 8 : Average user feedback on parks/picnic spots  
Attribute 9 : Average user feedback on beaches  
Attribute 10 : Average user feedback on theaters  
Attribute 11 : Average user feedback on religious institutions

Relevant Papers:

Renjith, Shini, A. Sreekumar, and M. Jathavedan. 2018. â€œEvaluation of Partitioning Clustering Algorithms for Processing Social Media Data in Tourism Domainâ€. In 2018 IEEE Recent Advances in Intelligent Computational Systems (RAICS), 127â€“31. IEEE.

Third data set regarding hotel prices based on the location, rating and season was also collected from kaggle and the details are as follows:

This dataset contains 120 observations and 9 variables.

The variables are:

* ID
* Hotel name
* Price(BAM): Price for one night In currency of Bosnia and Herzegovina
* Hotel star rating: Number of stars (1-5)
* Distance: The distance of hotel from city centre in meters
* Customer rating: Rating on booking
* Rooms: Number of rooms
* Squares: m^2
* City: City in which hotel is

(AMAR ALADŽUZ AND 1 COLLABORATOR · UPDATED 2 YEARS AGO)

# 5.1) Applicable Regulations

a. Patents on ML algorithms developed

b. Laws related to privacy for collecting data from users

c. Protection/ownership regulations

d. Ensuring open-source, academic and research community for an audit of Algorithms.

e. Review of existing work authority regulations.

# 5.2) Applicable constraints

1) It is extremely difficult to obtain a global dataset of flights, hotels and tourist destination combined.

2) It is usually inevitable to have more than one data set or to combine all the different datasets into one common data set which is extremely difficult.

3) Data to be obtained by the users for the process to work may be confidential and they usually may not be open to selling their data very easily and this would also include many patents and local government regulations.

4) A proper mapping of different features manually by the trip advisor maybe necessary.

# 6.0 Benchmarking alternate products:

Travel providers can help travellers find the best time to book a hotel or to buy a cheap ticket by leveraging machine learning. When the deal becomes available, the app will likely send notification to user.

For example, [Hopper app](https://www.hopper.com/), known for helping customers track best flight deals, has recently incorporated a functionality for choosing hotels; a ML algorithm implemented in the app will recommend whether to book a hotel or to wait for the price to drop, similar to how it works with plane tickets. Artificial intelligence and predictive analytics are at the heart of this drive. Using the color-coded calendar, which marks expensive with red, moderate with yellow, and the cheapest with green, Hopper lets users view which dates will be more expensive than others.

# 7.0 Business Opportunity

Tour planning has always been an extremely popular field. A lot of large scale MNC’s and TRIP planning Companies world over are prevalent in the modern day.

But the small scale Trip planners and advisors in remote pockets and major cities always provide a personal touch and tailor-made experience which most of the huge platforms on the internet fail to provide.

So it is extremely important to save such a dying industry crying for help.

Modern machine learning algorithms can provide a common platform for all such Personal Trip advisors the world over which would make their job easier and also more lucrative due to the personal experience and also the digitalised machine touch with wider scope and options which humans can’t provide manually.

# 8.0 Concept Selection

The main goal and the concept we wish to achieve is to generate a fully automated environment for all small scale personal Trip advisors around the globe.

The environment utilises machine learning algorithms and provides the basic functionality and predictive analysis for the trip advisors which enables them to provide a personalised travel experience for each of the individual customers based on the data of similar customers who visited the same places with identical budgets and a similar travel route with matching tastes.

The system should be able to predict the price of the Air travel, costs for hotel booking during various seasons across numerous tourist locations which the customers wish to visit, and also should predict the nearby attractions the customers could visit around the main stay area of the tourists.

With the help of all these data, the trip advisors can perfectly suggest the best cost effective timing, duration and location for their customers and also suggest the Airlines and class and standards of hotel stay based on the budget put forth by their individual customers.

# 9.0 Concept Generation

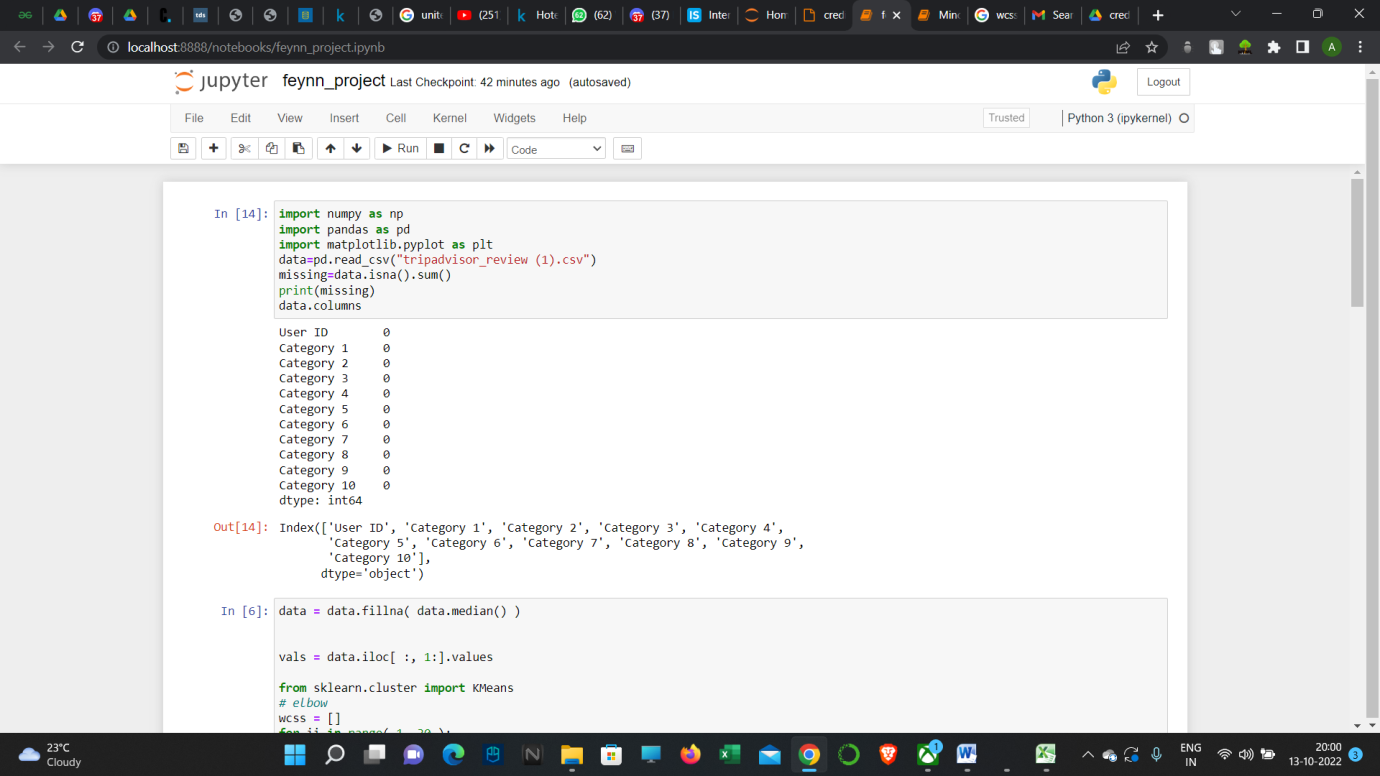
Initially, we apply clustering model on the trip advisor data set which we obtained (dataset 2).

We first import pandas to read and handle the trip\_advisor.csv file and access the data.

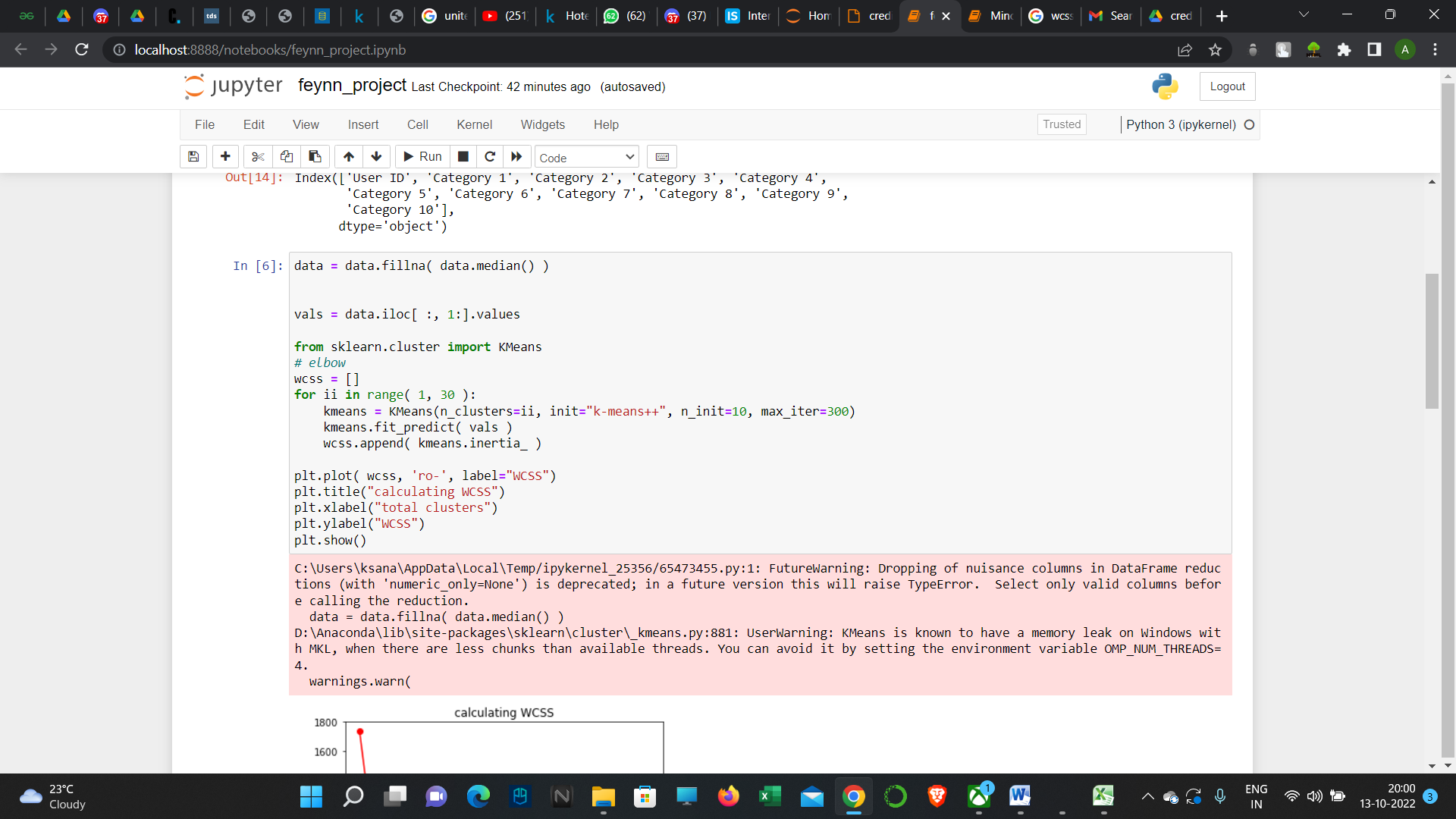
We also import matplotlib to plot the clustering model developed.

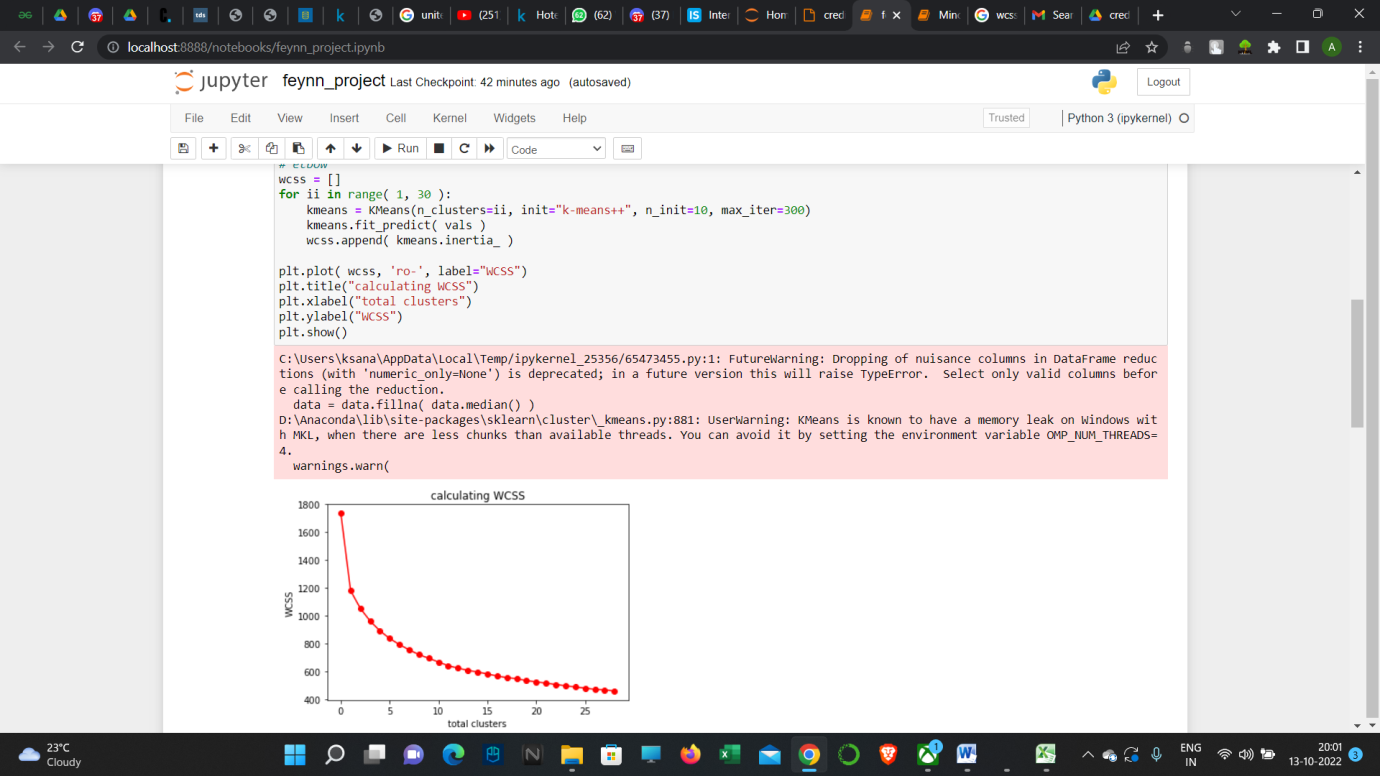
**Code implementation on small scale - 1:**

The columns and rows of the data set are checked for any null values, so that these can be filled with the mean or median values.

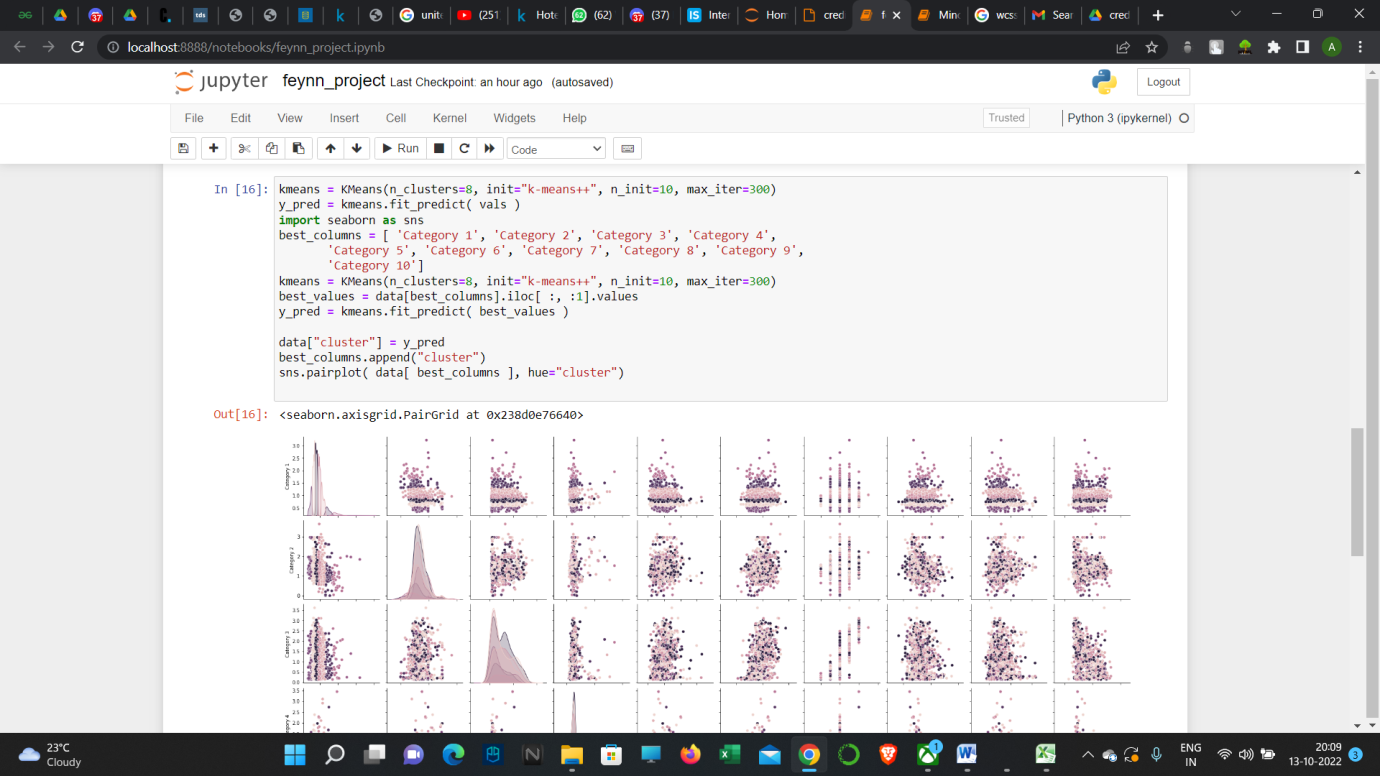


Secondly, we try to check the feasibility of the clustering model on the data set by applying the elbow method and plotting the Within Cluster Sum of Squares(WCSS).





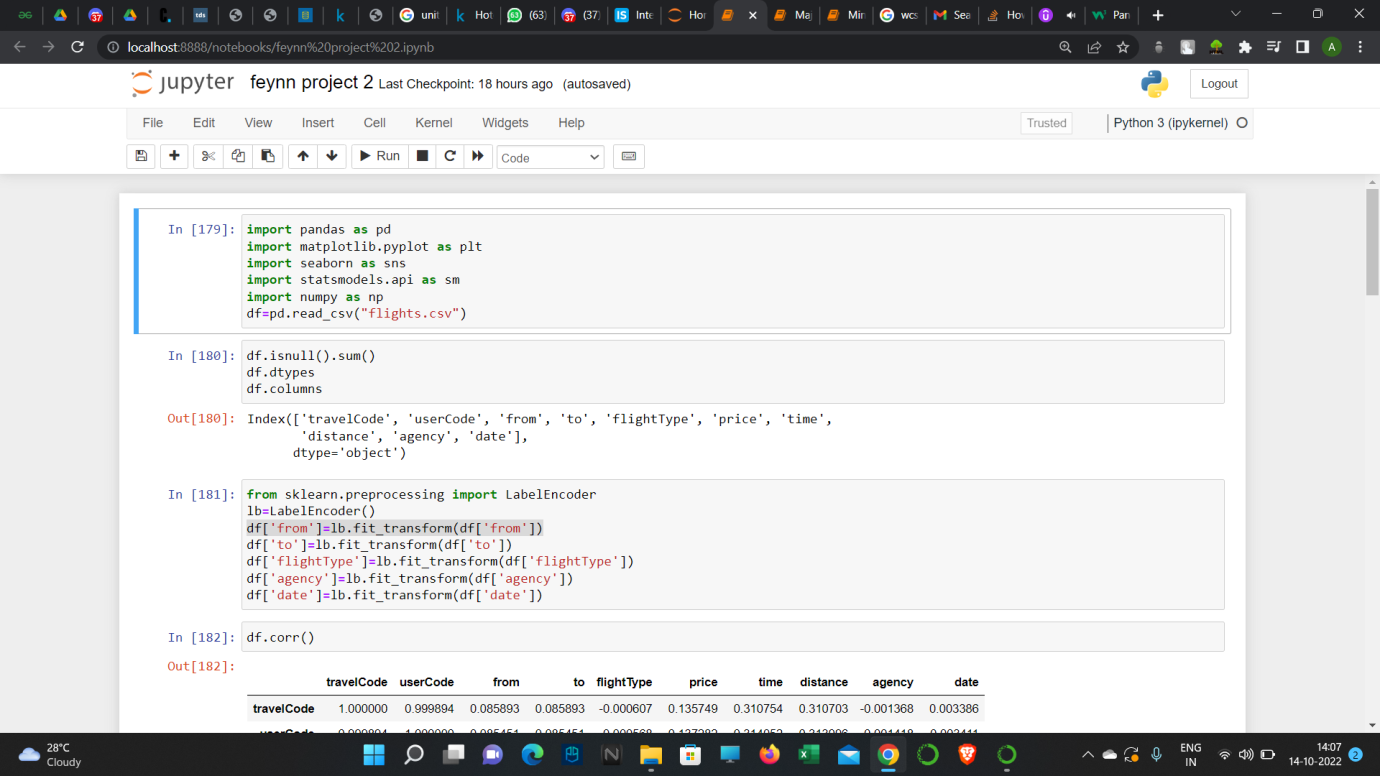
Once the feasibility is verified using the elbow method, we proceed to accessing all the columns of the dataset using the ‘iloc’ function and apply clustering to the dataset.



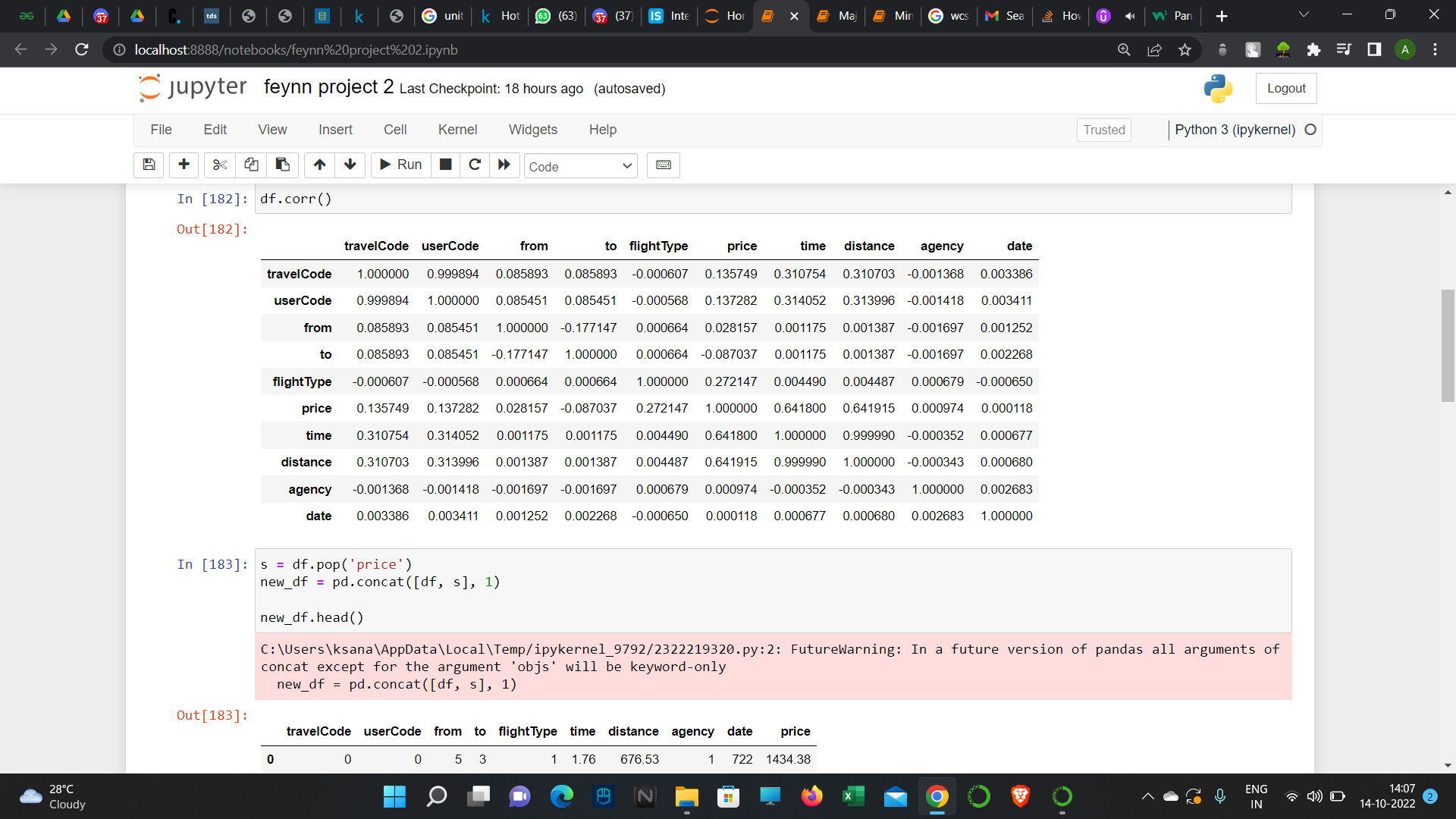
**Code Implementation on small scale - 2:**

Now we consider our first dataset, flights.csv and import the same modules as we did for the clustering on the previous implementation.

NOTE THAT FEATURE SCALING, NORMALISATION AND OTHER OPTIMISATION TECHNIQUES ARE NOT APPLIED IN THIS IMPLEMENTATION AS IT IS ON A SMALL SCALE AND THESE FURTHER PREPROCESSING OF THE DATA BEFORE THE APPLICATION OF THE REGRESSION MODEL MAY FURTHER IMPROVE THE ACCURACY AND SCORES.

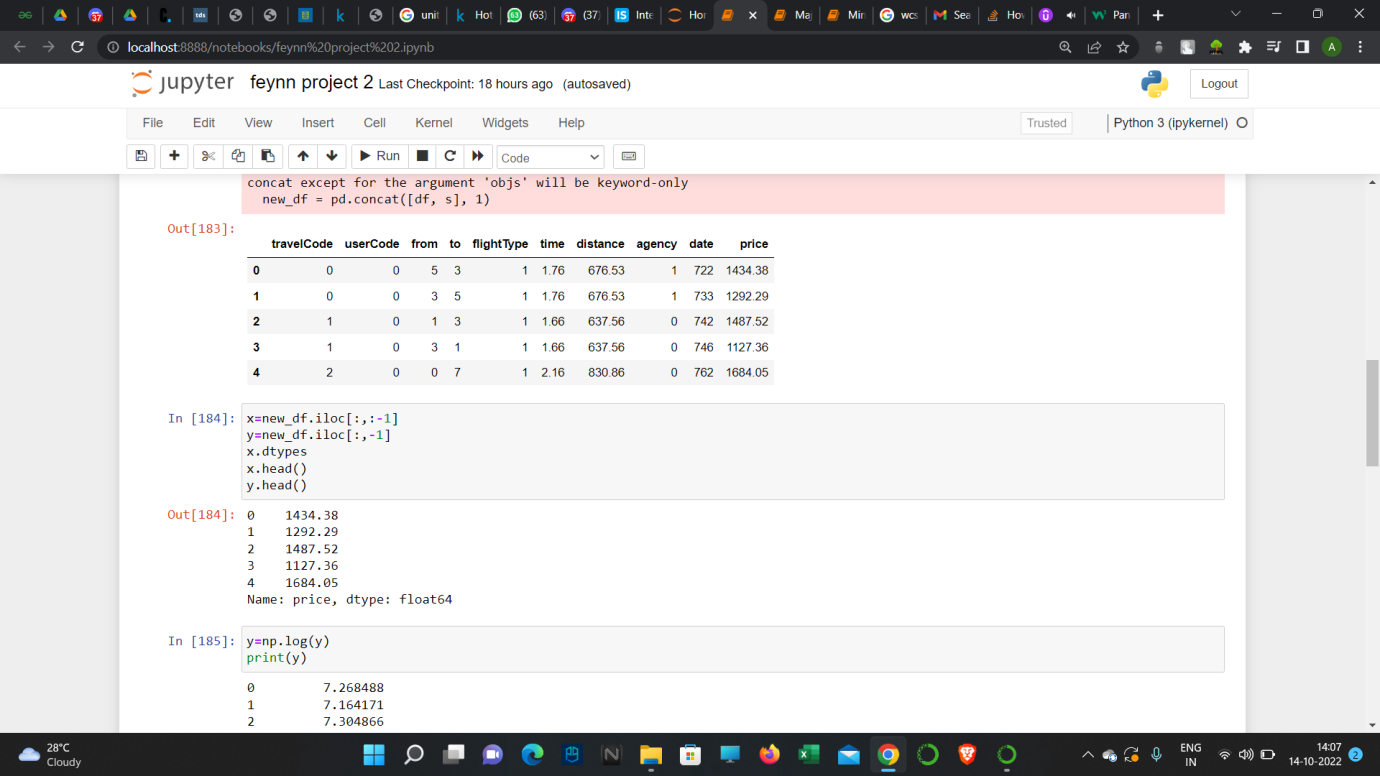


Since some of the attributes in the data set are ‘object’ type, we use LabelEncoder to convert these columns into continuous values which enables us to use the data further.

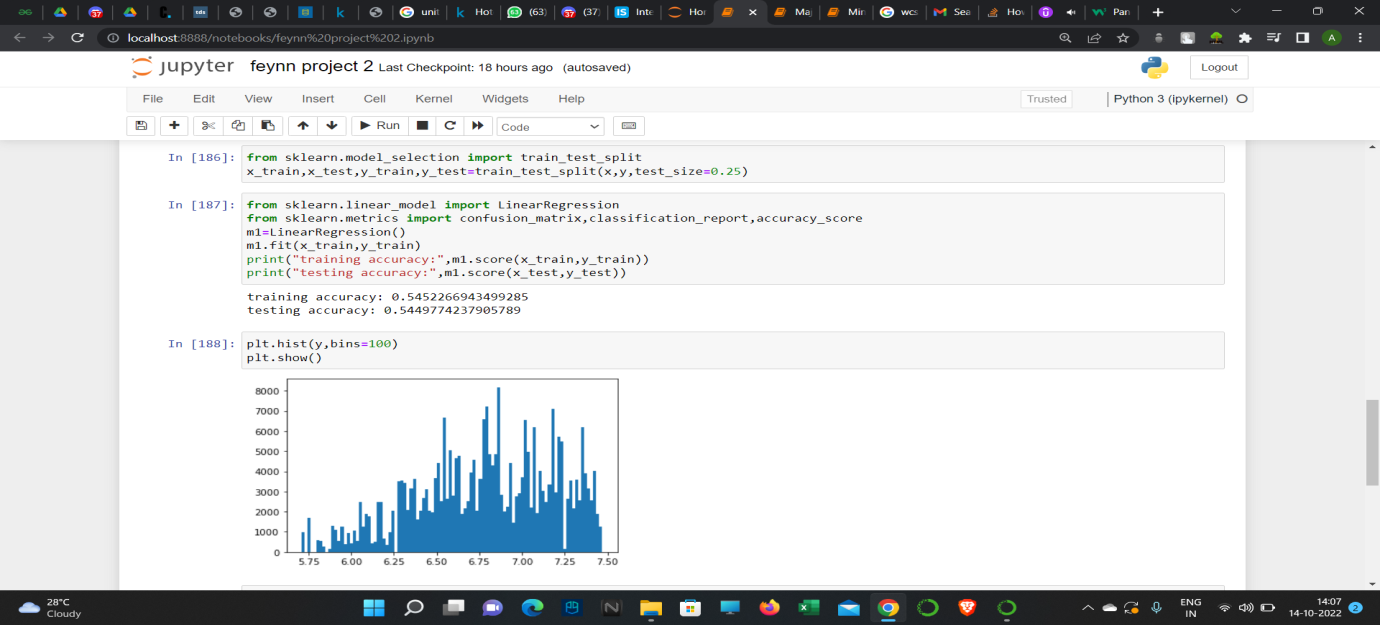


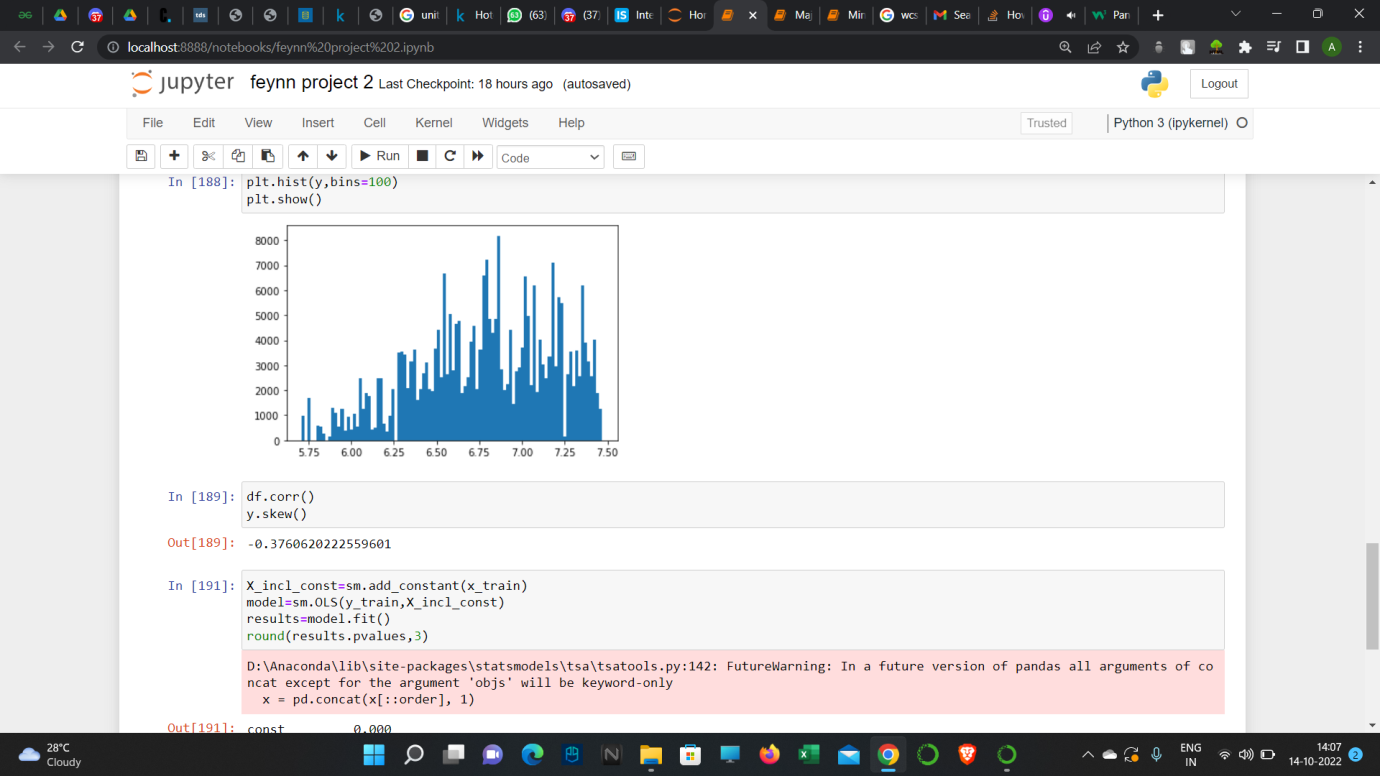
Further, the price column is moved to the last since it is the required output to be predicted.

This is done using the ‘pop’ and ‘concat’ functions which removes the column from a certain position and appends the same column at the end of the DataFrame which enables us to easily distinguish the price column as the required output.



Now we import the sci-kit learn module to train, test and split the data in order to fit it to the regression model.





Here we have only converted the output columns to log values to improve the skew, do not that the skew is still not ideal even after the application of log values.

PLEASE AGAIN NOTE THAT FEATURE SCALING, NORMALISATION AND OTHER OPTIMISATION TECHNIQUES ARE NOT APPLIED IN THIS IMPLEMENTATION AS IT IS ON A SMALL SCALE AND THESE FURTHER PREPROCESSING OF THE DATA BEFORE THE APPLICATION OF THE REGRESSION MODEL MAY FURTHER IMPROVE THE ACCURACY AND SCORES.

**10.0 Team required**

The team required for developing the environment includes:

1) Data scientists – 1 or 2 (depending on availability)

2) Front End engineers – to provide the required styling for the environment used by the travel agents.

3) Back end engineers to manage the data and also to improve upon the presently available data.

4) Network Engineers to provide and necessitate the workflow between the service providers-data scientists and the other engineers and also the customers which in this case would be the Independent personal travel agents to whom the environment is provided.

**11.0 Conclusion**

Initially, the purpose of the project was to device an environment or a platform for small scale independent personal travel agents and enable them to predict the rates and seasons of travel to destinations for their customers which would cut costs and give a personalized tailor-made experience of which places to visit during which time, which the machine learning model could predict.

On the small scale implementation of the model using the kaggle data sets we could obtain, we saw that even on small scale, we could get a decent prediction of prices both for hotels and flight fares.

The process of obtaining data on a large scale may yet prove cumbersome, but with a proper team dedicated to each field with data scientists to run machine learning algorithms and full stack developers to aid the data handling and presentation environments to the travel agents, achieving what this project set out to do could be accomplished.

**References:**

<https://towardsdatascience.com/how-machine-learning-and-ai-can-improve-travel-services-3fc8a88664c4>

([Alexandr Bulanov](https://medium.com/@alexandrbulanov?source=post_page-----3fc8a88664c4--------------------------------), Oct 3, 2018)

<https://smallbusiness.chron.com/there-honest-way-make-money-home-41320.html>

By Gregory Hamel

<https://archive.ics.uci.edu/ml/datasets/Travel+Reviews>

**Source:**

Shini Renjith, shinirenjith **'@'** gmail.com

<https://www.kaggle.com/datasets/leomauro/argodatathon2019?select=flights.csv>

**LEONARDO MAURO · UPDATED 3 YEARS AGO**