

NAME OF THE PROJECT

Submitted by:

YOUR NAME

**ACKNOWLEDGMENT**

This includes mentioning of all the references, research papers, data sources, professionals and other resources that helped you and guided you in completion of the project.

**INTRODUCTION**

* Business Problem Framing

In this modern everything comes under time , every hour every minute every seconds are important in daily life and people are trying to consume the tie by using different robotic technologies to consume time , the Transportation place a major role in consuming the time , like vacation business trip etc.., we do have transport like car, bus , train, bike etc., In which we see that flights are much faster that all other transportation for humans to travel around the world but it would seems like much expensive , everyone are willing to lower the cost on their travel in this part data scientist help the clients how can they book the flight in lower cost, with having an original data scientist can build a model and can help the client to book the ticket in lower cost

* Conceptual Background of the Domain Problem

The main concept behind the we hear proverbs like time is money but spending the money in travel which actually can lowered in looks like odd, so this project helps the client to book the ticket in lower cost as per their wish by following some criteria

* Review of Literature

The Research I have made for this projected with Wikipedia and with some the traveller blogs and vlog which added the advantage to get a result better

* Motivation for the Problem Undertaken

I would say my own experience which motivated me to build this model, we were planned to Goa but the ticket cost has around 10000 bucks, which made me to research in booking a flight n lower cost which is quite interesting

**Analytical Problem Framing**

* Mathematical/ Analytical Modelling of the Problem

The major analytics done in this project only with one source and destination where to do this project I have collected the data’s like Source destination, date of journey, Number of stops, Hours of travel all that which analytics helped me to get the better results in this project

* Data Sources and their formats

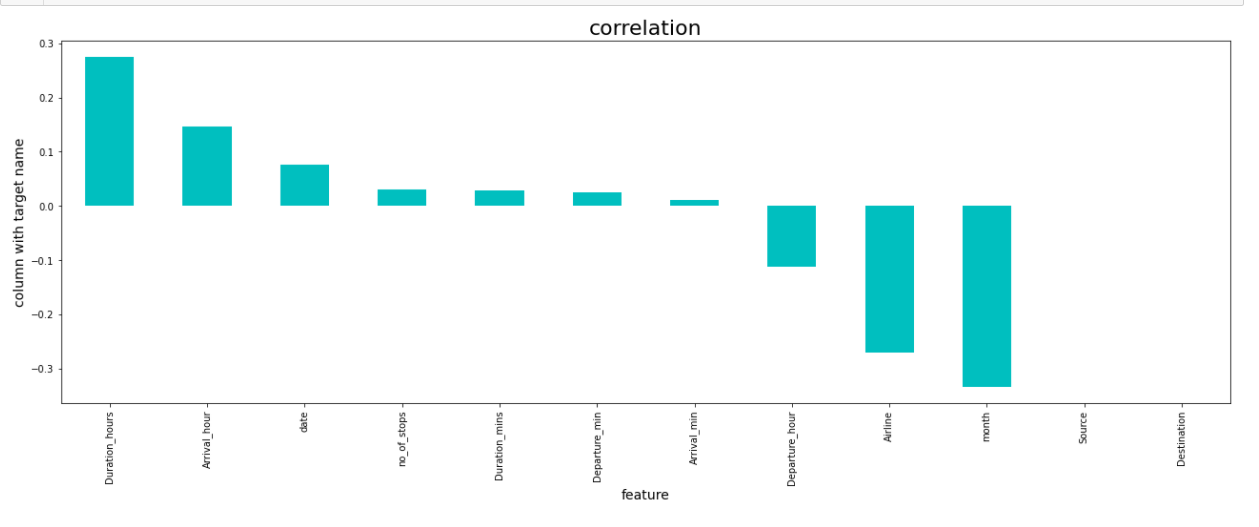
The Data sources are collected form the website like Sky Scanner, Make my trip, which quite lesser cost compared to the other website , and the format of the data’s are Categorical and numerical and date and time format

* Data Pre-processing Done

The data pre-processing has been done with the data which seems like higher cost which is more than 50000 bucks also the some of the outlier removed which are higher cost than expected

* Data Inputs- Logic- Output Relationships

The Data input are much correlated with the out where some of the data are more correlated with the output I do see that the airline, Number of stops and hour of journey has affected the output more also I have attached the picture for your reference

* State the set of assumptions (if any) related to the 
* Hardware and Software Requirements and Tools Used
* The software I used is Jupyter notebook

**The libraries I used are**

Pandas

Numpy

Seaborn

Matplotlib

PANDAS

Pandas is mainly used for data analysis. Pandas allows importing data from various file formats such as comma-separated values, JSON, SQL, and Microsoft Excel. Pandas allows various data manipulation operations such as merging, reshaping, selecting, as well as data cleaning, and data wrangling features.

NUMPY

Numpy is a Python library that provides a simple yet

Powerful data structure: the n-dimensional array. This is the

Foundation on which almost all the power of Python's data science

Toolkit is built

SEABORN

Seaborn is a library for making statistical graphics in Python.

Its plotting functions operate on data frames and arrays containing

Whole datasets and internally perform the necessary semantic

Mapping and statistical aggregation to produce informative plots.

MATPLOTLIB

Matplotlib is a Python library that is used often with

Jupyter Notebook. Matplotlib. ... ePyplot provides a MATLAB-like way

Of plotting.

**Model/s Development and Evaluation**

* Identification of possible problem-solving approaches (methods)
* Testing of Identified Approaches (Algorithms)
* The used algorithms are

Linear Regression

Knearest neighbour

Random forest regressor

Decision tree Regressor

Support vector Regressor

.

* Run and Evaluate selected models

Linear Regression

Linear Regression is a machine learning algorithm based on supervised learning. ... Linear regression performs the task to predict a dependent variable value (y) based on a given independent variable (x). So, this regression technique finds out a linear relationship between x (input) and y(output).

Knearest neighbour

It is a supervised machine learning algorithm. The algorithm can be used to solve both classification and regression problem statements. The number of nearest neighbours to a new unknown variable that has to be predicted or classified is denoted by the symbol 'K'.

Random Forest Regressor

Random Forest Regression is a supervised learning algorithm that uses ensemble learning method for regression. ... A Random Forest operates by constructing several decision trees during training time and outputting the mean of the classes as the prediction of all the trees.

Decision tree Regressor

Decision tree builds regression or classification models in the form of a tree structure. It breaks down a dataset into smaller and smaller subsets while at the same time an associated decision tree is incrementally developed. The final result is a tree with decision nodes and leaf nodes.

Support vector Regressor

Support Vector Regression is a supervised learning algorithm that is used to predict discrete values. Support Vector Regression uses the same principle as the SVMs. The basic idea behind SVR is to find the best fit line. In SVR, the best fit line is the hyper plane that has the maximum number of points.

* Key Metrics for success in solving problem under consideration

The statistics says that if the number of Stops increases the Price of flight ticket will increase also booking a ticket in last moment will increase the ticket price and Traveling in weekend will be little costlier , focusing on tis given a good result on this project

* Visualizations

I have done with

Distribution plot

Box plot

Swarm plot

Histogram plot

Heat map

Violin plot

Cat plot

In all this which mainly described of Number of stops increase price will increase, if booking a ticket in last minute price will be higher, le me add the heat map here which helps you to understand the correlation between the input and output



**CONCLUSION**

* Key Findings and Conclusions of the Study

The main thing I have noticed booking a ticket prior of week And booking a ticket with less number of stops will helps the passenger or client to book the ticket in lower cost

* Learning Outcomes of the Study in respect of Data Science

Really I would say the Visualization helped me lot to analysis and predict the things what I need to do next, which helped me to remove the outlier and helped to check the correlation between the each data