iNeuron Machine Learning Internship

Flight Price Prediction

**PROJECT REPORT**

TITLE: Flight Price Prediction

**AIM:**

The main objective of this project is to predict the flight price given certain factors that are proven to be effective on Flight Price. In Airline Business, flight Price is an important factor that drives customer base. Customers generally tend to go for budget friendly. So, airline company may use different factors to estimate flight price and adjust the pricing accordingly in real time. The use case of this application can be expanded to Flight Price Forecast and Flight Recommendation Applications.

# ABSTRACT:

As per the problem statement, we can make use of different machine learning regression algorithms to do the predictions. The data used for the algorithm was well structured In tabular format and proper data dictionary was available for the data. So, we proceeded with all the data processing and modeling activities.

# INTRODUCTION:

# The dataset used for this project was extracted from Kaggle.

# Link: https://www.kaggle.com/nikhilmittal/flight-fare-prediction-mh

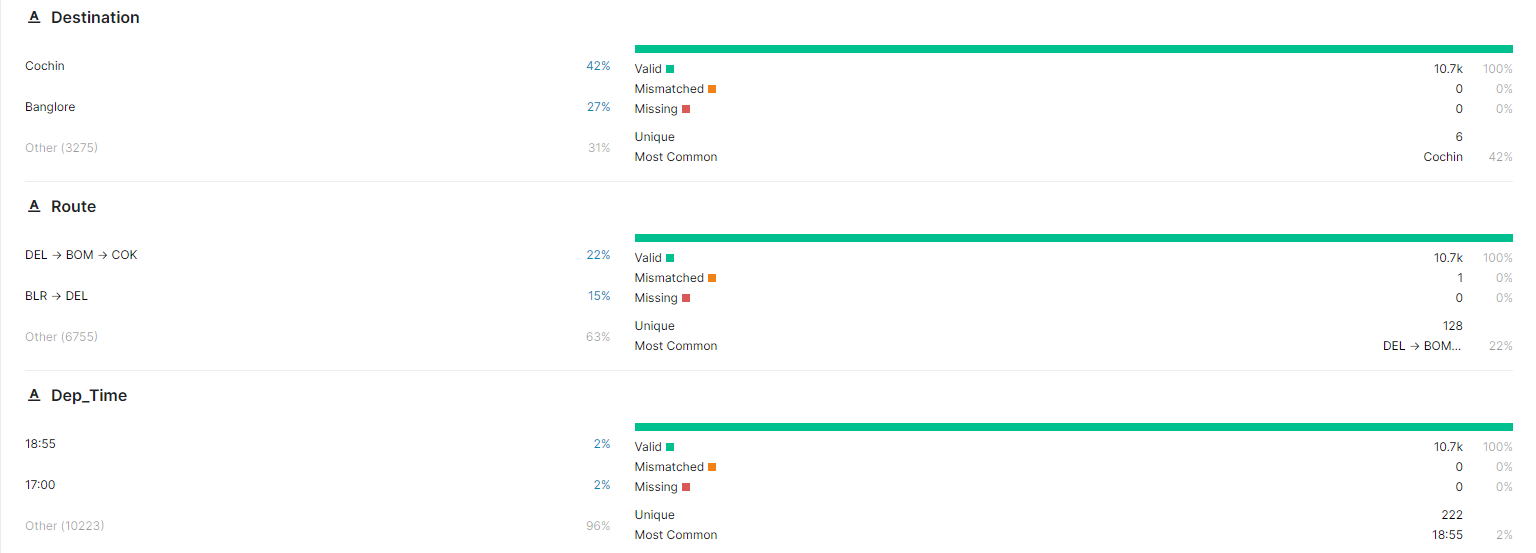
# Data Format: Excel

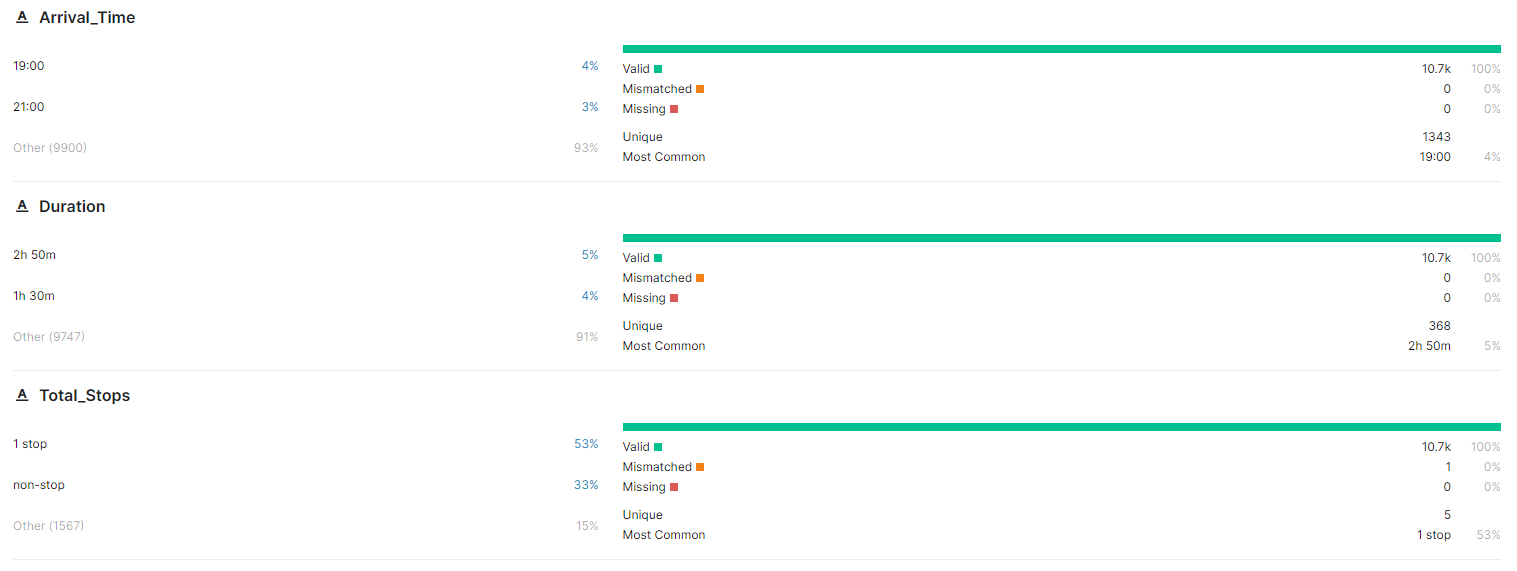
# Additional Info about Data Source:

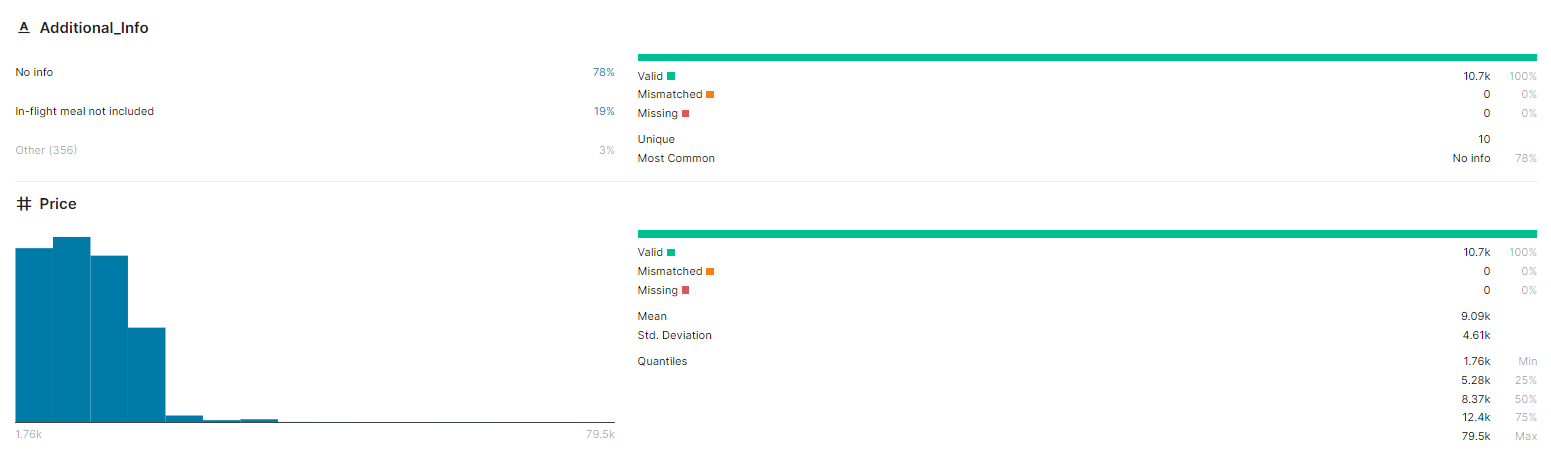
# 

# 

# 







# OVERVIEW:

* Data Segmentation and Data Cleaning
* Exploratory Data Analysis using python’s data visualization libraries.
* Training the model based on the historical data available.

# DATA UNDERSTANDING:

# 

# Columns:

* Airline: Name of the airline used for traveling
* Date\_of\_Journey: Date at which a person traveled
* Source: Starting location of flight
* Destination: Ending location of flight
* Route: This contains information on starting and ending location of the journey in the standard format used by airlines.
* Dep\_Time: Departure time of flight from starting location
* Arrival\_Time: Arrival time of flight at destination
* Duration: Duration of flight in hours/minutes
* Total\_Stops: Number of total stops flight took before landing at the destination.
* Additional\_Info: Shown any additional information about a flight
* Price: Price of the flight (**Target**)

# DATA SEGMENTATION AND DATA CLEANING:

* At first, we preformed AutoEDA on the raw data, for initial inference about the data to get overall data information, dimension, missing counts, correlations etc.
* The dataset had few missing values. We dropped those rows.
* We visualized categorical variables with count plot and numerical target variable using boxplot.
* Most of the features were numerical but given as object in terms of their representation, so manual processing was done.
* We manually handled date, time and datetime field using custom functions.
* We Extracted day and month as numerical value form date column.

# EXPLORATORY DATA ANALYSIS

# Using AutoEDA we analyzed raw data and generated some descriptive statistics and plots based on its data types.

# 

# Stats

# 

# 

# 

# 

# 

# 

# Plots:

# 

# 

# 

# 

# 

# 

# 

# 

# DATASET AFTER NECESSARY DATA CONVERSIONS

# After AutoEDA we performed necessary data conversions and plotted the data manually.

# 

# HANDLING CATEGORICAL DATA

# 

# 

# 

# 

# 

# 

# HANDLING NUMERICAL DATA

# 

# MODELLING USING RANDOM FOREST REGRESSOR

# Prior to modeling we performed train-test split in 4:1 ratio, and also cross checked if the train and test target distribution is similar to overall target distribution or not. After confirmation we went ahead with performing base modeling, feature selection using Sequential Feature Selector technique and again re-modelling using significant features .

# 

# SELECTED FAETURES BASED ON FEATURE SELECTION

# 

# TRAIN AND TEST SCORES OF BEST-FIT MODEL

# 

# 

# EXPORT OF MODEL FILE USING PICKLE

# 

# WEB APPLICATION FOLDER STRUCTURE

# 

# app.py

# 

# requiremnts.txt

# 

# WEB APPLICATION HOSTED USING STREAMLIT-SHARE

# Link: <https://share.streamlit.io/hassananjum913/flight-price-prediction-app/app.py>

# 

# MODEL BUILDING:

# Random Forest:

# DEPLOYMENT:

**STREAMLIT DEPLOYEMENT:**

# The random forest model is deployed using streamlit and streamlit-share into a real time website.

# The link to the website is:

**TEAM MEMBERS:**