Dr. Vishwanath Karad
MIT WORLD PEACE UNIVERSITY | PUNE
TECHNOLOGY, RESEARCH, SOCIAL INNOVATION & PARTNERSHIPS

Interim Report of TY Mini Project

On

Exploratory Data Analysis & Model Analysis on Aviation Dataset

By

Devesh Khanna (1032201115)
Chinmayi Railkar (1032201218)
Sharvari Dhamale (1032201837)

Under the guidance of

Prof. Jayshree Aher

MIT-World Peace University (MIT-WPU)

Faculty of Engineering
School of Computer Engineering & Technology

*** 2023-2024 ***

TY Miniproject, MinorCSE Submission:

Date: 15.05.2023

Sr. No.	Student Name	PRN No.	Sign
1	Chinmayi Railkar	1032201218	
2	Sharvari Dhamale	1032201837	
3	Devesh Khanna	1032201115	

Prof.Jayshree Aher
Coordinator, SCET, MinorCSE

Interim Report for TY Mini Project

Team ID: _____

1) PROBLEM STATEMENT

To do primary exploratory data analysis on the data, create various graphs to better understand the data, and apply various ML models, it is important to find a model that is appropriate for aviation data.

2) OBJECTIVE

The project objectives include:

1. Performing Exploratory Data Analysis
2. Understanding various ML algorithms
3. Understanding various performance parameters
4. Analyzing the models

3) INTRODUCTION

Exploratory Data Analysis (EDA) is a valuable tool used by data scientists to examine and study datasets. By employing various data visualization techniques, EDA enables data scientists to uncover patterns, detect anomalies, test hypotheses, and validate assumptions. It helps in gaining a deeper understanding of the variables within the dataset and their relationships. EDA goes beyond formal modeling or hypothesis testing, allowing for a comprehensive exploration of the data. Additionally, EDA assists in assessing the appropriateness of statistical methods considered for data analysis. The primary objective of EDA is to foster data analysis without making any assumptions. It aids in identifying errors, comprehending data patterns, detecting outliers, and revealing relationships between variables. The insights gained through EDA are reliable and directly applicable to business objectives and goals. EDA also benefits stakeholders by ensuring they are asking the right questions. It leverages measures such as standard deviations, categorical variables, and confidence intervals to facilitate understanding. As EDA progresses, it can be used for more sophisticated data analysis, such as employing machine learning techniques to enhance accuracy. Ultimately, EDA enables the formulation of conclusive findings based on the analyzed data. Thus we are going to perform exploratory data analysis on airfare data and perform model analysis and get accuracy of various machine learning models.

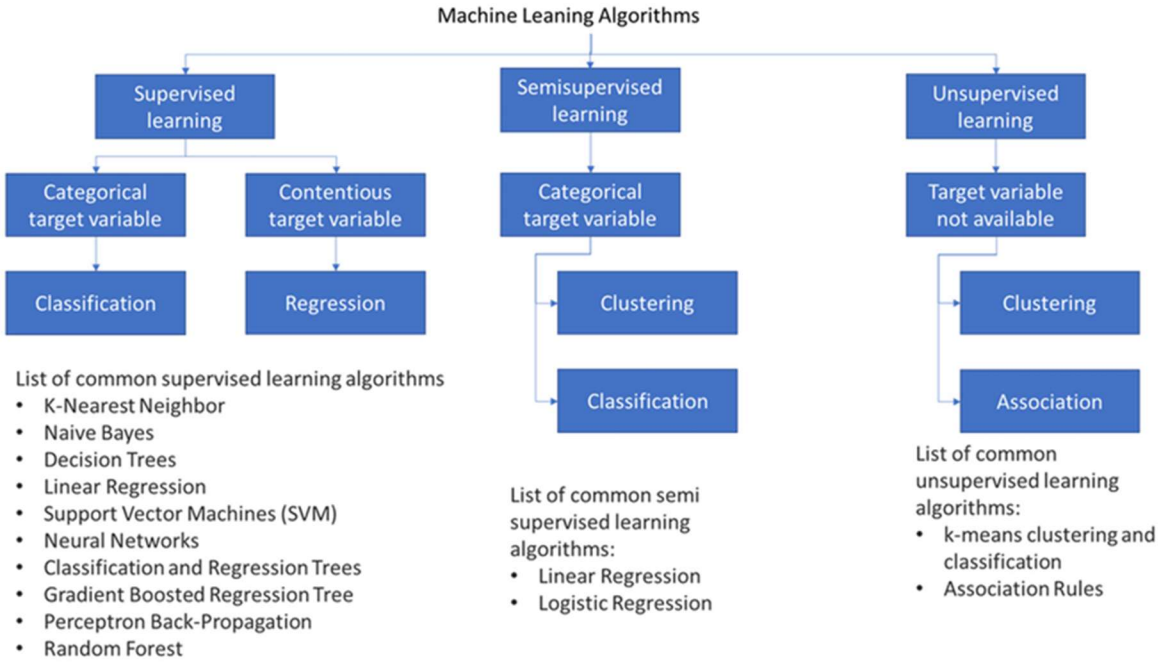
4) MOTIVATION

1. There are various apps available for purchasing airline tickets. If travelers desire to go from one place to another, they may not truly be aware of the cost of that particular airline.

2. There will be a system that allows customers to buy airplane tickets in accordance with their needs in order to save money and time.

5) METHODOLOGY

The project's methodology entails data analysis for forecasting airfares as well as data pre-processing and visualization. As there are several machine learning techniques (supervised and unsupervised), the model analysis of the algorithm is an important step.



The hardware and software part of the project includes a suitable python system and hardware devices suitable for implementing the code.

6) LITERATURE SURVEY (ML TECHNIQUES & DATA ANALYSIS)

A literature survey on various machine learning algorithms and data analysis would involve exploring a wide range of research papers, books, and articles that discuss these topics. Here is a summary of some popular machine learning algorithms and data analysis techniques commonly covered in the literature:

1. Supervised Learning Algorithms:

- Linear Regression
- Logistic Regression
- Decision Trees
- Random Forests
- Support Vector Machines (SVM)
- Gradient Boosting methods (e.g., XGBoost, AdaBoost)

2. Unsupervised Learning Algorithms:

- K-Means Clustering
- Hierarchical Clustering
- Principal Component Analysis (PCA)
- Association Rule Learning (e.g., Apriori algorithm)
- Self-Organizing Maps (SOM)
- Gaussian Mixture Models (GMM)

3. Deep Learning Algorithms:

- Artificial Neural Networks (ANN)
- Convolutional Neural Networks (CNN)
- Recurrent Neural Networks (RNN)
- Long Short-Term Memory (LSTM)
- Generative Adversarial Networks (GAN)
- Transformer models (e.g., BERT, GPT)

4. Dimensionality Reduction Techniques:

- Principal Component Analysis (PCA)
- t-Distributed Stochastic Neighbor Embedding (t-SNE)
- Non-negative Matrix Factorization (NMF)
- Autoencoders

5. Data Preprocessing and Feature Engineering:

- Missing Data Imputation
- Feature Scaling and Normalization
- One-Hot Encoding and Feature Encoding
- Feature Selection techniques
- Handling Imbalanced Datasets

6. Evaluation Metrics and Model Selection:

- Accuracy, Precision, Recall, F1 Score
- Receiver Operating Characteristic (ROC) Curve
- Cross-Validation
- Hyperparameter Tuning
- Model Selection techniques (e.g., Grid Search, Random Search)

7. Exploratory Data Analysis (EDA) Techniques:

- Data Visualization (e.g., scatter plots, histograms, box plots)
- Correlation Analysis

- Outlier Detection

- Data Cleaning and Data Wrangling

- Statistical Methods (e.g., hypothesis testing, confidence intervals)

This literature survey should provide you with a starting point to explore the vast body of knowledge on machine learning algorithms and data analysis. It's important to delve into specific research papers and resources for in-depth understanding and implementation details of each algorithm or technique.

7) ALGORITHM & TECHNIQUES

Supervised (Classification) Machine Learning Algorithms

KNN

K-Nearest Neighbor (KNN) is a straightforward supervised learning algorithm widely used in machine learning. It operates by comparing new cases with existing cases and assigning the new case to the category that closely resembles the existing categories. The algorithm categorizes a new data point by referencing the stored dataset, enabling swift and accurate classification.

While KNN is primarily employed for classification tasks, it can also be utilized for regression. Furthermore, KNN is considered a non-parametric technique. It falls under the category of lazy learning algorithms, as it does not actively learn from the training dataset but rather stores it for classification purposes.

The KNN algorithm follows a specific process to classify new data points:

1. Determine the number of neighbors (K).

2. Calculate the Euclidean distance between the new data point and the neighbors.

3. Select the K neighbors based on the computed Euclidean distance.

4. Count the data points in each category among the K neighbors.

5. Assign the new data point to the category with the highest neighbor count.

6. The model is then considered complete

By following these steps, the KNN algorithm effectively classifies new data points based on their similarity to the existing dataset.

Naïve Bayes Algorithm

The Naive Bayes algorithm is a supervised learning method used for classification tasks, leveraging the principles of Bayes' theorem. It is widely recognized as an efficient classification algorithm, enabling the development of accurate predictive machine learning models.

Naive Bayes is categorized as a probabilistic classifier and finds applications in various domains such as spam filtration, sentiment analysis, and article classification. The name "Naive Bayes" can be understood as follows:

- "Naive" refers to the assumption that the occurrence of one feature is independent of the occurrence of other features. For example, when identifying an apple based on its color, shape, and flavor, the naive assumption is that each characteristic contributes independently to the identification of an apple.

- "Bayes" indicates that the algorithm is based on Bayes' theorem, a fundamental concept in probability theory.

Naive Bayes encompasses different models, one of which is the Gaussian model. In the Gaussian model, the distribution of data points follows a normal distribution. If the predictor values take continuous values, they are treated as samples from a Gaussian (normal) distribution.

By utilizing these principles, the Naive Bayes algorithm facilitates accurate classification by making use of probabilistic reasoning and the assumption of feature independence.

SVM (Support Vector Machine)

Support Vector Machine (SVM) is a versatile machine learning algorithm used for both classification and regression problems, although it is predominantly used for classification tasks. The main objective of SVM is to find the optimal hyperplane that separates data points of different classes in an n-dimensional space.

Here are some key points about SVM:

1. Hyperplane and Support Vectors: In SVM, the hyperplane represents the decision boundary that maximally separates the data points of different classes. The goal is to find the hyperplane with the largest margin between classes. Support vectors are the data points that lie closest to

the decision boundary and have the most influence on determining the position and orientation of the hyperplane.

2. Linear SVM: Linear SVM is used when the data is linearly separable, meaning it can be divided into two classes using a straight line (in 2D) or a hyperplane (in higher dimensions). Linear SVM seeks to find the best hyperplane that maximizes the margin between the two classes.

3. Non-linear SVM: In cases where the data cannot be separated by a linear boundary, non-linear SVM is used. Non-linear SVM applies a technique called the kernel trick, which transforms the original feature space into a higher-dimensional space, where the data becomes linearly separable. By mapping the data into this higher-dimensional space, a linear hyperplane can be found to classify the data. Common kernel functions used in non-linear SVM include polynomial kernels, radial basis function (RBF) kernels, and sigmoid kernels.

SVM is known for its ability to handle high-dimensional data and its robustness against overfitting. It is widely used in various applications such as text classification, image recognition, and bioinformatics. The choice between linear and non-linear SVM depends on the nature of the data and the complexity of the classification problem at hand.

Unsupervised Machine Learning Algorithms

K – means Clustering

The K-Means Clustering is a type of Unsupervised algorithm that groups the unlabeled dataset into clusters. K defines the number of predetermined clusters , for example if K=2, there will be two clusters, K=3 there will be three clusters used.

K- Means Clustering is a centroid-based algorithm. Each cluster has a centroid. The aim is to minimize the distances between the data point and clusters.

The input is the unlabelled data that divides the dataset into k clusters. The process is repeated to get the best cluster. The value of k is predetermined in the algorithm.

The k-means algorithm performs two tasks:

- It determines the value for K points by an iterative process.
- It assigns each data point to the closest k-center which creates the cluster.

The working of the K-Means algorithm is as following:

- Step-1: The number K for the clusters is selected.
- Step-2: Random points for centroids selected.
- Step-3: Each data point is assigned to the closest centroid from the defined clusters.
- Step-4: The variance and position of each centroid is calculated.
- Step-5: Repeat the process to redesign the data points to the centroid to get the best cluster.
- Step-6: If any error occurs repeat the step-4 else step 7
- Step-7: The model completes the training.

Hierarchical Clustering

Hierarchical clustering is indeed a type of unsupervised machine learning algorithm used for grouping unlabeled datasets into clusters. It organizes the data points into a hierarchical structure, often visualized as a tree-like diagram called a dendrogram. There are two main approaches to hierarchical clustering:

1. Agglomerative: This bottom-up approach starts with each data point as an individual cluster and iteratively merges the closest clusters until a single cluster encompassing all the data points is formed. The distance between clusters is calculated based on a chosen linkage criterion, such as the Euclidean distance or the complete linkage distance.

2. Divisive: In contrast to agglomerative clustering, the divisive approach begins with a single cluster containing all the data points and recursively splits it into smaller clusters until each cluster consists of a single data point. This top-down process separates the data points based on certain criteria, such as maximizing the inter-cluster dissimilarity or minimizing the intra-cluster variance.

The steps involved in hierarchical clustering are similar to those in k-means clustering, but with some differences:

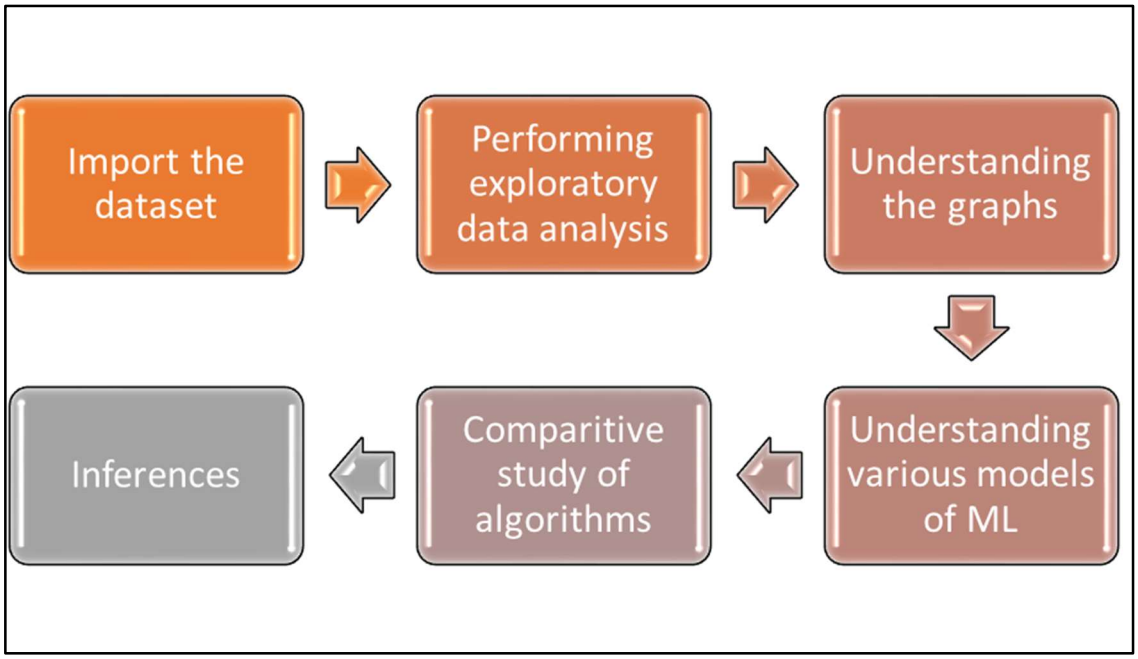
1. Data Pre-processing: As with any clustering algorithm, it is important to preprocess the data by handling missing values, normalizing or standardizing features, and addressing outliers.

2. Finding clusters using the Dendrogram: Hierarchical clustering generates a dendrogram, which illustrates the hierarchy of clusters. The dendrogram helps determine the number of clusters by identifying significant jumps in the dissimilarity or distance between data points or clusters.

3. Training the hierarchical clustering model: The hierarchical clustering algorithm does not require a predetermined number of clusters. Instead, it uses the dendrogram to create clusters based on the desired level of similarity or dissimilarity.

4. Visualizing the clusters: Once the clusters are determined, they can be visualized to gain insights into the structure of the data. Various techniques, such as scatter plots or heatmaps, can be employed to represent the clusters visually.

8) FLOW DIAGRAM



9) LITERATURE SURVEY

Paper Ref. (as per order of references)	Title of Paper	Year	Basic idea presented in paper	Dataset used	Method Used	Results achieved	Gaps / shortcomings / future scope
Ref[1]	Predictive analytics with aviation big data.	2013	Big Data analytics can be used in the airline industry to improve the performance of aviation operations	Evaluation of Aviation Data	-	The presented paper is a review paper hence results are in form of sentences	Using data analysis for improvement in aviation industry
Ref [2]	"Airfare price prediction using machine learning techniques" in European Signal Processing Conference	2017	Different numbers of features were used to train models to showcase how selection of features can change	Dataset consisting of 1814 data flights to train machine learning model.	Multilayer Perceptron (MLP), Generalized Regression Random Forest Regression Tree. o Regression Tree, Regression SVM and Linear	Various types of models with removing and adding different features from the dataset. Followed a typical data science life cycle. The best results came from	Using various features for the same dataset is questionable.,

			the accuracy of the model.		Regression used	the Bagging regression tree.	
Ref [3]	A survey on flight fare prediction using machine learning algorithm	2019	Performing research on a small dataset and deriving results based on various parameters.	Small dataset Delhi and Bombay flights	Algorithms like KNN, linear regression, support vector machine are used	The best model results were of the Decision Tree algorithm.	Carrying out the same procedure with bigger datasets
Ref [4]	"A Framework for airline price prediction: A machine learning approach"	2018	In this paper where two databases are combined together with macroeconomic data	Combination of 2 data sets	Machine learning algorithms such as support vector machines, XGBoost are used to model the average ticket price based on source and destination pairs.	The framework achieves a high prediction accuracy 0.869 with the adjusted R squared performance metrics	
Ref [5]	"A linear quantile mixed regression model for prediction of airline ticket prices"	2018	To find the best time for flight using the Linear Quantile Blended Regression methodology for San	Timid dataset for aviation.	Various machine learning algorithms like linear regression, Decision Tree, Random Forest, K-NN support vector machine, etc. Prediction algorithms like	Two features such as number of days for departure and whether departure is on weekend or weekday are considered to develop the model.	

			Francisco –New York course where each day airfares are given by online website.		Naïve Bayes and Stacked Prediction Model are used		
Ref [6]	"Exploratory Data Analysis of Airfare Dataset for Prediction using Machine Learning Techniques" by Gupta et al. (2018)	2018	The study employed various EDA techniques, including statistical analysis, data visualization, and correlation analysis, to explore an airfare prediction dataset. They analyzed factors such as departure and arrival airports, flight duration, airlines, and seasons.	Aviation data	Identified significant correlations between fare prices and factors such as flight duration, seasonality, and the presence of layovers. Discovered patterns indicating higher fares during peak travel seasons and for certain airlines.	Provided valuable insights into the relationships between airfare prices and different factors, aiding the development of accurate prediction models.	

Ref [7]	"Exploratory Data Analysis of Airfare Prediction Data using Big Data Analytics" by Das et al. (2019)	2019	The study used EDA techniques to analyze a large-scale airfare prediction dataset. They applied descriptive statistics, data visualization, and clustering analysis to identify patterns and trends in the data.		Uncovered price variations based on airlines, routes, and travel seasons. Identified clusters of similar fare patterns, enabling targeted pricing strategies for airlines.	Demonstrated the effectiveness of EDA in understanding complex airfare prediction data and provided insights to enhance pricing strategies for airlines.	
Ref [8]	"Exploratory Data Analysis of Airfare Datasets: Insights and Trends" by Sharma et al. (2020)	2020	The study conducted an extensive EDA of multiple airfare datasets, focusing on factors such as flight routes, airlines, departure time, and ticket class. They employed		Identified temporal trends in airfare prices, including price fluctuations based on specific days of the week and time of day. Discovered variations in prices across different airlines and ticket classes.	Provided valuable insights into the temporal and contextual factors affecting airfare prices, facilitating better pricing strategies and decision-making for airlines and travelers.	

			visualization techniques, statistical analysis, and outlier detection methods.				
Ref [9]	"Exploratory Data Analysis of Domestic Airfare Dataset for Predictive Analytics" by Patel et al. (2021)	2021	The study performed EDA on a domestic airfare dataset, using descriptive statistics, data visualization, and time series analysis. They focused on factors such as flight distance, flight duration, departure time, and airline popularity.		Revealed relationships between airfare prices and flight distance, flight duration, and departure time. Detected seasonal patterns and trends in airfare prices.	Provided insights into the relationships between airfare prices and various flight characteristics, enabling more accurate predictive analytics and pricing strategies.	
Ref [10]	"Predicting Airfare Price Trends using	2018	The study applied regression analysis		Obtained accurate price trend predictions	Coefficient of Determination (R-	

	Regression Analysis" by Patel et al. (2018)		techniques, specifically linear regression and polynomial regression, to predict airfare price trends. They considered factors such as departure and arrival airports, flight duration, airline popularity, and historical fare data.		using polynomial regression. Found that flight duration, departure time, and airline popularity significantly influenced airfare prices.	squared), Mean Absolute Percentage Error (MAPE), and Mean Squared Error (MSE) were used to evaluate the accuracy of the prediction models.	
Ref [11]	"Airfare Price Prediction using Machine Learning and Big Data Analytics" by Gohil et al. (2019)	2019	The study employed machine learning algorithms such as decision trees, gradient boosting, and random forests to predict airfare prices.		Achieved accurate airfare predictions using gradient boosting, which outperformed other algorithms. Seasonal factors, flight duration, and departure time were found to be crucial in	: Mean Absolute Percentage Error (MAPE) and Root Mean Squared Logarithmic Error (RMSLE) were used to assess the prediction accuracy.	

			They incorporated features like departure and arrival cities, flight duration, and seasonal information.		determining airfare prices.		
Ref [12]	"Airfare Prediction using Deep Learning and Ensemble Methods" by Lee et al. (2020)	2020	The study combined deep learning techniques, such as recurrent neural networks (RNNs), with ensemble methods, including gradient boosting and random forests, to predict airfare prices. They utilized features such as flight distance, departure time, and		Demonstrated that the ensemble of RNNs and gradient boosting achieved superior prediction accuracy. Identified departure time, flight distance, and airline as significant factors affecting airfare prices.	Mean Absolute Percentage Error (MAPE) and Root Mean Squared Error (RMSE) were used to evaluate the performance of the prediction models.	

			airline information.				
Ref [13]	Machine Learning Algorithms - A Review Batta Mahesh	2020	Study of various machine learning algorithms and their use cases.		No models and techniques used as it is a review paper	-	-

1) **Datasets to be used in your Mini Project:**

Dataset 1:
Name: The dataset of Airfare prediction

Description: The dataset consists of information about various airplanes in the country. The data information is as follows

```
df.info()
```

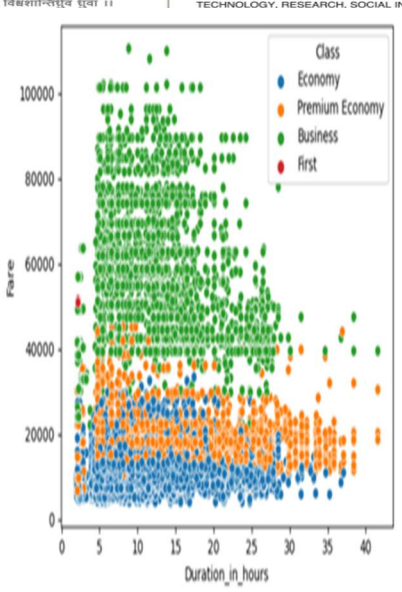
```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 405852 entries, 0 to 405851
Data columns (total 13 columns):
 #   Column              Non-Null Count   Dtype
---  ------              --------------   -----
 0   Date_of_journey     405852 non-null  object
 1   Journey_day         405852 non-null  object
 2   Airline             405852 non-null  object
 3   Flight_code         405852 non-null  object
 4   Class               405852 non-null  object
 5   Source              405852 non-null  object
 6   Departure           405852 non-null  object
 7   Total_stops         405852 non-null  object
 8   Arrival             405852 non-null  object
 9   Destination         405851 non-null  object
 10  Duration_in_hours   405851 non-null  float64
 11  Days_left           405851 non-null  float64
 12  Fare                405851 non-null  float64
dtypes: float64(3), object(10)
memory usage: 40.3+ MB
```

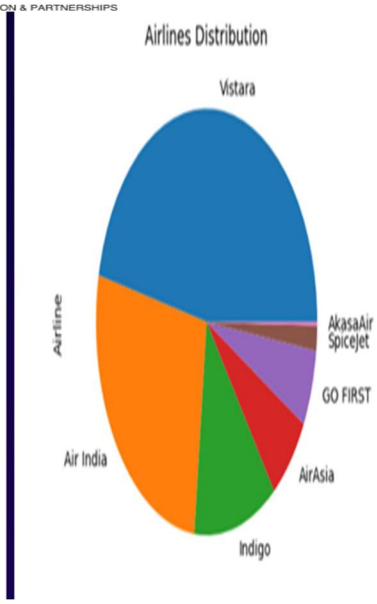
Screenshots of data:

	Date_of_journey	Journey_day	Airline	Flight_code	Class	Source	Departure	Total_stops	Arrival	Destination	Duration_in_hours	Days_left	Fare
0	2023-01-16	Monday	SpiceJet	SG-8169	Economy	Delhi	After 6 PM	non-stop	After 6 PM	Mumbai	2.0833	1.0	5335.0
1	2023-01-16	Monday	Indigo	6E-2519	Economy	Delhi	After 6 PM	non-stop	Before 6 AM	Mumbai	2.3333	1.0	5899.0
2	2023-01-16	Monday	GO FIRST	G8-354	Economy	Delhi	After 6 PM	non-stop	Before 6 AM	Mumbai	2.1667	1.0	5801.0
3	2023-01-16	Monday	SpiceJet	SG-8709	Economy	Delhi	After 6 PM	non-stop	After 6 PM	Mumbai	2.0833	1.0	5794.0
4	2023-01-16	Monday	Air India	AI-805	Economy	Delhi	After 6 PM	non-stop	After 6 PM	Mumbai	2.1667	1.0	5955.0
...	...	...	...	...	...	...	...	...	...	...	...	...	...
405847	2023-03-06	Monday	Indigo	6E-6815	Economy	Chennai	6 AM - 12 PM	1-stop	12 PM - 6 PM	Ahmedabad	5.7500	50.0	6733.0
405848	2023-03-06	Monday	Indigo	6E-605	Economy	Chennai	12 PM - 6 PM	1-stop	After 6 PM	Ahmedabad	6.0833	50.0	6733.0
405849	2023-03-06	Monday	Indigo	6E-6288	Economy	Chennai	6 AM - 12 PM	1-stop	12 PM - 6 PM	Ahmedabad	6.5833	50.0	6733.0
405850	2023-03-06	Monday	Indigo	6E-6892	Economy	Chennai	6 AM - 12 PM	1-stop	After 6 PM	Ahmedabad	7.6667	50.0	6733.0
405851	2023-03-06	Monday	Indigo	6E-847	Economy	Chennai	12 PM - 6 PM	1-stop	After	NaN	NaN	NaN	NaN

2) Performance metrics & Comparative Analysis with graphs/figures, observations/tables, etc

EDA plots:





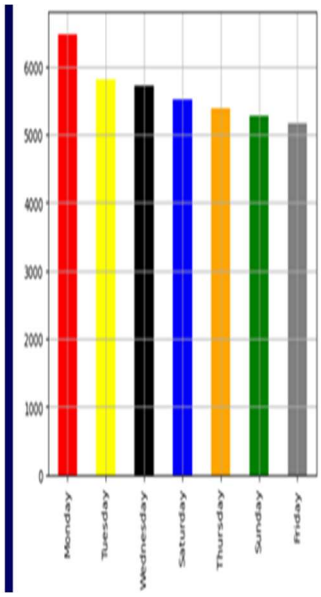
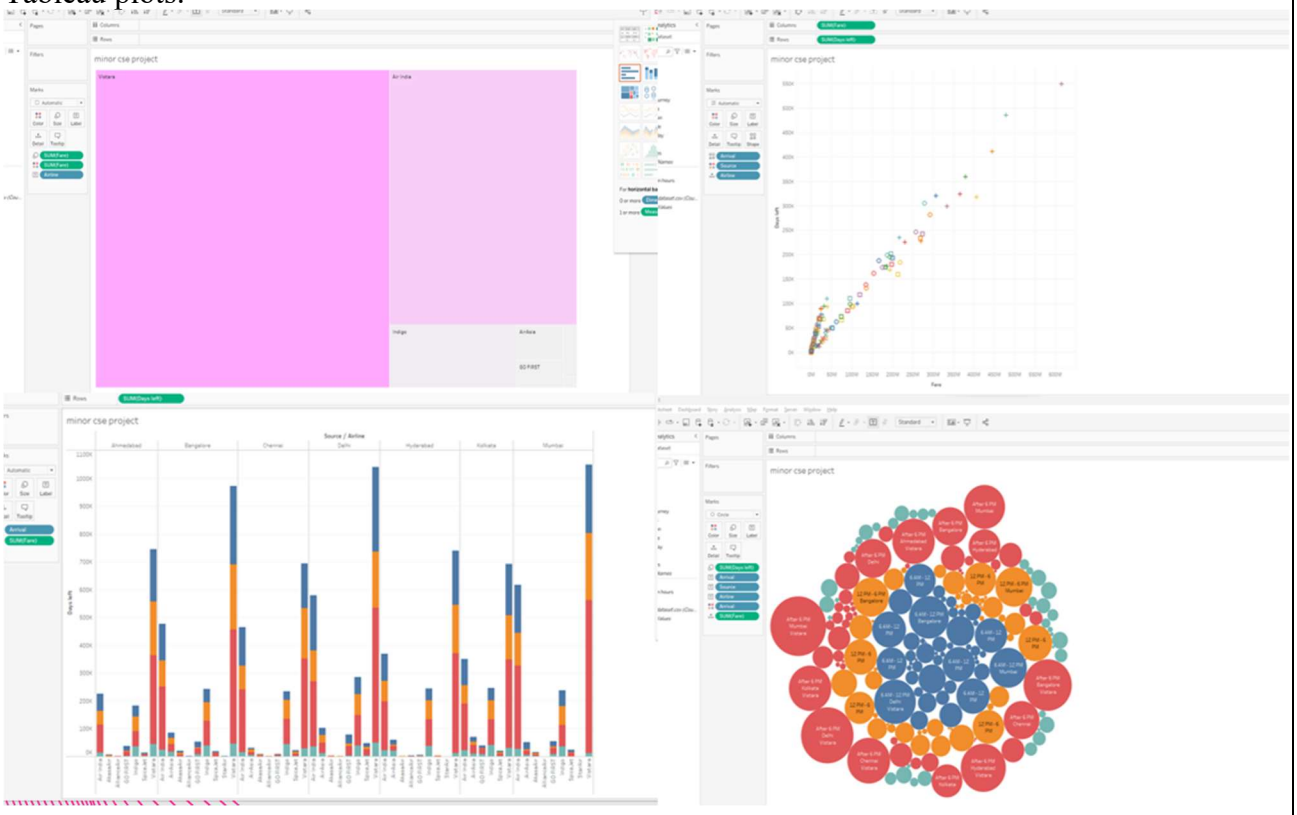
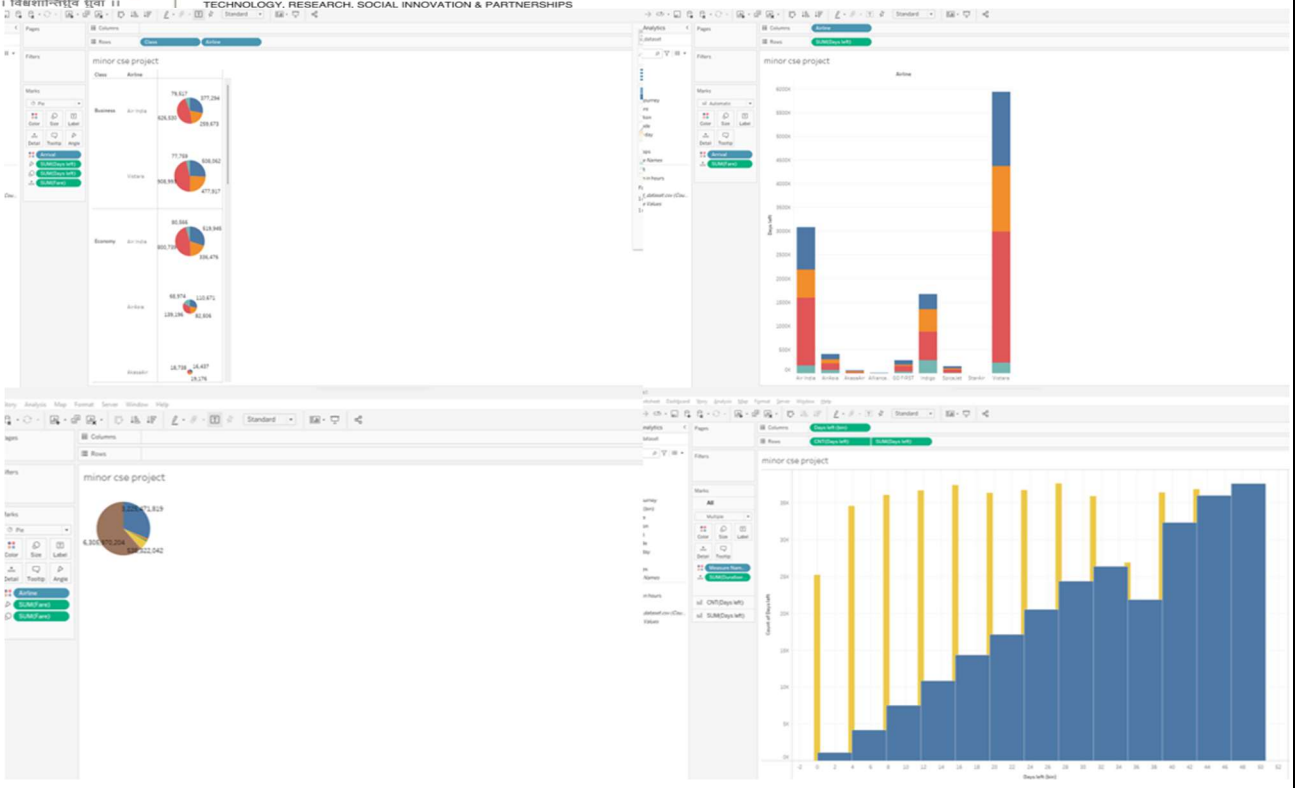


Tableau plots:





Machine Learning Analysis

	Naive Bayes	KNN	Logistic Regression	SVR	Lasso Regression	DBSCAN CLUSTERING	K-Means
Model	Naive Bayes	KNN	Logistic Regression	SVR	Lasso Regression	Density	K-Means
Precision	0.62	0.76	0.83	0.72	0.88	0.93	0.77
Scaling	Standard Data	Standard Data	Standard Data	Standard Data	Standard Data	Standard Data	Standard Data
Type	Gaussian	-	-	NaN	NaN	-	-

3) Individual contribution :

Sr.No	School Name (Major)	Name of Team member	Task done
1	School of Electrical	Chinmayi Railkar	1. Visualized the selected dataset 2. Implemented Kmeans and DBSCAN learning algorithm. 3. Created a precise presentation for the project.

2	School of Electrical	Devesh Khanna	1. Performed data pre processing and made it suitable for applying ML algorithms. 2. Implemented Lasso Regression & SVR 3. Tableau analysis of the selected data set.
3	School of Electrical	Sharvari Dhamale	1. Implemented Naive Bayes , KNN. 2. Comparative analysis of the algorithms implemented 3. Preparation of detailed project report.

4) Print of first page of Base paper

International Research Journal of Engineering and Technology (IRJET) e-ISSN: 2395-0056
Volume: 09 Issue: 05 | May 2022 www.irjet.net p-ISSN: 2395-0072

Airfare Price Prediction System

Rutuja Konde[1], Rutuja Somvanshi[2], Pratiksha Khaire[3], Prachi Zende[4], Kamlesh Patil[5]

[1234]*Student, Department of Information Technology, Bharati Vidyapeeth College of Engineering for Women, Pune, Maharastra, INDIA*
[5]*Professor, Dept. of Information Technology Engineering, Bharati Vidyapeeth College of Engineering for Women, Pune, Maharastra, INDIA*

Abstract - *Our actual idea about Flight prediction System is that we will going to predict flights prices with comparison of today to another day. The main purpose of out System is to predict the flight prices with comparison of today to another any day because of this customer can be book their tickets of Flight according to their comfortably, according to their affordability, Means whichever cheaper cost they want they can be easily choose .Before going to visit any place customer have to know what is actual prices for same destination which types of different Flights are available, if they don't know they will pay more money for tickets than the usual. This ticket prices based on various factors like winter, festival, summer, as well as number of tickets available for particular flights .Flight tickets may be vary during day and night. With consideration of some features like arrival time, departure time as well as time to purchase the ticket using these factors prices can be predict. due to this factors there may be change in airline fare prices and also detect how factors are related to being change of Flight ticket. By using the information above to build a system to predict the fare of the ticket priority using machine learning techniques like Random forest algorithm, that might be helpful for the passengers whether to buy a ticket or not. Nowadays, the airline corporations are using complicated strategies and strategies to assign airfare prices in a dynamic fashion. . It can be tough to wager the flight price price tag rate when we check it today compared to the other day. The tourists who want to visit a new place should know the fare price tag rate in order to get the cheapest and certain price tag rate with their needs. That's why we decided this project. In this project we will going to use machine Learning as back end. Flask as a python framework. Front end . Flask is*

complicated shape of state-of-the-art guidelines and mathematical fashions that pressure the pricing techniques of airfare . Although nevertheless in large part held in secret, research have located that those guidelines are widely recognized to be stricken by a whole lot of elements. Traditional variables consisting of distance, despite the fact that nevertheless gambling a extensive role, are now not the only thing that dictate the pricing approach. Elements associated with economic, advertising and social developments have performed growing roles in dictating the airfare charges. Nowadays, the airline groups are the usage of complicated techniques and strategies to assign airfare charges in a dynamic fashion. These techniques are taking into consideration numerous financial, advertising, business and social elements intently linked with the very last airfare charges. It may be tough to bet the airfare price tag fee whilst. We test it nowadays as compared to the alternative day. The vacationers who need to go to a brand new location need to realize the price tag fee to be able to get the most inexpensive and positive price tag fee with their needs. This whole thing brings the concept to make a prediction approximately the flight tickets to be able to make the vacationers simpler to book their tickets with their needs. Due to the excessive complexity of the pricing fashions implemented with the aid of using the airlines, it's miles very tough a client to buy an air price tag with inside the lowest fee, for the reason that fee modifications dynamically. For this cause a fixed of functions characterizing a normal flight is decided, supposing that those functions have an effect on the fee of an airfare price tag. Technology can convey an answer via the implementation of Machine studying strategies to enhance

5) Print of plagiarism report

References: (in IEEE format)

1. 1. Ayhan, S., Pesce, J., Comitz, P., Sweet, D., Bliesner, S. and Gerberick, G., 2013, April. Predictive analytics with aviation big data. In 2013 Integrated Communications, Navigation and Surveillance Conference (ICNS) (pp. 1-13). IEEE
2. 1. K. Tziridis T. Kalampokas G.Papakostas and K. Diamantaras "Airfare price prediction using machine learning techniques" in European Signal Processing Conference (EUSIPCO), DOI: 10.23919/EUSIPCO .2017.8081365L. Li Y. Chen and Z. Li" Yawning detection for monitoring driver fatigue based on two cameras" Proc. 12th Int. IEEE Conf. Intel. Transp. Syst. pp. 1-6 Oct. 2009.
3. 1. Supriya Rajankar, Neha sakhrakar and Omprakash rajankar "Flight fare prediction using machine learning algorithms" International journal of Engineering Research and Technology (IJERT) June 2019.
4. 1. Tianyi wang, samira Pouyanfar, haiman Tian and Yudong Tao "A Framework for airline price prediction: A machine learning approach"
5. 1. T. Janssen "A linear quantile mixed regression model for prediction of airline ticket prices"
6. "Exploratory Data Analysis of Airfare Dataset for Prediction using Machine Learning Techniques" by Gupta et al. (2018)
7. "Exploratory Data Analysis of Airfare Prediction Data using Big Data Analytics" by Das et al. (2019)
8. "Exploratory Data Analysis of Airfare Datasets: Insights and Trends" by Sharma et al. (2020)
9. "Exploratory Data Analysis of Domestic Airfare Dataset for Predictive Analytics" by Patel et al. (2021)
10. "Predicting Airfare Price Trends using Regression Analysis" by Patel et al. (2018)
11. "Airfare Price Prediction using Machine Learning and Big Data Analytics" by Gohil et al. (2019)
12. "Airfare Prediction using Deep Learning and Ensemble Methods" by Lee et al. (2020)
13. International Journal of Science and Research (IJSR) ISSN: 2319-7064 ResearchGate Impact Factor (2018): 0.28 | SJIF (2018): 7.426 Volume 9 Issue 1, January 2020 Machine Learning Algorithms - A Review Batta Mahesh
14. "Python Machine Learning" by Sebastian Raschka and Vahid Mirjalili: This book provides practical examples of using machine learning algorithms for data analysis, including exploratory data analysis. It covers topics like feature selection, dimensionality reduction, and clustering.

15. "Hands-On Machine Learning with Scikit-Learn and TensorFlow" by Aurélien Géron: This book demonstrates the application of machine learning algorithms for data analysis tasks, including EDA. It covers topics like data preprocessing, feature engineering, and model evaluation.
16. "Python Data Science Handbook" by Jake VanderPlas: This book explores various data analysis techniques using Python libraries, including machine learning approaches. It covers topics like data visualization, dimensionality reduction, and unsupervised learning for EDA.
17. "Applied Data Science" by Kelleher, Mac Namee, and D'Arcy: This book provides an overview of data science techniques, including the use of machine learning algorithms for EDA. It covers topics like clustering, association analysis, and anomaly detection.
18. "R for Data Science" by Hadley Wickham and Garrett Grolemund: This book focuses on using R for data analysis, including machine learning techniques for EDA. It covers topics like data wrangling, visualization, and modeling.
19. "Airfare Price Prediction Using Machine Learning Techniques and Big Data" by Aparna Kumari and Bhaskar Solanki
20. "Airfare Prediction using Big Data Techniques" by Sangeetha Vijayakumar and Dr. S. V. Kasmir Raja