DEVELOPING A FLIGHT DELAY PREDICTION MODEL USING MACHINE LEARNING

A Project Report

Submitted By

Team ID – PNT2022TMID45564

TEAM DETAILS

SARANYA.S(Team Leader) – Reg No : 812719106013
MANISHA.M(Team Member) – Reg No : 812719106008
KALAISELVI.P(Team Member) – Reg No : 812719106005
SNEHA.S(Team Member) - Reg No: 812719106015

TABLE OF CONTENTS:

1.INTRODUCTION:

1.1PROJECT OVERVIEW

Throughout the year 2015, there has been over 5,4 million domestic flights within the US. All of their metadata are recorded and saved in the Department of Transportation's (DOT) Bureau of Transportation Statistics.Flight delays cause significant financial and other losses to airlines, airports, and passengers. Their prediction is crucial during the decision-making process for all players of American aviation industry. Therefore, predicting the likelihood of delay based on flights' features bridges an important information asymmetry between airlines and passengers.The primary use case of the algorithm will be: predicting a potential delay, on a given day, for a given airport and airline.The dataset to be analysed consists of data about all domestic flights in the United States for the whole year 2015, for all airports and airlines. The data was collected and published by the DOT's Bureau of Transportation Statistics. The raw data file, with appropriate documentation, can be found on: https://www.kaggle.com/usdot/flight-delays .The whole set has over 5,400,000 examples. By default, every example has 31 features, although not all information is complete for all examples. It is nevertheless an extremely large and comprehensive dataset, which could allows for accurate statistical inferences.

For the prediction model building, we decided to use built-in functions within the sklearn kit.

The following table shows the types of algorithms we used, with their advantages and challenges with regard to our dataset:

Algorithm	Pros	Cons
Neural network (Multi-Layer Perceptron)	With a large amount of data and a limited number of features NN is likely to give good predictions.	Very slow, especially given that we're dealing with very large amounts of data.
Random Forest	Very slow, especially given that we're dealing with very large amounts of data.	Not as easy to visually interpret as decision trees.
Logistic Regression	Can provide probabilities, can be used with kernel methods,faster than neural network.	Some risk of underfit.

1.2 PURPOSE

Pauses the program for the amount of time (in milliseconds) specified as parameter. (There are 1000 milliseconds in a second.)Flight Tracking uses actual time global positioning data, **to help flight dispatchers and operators visualise their current flights.** Providing vital insights of flight status, on-time performance, environmental and weather factors, ensuring your airline is fully GADSS

.

2.LITERATURE SURVEY

2.1 EXISTING PROBLEM

While airlines want to get passengers to their destinations on time, there are many things that can – and sometimes do – make it difficult for flights to arrive on time. Some problems, like bad weather, air traffic delays, and mechanical issues, are hard to predict and often beyond the airlines' control.

[1] The main concern of the researchers and analysts is to predict the reasons for flight delays and for that they have put in their efforts on collecting data about flight and the weather. Mohamed et al.

[2] have studied the pattern of arrival delay for non-stop domestic flights at the Orlando International Airport. They focused primarily on the cyclic variations that happen in the air travel demand and theweather at that particular airport.In Shervin et al.'s work

[3], their motive of research is to propose an approach that improves
the operational performance without hampering or effecting the planned cost.Adrian et al.

[4] have created a data mining model which enables the flight delays by observing
the weather conditions. They have used WEKA and R to build their models by selecting different classifiers and choosing the one with the best results.Techniques like Naïve Bayes and Linear Discriminant Analysis classifier.Choi et al.

[5] have focused on overcoming the effects of the data imbalancing caused during
data training. They have used techniques like Decision Trees, AdaBoost, and K-Nearest Neighbors for predicting individual flight delays.A binary classification was performed by the model to predict the scheduled flight delay.Schaefer et al.

[6] have made Detailed Policy Assessment Tool (DPAT) that is used to stimulate
the minor changes in the flight delay caused by the weather changes.Bing Liu

[7] has done a sentiment analysis and opinion mining that analyzes people’s opinions, sentiments, and studies their behavior. The output of the research is a feature-based opinion summary.which is also known as sentiment classification.

Using techniques such as Natural Language Processing, Naïve Bayes, and Support Vector Machine, researchers built algorithms for analysis that helped them in extracting features in the model.Most of them focused on predicting overall flight delays. Our research concentrated mainly on predicting flight delays for a particular airport over a specific period of time. First, we used a regression model to examine the significance of each feature and then, a feature selection approach to examine the impact of feature combination. These two techniques determined the features to retain in the model. Instead of using the whole set, we sampled 5,000 records at a time to run through different machine learning models. The machine learning models implemented here were Random Forest classifier and Support Vector Machine (SVM) classifier. Further, we applied an approach 5 called One-Hot-Encoder to create a variant of the model for evaluating potential prediction performance.

2.2 REFERENCES

[1] A. B. Guy, "Flight delays cost $32.9 billion, passengers foot half the bill". [Online] Available https://news.berkeley.edu/2010/10/18/flight_delays/3/. [Accessed on June 2017].

[2] M. Abdel-Aty, C. Lee, Y. Bai, X. Li and M. Michalak, "Detecting periodic patterns of arrival delay",Journal of Air Transport Management,, Volume 13(6), pp. 355– 361, November, 2007.

[3] S. AhmadBeygi, A. Cohn and M. Lapp, "Decreasing Airline Delay Propagation By Re-Allocating Scheduled Slack", Annual Conference, Boston, 2008.

[4] A. A. Simmons, "Flight Delay Forecast due to Weather Using Data Mining", M.S. Disseration,University of the Basque Country, Department of Computer Science, 2015.

[5] S. Choi, Y. J. Kim, S. Briceno and D. Mavris, "Prediction of weatherinduced airline delays based on machine learning algorithms", Digital Avionics Systems Conference (DASC), 2016 IEEE/AIAA 35th,Sacramento, CA, USA, 2016.

[6] L. Schaefer and D. Millner, "Flight Delay Propagation Analysis With The Detailed Policy Assessment Tool", Man and Cybernetics Conference, Tucson, AZ, 2001.

[7] B. Liu "Sentiment Analysis and Opinion Mining Synthesis", Morgan & Claypool Publishers, p. 167,2012. 55

[8] Statistical Computing Statistical Graphics. [Online]. Available: http://statcomputing.org/dataexpo/2009/the-data.html. [Accessed on April 2017].

[9] FAA Operations & Performance Data. [Online]. Available: https://aspm.faa.gov/.[Accessed on April 2017].

[10] B. Bailey, "Data Cleaning 101". [Online]. Available: https://towardsdatascience.com/data-cleaning-101-948d22a92e4. [Accessed on March 2018].

[11] P. Panov, L. Soldatova and S. Džeroski, " OntoDM-KDD: Ontology for Representing the Knowledge Discovery Process", Discovery Science 2013, Volume 8140, pp. 126-140, 2013.

[12]Bureau of Transportation Statistics. [Online]. Available: https://www.transtats.bts.gov/carriers.asp. [Accessed on 2 April 2017].

[13] How to Predict Yes/No Outcomes Using Logistic Regression. [Online]. Available: https://blog.cleaarbrain.com/posts/how-to-predict-yesno-outcomesusing-logisticregression [Accessed on 3 Feubrary 2018].

[14] S. Polamuri, "How The Random Forest Algorithm Works In Machine Learning". [Online]. Available:https://medium.com/@Synced/how-randomforest-algorithmworks-in-machine learning-3c0fe15b6674.[Accessed January 2018].

[15] S. Ray, "Understanding Support Vector Machine algorithm". [Online]. Available: https://www.analyticsvidhya.com/blog/2017/09/understaing-supportvectormachine examplecode/.[Accessed November 2017].

[16] OneHotEncoder. [Online]. Available: http://scikitlearn.org/stable/modules/generated/sklearn.preprocessing.OneHotEn coder.html. [Accessed on March 2018]. [17] R. Vasudev, "Why and When do you have to use OneHotEncoder?".[Online]. Available:https://hackernoon.com/what-is-onehot-encoding-why-and-when-doyou-have-to-use-it-e3c6186d008f.[Accessed on March 2018].

[18] Twitter API Twitter. [Online].Available https://developer.twitter.com/en/docs.

[19] S. Loria , "TextBlob: Simplified Text Processing", 2016. [Online]. Available:http://textblob.readthedocs.io/en/dev/ [Accessed on December 12, 2017].

[20] A. Agarwal, B. Xie, I. Vovsha, O. Rambow and R. Passonneau, "Sentiment Analysis of Twitter Data,"Columbia University, New York, December, 2011.

[21] V. A. Kharde and S. Sonawane, "Sentiment Analysis of Twitter Data: A Survey of Techniques", International Journal of Computer Applications (0975 – 8887), Volume 139, no.11, p.11, April 2016.

2.3 PROBLEM STATEMENT DEFINITION:

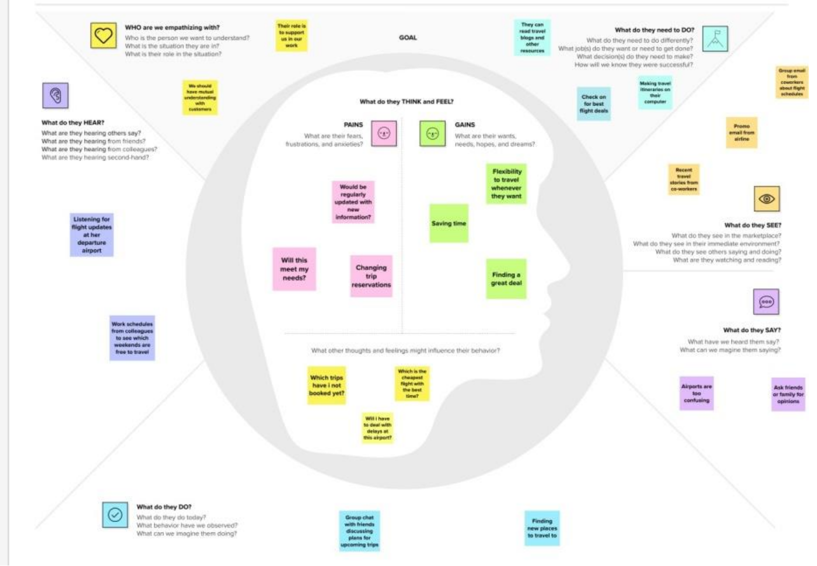
Our case study was about LaGuardia airport in Newyork,Logan international Airport in Bostan,San Fransisco Interational Airport in San Fransisco and O'Hare Interational Airport in Chicago,which are four major airports in the United States of America.But we focused the idea and Research and on LaGuardia International Airport.Compared with the data produced by all Airports in USA,the data which we gathered was very limited,but it gave as a great direction on how weather plays a part in flight delays.

3.IDEATION & PROPOSED SOLUTION

3.1EMPATHY MAP CANVAS

An empathy map is a simple, easy-to-digest visual that captures knowledge about a user's behaviours and attitudes.

It is a useful tool to helps teams better understand their users.
Creating an effective solution requires understanding the true problem and the person who is experiencing it. The exercise of creating the map helps participants consider things from the user's perspective along with his or her goals and challenges.

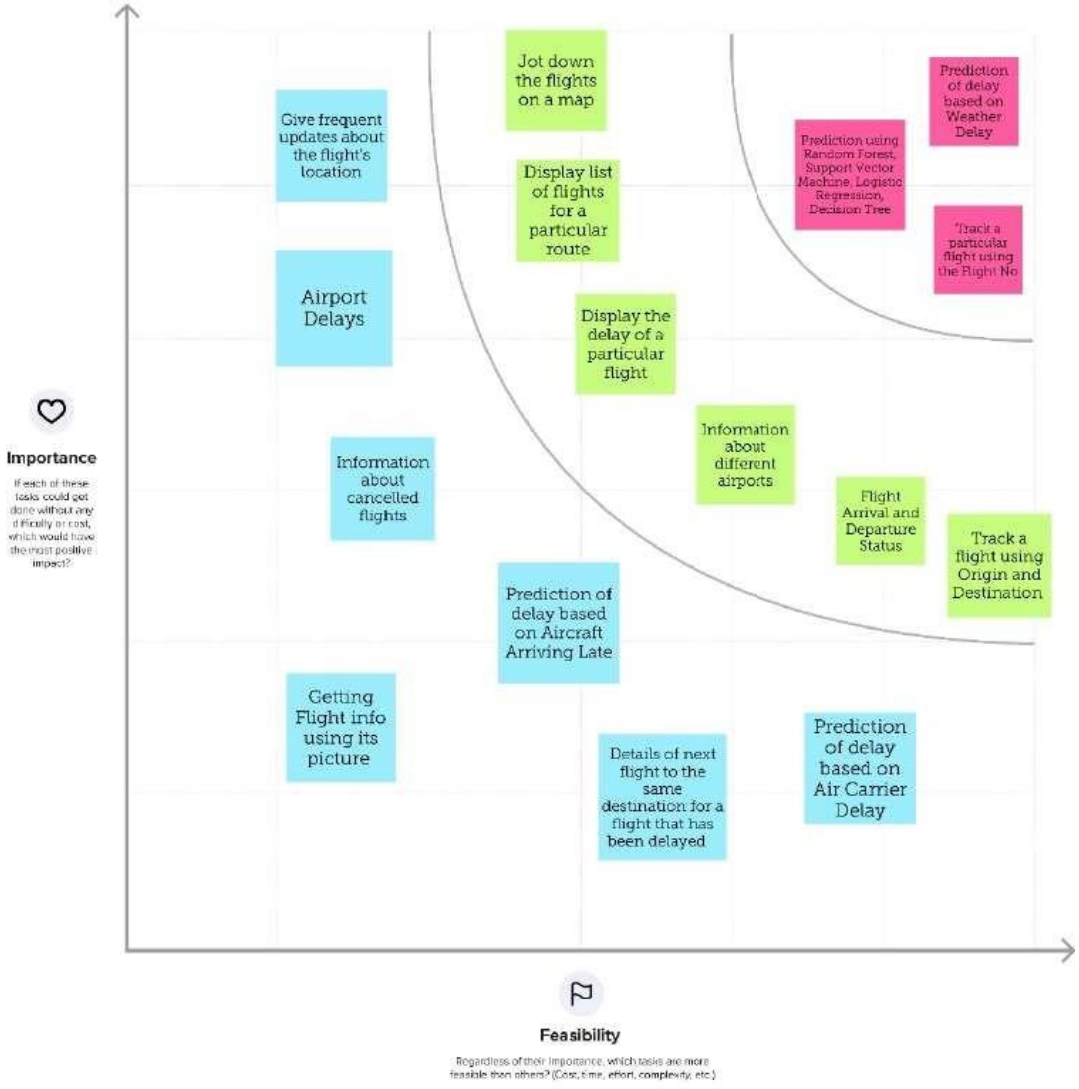


WHO are we empathizing with?
Who is the person we want to understand?
What is the situation they are in?
What is their role in the situation?
Their role is to support us in our work
GOAL
They can read travel blogs and other resources
What do they need to DO?
What do they need to do differently?
What job(s) do they want or need to get done?
What decision(s) do they need to make?
How will we know they were successful?
We should have mutual understanding with customers
Check on for best flight deals
Making travel itineraries on their computer
Group email from coworkers about flight schedules
What do they HEAR?
What are they hearing others say?
What are they hearing from friends?
What are they hearing from colleagues?
What are they hearing second-hand?
What do they THINK and FEEL?
PAINS
What are their fears, frustrations, and anxieties?
GAINS
What are their wants, needs, hopes, and dreams?
Promo email from airline
Would be regularly updated with new information?
Flexibility to travel whenever they want
Recent travel stories from co-workers
Listening for flight updates at her departure airport
Saving time
What do they SEE?
What do they see in the marketplace?
What do they see in their immediate environment?
What do they see others saying and doing?
What are they watching and reading?
Will this meet my needs?
Changing trip reservations
Finding a great deal
What do they SAY?
What have we heard them say?
What can we magine them saying?
Work schedules from colleagues to see which weekends are free to travel
What other thoughts and feelings might influence their behavior?
Which trips have i not booked yet?
Which is the cheapest flight with the best time?
Will i have to deal with delays at this airport?
Airports are too confusing
Ask friends or family for opinions
What do they DO?
What do they do today?
What behavior have we observed?
What can we imagine them doing?
Group chat with friends discussing plans for upcoming trips
Finding new places to travel to

3.2 IDEATION

Developing a Flight
Delay Prediction
Model using Machine
Learning

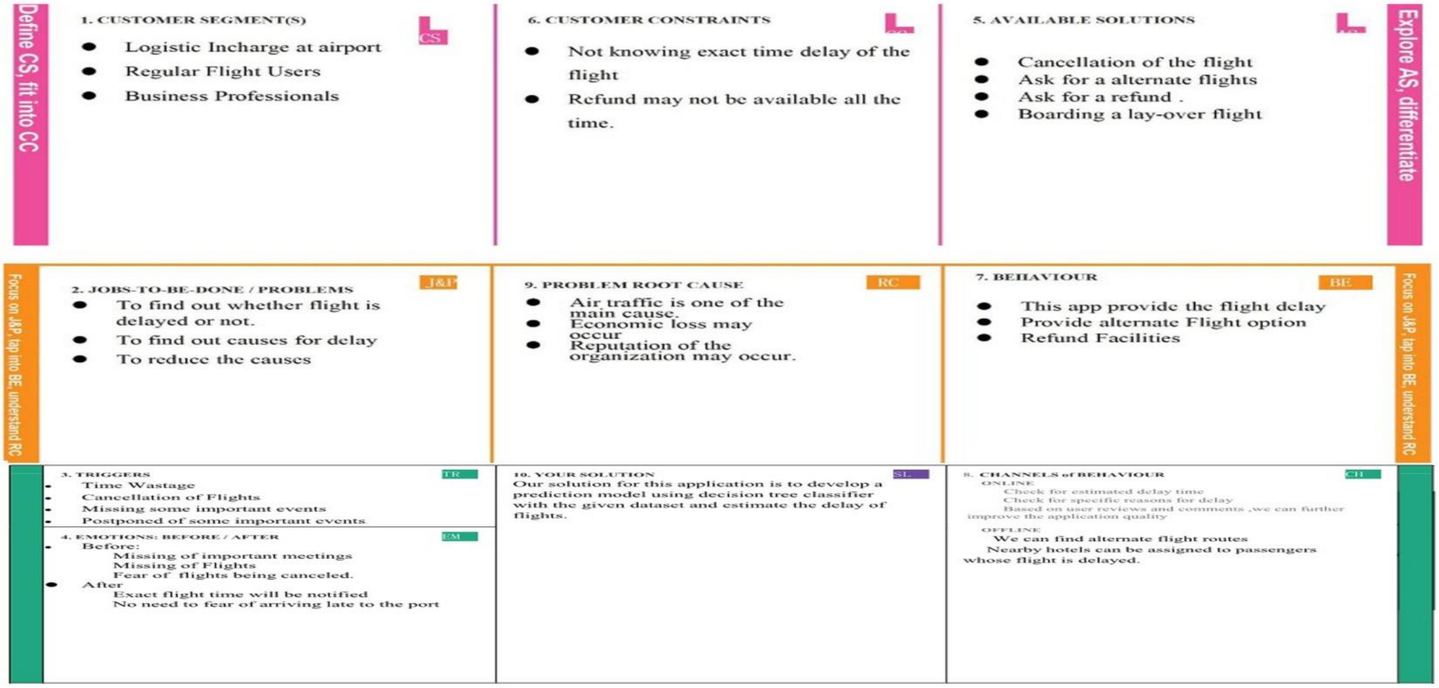
IDEATION :


Give frequent updates about the flight's location
Jot down the flights on a map
Prediction of delay based on Weather Delay
Prediction using Random Forest, Support Vector Machine, Logistic Regression, Decision Tree
Display list of flights for a particular route
Track a particular flight using the Flight No
Airport Delays
Display the delay of a particular flight
Importance
If each of these tasks could get done without any difficulty or cost, which would have the most positive impact?
Information about cancelled flights
Information about different airports
Flight Arrival and Departure Status
Track a flight using Origin and Destination
Prediction of delay based on Aircraft Arriving Late
Getting Flight info using its picture
Details of next flight to the same destination for a flight that has been delayed
Prediction of delay based on Air Carrier Delay
Feasibility
Regardless of their importance, which tasks are more feasible than others? (Cost, time, effort, complexity, etc.)

3.3 PROPOSED SOLUTION

S.No.	Parameter	Description
1.	Problem Statement (Problem to be solved)	Flight delays have been the most challenging area for airlines to improve. ● They have been affecting the air industry directly and indirectly causing unforeseen expenses thereby reducing the reputation of the industry and the airlines. ● Thus, knowing if a flight would be delayed beforehand can let passengers and airlines be prepared for the circumstances. ● This solution aims at making it possible by predicting arrival and departure delays using Machine learning
2.	Idea / Solution description	Building an application interface for customers(passengers and airlines) to know if a flight is delayed by implementing a machine learning based model to predict departure and arrival delays of an aircraft considering spatial, temporal and other dependencies causing the delay.
3.	Novelty / Uniqueness	● The solution takes into account all possible reasons for delay(crew delays, weather, air traffic, aircraft type) to provide an accurate prediction. ● Apart from predicting arrival delays, departure delays are also predicted in order forthe passengers to prepare accordingly and for the airline to make arrangements suitably.
4.	Social Impact / Customer Satisfaction	A lot of time and money can be saved for the customers and the loyalty and trust of customers towards the company increases.

		Improves airline operations by letting the company prepare in prior to adversaries (like crew illness, timeouts, rescheduling) leading to passenger satisfaction which will result positively on the economy and brand value.
5.	Business Model (Revenue Model)	Business to Consumer model ● The solution is a low-cost airline model planned to be created as an application withwhich the consumers can interact directly toknow the details of their flight. ● It follows a non-monetary revenue model where the consumers aren’t charged for whatthey get but are asked to provide their flight details and ratings which can be used to improve the model and shared with the airline in return for airline’s flight data
6.	Scalability of the Solution	● The present solution is drafted with the aim of experimenting with airlines based out of theUnited States of America. ● If there is a possibility to acquire data of a broader region (say North America, other continents), then the solution can be developed to benefit a wider range of people. ● International flight dependencies in both temporal and spatial focus can be derived from that data to provide more accurate predictions. ● Presence of ADS-B data can further increase the efficiency of system making it reach globalaudience and live time tracking of flights.


Define CS, fit into CC
CS
Explore AS, differentiate
Focus on J&P, tap into BE, understand RC
J&P
RC
BE
TR
EM
SL
CH

1. CUSTOMER SEGMENT(S)

- Logistic Incharge at airport
- Regular Flight Users
- Business Professionals

6. CUSTOMER CONSTRAINTS

- Not knowing exact time delay of the flight
- Refund may not be available all the time.

5. AVAILABLE SOLUTIONS

- Cancellation of the flight
- Ask for a alternate flights
- Ask for a refund .
- Boarding a lay-over flight

2. JOBS-TO-BE-DONE / PROBLEMS

- To find out whether flight is delayed or not.
- To find out causes for delay
- To reduce the causes

9. PROBLEM ROOT CAUSE

- Air traffic is one of the main cause.
- Economic loss may occur
- Reputation of the organization may occur.

7. BEHAVIOUR

- This app provide the flight delay
- Provide alternate Flight option
- Refund Facilities

3. TRIGGERS

- Time Wastage
- Cancellation of Flights
- Missing some important events
- Postponed of some important events

4. EMOTIONS: BEFORE / AFTER

- Before:
  Missing of important meetings
  Missing of Flights
  Fear of flights being canceled.
- After
  Exact flight time will be notified
  No need to fear of arriving late to the port

10. YOUR SOLUTION

Our solution for this application is to develop a prediction model using decision tree classifier with the given dataset and estimate the delay of flights.

8. CHANNELS of BEHAVIOUR

ONLINE

Check for estimated delay time
Check for specific reasons for delay
Based on user reviews and comments ,we can further improve the application quality

OFFLINE

We can find alternate flight routes
Nearby hotels can be assigned to passengers whose flight is delayed.

4.REQUIREMENT ANALYSIS

4.1FUNCTIONAL REQUIREMENTS

Following are the functional requirements of the proposed solution.

FR No.	Functional Requirement (Epic)	Sub Requirement (Story / Sub-Task)
FR-1	User Registration	Registration through Form Registration through Gmail Registration through LinkedIN
FR-2	User Confirmation	Confirmation via Email Confirmation via OTP
FR-3	Login	Log in using given credentials
FR-4	Delay Prediction	Requesting for details from user for making the predictions about their flight
FR-5	Feedback	Get feedback from customer about their experience
FR-6	Logout	Logout from the session

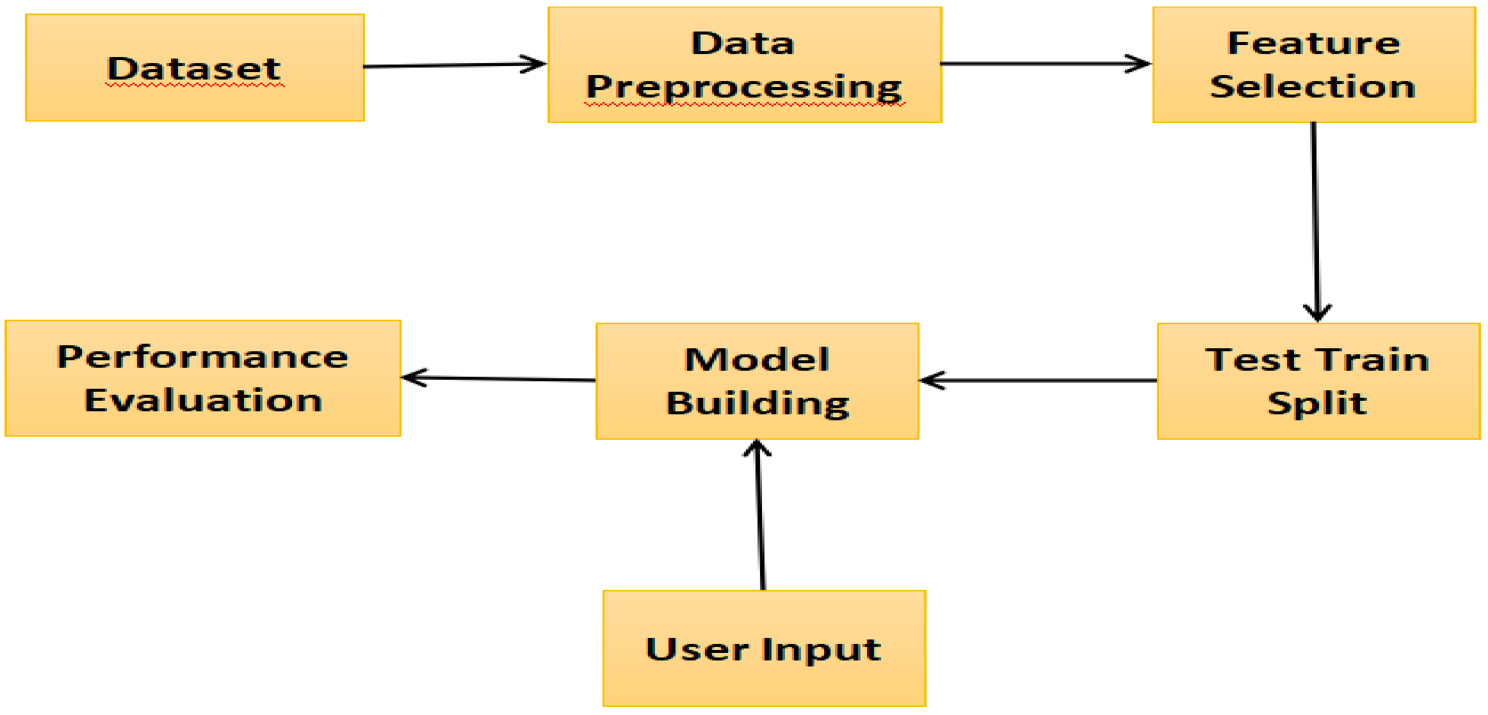
4.2 NON-FUNCTIONAL REQUIREMENTS

Following are the non-functional requirements of the proposed solution.

FR No.	Non-Functional Requirement	Description
NFR-1	**Usability**	The app must have such ease in usability. The UI should be simple and clear for the user tounderstand
NFR-2	**Security**	A 2 step verification should be implemented in theapplication All the confidential information about the flight andcustomer must be confidential
NFR-3	**Reliability**	The application must be able to work under any situation. It should be able to restore all its content even if thesystem fails
NFR-4	**Performance**	The application should perform in a faster way It shouldn't take much time for prediction of delayThe time should be less than 3 seconds.
NFR-5	**Availability**	The application should be available 24/7 . Even if any bug fix or any update happens it should perform normally and let the update or fixes happenin parallel. The customer should be able to get the prediction result even if the application is under update or anybug fix.
NFR-6	**Scalability**	The application should be capable of handling multiple requests from the user as millions of users will be using it at the same time.

5.PROJECT DESIGN

5.1DATA FLOW DIAGRAMS



5.2 USER STORIES

Use the below template to list all the user stories for the product.

User Type	Functional Requirement (Epic)	User Story Number	User Story / Task	Acceptance criteria	Priority	Release
Customer (Mobile user)	Registration	USN-1	As a user, canregisterfor theapplication by entering my email,password, and confirmingmy password.	I can access my account / dashboard	high	Sprint-1
		USN-2	As a user, I will receive confirmation email once I have registered for the application	I can receive confirmation email & click confirm	high	Sprint-1
		USN-3	As a user, I can register for the applicationthrough Facebook	I can register & access the dashboard with Facebook Login	low	Sprint-2
		USN-4	As a user, I can register for the application through Gmail	I can register successfully with gmail login	medium	Sprint-1
	Login	USN-5	As a user, I can log into the application byentering email & password	I can login using my emai land password	high	Sprint-1
	Dashboard	USN-6	As a user, I can use the dashboard to enter the details of my	I can enter the details	high	Sprint-2

			flight like arrival time, flight id,destination,source place,etc…	intothe textfield		
Customer(web user)	**Predict delay**	USN-7	As a user , I can use this button to calculate or predict the delay of my flight	I can push the button to get the result	**high**	Sprint-2
Customer care executive	**Feedback section**	USN-8	As a user , I can give my feedback in this section	I am able to give my feedback.	**medium**	Sprint-3
Administrator	**Maintenance**	USN-9	As an administrator,I can be able to monitor allthe tasks going on and can modify any changes	I can have the control toall section to modify anything.	**medium**	Sprint-4

6.PROJECT PLANNING & SCHEDULING

6.1 SPRINT PLANNING & ESTIMATION

S.NO	MILESTONE	DESCRIPTION	DURATION	WORKINGSTATUS
1	Prerequisites	Prerequisites are all the needs atthe requirementlevel needed for the execution of the different phases of a project.	1 WEEK	Completed
2	Data collection	IBM provides the dataset. It is the actual datasetused to train the model for performing various actions. In this activitylet’s focus on gathering the dataset	1 WEEK	Completed
3	Data Pre-processing	Data Pre processing is a technique that is used toconvertthe raw data into a clean data set.	3 WEEKS	Completed
4	Ideation phase	Ideation is the process whereyou generate ideasand solutions through sessions such as Sketching, Prototyping, Brainstorming,Brain writing, Worst PossibleIdea, and a wealth of other ideation techniques.	1 WEEK	Completed

5	Project design phases	Project design is an early phase of a project where the project's key features,structure, criteriafor success,and major deliverables are planned out. The aim is todevelop one or more designs that can be used to achieve the desired project goals.	1 WEEK	Completed
6	Train the model on IBM	Model training is the primarystep in machine learning, resulting in a working model that canthen be validated, tested and deployed.	1 WEEK	In process
7	Project planning phase	In the Planning Phase, theProject Manager workswith the project team to create the technical design, task list, resource plan, communications plan, budget, and initial schedule for the project, andestablishes the roles and responsibilities of the project team and its stakeholders.	1 WEEK	In process
8	Project Development Phase	Project development is theprocess of planningand allocating resources to fully develop a project or product from concept to go-live.	1 WEEK	In process
9	Application Building	A web application (or web app)is application software that runsin a web browser unlike software programs that runlocally and nativelyon the operating system (OS) of the device.	1 WEEK	In process

6.2 SPRINT DELIVERY SCHEDULE

Use the below template to create product backlog and sprint schedule.

Sprint	Functional Requirement (Epic)	User Story Number	User Story / Task	Story Points	Priority	Team Members
Sprint-1	Registration	USN-1	As a user , I can register for the application by enterin	2	High	

			g my email, password, and confirming my password.			
Sprint-1		USN-2	As a user, I will receive confirmation email once I have registered for the application	1	High	
Sprint-2		USN-3	As a user, I can register for the application through Facebook	2		
Sprint-1		USN-4	As a user, I can register for the application through Gmail	2	Medium	
Sprint-1	Login	USN-5	As a user, I can log into the application by entering email & passwor	1	High	

			d			
	Dashboard					

Project Tracker, Velocity & Burndown Chart: (4 Marks)

Sprint	Total Story Points	Duration	Sprint Start Date	Sprint End Date (Planned)	Story Points Completed (as on Planned End Date)	Sprint Release Date (Actual)
Sprint-1	20	6 Days	24 Oct 2022	29 Oct 2022	20	29 Oct 2022
Sprint-2	20	6 Days	31 Oct 2022	05 Nov 2022		
Sprint-3	20	6 Days	07 Nov 2022	12 Nov 2022		
Sprint-4	20	6 Days	14 Nov 2022	19 Nov 2022		

Velocity:

Imagine we have a 10-day sprint duration, and the velocity of the team is 20 (points per sprint). Let's calculate the team's average velocity (AV) per iteration unit (story points per day)

$$AV = \frac{sprint\ duration}{velocity} = \frac{20}{10} = 2$$

Burndown Chart:

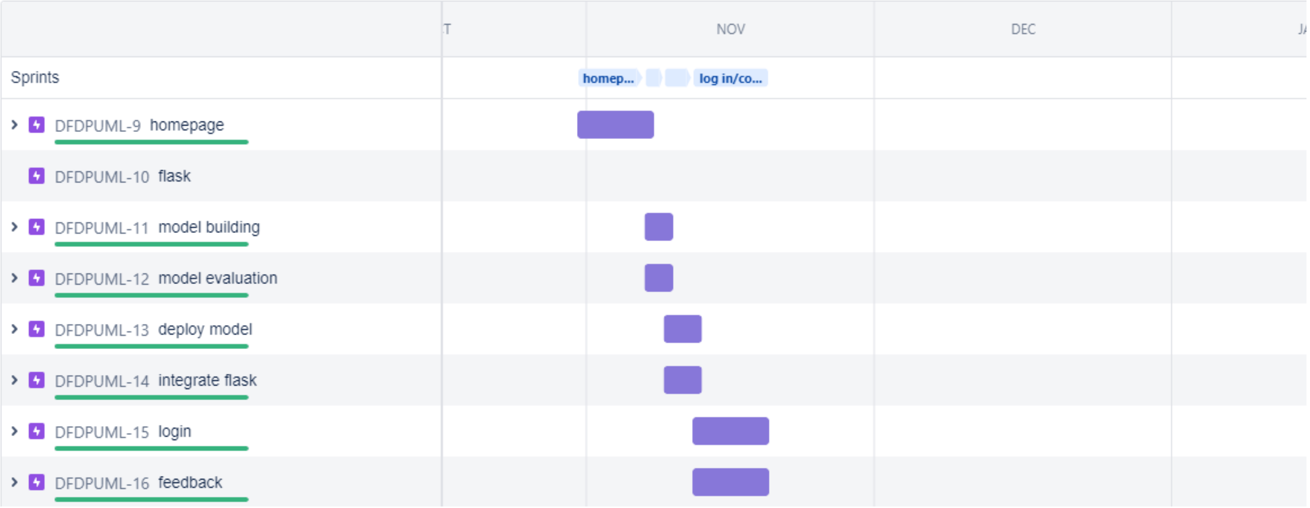
A burn down chart is a graphical representation of work left to do versus time. It is often used in agile software development methodologies such as Scrum. However, burn down charts can be applied to any project containing measurable progress over time.

https://www.visual-paradigm.com/scrum/scrum-burndown-chart/
https://www.atlassian.com/agile/tutorials/burndown-charts

Reference:
https://www.atlassian.com/agile/project-management
https://www.atlassian.com/agile/tutorials/how-to-do-scrum-with-jira-software
https://www.atlassian.com/agile/tutorials/epics
https://www.atlassian.com/agile/tutorials/sprints
https://www.atlassian.com/agile/project-management/estimation
https://www.atlassian.com/agile/tutorials/burndown-charts

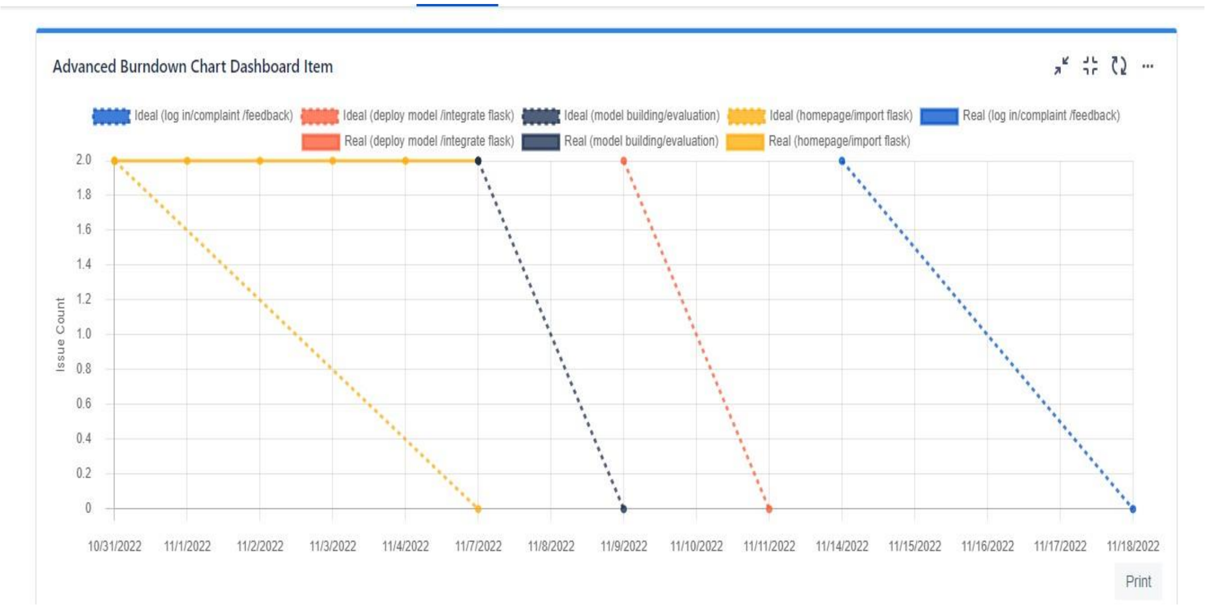
6.3 REPORTS FROM JIRA

ROADMAP:



	T	NOV	DEC	JA
Sprints		homep... log in/co...		
DFDPUML-9 homepage				
DFDPUML-10 flask				
DFDPUML-11 model building				
DFDPUML-12 model evaluation				
DFDPUML-13 deploy model				
DFDPUML-14 integrate flask				
DFDPUML-15 login				
DFDPUML-16 feedback				

BURNDOWN GRAPH:



7.CODING & SOLUTIONING

Coding:

We use the jupyter notebook and to insert the datas to coding the flight delay prediction.

Solutioning:

Using airport operators and airlines' existing security and observation cameras, IntellAct's artificial intelligence (AI) software system can automatically detect delays in these turnaround services. It can then highlight the problem to airport staff or ground crew and suggest a mitigation plan in real time.

Among these are: required diversion of some traffic to reliever airports, more balanced use of metropolitan air carrier airports, restriction of airport access by aircraft type or use, and establishment of quotas (either on the num- ber of operations or on passenger enplanements)

7.1 FEATURES

The results show that **adverse weather conditions, low ceilings, and low visibility conditions** strongly influence flight delays. Similarly, Asfe et al. [2] investigated the major causal factors of flight delays by ranking different factors using the analytical hierarchical process.

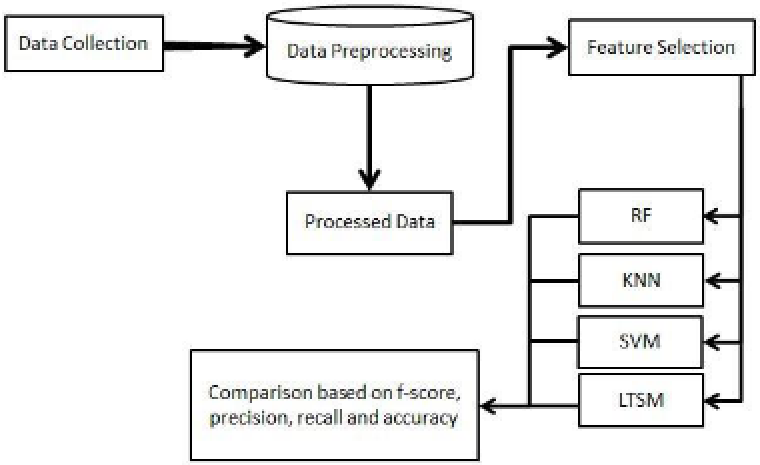
There are no federal laws requiring airlines to provide passengers with money or other compensation when their flights are delayed. Each airline has its own policies about what it will do for delayed passengers.01-Sept-2022

You're legally entitled to get compensation if the cancellation is the airline's responsibility and both the following apply: the replacement flight delays your arrival by **2 or more hours**. your flight was cancelled less than 14 days before departure.

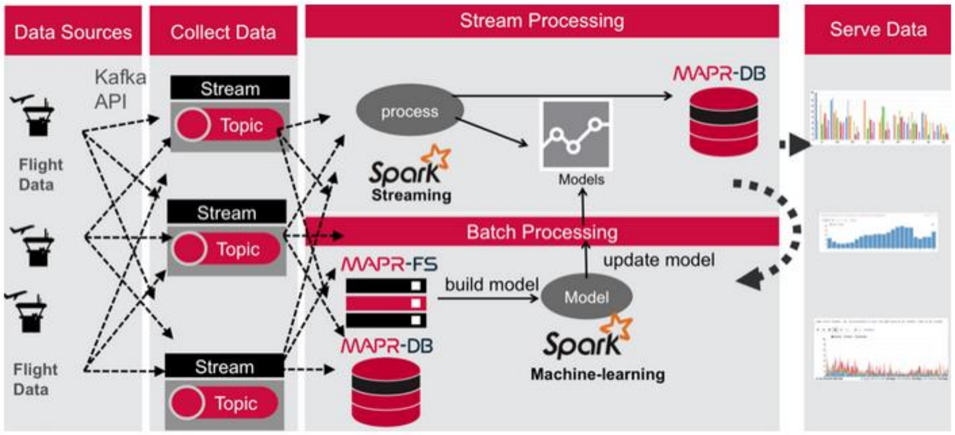
The flight delay Prediction is very useful to people to find the any issues to book or cancellation process.

If an airline consistently reschedules passengers due to overbooking flights and does not compensate them for the delay, **passengers may be able to bring a class action lawsuit against the airline**.

7.2 FEATURE SELECTION

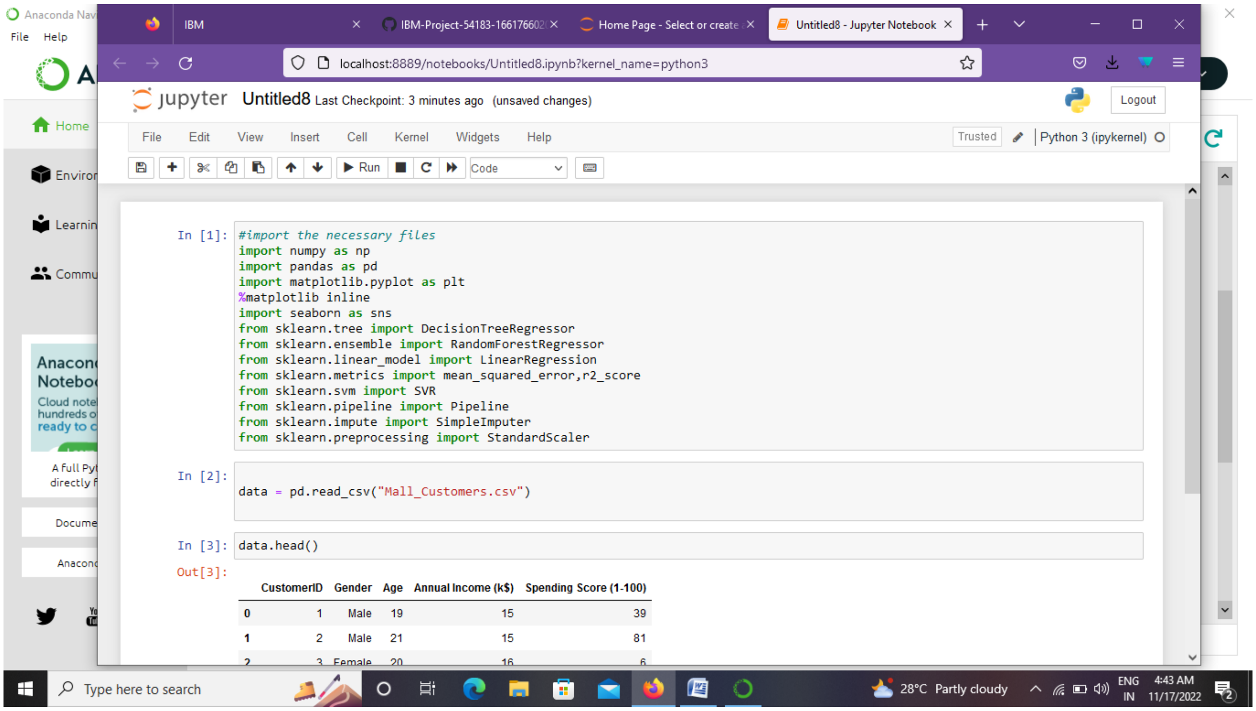


7.3 DATABASE



8.TESTING

8.1 TEST CASES


IBM
IBM-Project-54183-166176602
Home Page - Select or create
Untitled8 - Jupyter Notebook
localhost:8889/notebooks/Untitled8.ipynb?kernel_name=python3
jupyter Untitled8 Last Checkpoint: 3 minutes ago (unsaved changes)
Logout
File Edit View Insert Cell Kernel Widgets Help
Trusted
Python 3 (ipykernel)
Run
Code
Type here to search
28°C Partly cloudy
ENG IN
4:43 AM 11/17/2022

```
In [1]: #import the necessary files
        import numpy as np
        import pandas as pd
        import matplotlib.pyplot as plt
        %matplotlib inline
        import seaborn as sns
        from sklearn.tree import DecisionTreeRegressor
        from sklearn.ensemble import RandomForestRegressor
        from sklearn.linear_model import LinearRegression
        from sklearn.metrics import mean_squared_error,r2_score
        from sklearn.svm import SVR
        from sklearn.pipeline import Pipeline
        from sklearn.impute import SimpleImputer
        from sklearn.preprocessing import StandardScaler

In [2]: 
        data = pd.read_csv("Mall_Customers.csv")

In [3]: data.head()
Out[3]:
           CustomerID  Gender  Age  Annual Income (k$)  Spending Score (1-100)
        0           1    Male   19                  15                      39
        1           2    Male   21                  15                      81
        2           3  Female   20                  16                       6
```

8.2 USER ACCEPTANCE TESTING

1. Purpose of Document

The purpose of this document is to briefly explain the flight delay prediction and open issues of the Predicting the delay output of flights based on weather condition project at thetime of the release to User Acceptance Testing (UAT).

2. Defect Analysis

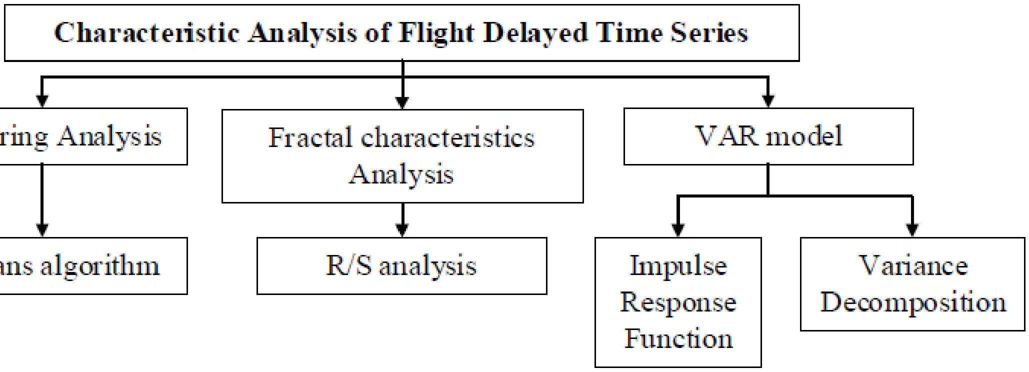
This report shows the delay prediction level and howthey were resolved and have Features.

The contributions are as follows.

i. The K-means algorithm is used to cluster the sequences. Based on the clustering center, the series I are divided into five categories, and the delay degree of each category is analyzed respectively.
ii. The fractal characteristics of time series are explored by using the fractal theory.
iii. Vector Auto Regression (VAR) model is designed. The Impulse Response Function (IRF) is used to reflect the dynamic influence of each delay rate time series. Then, Variance Decomposition is used to explore the specific sources of variance generation.

The rest of this article is organized as follows.

The section II is related works, mainly analyzing the research results in this field. Section III is the characteristics analysis of flight delay time series based on fractal theory. Section V is experiment and results analysis, including three test contents and comparative analysis. Section IV concludes this paper with research work summary.


Characteristic Analysis of Flight Delayed Time Series
Clustering Analysis
Fractal characteristics Analysis
VAR model
K-means algorithm
R/S analysis
Impulse Response Function
Variance Decomposition

3.Test case Analysis

This report shows the number of test cases that have passed, failed, and untested.

Section	Total cases	Not Tested	Fail	Pass
Index	1	0	0	1
Predict	1	0	0	1
Security	2	0	0	2
Weather Details	20	0	0	20
Exception Reporting	9	0	0	9
Flight Delay Output	4	0	0	4

9.RESULTS

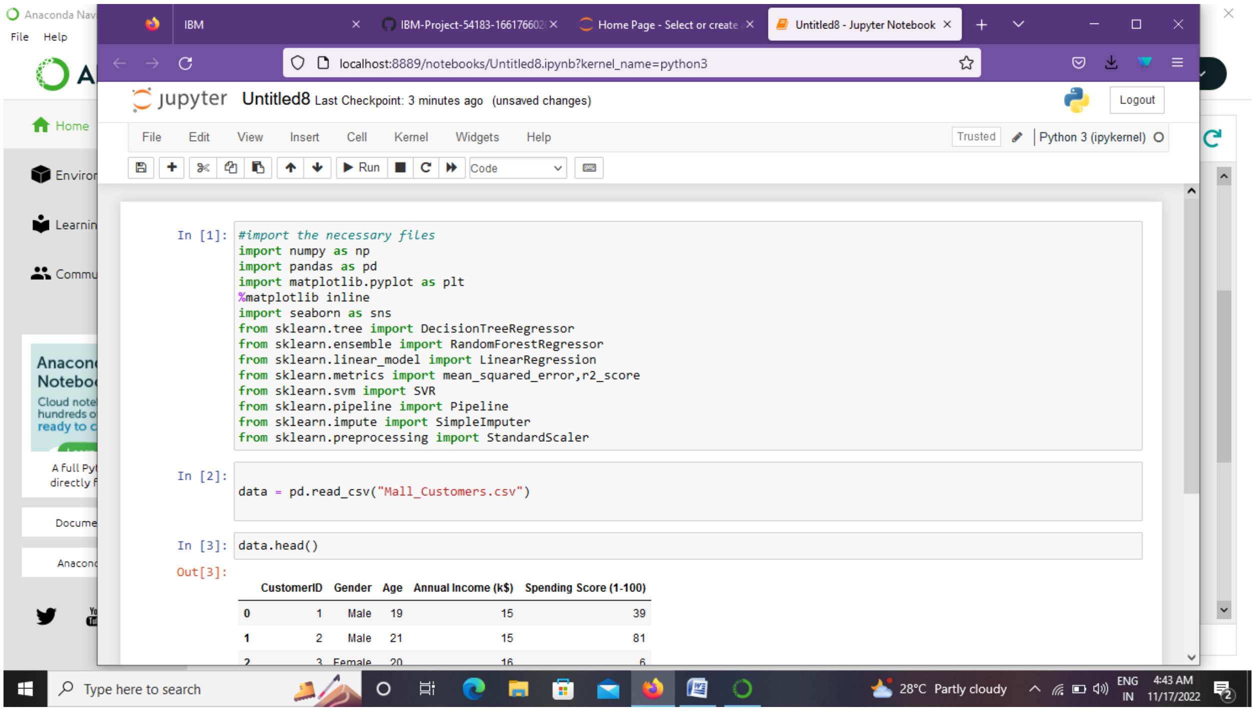
On Time Performance of an airline network is an important metric for the airline. It is calculated as **the percentage of flights which are delayed by more than 14 minutes while the aircraft arrives at the gate**

A flight is counted as "on time" **if it operated less than 15 minutes later than the scheduled time shown in the carriers' Computerized Reservations Systems (CRS)**. Arrival performance is based on arrival at the gate. Departure performance is based on departure from the gate

OTP is a valuable metric for airlines and is **an outward demonstration of reliability which can affect brand loyalty and ticket sales**. Customer satisfaction is influenced by customer expectations and a flight that arrives after the scheduled arrival time can be a stressful experience for passengers.

Used to describe **the ability of an aircraft to accomplish certain things that make it useful for certain purposes**. For example, the ability of an aircraft to land and take off in a very short distance is an important factor to the pilot who operates in and out of short, unimproved airfields.

9.1 PERFORMANCE METRICS


Anaconda Nav
File Help
Home
Enviror
Learnin
Commu
Anacon
Notebo
Cloud note
hundreds o
ready to c
A Full Py
directly f
Docume
Anacond
IBM
IBM-Project-54183-166176602
Home Page - Select or create
Untitled8 - Jupyter Notebook
localhost:8889/notebooks/Untitled8.ipynb?kernel_name=python3
jupyter Untitled8 Last Checkpoint: 3 minutes ago (unsaved changes)
Logout
File Edit View Insert Cell Kernel Widgets Help
Trusted
Python 3 (ipykernel)
Run
Code
Type here to search
28°C Partly cloudy
ENG IN
4:43 AM 11/17/2022

```
In [1]: #import the necessary files
        import numpy as np
        import pandas as pd
        import matplotlib.pyplot as plt
        %matplotlib inline
        import seaborn as sns
        from sklearn.tree import DecisionTreeRegressor
        from sklearn.ensemble import RandomForestRegressor
        from sklearn.linear_model import LinearRegression
        from sklearn.metrics import mean_squared_error,r2_score
        from sklearn.svm import SVR
        from sklearn.pipeline import Pipeline
        from sklearn.impute import SimpleImputer
        from sklearn.preprocessing import StandardScaler

In [2]: data = pd.read_csv("Mall_Customers.csv")

In [3]: data.head()

Out[3]:
```

```
   CustomerID  Gender  Age  Annual Income (k$)  Spending Score (1-100)
0           1    Male   19                  15                      39
1           2    Male   21                  15                      81
2           3  Female   20                  16                       6
```

10.ADVANTAGES AND DISADVANTAGES

ADVANTAGES:

Predicting flight delays can **improve airline operations and passenger satisfaction**, which will result in a positive impact on the economy. In this study, the main goal is to compare the performance of machine learning classification algorithms when predicting flight delays.

Flight delays not only irritate air passengers and disrupt their schedules but also cause a decrease in efficiency, an increase in capital costs, reallocation of flight crews and aircraft, and additional crew expenses (Britto et al., 2012; Yablonsky et al., 2014).

More than 1 in 5 domestic flights delayed early in 2022

"The shortage of pilots and airport workers left over from pandemic-era employee cuts has left many airlines struggling to match the high demand brought on by summer travel," says Sophia Mendel, ValuePenguin travel expert.

DISADVANTAGES:

The most common reasons for flight delays

- Air Traffic Control (ATC) restrictions. ...
- Adverse weather conditions. ...
- Bird strikes. ...
- Knock-on effect due to a delayed aircraft. ...
- Strikes. ...
- Waiting for connecting passengers. ...
- Waiting for connecting bags. ...
- Waiting for cargo

Simply search for your flight as a new (paid) reservation. If the flight doesn't appear in the search results, the airline is likely to cancel it. Alternatively, you can use ExpertFlyer and check to see if your flight has been "zeroed out" across all fare classes. If so, just sit back and wait.

The four forces acting on an aircraft in straight-and-level, unaccelerated flight are thrust, drag, lift, and weight. They are defined as follows: Thrust—the forward force produced by the powerplant/ propeller or rotor.The primary factors most affected by performance are the takeoff and landing distance, rate of climb, ceiling, payload, range, speed, maneuverability, stability, and fuel economy

11.CONCLUSION

In this project, we use flight data, weather, and demand data to predict flight parture delay. Our result shows that the Random Forest method yields the best performance compared to the SVM model. Somehow the SVM model is very time consuming and does not necessarily produce better results. In the end, our model correctly predicts 91% of the non-delayed flights. However, the delayed flights are only correctly predicted 41% of time. As a result, there can be additional features related to the causes of flight delay that are not yet discovered using our existing data sources.

In the second part of the project, we can see that it is possible to predict flight delay patterns from just the volume of concurrently published tweets, and their sentiment and objectivity. This is not unreasonable; people tend to post about airport delays on Twitter; it stands to reason that these posts would become more frequent, and more profoundly emotional, as the delays get worse. Without more data, we cannot make a robust model and find out the role of related factors and chance on these results.However, as a proof of concept, there is potential for these results. It may be possible to routinely use tweets to ascertain an understanding of concurrent airline delays and traffic patterns, which could be useful in a variety of circumstances.

12.FUTURE SCOPE

Keyword would not be very helpful. Instead, we're going to find these tweets through sentimentanalysis.The package we use, TextBlob, provides two separate assessments about a string. The first is polarity: this measures happiness versus unhappiness, with a scale from -1 (most unhappy) to 1 (most happy). The second is subjectivity, on a scale from 0 (most objective) to 1 (most subjective). 8.6 Tweets for Negative Sentiment First, we have analyzed the tweets for negative sentiment. To begin, we need to decide how extreme the negative tweets should be in order to be included in our set [20]. Therefore, we need to choose a threshold for polarity, between -1 and 0, such that only tweets below that threshold get included in the set. Let's plot the volume of tweets at each threshold.Tweets Volume with Sentiment <-0.1 In Figure 24, we still don't have a vast number of tweets per hour but was enough for a machine learning algorithm Given that we are not working with a massive volume of tweets, rather than making our set more selective, we have left it at -0.1 8.7 Tweets for Hgh Subjectivity Next, we have analyzed the tweets for high subjectivity.As we did above, first comparing the volume of tweets at each subjectivity threshold. In Figure 25, possible subjectivity ranges from 0 (entirely factual) to 1 (thoroughly opinionated) [21]. At each limit, we have considered the size of the set of tweets which are at or above that threshold. Figure25:Tweets Volume with Subjectivity >0.5 At this threshold level, we have enough tweets to have much variation hour-to-hour. We have kept this threshold for the rest of the analysis. subjectivity seems to track much more closely to the delay patterns than the graph for negative sentiment. Both analyses, plus the total tweet volume, may be useful for the machine learning algorithms.Now, to try the proof of concept, we are going to try to predict delays using machine learning. We have assembled a matrix with the three relevant factors for each hour -- tweet volume, number of tweets with negative sentiment, and number of subjective tweets.We have a concise range of time for which we have both tweets and data. Therefore, we have set a small goal: Using the tweets and delay data from December 4 as a training set, how well can we predict delays for December 5? First, we have created a Data Frame with the same results as before. After creating a Data Frame, we have added the average delay time per hour, and finally we removed training set from the variable,wepredicted delays at the point that they become problematic -- perhaps whether they exceed 20 minutes. Since we have worked with each hour of the day separately, we have tried to predict whether the average delay time in a given hour will be more considerable or less than 20 minutes 8.8 Random Forest.

```
Y=pd.DataFrame.transpose(pd.DataFrame([pd.concat([pd.DataFrame([Y_test])[i] for i
in pd.DataFrame([Y_test])], ignore_index = True), forest_predictions]))
true1 = 0
false1 = 0
true0 = 0
false0 = 0
for x in range(0, len(Y)):
```

```
if Y.ix[x,0] == 0:
if Y.ix[x,1] == 0:
true0 = true0 + 1
else:
false1 = false1 + 1
else:
if Y.ix[x,1] == 1:
true1 = true1 + 1
else:
false0 = false0 + 1
print ' Confusion Matrix'
print ' '
print ' ', 'No Delay (Predicted) ', 'Delay (Predicted)'
print 'No Delay ', true0, ' ', false1
print 'Delay ', false0, ' ', true1
```

Figure 26: Random Forest Confusion Matrix

In Figure 26, Random Forest algorithm had some mistakes. It predicted the majority of the delayed hours correctly, but only one of the two instances of the on time hours.8.9 Support Vector Machine

Figure 27: Support Vector Machine Confusion Matrix

In Figure 27, the results are slightly better using the SVM. From the perspective of delays: there were no false positives, one false negative, fourteen true positives and two true notes.

13.APPENDIX

SOURCE CODE

FLIGHT DELAY PREDICTION

Flight landing

LANDING

FLIGHT DATA:

login.php

```
<?php
include 'config.php';

use PHPMailer\PHPMailer\PHPMailer;use

PHPMailer\PHPMailer\SMTP;

use PHPMailer\PHPMailer\Exception;

require 'vendor/autoload.php';

session_start();

error_reporting(0);
if (isset($_SESSION["user_id"]))

    {header("Location:

    welcome.php");

}

if (isset($_POST["signup"])) {

    $full_name = mysqli_real_escape_string($conn, $_POST["signup_full_name"]);
    $email = mysqli_real_escape_string($conn, $_POST["signup_email"]);

    $password = mysqli_real_escape_string($conn, md5($_POST["signup_password"]));

    $cpassword = mysqli_real_escape_string($conn, md5($_POST["signup_cpassword"]));
    $token = md5(rand());
  $check_email = mysqli_num_rows(mysqli_query($conn, "SELECT email FROM users
WHERE email='$email'"));
```

```
        if ($password !== $cpassword) {
          echo "<script>alert('Password did not match.');</script>";

        } elseif ($check_email > 0) {
          echo "<script>alert('Email already exists in out database.');</script>";
        } else {

    $sql = "INSERT INTO users (full_name, email, password, token, status) VALUES
('$full_name', '$email', '$password', '$token', '0')";

          $result = mysqli_query($conn, $sql);

          if ($result) {

            $_POST["signup_full_name"] = "";
            $_POST["signup_email"] = "";
            $_POST["signup_password"] = "";

            $_POST["signup_cpassword"] = "";
            $to = $email;
            $subject = "Email verification - Find Me";

            $message = "

            <html>

            <head>

            <title>{$subject}</title>

            </head>

            <body>
            <p><strong>Dear {$full_name},</strong></p>

      <p>Thanks for registration! Verify your email to access our website. Click below link toverify
your email.</p>

            <p><a href='{$base_url}verify-email.php?token={$token}'>Verify Email</a></p>
            </body>

            </html>

            ";
            //Create an instance; passing `true` enables exceptions
            $mail = new PHPMailer(true);try

            {

              //Server settings

              $mail->SMTPDebug = 0;                     //Enable verbose debug output

              $mail->isSMTP();                                   //Send using SMTP
              $mail->Host       = $smtp['host'];                 //Set the SMTP server to send through
              $mail->SMTPAuth   = true;                               //Enable SMTP authentication
              $mail->Username   = $smtp['user'];                    //SMTP username
              $mail->Password   = $smtp['pass'];                           //SMTP password
        $mail->SMTPSecure = PHPMailer::ENCRYPTION_SMTPS;                       //Enable implicit TLS
encryption

        $mail->Port           = $smtp['port'];                             //TCP port to connect to; use 587
ifyou have set `SMTPSecure = PHPMailer::ENCRYPTION_STARTTLS`

              //Recipients
              $mail->setFrom($my_email);

              $mail->addAddress($email, $full_name);      //Add a recipient
              //Content
              $mail->isHTML(true);                               //Set email format to HTML

              $mail->Subject = $subject;

              $mail->Body     = $message;
              $mail->send();
              echo "<script>alert('We have sent a verification link to your email - {$email}.');</script>";

            } catch (Exception $e) {
              echo "<script>alert('Mail not sent. Please try again.');</script>";

            }
          } else {
            echo "<script>alert('User registration failed.');</script>";
```

```
          }
        }

}



if (isset($_POST["signin"])) {

        $email = mysqli_real_escape_string($conn, $_POST["email"]);
        $password = mysqli_real_escape_string($conn, md5($_POST["password"]));



  $check_email = mysqli_query($conn, "SELECT id FROM users WHERE email='$email' AND
password='$password' AND status='1'");

        if (mysqli_num_rows($check_email) > 0) {
          $row = mysqli_fetch_assoc($check_email);

          $_SESSION["user_id"] = $row['id'];

          header("Location: welcome.php");

        } else {

          echo "<script>alert('Login details is incorrect. Please try again.');</script>";

        }

}

?>

<!DOCTYPE html>

<html lang="en">
<head>

        <meta charset="UTF-8" />
        <meta name="viewport" content="width=device-width, initial-scale=1.0" />
        <link rel="stylesheet" href="style.css" />

        <title>Find Me</title>
</head>
<body>
```

```
    <div class="container">
      <div class="forms-container">

        <div class="signin-signup">
          <form action="" method="post" class="sign-in-form">
            <h2 class="title">Sign in</h2>

            <div class="input-field">
              <i class="fas fa-user"></i>

              <input type="text" placeholder="Email Address" name="email" value="<?php echo
$_POST['email']; ?>" required />

            </div>
            <div class="input-field">
              <i class="fas fa-lock"></i>
              <input type="password" placeholder="Password" name="password" value="<?php echo
$_POST['password']; ?>" required />

            </div>
            <input type="submit" value="Login" name="signin" class="btn solid" />

      <p style="display: flex;justify-content: center;align-items: center;margin-top: 20px;"><a
href="forgot-password.php" style="color: #4590ef;">Forgot Password?</a></p>

          </form>

          <form action="" class="sign-up-form" method="post">
            <h2 class="title">Sign up</h2>

            <div class="input-field">
              <i class="fas fa-user"></i>
        <input type="text" placeholder="Full Name" name="signup_full_name" value="<?php
echo $_POST["signup_full_name"]; ?>" required />

            </div>

            <div class="input-field">

              <i class="fas fa-envelope"></i>
        <input type="email" placeholder="Email Address" name="signup_email" value="<?phpecho
$_POST["signup_email"]; ?>" required />
            </div>
            <div class="input-field">
              <i class="fas fa-lock"></i>
        <input type="password" placeholder="Password" name="signup_password"
value="<?php echo $_POST["signup_password"]; ?>" required />

            </div>
            <div class="input-field">

              <i class="fas fa-lock"></i>
        <input type="password" placeholder="Confirm Password" name="signup_cpassword"
value="<?php echo $_POST["signup_cpassword"]; ?>" required />

            </div>

            <input type="submit" class="btn" name="signup" value="Sign up" />

          </form>
        </div>

      </div>



      <div class="panels-container">
        <div class="panel left-panel">
          <div class="content">

            <h3>New here ?</h3>

            <br>



            <button class="btn transparent" id="sign-up-

              btn">Sign up

            </button>

          </div>
          <img src="img/download.png" class="image" alt="" />
        </div>

        <div class="panel right-panel">
          <div class="content">
            <h3>One of us ?</h3>

            <br>

                       <button class="btn transparent" id="sign-in-

              btn">Sign in

            </button>

          </div>
          <img src="img/image.png" class="image" alt="" />
        </div>
      </div>
    </div>

<script src="https://kit.fontawesome.com/64d58efce2.js" crossorigin="anonymous"></script>
    <script src="app.js"></script>
</body>
</html>
```

logout.php

```
<?php

session_start();

session_unset();

session_destroy()

;

header("Location: login.php");

?>
```

Verify-email.php

```
<?php

session_start();

if (isset($_GET["token"]))

        {include 'config.php';
        $sql = "UPDATE users SET status='1' WHERE token='{$_GET["token"]}'";

        mysqli_query($conn, $sql);

    $showUserId = mysqli_fetch_assoc(mysqli_query($conn, "SELECT id FROM users WHERE
token='{$_GET["token"]}'"));

        $_SESSION["user_id"] = $showUserId['id'];

        header("Location: welcome.php");

} else {

        header("Location: login.php");

}

?>
```

reset-password.php

```
<?php



include 'config.php';



error_reporting(0);



session_start();



if (isset($_SESSION["user_id"]))

      {header("Location:

      welcome.php");

}
```

```
if (isset($_POST["resetPassword"])) {
        $password = mysqli_real_escape_string($conn, md5($_POST["new_password"]));

        $cpassword = mysqli_real_escape_string($conn, md5($_POST["cnew_password"]));

        if ($password === $cpassword) {

          $sql = "UPDATE users SET password='$password' WHERE token='{$_GET["token"]}'";

          mysqli_query($conn, $sql);

          header("Location: login.php");

        } else {

          echo "<script>alert('Password not matched.');</script>";

        }

}



?>
<!DOCTYPE html>
<html lang="en">

<head>
        <meta charset="UTF-8" />
        <meta name="viewport" content="width=device-width, initial-scale=1.0" />
        <link rel="stylesheet" href="style.css" />
        <title>Find Me</title>
</head>

<body>

        <div class="container">
          <div class="forms-container">

            <div class="signin-signup">
              <form action="" method="post" class="sign-in-form">
                <h2 class="title">Reset Password</h2>
        <div class="input-field">
         <i class="fas fa-lock"></i>

      <input type="password" placeholder="New Password" name="new_password"
value="<?php echo $_POST['new_password']; ?>" required />

        </div>
        <div class="input-field">

         <i class="fas fa-lock"></i>
      <input type="password" placeholder="Confirm New Password" name="cnew_password"
value="<?php echo $_POST['cnew_password']; ?>" required />

        </div>

        <input type="submit" value="Reset Password" name="resetPassword" class="btn solid" />
       </form>
      </div>

     </div>



     <div class="panels-container">

      <div class="panel left-panel">
       <div class="content">
        <h3>Reset Password ?</h3>
       </div>

       <img src="img/log.svg" class="image" alt="" />
      </div>
     </div>

    </div>
<script src="https://kit.fontawesome.com/64d58efce2.js" crossorigin="anonymous"></script>
    <script src="app.js"></script>

</body>
</html>
```

reset-password.php

```
<?php

include 'config.php';

error_reporting(0);

session_start();

if (isset($_SESSION["user_id"]))
        {header("Location:

        welcome.php");

}

if (isset($_POST["resetPassword"])) {
        $password = mysqli_real_escape_string($conn, md5($_POST["new_password"]));
        $cpassword = mysqli_real_escape_string($conn, md5($_POST["cnew_password"]));

        if ($password === $cpassword) {

          $sql = "UPDATE users SET password='$password' WHERE token='{$_GET["token"]}'";

          mysqli_query($conn, $sql);

          header("Location: login.php");

        } else {
          echo "<script>alert('Password not matched.');</script>";

        }
}

?>
<!DOCTYPE html>

<html lang="en">
<head>
        <meta charset="UTF-8" />

        <meta name="viewport" content="width=device-width, initial-scale=1.0" />
        <link rel="stylesheet" href="style.css" />
        <title>Find Me</title>
```

```
</head>
<body>

        <div class="container">
          <div class="forms-container">
            <div class="signin-signup">

              <form action="" method="post" class="sign-in-form">
                <h2 class="title">Reset Password</h2>
                <div class="input-field">

                  <i class="fas fa-lock"></i>

            <input type="password" placeholder="New Password" name="new_password" value="<?php echo $_POST['new_password']; ?>" required />

                </div>

                <div class="input-field">
                  <i class="fas fa-lock"></i>
            <input type="password" placeholder="Confirm New Password" name="cnew_password" value="<?php echo $_POST['cnew_password']; ?>" required />

                </div>

                <input type="submit" value="Reset Password" name="resetPassword" class="btn solid" />
              </form>
            </div>

          </div>
          <div class="panels-container">
            <div class="panel left-panel">

              <div class="content">
                <h3>Reset Password ?</h3>
              </div>
              <img src="img/log.svg" class="image" alt="" />
            </div>
          </div>
        </div>

        <script src="https://kit.fontawesome.com/64d58efce2.js" crossorigin="anonymous"></script>
        <script src="app.js"></script>
</body>

</html>
```

forgot-password.php

```
<?php



include 'config.php';

use PHPMailer\PHPMailer\PHPMailer;use

PHPMailer\PHPMailer\SMTP;

use PHPMailer\PHPMailer\Exception;

require 'vendor/autoload.php';

session_start();

error_reporting(0);

if (isset($_SESSION["user_id"]))

        {header("Location:

        welcome.php");

}
if (isset($_POST["resetPassword"])) {

        $email = mysqli_real_escape_string($conn, $_POST["email"]);

        $check_email = mysqli_query($conn, "SELECT * FROM users WHERE email='$email'");if

        (mysqli_num_rows($check_email) > 0) {

          $data = mysqli_fetch_assoc($check_email);

          $to = $email;
          $subject = "Reset Password - Find Me";
          $message = "
          <html>

          <head>
          <title>{$subject}</title>
          </head>

          <body>
          <p><strong>Dear {$data['full_name']},</strong></p>
          <p>Forgot Password? Not a problem. Click below link to reset your password.</p>

     <p><a href='{$base_url}reset-password.php?token={$data['token']}'>Reset
Password</a></p>

          </body>
          </html>

          ";

     $mail = new PHPMailer(true);try {

               $mail->SMTPDebug = 0;
            $mail->isSMTP();
            $mail->Host       = $smtp['host'];
            $mail->SMTPAuth  = true;

            $mail->Username   = $smtp['user'];
            $mail->Password   = $smtp['pass'];
            $mail->SMTPSecure = PHPMailer::ENCRYPTION_SMTPS;

            $mail->Port       = $smtp['port'];
            $mail->setFrom($my_email);
            $mail->addAddress($email, $data['full_name']);

            $mail->isHTML(true);

            $mail->Subject = $subject;

            $mail->Body     = $message;
```

```
      $mail->send();

      echo "<script>alert('We have sent a reset password link to your email -
{$email}.');</script>";
     } catch (Exception $e) {
      echo "<script>alert('Mail not sent. Please try again.');</script>";

     }

    } else {
     echo "<script>alert('Email not found.');</script>";

    }

}

?>
<!DOCTYPE html>

<html lang="en">
<head>
    <meta charset="UTF-8" />
    <meta name="viewport" content="width=device-width, initial-scale=1.0" />

    <link rel="stylesheet" href="style.css" />
    <title>Find Me</title>

</head>
<body>
    <div class="container">
     <div class="forms-container">

      <div class="signin-signup">
       <form action="" method="post" class="sign-in-form">
        <h2 class="title">Reset Password</h2>
        <div class="input-field">
         <i class="fas fa-user"></i>
```

```
                  <input type="text" placeholder="Email Address" name="email" value="<?php echo
$_POST['email']; ?>" required />

                </div>
          <input type="submit" value="Send Verification Link" name="resetPassword" class="btnsolid" />

              </form>

            </div>
          </div>
          <div class="panels-container">

            <div class="panel left-panel">
              <div class="content">
                <h3>Forgot Password ?</h3>
</div>

              <img src="img/img1.png" class="image" alt="" />
            </div>
          </div>
        </div>
        <script src="https://kit.fontawesome.com/64d58efce2.js" crossorigin="anonymous"></script>

        <script src="app.js"></script>
</body>
</html>
```

config.php

```
<?php

$hostname = "localhost";

$username = "root";
$password = "";
$database = "test";
```

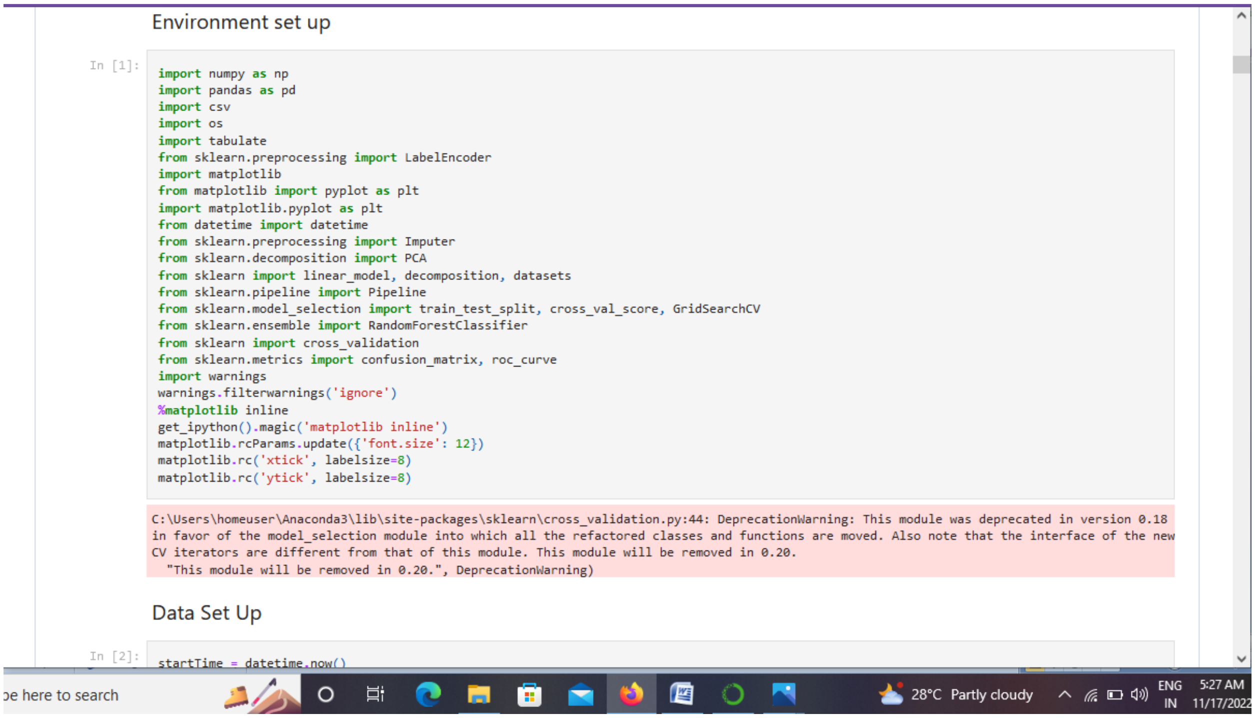
```
$conn = mysqli_connect($hostname, $username, $password, $database) or die("Database
connection failed");
$base_url = "http://localhost/Find-Me/";
$my_email = "ENTER YOUR MAIL";
$smtp['host'] = "smtp.gmail.com";
$smtp['user'] = " ENTER YOUR MAIL ";

$smtp['pass'] = "ENTER YOUR PASSWORD";
$smtp['port'] = 465;
```

db.sql

```
SET SQL_MODE = "NO_AUTO_VALUE_ON_ZERO";START

TRANSACTION;

SET time_zone = "+00:00";


CREATE TABLE `users` (
        `id` int(11) NOT NULL,
        `full_name` varchar(50) NOT NULL,

        `email` varchar(100) NOT NULL,

        `password` varchar(255) NOT NULL,
        `token` varchar(255) NOT NULL,

        `status` int(11) NOT NULL,
        `photo` varchar(255) NOT NULL
) ENGINE=InnoDB DEFAULT CHARSET=utf8mb4;


INSERT INTO `users` (`id`, `full_name`, `email`, `password`, `token`, `status`, `photo`)
VALUES
```

```
(1, 'Prasanna', 'jeyaprasanna630@gmail.com',
'bedebb62a1716ada5fa7203f46c02723', '0cb5e63a18e45aadc68a9e894d88618d',
1, '907182248Screenshot 2021-05-17 115128.png');


ALTER TABLE `users`
      ADD PRIMARY KEY (`id`);


AUTO_INCREMENT for dumped

tablesAUTO_INCREMENT for

table `users` ALTER TABLE

`users`

          MODIFY `id` int(11) NOT NULL AUTO_INCREMENT, AUTO_INCREMENT=3;COMMIT;
```

PREDICTION PROGRAM:



Environment set up

In [1]:

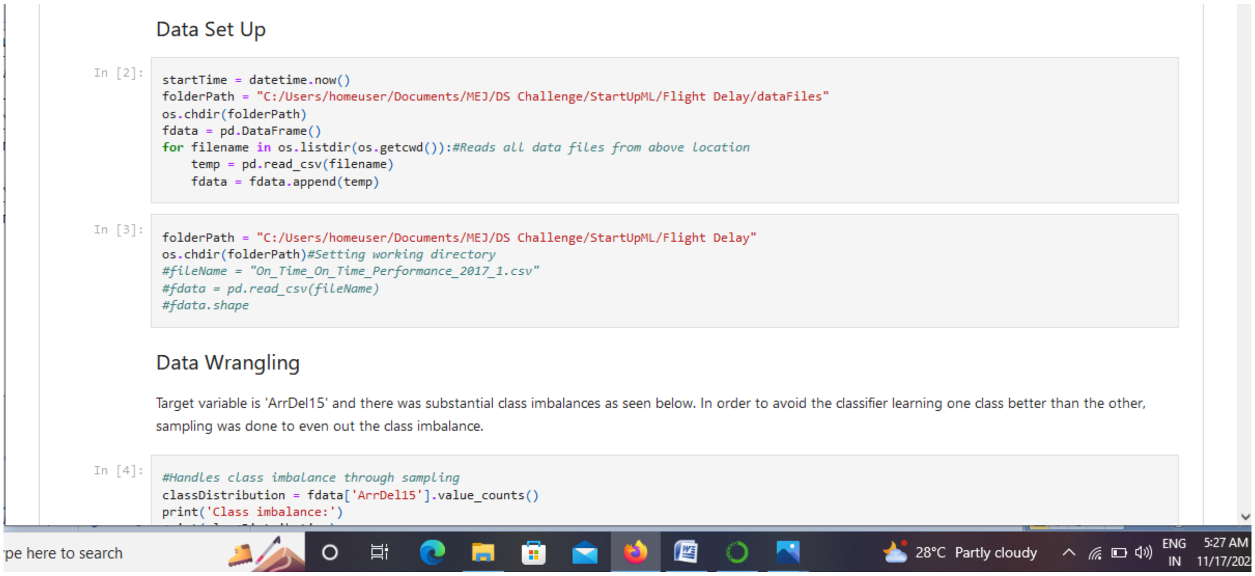
```
import numpy as np
import pandas as pd
import csv
import os
import tabulate
from sklearn.preprocessing import LabelEncoder
import matplotlib
from matplotlib import pyplot as plt
import matplotlib.pyplot as plt
from datetime import datetime
from sklearn.preprocessing import Imputer
from sklearn.decomposition import PCA
from sklearn import linear_model, decomposition, datasets
from sklearn.pipeline import Pipeline
from sklearn.model_selection import train_test_split, cross_val_score, GridSearchCV
from sklearn.ensemble import RandomForestClassifier
from sklearn import cross_validation
from sklearn.metrics import confusion_matrix, roc_curve
import warnings
warnings.filterwarnings('ignore')
%matplotlib inline
get_ipython().magic('matplotlib inline')
matplotlib.rcParams.update({'font.size': 12})
matplotlib.rc('xtick', labelsize=8)
matplotlib.rc('ytick', labelsize=8)
```

```
C:\Users\homeuser\Anaconda3\lib\site-packages\sklearn\cross_validation.py:44: DeprecationWarning: This module was deprecated in version 0.18
in favor of the model_selection module into which all the refactored classes and functions are moved. Also note that the interface of the new
CV iterators are different from that of this module. This module will be removed in 0.20.
  "This module will be removed in 0.20.", DeprecationWarning)
```

Data Set Up

In [2]:

```
startTime = datetime.now()
```



Data Set Up

In [2]:

```
startTime = datetime.now()
folderPath = "C:/Users/homeuser/Documents/MEJ/DS Challenge/StartUpML/Flight Delay/dataFiles"
os.chdir(folderPath)
fdata = pd.DataFrame()
for filename in os.listdir(os.getcwd()):#Reads all data files from above location
    temp = pd.read_csv(filename)
    fdata = fdata.append(temp)
```

In [3]:

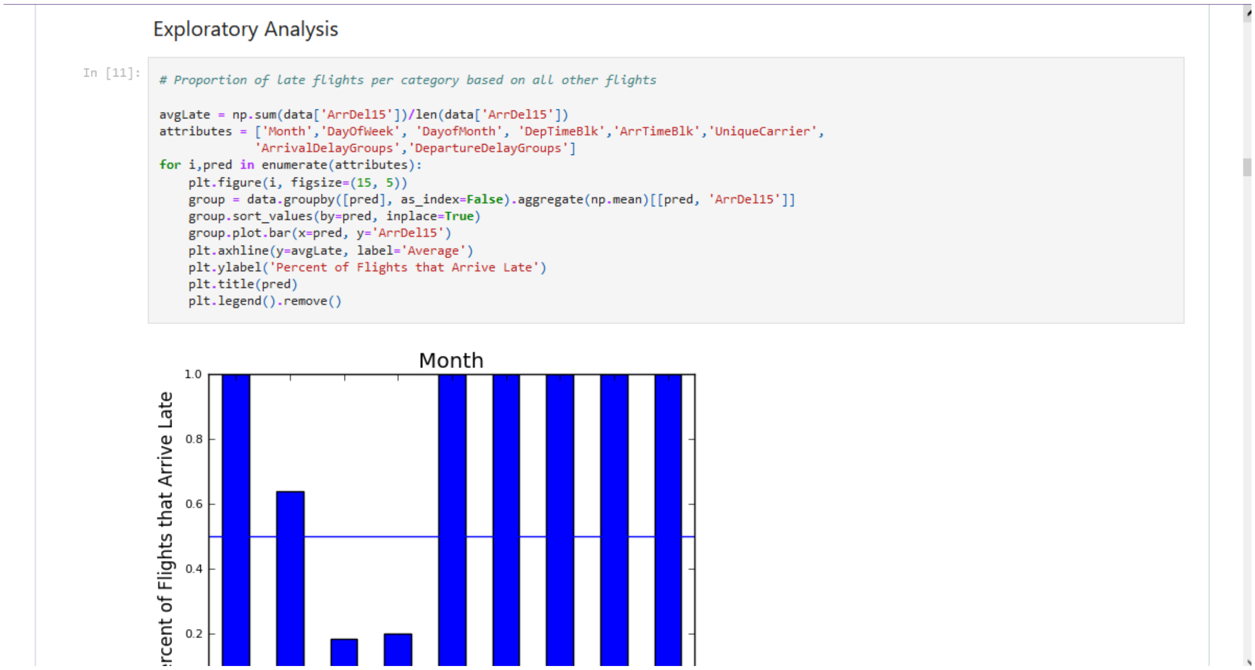
```
folderPath = "C:/Users/homeuser/Documents/MEJ/DS Challenge/StartUpML/Flight Delay"
os.chdir(folderPath)#Setting working directory
#fileName = "On_Time_On_Time_Performance_2017_1.csv"
#fdata = pd.read_csv(fileName)
#fdata.shape
```

Data Wrangling

Target variable is 'ArrDel15' and there was substantial class imbalances as seen below. In order to avoid the classifier learning one class better than the other, sampling was done to even out the class imbalance.

In [4]:

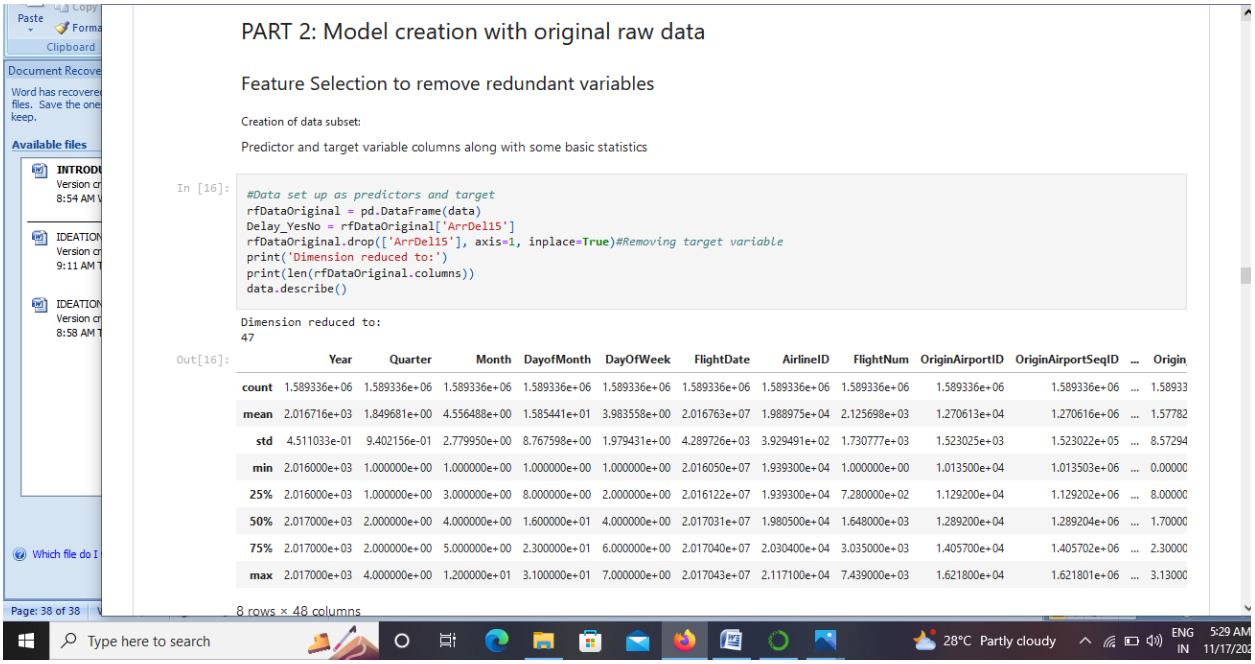
```
#Handles class imbalance through sampling
classDistribution = fdata['ArrDel15'].value_counts()
print('Class imbalance:')
```



Exploratory Analysis

In [11]:

```
# Proportion of late flights per category based on all other flights

avgLate = np.sum(data['ArrDel15'])/len(data['ArrDel15'])
attributes = ['Month','DayOfWeek', 'DayofMonth', 'DepTimeBlk','ArrTimeBlk','UniqueCarrier',
              'ArrivalDelayGroups','DepartureDelayGroups']
for i,pred in enumerate(attributes):
    plt.figure(i, figsize=(15, 5))
    group = data.groupby([pred], as_index=False).aggregate(np.mean)[[pred, 'ArrDel15']]
    group.sort_values(by=pred, inplace=True)
    group.plot.bar(x=pred, y='ArrDel15')
    plt.axhline(y=avgLate, label='Average')
    plt.ylabel('Percent of Flights that Arrive Late')
    plt.title(pred)
    plt.legend().remove()
```



PART 2: Model creation with original raw data

Feature Selection to remove redundant variables

Creation of data subset:

Predictor and target variable columns along with some basic statistics

In [16]:

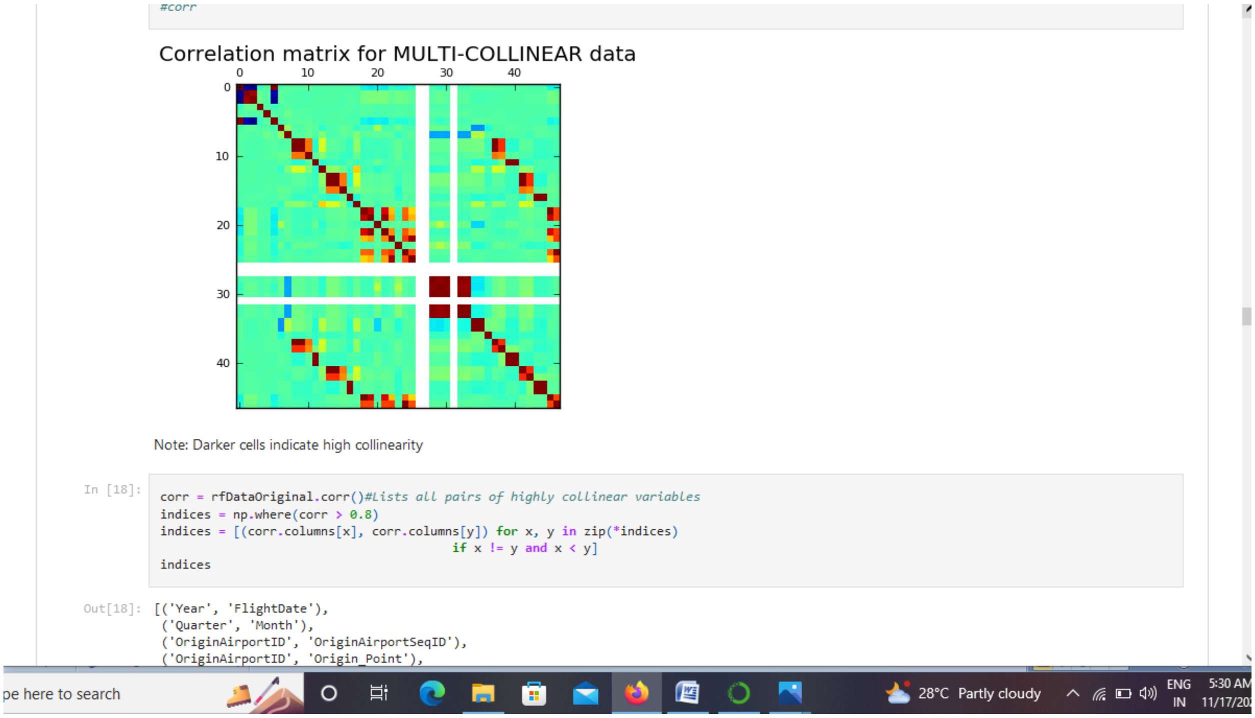
```
#Data set up as predictors and target
rfDataOriginal = pd.DataFrame(data)
Delay_YesNo = rfDataOriginal['ArrDel15']
rfDataOriginal.drop(['ArrDel15'], axis=1, inplace=True)#Removing target variable
print('Dimension reduced to:')
print(len(rfDataOriginal.columns))
data.describe()
```

```
Dimension reduced to:
47
```

Out[16]:

	Year	Quarter	Month	DayofMonth	DayOfWeek	FlightDate	AirlineID	FlightNum	OriginAirportID	OriginAirportSeqID	...	Origin
count	1.589336e+06	1.589336e+06	1.589336e+06	1.589336e+06	1.589336e+06	1.589336e+06	1.589336e+06	1.589336e+06	1.589336e+06	1.589336e+06	...	1.58933
mean	2.016716e+03	1.849681e+00	4.556488e+00	1.585441e+01	3.983558e+00	2.016763e+07	1.988975e+04	2.125698e+03	1.270613e+04	1.270616e+06	...	1.57782
std	4.511033e-01	9.402156e-01	2.779950e+00	8.767598e+00	1.979431e+00	4.289726e+03	3.929491e+02	1.730777e+03	1.523025e+03	1.523022e+05	...	8.57294
min	2.016000e+03	1.000000e+00	1.000000e+00	1.000000e+00	1.000000e+00	2.016050e+07	1.939300e+04	1.000000e+00	1.013500e+04	1.013503e+06	...	0.00000
25%	2.016000e+03	1.000000e+00	3.000000e+00	8.000000e+00	2.000000e+00	2.016122e+07	1.939300e+04	7.280000e+02	1.129200e+04	1.129202e+06	...	8.00000
50%	2.017000e+03	2.000000e+00	4.000000e+00	1.600000e+01	4.000000e+00	2.017031e+07	1.980500e+04	1.648000e+03	1.289200e+04	1.289204e+06	...	1.70000
75%	2.017000e+03	2.000000e+00	5.000000e+00	2.300000e+01	6.000000e+00	2.017040e+07	2.030400e+04	3.035000e+03	1.405700e+04	1.405702e+06	...	2.30000
max	2.017000e+03	4.000000e+00	1.200000e+01	3.100000e+01	7.000000e+00	2.017043e+07	2.117100e+04	7.439000e+03	1.621800e+04	1.621801e+06	...	3.13000

8 rows × 48 columns



```
#corr
```

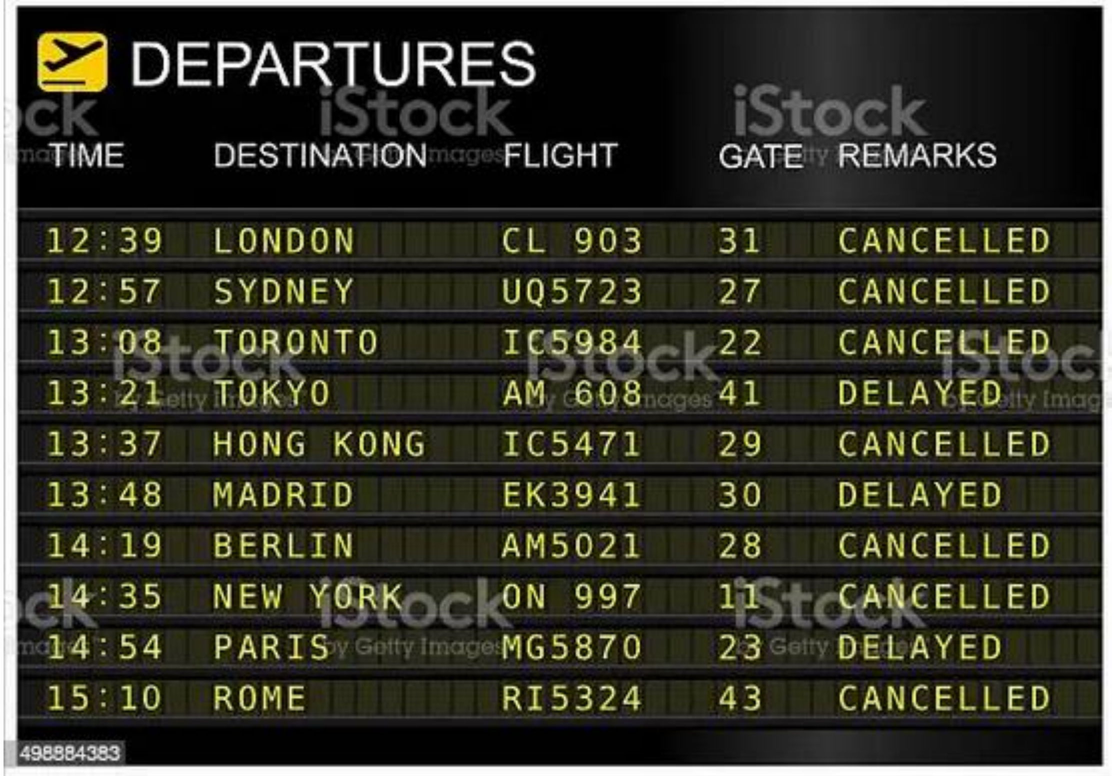
Correlation matrix for MULTI-COLLINEAR data

Note: Darker cells indicate high collinearity

```
In [18]:  corr = rfDataOriginal.corr()#Lists all pairs of highly collinear variables
           indices = np.where(corr > 0.8)
           indices = [(corr.columns[x], corr.columns[y]) for x, y in zip(*indices)
                                                   if x != y and x < y]
           indices
```

```
Out[18]: [('Year', 'FlightDate'),
          ('Quarter', 'Month'),
          ('OriginAirportID', 'OriginAirportSeqID'),
          ('OriginAirportID', 'Origin_Point'),
```

TESTING OF FLIGHT DELAY



DEPARTURES

TIME	DESTINATION	FLIGHT	GATE	REMARKS
12:39	LONDON	CL 903	31	CANCELLED
12:57	SYDNEY	UQ5723	27	CANCELLED
13:08	TORONTO	IC5984	22	CANCELLED
13:21	TOKYO	AM 608	41	DELAYED
13:37	HONG KONG	IC5471	29	CANCELLED
13:48	MADRID	EK3941	30	DELAYED
14:19	BERLIN	AM5021	28	CANCELLED
14:35	NEW YORK	ON 997	11	CANCELLED
14:54	PARIS	MG5870	23	DELAYED
15:10	ROME	RI5324	43	CANCELLED

ON TIME
DELAYED
CANCELLED

ON TIME
C 03
DELAYED
16
ON TIME
B 07
ON
iStock by Getty Images
845778956

GITHUB LINK:

https://github.com/IBM-EPBL/IBM-Project-54183-1661766020

YOUTUBE LINK:

https://youtu.be/qTDfbSvdVwo