Flight Delay Prediction Aviation Industry Using Machine Learning

CONTENTS

CONTENTS OF TABLE

S.NO	CONTENTS	PAGE.NO
1	1.INTRODUCTION	
	1.1 Overview	
	1.2 Purpose	
2	2.Problem Definition & Design Thinking	
	2.1 Empathy Map	
	2.2 Ideation & Brainstorming Map	
3	3.RESULT	
4	4.ADVANTAGES & DISADVANTAGES	
5	5.APPLICATIONS	
6	6.CONCLUSION	
7	7.FUTURE SCOPE	
8	8.APPENDIX	

1.INTRODUCTION

1.1 OVERVIEW

Flight delay prediction using machine learning is a predictive modeling approach that uses algorithms to identify variables and patterns that influence flight delays. This approach involves collecting historic flight data and using it to create models that predict the likelihood of a flight delay.

Machine learning algorithms are used to analyze large data sets such as flight data, weather data, flight schedules, and historical flight data to identify trends, patterns and other variables

that may influence flight delays. The system then employs these variables to make a prediction on whether or not a flight is likely to be delayed. This helps airlines and passengers to manage their schedules better and take steps to minimize the impact of delays. Machine learning algorithms used for flight delay prediction include random forest, decision tree, neural networks, and support vector machinesThese algorithms can be trained on a large dataset of historic flight data that contains features such as flight route, airline, average flight time, weather conditions, and previous flight delay data. Once the algorithm has been trained, it can be used to predict the probability of a flight delay given new data points.

There are several benefits to using machine learning for flight delay prediction. For airlines, it can help improve on-time performance and reduce the cost of disruptions, such as having to rebook passengers on a later flight. For passengers, it can help them plan their travel more effectively and avoid the inconvenience of a delayed flight.

There are also challenges to using machine learning for flight delay prediction. One challenge is obtaining accurate and timely data, particularly weather data, which...can have a significant impact on flight delays. Collecting and analyzing large amounts of data can also be time-consuming and require expertise in data science and machine learning.

Another challenge is the unpredictability of flight delays due to unforeseen events such as equipment malfunctions, air traffic control issues, and security concerns. Machine learning models may not always account for these unexpected events and may provide inaccurate predictions.

Despite these challenges, many airlines and aviation companies are investing in machine learning for flight delay prediction as part of their overall strategy to improve operational efficiency and provide a better customer experience. As machine learning technology continues to advance, we can expect to see even more sophisticated and accurate predictions of flight delays in the future.

Flight delay prediction using machine learning is a crucial tool in the aviation industry. Machine learning models can analyze a vast amount of data from various sources, such as weather conditions, historical flight data, and air traffic control data, to predict flight delays accurately. Accurate predictions can enable airlines to optimize their operations, improve customer satisfaction, reduce costs, and maintain a competitive advantage.

The use of machine learning in flight delay prediction is still in its early stages, and there are various challenges to overcome, such as data quality, complexity, interpretability, and ethical concerns. However, as the technology continues to evolve, it has the potential to revolutionize the aviation industry, improving the efficiency and safety of air travel and enhancing the customer experience.

The future scope for flight delay prediction using machine learning is vast, with potential developments in data integration, real-time data, interpretability, explainable AI, unsupervised learning, system integration, and ethical considerations. As airlines continue to adopt and leverage these models, they can remain competitive in the industry and provide a better travel experience for their passengers.

1.2 PURPOSE

PURPOSE

The purpose of flight delay prediction using machine learning is to provide accurate and timely information to airline companies, passengers, and airports regarding potential delays. By analyzing historical flight data, weather patterns, air traffic, and other factors using machine learning algorithms, predictions can be made about the likelihood and duration of flight delays.

This allows airline companies to proactively adjust schedules, optimize resources, and manage customer expectations, ultimately enhancing the overall travel experience. Passengers can also benefit from this information by being alerted of potential delays and having the ability to make alternative plans if necessary. Additionally, airports can use this information to optimize their operations, improving efficiency and reducing congestion.

Another benefit of flight delay prediction using machine learning is that it can help reduce the economic costs of flight delays. According to a report by the US Federal Aviation Administration, flight delays cost the US economy around $32.9 billion annually. By providing

accurate predictions and enabling proactive measures, airline companies can reduce the impact of delays and potentially save millions of dollars in compensations to passengers, lost revenue, and operational costs.

Furthermore, flight delay prediction models can be integrated into other travel-related applications, such as mobile apps and airport information displays, to help passengers plan their travel and avoid unnecessary stress.

However, there are some challenges associated with developing accurate flight delay prediction models using machine learning. One of the main challenges is the complexity of the aviation industry and the…
variability of factors affecting flight delays. The machine learning algorithms need to take into account a range of variables, such as weather conditions, air traffic congestion, mechanical issues, and crew availability, among others. Moreover, the data used to train the models In conclusion, the use of machine learning in predicting flight delays has the potential to greatly improve the efficiency of the aviation industry. However, it also presents several challenges that must be addressed for it to be successful.

Overcoming these challenges will require a collaborative effort between airlines, technology firms, and regulators to ensure that machine learning is developed and implemented in a responsible and effective manner.needs to be diverse and large enough to cover all possible scenarios, which can be a daunting task.

Another challenge is the continuous evolution of the aviation industry, as new routes, airlines, and technologies are introduced. This means that the models need to be regularly updated to ensure they remain accurate and relevant.

Finally, privacy concerns need to be addressed when collecting and processing personal data related to flight passengers. The airlines need to be transparent about their data handling procedures and obtain explicit consent from passengers before using their data.

2.PROBLEM DEFINITION & DESIGN THINKING

Problem Definition:

The problem we want to solve is predicting flight delays using machine learning. Flight delays can cause a lot of inconvenience to passengers and airlines alike. For passengers, it can lead to missed connections, delays in reaching their destination, and other inconveniences. For airlines, it can lead to decreased customer satisfaction and increased costs due to missed connections and scheduling conflicts. Hence, accurate prediction of flight delays can help airlines and passengers make informed decisions and minimize the negative impact of delays.

Design Thinking Approach:

Design thinking is a problem-solving approach that focuses on understanding the users' needs, ideating possible solutions, prototyping and testing those solutions to ensure they meet the users' requirements.

Here are the steps we can follow to apply design thinking to our flight delay prediction problem:

1. **Empathize: The first step is to empathize with the users, which in this case are the passengers and airlines. We need to understand their pain points, their needs, and how they would like to benefit from a flight delay prediction system. We can do this by conducting interviews, surveys, and analyzing user feedback.**
2. **Define: Based on our research and understanding of user needs, we can define the problem and the requirements for the flight delay prediction system. For instance, we may define the problem as predicting flight delays accurately, with a high degree of confidence, and providing actionable recommendations to airlines and passengers.**
3. **Ideate: In this step, we generate ideas for the flight delay prediction system. We can brainstorm and generate multiple solutions that meet the defined requirements. For example, we can use machine learning algorithms to predict flight delays, combine weather data, airport congestion, and airline operational data to improve the accuracy of predictions, and use visualizations and recommendations to make predictions more actionable.**
4. **Prototype: Once we have some ideas, we can create prototypes of the flight delay prediction system. We can create a minimum viable product (MVP) that includes only the essential features required to test the system's effectiveness. For instance, we can create a simple web application that takes in a flight number and date and returns the predicted delay time and possible recommendations.**
5. **Test: Finally, we test the prototype with real users to validate our assumptions and refine the design based on feedback. We can use A/B testing to compare the effectiveness of different machine learning algorithms, user testing to understand how users interact with the system, and collect feedback from users to make further improvements.**

By following this design thinking approach, we can develop a flight delay prediction system that

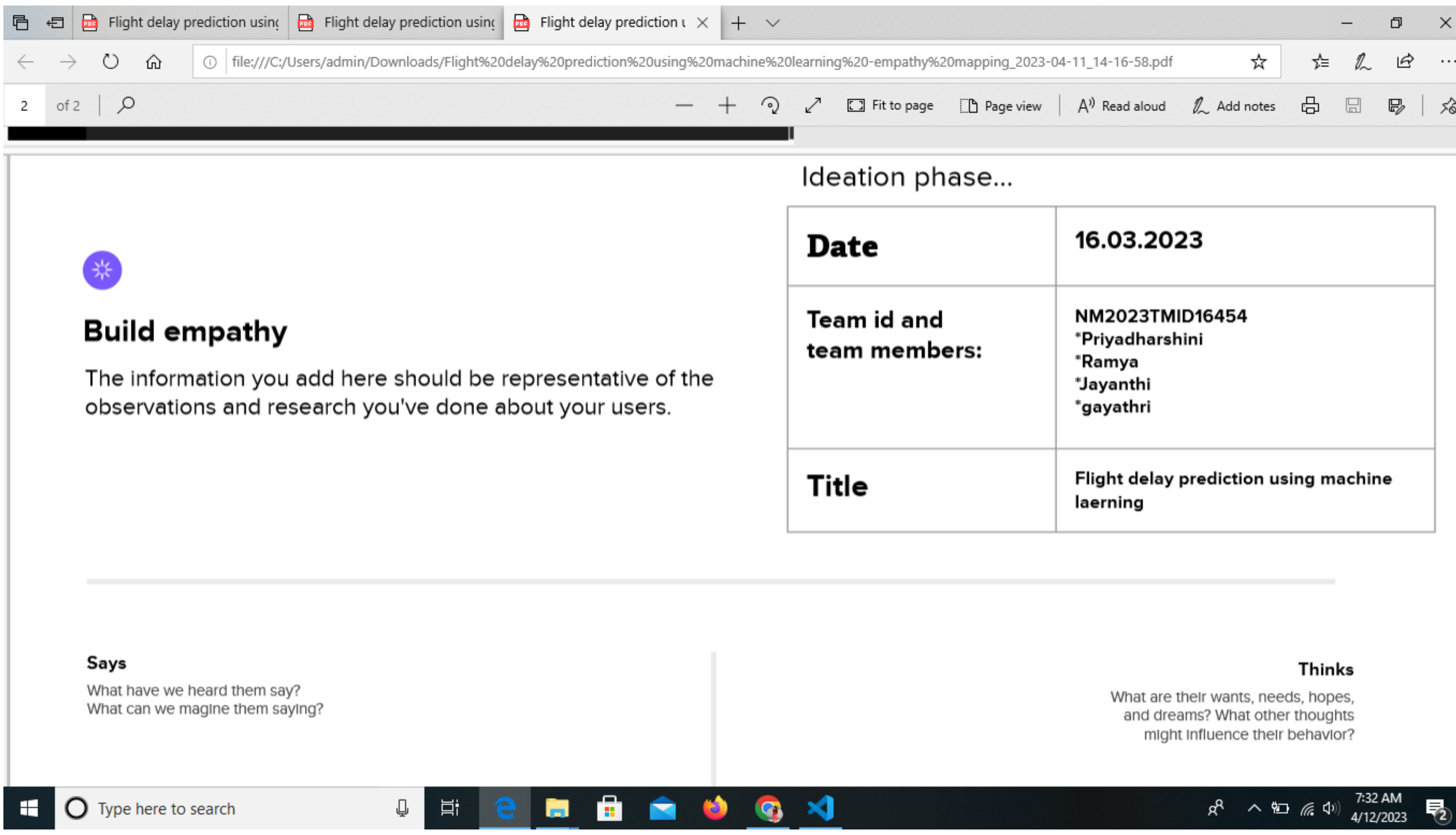
meets the users' needs, is effective, and improves the overall user experience for passengers and airlines.

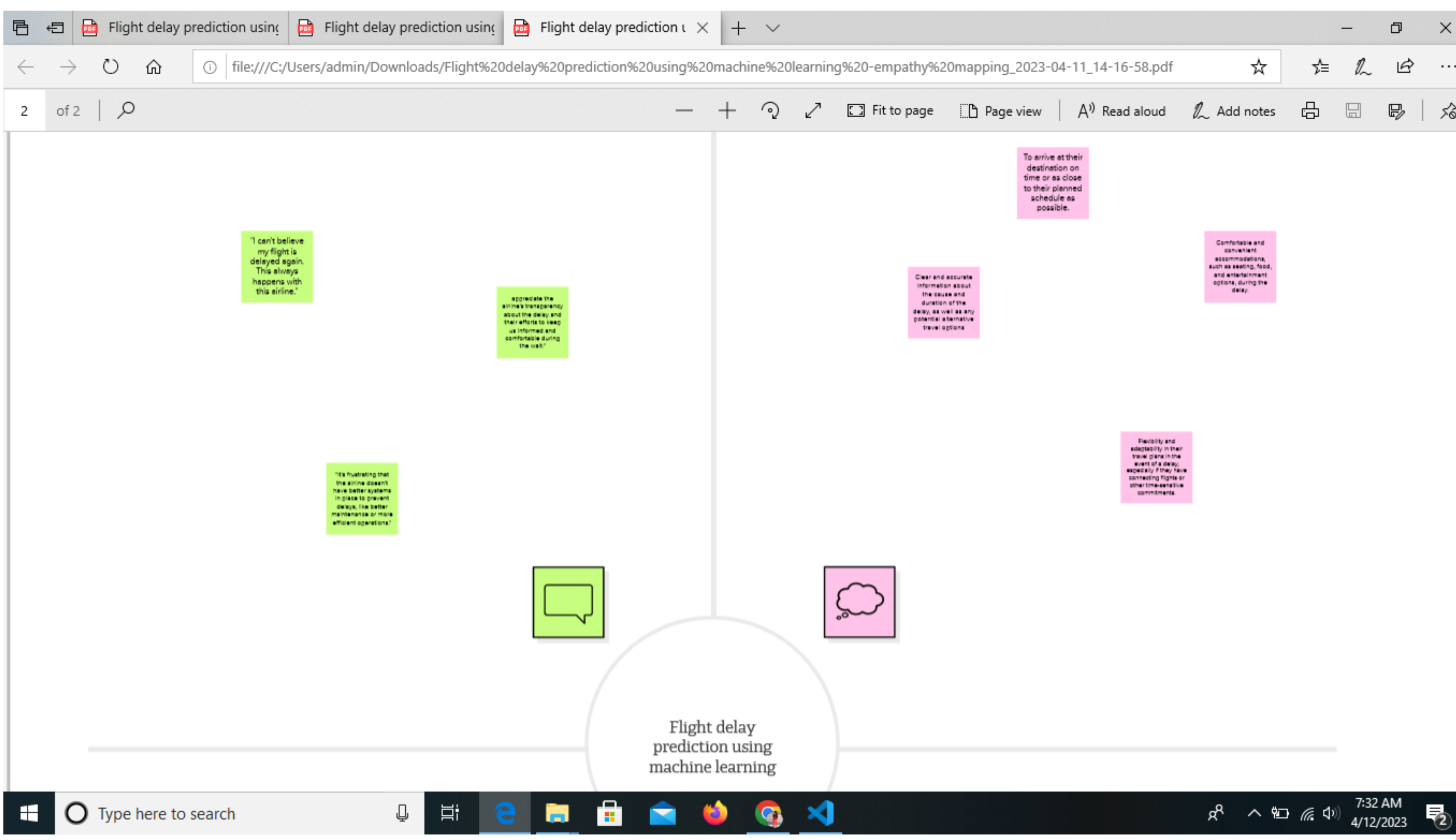
2.1 EMPATHY MAP

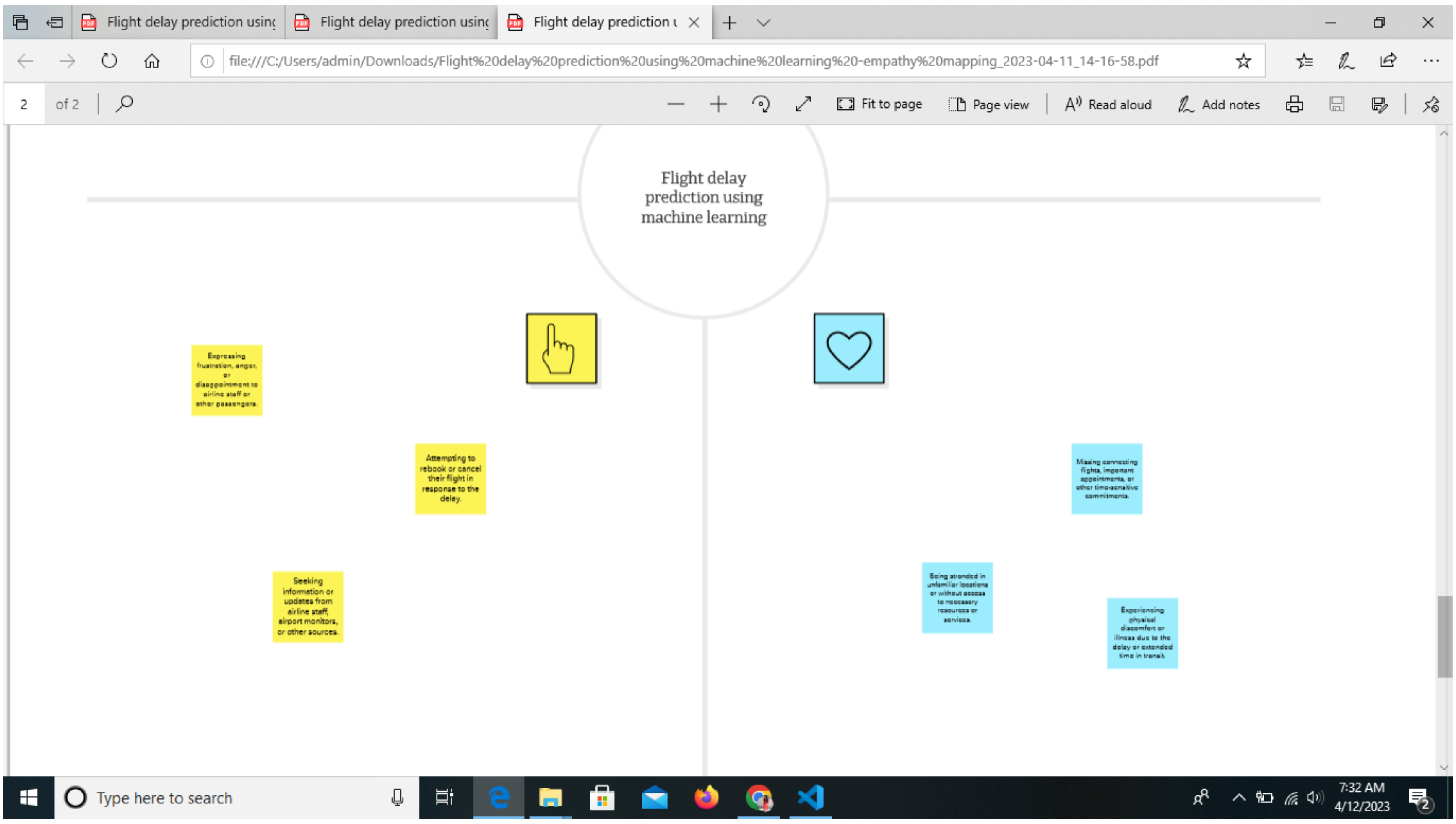
Template

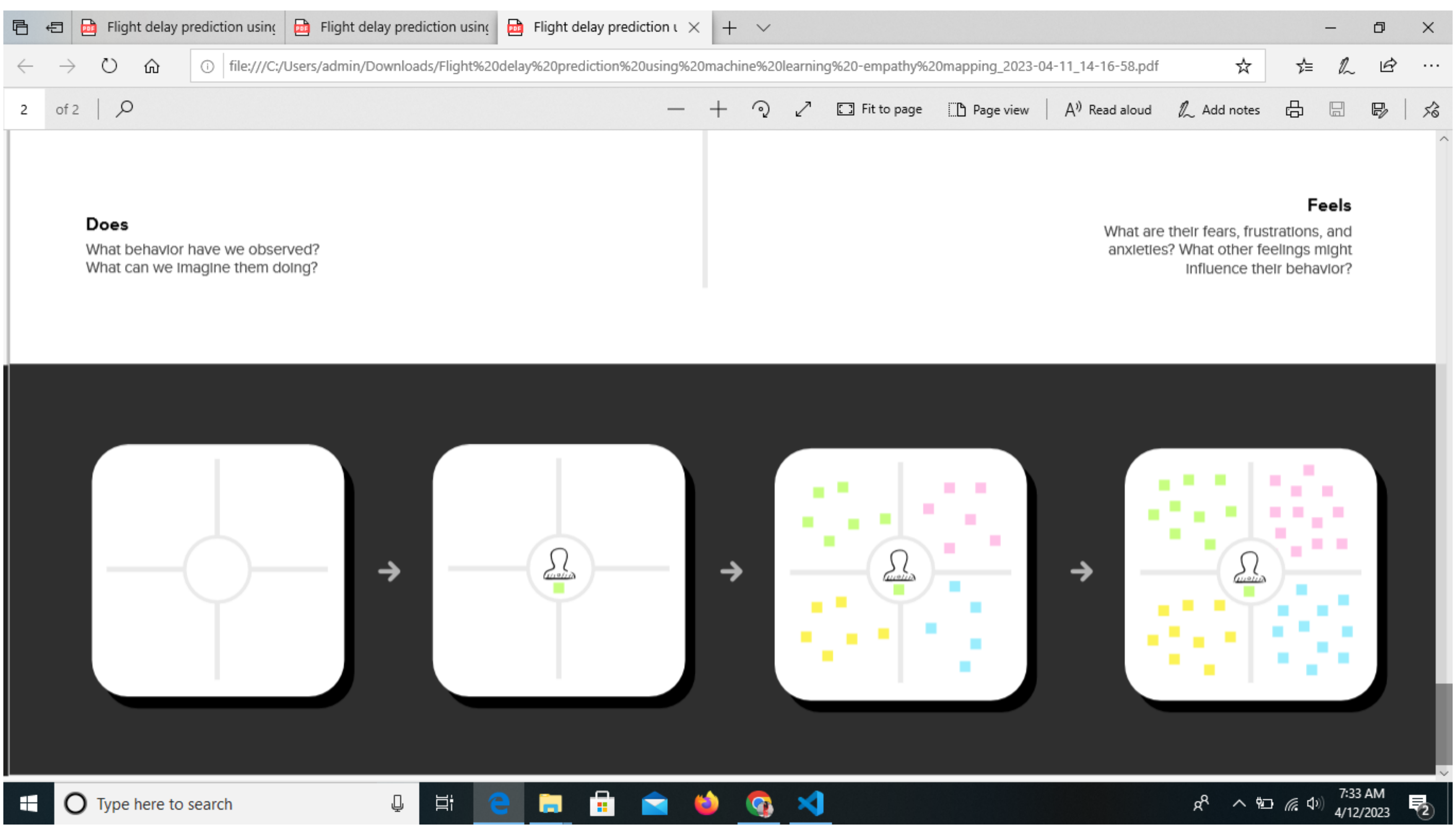
Empathy map

Use this framework to develop a deep, shared understanding and empathy for other people. An empathy map helps describe the aspects of a user's experience, needs and pain points, to quickly understand your users’ experience and mindset.


Ideation phase...
Build empathy
The information you add here should be representative of the observations and research you've done about your users.
Date
16.03.2023
Team id and team members:
NM2023TMID16454
*Priyadharshini
*Ramya
*Jayanthi
*gayathri
Title
Flight delay prediction using machine laerning
Says
What have we heard them say?
What can we magine them saying?
Thinks
What are their wants, needs, hopes, and dreams? What other thoughts might influence their behavior?


"I can't believe my flight is delayed again. This always happens with this airline."
appreciate the airline's transparency about the delay and their efforts to keep us informed and comfortable during the wait."
"It's frustrating that the airline doesn't have better systems in place to prevent delays, like better maintenance or more efficient operations."
To arrive at their destination on time or as close to their planned schedule as possible.
Clear and accurate information about the cause and duration of the delay, as well as any potential alternative travel options
Comfortable and convenient accommodations, such as seating, food, and entertainment options, during the delay.
Flexibility and adaptability in their travel plans in the event of a delay, especially if they have connecting flights or other time-sensitive commitments.
Flight delay prediction using machine learning





2.2 IDEATION & BRAINSTORMING MAP

Template

Brainstorm & idea prioritization

Use this template in your own brainstorming sessions so your team can unleash their imagination and start shaping concepts even if you're not sitting in the same room.

- **10 minutes** to prepare
- **1 hour** to collaborate
- **2-8 people** recommended

Date	16.03.2023
Team id and team members:	NM2023TMID16454 *Priyadharshini *Ramya *Jayanthi *gayathri
Title	Flight delay prediction using machine laerning

Before you collaborate

A little bit of preparation goes a long way with this session. Here's what you need to do to get going.

10 minutes

A **Team gathering**

Define who should participate in the session and send an invite. Share relevant information or pre-work ahead.

B **Set the goal**

Think about the problem you'll be focusing on solving in the brainstorming session.

C **Learn how to use the facilitation tools**

Use the Facilitation Superpowers to run a happy and productive session.

Open article →

1

Flight delay prediction problem statement

What problem are you trying to solve? Frame your problem as a How Might We statement. This will be the focus of your brainstorm.

5 minutes

PROBLEM STATEMENT

The goal is to provide airlines and passengers with reliable predictions of flight delays, which can help them to adjust their schedules accordingly and minimize the impact of delays.

Key rules of brainstorming

To run an smooth and productive session

- Stay in topic.
- Encourage wild ideas.
- Defer judgment.
- Listen to others.
- Go for volume.
- If possible, be visual.

3

Group ideas

Take turns sharing your ideas while clustering similar or related notes a
sticky notes have been grouped, give each cluster a sentence-like lab
bigger than six sticky notes, try and see if you and break it up into sma

20 minutes

accurate predicting: these flight delays allows passengers to be well prepared for the deterrent caused to their journey and enables airlines to respond to the potential causes of the flight delays in advance to diminish the negative imp

Gather dataset: The project, we look at using Python based Logistic Regression along with Support Vector Machine and then plugging the dataset into our classifier for results. In the second part of the project, we primarily focus on gathering a dataset fro

Data preprocessing: What steps should be taken to preprocess and clean the data before feeding it into machine learning models? Possible steps could include filling in missing data, scaling and normalizing features, and identifying and removing outliers.

Model evaluation: How should the performance of machine learning models be evaluated? Possible metrics could include accuracy, precision, recall, F1 score, and receiver operating characteristic (ROC) curve.

Real-time prediction: How can machine learning models be deployed to predict flight delays in real-time? Possible solutions could include integrating machine learning models into airline or airport operations systems, or building a separate application for predicting flight delays.

Interpretability: How can machine learning models be made more interpretable to stakeholders? Possible solutions could include using techniques like feature importance, partial dependence plots, and model visualization.

ur ideas while clustering similar or related notes as you go. Once all
›n grouped, give each cluster a sentence-like label. If a cluster is
notes, try and see if you and break it up into smaller sub-groups.

Gather dataset:
The project, we look at using Python based Logistic Regression along with Support Vector Machine and then plugging the dataset into our classifier for results. In the second part of the project, we primarily focus on gathering a dataset fro

Data preprocessing: What steps should be taken to preprocess and clean the data before feeding it into machine learning models? Possible steps could include filling in missing data, scaling and normalizing features, and identifying and removing outliers.

Model update:
How can machine learning models be updated to incorporate new data and adapt to changing conditions? Possible solutions could include using online learning techniques or periodically retraining the model on updated data.

Real-time prediction:
How can machine learning models be deployed to predict flight delays in real-time? Possible solutions could include integrating machine learning models into airline or airport operations systems, or building a separate application for predicting flight delays.

Interpretability:
How can machine learning models be made more interpretable to stakeholders? Possible solutions could include using techniques like feature importance, partial dependence plots, and model visualization.

4

Prioritize

Your team should all be on the same page about what's important moving forward. Place your ideas on this grid to determine which ideas are important and which are feasible.

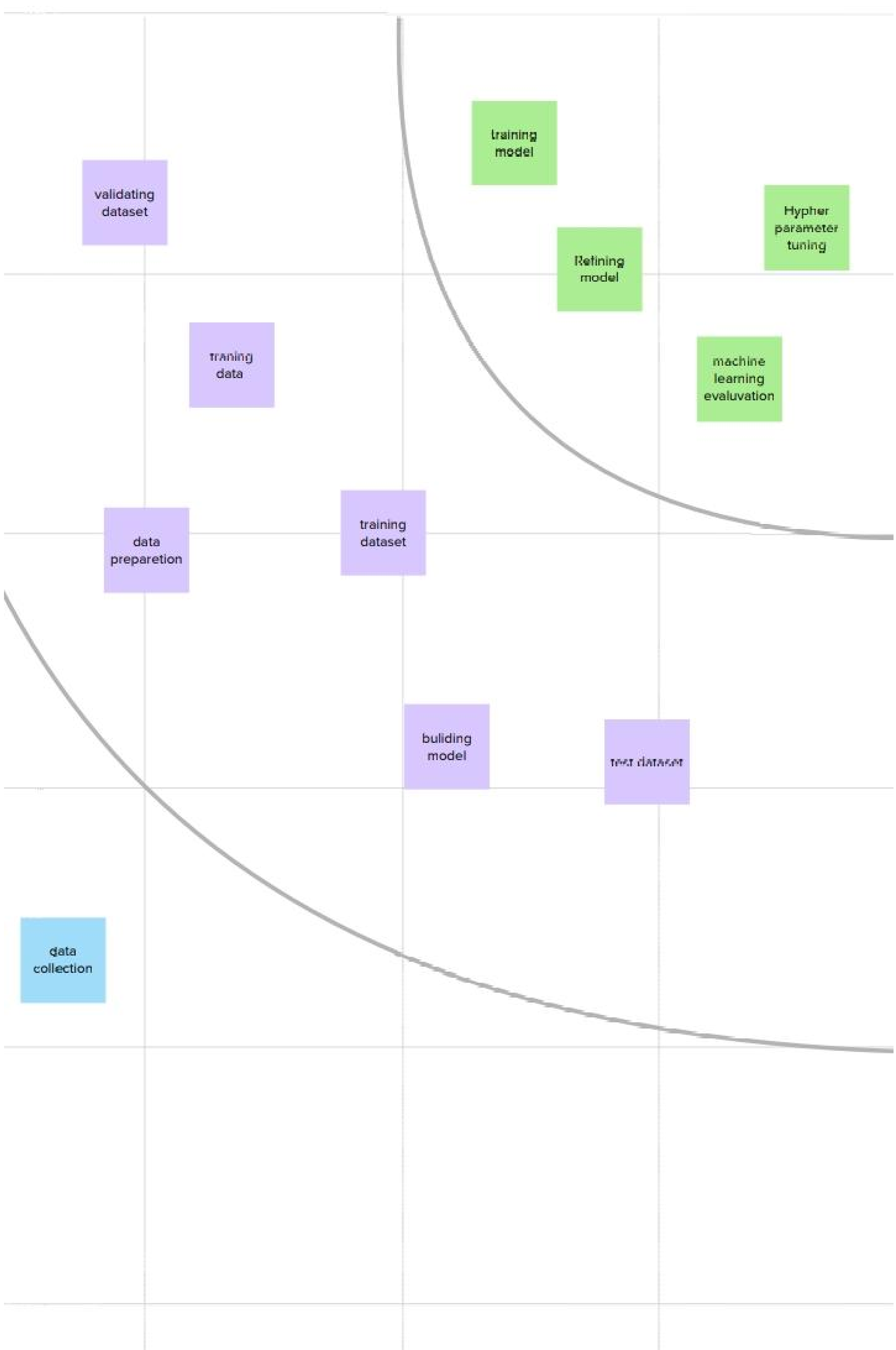
20 minutes

Importance

If each of these tasks could get done without any difficulty or cost, which would have the most positive impact?

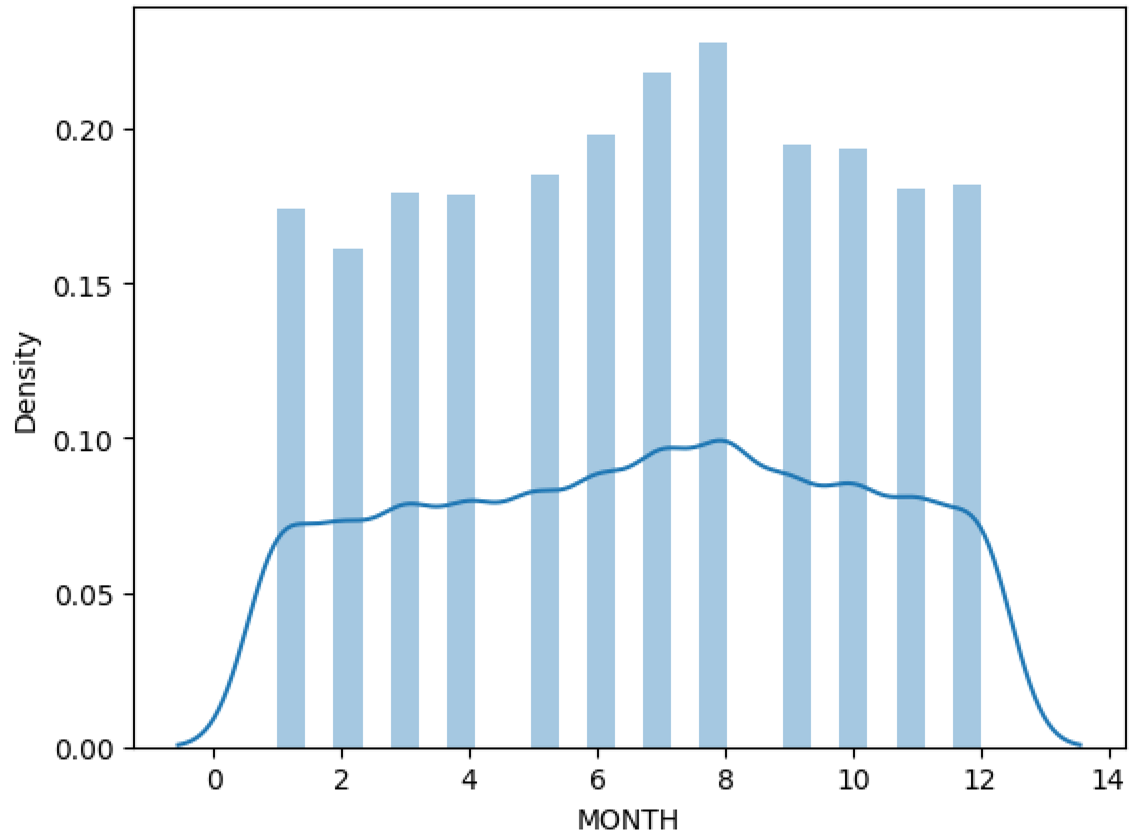
TIP

Participants can use their cursors to point at where sticky notes should go on the grid. The facilitator can confirm the spot by using the laser pointer holding the **H key** on the keyboard.

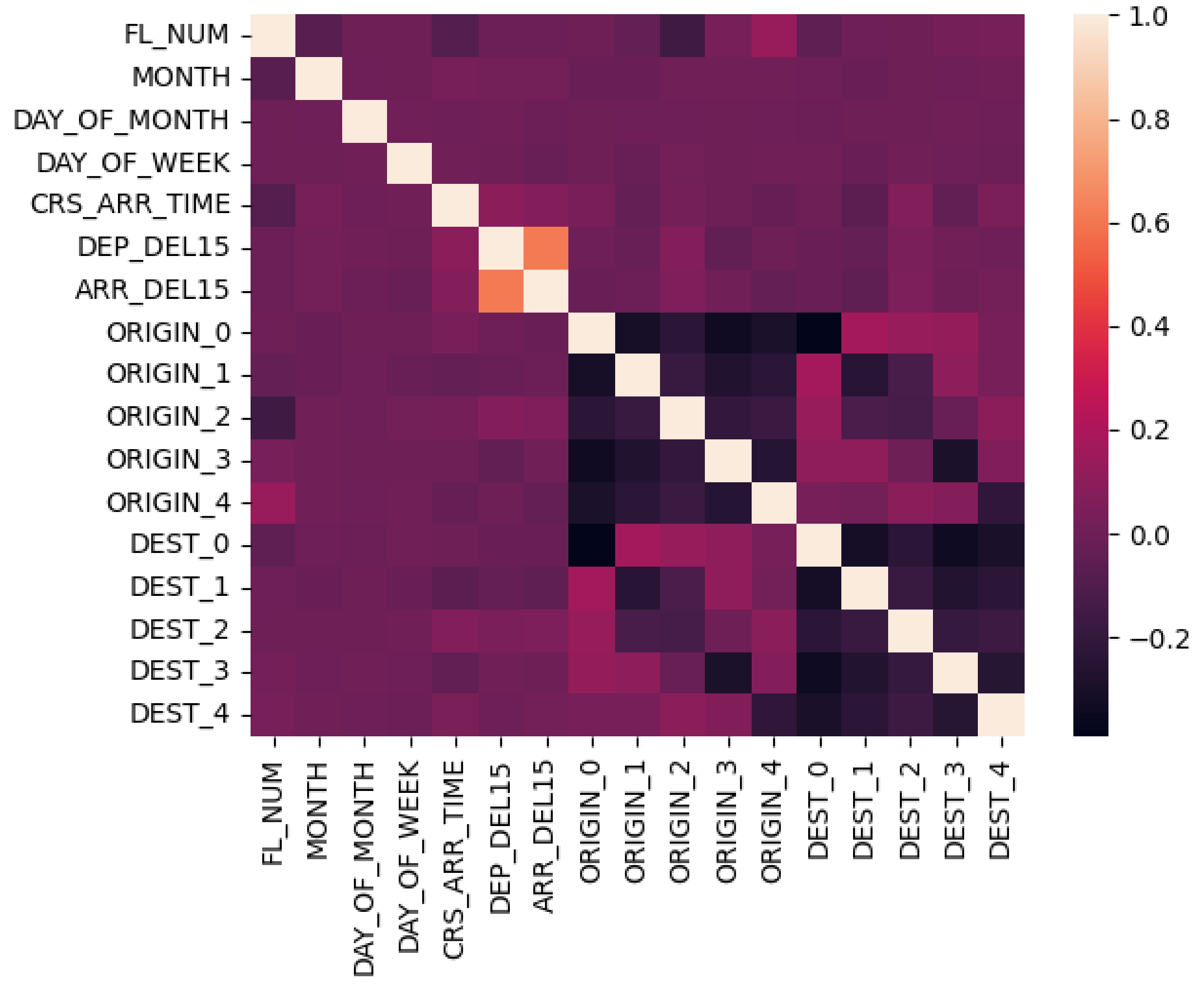

training
model
validating
dataset
Refining
model
Hypher
parameter
tuning
traning
data
machine
learning
evaluvation
data
preparetion
training
dataset
buliding
model
data
collection

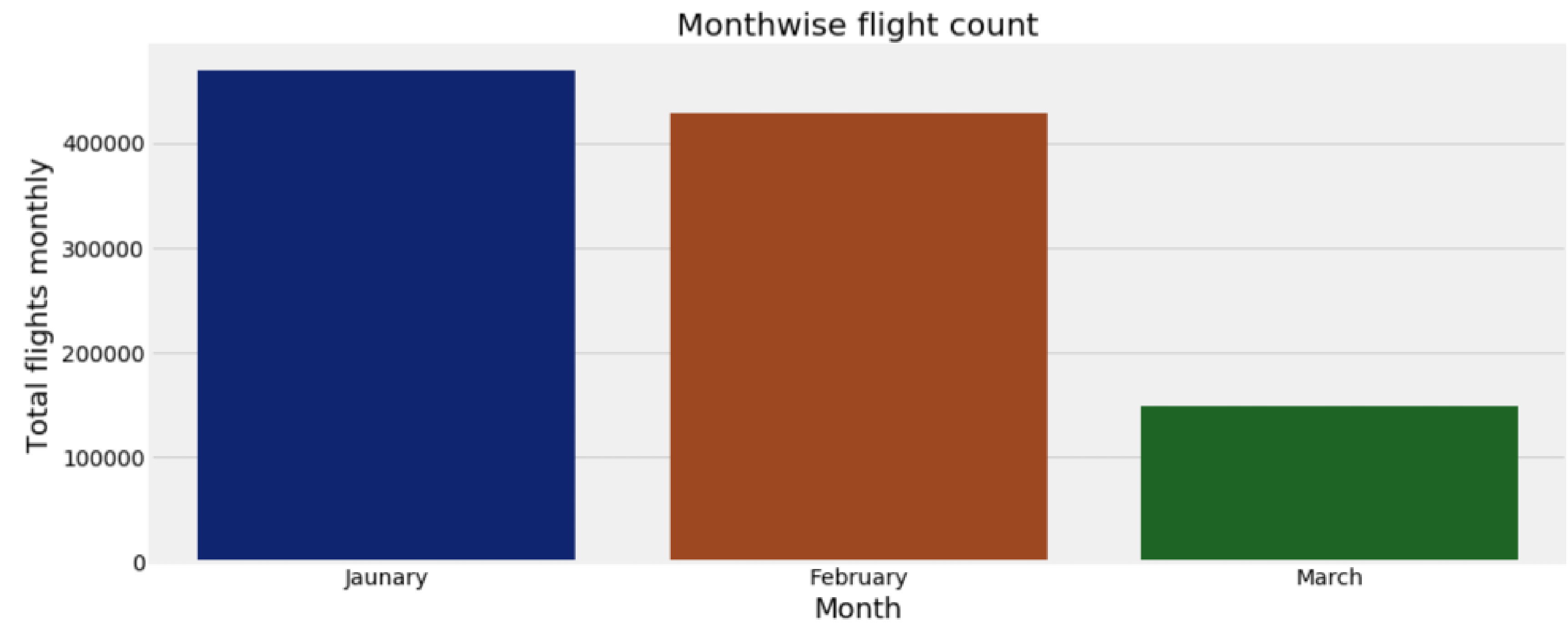
3.RESULT

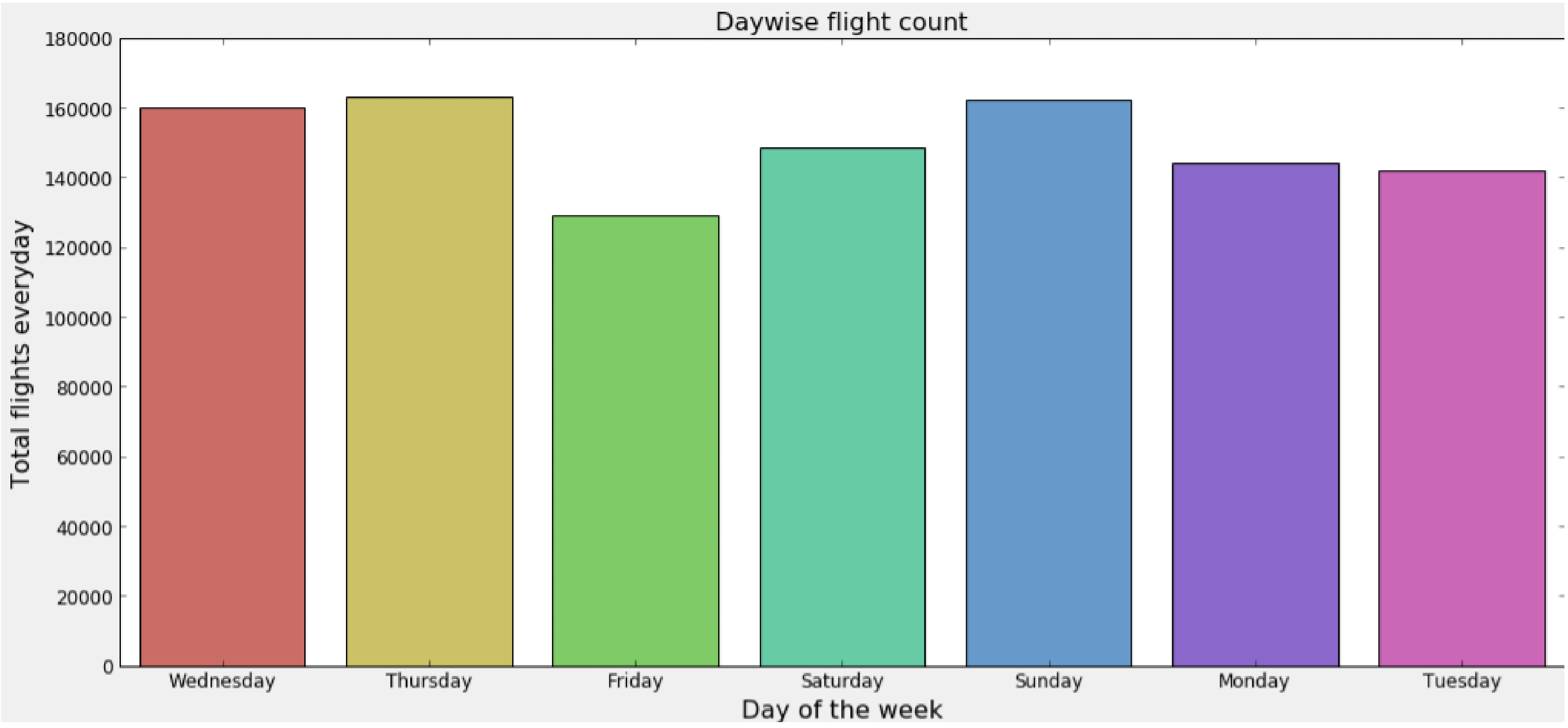
sns.distplot(dataset.MONTH)
<Axes: xlabel='MONTH', ylabel='Density'>

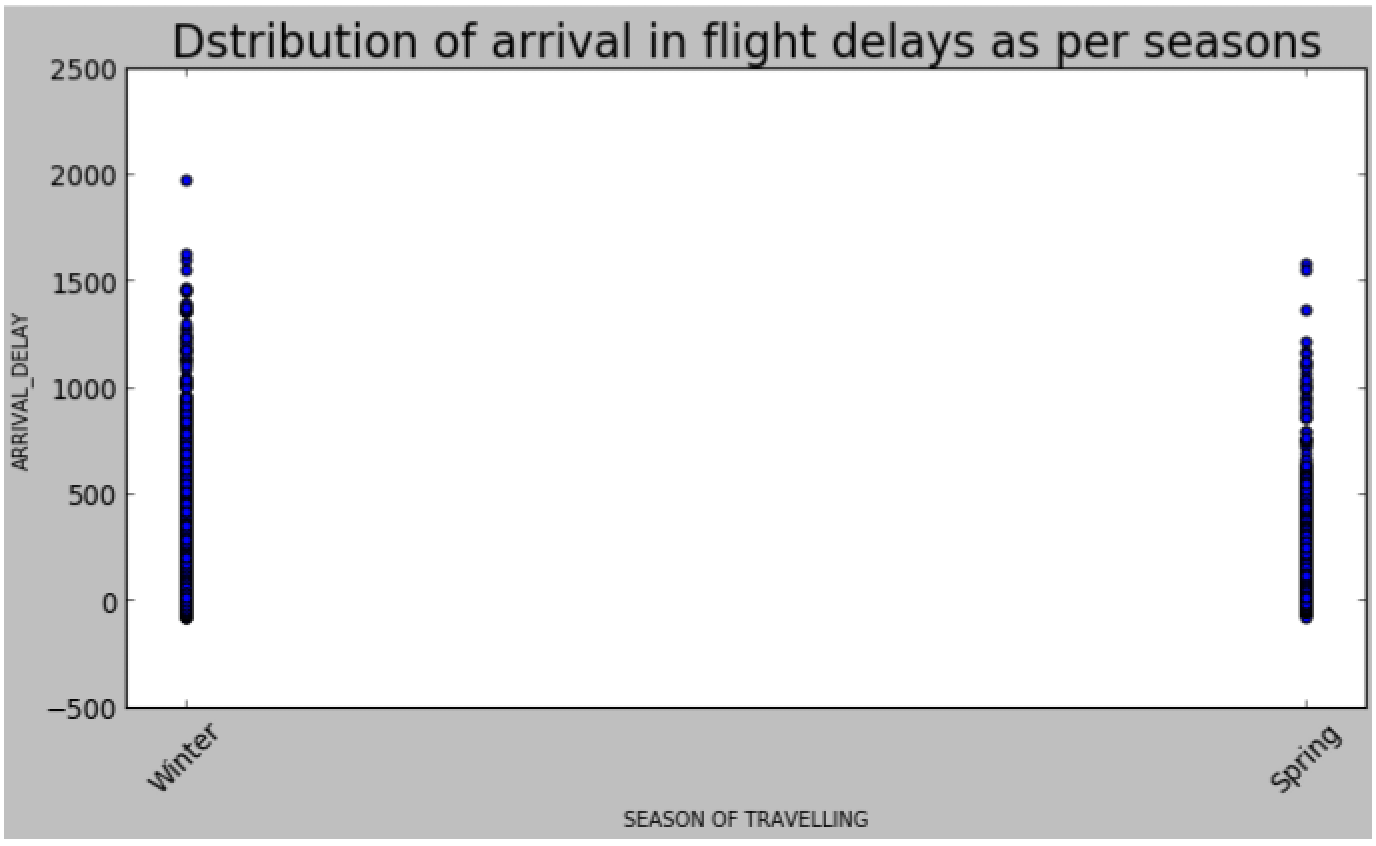


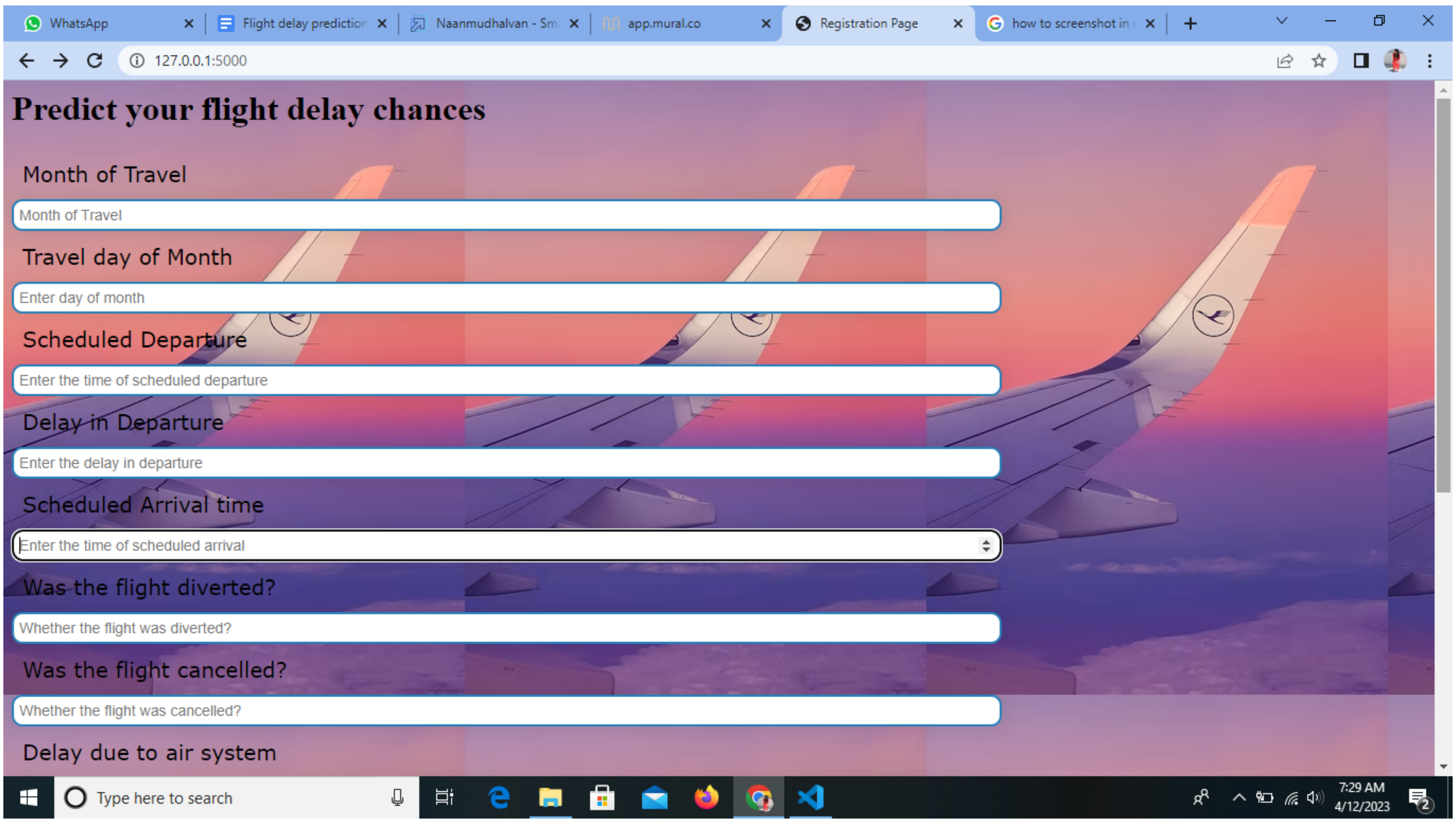
<Axes: >

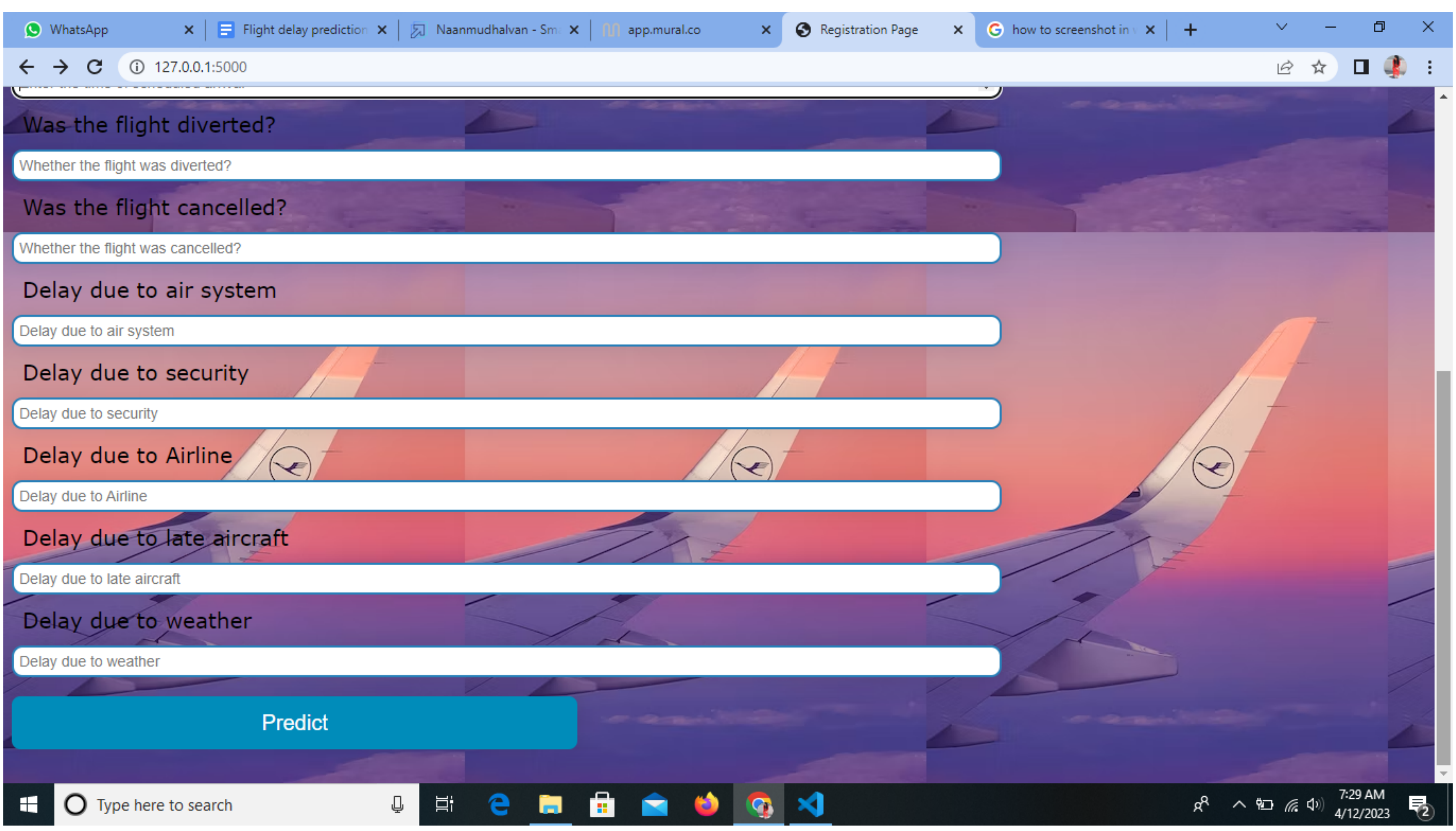



Monthwise flight count
Total flights monthly
0
100000
200000
300000
400000
Jaunary
February
March
Month


Daywise flight count
Total flights everyday
0
20000
40000
60000
80000
100000
120000
140000
160000
180000
Wednesday
Thursday
Friday
Saturday
Sunday
Monday
Tuesday
Day of the week


Dstribution of arrival in flight delays as per seasons
ARRIVAL_DELAY
−500
0
500
1000
1500
2000
2500
Winter
Spring
SEASON OF TRAVELLING





RESULTS

	YEAR	QUARTER	MONTH	DAY_OF_MONTH	DAY_OF_WEEK	UNIQUE_CARRIER	TAIL_NUM	FL_NUM	ORIGIN_AIRPORT_ID	ORIGIN	...	CRS_ARR_TIME	ARR_TIME	ARR_DELAY	ARR_DEL15	CANCELLED	DIVERTED	CRS_ELAPSED_TIME	ACTUAL_ELAPSED_TIME	DISTANCE	Unnamed: 25
0	2016	1	1	1	5	DL	N836	1399	103	ATL	...	2143	2102.0	-41.0	0.0	0.0	0.0	338.0	295.0	2182.0	NaN

							DN		97												
1	2016	1	1	1	5	DL	N964DN	1476	11433	DTW	...	1435	1439.0	4.0	0.0	0.0	0.0	110.0	115.0	528.0	NaN
2	2016	1	1	1	5	DL	N813DN	1597	10397	ATL	...	1215	1142.0	-33.0	0.0	0.0	0.0	335.0	300.0	2182.0	NaN
3	2016	1	1	1	5	DL	N587NW	1768	14747	SEA	...	1335	1345.0	10.0	0.0	0.0	0.0	196.0	205.0	1399.0	NaN
4	2016	1	1	1	5	DL	N836DN	1823	14747	SEA	...	607	615.0	8.0	0.0	0.0	0.0	247.0	259.0	1927.0	NaN

5 rows × 26 columns

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 11231 entries, 0 to 11230
Data columns (total 26 columns):
 #   Column               Non-Null Count  Dtype
---  ------               --------------  -----
 0   YEAR                 11231 non-null  int64
 1   QUARTER              11231 non-null  int64
 2   MONTH                11231 non-null  int64
 3   DAY_OF_MONTH         11231 non-null  int64
 4   DAY_OF_WEEK          11231 non-null  int64
 5   UNIQUE_CARRIER       11231 non-null  object
 6   TAIL_NUM             11231 non-null  object
 7   FL_NUM               11231 non-null  int64
 8   ORIGIN_AIRPORT_ID    11231 non-null  int64
 9   ORIGIN               11231 non-null  object
 10  DEST_AIRPORT_ID      11231 non-null  int64
 11  DEST                 11231 non-null  object
 12  CRS_DEP_TIME         11231 non-null  int64
 13  DEP_TIME             11124 non-null  float64
 14  DEP_DELAY            11124 non-null  float64
 15  DEP_DEL15            11124 non-null  float64
 16  CRS_ARR_TIME         11231 non-null  int64
 17  ARR_TIME             11116 non-null  float64
 18  ARR_DELAY            11043 non-null  float64
 19  ARR_DEL15            11043 non-null  float64
 20  CANCELLED            11231 non-null  float64
 21  DIVERTED             11231 non-null  float64
 22  CRS_ELAPSED_TIME     11231 non-null  float64
 23  ACTUAL_ELAPSED_TIME  11043 non-null  float64
 24  DISTANCE             11231 non-null  float64
 25  Unnamed: 25          0 non-null      float64
dtypes: float64(12), int64(10), object(4)
```

memory usage: 2.2+ MB

0s

```
dataset=dataset.drop('Unnamed: 25',axis=1)

dataset.isnull().sum()
```

```
YEAR                 0
QUARTER              0
MONTH                0
DAY_OF_MONTH         0
DAY_OF_WEEK          0
UNIQUE_CARRIER       0
TAIL_NUM             0
FL_NUM               0
ORIGIN_AIRPORT_ID    0
ORIGIN               0
DEST_AIRPORT_ID      0
DEST                 0
CRS_DEP_TIME         0
DEP_TIME           107
DEP_DELAY          107
DEP_DEL15          107
CRS_ARR_TIME         0
ARR_TIME           115
ARR_DELAY          188
ARR_DEL15          188
CANCELLED            0
DIVERTED             0
CRS_ELAPSED_TIME     0
ACTUAL_ELAPSED_TIME  188
DISTANCE             0
dtype: int64
```

```
FL_NUM             0
MONTH              0
DAY_OF_MONTH       0
DAY_OF_WEEK        0
ORIGIN             0
DEST               0
CRS_ARR_TIME       0
DEP_DEL15        107
ARR_DEL15        188
dtype: int64
```

```
array([0, 1, 4, 3, 2])
```

```
array([[1.399e+03, 1.000e+00, 1.000e+00, ..., 0.000e+00, 0.000e+00,
        1.000e+00],
       [1.476e+03, 1.000e+00, 1.000e+00, ..., 0.000e+00, 0.000e+00,
        0.000e+00],
       [1.597e+03, 1.000e+00, 1.000e+00, ..., 0.000e+00, 0.000e+00,
        1.000e+00],
       ...,
       [1.823e+03, 1.200e+01, 3.000e+01, ..., 0.000e+00, 0.000e+00,
        0.000e+00],
       [1.901e+03, 1.200e+01, 3.000e+01, ..., 0.000e+00, 0.000e+00,
        1.000e+00],
       [2.005e+03, 1.200e+01, 3.000e+01, ..., 0.000e+00, 0.000e+00,
        1.000e+00]])
```

```
array([[0., 0., 0., ..., 1., 0., 0.],
       [0., 0., 0., ..., 0., 0., 0.],
       [0., 0., 0., ..., 0., 0., 0.],
       ...,
       [0., 0., 0., ..., 0., 1., 0.],
```

```
       [0., 0., 0., ..., 0., 0., 0.],
       [0., 0., 0., ..., 0., 0., 0.]])
array([[1., 0.],
       [1., 0.],
       [1., 0.],
       ...,
       [1., 0.],
       [1., 0.],
       [1., 0.]])
(2247, 8)

(8984, 8)

(2247, 1)

(8984, 1)

array([1, 0, 0, ..., 1, 0, 0], dtype=uint8)

0.9893190921228304

array([[0.9984446 ],
       [0.00276279],
       [0.        ],
       ...,
       [0.6241991 ],
       [0.        ],
       [0.96088314]], dtype=float32)
```

---

[55]

0s

```
y_pred = (y_pred > 0.5)

y_pred

array([[0.9984446 ],

       [0.00276279],

       [0.        ],

       ...,

       [0.6241991 ],

       [0.        ],
       [0.96088314]], dtype=float32)
```

```
y_pred = (y_pred > 0.5)

y_pred

array([[ True],

    [False],

    [False],

    ...,

    [ True],

    [False],
    [ True]])
```

```
array([[1773,  29],
    [ 234,  211]])
```

```
0.9893190921228304
```

```
flights=pd.read_csv('flights.csv',low_memory=True)
flights
```

	YEAR	MONTH	DAY	DAY_OF_WEEK	AIRLINE	FLIGHT_NUMBER	TAIL_NUMBER	ORIGIN_AIRPORT	DESTINATION_AIRPORT	SCHEDULED_DEPARTURE	...	ARRIVAL_TIME	ARRIVAL_DELAY	DIVERTED	CANCELLED	CANCELLATION_REASON	AIR_SYSTEM_DELAY	SECURITY_DELAY	AIRLINE_DELAY	LATE_AIRCRAFT_DELAY	WEATHER_DELAY
0	2015	1	1	4	AS	98	N407AS	ANC	SEA	5	...	408.0	-22.0	0	0	NaN	NaN	NaN	NaN	NaN	NaN
1	2015	1	1	4	AA	2336	N3KUAA	LAX	PBI	10	...	741.0	-9.0	0	0	NaN	NaN	NaN	NaN	NaN	NaN
2	2015	1	1	4	US	840	N171US	SFO	CLT	20	...	811.0	5.0	0	0	NaN	NaN	NaN	NaN	NaN	NaN
3	2015	1	1	4	AA	258	N3HYAA	LAX	MIA	20	...	756.0	-9.0	0	0	NaN	NaN	NaN	NaN	NaN	NaN
4	2015	1	1	4	AS	135	N527AS	SEA	ANC	25	...	259.0	-21.0	0	0	NaN	NaN	NaN	NaN	NaN	NaN
...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...
1048570	2015	3	10	2	EV	4122	N11191	RDU	EWR	1013	...	1133.0	-16.0	0	0	NaN	NaN	NaN	NaN	NaN	NaN
1048571	2015	3	10	2	UA	1018	N79279	LGA	IAH	1013	...	1335.0	-2.0	0	0	NaN	NaN	NaN	NaN	NaN	NaN
1048572	2015	3	10	2	UA	1260	N76508	SAN	ORD	1013	...	1627.0	3.0	0	0	NaN	NaN	NaN	NaN	NaN	NaN
1048573	2015	3	10	2	EV	4349	N14158	MSY	ORD	1013	...	1229.0	-13.0	0	0	NaN	NaN	NaN	NaN	NaN	NaN
104	20	3	10	2	MQ	2916	N539MQ	CID	ORD	10	...	NaN	NaN	0	1	B	NaN	NaN	NaN	NaN	NaN

8 5 7 4	1 5	1 3

1048575 rows × 31 columns

```
airports=pd.read_csv('airports.csv')
airports
```

IATA_CODE	AIRPORT	CITY	STATE	COUNTRY	LATITUDE	LONGITUDE	
0	ABE	Lehigh Valley International Airport	Allentown	PA	USA	40.65236	-75.44040
1	ABI	Abilene Regional Airport	Abilene	TX	USA	32.41132	-99.68190
2	ABQ	Albuquerque International Sunport	Albuquerque	NM	USA	35.04022	-106.60919
3	ABR	Aberdeen Regional Airport	Aberdeen	SD	USA	45.44906	-98.42183
4	ABY	Southwest Georgia Regional Airport	Albany	GA	USA	31.53552	-84.19447
...	...	...	...	...	...	...	...
317	WRG	Wrangell Airport	Wrangell	AK	USA	56.48433	-132.36982
318	WYS	Westerly State Airport	West Yellowstone	MT	USA	44.68840	-111.11764
319	XNA	Northwest Arkansas Regional Airport	Fayetteville/Springdale/Rogers	AR	USA	36.28187	-94.30681
320	YAK	Yakutat Airport	Yakutat	AK	USA	59.50336	-139.66023
321	YUM	Yuma International Airport	Yuma	AZ	USA	32.65658	-114.60597

322 rows × 7 columns

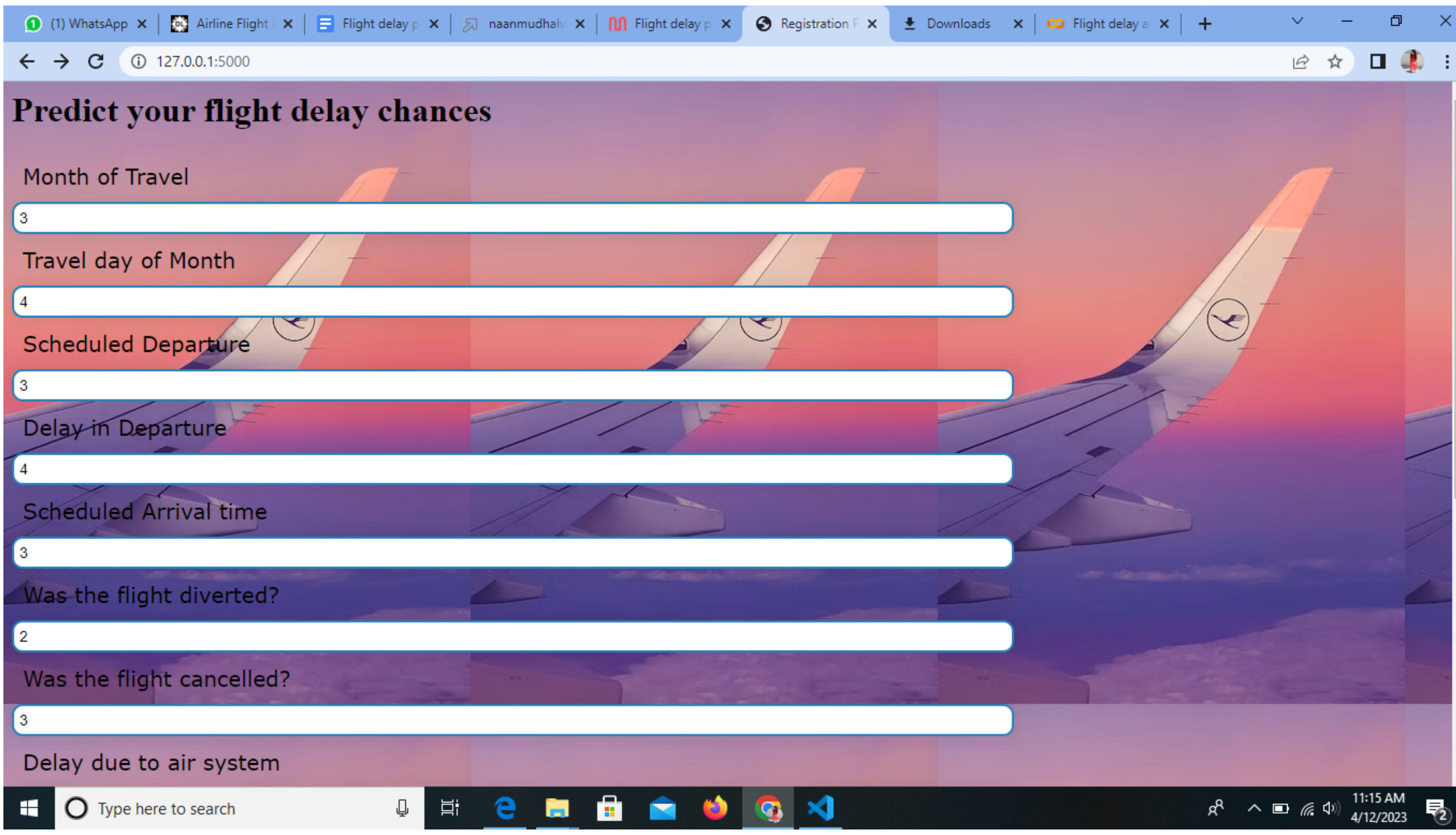
```
airlines=pd.read_csv('airlines.csv')
airlines
```

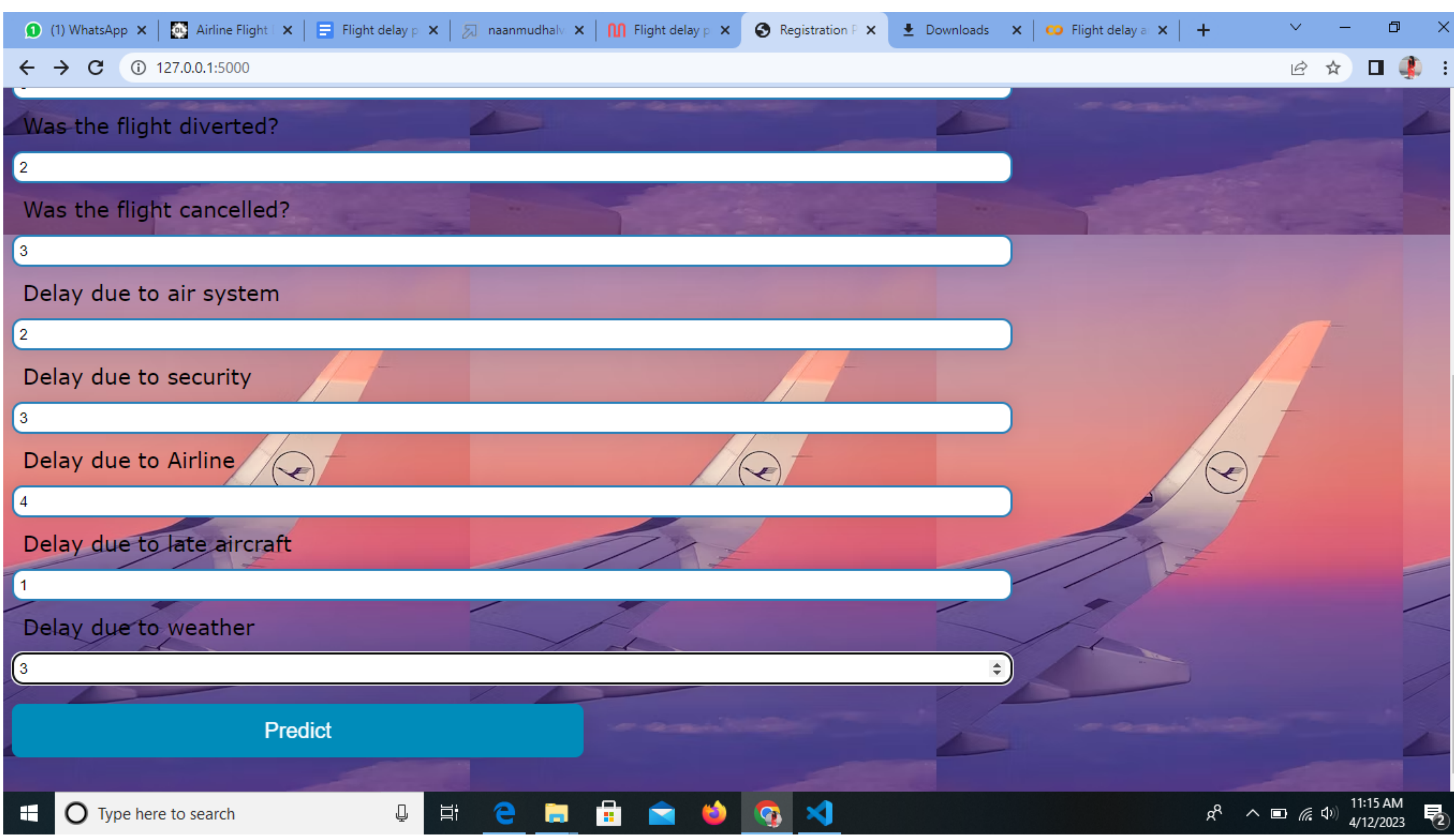
IATA_CODE	AIRLINE	
0	UA	United Air Lines Inc.
1	AA	American Airlines Inc.
2	US	US Airways Inc.
3	F9	Frontier Airlines Inc.
4	B6	JetBlue Airways
5	OO	Skywest Airlines Inc.
6	AS	Alaska Airlines Inc.
7	NK	Spirit Air Lines
8	WN	Southwest Airlines Co.
9	DL	Delta Air Lines Inc.

10	EV	Atlantic Southeast Airlines
11	HA	Hawaiian Airlines Inc.
12	MQ	American Eagle Airlines Inc.
13	VX	Virgin America

IATA_CODE 0

```
airports.isnull().sum()
IATA_CODE     0
AIRPORT       0
CITY          0
STATE         0
COUNTRY       0
LATITUDE      3
LONGITUDE     3
dtype: int64

airports.dtypes

IATA_CODE      object
AIRPORT        object
CITY           object
STATE          object
COUNTRY        object
LATITUDE      float64
LONGITUDE     float64
dtype: object

columns=airports.loc[:,['LATITUDE','LONGITUDE']]

airports.dropna(inplace=True)

airports.isnull().sum()

IATA_CODE     0
AIRPORT       0
CITY          0
STATE         0
COUNTRY       0
LATITUDE      0
LONGITUDE     0
dtype: int64

array([[1773,  29],
   [ 234, 211]])

array([[1719,   83],
        [ 159,  286]])
```


(1) WhatsApp
Airline Flight
Flight delay
naanmudhalv
Flight delay
Registration
Downloads
Flight delay
127.0.0.1:5000
Predict your flight delay chances
Month of Travel
3
Travel day of Month
4
Scheduled Departure
3
Delay in Departure
4
Scheduled Arrival time
3
Was the flight diverted?
2
Was the flight cancelled?
3
Delay due to air system
Type here to search
11:15 AM
4/12/2023


(1) WhatsApp
Airline Flight
Flight delay
naanmudhalv
Flight delay
Registration
Downloads
Flight delay
127.0.0.1:5000
Was the flight diverted?
2
Was the flight cancelled?
3
Delay due to air system
2
Delay due to security
3
Delay due to Airline
4
Delay due to late aircraft
1
Delay due to weather
3
Predict
Type here to search
11:15 AM
4/12/2023

4.ADVANTAGES & DISADVANTAGES

ADVANTAGES AND DISADVANTAGES

Advantages of Flight Delay Prediction using Machine Learning:

1. Improved accuracy: Machine learning algorithms can process large amounts of data and learn from patterns, which can improve the accuracy of flight delay predictions.
2. Real-time predictions: Machine learning models can be trained on real-time data, enabling airlines to make informed decisions about delays as they happen.
3. Cost savings: By predicting flight delays, airlines can optimize their resources, reduce costs associated with operational inefficiencies, and avoid potential financial losses from missed connections or cancellations.
4. Improved customer satisfaction: Accurate prediction of flight delays can allow airlines to communicate more effectively with passengers, reducing frustration and improving overall customer satisfaction.
5. Competitive advantage: Airlines that offer accurate flight delay predictions can differentiate themselves from their competitors and attract more customers.

Disadvantages of Flight Delay Prediction using Machine Learning:

1. Data quality: The accuracy of machine learning predictions is highly dependent on the quality of the data used to train the algorithms. If the data is incomplete or inaccurate, the model's predictions may be unreliable.
2. Complexity: Machine learning algorithms can be complex and require significant computational resources to train and run, making it challenging for airlines to develop and deploy these systems.
3. Overfitting: Machine learning models can overfit to the training data, leading to inaccurate predictions on new data. This can be mitigated by using proper training and testing techniques.
4. Interpretability: Machine learning models can be difficult to interpret, which can make it challenging to identify the factors contributing to flight delays and make informed decisions.
5. Ethical concerns: Machine learning models may inadvertently perpetuate bias if the training data is not representative of the entire population. This can lead to unfair outcomes and raise ethical concerns.

In summary, flight delay prediction using machine learning offers several advantages, including improved accuracy, real-time predictions, cost savings, improved customer satisfaction, and a competitive advantage. However, it also comes with challenges, including data quality, complexity, overfitting, interpretability, and ethical concerns that need to

be addressed to ensure reliable and unbiased predictions.

5.APPLICATIONS

APPLICATIONS

There are several applications of flight delay prediction using machine learning. Some of the most common applications include:

1. **Flight operations management: Machine learning models can be used to predict flight delays, enabling airlines to optimize their resources, reduce costs, and improve operational efficiency.**
2. **Passenger information and communication: Accurate predictions of flight delays can allow airlines to communicate more effectively with passengers, reducing frustration and improving overall customer satisfaction.**
3. **Resource allocation: By predicting flight delays, airlines can allocate resources more effectively, such as gate and crew assignments, to minimize the impact of delays.**
4. **Maintenance and repair scheduling: Machine learning models can predict when aircraft will require maintenance or repairs, allowing airlines to schedule maintenance and repairs in advance and minimize the impact on operations.**
5. **Revenue management: Accurate predictions of flight delays can enable airlines to adjust pricing and optimize their revenue management strategies, such as overbooking, to maximize profits.**
6. **Route optimization: By predicting flight delays, airlines can optimize their route planning, avoiding congestion and minimizing the impact of weather conditions.**
7. **Safety management: Machine learning models can be used to predict safety-related incidents, such as turbulence or air traffic control disruptions, enabling airlines to take proactive measures to ensure the safety of passengers and crew.**

In summary, flight delay prediction using machine learning has a wide range of applications, from flight operations management to revenue management and safety management. Accurate predictions can help airlines optimize their resources, reduce costs, improve customer satisfaction, and maintain a competitive advantage in the industry.

Flight delay prediction using machine learning can be applied in various areas, such as:

1. **Airline Operations: Accurate predictions of flight delays can enable airlines to optimize their operations, minimize delays, and improve overall efficiency. This includes crew scheduling, resource allocation, and route planning.**
2. **Airport Management: Machine learning can be used to predict flight delays and congestion, allowing airport authorities to optimize their resources, improve ground handling, and reduce wait times.**
3. **Air Traffic Management: Machine learning can be used to predict air traffic congestion and delays, enabling air traffic controllers to manage traffic flow and reduce delays.**
4. **Aviation Safety: Accurate predictions of safety-related incidents, such as turbulence or severe weather conditions, can enable airlines to take proactive measures to ensure the safety of passengers and crew.**
5. **Passenger Experience: Accurate predictions of flight delays can allow airlines to communicate more effectively with passengers, reducing frustration and improving overall customer satisfaction.**
6. **Revenue Management: Machine learning can be used to predict flight delays and optimize revenue management strategies, such as overbooking, to maximize profits.**
7. **Maintenance and Repair: Machine learning can predict when aircraft will require maintenance or repairs, enabling airlines to schedule maintenance and repairs in advance and minimize the impact on operations.**

In summary, flight delay prediction using machine learning has numerous applications in various areas of the aviation industry, from airline operations to airport management, air traffic management, aviation safety, passenger

experience, revenue management, and maintenance and repair. Accurate predictions can help improve operational efficiency, reduce costs, and maintain a competitive advantage in the industry.

6.CONCLUSION

CONCLUSION

In conclusion, flight delay prediction using machine learning has become an essential tool for airlines and the aviation industry in general. Accurate predictions can enable airlines to optimize their operations, improve resource allocation, minimize delays, improve customer satisfaction, and reduce costs.

However, the use of machine learning models in flight delay prediction comes with challenges such as data quality, complexity, overfitting, interpretability, and ethical concerns that need to be addressed to ensure reliable and unbiased predictions.

Despite the challenges, flight delay prediction using machine learning has numerous applications in various areas of the aviation industry, such as airline operations, airport management, air traffic management, aviation safety, passenger experience, revenue management, and maintenance and repair.

Overall, flight delay prediction using machine learning is a promising technology that can improve the efficiency and safety of air travel while enhancing the customer experience. As the technology continues to evolve, it will become increasingly important for airlines to adopt and leverage these models to remain competitive in the industry.

The paper performed a prediction of the occurrence of flight delays by adapting it into a machine learning problem. A supervised machine learning approach in the form of binary classification was used for the prediction. Seven algorithms were used for delay prediction, and four measures were used for algorithms performance evaluation. Due to the imbalanced nature of the data set, evaluation measures were weighted to eliminate the dominant effect of non-delayed flights over delayed flights. After applying classifiers to the delay prediction, the values of their four measures were compared to evaluate the performance of each model.

7.FUTURE SCOPE

FUTURE SCOPE

The future scope for flight delay prediction using machine learning is vast and promising. Here are some potential future developments:

1. **Integration with other data sources: Machine learning models can be trained on a wide range of data sources, including weather, air traffic control, and social media data, to improve the accuracy of predictions.**
2. **Use of real-time data: The use of real-time data, such as data from sensors on aircraft and airports, can enable machine learning models to provide more accurate and up-to-date predictions.**
3. **Improved interpretability: The development of more interpretable machine learning models can enhance the understanding of predictions, enabling airlines to make better-informed decisions.**
4. **Adoption of explainable AI: Explainable AI can provide clear explanations of how machine learning models make predictions, enabling airlines to understand the rationale behind predictions and improve transparency.**
5. **Use of unsupervised learning: Unsupervised learning can be used to identify new patterns and anomalies in data, enabling airlines to detect and prevent delays before they occur.**
6. **Integration with other systems: Integration with other systems, such as flight planning and scheduling software, can enable airlines to automate decision-making processes and improve overall efficiency.**
7. **Adoption of ethical considerations: The development of ethical frameworks for machine learning models can help ensure that predictions are fair, unbiased, and do not discriminate against specific groups.**

In summary, the future scope for flight delay prediction using machine learning is promising, with potential developments in data integration, real-time data, interpretability, explainable AI, unsupervised learning, system integration, and ethical considerations. These developments can help airlines optimize their operations, improve customer satisfaction, and maintain a competitive advantage in the industry.

8.APPENDIX

A.SOURCE CODE

```
import gmplot
latitudes=airports.loc[:,'LATITUDE']
longitudes=airports.loc[:,'LONGITUDE']
gmap=gmplot.GoogleMapPlotter(35,102,2)
gmap.scatter(latitudes,longitudes,'red',size=5)
gmap.draw('map/gmplot.html')


from IPython.display import IFrame
IFrame(src='map/gmplot.html',width=900, height=600)

from flask import Flask,render_template,request
import pickle
app=Flask(__name__)

@app.route('/',methods=['GET'])
def HomePage():
 return render_template('index.html')


@app.route('/predict',methods=['GET','POST'])
def index():
 if request.method=='POST':
  try:
   month = int(request.form['month'])
```

```
    day    = int(request.form['day'])
    schdl_dep = float(request.form['schdl_dep'])
    dep_delay = float(request.form['dep_delay'])
    schdl_arriv = float(request.form['schdl_arriv'])
    divrtd = int(request.form['divrtd'])
    cancld = int(request.form['cancld'])
    air_sys_delay = float(request.form['air_sys_delay'])
    secrty_delay = float(request.form['secrty_delay'])
    airline_delay = float(request.form['airline_delay'])
    late_air_delay = float(request.form['late_air_delay'])
    wethr_delay  = float(request.form['wethr_delay'])

    print('HI')
    filename = 'finalized_model.sav'
    loaded_model = pickle.load(open(filename, 'rb'))


    import numpy as np

prediction=loaded_model.predict([[month,day,schdl_dep,dep_delay,schdl_arriv,divrtd,cancld,air_sys_dela
y,secrty_delay,
                                     airline_delay,late_air_delay,wethr_delay]])
    for i in prediction:
     if i==1:
      prediction='will be'
     else:
      prediction='wont get'


    return render_template('results.html',prediction=prediction)
   except Exception as e:
     print('The Exception message is: ',e)
     return 'something is wrong'
  else:
   return render_template('index.html'),

if __name__ == "__main__":
    app.run()

<html>
<head>
    <title>Registration Page</title>
    <link rel="stylesheet" href="{{url_for('static', filename='/main.css')}}" type="text/css">
</head>
<body style="background-image:
url(https://images.unsplash.com/photo-1634981239781-b313c21d100e?ixlib=rb-4.0.3&ixid=MnwxMjA3fDB8MHxwa
G90by1wYWdlfHx8fGVufDB8fHx8&auto=format&fit=crop&w=435&q=80);">
    <h1>Predict your flight delay chances</h1>

        <form action="/predict" method="POST">
            <label>Month of Travel</label>
            <input type="number" name="month" id="month" placeholder="Month of Travel">

            <label>Travel day of Month</label>
            <input type="number" name="day" id="day" placeholder="Enter day of month ">

            <label>Scheduled Departure</label>
            <input type="number" name="schdl_dep" id="schdl_dep" placeholder="Enter the time of
scheduled departure">

            <label>Delay in Departure</label>
            <input type="number" name="dep_delay" id="dep_delay" placeholder="Enter the delay in
departure" step=".01">

            <label>Scheduled Arrival time</label>
            <input type="number" name="schdl_arriv" id="schdl_arriv" placeholder="Enter the time of
scheduled arrival">
```

```
            <label>Was the flight diverted?</label>
            <input type="number" name="divrtd" id="divrtd" placeholder="Whether the flight was
diverted?">

            <label>Was the flight cancelled?</label>
            <input type="number" name="cancld" id="cancld" placeholder="Whether the flight was
cancelled?">

            <label>Delay due to air system</label>
            <input type="number" name="air_sys_delay" id="air_sys_delay"placeholder="Delay due to air
system" step=".01">

            <label>Delay due to security</label>
            <input type="number" name="secrty_delay" id="secrty_delay" placeholder="Delay due to
security" step=".01">

            <label>Delay due to Airline</label>
            <input type="number" name="airline_delay" id="airline_delay" placeholder="Delay due to
Airline" step=".01">

            <label>Delay due to late aircraft</label>
            <input type="number" name="late_air_delay" id="late_air_delay"placeholder="Delay due to
late aircraft" step=".01">

            <label>Delay due to weather</label>
            <input type="number" name="wethr_delay" id="wethr_delay"placeholder="Delay due to weather"
step=".01">


            <input type="submit" value="Predict" class="sbt_btn">
        </form>
</body>
</html>

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

<head>
  <meta charset="UTF-8">
  <title>Result Page</title>

    <link rel="stylesheet" href="{{ url_for('static', filename='/main.css') }}">


</head>

<body style="background-image:
url(https://images.unsplash.com/photo-1634981239781-b313c21d100e?ixlib=rb-4.0.3&ixid=MnwxMjA3fDB8MHxwa
G90by1wYWdlfHx8fGVufDB8fHx8&auto=format&fit=crop&w=435&q=80);">
    <h1>Flight Delay Prediction</h1>
    <p>Your flight {{prediction}} delayed </p>
</body>
</html>
```