# Flight Delay Prediction

## 1.Introduction:

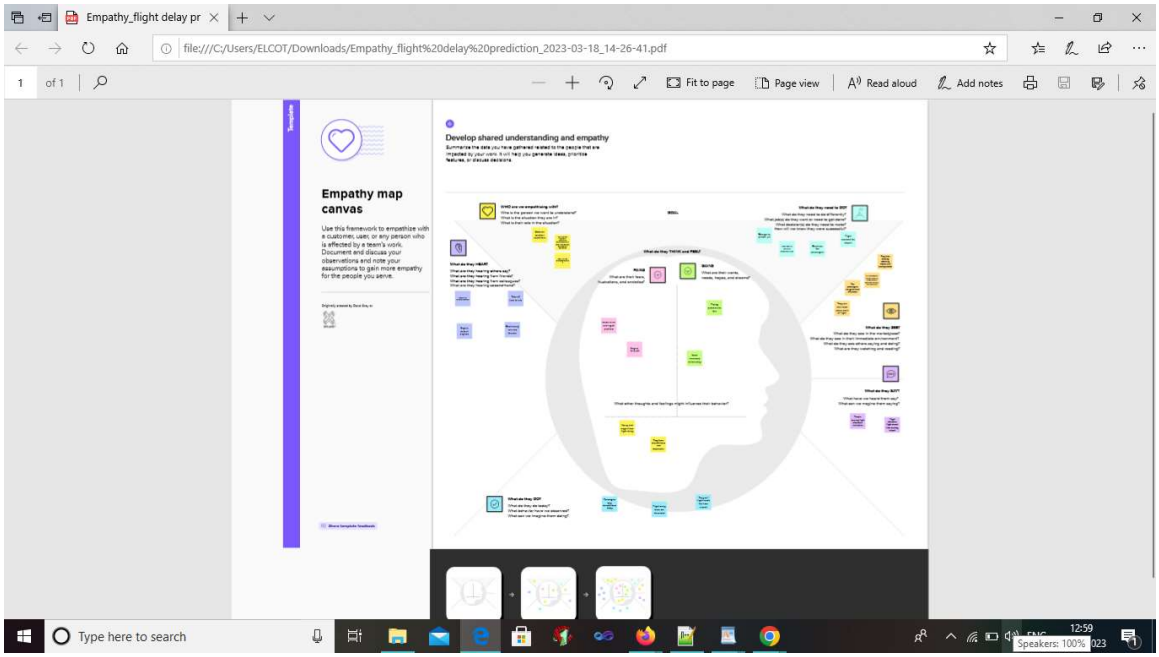
My flight was **delete as appropriate** delayed/cancelled and as a result I was delayed by (add number of hours) to my final destination.

purpose:

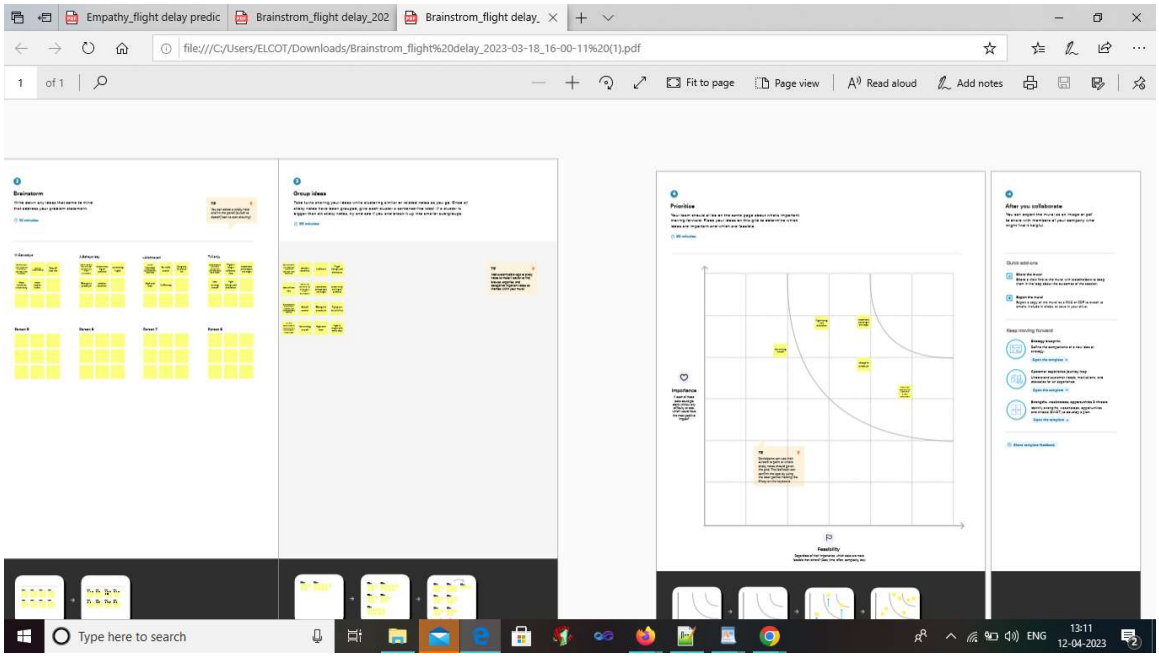
There are plenty of other reasons why passengers face flight delays, such as time for fueling, boarding passengers, aircraft cleaning, etc. Airlines allow for a little bit of flexibility, and it's important for passengers to understand such so they can fit some flexibility into their schedule, too.

## 2.Problem Definition & Design Thinking

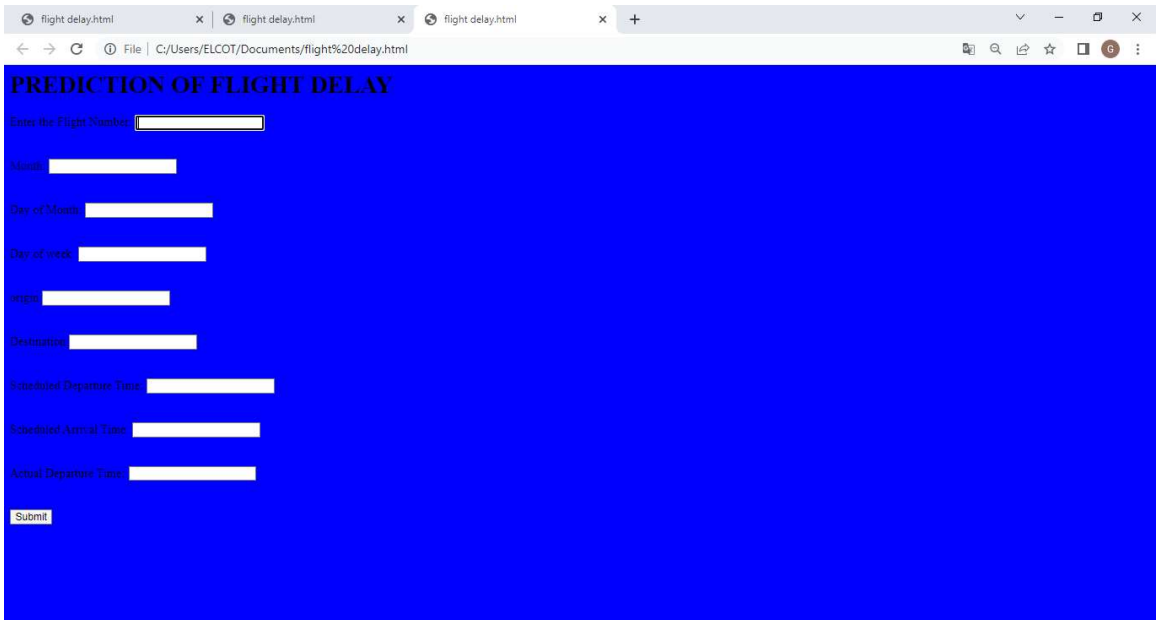
# Empathy map




# Ideation & Brainstorming Map



# 3.Result




# 4.Advantage

Functional cookies. These cookies are necessary for the website to function; for example for chatting with our employees.

Analytical cookies. Google Analytics collects anonymized data about surfing behavior. This way we can see how visitors use the website and improve the website based on

that.

Trackingcookies. These cookies are used by third parties to show customized EUclaim advertisements on third party websites based on previous interactions

Disadvantage:

One of the top reasons -- because of the weather conditions. It can be due to bad weather at the departure airport, on the way to your destination or at the destination. Weather is one of the top reasons why flights get delayed.

The most important thing to understand here is that your flight may get delayed or cancelled due to bad weather, even if it’s a perfect weather outside the window. And you may be even thinking why is my flight

delayed, if the weather is so nice? But that’s because the airline looks at the whole journey and makes sure that there isn’t something along the way that could endanger the airplane, crew or passengers.

Since flight delays due to weather conditions are out of the airline's control, it isn't considered their responsibility, if that causes any problems. And if your flight gets delayed or cancelled due to bad and unpredictable, severe weather patterns, you can't get compensation from the airline.

Security issues

There is a lot of things going on at the airport.

A lot of things you and I don’t know about,

all to make the airports as safe place as possible. And when the security personnel find out that something is possibly wrong, with for example, a piece of luggage loaded on your plane, they will demand it being unloaded.

And your flight may get delayed because of that

Technical issues with the aircraft

Things like these happen very often.

If during pre-flight inspection engineers find out that it isn’t safe for the particular aircraft to fly, your flight may get delayed or cancelled. What you should know is that mostly it doesn't fall under the term “extraordinary circumstances” and isn't considered something out of airline’s

control. Quite the opposite - in most situations technical problems are considered airline’s fault.

## Employee strikes

Strikes happen a lot in Europe.

In fact, airport and airline employee strikes is one of the top reasons why flights get delayed and cancelled in Europe. A few years ago Ryanair experienced its biggest employee strike to the day, tens of flights were cancelled affecting tens of thousands of passengers. In 2022 summer alone, there have been several similar strikes.

Compensation-wise airport employee strikes often are considered “extraordinary circumstances” and you can’t get compensation. But there is a "but". Not all

airport staff strikes are the same. If it's an airport employee strike, mostly you can't get compensation. But if it's an airline employee strike -- very often you can get compensation.

Often you can get both compensation and full refund.

Waiting for connecting passengers

It's something most travellers have experienced.

Mostly it causes only insignificant flight delays, like 20 or 30 minute delays. Unless you are travelling at peak times, when it may also cause a long flight delay. And if it's a long delay, you may also get compensation from airline.

Why do planes get delayed because of things like that? Don't they have to be running on time? It's up to your airline and the captain of the airplane to decide, whether it's worth waiting or not. If you are flying to or from Europe, or with a European airline, and if you reach your destination 3 or more hours later, then you may be entitled to flight delay compensation.

Airport congestion, or not enough staff

Yes, that's a thing lately.

With the number of air passengers and flights growing, airport congestion is a worldwide trend nowadays. And, if for example, captain of the plane decides to wait for connecting passengers during peak hours, it may lead to a rather long delay in

the end. Simply because it may not be possible to leave the airport that fast, when all the passengers have arrived. Because of other planes coming in and leaving the airport at the same time.

## 5.Application:

In order to verify the effectiveness and practicability of our method, we collect additional six-month data of flight operation in the first half of 2018. We will use the method of this article to adjust the flight planning and compare the number of delayed flights before and after adjustment for the first six months of 2018. From the previous statistical results, we know that the 10 airports with the busiest and the highest delay ratio are SFO, EWR, JFK, LGA, MIA, PBI, ORD, BOS, LAX and SEA. We assume that if one flight departs from one

of these 10 airports, it will generate newly formed delay and cause another flight immediately after it with the same aircraft also to delay. However, strictly speaking, the latter flight delay may be not merely attributed by a late arrival of the flight immediately preceding it, but also be attributed by one or more of other factors. In other words, sometimes the actual departure delay is hard to predict when we change the flight plan by our method, while the delay only caused by PF is not. Therefore, to simplify the prediction of current flight delays in the present work, we do not take into account the newly formed delay when the last flight by the same aircraft departed from one of the 10 highest delay ratio airports.

## 6.Conclusion:

By data mining and statistical analysis,

we study the distribution characteristics and inherent mechanism of flight departure delay for DL. From the statistical results, we find that the distribution of flight departure delay follows SPL, and when we eliminated effects of PF, the distribution of departure delay exhibit an obvious PL feature instead of SPL. The queue model which executes the highest-priority item on its list helps to understand mechanism of PL feature. We consider that the mechanism of SPL distribution of flight departure delay is the results of aircraft queue for take-off due to airports congestion and propagation delay caused by late-arriving aircraft.

Based on the above mechanism, we develop a specific measure to mitigate propagation delay without increasing operational costs. Specifically, if one aircraft

takes off from an airport with a higher delay ratio, we delayed the schedule departure time of the next flight operated by the same aircraft, which is equal to the schedule buffer time between the next flight and the subsequent flight. It is proved that our approach is pretty effective in reducing flight delay, although it is not significant for flights with larger delay.

In addition, our approach is based on the predictability of propagation delays and mathematical induction, which provides a new way to optimize flight schedules. Although this is by no means intended as a exhaustive study, it nonetheless provides a starting point to motivate future research, which is more accurate forecasting of the newly formed delays and finding the optimal amount of slack that we

redistributed.

7.Future Scope:

Bad weather, air traffic delays, and mechanical issues can be difficult to predict and sometimes outside of the control of the airline. With that said, passengers are not required to be compensated by the airline if your flight is delayed or canceled for these bad weather, air traffic delays, or mechanical issues.

If you find yourself with a delayed flight due to one of these reasons, ask the original airline if it will pay for a ticket on another airline. The DOT does not require the airline to offer compensation, but it doesn’t hurt to ask.

With no federal compensation requirement

for delayed passengers, refer to the policies of the operating airline to determine what compensation the airline will offer. If a significant delay occurs, ask the airline if it will compensate you for meals during the delay.

If the airline doesn't immediately offer you compensation for your meals or expenses incurred during the delay, you may be able to get reimbursed for expenses incurred under Article 19 of the Montreal Convention by filing a claim with the airline. If that claim is denied, you could also pursue reimbursement in court.

For travelers with a canceled flight, the airline should re-book you on its first flight with available space to your destination at no additional charge. If the rebooked flight

requires a significant delay, ask the original airline if it will pay for a ticket on another airline.

8.Appendix:

```
 import pandas as pd

import numpy as np

import pickle

import matplotlib.pyplot as plt

%matplotlib inline

import seaborn as sns

import sklearn

from sklearn.tree import
DecisionTreeClassifier

from sklearn.ensemble import
GradientBoostingClassifier,
RandomForestClassifier

from sklearn.neighbors import
```

```
KNeighborsClassifier

from sklearn.model_selection import
RandomizedSearchCV

import imblearn

from sklearn.model_selection import
train_test_split

from sklearn.preprocessing import
StandardScaler

from sklearn.metrics import
accuracy_score, classification_report,
confusion_matrix, f1_score

#dataset= pd.read_csv("flightdata.csv")

#dataset.head()

#dataset.info()

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

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

```
#filter The dataset to climinate columns that aren't relevant to a predictive model_

#dataset - dataset[["FL-NLM", "MONTH", "DAY-OF-MONTH", "DAY-OF-WEEK", "DRIGIN", "DEST", "CRS-ARR-TIME", "DEP-DELIS", "ARR-SELIS"]]

#dataset.isnull().sum()

#dataset[dataset.isnull().any(axis-1)].head(10)

#dataset['DEP-DELIS'].mode()

#replace the missing values with is.

#dataset - dataset.fillna({'ARR-DELIS': 1})

#dataset - dataset.fillna({'DEP-DELIS': 0})

#dataset.iloc[177:185]

import math

#for index, row in dataset.iterrows():

#dataset.loc[index, 'CRS-ARR-TIME'] =
```

```
math.floor(row['CRS-ARR-TIME'] / 100)

#dataset.head()

from sklearn.preprocessing import
LabelEncoder

le = LabelEncoder()

#dataset['DEST'] =
le.fit_transform(dataset['DEST'])

#dataset['ORIGIN'] =
le.fit_transform(dataset['ORIGIN'])

#dataset.head(5)

#dataset['ORIGIN'].unique()

from sklearn.preprocessing import
OneHotEncoder

oh = OneHotEncoder()

#z=oh.fit_transform(x[:,4:5]).toarray()

#t=oh.fit_transform(x[:,5:6]).toarray()

#x=np.delete(x,[4,7],axis=1)
```

```
#x=np.delete(x,[4,5],axis=1)
 #print(y_pred)
#(y_pred)
#y_pred =
rfc.predict([[129,99,1,0,0,1,0,1,1,1,0,1,1,1,1,
1,]])
#print(y_pred)
#(y_pred)
#classification.saver('flight.hs')
#y_pred = classification.predict(x_test)
#y_pred
#y_pred = (y_pred > 0.5)
def predict_exit(sample_value):
  sample_value = np.array(sample_value)
  sample_value =
```

```
sample_value.reshape(1,-1)
  sample_value =
sc.transform(sample_value)
#return classifier.predict(sample_value)
#test=classification.predict([[1,1,121.00000
0,36.0,0,0,1,0,1,1,1,1,1,1,1,1]])
#if test==1:
  print('prediction: Chance ofdelay')
#else:
#print('pedicition: No chance of delay.')
            from sklearn import
model_selection
from sklearn.neural_network import
MLPClassifier
dfs = []
```

```
models = [
        #('RF', RandoForestCLassifier()),

('DecisionTree',DecisionTreeClassifier()),
        ('ANN',MLPClassifier())
       ]
result = []
names = []
scoring =['accuracy',
'precision_weighted','recall_weighted', 'f1
_weighted', 'roc_auc']
target_names = ['no delay', 'delay']
#for name, model im model:
  #kfold = model_selection.Kfold(n_splits=5,
shuffle=True, random_state=90210)
```

```
  #cv_results =
model_selection.cross_validate(model,
x_train, y_train, cv=kfold,scoring=scoring)
  #c1f = model.fit(x_train, y_train)
  #y_pred = c1f.predict(x_test)
  #print(name)
  #print(classification_report(y_test,
y_pred, target_names=target_names))
  #results.append(cv_result)
  #names.append(name)
  #this_df = pd.DataFrame(cv_results)
  #this_df['model'] = name
  #dfs.append(this_df)
  #final = pd.concat(dfs,ignore_index=True)
  #return final
```

```
  #print('Training accuracy:
',accuracy_score(y_train,y_predict_train))
  #print('Testing accuracy:
',accuracy_score(y_test,y_predict))
  #from sklearn.metrics import
confusion_matrix
  #cm = confusion_matrix(y_test, y_predict)


from sklearn.metrics import accuracy_score
#desacc =
accuracy_score(y_test,decisiontree)
from sklearn.metrics import
confusion_matrix
#cm =
confusion_matrix(y_test,decisiontree)
from sklearn.metrics import
```

```
accuracy_score,classification_report

#score = accuracy_score(y_pred,y_test)

#print('The accuracy for ANN model is: {}%'.format(score*100))

from sklearn.metrics import confusion_matrix

#cm = confusion_matrix(y_test, y_pred)


parameter = {


'estimators' :[1,20,30,55,68,74,90,120,115],

            'criterion' :['gini','entropy'],

            'max_features' :["auto", "sqrt", "log2"],

      'max_depth' : [2,5,8,10], 'verbose' :
```

```
[1,2,3,4,5,6,8,9,10]

}

#RCV =
RandomizedSearchCV(estimator=rf,param_
distributions=parameters,cv=10,n_iter=4)

#RCV.fit(x_train,y_train)

model = RandomForestClassifier(verbose=
10, n_estimators= 120,max_features=
'log2',max_depth= 10,criterion= 'entropy')

#RCV.fit(x_train,y_train)

#y_predict_rf = RCV.predict(x_test)

#RFC=accuracy_score(y_test,y_predict_rf)


import pckle

#pickle.dump(RCV,open('flight.pk1','wb'))
```

```
from flask import
Flask,request,render_template

import numpy as np

import pandas as pd

import pickle

import os

#model = pickle.loaid(open('flight.pk1','rb'))

app = Flask(__name__)

@app.route('/')

def home():

  return render_template("index.html")


  @app.route('/prediction',methods =
['POST'])
```

```
  def predict():
    name = request.form['name']
    month = request.form['month']
    dayofmonth =
request.form['dayofmonth']
    dayofweek = request.form['dayofweek']
    origin = request.form['origin']
    if(origin =="msp"):
        originl,origin2,origin3,origin4,origin5 =
0,0,0,0,1
    if(origin =="dtw"):
        origin1,origin2,origin3,origin4,origin5 =
1,0,0,0,0
    if(origin == "jfk"):
        origin1,origin2,origin3,origin4,origin5 =
```

```
0,0,1,0,0
    if(origin == "sea"):
        origin1,origin2,origin3,origin4,origin5 =
0,1,0,0,0
    if(origin == "alt"):
        origin1,origin2,origin3,origin4,origin5 =
0,0,0,1,0
#destination = request.form['destination']
#if(destination == "msp"):

destination1,destination2,destination3,dest
ination4,destination5 = 0,0,0,0,1
#if(destination == "dtw"):

destination1,destination2,destination3,dest
```

```
ination4,destination5 = 1,0,0,0,0

#if(destination == "jfk"):


destination1,destination2,destination3,disti
nation4,destination5 = 0,0,1,0,0

#if(destination == "sea"):


destination1,destination2,destination3,dest
ination4,destination5 = 0,1,0,0,0

#if(destination == "alt"):


destination1,destination2,destination3,dest
ination4,destination5 = 0,0,0,1,0

#dept = request.form['dept']

#arrtime = request.form['arrtime']
```

```
#actdept = request.form['actdept']

#dept15=int(dept)-int(actdept)

#total =
[(name,month,dayofmonth,dayofweek,origi
na1,origin2,origin3,origin4,origin5,destinati
on1,destination2,destination3,destination4,
destination5)]

#y_pred = model.predict(total)

#print(y_pred)

#if(y_pred==[0.]):

  #ans="The Flight Will be on time"

#else:

  #ans="The Flight Will be delayed"

  #return
render_template("index.html",showcase =
```

```
ans)
#if__name__=='__main__':
  #app.run(debug = True)
```