Predict visibility distance based on different climatic conditions as used by ATC

BY:-

Kushagra Sharma
Parth Gupta
Raunak Shukla
Arya BS
Kovuri Mohan Kumar

Problem Statement

In this project we have to predict visibility distance based on different climatic conditions as used by ATC for their purpose of landing or taking off the aeroplane.

In this report we predicts the visibility distance based on the different indicators as VISIBILITY, DRYBULBTEMPF, WETBULBTEMPF, DewPointTempF, RelativeHumidity, WindSpeed, WindDirection, StationPressure, SeaLevelPressure, Precip with this we providing a new visibility return forecast for weather forecast personnel so as to improve the visibility of the level of visibility to ensure the safe and stable operation of the airport.

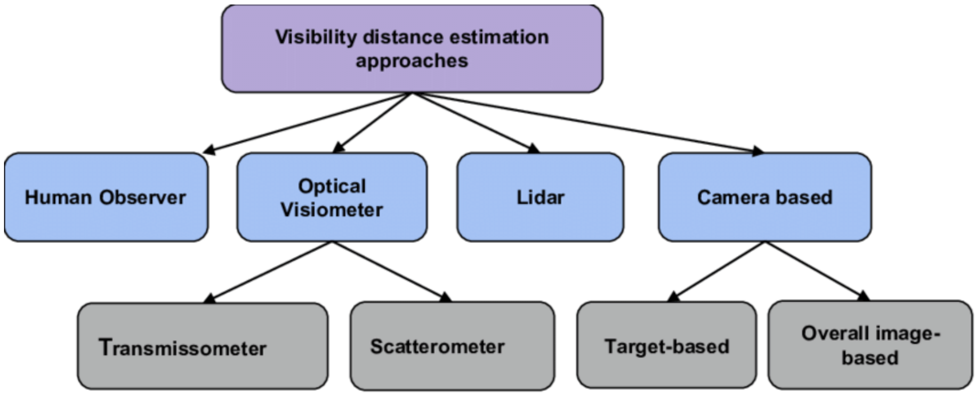
BUSINESS NEED ASSESSMENT

- With the rapid development of the national economy and the increasing popularity of civil aviation transport, airport operation on the visibility is becoming increasingly prominent. A long, low-visibility weather caused by fog, haze and other weather can cause a wide range of airport delays and cancellations.
- This not only has brought huge losses for the airlines and the airport, but also affects the public travel. At the same time visibility and flight safety are closely related. Low visibility is also one of the most common causes of flight accidents.
- Most of the low visibility days occurred in the winter half (November to March), up to 57 days. The weather phenomenon that causes low visibility is mainly fog and smoke.

TARGET SPECIFICATIONS AND CHARACTERIZATIONS:

- Improving the level of visibility is an important measure to ensure the safe and stable operation of the airport. At present, the low visibility forecast for the smoke, fog and other weather.
- The expected departure time (EDT) was a prerequisite for determining the departure of the flight strictly, and it could be regarded as the initial departure qualification in our model.
- The actual departure times of flights might have deviated from the scheduled times due to the implementation of GDPs or the reasons of the airlines. The inaccessibility and uncertainty of these factors made it difficult to predict the expected departure time of the flights accurately.
- Future air traffic management concepts envisage shared decision-making responsibilities between controllers and pilots, necessitating that controllers be supported by automated decision aids.
- Even as automation tools are being introduced, however, their impact on the air traffic controller is not well understood. The present experiments examined the effects of an aircraft-to-aircraft conflict decision aid on performance and mental

workload of experienced, full-performance level controllers in a simulated Free Flight environment.



EXTERNAL SEARCH (INFORMATION SOURCES/REFERENCES):

REFERENCES:

1. Weather Visibility Prediction Based on Multimodal Fusion
   https://www.researchgate.net/publication/333643589_Weather_Visibility_Prediction_Based_on_Multimodal_Fusion

2. Air traffic control
   https://www.britannica.com/technology/traffic-control/Air-traffic-control

DATASET:

1. Referenced Small dataset from Kaggle.
   https://www.kaggle.com/datasets/zhaodianwen/noaaweatherdatajfkairport

2. Used dataset in the project
   https://docs.google.com/spreadsheets/d/1NAr3H9av_IiBFYATt9ghPRhKyJw-509C/edit?usp=sharing&ouid=100120693750941465424&rtpof=true&sd=true

```python
import pandas as pd
import numpy as np
import matplotlib.pyplot as plt
import seaborn as sns
```

Getting Data

```python
df = pd.read_csv('/content/drive/MyDrive/data/ATC_Plane/jfk_weather_cleaned.csv')
df.head()
```

	DATE	VISIBILITY	DRYBULBTEMPF	WETBULBTEMPF	DewPointTempF	RelativeHumidity	WindSpeed	WindDirection	StationPressure	SeaLevelPressure	Precip
0	01-01-2010 00:51	6.0	33	32	31	92	0	0	29.97	29.99	0.01
1	01-01-2010 01:51	6.0	33	33	32	96	0	0	29.97	29.99	0.02
2	01-01-2010 02:51	5.0	33	33	32	96	0	0	29.97	29.99	0.02
3	01-01-2010 03:51	5.0	33	33	32	96	0	0	29.95	29.97	0.02
4	01-01-2010	5.0	33	32	31	92	0	0	29.93	29.96	0.02

The above dataset will enable us to show that even with little data we will be able to bring them profits.

```python
df.describe().T
```

	count	mean	std	min	25%	50%	75%	max
VISIBILITY	75083.0	9.211896	2.202311	0.00	10.00	10.00	10.00	14.00
DRYBULBTEMPF	75083.0	55.355527	17.394334	1.00	42.00	56.00	70.00	102.00
WETBULBTEMPF	75083.0	49.327544	16.182867	-1.00	36.00	50.00	64.00	85.00
DewPointTempF	75083.0	42.424024	19.577957	-19.00	27.00	44.00	59.00	84.00
RelativeHumidity	75083.0	64.812075	19.898962	8.00	49.00	66.00	82.00	100.00
WindSpeed	75083.0	11.253240	6.101048	0.00	7.00	10.00	15.00	53.00
WindDirection	75083.0	196.550751	107.692804	0.00	110.00	200.00	290.00	360.00
StationPressure	75083.0	30.005579	0.235172	28.52	29.86	30.00	30.15	30.83
SeaLevelPressure	75083.0	30.026049	0.234069	28.54	29.88	30.02	30.17	30.85
Precip	75083.0	0.005478	0.036161	0.00	0.00	0.00	0.00	2.41

BENCHMARKING ALTERNATE PRODUCTS:

In present, benchmarking has become an increasingly common tool used in wide range industries and helpful in identifying problems and enhances performance. This paper will discuss the application of benchmarking to predict visibility distance based on different climatic conditions as used by ATC, they are airport services charges and airport productivity. Airport as the major firm in entire aviation industry, which performs as a supply chain and provides infrastructure services to airlines and passengers, therefore, airport productivity is consisting as an important essential feature for airline, passengers and itself. But according to International Air Transport Association (IATA), it stated challenges associated with benchmarking actually exist and willing to directly impact benchmark sound effects, they are availability and quality of data, and different factors effecting differences between two organizations (Tretheway and Kincaid, 2010). In this will discuss the different users and their purpose of using benchmarking, and introduce Southwest Airline as example for represent airport user, to illustrate how airport's productivity will impact their users and how they achieved their goals through benchmark other practices in the industry.

APPLICABLE PATENTS:

Weather data selection relative to an aircraft trajectory by Frank Saggio, IIIAna Isabel Del Amo Blanco (https://patents.google.com/patent/US8433506).

APPLICABLE REGULATIONS:

- Data protection and privacy rules.
- License for the open-source codes that might be used in the model implementation.
- Laws related to AI.

APPLICABLE CONSTRAINTS:

- Data collection from the customer.
- The customer should know about the time, money and scope of the project before it starts.
- Transparent use of the data obtained from the customer.

BUSINESS OPPORTUNITIES:

- The altitude or flight level instructions in an ATC clearance normally require that a pilot "MAINTAIN" the altitude or flight level at which the flight will operate when in controlled airspace. Altitude or flight level changes while en route should be requested prior to the time the change is desired.
- When possible, if the altitude assigned is different from the altitude requested by the pilot, ATC will inform the pilot when to expect climb or descent clearance or to request altitude change from another facility. If this has not been received prior to

crossing the boundary of the ATC facility's area and assignment at a different altitude is still desired, the pilot should reinitiate the request with the next facility.

- The term “cruise” may be used instead of “MAINTAIN” to assign a block of airspace to a pilot from the minimum IFR altitude up to and including the altitude specified in the cruise clearance. The pilot may level off at any intermediate altitude within this block of airspace. Climb/descent within the block is to be made at the discretion of the pilot. However, once the pilot starts descent and verbally reports leaving an altitude in the block.
- TIS provides proximity warning only, to assist the pilot in the visual acquisition of intruder aircraft. No recommended avoidance maneuvers are provided nor authorized as a direct result of a TIS intruder display or TIS alert. It is intended for use by aircraft in which TCAS is not required.
- TIS does not alter or diminish the pilot's basic authority and responsibility to ensure safe flight. Since TIS does not respond to aircraft which are not transponder equipped, aircraft with a transponder failure, or aircraft out of radar coverage, TIS alone does not ensure safe separation in every case.
- At this time, no air traffic service nor handling is predicated on the availability of TIS equipment in the aircraft.
- Presently, no air traffic services or handling is predicated on the availability of an ADS-B cockpit display. A “traffic-in-sight” reply to ATC must be based on seeing an aircraft out-the-window, NOT on the cockpit display.

CONCEPT GENERATION:

Visibility is an indicator of atmospheric transparency. Visibility was first defined for meteorological purposes as a quantity to be estimated by a human observer, and observations made in that way are widely used. The presence of atmospheric particles always causes a reduction of visibility. As air pollution has become increasingly serious in recent years, visibility forecasts have directly affected work and life. On the morning of 27 December 2019, a plane of Kazakhstan Bek Airlines crashed after take-off at Almaty Airport because of the heavy fog. Therefore, the accuracy of low-visibility forecasts is directly related to aviation safety. The factors that affect visibility are complex, related to the degree of atmospheric aerosol pollution and atmospheric circulation, near-surface water vapor, temperature inversion, and human activities.

In the early days, visibility forecasts mainly used extrapolation methods to study the occurrence and development of weather systems. However, weather systems are nonlinear, and many important parameters are ignored after linearization. With the development of numerical weather prediction technology, researchers have gradually used numerical weather prediction technology to forecast visibility, but visibility changes have a strong locality. Numerical weather forecasting cannot meet airports’ need for high accuracy and high time resolution of visibility. With the development of AI/ML, researchers have increasingly used it to predict visibility.

CONCEPT DEVELOPMENT:

We must first understand the environment before we start working on a model and the to predict visibility distance based on different climatic conditions as used by ATC for their purpose of landing or taking off the aeroplane. After gaining sufficient knowledge about the environment we have to start collecting data. After collecting the data, we have to perform EDA which is used to identify patterns in the dataset and it will help us zone in on the areas that are leaking money. Visualization will help a lot here. Once we have found the trend and outliers, the next step is to use the basic regression models and time-series models , in which we will fit our training dataset and see what sort of results we will be getting. After analysing various parameters like squared-error, etc we will know what type of model to use and what type of model should our model be based around. The models will be regression models and time-series models.

CODE IMPLEMENTATION (SMALL SCALE):

GitHub links:

1.https://github.com/reesejohn900/PROJECTS/blob/main/JFK_Weather_EDA.ipynb

2.https://github.com/Raunak017/Feynn-Projects/blob/main/ATC/ATC_Regression.ipynb

3. https://github.com/Kushagra-Sharma26/Projects/blob/main/visibilityPrediction.ipynb

Data Preprocessing

```python
for i in df.columns:
  if df[i].dtype !='O':
    print('Distribution of '+i)
    print(df[i].value_counts())
    print('-'*50)
```

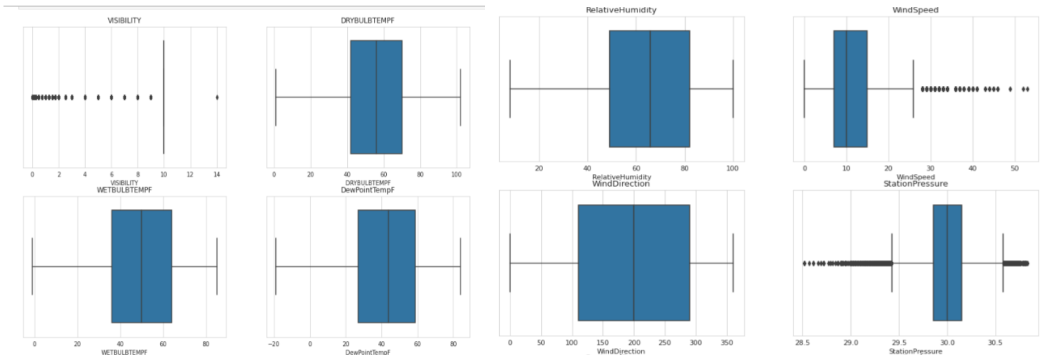
```
Distribution of VISIBILITY
10.00    63371
9.00      1748
8.00      1582
7.00      1375
6.00      1210
5.00       934
3.00       694
2.00       595
4.00       536
0.25       521
0.50       512
1.00       435
2.50       381
1.50       340
0.12       336
0.00       245
0.75       202
1.75        34
1.25        31
14.00        1
Name: VISIBILITY, dtype: int64
--------------------------------------------------
Distribution of DRYBULBTEMPF
73    1690
```

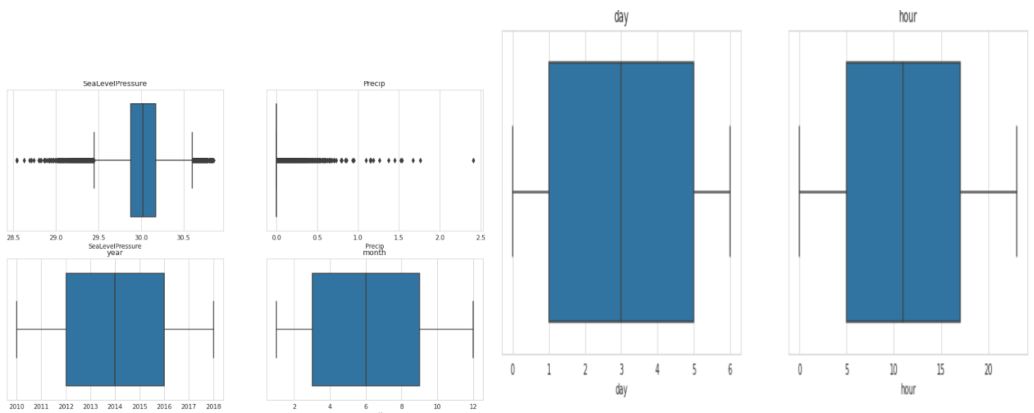
```python
# extraction of date
df['DATE'] = pd.to_datetime(df['DATE'])

df['year'] = df['DATE'].dt.year
df['month'] = df['DATE'].dt.month
df['day'] = df['DATE'].dt.dayofweek
df['hour'] = df['DATE'].dt.hour
```

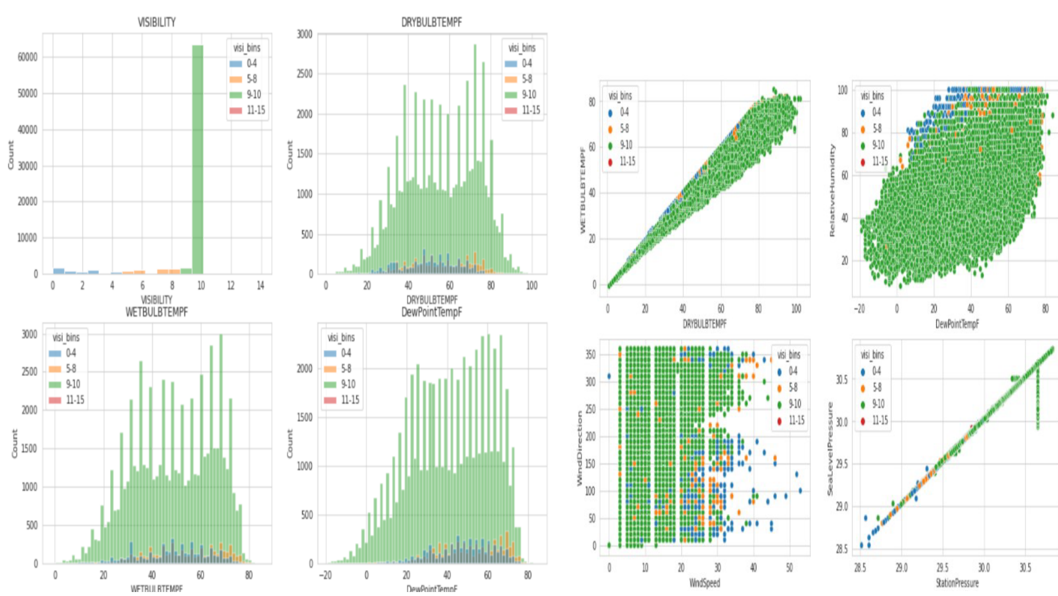
```python
l = ['year','month','day','hour']

for i in l:
  print('Distribution of '+i)
  print(df[i].value_counts())
  print('-'*50)
```

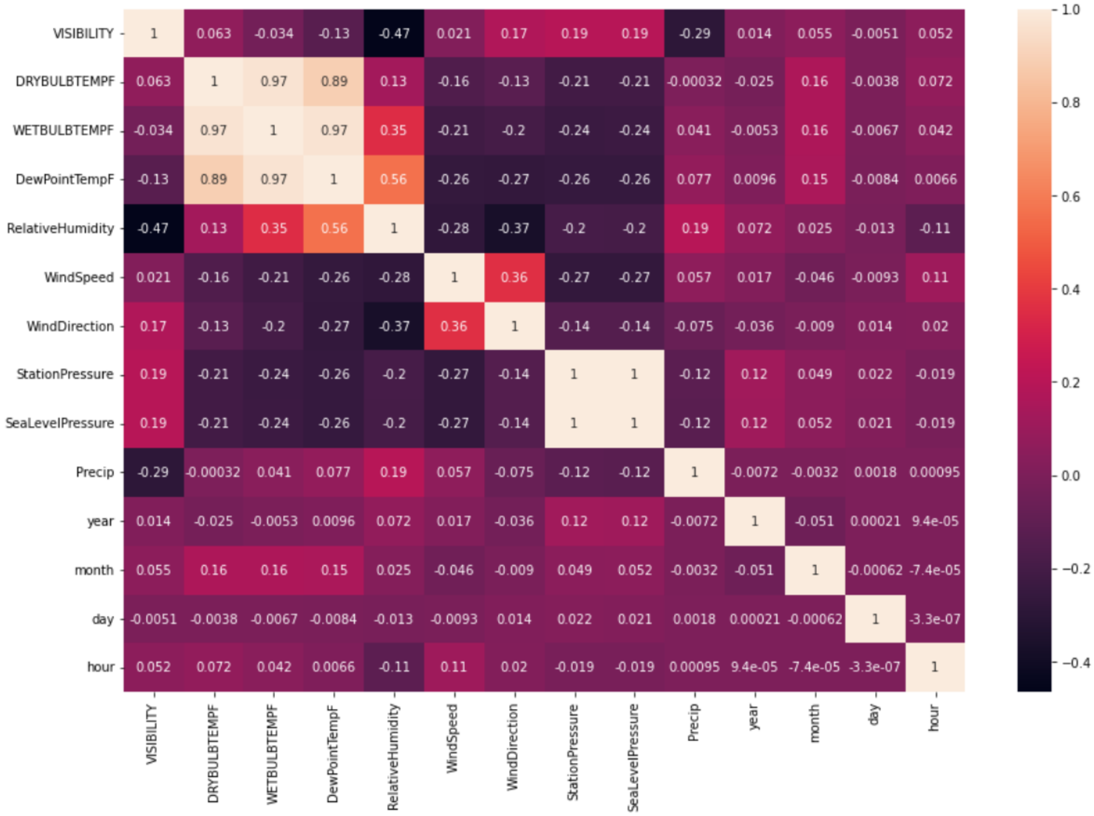
```
Distribution of year
2012    8784
2016    8780
2011    8762
2010    8756
2014    8755
2015    8754
2017    8754
2013    8753
2018    4985
Name: year, dtype: int64
--------------------------------------------------
Distribution of month
1     6574
5     6573
3     6565
7     6475
6     6358
```

Data Visualization:

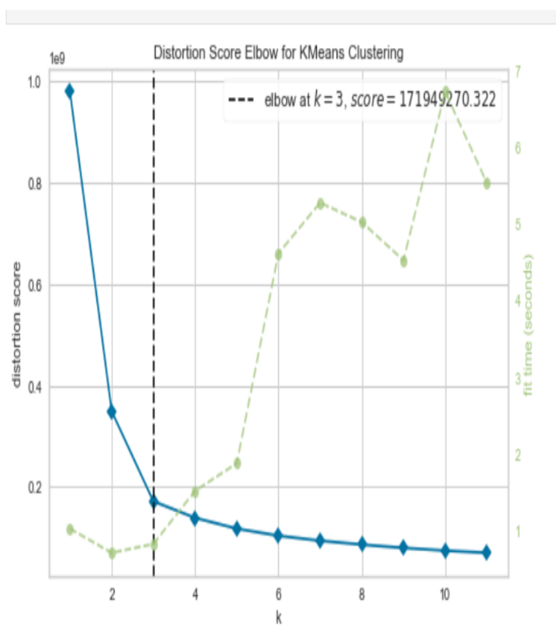


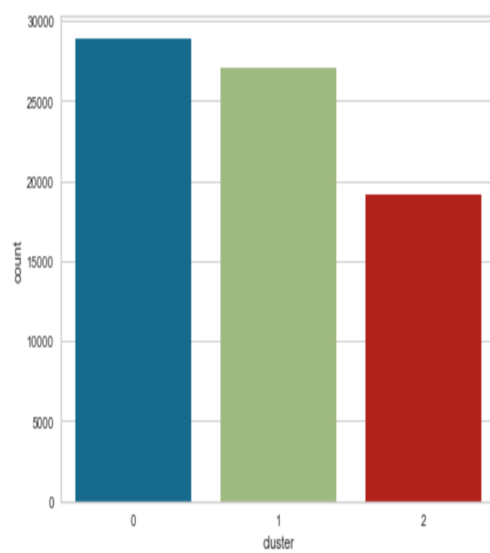


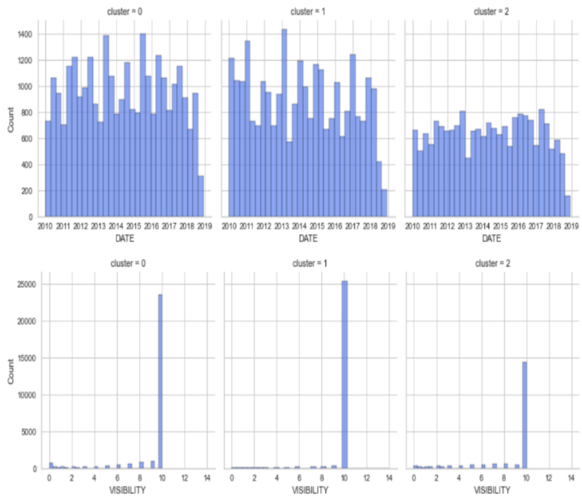


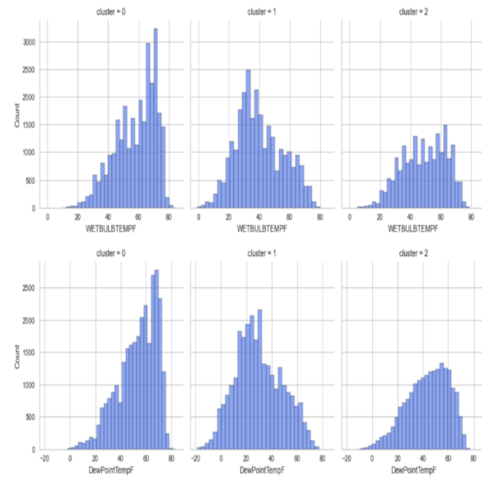


KMeans clustering









Train and Test

```
       DRYBULBTEMPF  RelativeHumidity  WindSpeed  WindDirection  \
61271            33                58          0              0
51590            45                46         11             30
68967            43                76          9             20
12362            77                58          9             30
9986             31                85          8             70
...             ...               ...        ...            ...
37374            48                77         18             90
52530            46                79          6             60
39860            74                64         14            100
43599            47                90         15             60
11857            56                47         15             10

       SeaLevelPressure  Precip  year  month  day  hour  cluster
61271             30.16    0.00  2016     12    2    23        2
51590             30.29    0.00  2015     11    5     7        2
68967             30.32    0.00  2017     11    1    18        2
12362             30.18    0.00  2011      5    1     6        2
9986              29.81    0.01  2011      2    0     6        2
...                 ...     ...   ...    ...  ...   ...      ...
37374             29.96    0.03  2014      7    4    16        2
52530             30.26    0.00  2015     12    2    14        2
39860             30.16    0.00  2014      7    6    10        2
43599             30.12    0.00  2014     12    1     5        2
11857             30.00    0.00  2011     10    2     5        2

[13385 rows x 11 columns]
61271    10.0
51590    10.0
68967    10.0
12362    10.0
9986      2.0
         ...
37374     6.0
52530    10.0
39860    10.0
43599    10.0
11857    10.0
Name: VISIBILITY, Length: 13385, dtype: float64
       DRYBULBTEMPF  RelativeHumidity  WindSpeed  WindDirection  \
74583            64                58         14             10
74181            70                55          9             40
52526            43                86         10             70
18823            40                86         10             70
47532            57                72          7             80
...             ...               ...        ...            ...
66266            64                87          9             40
32024            86                49         16             10
22947            76                72          0              0
40839            65                76          0              0
73946            70                41          9            110

       SeaLevelPressure  Precip  year  month  day  hour  cluster
74583             30.25    0.00  2018      7    5     2        2
74181             29.91    0.00  2018      6    2     6        2
52526             30.32    0.00  2015     12    2    10        2
18823             29.67    0.01  2012      2    4     9        2
47532             30.14    0.00  2015      5    2     5        2
...                 ...     ...   ...    ...  ...   ...      ...
66266             30.02    0.00  2017      7    1     4        2
32024             29.85    0.00  2013      8    1    14        2
22947             29.97    0.00  2012      8    1     6        2
40839             30.28    0.00  2014      8    5     5        2
73946             30.09    0.00  2018     10    5    11        2

[5737 rows x 11 columns]
74583    10.0
74181    10.0
52526     8.0
18823     7.0
47532    10.0
         ...
66266    10.0
32024    10.0
```

```python
from sklearn.preprocessing import StandardScaler
sc = StandardScaler()
X_train = sc.fit_transform(X_train)
X_test = sc.transform(X_test)

print(X_train)

print(X_test)
```

```
[[-1.37420403 -0.75223945 -1.3541928  ... -0.48499618  1.72537278
   0.        ]
 [-0.58453388 -1.4026984   0.44735568 ...  1.00266874 -0.49336178
   0.        ]
 [-0.71614557  0.22344897  0.11980141 ... -0.98088448  1.03201823
   0.        ]
 ...
 [ 1.32383564 -0.42700998  0.93868709 ...  1.49855705 -0.07734905
   0.        ]
 [-0.45292219  0.98231774  1.10246422 ... -0.98088448 -0.7707036
   0.        ]
 [ 0.13933042 -1.34849349  1.10246422 ... -0.48499618 -0.7707036
   0.        ]]
[[ 0.66577718 -0.75223945  0.93868709 ...  1.00266874 -1.18671633
   0.        ]
 [ 1.06061226 -0.91485419  0.11980141 ... -0.48499618 -0.63203269
   0.        ]
 [-0.71614557  0.76549809  0.28357855 ... -0.48499618 -0.07734905
   0.        ]
 ...
 [ 1.45544733  0.00662932 -1.3541928  ... -0.98088448 -0.63203269
   0.        ]
 [ 0.73158303  0.22344897 -1.3541928  ...  1.00266874 -0.7707036
   0.        ]
 [ 1.06061226 -1.67372296  0.11980141 ...  1.00266874  0.06132186
   0.        ]]
```

The results after running various base Decision Tree Regressor, XG Boosting, Stacking are available in the above picture. Here we can see that even for a dataset with a lot features the basic regression models perform well without hyperparameter tuning. All the models perform well but for a real world dataset there will be a lot of additional features, correlations and patterns to consider. There is a lot of EDA done on the notebook present in the GitHub link that has not been given here.

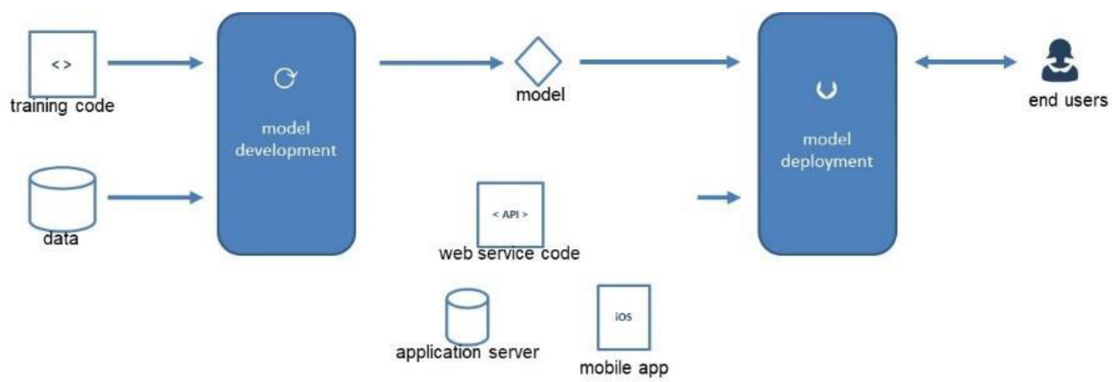
```
In [57]:
print("Evaluation Metrics")
MAE_list = {
    'Decision Tree':regt_mae,
    'XGB':xgb_mae,
    'Stack':sr_mae
}

MSE_list = {
    'Decision Tree':regt_mse,
    'XGB':xgb_mse,
    'Stack':sr_mse
}

RMSE_list={
    'Decision Tree':regt_rmse,
    'XGB':xgb_rmse,
    'Stack':sr_rmse

}

R2_list={
    'Decision Tree':regt_r2,
    'XGB':xgb_r2,
    'Stack':sr_r2

}
MAE_df_test = pd.DataFrame.from_dict(MAE_list,orient="index",columns=['MAE'])
MSE_df_test = pd.DataFrame.from_dict(MSE_list,orient="index",columns=['MSE'])
RMSE_df_test = pd.DataFrame.from_dict(RMSE_list,orient="index",columns=['RMSE'])
R2_df_test = pd.DataFrame.from_dict(R2_list,orient="index",columns=['R2'])
df_test = pd.concat([MAE_df_test,MSE_df_test,RMSE_df_test,R2_df_test],axis = 1)
df_test
```

```
Evaluation Metrics
Out[57]:
```

	MAE	MSE	RMSE	R2
Decision Tree	1.175886	4.317574	2.077877	0.424308
XGB	0.712531	1.810401	2.077877	0.758607
Stack	0.697952	1.789628	2.077877	0.761377

FINAL PRODUCT PROTOTYPE:

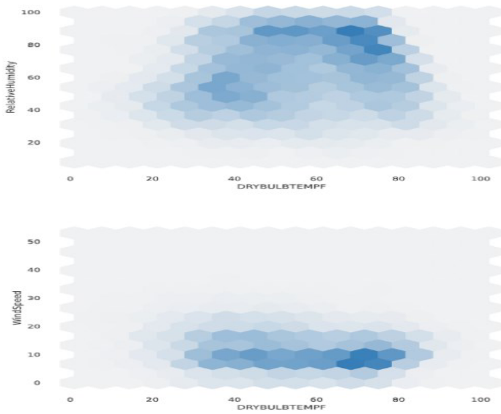
The final product should be preferably a web app specially customized for the customer so that he/she could work with the model easily.
The product follows the process given in the following picture.



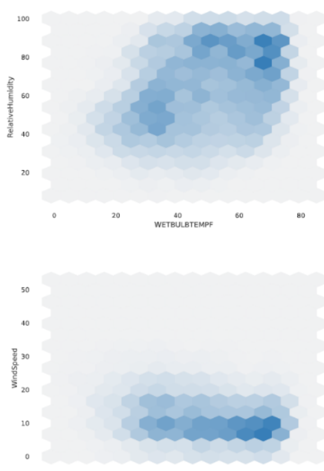
Here I have given the photos of implementation.

VISIBILITY

DRYBULBTEMPF



WETBULBTEMPF



There is a lot of Images related (DewPointTempF, RelativeHumidity, WindSpeed,WindDirection, StationPressure, SeaLevelPressure,Precip) done on the notebook present in the GitHub link that has not been given here.

The final model should be fitted regularly on the incoming data and we should always be on the look out for new features that might be affecting the sales.

Back-end:

- This involves data collection, pre-processing and integrating the model with the web app.
- The data entered by the customer should also be collected with the customer's permission.

Front-end:

- Front-end plays a crucial role as it is the interface with which the customer will be working.
- It should be highly user-friendly, otherwise they might enter some wrong data and the predicted sales will be totally wrong.

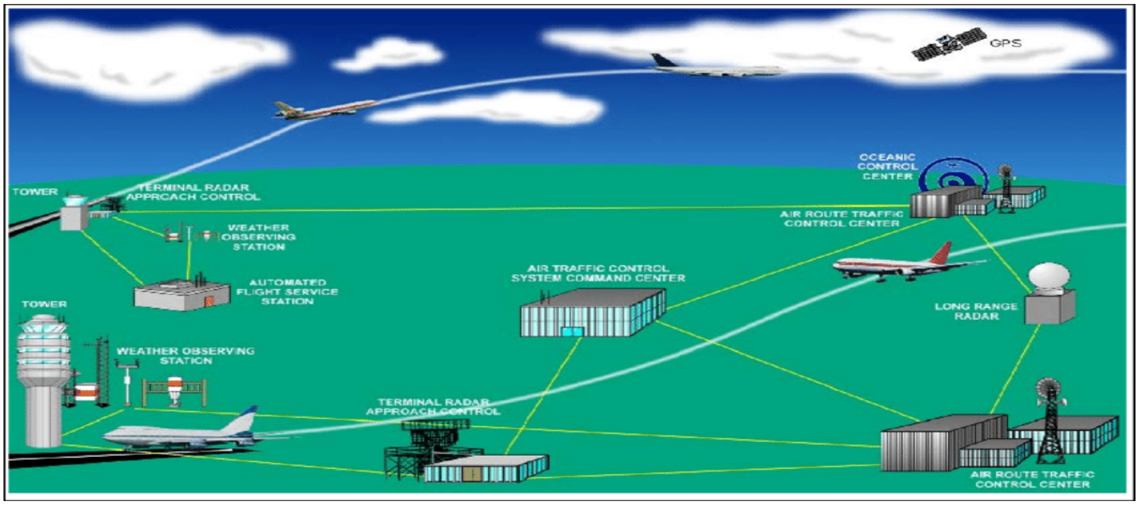
PRODUCT DETAILS:

- The model will be a web app in which the customer will be entering his/her predicted input features for the future (like 5-10 days) and the model will be forecasting the sales.
- The model requires data from the customer (as much as they can give) and we should first weed out the unnecessary features.
- EDA has to be done in a thorough manner and it should be done by a person with some knowledge about the industry.
- Next step is to use ML models. It can either be developed by us from the scratch or we could use the models available in the open-source libraries with a proper license.
- Pandas, seaborn, scikit-learn, python and jupyter notebook are the tools needed for performing EDA and model development.

CONCLUSION:

From the points that I have made above and the analysis that I had done with a somewhat real world dataset shows that there is indeed a lot of potential for AI/ML predict visibility distance based on different climatic conditions as used by ATC. The uses of AI in this industry are limitless and it is only a matter of time for AI to fully come into this industry. Big companies have already started integrating AI with their business. The models implemented by the big companies has been a game-changer for them and AI has helped them, stay ahead of the curve. Implementing this sort of technology will not only profit them but it will also open a lot of areas where AI can be used. I can totally see the using of AI in a few years as AI has already penetrated into a lot of other fields.

BUSINESS MODEL

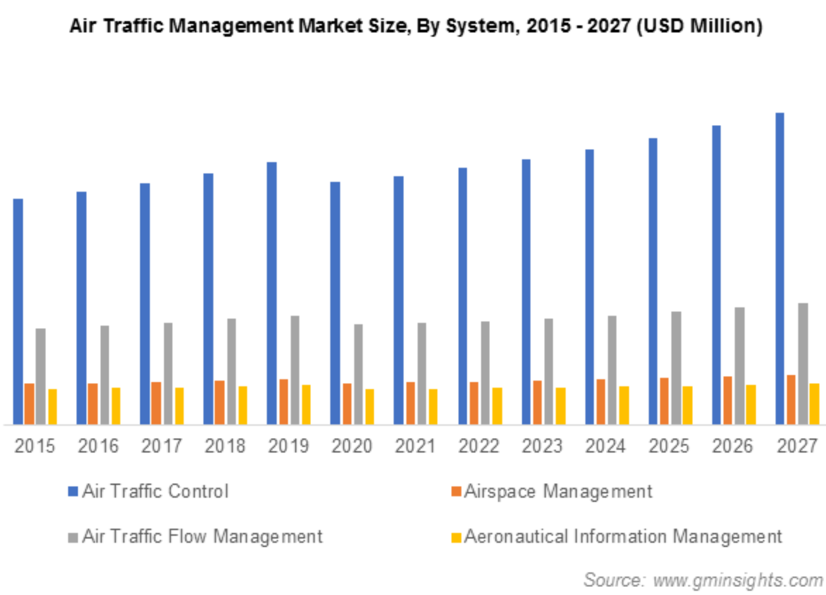


In this part of the report, we will look at the business model suggested for the idea presented earlier. There are many business models available but we have chosen the 'Consulting Business Model' which is the one suited for our idea.

BUSINESS MODEL:

"Consulting or agency business model provides a specific set of services for a fee. While this business model can be applied to almost any industry, the main drawback of consulting business models is that you should have expertise and authority around your brand."
We have to follow the Consulting Business Model because for each client of ATC the features will be different, and we have to consult them based on those features. But before following this model the ML engineers of the company have to get acclimatized with the ATC industry because this model requires an in-depth knowledge of the industry.

FINANCIAL EQUATION:



Air Traffic Management Market size valued at USD 8 billion in 2020 and is expected to grow at over 3% CAGR between 2021 and 2027. The growing trends of strategic partnerships & collaborations between industry players is enhancing the industry growth of air traffic management systems.

Increasing technological innovations and advancements will augment the software segment growth

The software segment is projected to witness high growth of around 3.5% CAGR through 2026 owing to rising technological innovation in air traffic management solutions. Technology providers are focusing on offering innovative solutions to enhance airport efficiency and allow smooth operations of the airlines, contributing to the growing market penetration.

Y = X * (1 + r)t
Y = (X)* (3.5)t
Y = Profit over time, X = Price of our Product, r = growth rate, t = time interval

1+r = 1 + 3.5% = 1.035