

Saudi Arabia Weather Forecasting

Black Belts



Abdulaziz Otaif

Reem Alsabti

Mohammed Alhaddad

Esraa AlZahrani

Elham AlGhamdi

Norah Jadkarim

Table of Contents

1
Introduction

2
**Dataset
overview**

3
Preprocessing

4
EDA

5
Dashboards

6
Regression

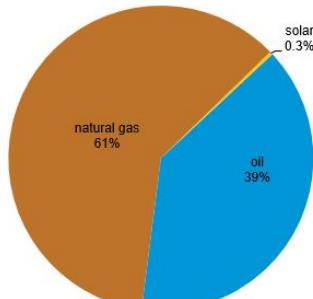
7
Classification

8
Future work

VISION 2030

- Saudi Vision 2030 aims to achieve a renewable and sustainable energy supply of 9.5 GW by 2030.
- The contribution of renewable energy to the overall energy mix will reach up to 50% of all energy supply in Saudi Arabia.

Figure 8. Saudi Arabia's electric power generation by fuel, 2020



Source: Graph by the U.S. Energy Information Administration, based on data from
BP Statistical Review of World Energy 2021

VISION 2030



Renewable energy sources are natural forces that are strongly dependent on weather conditions.

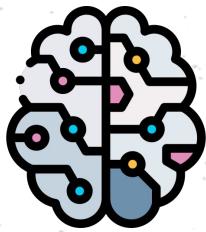


Bad weather conditions such as heavy clouds, rain, and sandstorms will reduce solar panels' energy supply significantly.



If we can predict these bad weather events beforehand, we can prepare hours, days, or even weeks in advance.

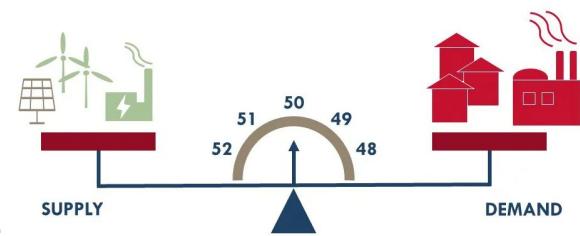
Goal



Leverage machine learning and data to predict weather conditions.



Weather forecasts can be used to increase the efficiency and reliability of renewable energy sources significantly.



Estimate how much energy is likely to be produced from renewable energy technologies and how much is likely to be needed.

Importance of Weather Forecasting



Renewable Energy



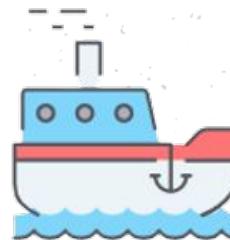
Aviation



Agriculture

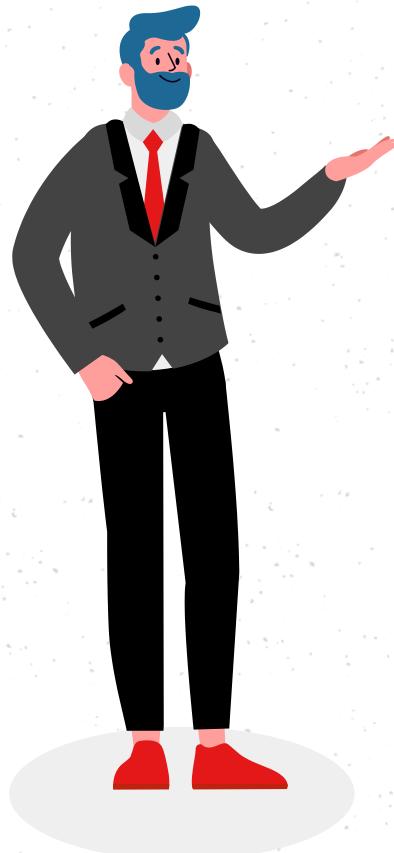


Logistics & Transportation



Marine

Dataset Overview



Overview

	city	date	time	year	month	day	hour	minute	weather	temp	wind	humidity	barometer	visibility	
0	Qassim	1 January 2017	00:00	2017		1	1	24	0	Clear	17	11	64.0	1018.0	16
1	Qassim	1 January 2017	01:00	2017		1	1	1	0	Clear	17	6	64.0	1018.0	16
2	Qassim	1 January 2017	03:00	2017		1	1	3	0	Clear	15	11	72.0	1019.0	16
3	Qassim	1 January 2017	04:00	2017		1	1	4	0	Clear	15	11	72.0	1019.0	16
4	Qassim	1 January 2017	05:00	2017		1	1	5	0	Clear	15	9	72.0	1019.0	16

Description

	city	date	time	year	month	day	hour	minute	weather	temp	wind	humidity	barometer	visibility
0	Qassim	1 January 2017	00:00	2017	1	1	24	0	Clear	17	11	64.0	1018.0	16

- The attributes can be classified into three subcategories:

Place of collection

City

Time and Date

time, year, month,
day, hour, minute

Condition & Metrics

Weather, temp, wind,
humidity, barometer,
visibility

Description

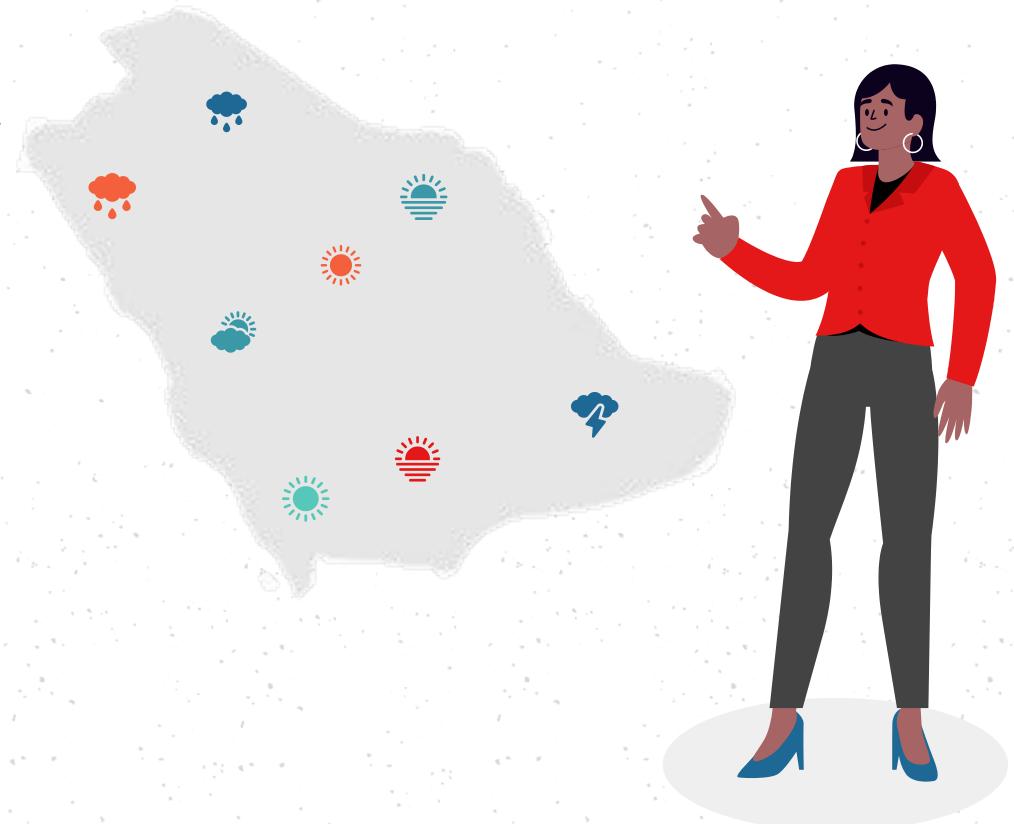
```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 249023 entries, 0 to 249022
Data columns (total 14 columns):
 #   Column      Non-Null Count  Dtype  
--- 
 0   city        249023 non-null   object 
 1   date        249023 non-null   object 
 2   time        249023 non-null   object 
 3   year        249023 non-null   int64  
 4   month       249023 non-null   int64  
 5   day         249023 non-null   int64  
 6   hour        249023 non-null   int64  
 7   minute      249023 non-null   int64  
 8   weather     249023 non-null   object 
 9   temp         249023 non-null   int64  
 10  wind         249023 non-null   int64  
 11  humidity    249006 non-null   object 
 12  barometer   248951 non-null   float64
 13  visibility  249023 non-null   int64  
dtypes: float64(1), int64(8), object(5)
memory usage: 26.6+ MB
```

	city	0
date	0	0
time	0	0
year	0	0
month	0	0
day	0	0
hour	0	0
minute	0	0
weather	0	0
temp	0	0
wind	0	0
humidity	17	0
barometer	72	0
visibility	0	0
dtype: int64		

Columns

Missing

Preprocessing



Preprocessing

Drop unnecessary columns such as time, minute, ID and NA values

1 Dropping

convert humidity into floats after removing the percent sign

2 Converting

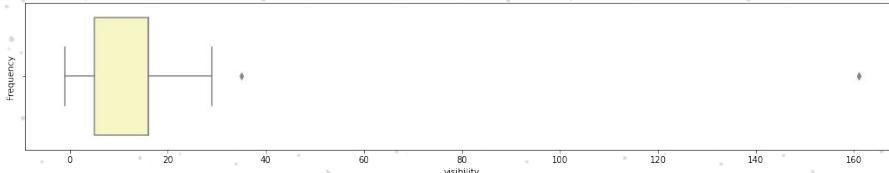
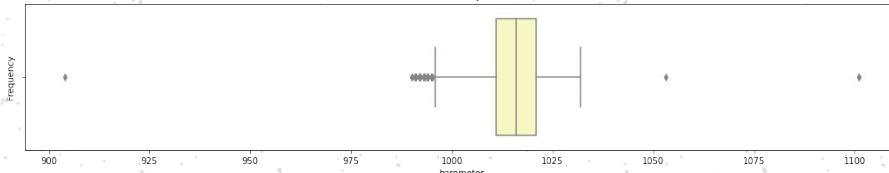
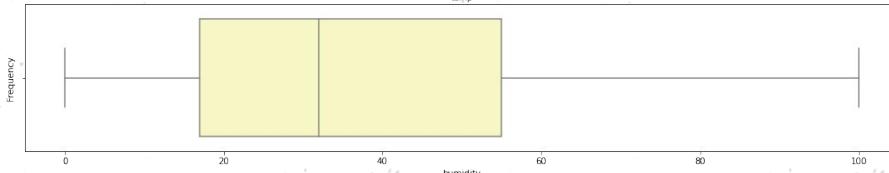
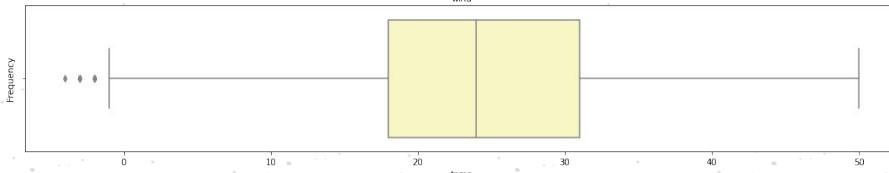
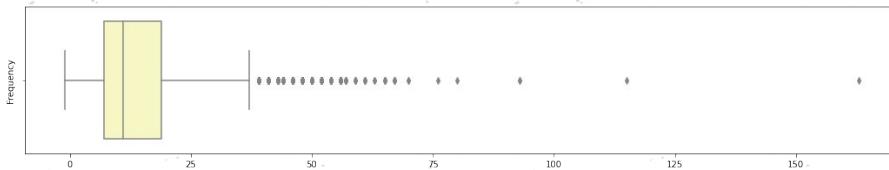
mapped the 81 weather conditions into only 9 conditions.

3 Mapping

applied the IQR method to Outliers on 5 different features

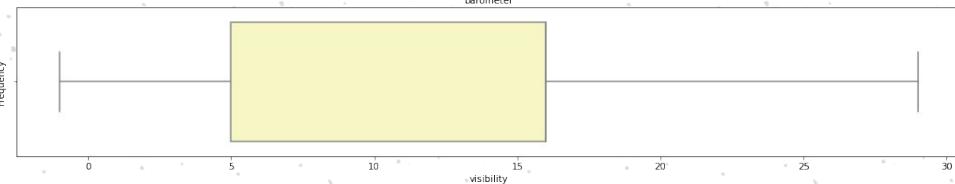
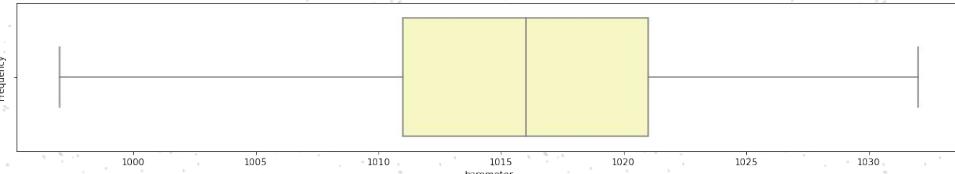
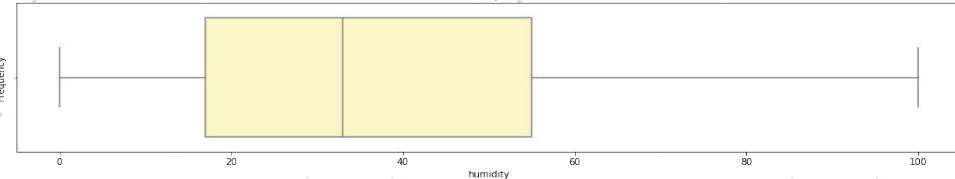
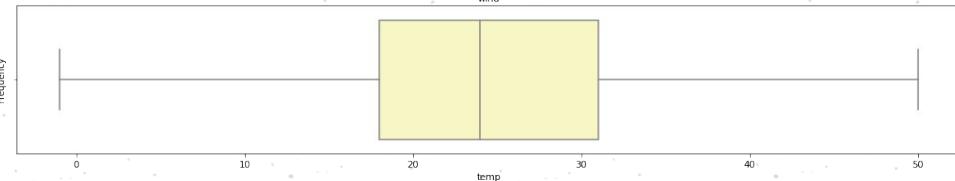
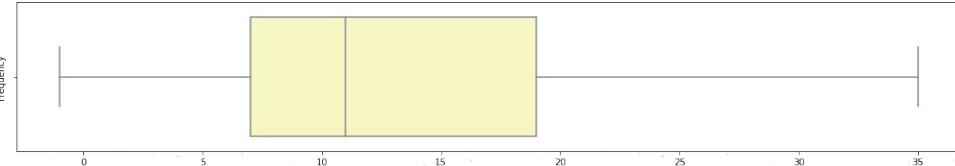
4 Outliers

Detect the outliers of different weather metrics and attributes (Before removing)



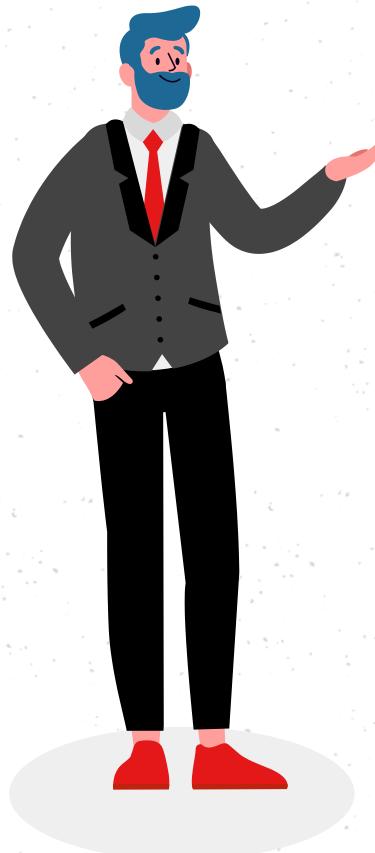
Before Removing

Detect the outliers of different weather metrics and attributes (After removing)



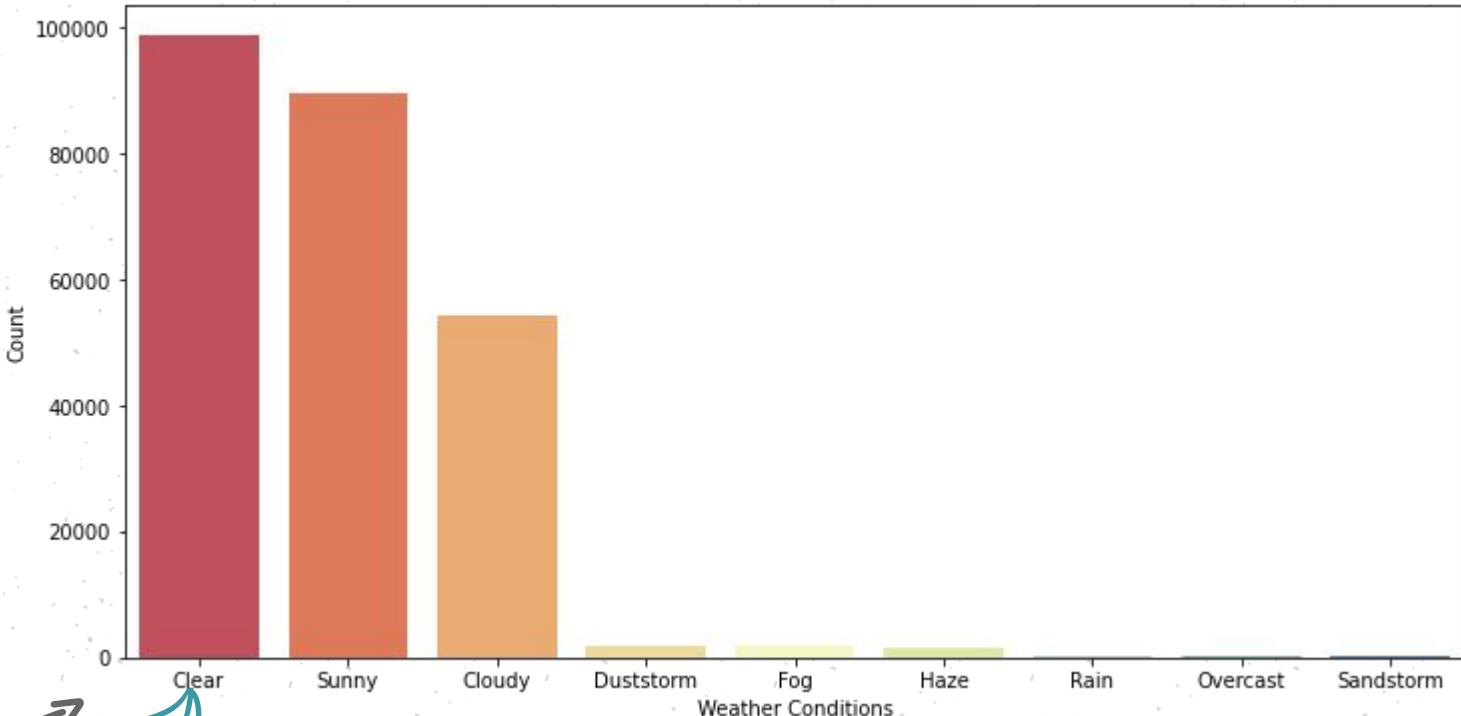
After Removing

Exploratory Data Analysis



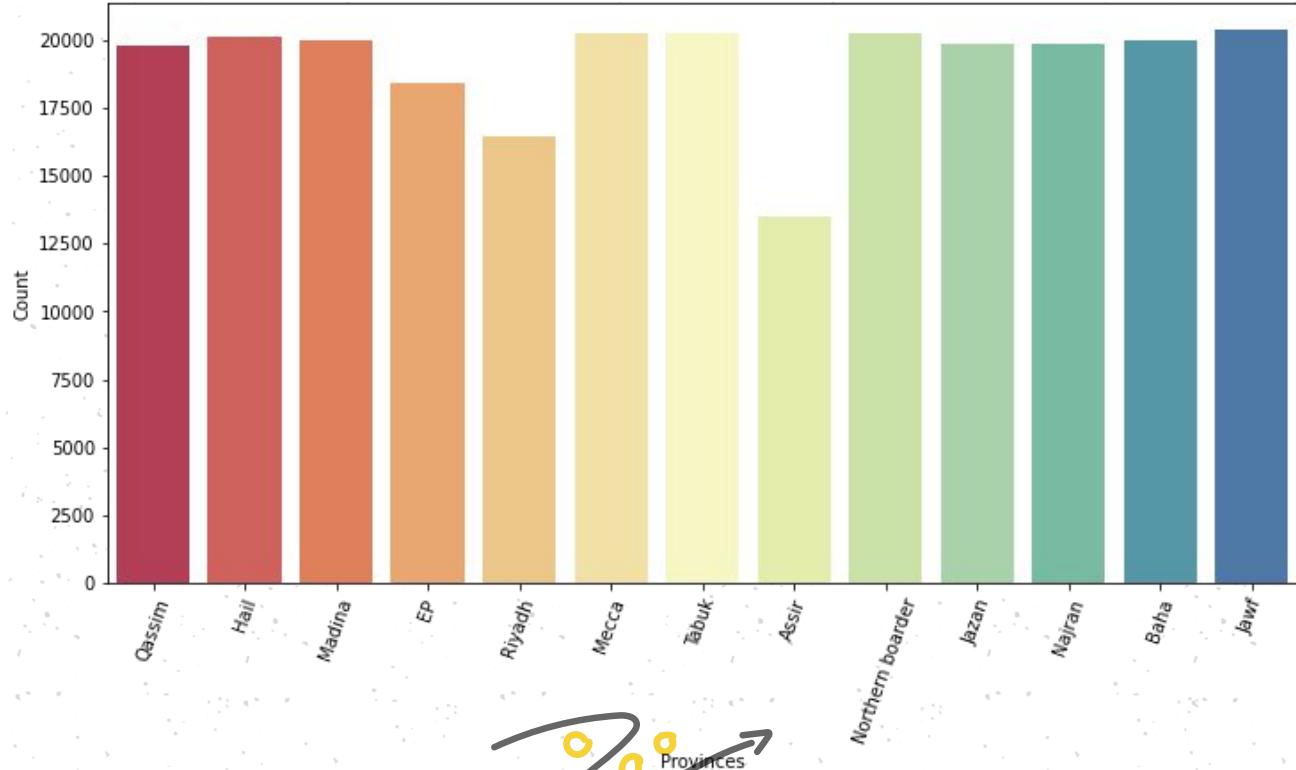
Weather Conditions

Most Frequent Weather Conditions

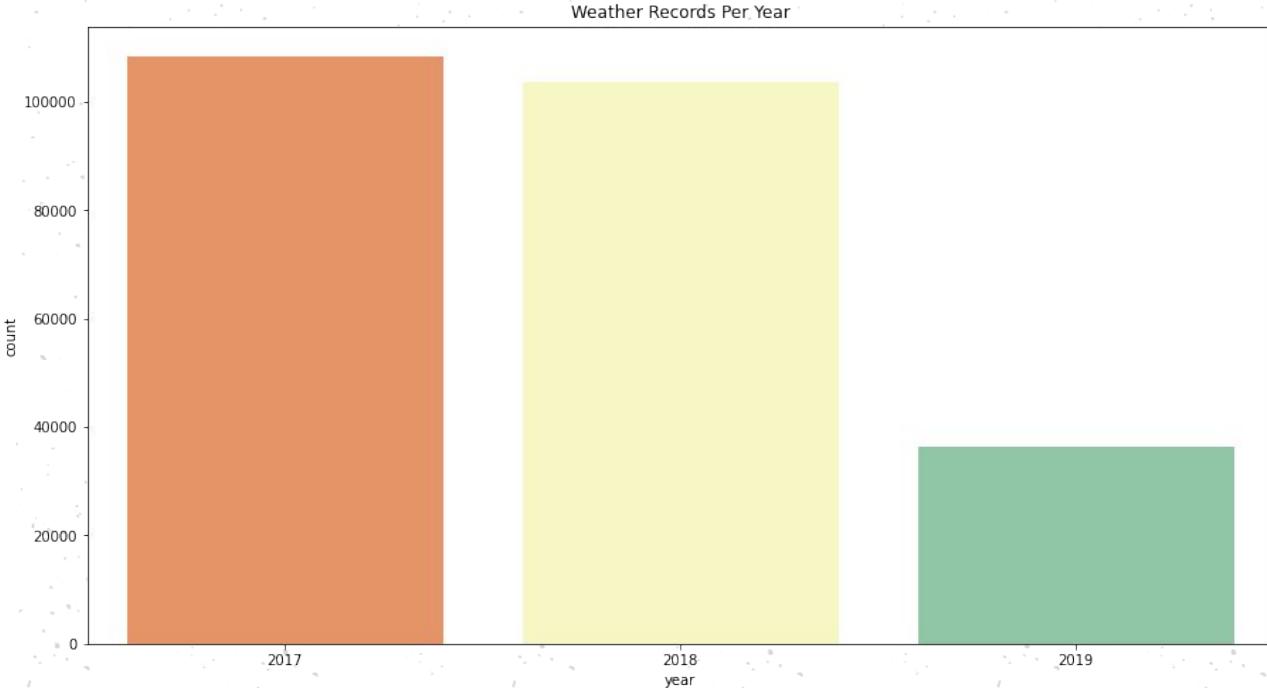


Provinces

Provinces distribution in the dataset

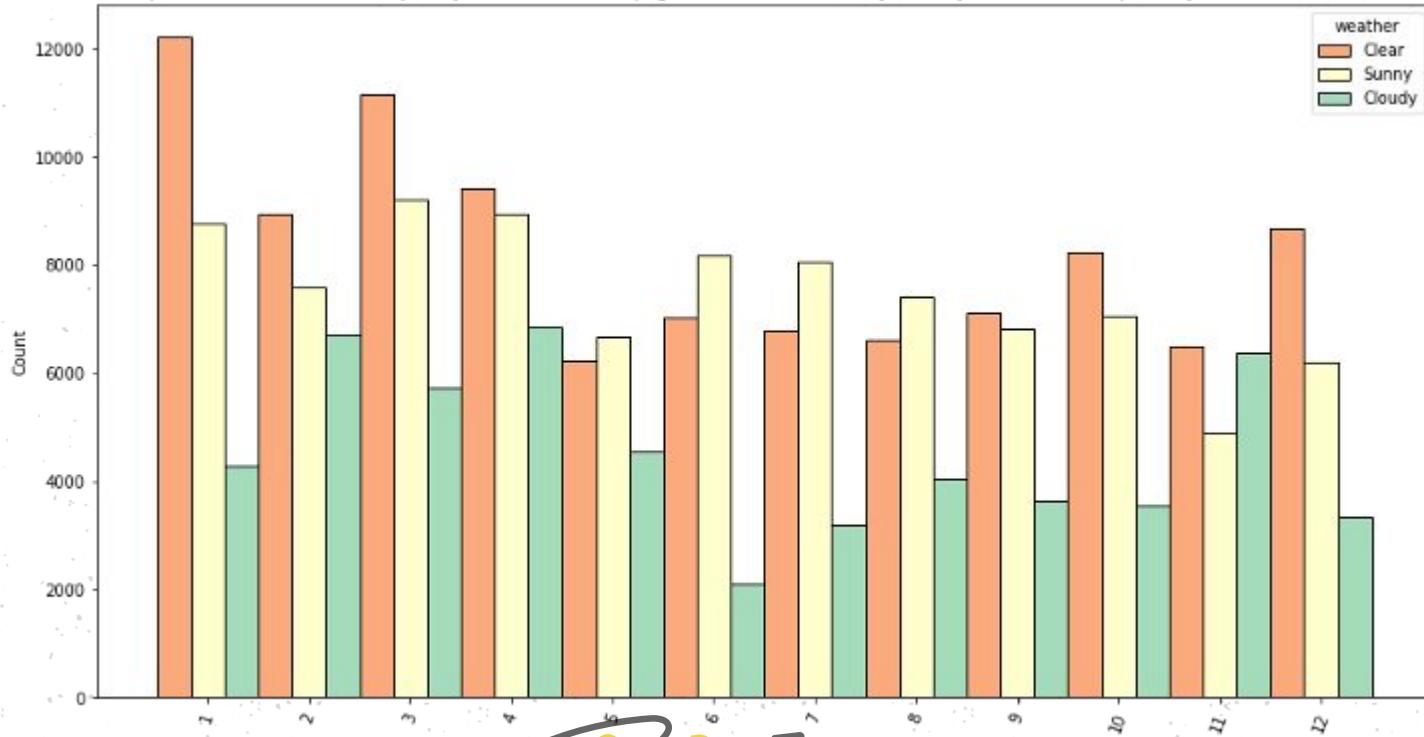


Years



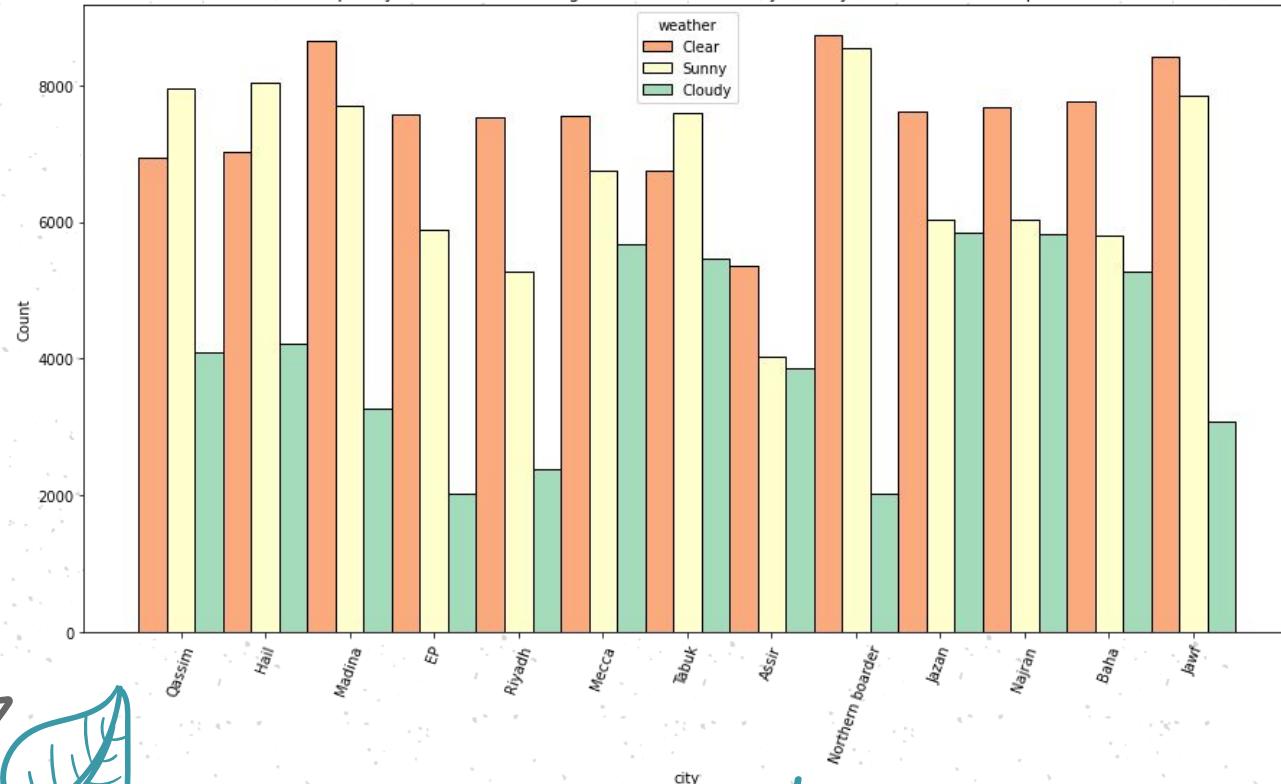
Weather Conditions X Months

the frequency of the most occuring weather (Clear, Sunny, Cloudy)conditions in each city



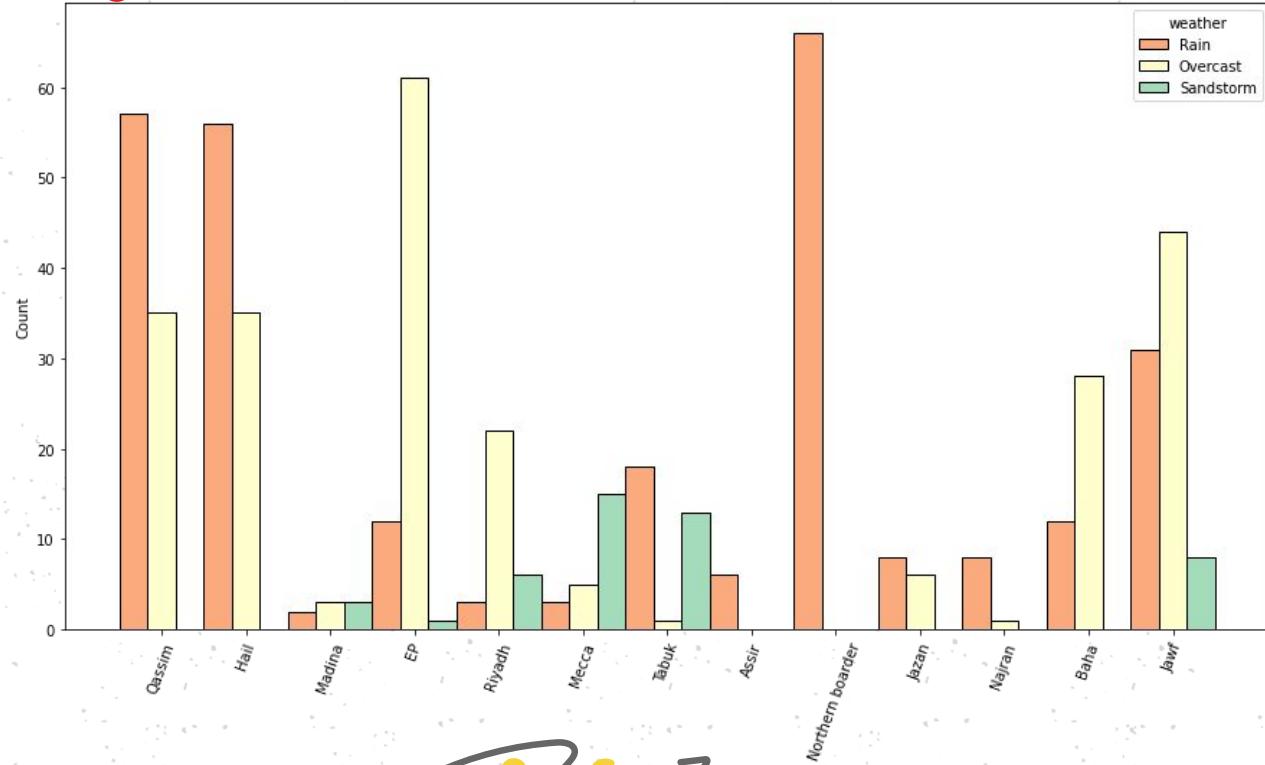
Weather Conditions X Provinces

the frequency of the most occuring weather (Clear, Sunny, Cloudy) conditions in each province



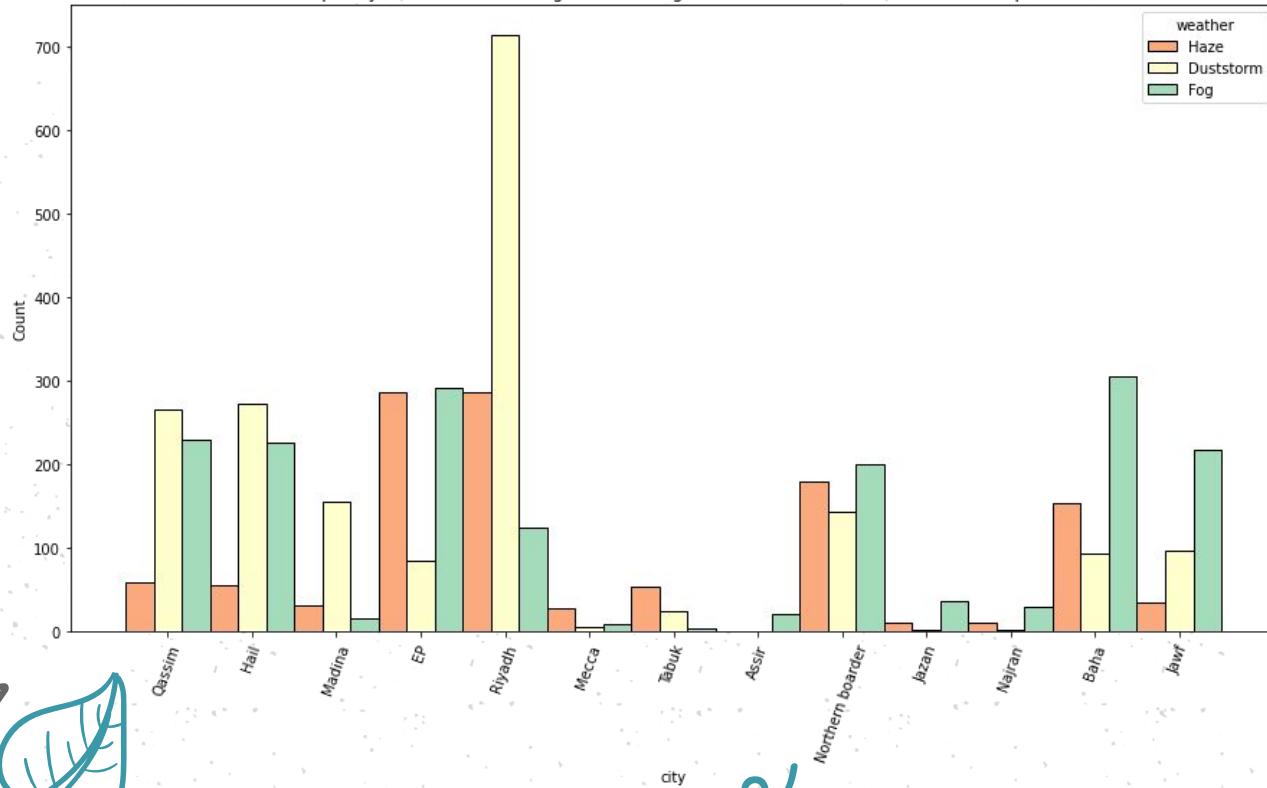
Weather Conditions X Provinces

the frequency of the most occuring weather (Sandstorm, Overcast, Rain) conditions in each province

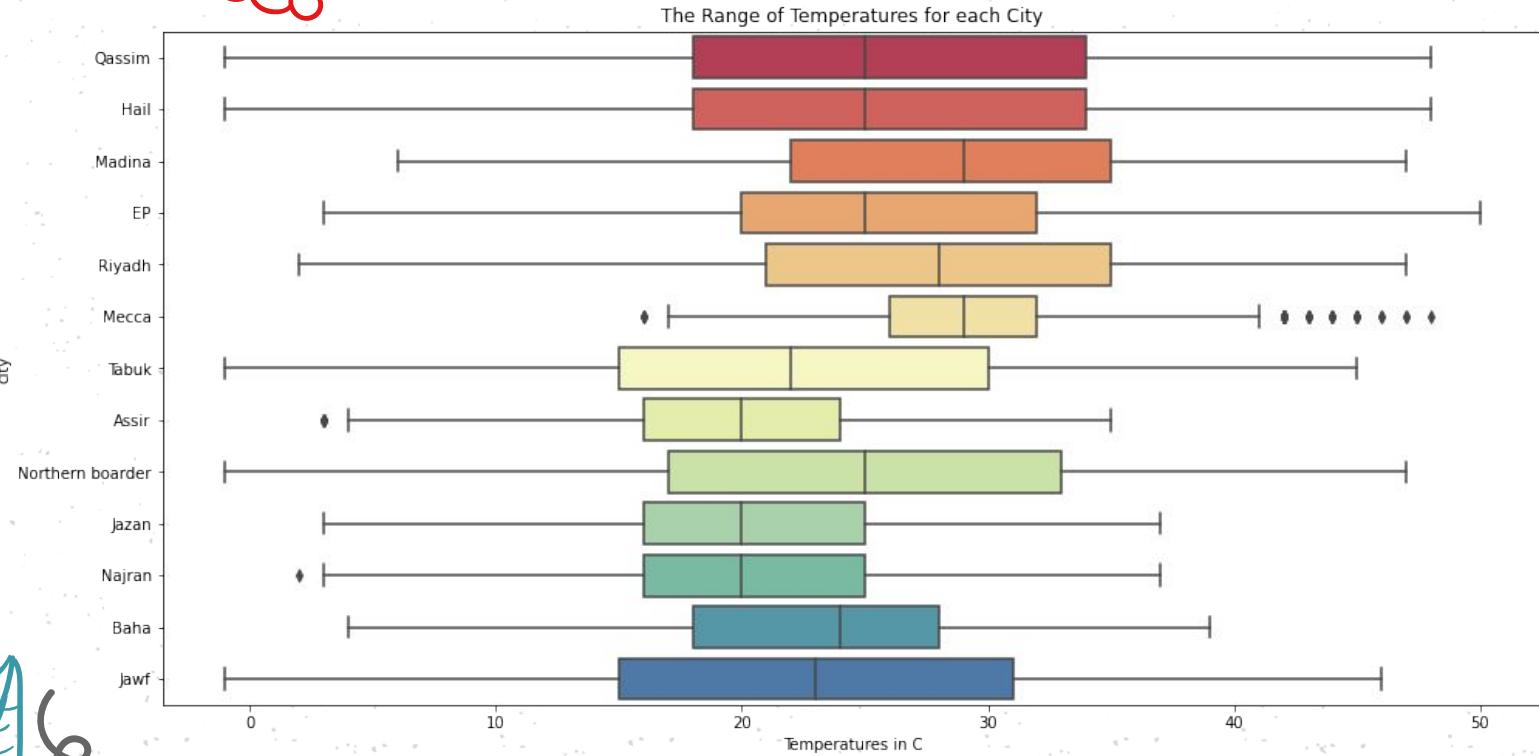


Weather Conditions X Provinces

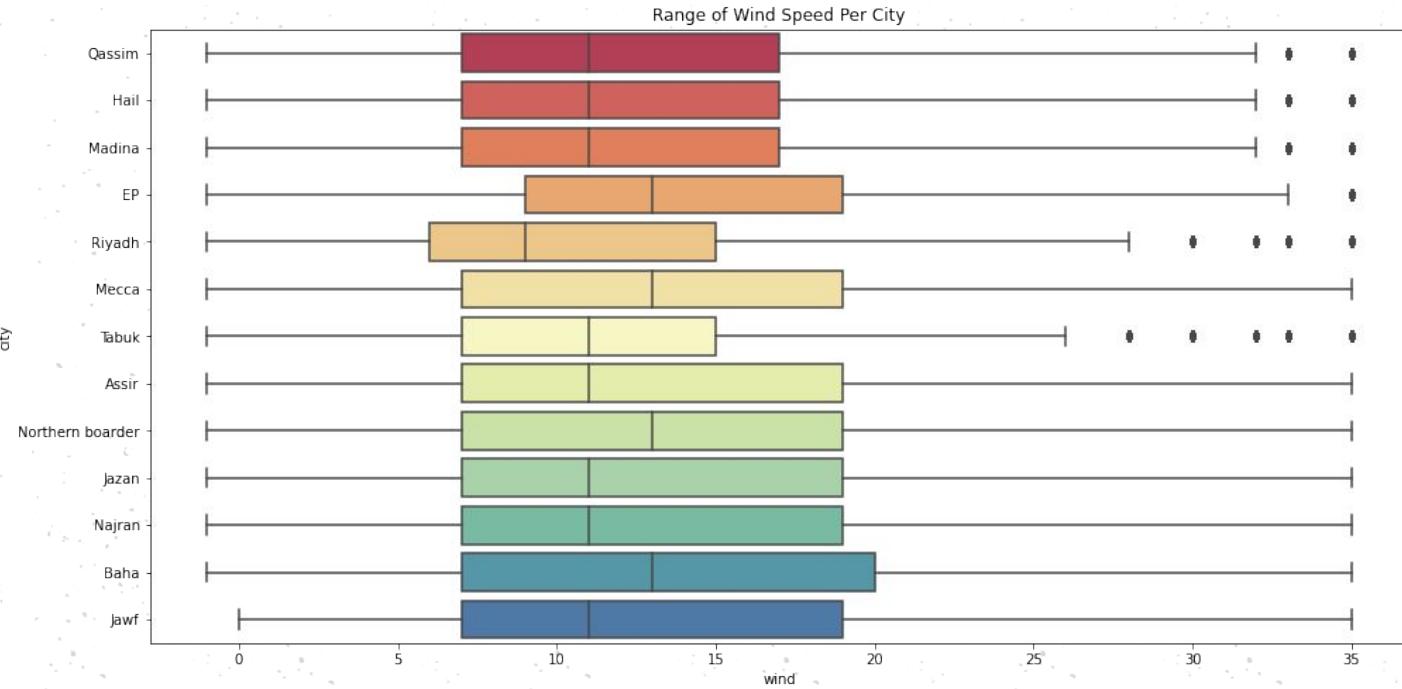
the frequency of the most occurring weather (Fog, Haze, Duststorm) conditions in each province



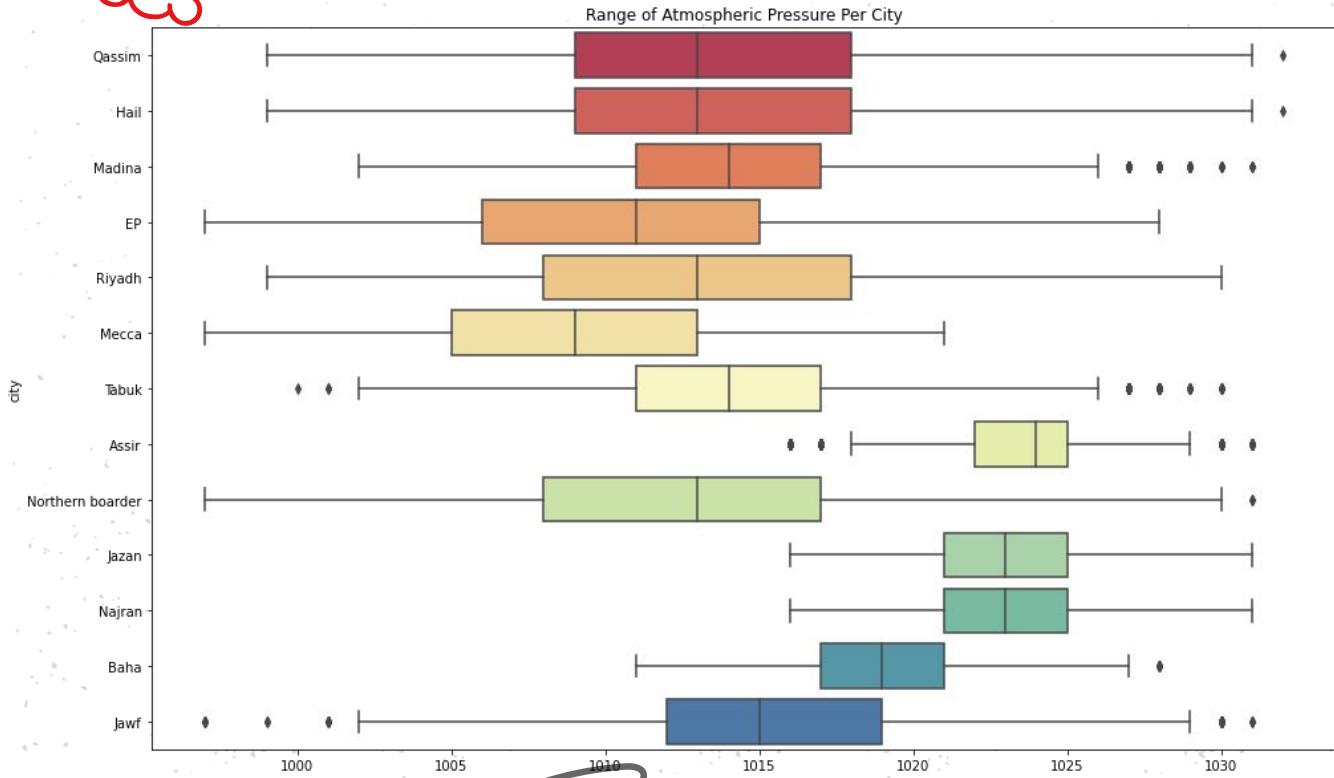
Provinces X Temperature



Provinces X Wind Speed



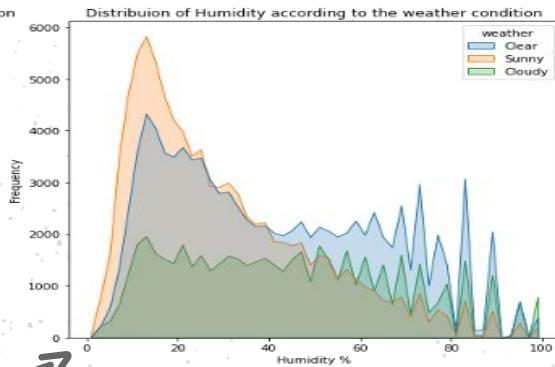
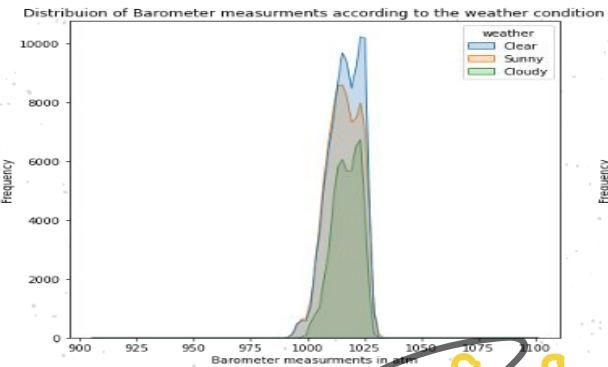
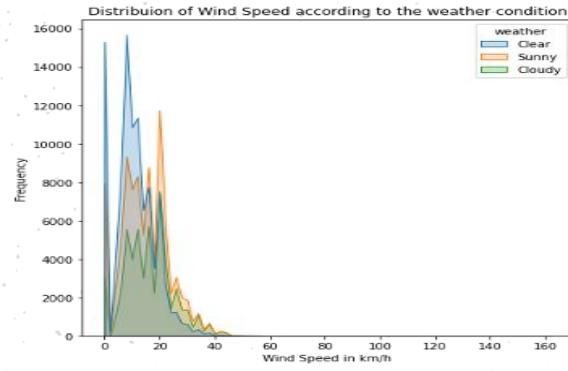
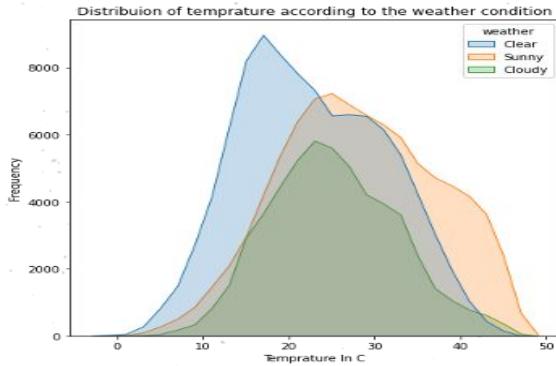
Provinces X Atmospheric Pressure



Weather Conditions X

Temperature, Atmospheric Pressure, Wind and Humidity

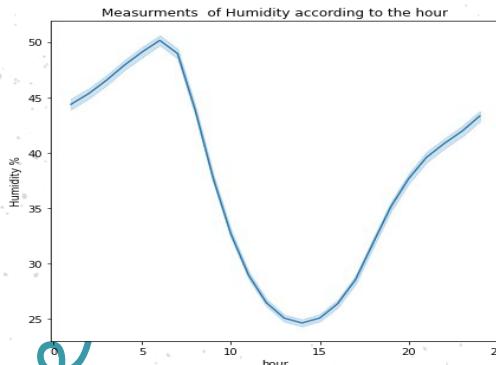
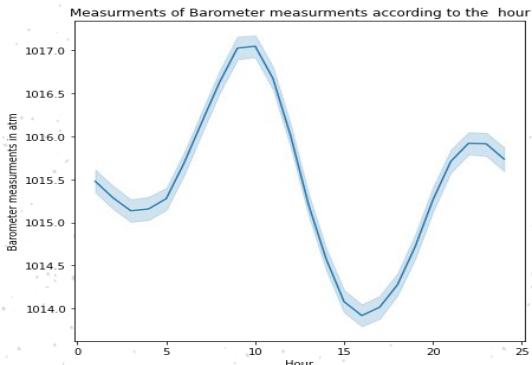
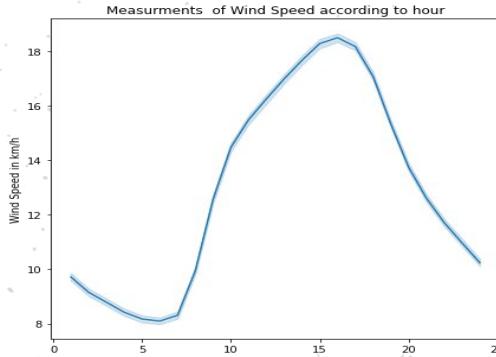
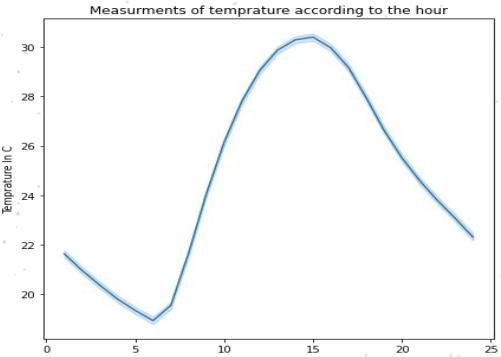
Distribution of different weather metrics & attributes according to the most occurring weather conditions



Hours X

Temperature, Atmospheric Pressure, Wind and Humidity

Measurment of weather metrics & attributes according to the hour



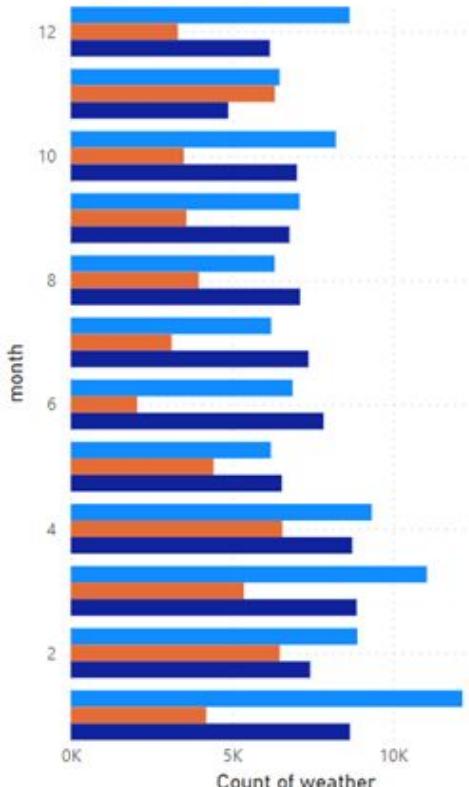
Dashboards



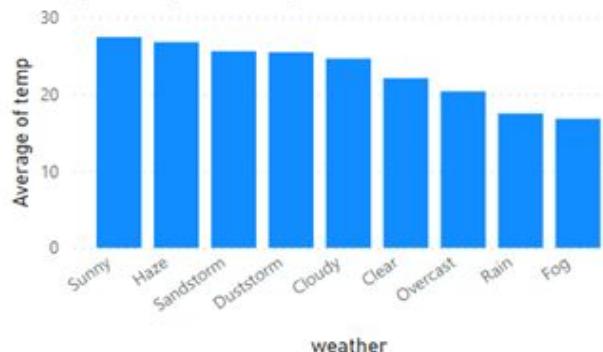
Saudi Arabia Weather Forecasting Dashboard

Count of weather by month and top 3 weather conditions

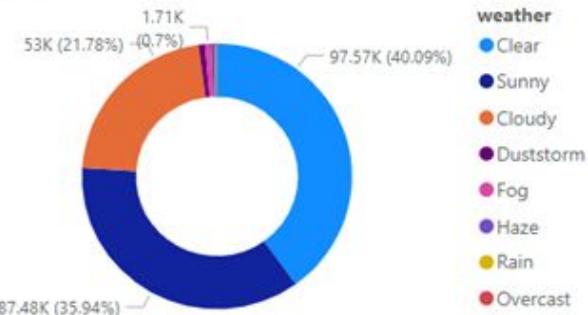
weather ● Clear ● Cloudy ● Sunny



Average of temperature by weather conditions

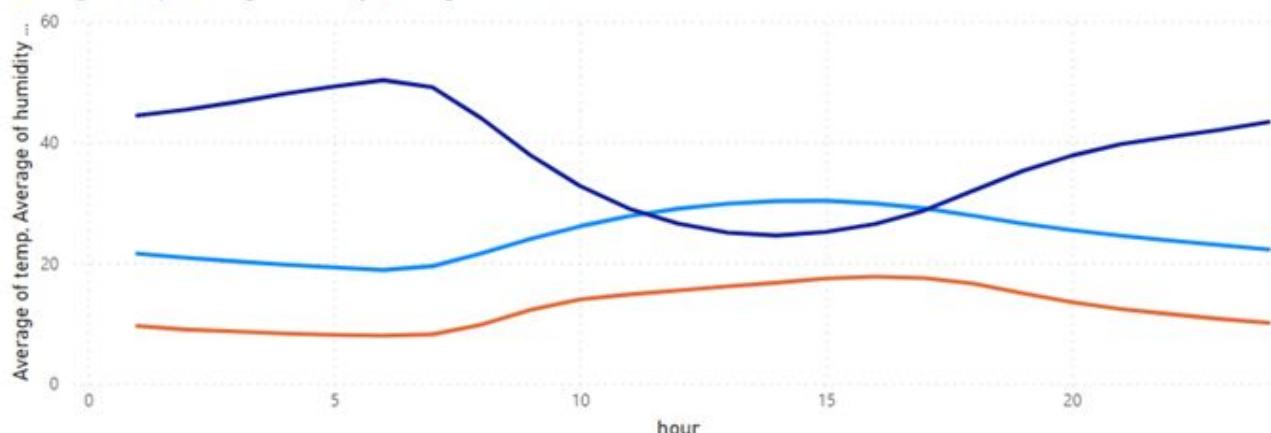


The proportions of weather conditions



The average temperature, humidity, and wind speed by hour

● Average of temp ● Average of humidity ● Average of wind



First Dashboard

This dashboard consists of 4 different charts that depicts Saudi Arabia weather forecasting dataset.

Chart 1: Clustered bar chart

The chart depicts the count of the top 3 weather conditions clear, cloudy, and sunny by month.

Chart 2: Stacked column chart

The chart shows the average temperature by weather conditions. As shown, sunny has the highest temperature and fog has lowest temperature.

First Dashboard Cont.

Chart 3: Donut chart

The chart represents the proportions of weather conditions. As shown, clear, cloudy, and sunny are the most dominant weather conditions.

Chart 4: Line chart

The chart depicts the average temperature, humidity, and wind speed by hour.

Saudi Arabia Weather Forecasting Maps Dashboard

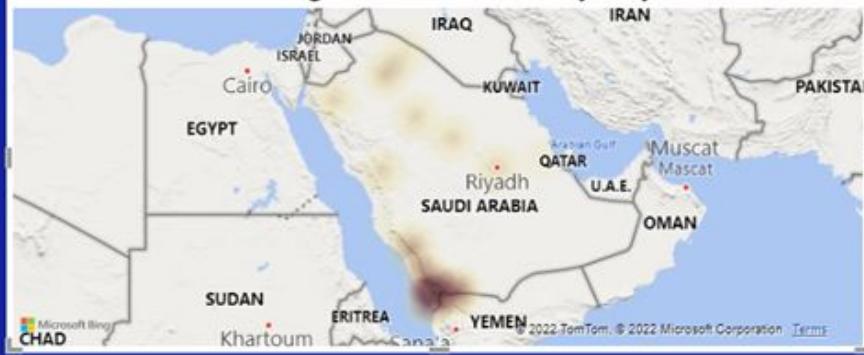
Count of city



Average of temperature by city



Average of barometer by city



Average of visibility by city



Second Dashboard

This dashboard shows 4 different map visualizations of the Saudi Arabia weather forecasting dataset.

Chart 1: The first map depicts the count of all 13 cities with green bubbles where larger bubbles represent higher counts and smaller bubbles represent fewer counts.

Chart 2: The second map depicts the average of temperature by city with the heat map option enabled where blue represents the low temperatures and red represents high temperatures.

Second Dashboard Cont.

Chart 3: The third map depicts the average barometer (atmospheric pressure) by city with the heat map option enabled where yellow represents the low barometers and brown represents high barometers.

Chart 4: The fourth map depicts the average visibility by city with purple bubbles where larger bubbles represent higher visibilities and smaller bubbles represent fewer visibilities

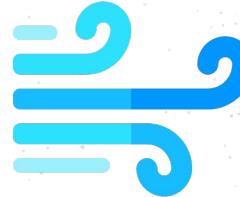
Regression



Targets



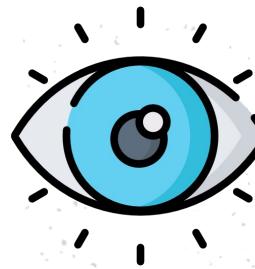
Temperature



Wind Speed

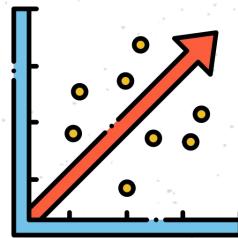


Humidity

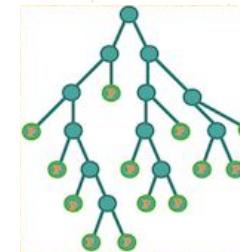


Visibility

Modeling Algorithms



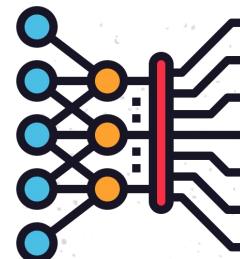
Linear Regression



Random Forest

XGBoost

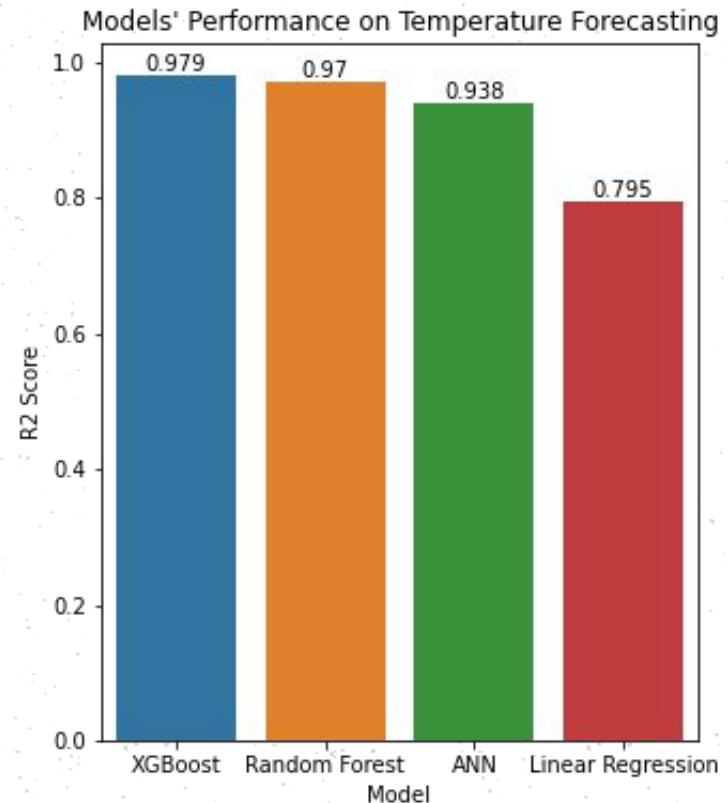
XGBoost



Artificial Neural Networks

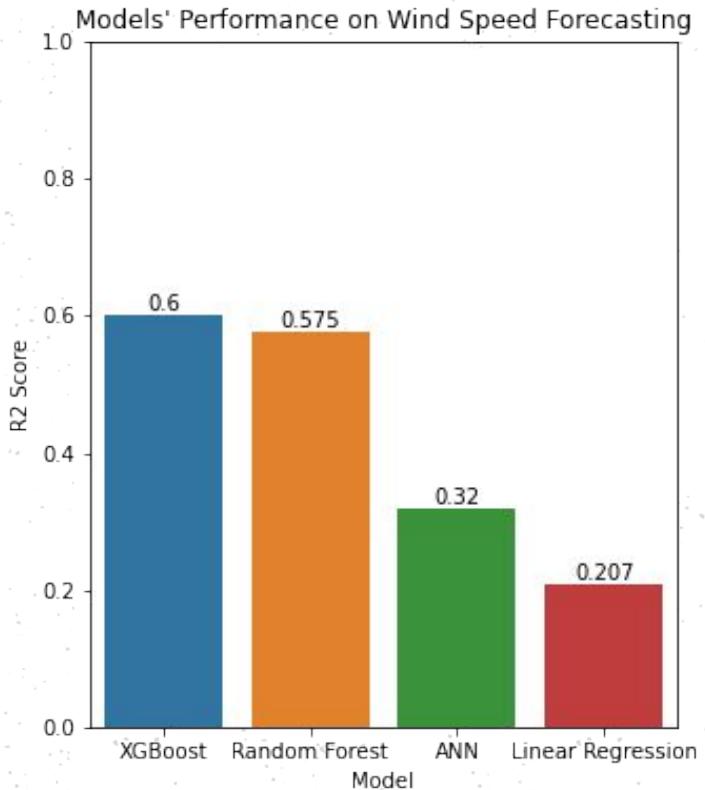
Evaluation: Temperature

	R2	MAE	MSE
Linear Regression	0.795	3.16	16.01
Random Forest	0.97	1.05	2.32
XGBoost	0.979	0.93	1.61
ANN	0.938	1.65	4.79



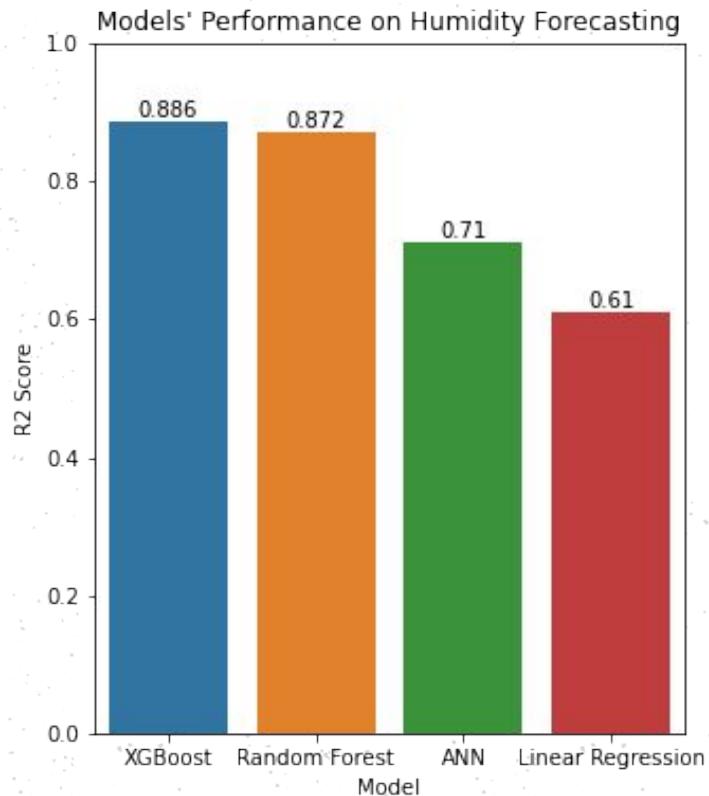
Evaluation: Wind

	R2	MAE	MSE
Linear Regression	0.207	6.10	60.85
Random Forest	0.575	4.20	32.58
XGBoost	0.6	4.15	30.69
ANN	0.32	5.48	52.13



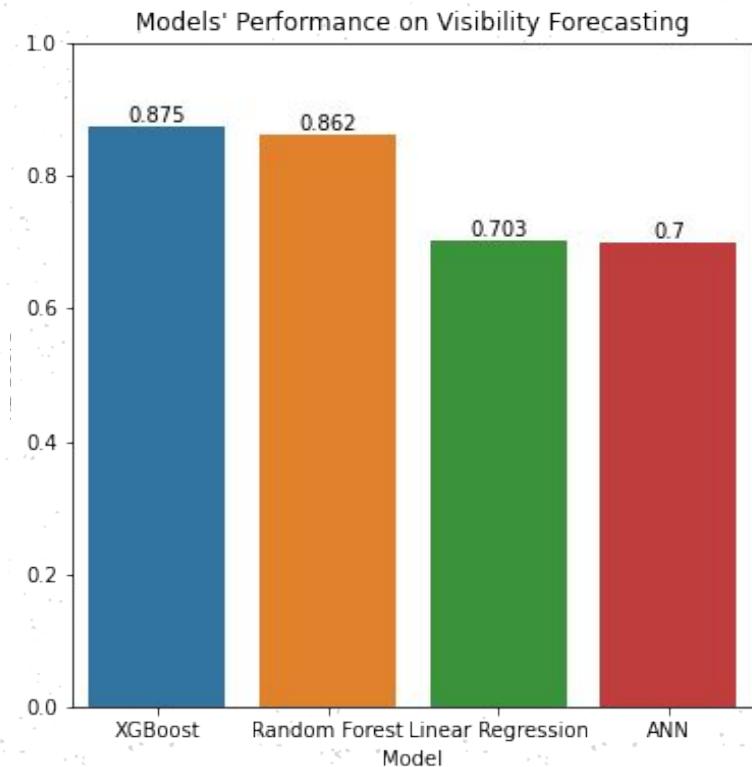
Evaluation: Humidity

	R2	MAE	MSE
Linear Regression	0.61	11.64	217.67
Random Forest	0.872	5.68	71.09
XGBoost	0.886	5.62	63.62
ANN	0.71	9.07	161.41

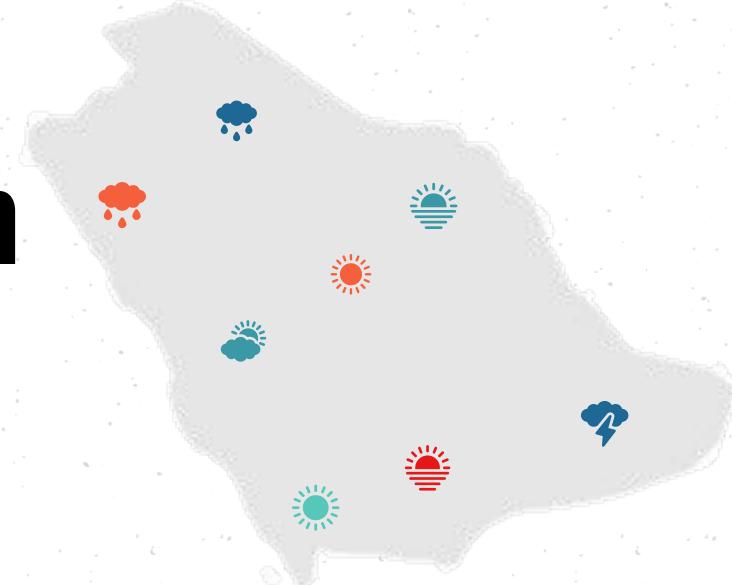


Evaluation: Visibility

	R2	MAE	MSE
Linear Regression	0.703	2.56	14.89
Random Forest	0.862	1.29	6.88
XGBoost	0.875	1.53	6.23
ANN	0.7	1.43	15.03



Classification



Machine Learning Steps

1
Data Preprocessing and Cleaning

2
EDA and data visualization

3
Feature Engineering

4
Split the data

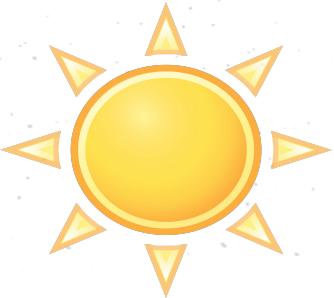
5
Models Building

6
Model Evaluation

7
Hyperparameter Tuning

8
Creating Pipeline

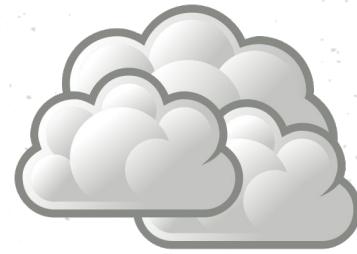
Target: Weather



Sunny



Clear



Cloudy

Baseline & Feature Engineering & Data Splitting

Baseline Model

```
df['weather'].value_counts(normalize=True)
```

```
Clear      0.409873
Sunny      0.367469
Cloudy     0.222658
Name: weather, dtype: float64
```

Feature Engineering, Feature Selection, and Split Data

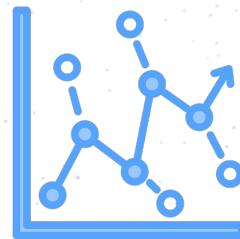
```
X = df.loc[:, df.columns != 'weather']
y = df["weather"]
```

```
le = LabelEncoder()
X['city'] = le.fit_transform(X['city'])
```

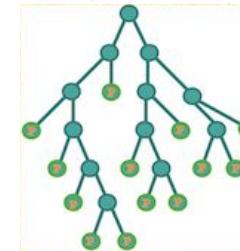
```
X_train, X_test, y_train, y_test = train_test_split(X, y, test_size=0.25, random_state=42)
```

```
sc = StandardScaler()
X_train_scaled = sc.fit_transform(X_train)
X_test_scaled = sc.transform(X_test)
```

Modeling Algorithms



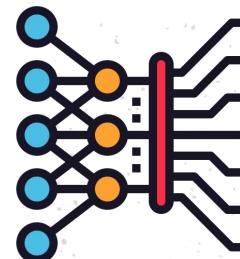
Logistic Regression



Random Forest

XGBoost

XGBoost

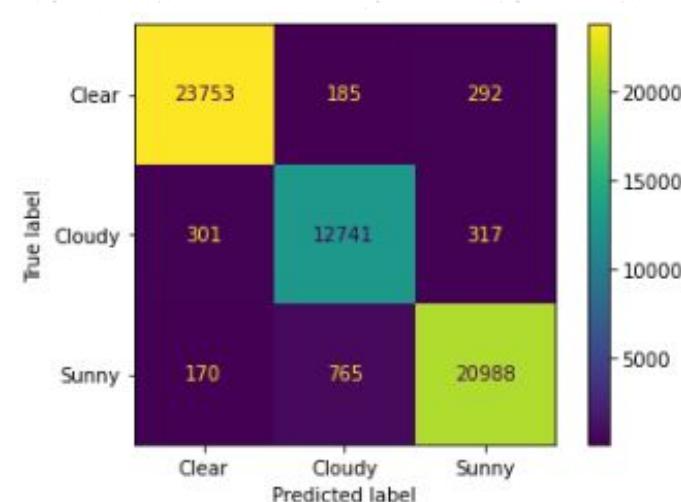


Artificial Neural Networks

Random Forest (Best ML Model)

```
class_forest = RandomForestClassifier(n_estimators = 200, criterion = 'gini', random_state = 42)
class_forest.fit(X_train_scaled, y_train)
forest_class = class_forest.predict(X_test_scaled)
```

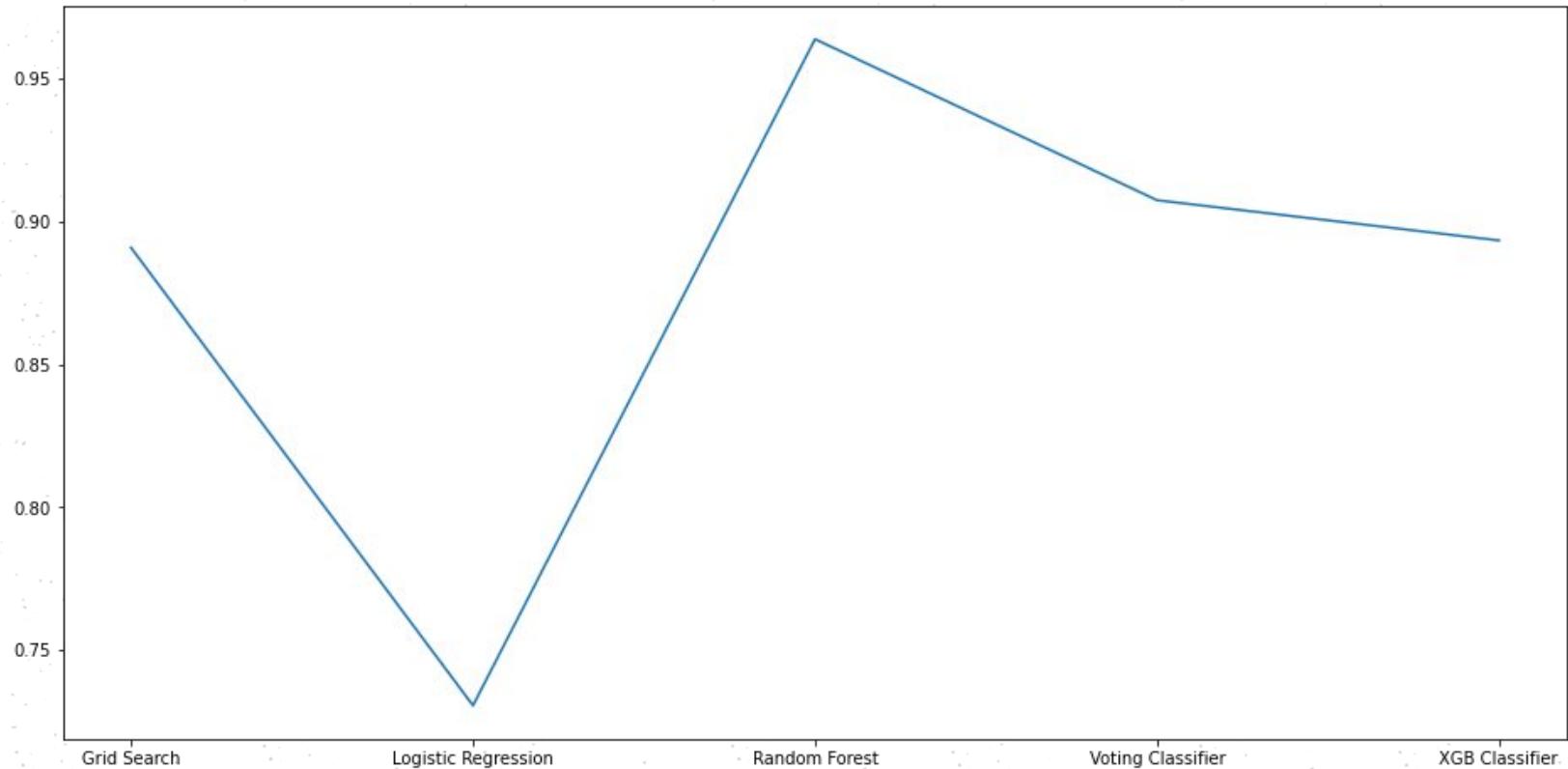
	precision	recall	f1-score	support
Clear	0.98	0.98	0.98	24230
Cloudy	0.93	0.95	0.94	13359
Sunny	0.97	0.96	0.96	21923
accuracy			0.97	59512
macro avg	0.96	0.96	0.96	59512
weighted avg	0.97	0.97	0.97	59512



Modeling AVG Evaluation: Weather

	Accuracy	Precision	Recall	F1
Logistic Regression	0.71	0.72	0.73	0.72
Random Forest	0.97	0.96	0.96	0.96
XGBoost	0.90	0.89	0.89	0.89
ANN	0.92	0.92	0.90	0.91

Balanced Score for ML Models: Weather



Hyperparameter Tuning : Grid Search

```
param_grid = {  
    "n_estimators": [10,20,30,45,50],  
    "criterion": ["gini", "entropy"],  
    "max_depth": [2,4,6]  
}  
grid = GridSearchCV(  
    class_forest,  
    param_grid,  
    cv = 5,  
    n_jobs=-1,  
    verbose=1  
)  
grid.fit(X_train, y_train)  
***  
***  
grid.score(X_test, y_test)  
0.9004066406775104  
grid.best_params_  
{'criterion': 'gini', 'max_depth': 6, 'n_estimators': 50}
```

Hyperparameter Tuning : Voting Classifier

```
from sklearn.ensemble import VotingClassifier
from sklearn.metrics import accuracy_score

VotingModel = VotingClassifier(
    estimators=[('Logistic Regression', lr), ('Random Forest', class_forest),
                ('Grid Search', grid), ('XGB Classifier', class_xgb)], voting='hard')

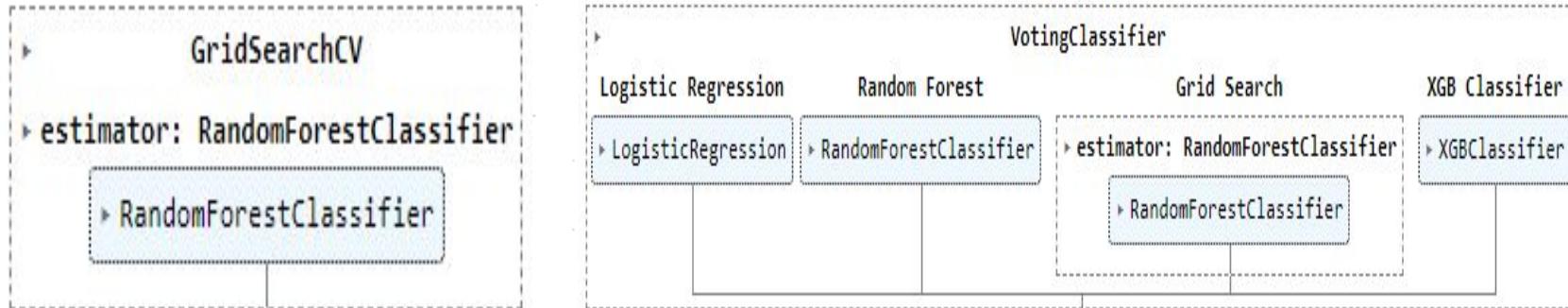
VotingModel.fit(X_train,y_train)

***

predictVoting = VotingModel.predict(X_test)
accuracy_score(y_test, predictVoting)

0.915832101088856
```

Model Optimization Results:



	Accuracy
Grid Search	0.90
Voting Classifier	0.91

ML Pipeline for Best Model

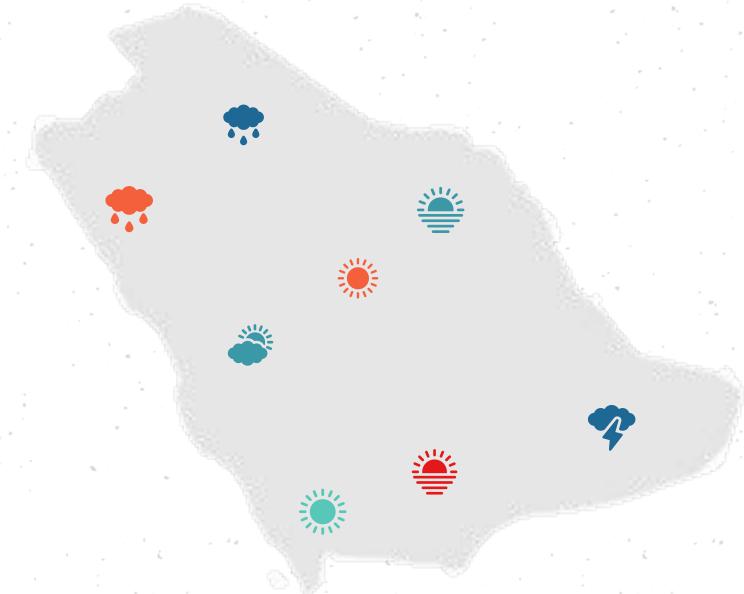
```
X_train_n = X_train.select_dtypes(exclude=["category", "object"])
X_test_n = X_test.select_dtypes(exclude=["category", "object"])

# Create a Pipeline for our model
pipe = make_pipeline(
    SimpleImputer(), # Impute missing values
    StandardScaler(), # scale columns

    # apply the model with best parameter we got from Grid Search
    RandomForestClassifier(n_estimators = 50, criterion = 'gini', max_depth= 6, random_state = 42)
)

pipe.fit(X_train_n,y_train)
pipe.score(X_test_n, y_test)
```

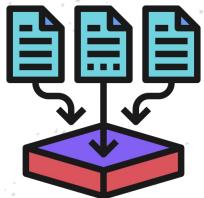
Future Work



Recommendations For Future Work



Advanced time series forecasting algorithms such as ARIMA, Prophet, LSTMs, and Transformers can be used to build more robust weather forecasting models.



Adding more data and more features such as precipitation and wind direction will result in more accurate weather forecasts.



Try building models for different time frames.

Thank You!

Any Questions ?