Bashayer Alahmari   Follow

Sep 2 · 4 min read · ▶ Listen

Save   🐦   ⓕ   in   🔗

Predicting Weather in Saudi Arabia by Using Machine Learning

Contents:

1- Business problem

2- Data description

3- Exploratory Data Analysis (EDA)

4- The approach used to analyze the data and extract the results

5- Power BI Dashboard

1- Business problem:

customers. Transportation departments need to know how much weather will affect the traveling public — whether it's sunny, rain, flooding or other hazards, also in the tourism sector it's very important. These are typical users of accurate weather intelligence to enable critical decisions, but other less-obvious industries also rely on weather forecasts and guidance from professional meteorologists.

For this reason, and given the importance of the weather in the tourism sector and other sectors of the Kingdom's Vision 2030 at the present time, we chose in our project to address the Weather.

2- Data description

The data was downloaded from the Kaggle website https://www.kaggle.com/datasets/esraamadi/saudi-arabia-weather-history. provide historical weather data for all Saudi Arabia cities from 2017 to 2019 in different City of the Kingdom of Saudi Arabia, as it contains 249023 rows, 14 columns giving (temperature, wind, humidity, barometer, and visibility).

3- Exploratory Data Analysis (EDA)

In this part, we created a visualization to help us explore the data and determine the directions of research and further investigation.
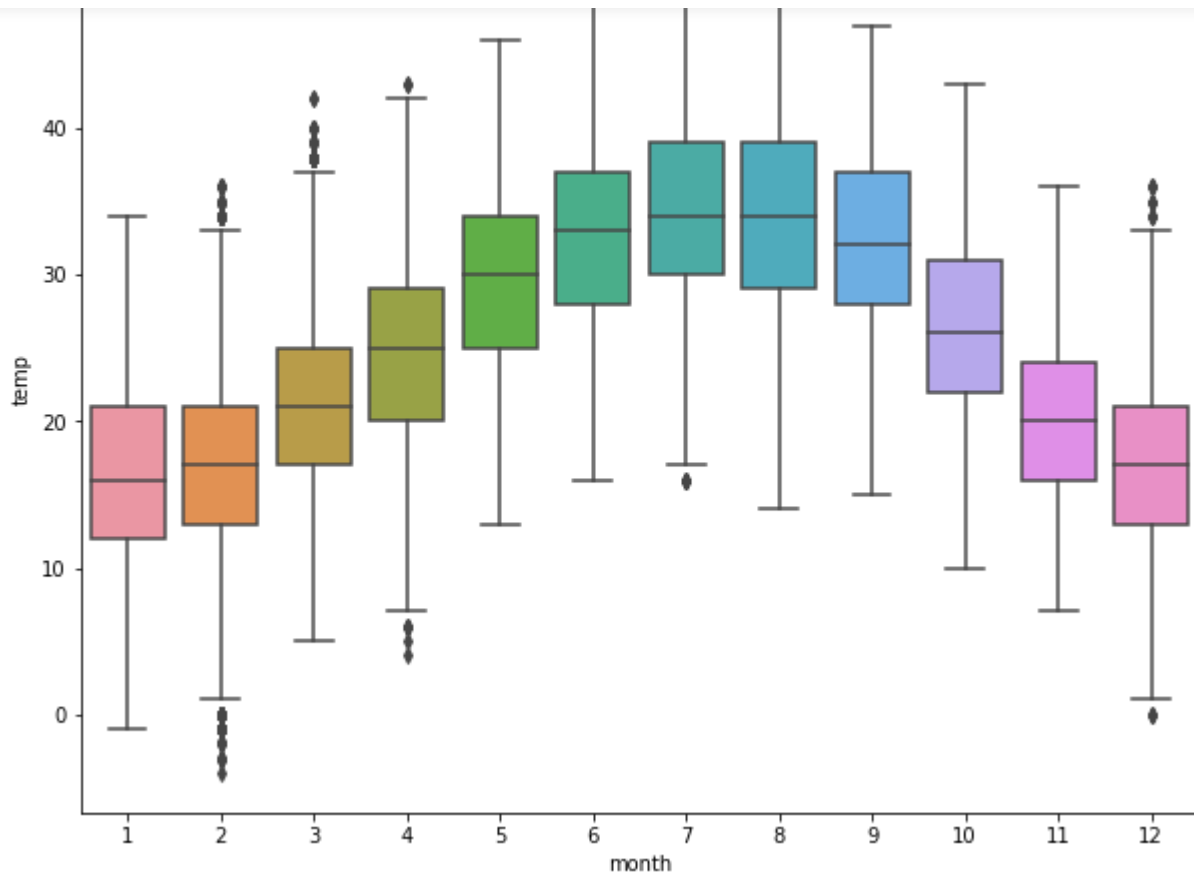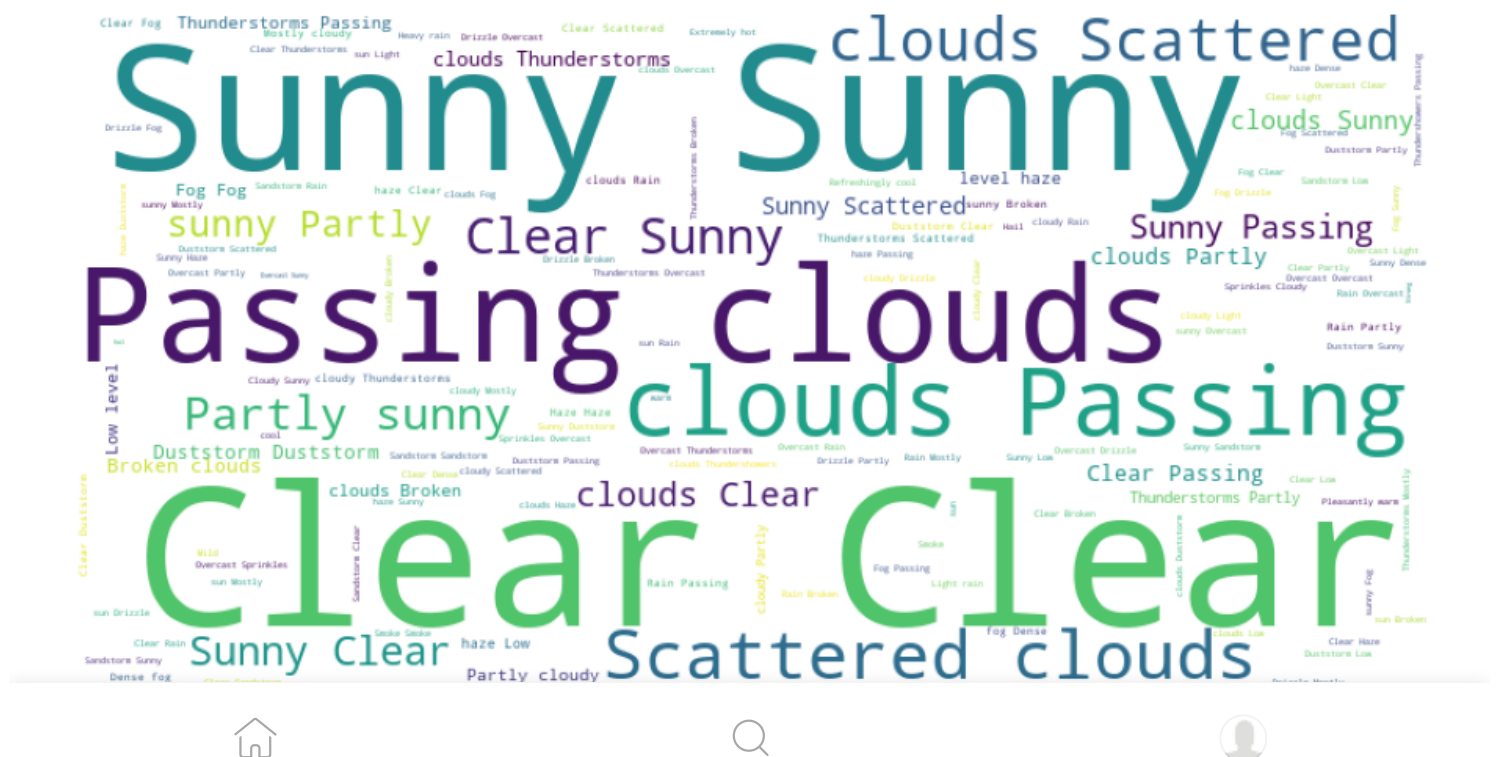
Distribution of temperature
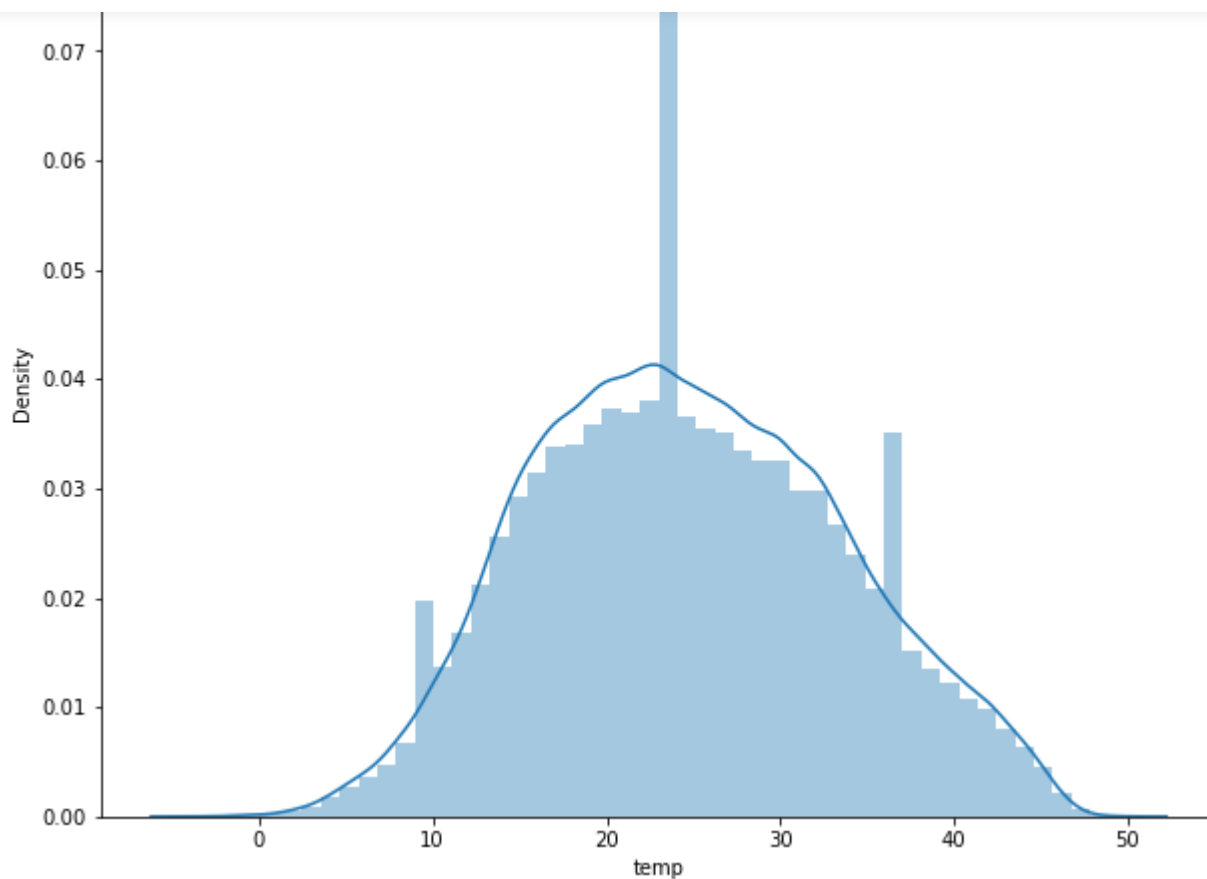
Monthly temperature

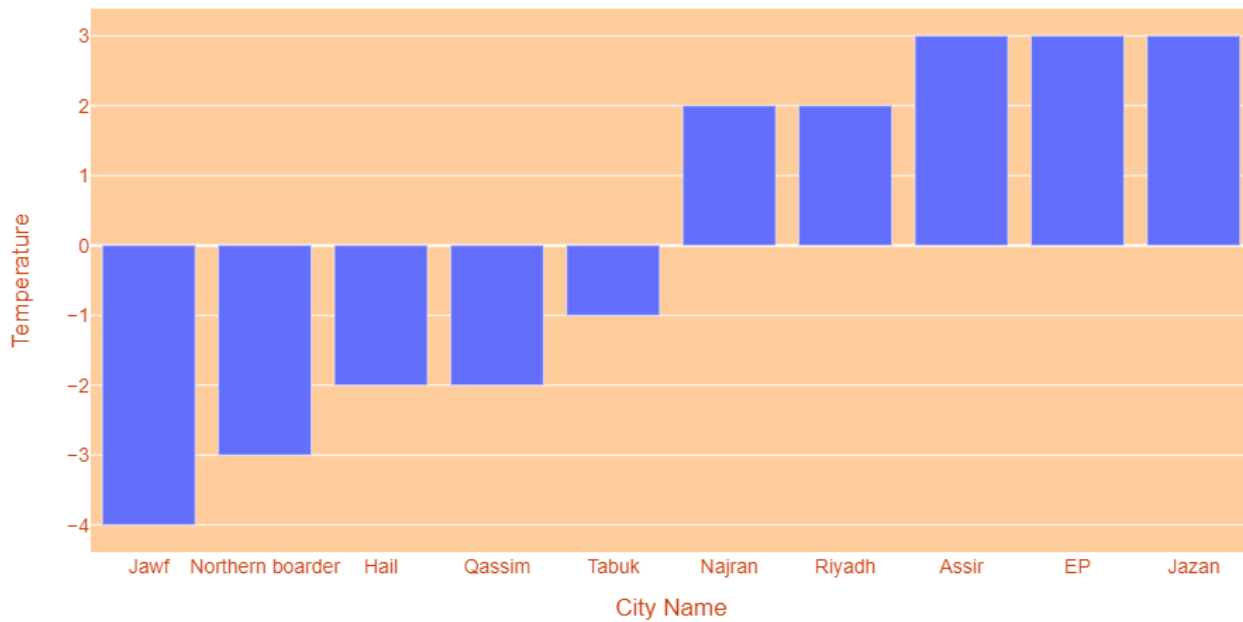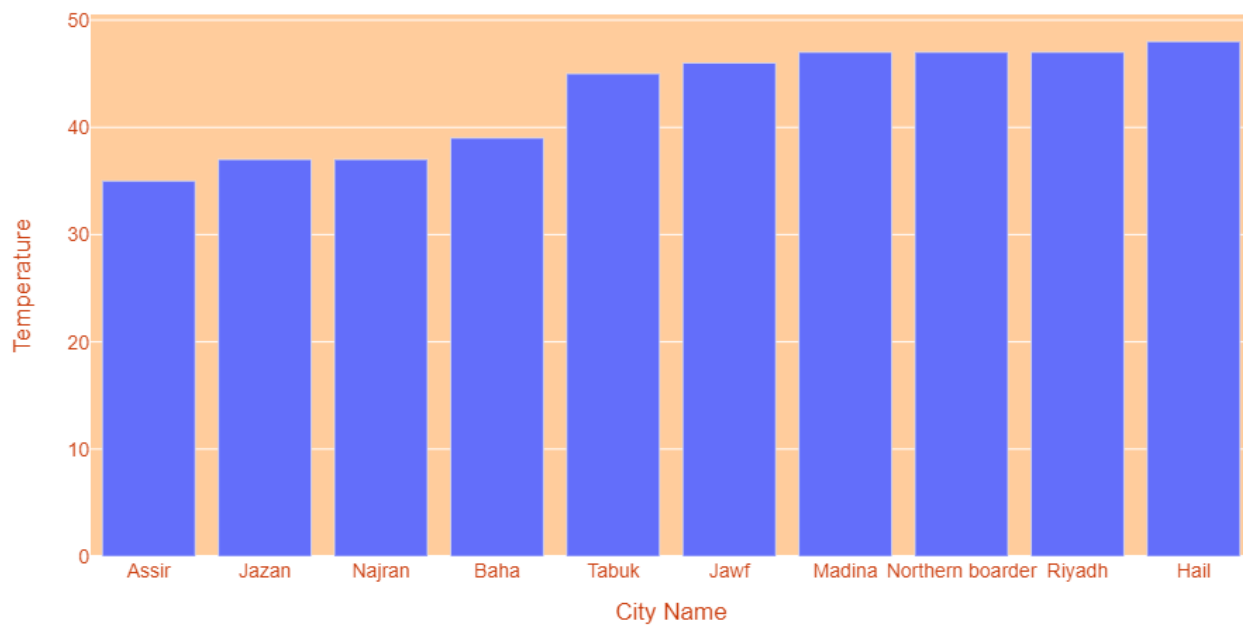## The Most Appeared Climate
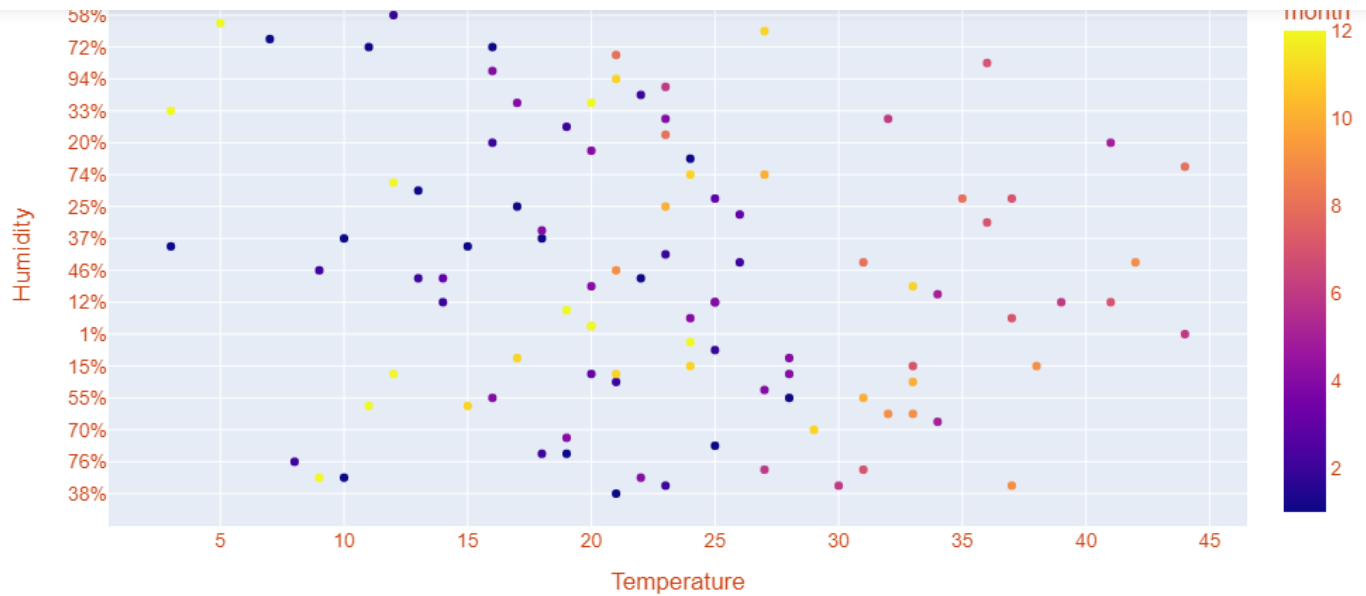
Top 10 cities which the lowest cities in KSA

Top 10 hottest cities in KSA from 2018 till 2019

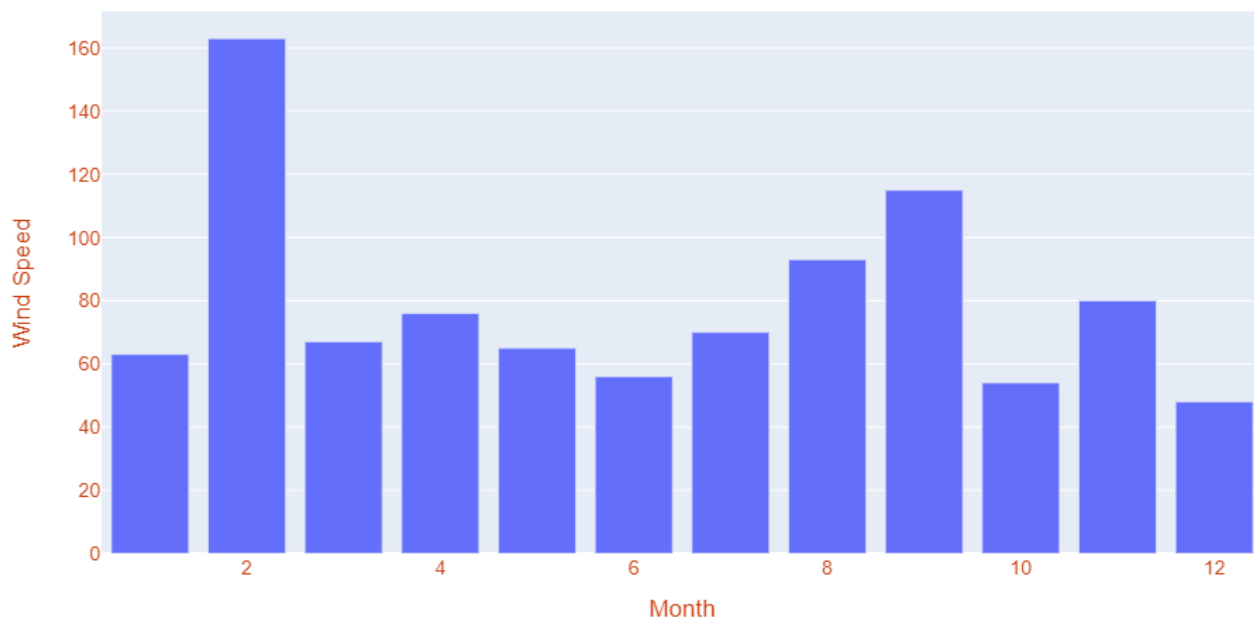Top 10 hottest cities in Saudi Arabia from 2018 till 2019

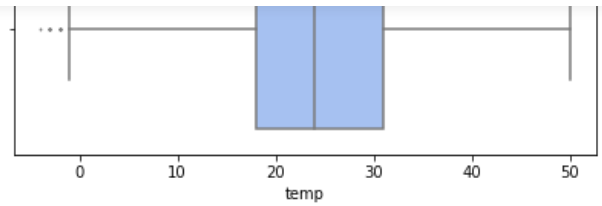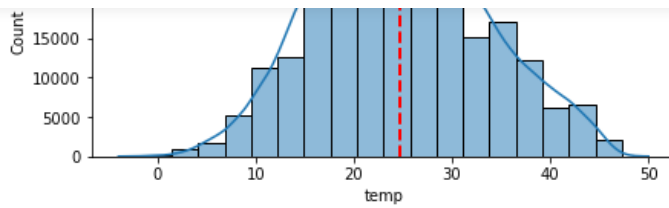
Temperature vs Humidity

Open in app　　Get started

Months with wind speed

Months with wind speed

Temperature vs visibility

Max tempture per city

Max tempture per city
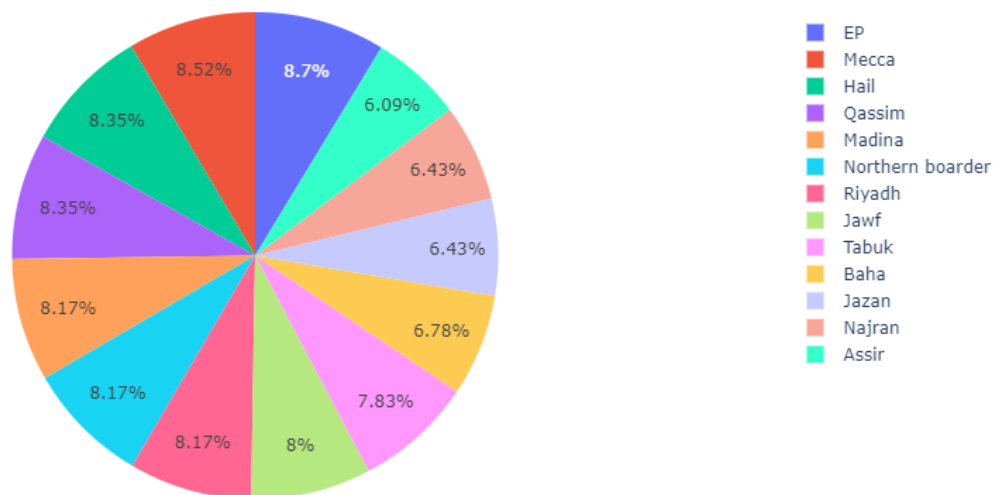
4- The approach used to analyze the data and extract the results

By using python and machine learning we studied and analyzed Weather in Saudi Arabia and we compared them between the city of the Kingdom of Saudi Arabia based on the Temperature. This was done by:

- Cleaning (wrangling) the data:

Data cleaning is the process of fixing or removing incorrect, corrupted, incorrectly formatted, duplicate, or incomplete data within a dataset. Clean data will boost overall productivity and assist you to make better decisions by allowing you to use the best information possible. This was done by:

last value. according to what is appropriate for each subset of the data frame.

- Use pandas.to_datetime() to convert string to datetime format.

- Machine learning models

Machine learning methods assist individuals in exploring, analyzing, and deducing meaning from large data. We followed these steps in our analysis:

- split the data:

In machine learning, it is a common practice to split your data into two different sets. These two sets are the training set and the testing set which works well with large datasets.

Classification

- **Decision Tree**

The goal is to create a model that predicts the value of a target variable by learning simple decision rules inferred from the data features, the result of the Decision Tree Training Accuracy: 86.0% || Test Accuracy: 64.0%
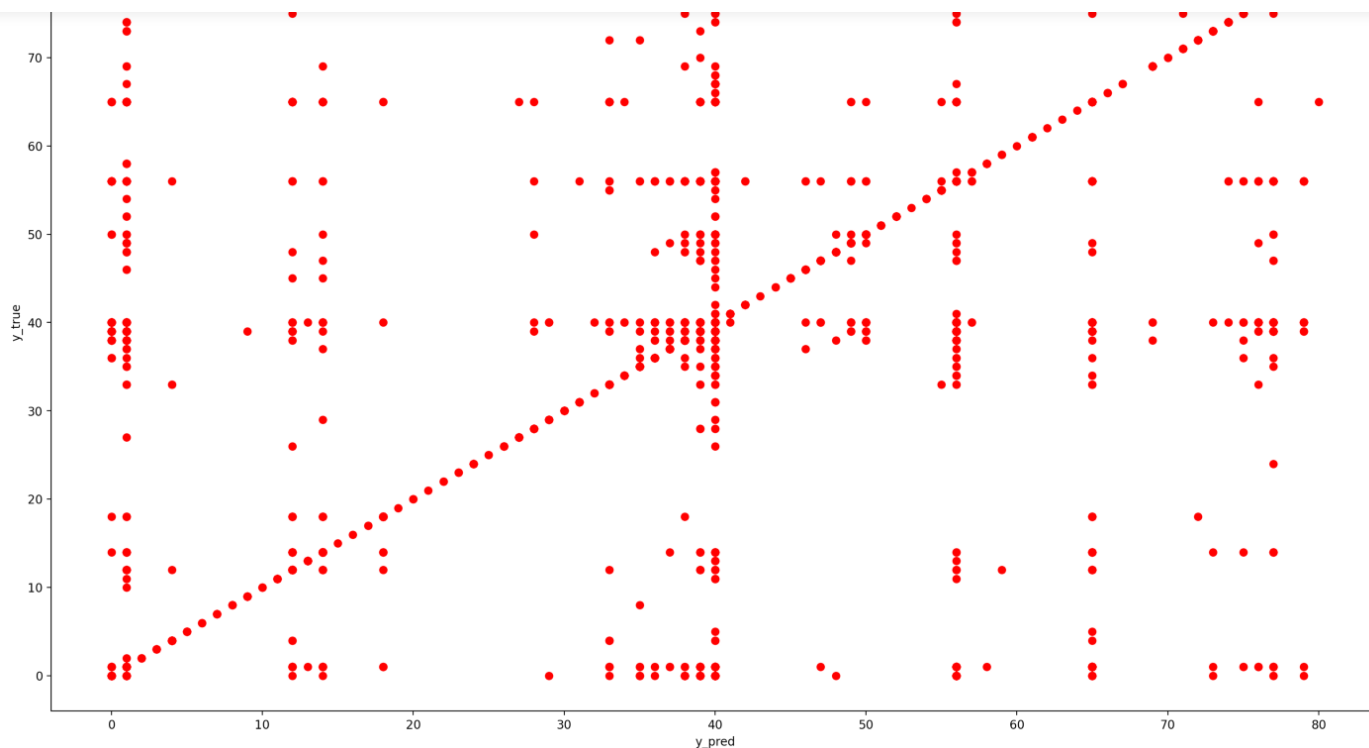
- **Random Forest:**

Random forest is a learning algorithm that works on the principle of Random Forest Regressor Result to decide the final outcome. In this step we made a calculated for some functions as (MSE).
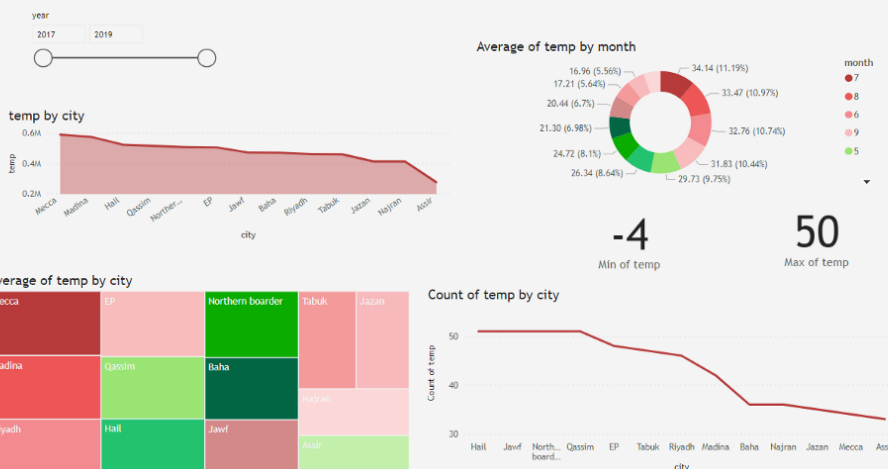
- **Grid Search for Regression**

we have defined the hyperparameters of the model which we want to pass through GridSearchCV for getting the best parameters. Also, we have calculated grid best_ score and grid best_params. Finally, we printed the tree.

5- Power BI Dashboard

Open in app

Get started

About   Help   Terms   Privacy

Get the Medium app