ZAWADI ABIGAEL MUKURU – DATA ENGINEER, ML ENGINEER

WEATHER EXPLORATORY ANALYSIS REPORT

11TH August 2024

**INTRODUCTION**

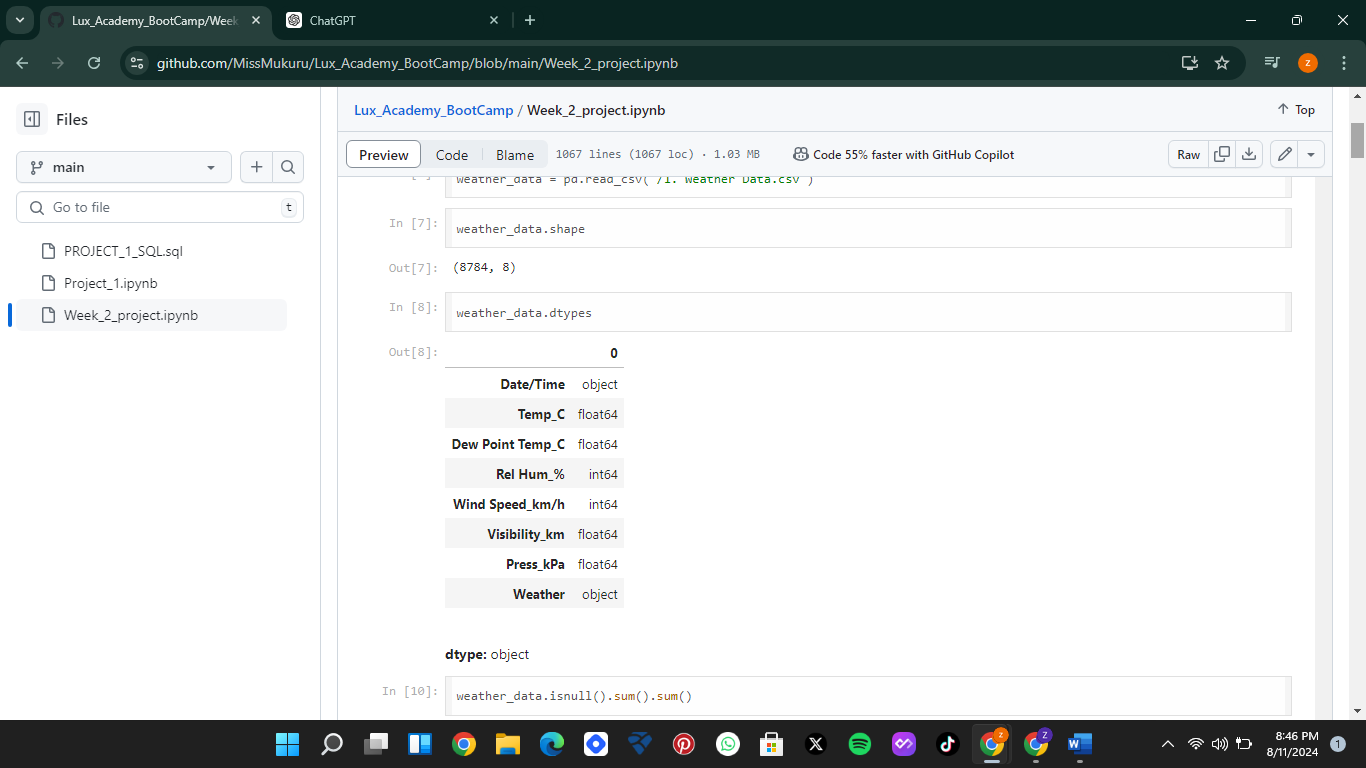
The purpose of this report is to provide insights on the weather dataset.

The dataset was obtained from Kaggle , named 1.Weather Data.csv, it contains 8 columns and 8784 rows.

Some of the major objectives of this report is to;

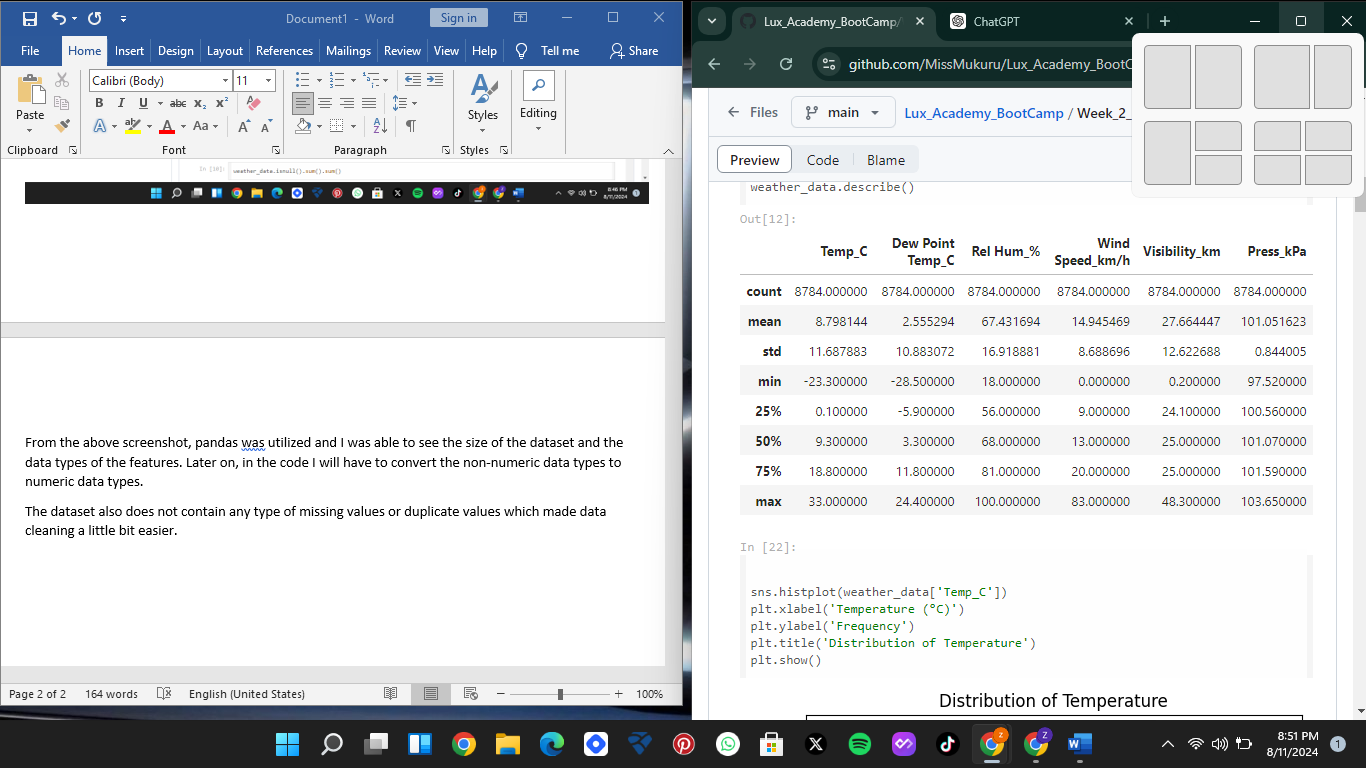
* Explain the trends in the data.
* Figure out the patterns in the data.
* Explain the visualizations I came up with.
* Suggest other areas for further analysis.
* Summarize the key insights gained from the Eda.

**DATA DESCRIPTION**



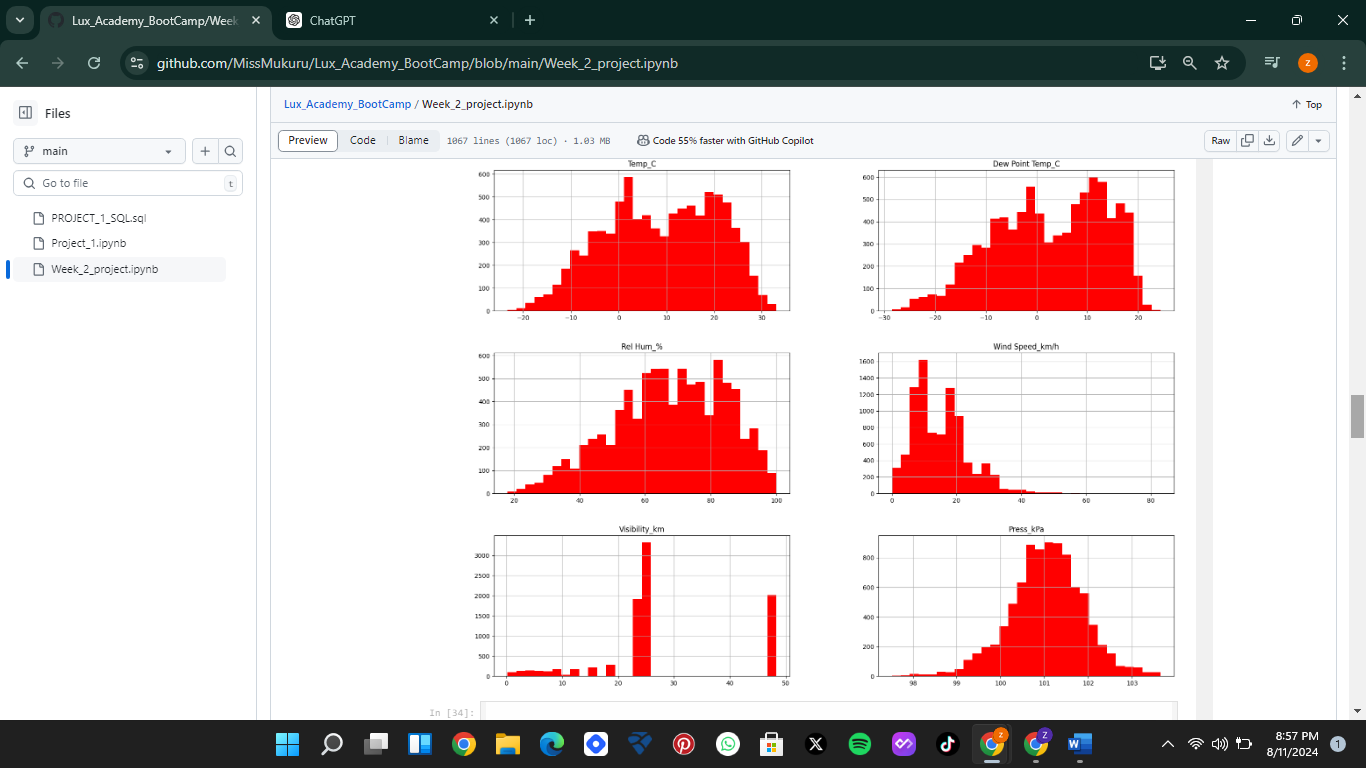
From the above screenshot, pandas was utilized and I was able to see the size of the dataset and the data types of the features. Later on, in the code I will have to convert the non-numeric data types to numeric data types.

The dataset also does not contain any type of missing values or duplicate values which made data cleaning a little bit easier.



Also as seen from the screenshot above I was able to derive a summarized statistical overview of the features/columns in the data set.

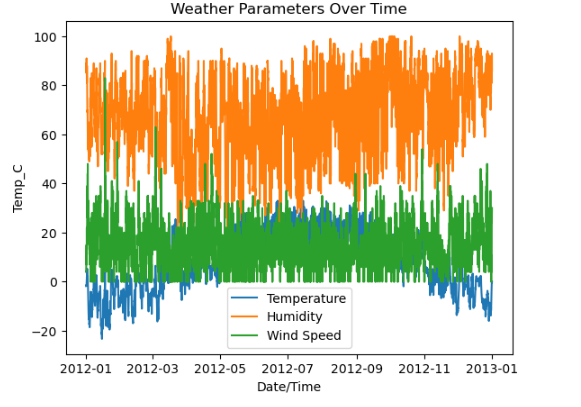
**EXPLORATORY DATA ANALYSIS ON THE DATA**

1. Distribution of the key weather parameters

We can see from the above distribution that

1. Temperature (Temp\_C), Press KPA – seems to be a rough bell curve showing normal distribution where the mean is perfectly symmetrical. It starts with a low value, rises to the mean and then falls back to the same low level. Also, the mean (average), median (middle value), and mode (most frequent value) are all equal and located at the center of the distribution.
2. Rel Hum\_%,Dew Point Temp\_C – These two seem to be leftly skewed, that means the peak of the distribution is to the right of the center, and most data points are stores in in the right side. In a left-skewed distribution, the mean is typically less than the median. The mean is pulled to the left by the lower values in the left tail. This means that the data may have very low values
3. Wins Speed\_Kmh – This appears to be rightly skewed and the tail on the left side is longer, the mean is also less than the median. Data may have a few very high values that pull the mean to the right.

Now in relation to the distribution of the weather features I drew up the following conclusion



Temperature

* Started low between January and March, with the highest being slightly above zero, and the lowest being at around -20 degrees
* The temperatures increased between March and November they sore a little higher to a mean of around 8 degrees
* They sored lower between November and December with the highest being around degrees and the lowest at around -20 degrees
* From around April to August the temperatures increased a bit but then from September they dropped drastically
* This suggest that there was a change in seasons in the year attributing to the extreme temperature changes.

**Humidity**

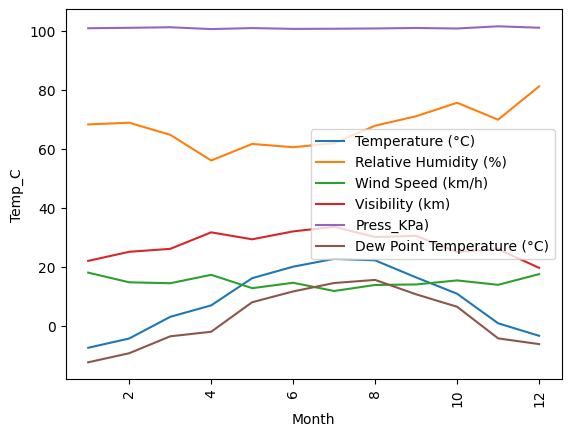
* Humidity seems averagely high in 2012, with the highest around 83 and the lowest at around 60
* The lowest month being April, and the highest being in December
* This may suggest the area the data was recorded from a relatively humid area.

**Wind Speed**

* The wind speed is relatively low but seemed to sore about February.
* Then we see that the wind speed that year was not as erratic as the temperature.

**Correlation of the key weather parameters to each other.**

* An increase in temperature causes a slight increase in wind and the vice versa is true.
* An increase in temperature also causes and increase in Dew point temperature. This means that they are positively correlated.
* Visibility increases as wind speed decreases which means they are negatively correlated.



While the relationships observed are informative, further research could explore the causative factors behind these correlations. For example, investigating how temperature changes specifically influence wind patterns or how different types of particles affect visibility could provide deeper insights.

The positive correlation between temperature and dew point temperature provides useful insights for understanding humidity levels and weather forecasting. The relationship between temperature and wind speed might also be considered in weather models to predict wind patterns based on temperature changes.