5CS037 – CONCEPT AND TECHNOLOGY OF AI

Name: SUMIT SHAH

Group: L5CG17

Student ID: 2358830

Tutor

Mr. NABIN ACHARYA

Contents

DATA MANAGEMENT WITH AI

Introduction:

The following report is about the data analysis and statical representation of temperature and humidity

The dataset is extracted from Kaggle-a datata science community with various variable data .

We import the csv files into the google colab and perform all the tasks. We mainly used these tools and packages for our assignment: pandas, matplotlib.pyplot, and seaborn.

Data Pre-processing and Statistical Interpretation:

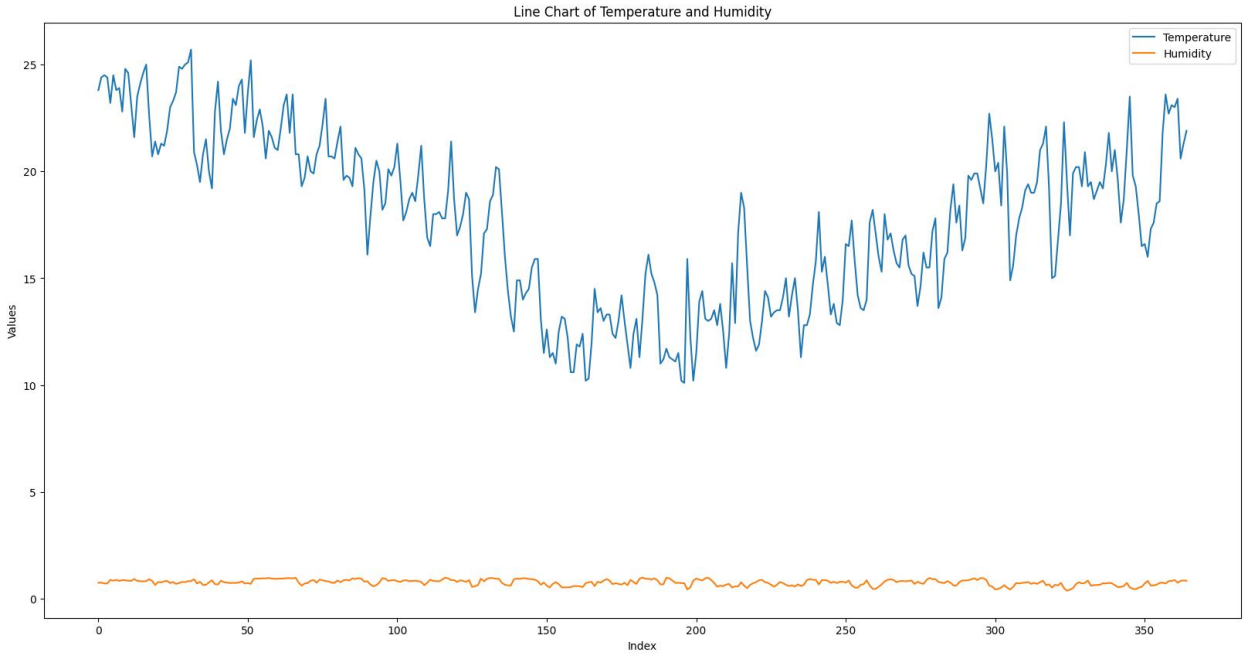
Firstly, csv file of data was imported. Then printed all the data present in our csv file. Seeing all the data of our csv file like description, information, size of our data frame, variables used for the different data types. These were done to check all the null values (missing value) and duplicate (repeated value) or faulty values and check if anything was wrong.

Data cleaning:

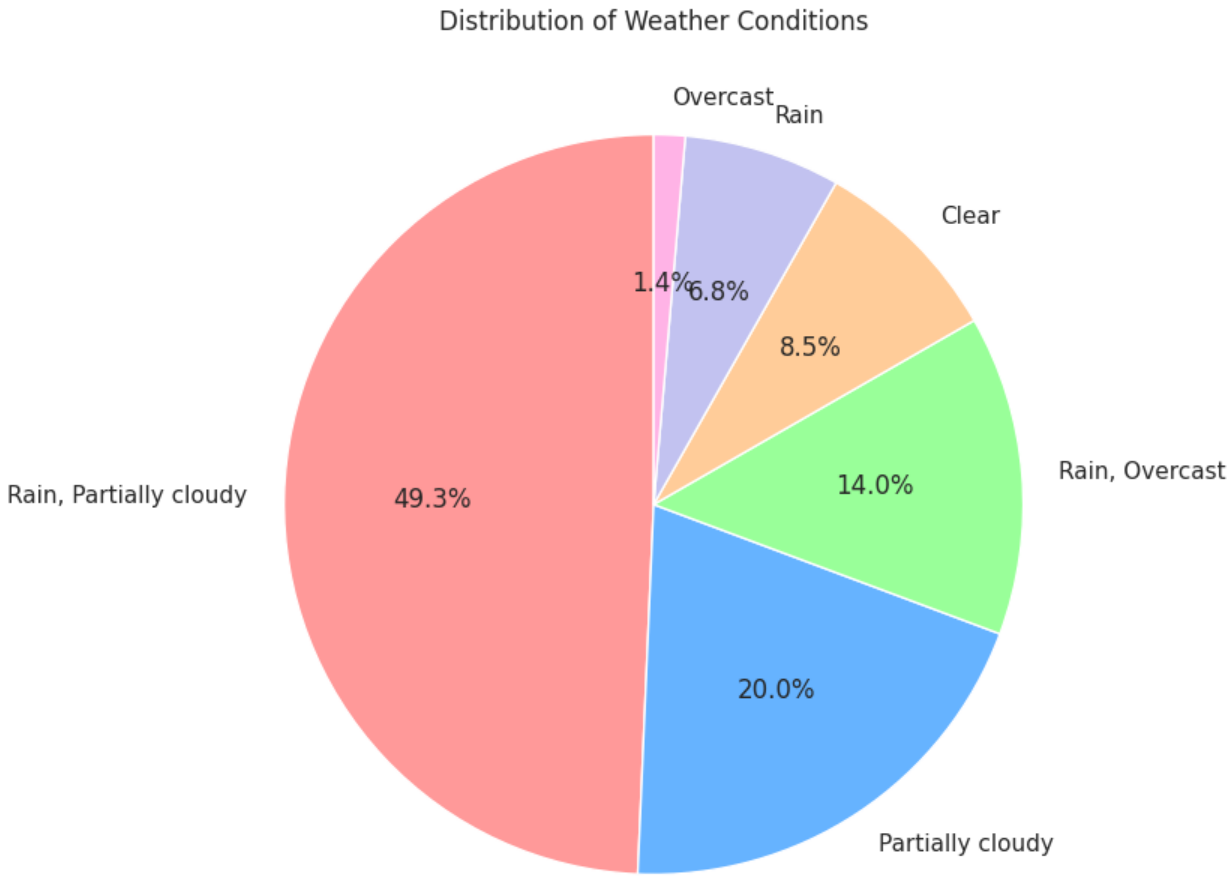
We clean the data like there are some missing data, there are some missing data and there are some duplicate data which can make a difference in our real data representation. After cleaning all the faults in data and errors we can proceed for further data representation for better results.

DATA VISUALIZATION AND EXPLORATIONS:

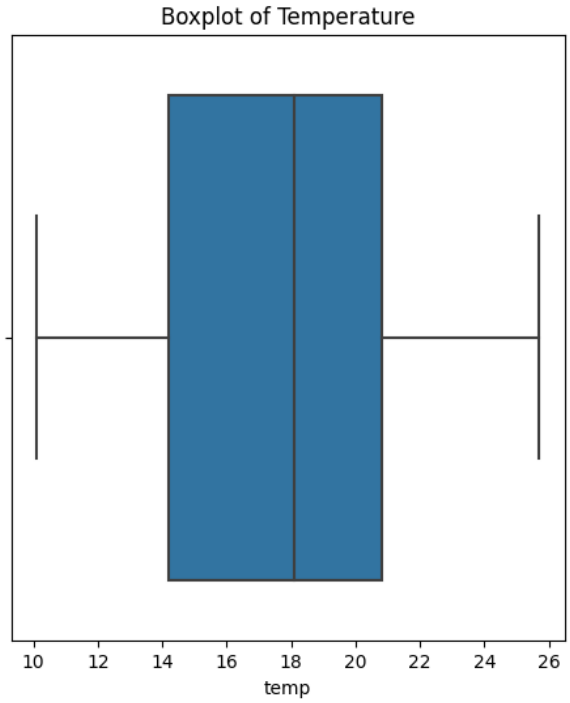
Univariate Analysis :

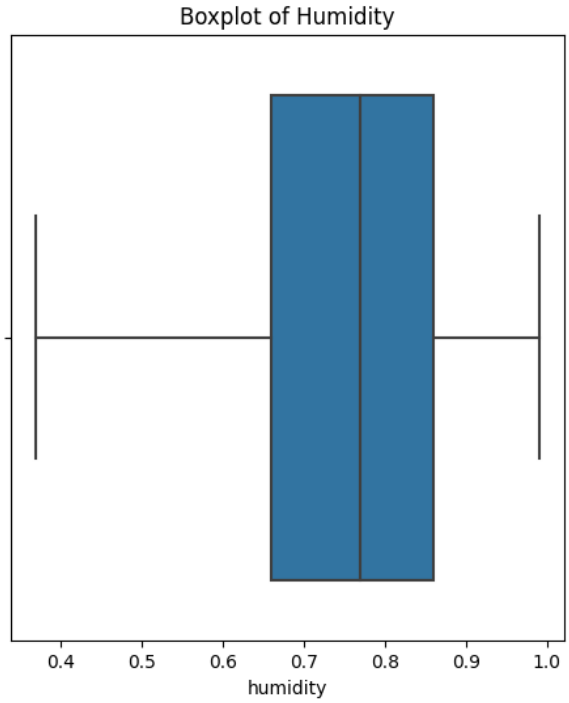


The above figure shows a Line chart of Temperature and Humidity. The graph uses line to represent the changes that occur with respect to the time. It also shows clearly the data which we can not count easily manually. The numeric value shows how it changes with one variable and the other variable with that variable.

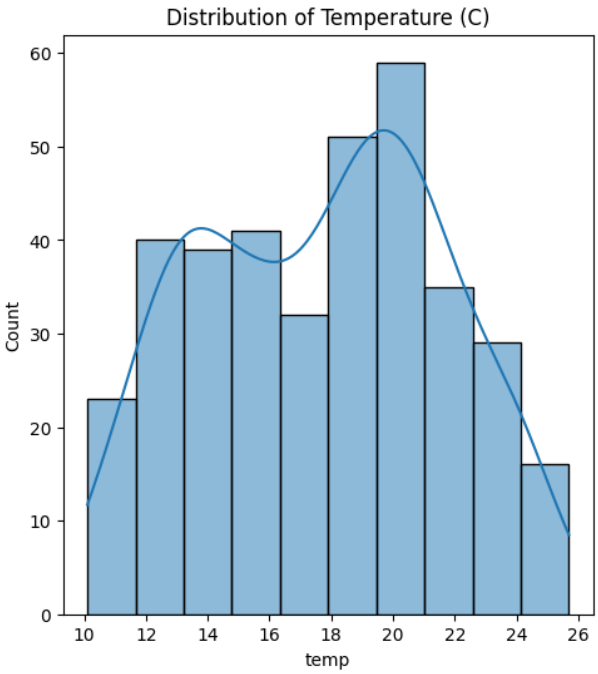


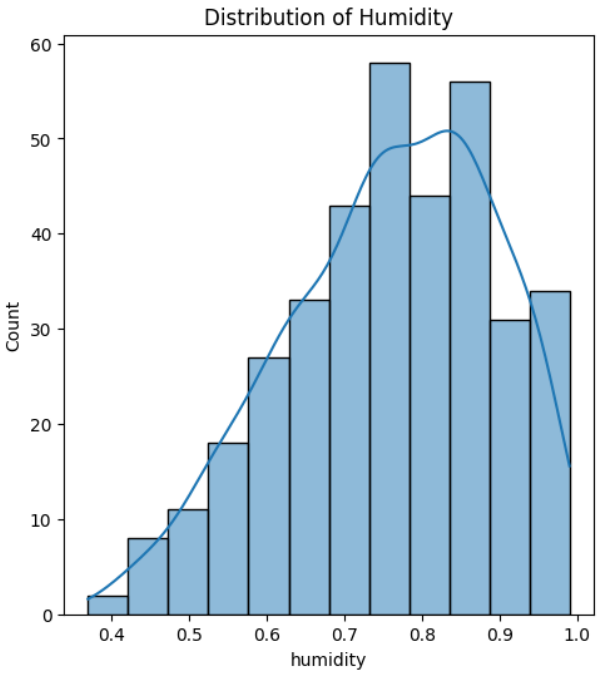
A pie chart is a simple representation of a slice of data with different values to represent the data. There is a percentage label that provides how much percent it is and the title gives us what the data is all about.



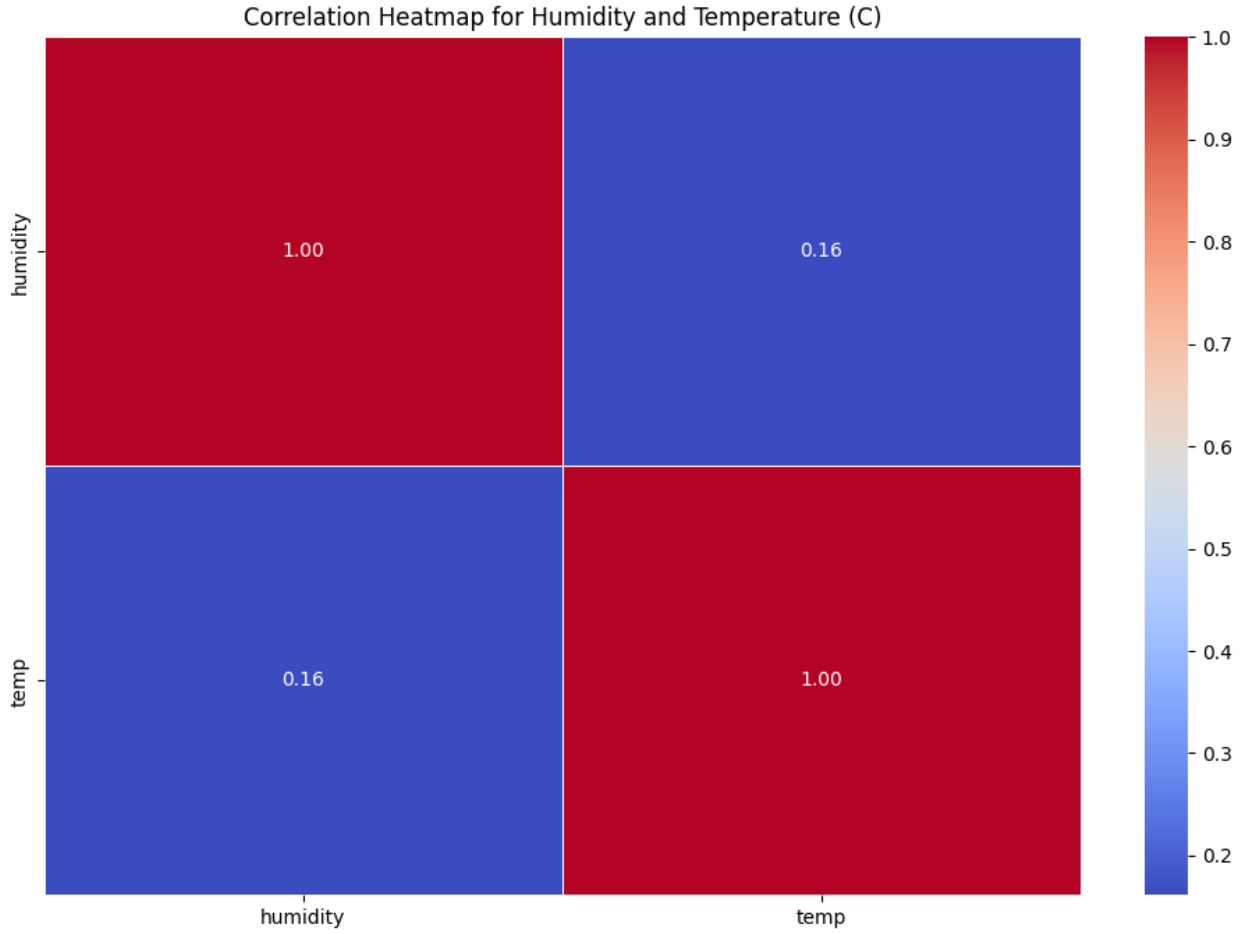


The above figure is a box plot. Which is a graphical representation of a dataset showing the distribution of the statistic values. Statistic value includes minimum, first quartile, median, third quartile, and maximum values. which makes the comparison of data simpler and easier to understand





The above figure is a box plot Histogram. It uses a rectangular bar to represent the frequency or count of the data. It summarizes discrete or continuous data that are measured on an interval of scale. There are no gaps between the bars because it represent continues and falls with specified intervals. The width of the bins is equal for consistency in data representation.



The above figure is a box plot Heat Map. In heat map, different color codes represent different values. Which enables quick distribution of statistical or data-driven information.

Conclusion :

This Assignment helped us to learn more about the management of data with the help of AI . From storing to cleaning of data to performing different kinds of filters and the representation of different categories of data. Now we can play with any kind of data like this.