CSCI-6401-01 **BLACK SHARKS**

WEATHER PREDICTION

Team name: **Black Sharks**

Team members:

1. Sudheer Kumar Reddy Reddymalla sredd1@unh.newhaven.edu
   00719922
2. Durga Guna Sekhar Reddy Savanam dsava2@unh.newhaven.edu
   00755099
3. Vamsi Krishna Regalla vrega1@unh.newhaven.edu
   00730353

Research Question:

To predict the weather based on the data previously collected using data mining techniques.

Dataset:

The dataset we have chosen is "Weather Prediction". This data set is taken from Kaggle website (https://www.kaggle.com/datasets/ananthr1/weather-prediction). This dataset is about the chance of Weather and what factors effect like rain, sun, etc.,

Exploration techniques:

No-one knows how to handle raw data. Need to use data algorithm techniques to handle raw data to develop insights. So, we are using data exploration techniques on our dataset.

Our dataset:

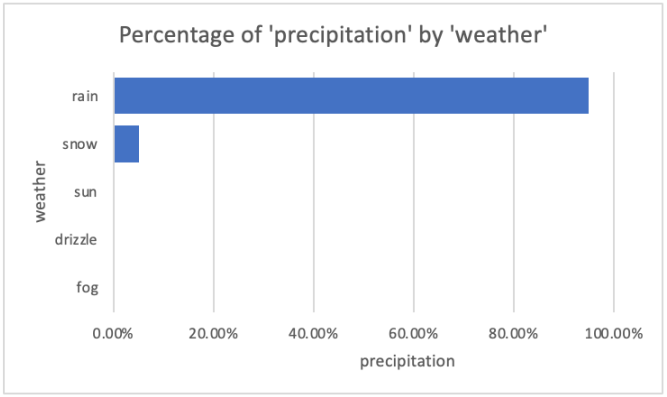
	A	B	C	D	E	F
1	date	precipitation	temp_max	temp_min	wind	weather
2	1/1/12	0	12.8	5	4.7	drizzle
3	1/2/12	10.9	10.6	2.8	4.5	rain
4	1/3/12	0.8	11.7	7.2	2.3	rain
5	1/4/12	20.3	12.2	5.6	4.7	rain
6	1/5/12	1.3	8.9	2.8	6.1	rain
7	1/6/12	2.5	4.4	2.2	2.2	rain
8	1/7/12	0	7.2	2.8	2.3	rain
9	1/8/12	0	10	2.8	2	sun
10	1/9/12	4.3	9.4	5	3.4	rain
11	1/10/12	1	6.1	0.6	3.4	rain
12	1/11/12	0	6.1	-1.1	5.1	sun
13	1/12/12	0	6.1	-1.7	1.9	sun
14	1/13/12	0	5	-2.8	1.3	sun
15	1/14/12	4.1	4.4	0.6	5.3	snow
16	1/15/12	5.3	1.1	-3.3	3.2	snow
17	1/16/12	2.5	1.7	-2.8	5	snow
18	1/17/12	8.1	3.3	0	5.6	snow
19	1/18/12	19.8	0	-2.8	5	snow
20	1/19/12	15.2	-1.1	-2.8	1.6	snow
21	1/20/12	13.5	7.2	-1.1	2.3	snow
22	1/21/12	3	8.3	3.3	8.2	rain

Data Exploration:

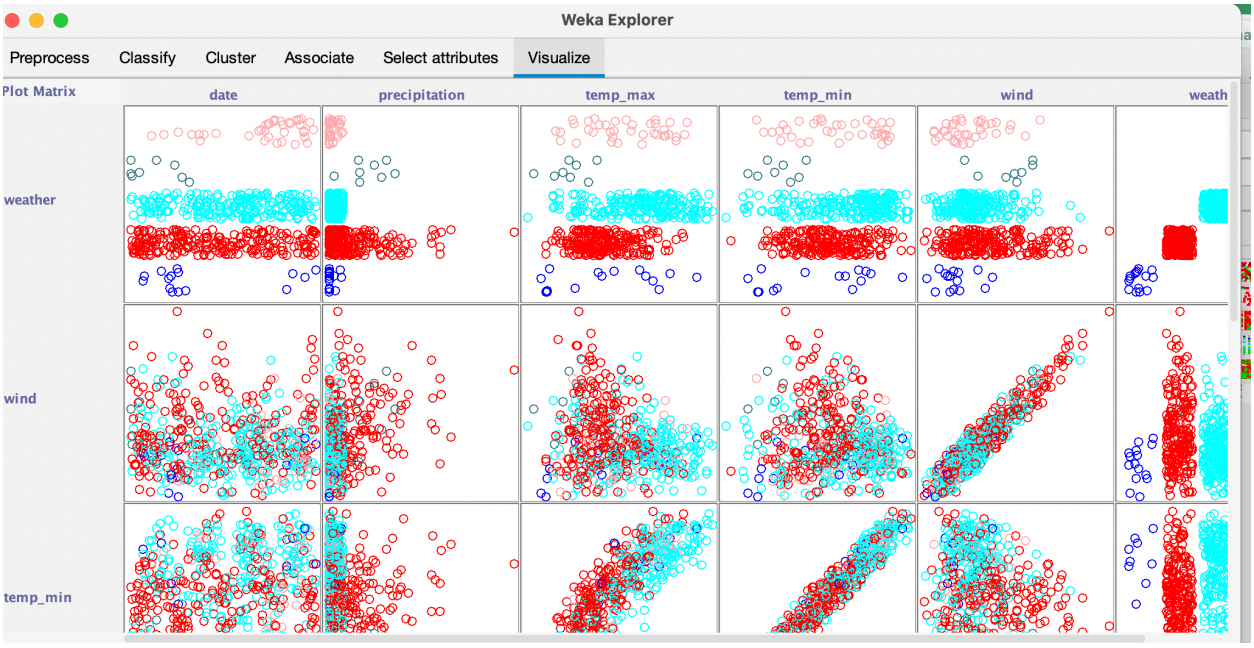
Unique value count: Examining how many distinct values there are in categorical columns is one of the first things that might be helpful during data exploration. This provides insight into the data's subject matter. A unique value count of categorical columns in the automobile's dataset is presented here.

% of total 'precipitation' by 'weather'

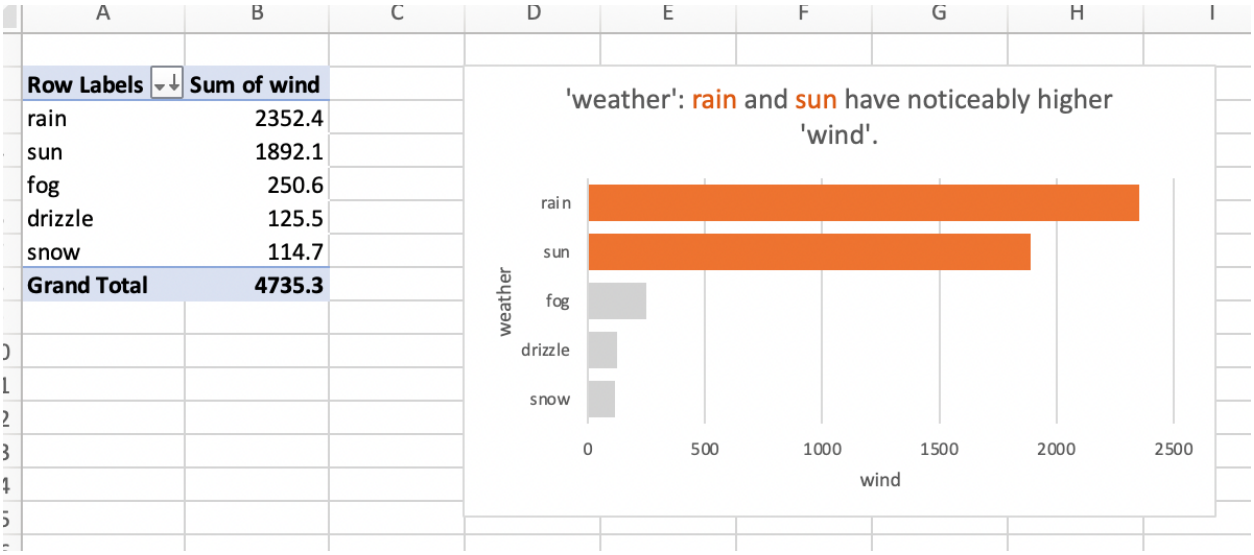
Row Labels	Sum of precipitation
rain	94.98%
snow	5.02%
sun	0.00%
drizzle	0.00%
fog	0.00%
Grand Total	**100.00%**



Scatter plot: It has points that shows the relationship between weather and wind looks as:

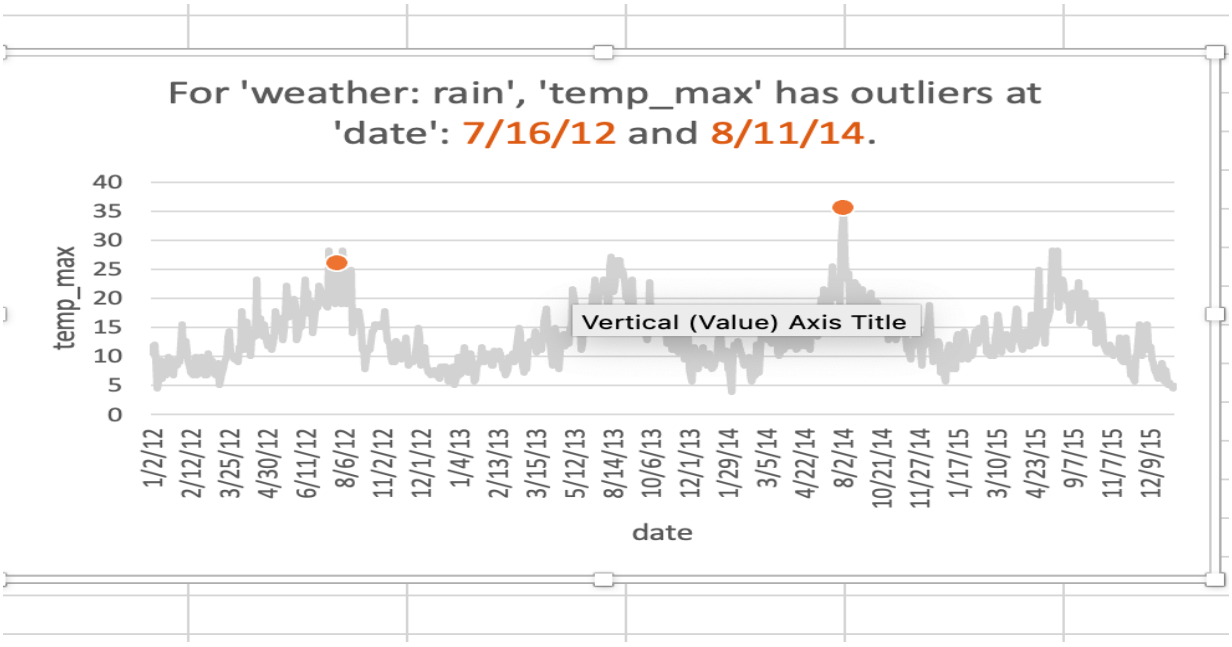


Histogram: The method determines how frequently different values appear in a column. Here, let us compare rain and sun with the wind.



Row Labels	Sum of wind
rain	2352.4
sun	1892.1
fog	250.6
drizzle	125.5
snow	114.7
Grand Total	**4735.3**

Pareto analysis: A unique method for concentrating on what matters is Pareto analysis. The Pareto 80-20 rule is a useful tool for data exploration. We may use Pareto analysis in this dataset in between rain and maximum temperatures.



Conclusion:

From the above data exploration techniques, we can easily extract the insights in no time with the help of them. These techniques help a lot to understand the data much better. Exploratory data analysis has always taken a very graphical approach. Histograms, box plots, scatter plots, and many others are examples of common plots used to discover distributions, correlations, outliers, trends, and other data features

GitHub: https://github.com/sudheerredde/BLACKSHARKS.git