# **Python/DL Project Proposal**

**Team 10**:

1. Sai Rohith Guntupally (8) – (33.3%)
2. Chaitanya Mallepudi (14) – (33.3%)
3. Madhu Varma Rudraraju (20) – (33.3%)

**Project title**:

**Analyzing and Visualizing the US Accidents Dataset** (A country wide traffic accident dataset)

**Goals and Objectives:**

Since road accidents is one of the major concerns in the United States, we would like to analyze and discover what are the causes for accidents such as the impact of precipitation or other environmental factors.

The main objective is to utilize the visualization to predict the car accidents zones, locations and hotspots, factors effecting the accident severity.

**Dataset Description:**

This is a countrywide car accident dataset, which covers **49 states of the United States**. The accident data are collected from **February 2016 to December 2019**, using several data providers, including two APIs that provide streaming traffic incident data.

These APIs broadcast traffic data captured by a variety of entities, such as the US and state departments of transportation, law enforcement agencies, traffic cameras, and traffic sensors within the road-networks. Currently, there are about **3.0 million** accident records in this dataset.

**Increment – 1:**

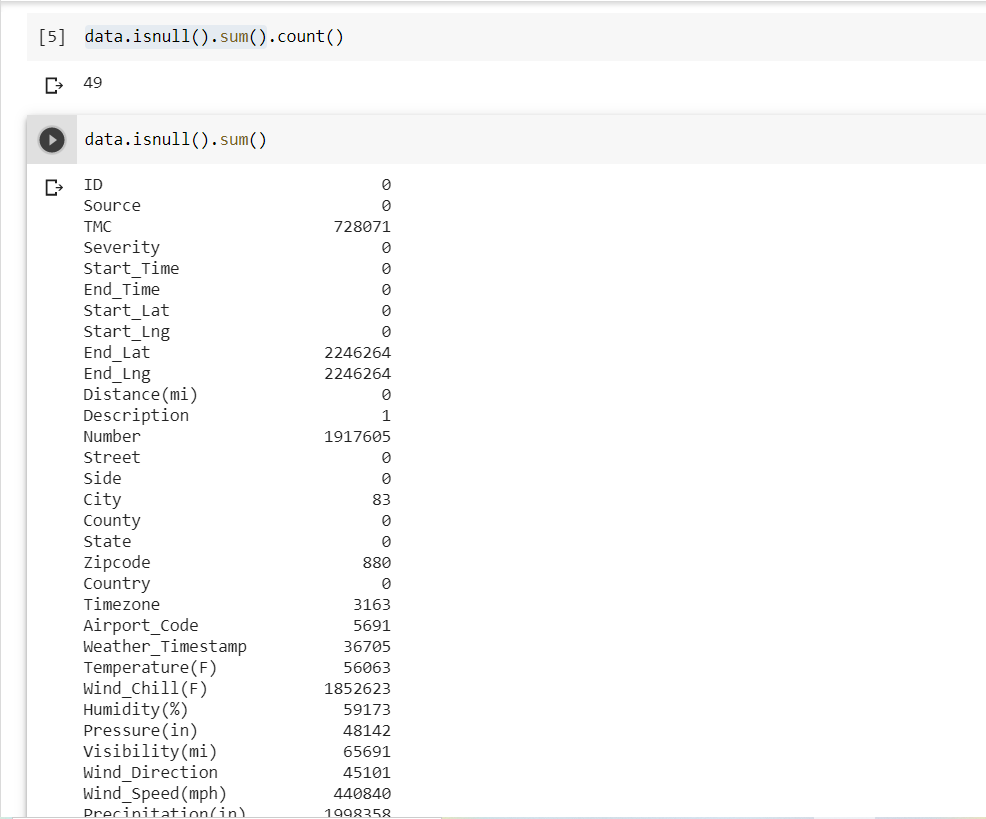
Loaded the dataset and familiarized the records in the dataset.



Performed the EDA by finding out the number of null values in the columns and fill them with the mean values.

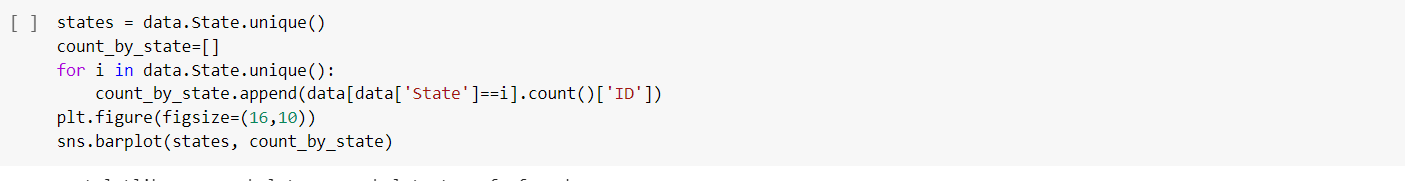


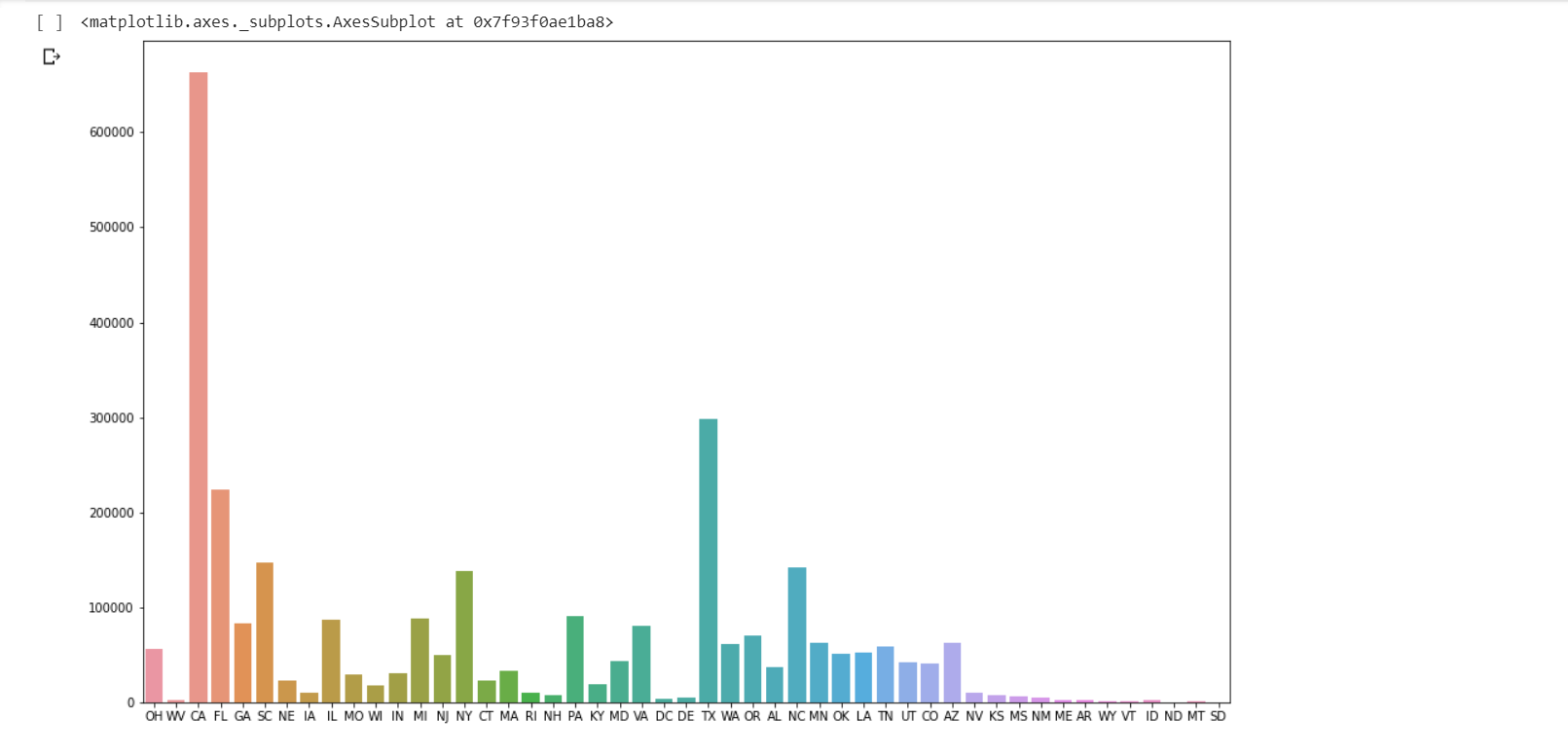
After performing the EDA; we can see there are no null values in the respective columns.



With the available data, we have visualized the following:

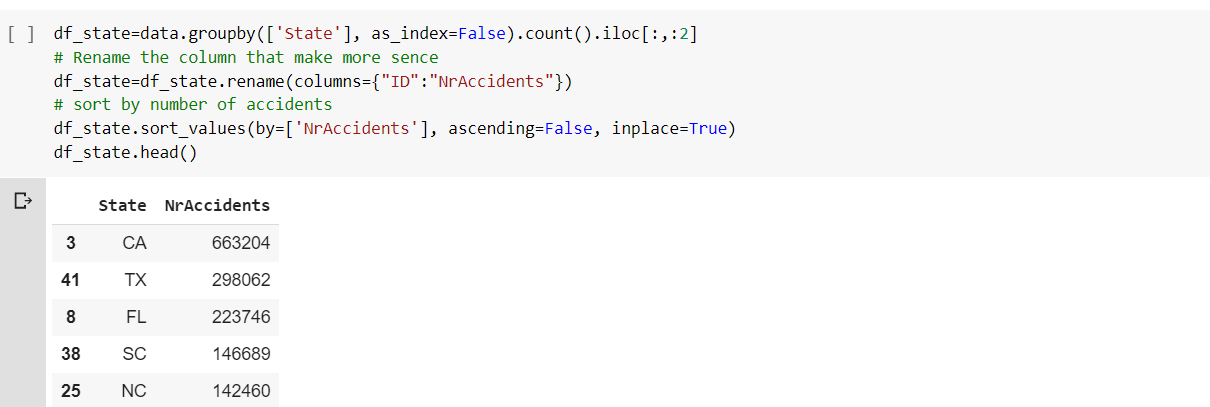
1. State wise accidents count



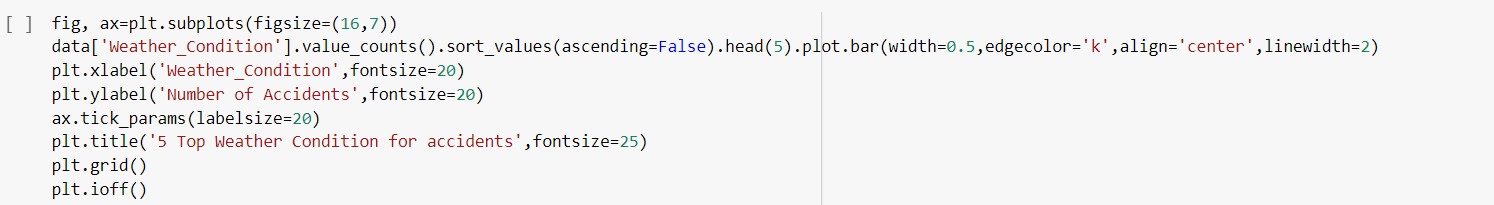


From the above we cannot infer everything clearly.

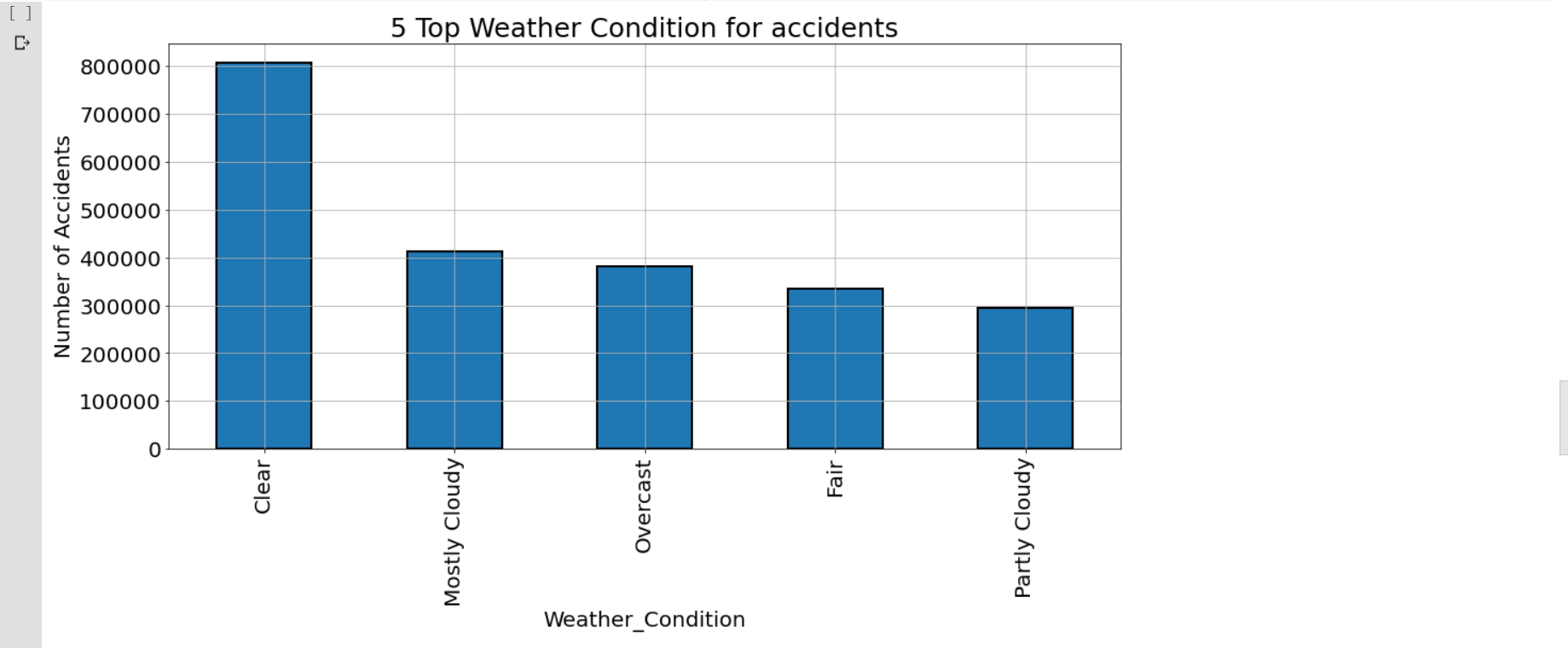
So, we have extracted the top 5 cities with most number of accidents.



The next thing we done is to find the what kind of weather conditions are influencing the accidents.

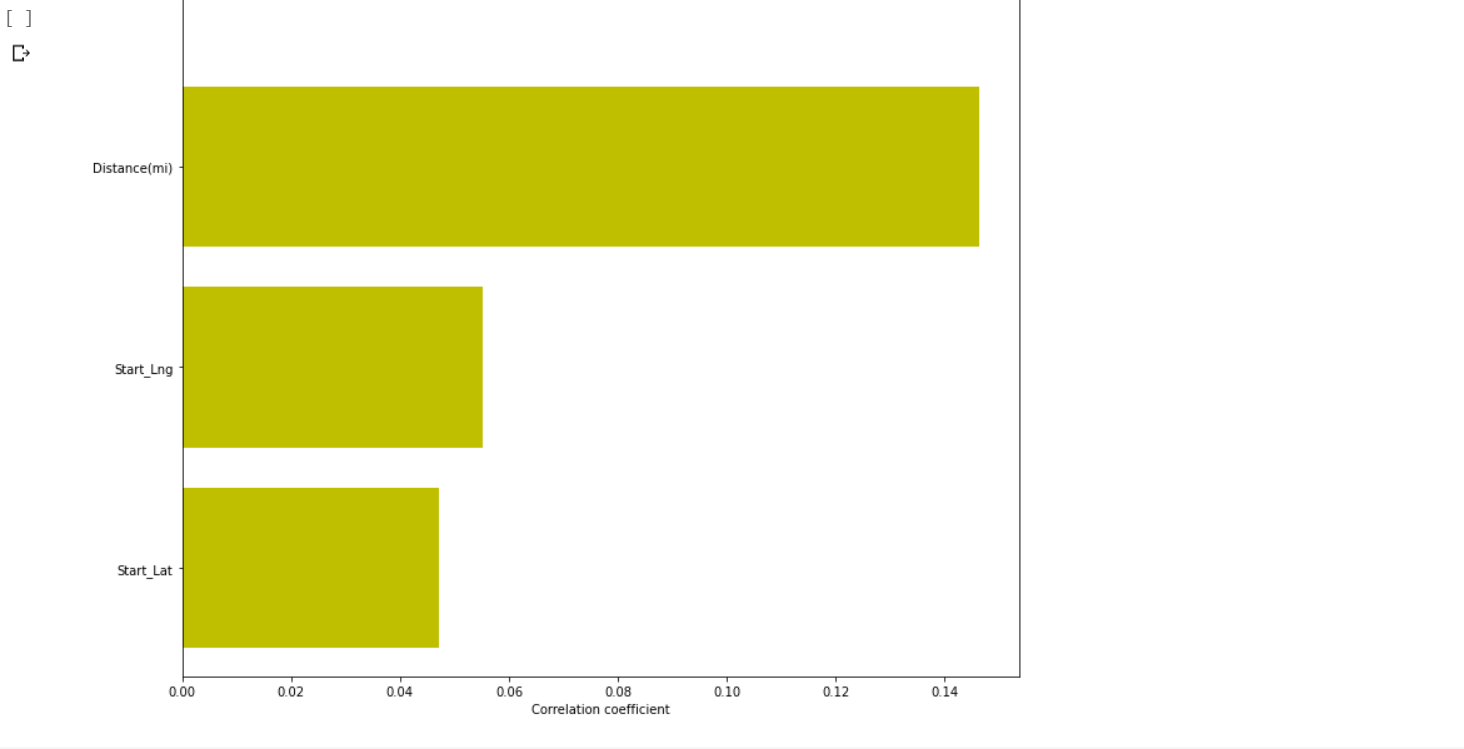


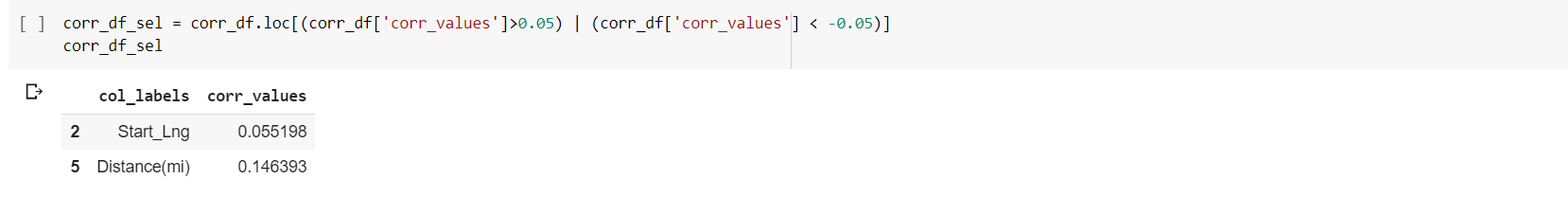
We have observed the following:



Next we have found the correlation between the features and accident cause; if the correlation coefficient is high, it means that it is the most affecting factor for the cause of accident.







for **Increment 2:**

Classify the data and predict the factors and severity affecting accidents and if possible to display the analysis dynamically by developing a website.

We are combinedly working on project to get familiar with each and every aspect of the project. So all of us are involved in everything and discussions.

**Languages and Libraries:**

* Python
* Matplotib
* Ski-kit learn
* Numpy
* Pandas.

**Software and platforms:**

* Jupyter Notebooks
* Google colab.

**References**:

<https://www.kaggle.com/us-accidents>