**Lovely Professional University**



**Academic Task of INT-375**

**(Data Science Toolbox: Python Programming)**

**COMPUTER SCIENCE AND ENGINEERING**

**INT375-Data Science Toolbox: Python Programming**

**PROJECT REPORT**

(Project Semester January-April 2025)

**Title of the Project:**

“Analyzing Road Traffic Accidents in Urban Areas

using

Python Toolkits"

Submitted By: Vansh Chhabra

Registration No. 12306383

Section: K23FK

Course Code: INT-375

**Under the Guidance of**

Dr. Karan Bajaj (32130)

**Discipline of CSE**

**Lovely School of Computer Science and Engineerin**

**Declaration**

I, Vansh Chhabra , student of BTech (Computer Science and Engineering) under CSE Discipline at, Lovely Professional University, Punjab, hereby declare that all the information furnished in this project report is based on my own intensive work and is genuine.

Date: 12-04-2025

Registration No. 12306383 Vansh Chhabra

**Certificate**

This is to certify that Vansh Chhabra bearing Registration no. 12306383 has completed INT-375 “**Data Science Toolbox: Python Programming”** project titled, **“Analyzing Road Traffic Accidents in Urban Areas using Python Toolkits”** under my guidance and supervision. To the best of my knowledge, the present work is the result of his original development, effort and study.

**Dr. Karan Bajaj**

**(UID: 32130)**

**School of Computer Science and Engineering.**

Lovely Professional University

Phagwara, Punjab.

Date: 12-04-2025

**Acknowledgement**

I would like to express my sincere gratitude to my project guide for their valuable guidance, encouragement, and continuous support throughout the development of my project titled “**Analyzing Road Traffic Accidents in Urban Areas using Python Toolkits**”, undertaken as a part of the course **INT-375: Data Science Toolbox: Python Programming.**

I am thankful for the opportunity to work on a real-world problem, which allowed me to apply theoretical knowledge and enhance my practical skills in data analysis and programming. The insightful feedback and suggestions provided at various stages greatly helped in shaping the outcome of this project.

I would also like to acknowledge the resources, tools, and datasets that contributed to my research and analysis. Lastly, I extend my appreciation to my peers and faculty members who provided their support and encouragement during the project.

Vansh Chhabra

Reg No. 12306383

**TABLE OF CONTENT**

|  |  |
| --- | --- |
| **Sr No.** | **Description** |
| **1** | **INTRODUCTION** |
| **2** | **Source of Dataset** |
| **3** | **EDA**  **Exploratory Data Analysis** |
| **4** | **Objectives** |
| **5** | **CONCLUSION** |
| **6** | **FUTURE SCOPE** |
| **7** | **References** |

**INTRODUCTION**

* Road traffic accidents represent a significant public safety issue, contributing to injury, death, and economic loss worldwide. Understanding the underlying causes and patterns of such accidents is crucial for implementing effective road safety measures. This study utilizes a comprehensive dataset of road traffic accidents, focusing on key variables such as time of occurrence, weather conditions, lighting, road surface, vehicle type, location, and severity.
* The dataset provides valuable insights into how various environmental and situational factors contribute to accidents in urban areas. By analyzing patterns in this data, we can identify high-risk zones, understand the correlation between road conditions and accident severity, and recommend improvements in traffic management and infrastructure planning. This analysis supports data-driven decision-making by traffic authorities, urban planners, and policymakers.
* With a structured and detailed approach to this data, the study aims to enhance public safety by contributing to the development of informed strategies for accident prevention and efficient emergency response systems.

**Source of Dataset**

**Link:**

<https://github.com/Tanvir-Taushif/free-datasets-for-learning/tree/main/Road%20Accident%20Dataset>

**EDA**

**Exploratory Data Analysis**

**Key Points:**

**Overview:**  
The dataset consists of detailed records of road traffic accidents, including date, time, severity, weather, road surface, number of vehicles involved, and geographic coordinates.

**MissingValue Check:**  
Missing data is identified and visualized using heatmaps. Columns such as Weather Conditions, Road Surface Conditions, or Carriageway Hazards may contain missing entries.

**Accident Severity Distribution:**  
The most common accident severity is visualized using a pie chart. Most accidents are found to be *Slight*, followed by *Serious* and *Fatal*.

**Summary Statistics:**  
Descriptive statistics such as mean, median, and standard deviation for numerical columns like Speed limit, Number of Vehicles, and Number of Casualties help understand data spread and central tendency.

**Road Surface Condition Impact:**  
Wet or damp road conditions are often associated with a higher number of serious accidents, implying a need for improved traction measures.

**Weather Condition Influence:**  
Poor weather conditions like fog, rain, and snow correlate with higher severity accidents, indicating visibility and road grip issues.

**Accidents by Speed Limit Zones:**  
Higher severity is seen in high-speed zones (e.g., 60–70 mph), reinforcing the importance of speed regulations.

**Vehicle Type Impact:**  
Accidents involving motorcycles or heavy goods vehicles tend to have higher severity levels compared to standard cars.

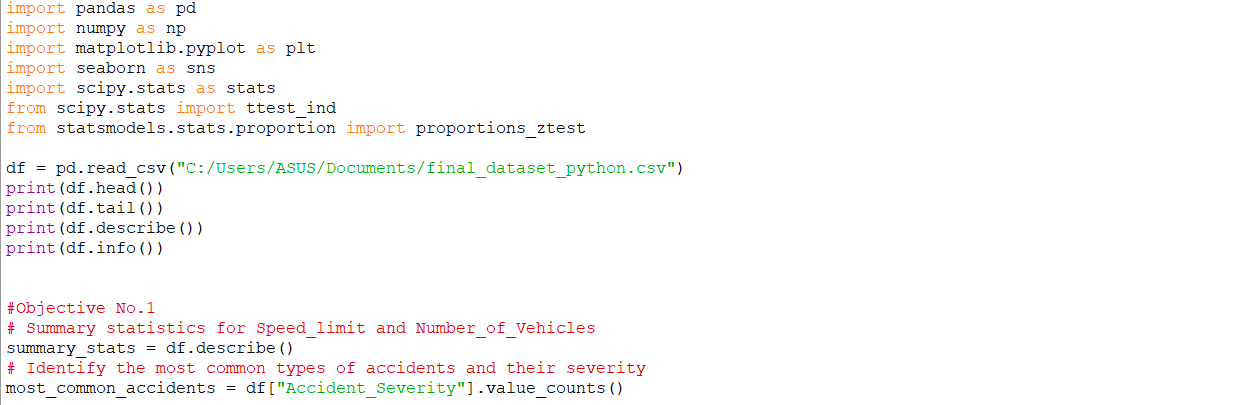
* **Driver Impairment Factors:**  
  Although not always captured, inference based on timing (e.g., late night or early morning) may point to fatigue or impairment-related causes.
* **Correlation Heatmap:**  
  A correlation matrix visually highlights relationships among numerical variables like Speed\_limit, Number\_of\_Vehicles, and Number\_of\_Casualties.
* **Average Speed vs Vehicle Count:**  
  A bar chart comparing the average speed limit and number of vehicles involved gives insight into whether speed might influence vehicle involvement.
* **Time of Day Analysis:**  
  Accidents peak during morning and evening rush hours, typically between 7–9 AM and 4–7 PM.

**Objectives**

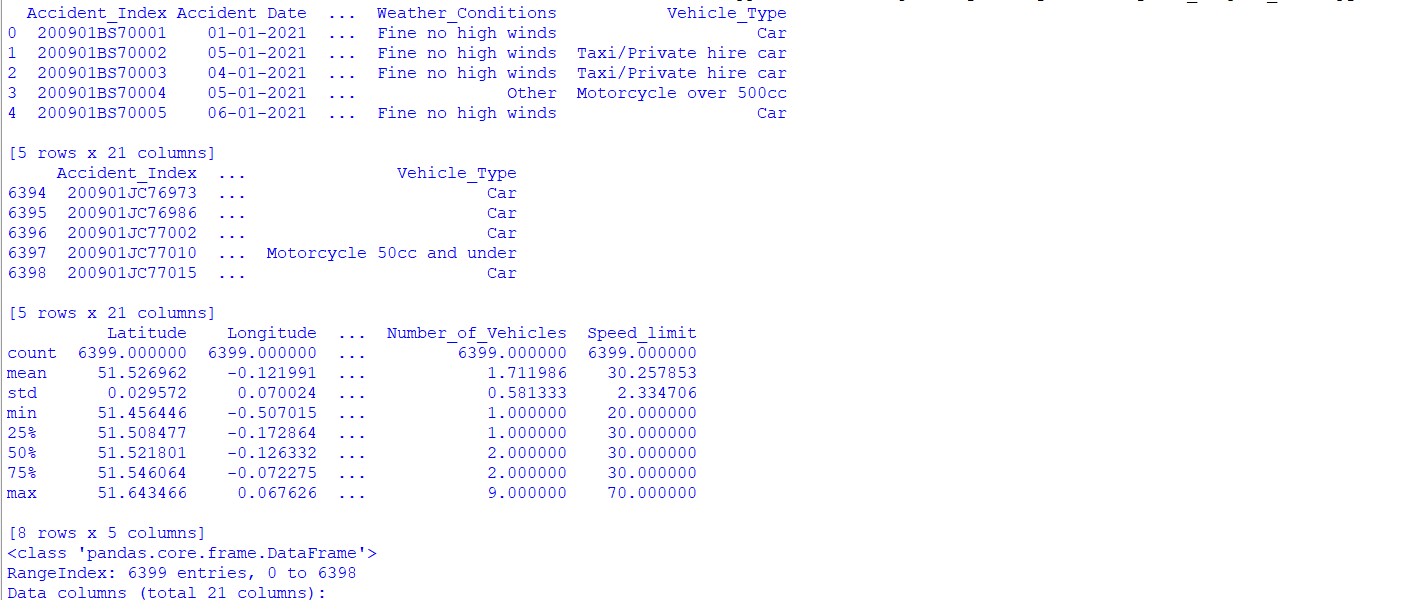
Objective 1:

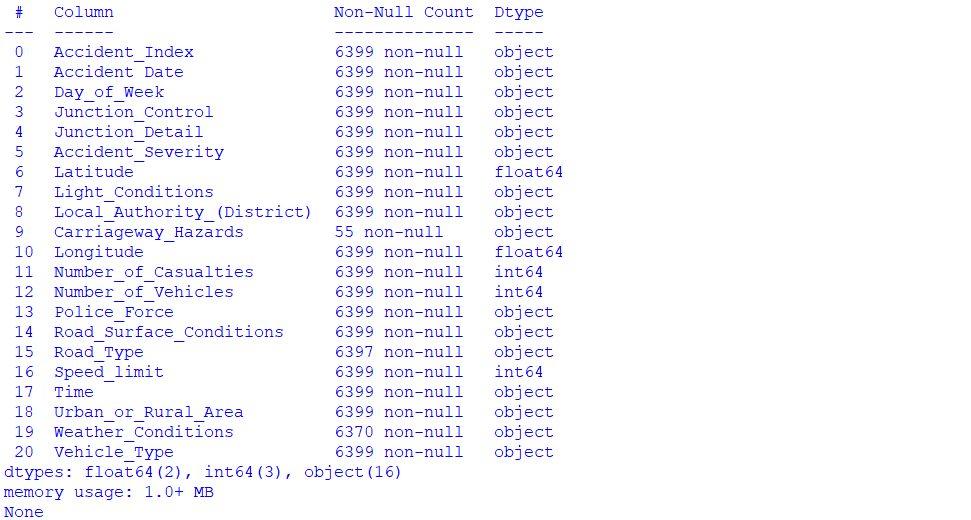
Data Exploration and Summary Statistics: Generate summary statistics for numeric columns (e.g., Speed\_limit, Number\_of\_Vehicles). Also, Identify the most common types of accidents and their severity.

Coding Part:



Implementation:

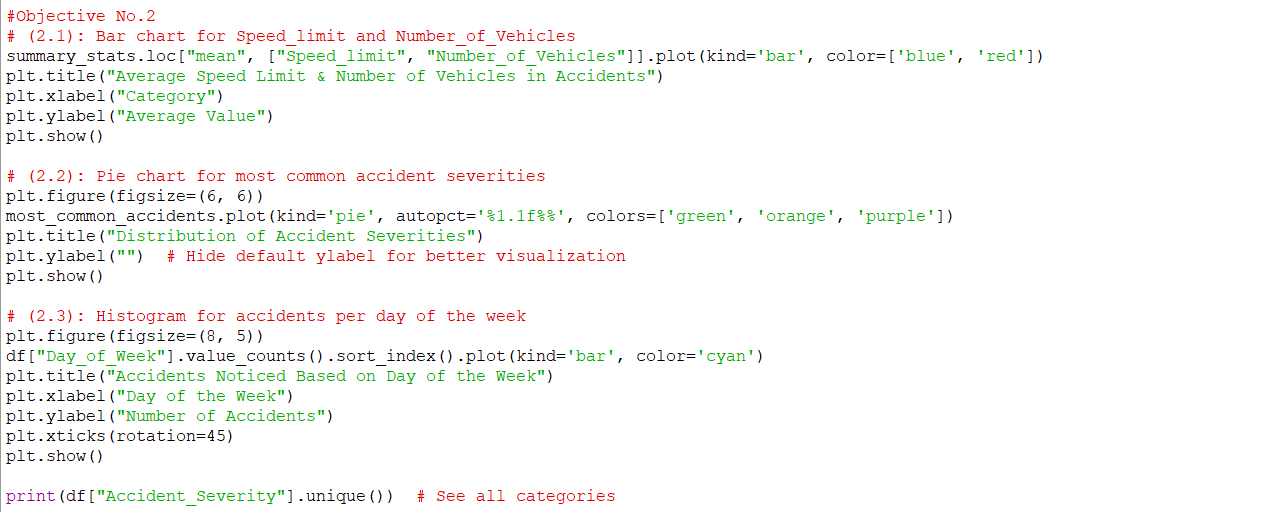




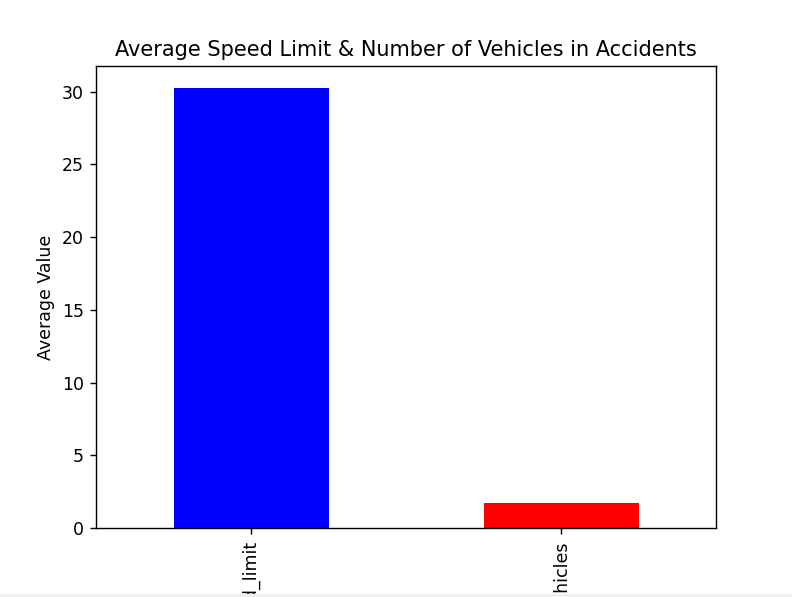
Objective 2:

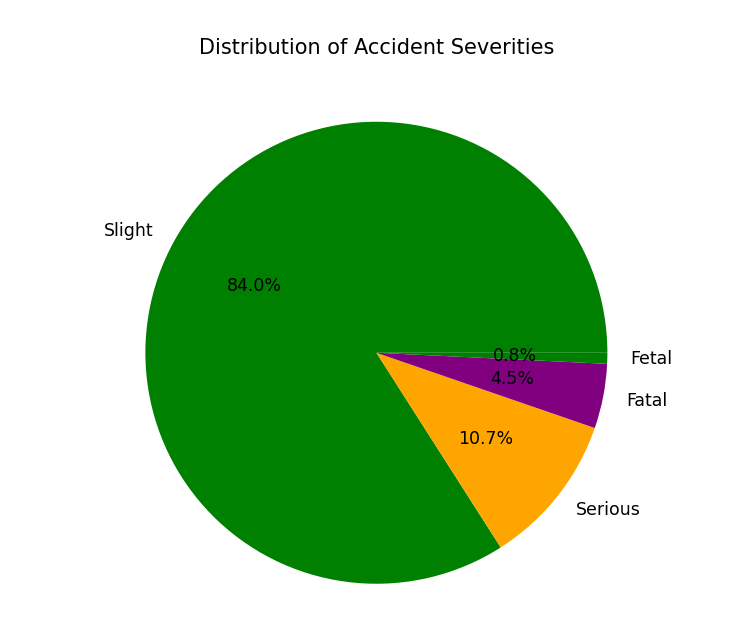
Visualization of Accident Trends: Create graphs (e.g., bar charts, line plots, histogram) to analyze the frequency of accidents by Day\_of\_Week or Accident\_Severity.

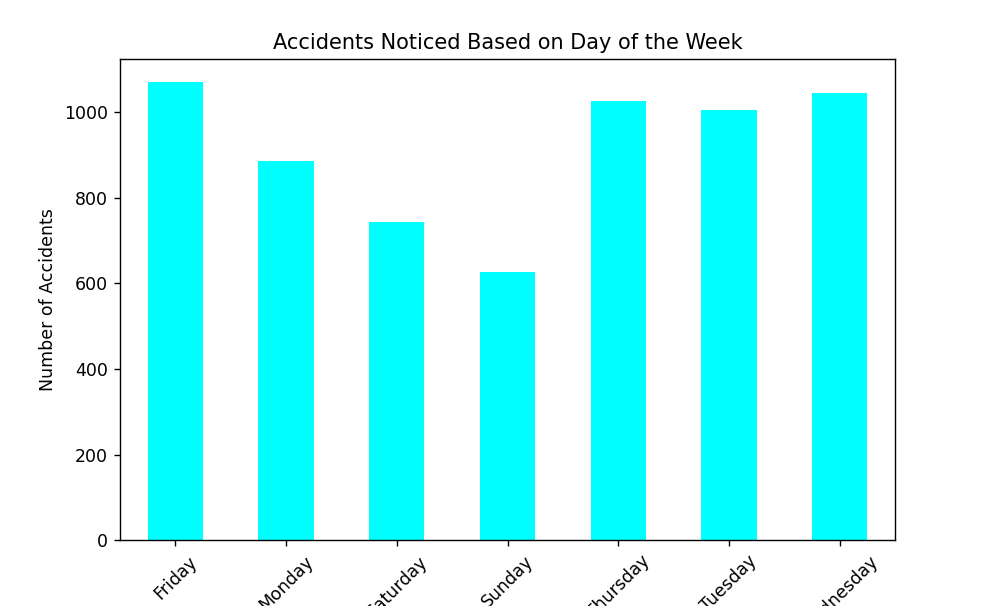
Coding Part:



Implementation:



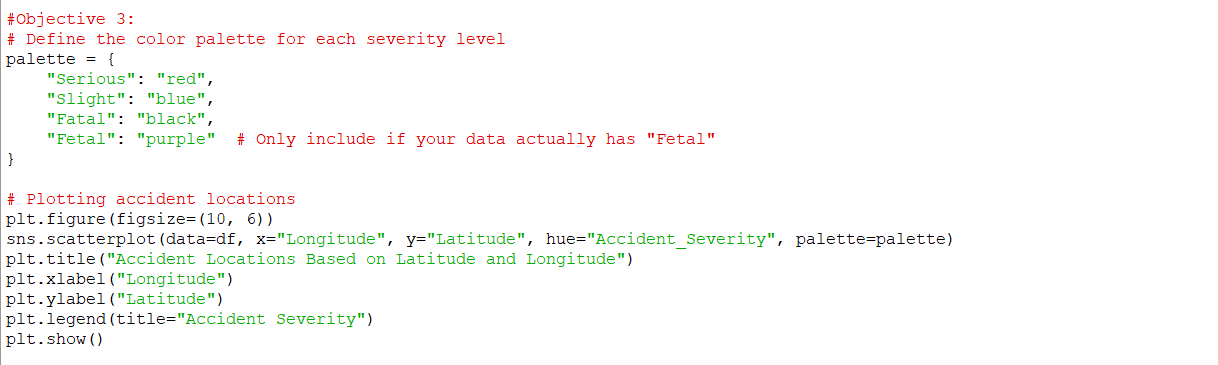




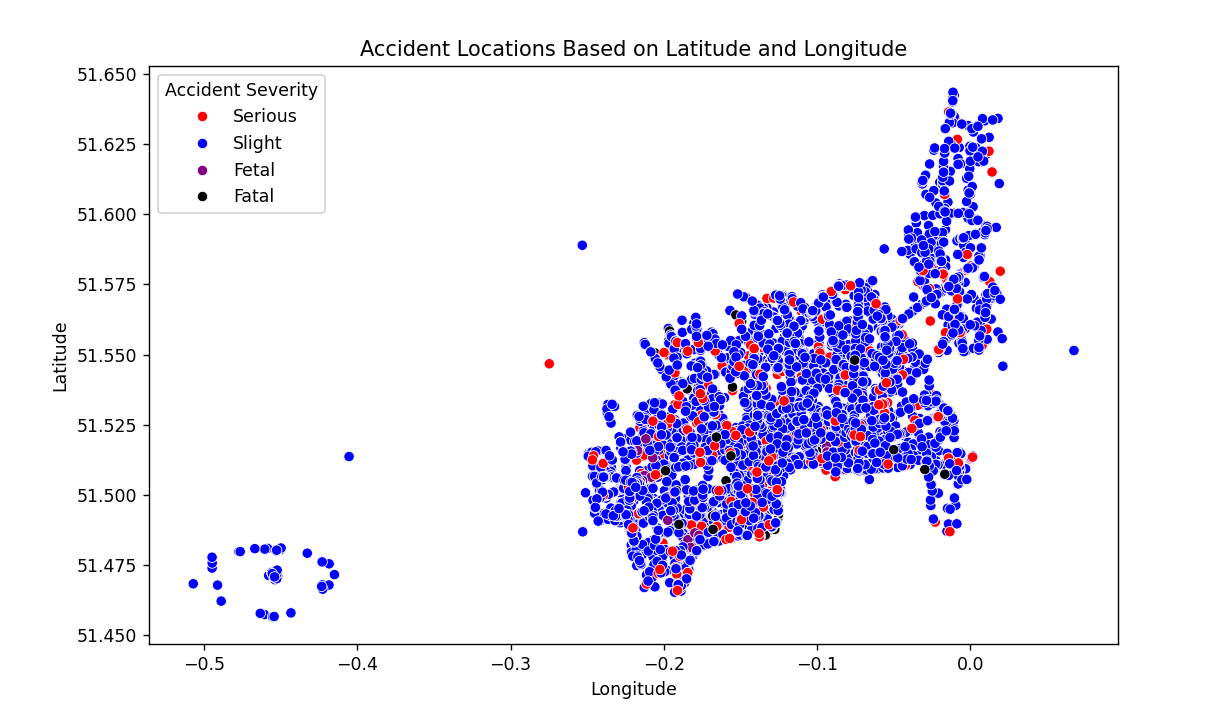
Objective 3:

Mapping Accident Locations: Plot accident locations using latitude and longitude on a map (using matplotlib).

Coding Part:



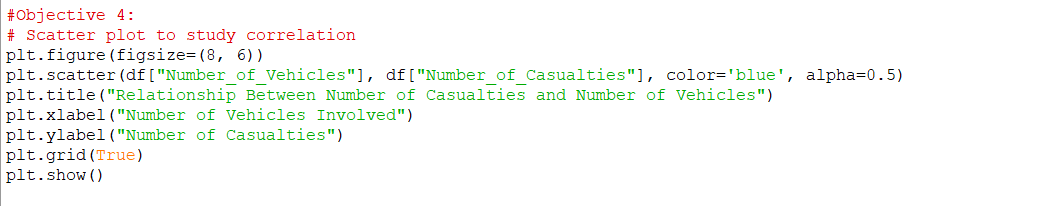
Implementation:



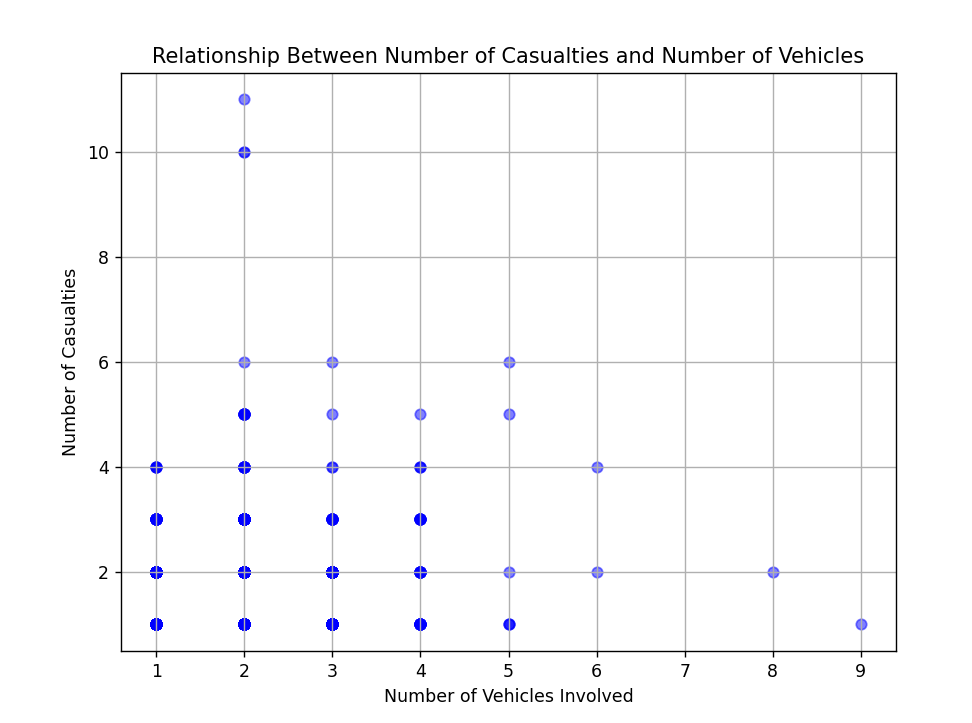
Objective 4:

Casualties and Vehicles Relationship: Study the correlation between Number\_of\_Casualties and Number\_of\_Vehicles.

Coding Part:



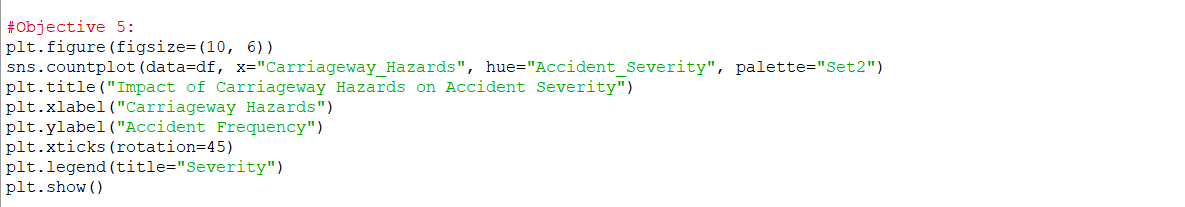
Implementation:



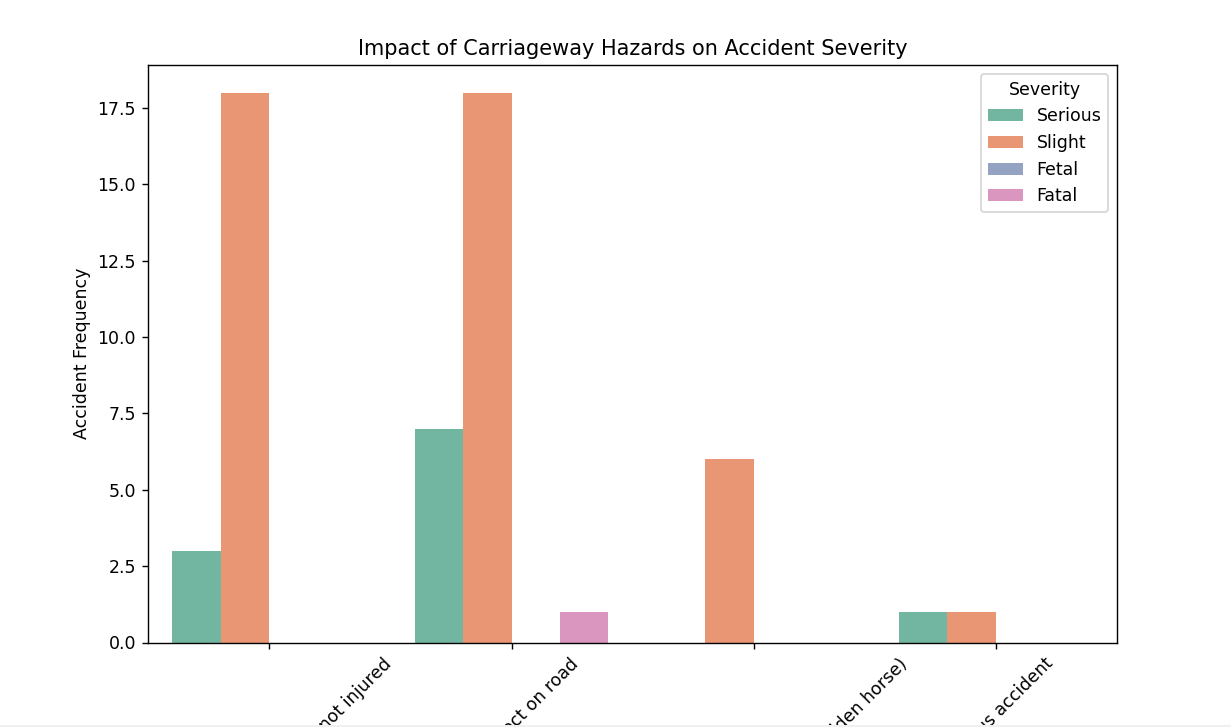
Objective 5:

Analyze if Carriageway\_Hazards impact accident severity or frequency: Road Surface Condition Impact and Explore how Road\_Surface\_Conditions influence accident severity (e.g., Wet or Damp vs Dry).

Coding Part:



Implementation:

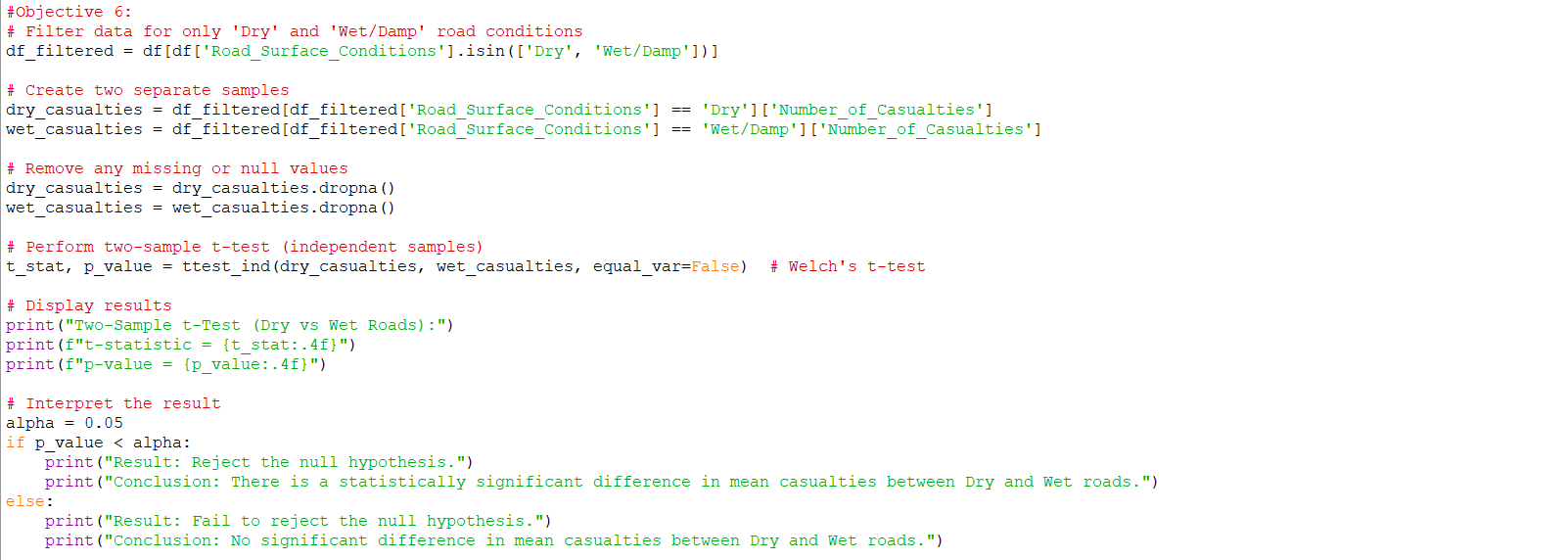


Objective 6:

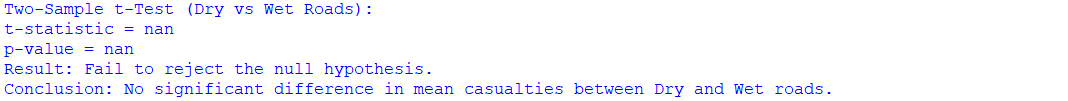
(t-test) Use a two-sample t-test to compare mean Number\_of\_Casualties for accidents occurring on Dry vs Wet roads.

* Hypothesis:
  + Null: Mean casualties on dry and wet roads are the same.
  + Alternate: Mean casualties are different.

Coding Part:



Implementation:

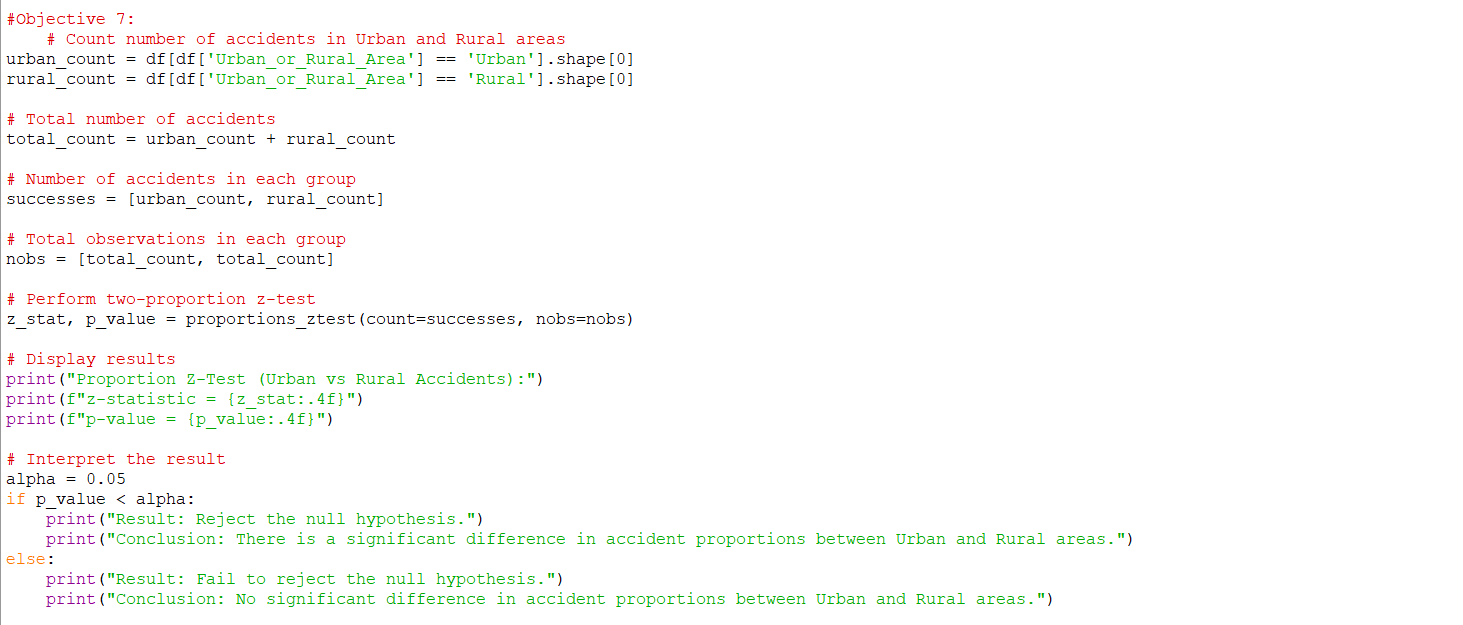


Objective 7:

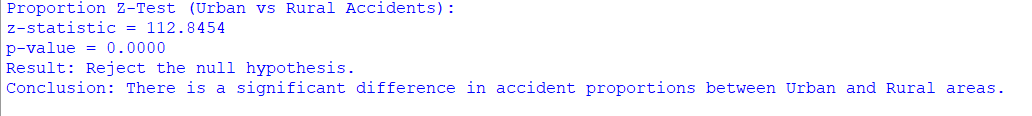
Conduct a proportion z-test to compare the accident frequency in urban vs rural areas.

* Hypothesis:
  + Null: Accident proportions are equal in both areas.
  + Alternate: Accident proportions are not equal.

Coding Part:



Implementation:



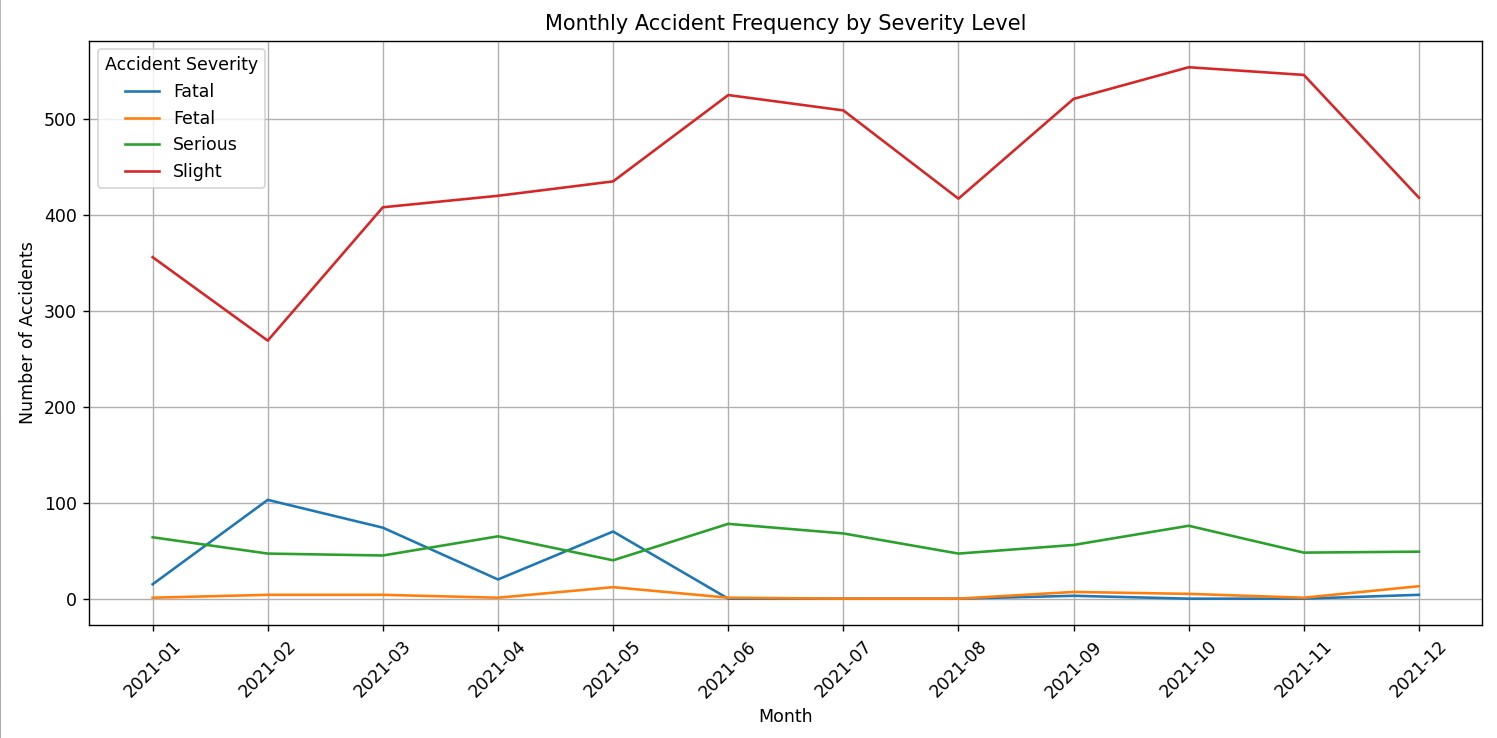
Objective 8:Convert Accident Date to datetime and analyze the accident frequency over time (weekly or monthly).

* Plot line graphs for different severity levels.

Coding Part:



Implementation:



**CONCLUSION**

In this project, I successfully leveraged Python and its data science toolkits to analyze road traffic accidents in urban areas, uncovering key insights from patterns in accident data.

Through thorough exploratory data analysis, visualizations, and statistical tests, we identified significant relationships between environmental conditions, road surface types, vehicle involvement, and accident severity. The findings emphasize the influence of factors such as weather conditions, speed limits, and time of day on accident frequency and outcomes.

By applying statistical techniques like t-tests and proportion z-tests, we were able to validate assumptions and draw meaningful comparisons between different accident scenarios. This analysis not only enhances our understanding of urban traffic dynamics but also supports the development of data-driven policies aimed at reducing road accidents and improving public safety.

The project demonstrates the power of Python in transforming raw data into actionable knowledge, paving the way for smarter urban planning and traffic management solutions.

**FUTURE SCOPE**

The analysis conducted in this project lays the foundation for more advanced research and real-world applications in the domain of road safety and urban planning. Future work can be extended in the following directions:

1.**Machine Learning for Accident Prediction:** Integrate machine learning models such as logistic regression, decision trees, or random forests to predict accident severity or likelihood based on input variables like weather, time, and road conditions.

2.**Real-time Accident Monitoring:** Incorporate real-time traffic and weather data to build a dynamic accident alert system that warns drivers and authorities of high-risk conditions.

3.**Geospatial Analysis with Advanced GIS Tools:** Use more sophisticated geospatial libraries (e.g., GeoPandas, QGIS) for deeper spatial clustering of accident-prone zones and hotspot analysis.

4.**Integration with IoT and Smart City Data:** Combine data from traffic cameras, vehicle sensors, and smart city infrastructure to enhance situational awareness and enable proactive traffic control.

**5.Policy Simulation:** Use the analyzed data to simulate the impact of proposed policy changes, such as speed limit adjustments or improved road lighting, and assess their effectiveness in accident reduction.

**REFERENCES:**

1.Dataset Link:

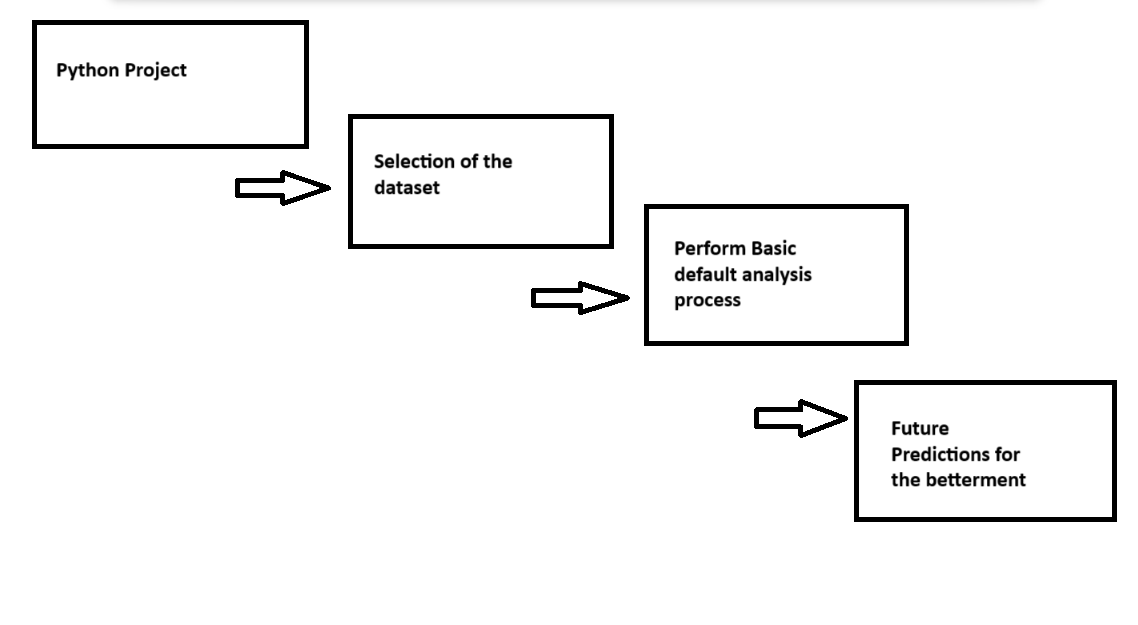
<https://github.com/Tanvir-Taushif/free-datasets-for-learning/tree/main/Road%20Accident%20Dataset>

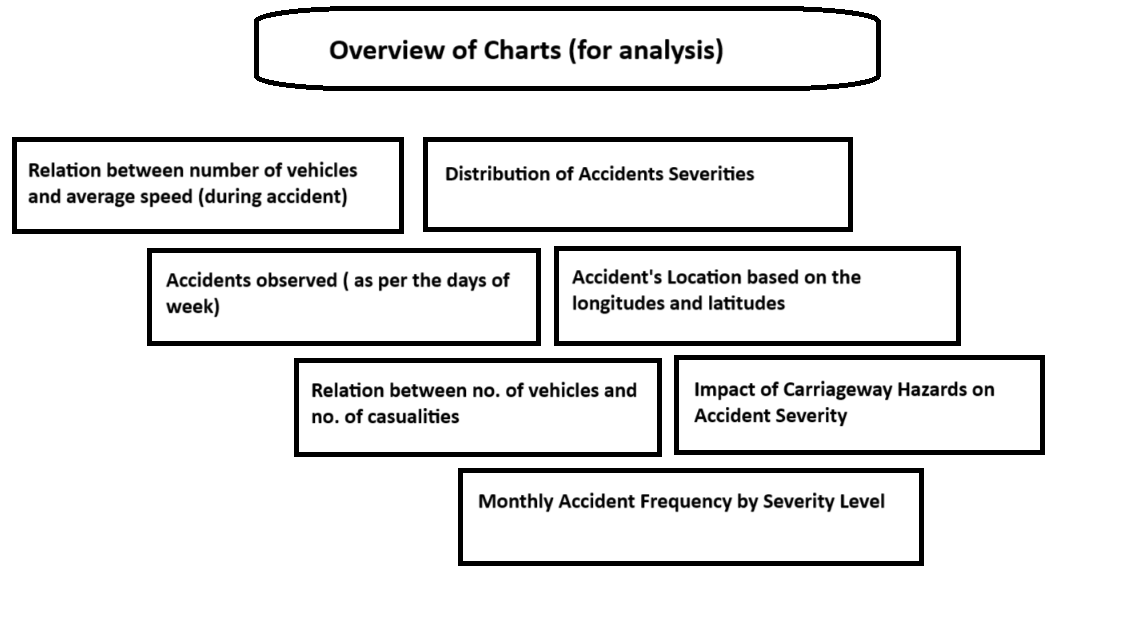
2.Python Libraries:

* **NumPy Developers. (2023).** *NumPy: Fundamental package for scientific computing with Python*. Retrieved from: <https://numpy.org/>
* **pandas development team. (2023).** *Pandas: Python Data Analysis Library*. Retrieved from: <https://pandas.pydata.org/>
* **Matplotlib Developers. (2023).** *Matplotlib: Visualization with Python*. Retrieved from: <https://matplotlib.org/>
* **Seaborn Developers. (2023).** *Seaborn: Statistical data visualization*. Retrieved from: <https://seaborn.pydata.org/>
* **SciPy Community. (2023).** *SciPy: Scientific Computing Tools for Python*. Retrieved from: <https://scipy.org/>

**Glimpse ot the project**

**Design and Architecture:**

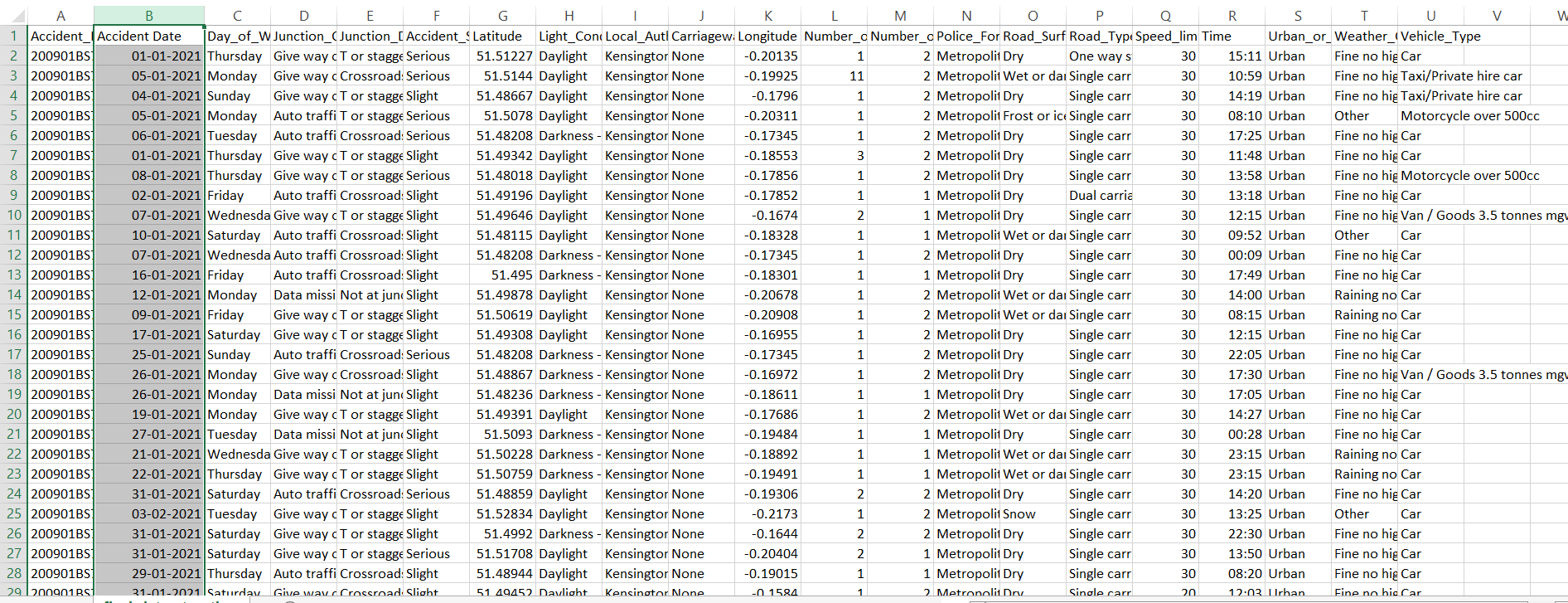
****

****

**Dataset Description:**

Here, I am attaching a screenshot of my dataset (as with the help of this screenshot you’ll get to know how actually my dataset actually looks like)

And with the help of this you’ll also get a glimpse of my dataset.

****

**Objective’s Description:**

1. Data Exploration and Summary Statistics: Generate summary statistics for numeric columns (e.g., Speed\_limit, Number\_of\_Vehicles). Also, Identify the most common types of accidents and their severity.
2. Visualization of Accident Trends: Create graphs (e.g., bar charts, line plots, histogram) to analyse the frequency of accidents by Day\_of\_Week or Accident\_Severity.
3. Mapping Accident Locations: Plot accident locations using latitude and longitude on a map (using matplotlib).
4. Casualties and Vehicles Relationship: Study the correlation between Number\_of\_Casualties and Number\_of\_Vehicles.
5. Analyse if Carriageway\_Hazards impact accident severity or frequency: Road Surface Condition Impact and Explore how Road\_Surface\_Conditions influence accident severity (e.g., Wet or Damp vs Dry).
6. (t-test) Use a two-sample t-test to compare mean Number\_of\_Casualties for accidents occurring on Dry vs Wet roads.
7. Conduct a proportion z-test to compare the accident frequency in urban vs rural areas.
8. Convert Accident Date to datetime and analyze the accident frequency over time (weekly or monthly).

**LinkedIn Likes and Comments:**

**Design and Architecture (Part2 )**