**Project Overview: Traffic and Accident Data Analysis**

**Introduction**

This document outlines the process of analyzing traffic volumes and road accidents in Israel from 2015 to 2021. The primary objective was to merge traffic count data with accident data across these years to identify potential patterns and insights. The final output is a set of data files and visualizations that can be used for further research.

**Step 1: Understanding and Analyzing the Data Structure**

**Initial Data Exploration**

* **Traffic Count Data**: We began by loading traffic count data, which included columns such as road number (kvish), segment (keta), day of the week (yom), and traffic volume (kamut\_kle\_rehev).
* **Accident Data**: The accident data contained columns like road number (KVISH1 and KVISH2 for the two possible roads involved in an accident), accident severity (HUMRAT\_TEUNA), and geographic coordinates (X, Y).

**Challenges**

* The datasets for different years had varied structures, with inconsistent column names and formats.
* A critical challenge was unifying this scattered data from 2015 to 2021 into a standardized format that could be merged and analyzed collectively.

**Step 2: Standardizing the Data**

**Accident Data**

* **Renaming Columns**: We standardized the column names for consistency across datasets. For instance, yom\_bashavua was renamed to yom to align with the traffic data.
* **Uniform File Creation**: We created a unified dataset for accidents across all years by aligning the columns and data types, ensuring all entries were comparable.

**Traffic Data**

* **Aggregating Data**: The traffic data was aggregated to calculate the average vehicle count (kamut\_kle\_rehev) for each road segment (kvish) on every day of the week (yom). This aggregation ignored the month to generalize the data across all years.
* **Uniform File Creation**: A standardized file for traffic counts was created by consolidating data across all years.

**Step 3: Merging Accident and Traffic Data**

**Objective**

To merge the standardized accident data with the aggregated traffic data so that each accident record had the corresponding average traffic count for the road and day of the week.

**Method**

* **Matching Logic**: We attempted to match each accident record’s kvish1 or kvish2 with the corresponding kvish in the traffic data. The match was based on the road (kvish) and the day of the week (yom).
* **Result**: The result was a new dataset where each accident record included an additional column for the average traffic count on that road and day across all years.

**Step 4: Data Cleaning**

**Objective**

To ensure data quality, we cleaned the merged dataset by removing any accident records that did not have a corresponding traffic count.

**Outcome**

This step resulted in a clean, final dataset where every accident had an associated average traffic count, ensuring that our analysis would be accurate and reliable.

**Step 5: Validation using QGIS**

**Objective**

To validate our data processing and merging, we imported the final dataset into QGIS.

**Validation**

By visualizing the data on a map, we checked the spatial distribution of accidents and traffic counts, ensuring that the merging process was correctly performed and that the data made sense geographically.

**Step 6: Creating Visualizations in QGIS**

**Heatmaps**

We created several heatmaps to visualize the relationship between traffic volumes and accident severity:

* **Traffic Heatmap**: Highlighted areas with high traffic counts.
* **Accident Severity Heatmaps**: Separate layers for fatal, severe, and minor accidents, overlaid with the traffic heatmap.

**Key Insights**

* **Easy Accidents**: Occur frequently regardless of traffic volume.
* **Severe Accidents**: More common on clearer roads.
* **Fatal Accidents**: Occur on both high and low traffic roads, suggesting a more complex relationship.

**Exported Files**

**Files Produced**

1. **cleaned\_accident\_data\_no\_null\_columns.csv**: This file contains the standardized and cleaned accident data across all years.
2. **cleaned\_traffic\_counts\_no\_null\_columns.csv**: This file contains the aggregated traffic counts across all years, showing the average vehicle count per road and day of the week.
3. **accident\_data\_with\_traffic.csv**: The final file merging the accident data with the average traffic counts provides a comprehensive view of accidents with contextual traffic information.

**Final Analysis in QGIS**

* We mapped all accidents where we had traffic count data across all years.
* Created separate layers for fatal, severe, and minor accidents.
* Produced heatmaps for each category of accidents and compared them with traffic heatmaps to identify patterns.

**Conclusion**

* **Correlation Findings**: While the data does not show strong correlations, further steps could include mapping intersections, roundabouts, and traffic lights to gain deeper insights.
* **Next Steps**: The provided dataset can serve as a foundation for further research, potentially expanding the scope to include additional factors like weather conditions or road types.

**Special Thanks**

* **Professor Michael Frenkel**: For guidance throughout the project.
* **Teammates**: Elisheva Klapisch and Lidor Eliyahu Shelef, for their collaborative efforts.