**Assignment 4: Solution Development**

**Safe LA: Enhancing Traffic Safety Through Data-Driven Insights**

**Team 8**

Alexandra Gladkova

Christina Saju

Hao Lun Rong

Kushwanth Sai Kolli

Ladan Asempour

**Capstone Project**

**BIA-5450-0LA**

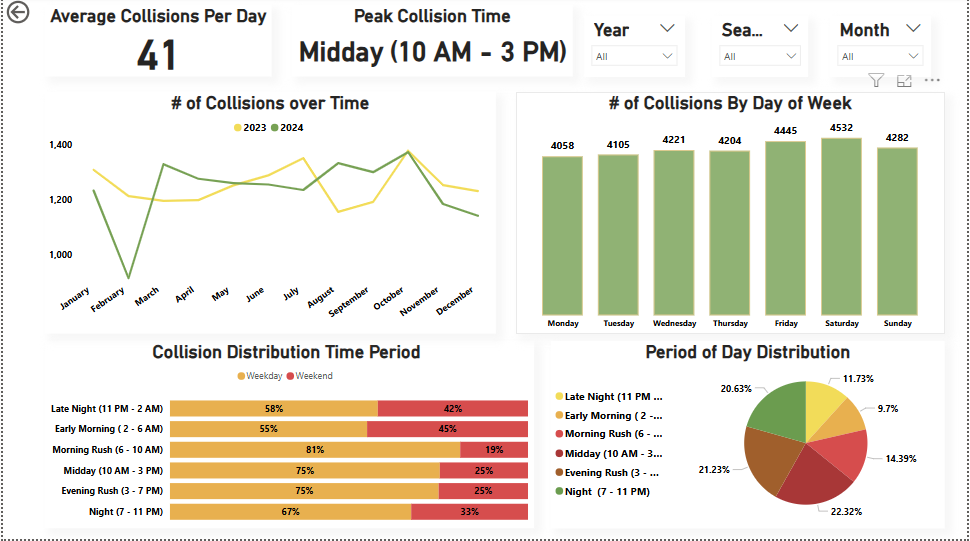
**Samer Al-Obaidi**

**Group Member Contribution**

|  |  |
| --- | --- |
| **Topic** | **Contributor** |
| Implement the Solution | Christina Saju – Hao Lun Rong |
| Outcome Testing and Reviewing | Kushwanth Sai Kolli |
| Optimization | Alexandra Gladkova - Ladan Asempour |

**Implement the solution**

The Solution Design for this project consists of multiple dashboards composed of various visualizations including Fact Cards, Bar Charts, Stacked Column, Line Graphs, Tree Maps among others. In this section, each dashboard, which answers one of our analytical questions, along with the components within each, will be reviewed.   
  
*Analysis of Chronologic Factors*



The solution design of this dashboard consists of multiple cards and visualizations which assess various time-dependant factors. These include time of day, days of the week, and month, as well as the distribution of collisions comparison per years. The dashboard highlights key insights such as the average collisions per day as well as the distribution of collisions by time period. Several trends such as monthly distribution of collisions, comparisons between the weekdays and weekends were also visualized. The dashboard is complimented with various slicers allowing for selection of different months, seasons, and years. The visualizations created are listed as follows:

Average Collisions per Day Card: This visualization emphasizes the amount of car collisions recorded by the LAPD daily throughout the past 2 years. It was created by utilizing the card visualization and inputting a DAX measure in the field, which takes the count of collisions divided by the number of days listed over the 2-year span.

Peak Collision Time Card: Insight into the time period which records the most collisions is given by this card. This was created by using the card visual and inputting a DAX measure that returns the time period with the highest record of collisions.

Collisions Over Time Chart: The line chart shows the count of collisions recorded each month, with separate lines representing each year, respectively, to highlight the contrast in recorded collisions. This was created by utilizing the line chart, inputting the Month and Collisions into the X and Y axes, respectively, and inputting the Year into the legend.

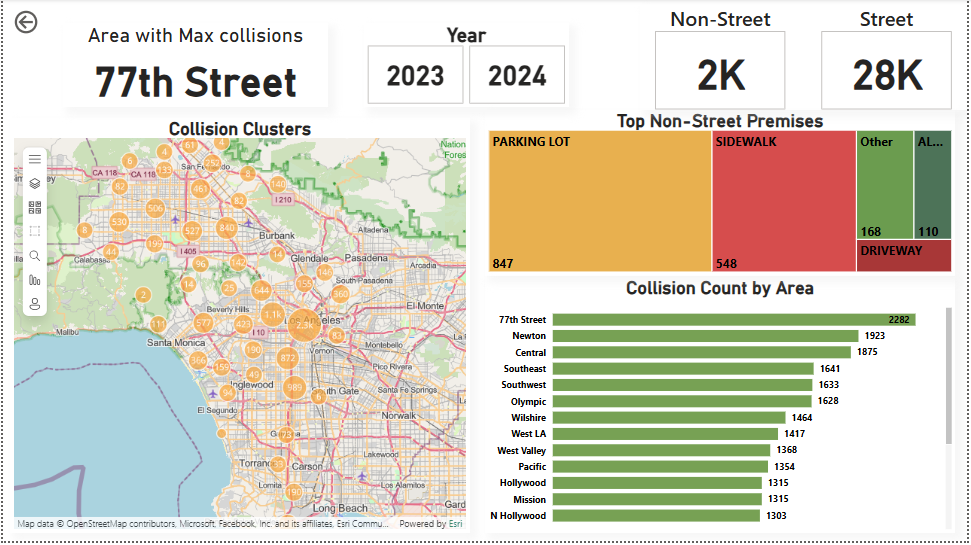
# of Collisions by Day of Week: This visualization shows the total records of collisions grouped by the day of the week they are recorded, providing insight into which days tend to have more collision occurrences throughout the week. This was created by utilizing the column chart, and inputting the Day of the Week and collision counts as the X and Y axes respectively.

Collision Distribution by Time Period: This visualization compares the proportion of collisions during weekends and weekdays, falling within different time periods. It highlights day-specific trends, such as weekend nights being most prone to vehicle collisions, while weekdays show higher collision frequencies during the daytime and evening rush hour periods. The visualization was created by using a stacked bar chart, with the time period and collision count as the X and Y axes, respectively, and the weekend or weekday classification used as the legend.

Period of Day Distribution: This visualization shows the percentage proportion of collisions categorized by the time period in which the collision was reported. It was created using a donut chart, with the Period of the Day as the legend and collision counts as values.

Year, Season and month Slicers: These slicers are meant to create interactive filters by utilizing the slicer visual and inputting the respective category into the Field section.

*Analysis of Geographic Factors*



The solution design of this dashboard consists of multiple cards and visualizations that assess various geographical and location-based factors. These include information on neighborhoods, premises, or surrounding descriptions, as well as maps containing clusters reflecting the frequency of collisions. The dashboard highlights key areas where collisions occur more frequently and provides descriptions of the environment or premises where these collisions take place. The visualizations created are listed as follows:

Area with most Collisions Card: This visual gives information on the area with the most collisions recorded. It was created by using the card visual and inputting a DAX measure that returns the area name with the highest collision count value.

Non-Street and Street Premise Card: This visual gives information on the count of collisions which happened on the road versus the count of collisions on other premises such as parking lots or sidewalks. It was created by using the card visual and inputting the Collision Count grouped by Location as the Data, and a Dax code which categorizes all the street premise and filters the rest as “other” into the categories input.

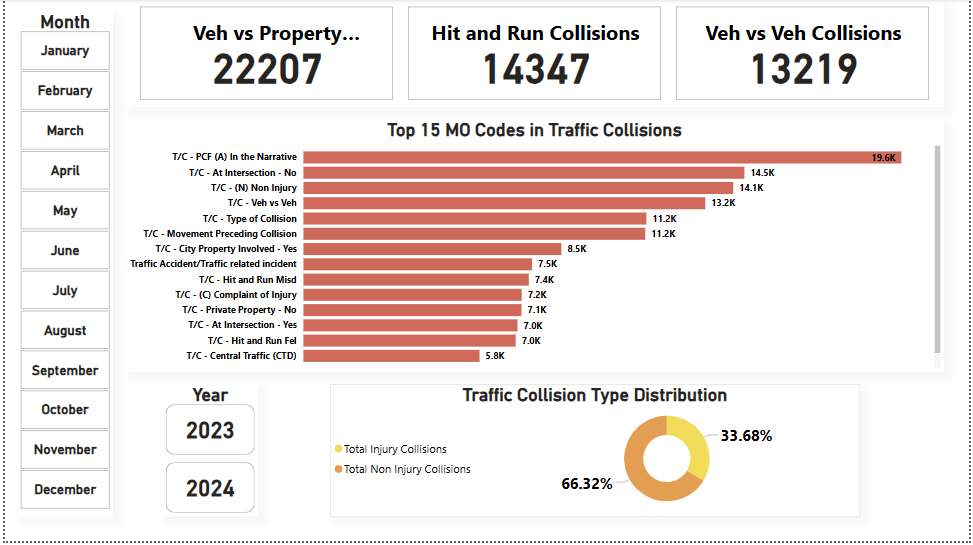
Top Non-Street Collisions Premises: This visual shows the premises that most commonly report collisions, excluding streets, as streets have the highest record of collisions on premises by a large margin. This provides more value into other environments or scenarios where people may be at risk of collisions. It was created by inputting a DAX code that reports the Top 4 non-street premises, with the rest grouped as "Other" into the Category box. The Value box was filled with the Collision Count Per Location.

Collision Clusters Map: This visual shows the collision density for the recorded collisions across LA. The size of the cluster indicates a higher density of collisions within the area. The visual was created by using the ArcGIS visual and inputting the Latitude and Longitude data into it respectively.

Collision Count by Area: This visual shows the number of recorded collisions based on the area or district name. It was created by utilizing a bar chart, with the Y axis representing the Area Name and the X axis representing the Collision Count per Location.

Year Slicers: These slicers are meant to create interactive filters by utilizing the button slicer visual and inputting the year into the Field box.

*Analysis of Crime Report Factors*



The solution design of this dashboard consists of multiple cards and visualizations that assess the factors and descriptions that were entailed in the collision. This gives a closer look at the MO codes that were listed in the collision report. As well as assessing the traffic types during the time of collision. The visualizations created are listed as follows:

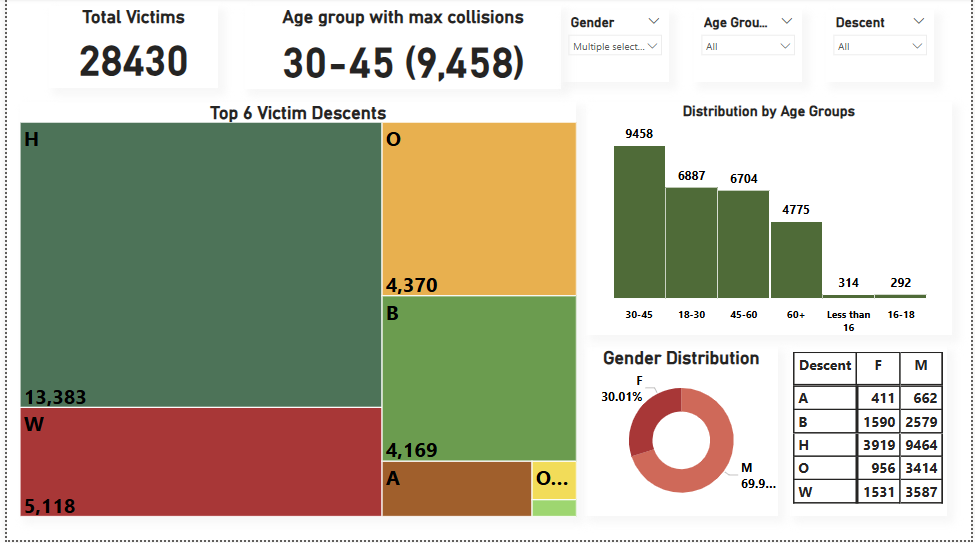
Collision Type Cards: The top 3 visualizations, list the count of each collision type. Which is done by listing the Total Vehicle vs. Property, Total Hit and Run, and Total Vehicle vs Vehicle. This was done by using the card visual and putting the DAX measure to return the filtered data for each.

Top 15 MO codes in Traffic Collisions: This shows the 15 most common MO codes used by the police when recording collision incidents. This was done by using a bar chart and inputting the Description of the MO code, and count of the MO codes for the Y and X axes, respectively.

Traffic Collision Type Distribution: This visualization shows the proportion of collisions where people faced injuries versus collisions with no reported injuries. It was created by utilizing the donut chart, and inputting DAX codes into the Values box, counting the collisions where injury was reported as well as another DAX code where injury was not reported.

Button Slicers: Numerous slicers were created for the months and year to enable interactive filters, using the button slicer visual and inputting the respective category into the Field section.

*Analysis of Demographic Factors*



The solution design of this dashboard consists of multiple cards and visualizations which assess various demographic factors consisting of the Age, Gender, and Ethnicity. The visualizations help identify the most affected age groups, the most affected ethnicities, and compares the gender distribution between them. Slicers are implemented to allow users to filter by specific groups for interactive visualization. The visualizations included are listed below:

Total Victims Card: This visualization displays the information about the total amount of collisions victims within our dataset, and it is done by taking the card visual and inputting the count of the collision records.

Age Group with Max Collisions Card: This visualization displays the age brackets into which the victims' ages have been categorized, and the age bracket with the highest number of recorded collisions is selected. This was done by creating a measure which counts the collisions grouped by the age column and selecting the one the highest value of counts.

Top 6 Victim Descents: This tree map shows the 6 highest ethnicities/descents which are involved in collisions, with the others being grouped together as "other." This visualization was created by using the tree map and inputting a DAX code to return the 6 highest descents and returning "other" for anything lower than the top 6 values.

Distribution by Age Groups: This visualization shows the number of recorded collisions corresponding to the age bracket the victims are in. The visualization was created by utilizing the column chart and inputting the age group and collision record columns in the X and Y axes respectively.

Victim Profile Matrix: This visualization shows the 5 most common ethnicities that are involved in vehicle collisions and are separated by gender to show the distribution of each ethnicity. This was done by utilizing the matrix table, and inputting descent into the rows, victim gender into the columns, and the collision records as the values. The filter was done manually by using the filters tab and selecting top 5 for the filter type.

Gender Distribution: The ratio of men to women involved in collisions is shown in this visualization using our dataset. It was created by utilizing the donut chart, using the victim gender as the legend and the count of collision records in percentage as the value.

Gender, Age Group, and Descent Slicers: These slicers are meant to create interactive filters by utilizing the slicer visual and inputting the respective category into the Field section.

**Outcome testing and reviewing**

During the implementation process, the dashboards were tested to make sure they displayed the data correctly and met the project's goals. The accuracy of each visualization was first checked. For example, the values for the "Average Collisions per Day" and "Peak Collision Time" cards were manually calculated and compared to confirm that the visualizations displayed the correct results. Different methods for building the visualizations were tested, such as using DAX measures for real-time calculations compared to pre-aggregated data. The effectiveness of different chart types was also checked to see which one presented the data most clearly.

The dashboards were also compared to the original project goals. It was confirmed that the visualizations answered the key questions and that the interactive filters made it easy for users to explore the data. In the end, the dashboards successfully communicated the insights.

**Optimization**

Several strategies were employed throughout the process of refining our dashboards from the previous solution design report, along with additional implementations that can be made. The first area was the Data Model, as we had previously not implemented a star schema for the dataset. Multiple tables were separated and then connected with various primary and foreign keys to make the relationships of the tables. This optimized performance when doing our visualizations as we are creating clear and concise dashboards which contain only information from specific tables for the designated dashboard.

DAX measures were also implemented to increase performance time for the visualizations. Multiple DAX formulas were reused across separate charts in a dashboard, rather than creating the same calculations multiple times throughout the table.

Visualizations were also improved by simplifying or narrowing down the options to improve visual clarity. Many graphs were limited to the top “X” amount of a column or constraint, to improve visual clarity from potentially 40 or 50 classes to about the top 6 or 7. The questions we aimed to answer were also iteratively refined as additional visualizations were created throughout the dashboard. This approach allowed each visualization to address a broader scope of questions without becoming repetitive. Furthermore, these visualizations included interactivity features such as the slicers to allow for simple yet effective analysis for potential stakeholders and other viewers.  
Future optimization for the project includes technological improvements, particularly in data collection and the data model. The analysis could be further enhanced by updating the database, including real time integration of the information as well as proposing that the LAPD expands their police reports to include several other factors as previously mentioned such as weather or road conditions during the incident.