Road Accident Report

**Prepared by:** [Sebastian Ramirez Orozco]

**Date:** [6/19/2024]

**Prepared for:** [Recruiters and Hiring Managers]

## Report Summary

Our client wants to create a dashboard on Road Accidents for year 2021 and 2022 and gain some insights on the subject. Following these, the most relevant insights are:

1. During 2021 and 2022 almost half a million casualties happened (417 K), where 84.10% represented slight, 14.19% serious and 1.71% fatal casualties.
2. Most casualties by vehicle were by car, in total 333,484 which represents the 79.80% of the total.
3. Compering casualties from 2021 to 2022 there is a reduction of 26,408, being 2022 the one with less casualties. Monthly there was a difference from 156 to 5376.
4. The top three casualties by road type belong to Single carriageway with 309.7 K and representing 74.11%, followed by Dual carriageway with 67 K and 16.42% and roundabout with 26 K and 6.12%.

Based on what we uncovered I highly recommend the following for the betterment of safety and reduction of casualties:

1. The Public Transportation Ministry should start a road safety and awareness campaign on multimedia platforms. Campaign should focus on general safe driving, distractions during driving, what to do after a traffic accident and obeying traffic laws.
2. Reduce the speed limit on certain areas, especially high pedestrian areas and residential areas. A vehicle traveling 35 mph doubles the likelihood of killing the victim.
3. Redesign streets to include landscaped medians, limited left turns, buffered bike lanes, enhanced bus transit, and additional crosswalks. Many of the road traffic systems are out of date or were never designed to reduce accidents. Implementing these changes can reduce accidents by more than 40 percent.

## Introduction

Did you know the most common causes of car accidents are distractions, fatigue, environmental conditions and medical or emotional impairment? Or maybe that almost half of all fatal car accidents involved alcohol! Facts like this make you aware that something as common as driving can change your life for the worst if you are not careful.

Now that we know, we are analyzing a [Road Accident Dataset](https://docs.google.com/spreadsheets/d/1R_uaoZL18nRbqC_MULVne90h3SdRbAyn/edit?usp=sharing&ouid=116890999875311477003&rtpof=true&sd=true) so we can understand better these tragic accidents. The data was obtained through the guided project of [Data Tutorials in YouTube](https://www.youtube.com/watch?v=XeWfLNe3moM&list=WL&index=30). The project is undertaken as a guided project to familiarize myself with data analysis and reporting.

Our main goal is to answer the business questions from our stakeholders and define our KPI’s based on that. Some examples are the Public Transportation Ministry and the Public.

Business queries:

* Primary KPI’s:

1. What are the total casualties that have taken after an accident?
2. What are the total casualties and percentage of total with respect to accident severity and maximum casualties by car?

* Secondary KPI’s:

1. What are the total casualties with respect to vehicle type?
2. We would like to see the monthly trend showing comparison of casualties for current year (2022) and previous year (2021).
3. What are the maximum casualties by road type?
4. We need the distribution of total casualties by road surface.
5. Is there a relation between casualties by area/location and by day/night?

Answering these questions will allow our stakeholders on business decision making and what strategy to follow, which can bring a betterment in road safety and less road casualties.

## Data analysis

Project Summary

The main goal of the project is to perform data analysis and create a dashboard on Excel to provide an answer to stakeholder’s inquiries.

The project has the following steps:

1. Data Sources and Data Gathering
2. Assessing and Data Cleaning
3. Data Analysis
4. Data Visualization

Project Environment

This project is run entirely through Excel. All calculations, cleaning and visualization are going to be perform using Excel and/or Power Query. Dataset must be downloaded from [link](https://docs.google.com/spreadsheets/d/1R_uaoZL18nRbqC_MULVne90h3SdRbAyn/edit?usp=sharing&ouid=116890999875311477003&rtpof=true&sd=true) and open in Excel.

Step 1. Data Sources and Data Gathering

**Data Sources**

* Road Accident Data.

**Data Gathering**

The dataset is obtained from Google Docs. The dataset includes 21 columns and 3.07 million rows.

Step 2. Assessing and Data Cleaning

**Importing Data**

Data is just open through the Excel File.

**Checking For Blanks And Spelling Errors**

For these steps we will be adding a filter for each header and check one by one blanks and errors. We are selecting the whole header range and on the Home tab, Sort and Filter button we select Filter.

On the Junction Control column, we find Auto traffic signal and Auto traffic sigl. To fix the error, we will go to Find and Select, replace, and change sigl to signal while selecting the whole column.

If we check the Accident column, among the values we will find Fatal and Fetal. Again, we are using Find and Select, replace, and change Fetal to Fatal.

After fixing spelling errors, we find a large quantity of blank values on some columns. The way to proceed is changing “” for Not specified, since it will be easier and faster than imputing by Frequent Categorical Imputation. So, no assumptions.

**Adding Columns**

We are adding to columns based on the Accident Date columns. One for month using the function TEXT(XX, “MMM”), and year TEXT(XX, “YYYY”). These columns help us to answer our business queries.

**Identify And Remove Duplicate Records**

As the last step we will search for duplicates and remove them from our tables. In this case only one duplicated record was found.

Step 3. Data Analysis

**Convert To Table**

The data we will be using for our analysis, will be converted as a table, and named Road Accident Data. This will allow us to work with pivot tables and pivot charts later.

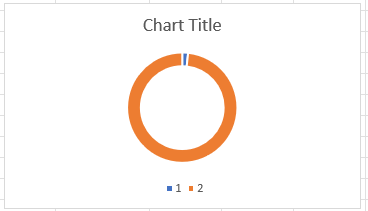
**Pivot Tables**

* A pivot table named Total Casualties will be created on a new sheet based on our Road Accident Data table. New sheet is named KPI’s. To create this table, just drag Number of casualties to values and it will Sum al casualties for us.
* For our second pivot table, we are copying the first one on the same sheet and named it Total Casualties By Severity. We just add Accident Severity to row values.
* Since we need percentage of total with respect to accident severity, we will create small tables to achieve these.   
  For Fatal Severity: Here using the Total Casualties By Severity pivot table, we get the total casualties by fatal and the other two together to get the percentages of each one.

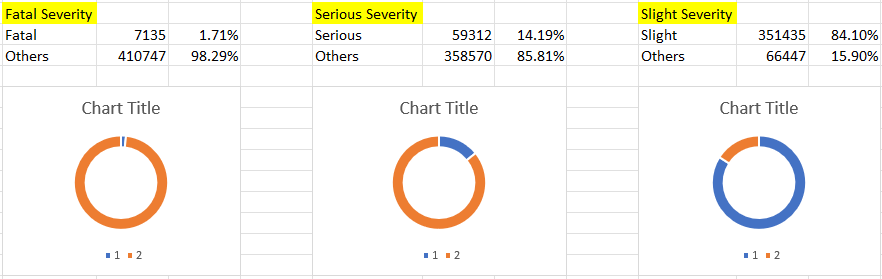
A white grid with black numbers

Description automatically generated

We fixed the F8 and F9 cells so is easier to drag the formula to the row below. Also, a doughnut chart using said values is created. Is important to change the format of our calculations to percentage and increase the decimals using the home tab, number format option.



We repeat the same process for Serious and Slight Severity. Note. The charts are going to be formatted later.



**Dashboard**

Even though we are not done with our pivot tables, the building of our dashboard is necessary. To start, a new sheet named Dashboard will be created.

* We will insert some shapes, format them, and add titles to begin with. In here, we are inserting our primary KPI’s.

A screen shot of a graph

Description automatically generated

Note. You can drag shapes and text boxes using alt to fit better on the cells.

* Next step is cut and paste our doughnut charts next to the primary KPI’s and format them. We also include de percentages using text box.



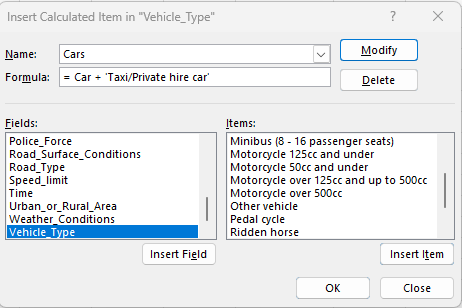
Is also important to notice that if change the format number so is easier to read for the audience.

* For the fourth primary KPI we will create a new pivot table. To create it, we are copying the Total Casualties By Severity table and replacing the accident severity row for vehicle type. This new table will be named Total Casualties By Vehicle.

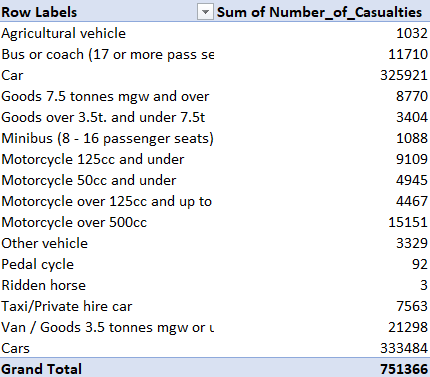
A screenshot of a computer

Description automatically generated

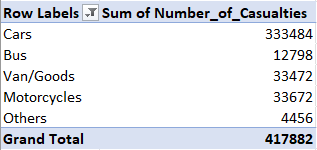
As we can see, there are too many categories. To address this, using the fields, items and sets bottom on the pivot table analyze tab we will merge some categories. Selecting calculated items, we are merging car and taxi/private hire car into a new category named cars as shown below.



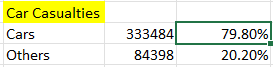
As we can see, a new category named cars is created and we just need to uncheck the car and taxi/private hire car from the filter in our pivot table.



We repeat the same process for the rest of the categories until we got a table like the one below.



Our last step is creating a manual table to see the percentage of car casualties and just repeat the same process as the previous doughnut charts.



* For the secondary KPI’s we need the total casualties by each vehicle, so we are inserting icons using the icon button and the totals from a function using the Total Casualties By Vehicle. We cannot directly use the values from the pivot table since an error will occur.



A screenshot of a number

Description automatically generated

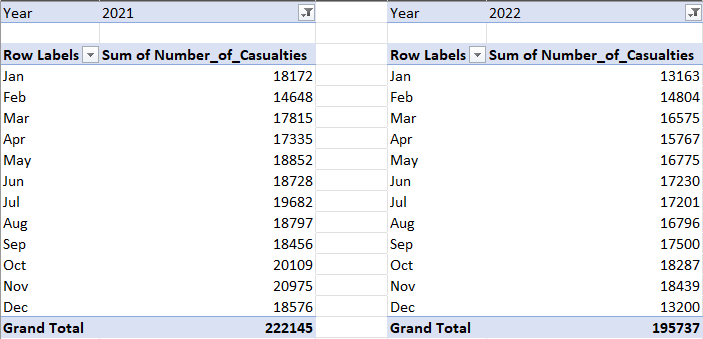
Here we can see the result:

A screenshot of a cell phone

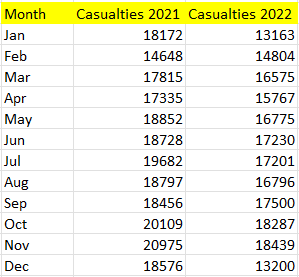
Description automatically generated

* For the secondary KPI’s, the customer is asking for the monthly trend of casualties of both years, so we are creating a new pivot table in a new sheet and name it Casualty Monthly Trend 2021 and a copy named Casualty Monthly Trend 2022.

A screenshot of a computer

Description automatically generated 

To create the line chart, we will create a manual table that includes the casualties from 2021 to 2022 as shown below.



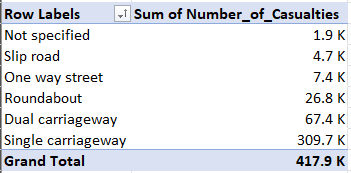
And now we insert the chart and format it.

A graph showing the number of trends

Description automatically generated with medium confidence

* Next, we are required by the customer to display the maximum casualties by road type. For this task we will create a new pivot table from out dataset Road Accident Data and insert a bar chart using the new table.

The new pivot table is named Maximum Casualties By Road Type and the sheet is also named the same. The table is created with the Road Type on the row value and Sum of Number of casualties on values, we also sort it from smallest to largest and format the numbers.



Once we create the bar char, we format it and add it to our dashboard.

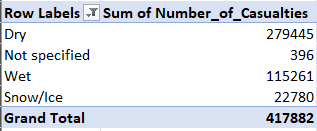
A screenshot of a computer

Description automatically generated

* Now we are asked for the total casualties by road surface. Again, we create a new pivot table and named it Casualties By Road Surface. For said table we use Road Surface Conditions and Number of casualties. There is certain redundance in our categories so we will merge some categories using the the fields, items and sets bottom on the pivot table analyze tab. Here is a before and after merging our categories.

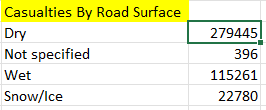
A screenshot of a computer

Description automatically generated



Now using this we will create a tree map, but since is not possible to create it suing the pivot table, we will copy outside the data.





Here is the tree map formatted.

A screenshot of a computer

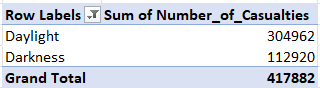
Description automatically generated

* As our final step, customer wants to see if there is a relation between casualties by area/location and by day/night. To accomplish these, we are creating two pivot tables in a new sheet and named it Area Light Casualties. First one is for area casualties, and the second one for light condition casualties.

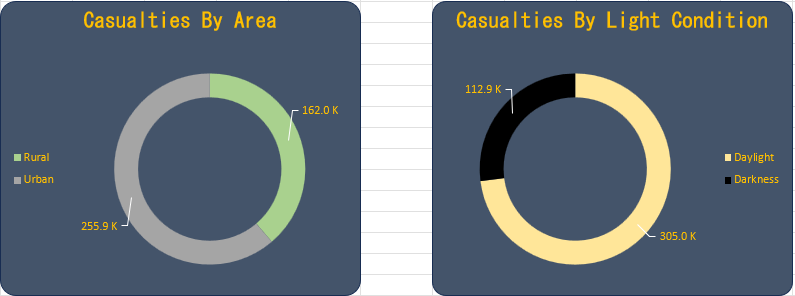
The Area Casualties does not need any changes, but Area Light Casualties. Area Light Casualties has redundance as shown below, so we will merge again some categories using the fields, items and sets bottom on the pivot table analyze tab.

A screenshot of a computer

Description automatically generated



With these two tables done, we proceed to create for each a doughnut chart and format it.



* With all the KPI’s included on our dashboard, we proceed to add some slicers to filter our data. We are adding one for the accident date and another for urban and rural area. To do this, just select any pivot table from the KPI’s sheet and add the slicers from the insert slicer button from the PivotTable Analyze tab.

Is important to format these slicers so they look well in our dashboard, and make sure to report the connections between all tables in our slicers.

A screenshot of a phone

Description automatically generated

Note. For the final touch we just adjust the columns and rows width and height and disable gridlines. Optionally disabling the formula bar and heading will make your dashboard look cleaner.

Step 4. Data Visualization

A screenshot of a computer dashboard

Description automatically generated

**Business Queries**

* Primary KPI’s:

1. What are the total casualties that have taken after an accident?

The total casualties from 2021 to 2022 are 417,882. Data is obtained through a pivot table, where it sums the total of casualties from each individual record.

A blue background with yellow text

Description automatically generated

1. What are the total casualties and percentage of total with respect to accident severity and maximum casualties by car?
2. Fatal casualties represent 1.71% of all casualties, with a total of 7,135. Is noticeable that fatal represent the less frequent among all casualties.

A blue background with yellow text

Description automatically generated

1. Serious casualties represent 14.19% of all casualties, with a total of 59,312. Second most common.



1. Slight casualties represent 84.10% of all casualties, with a total of 351,435. This category represents the most common of all casualties registered.



1. Casualties by car represent 79.80% of all casualties, with a total of 333,484. Cars are the vehicles most involved in casualties in our data. The other vehicles only represent 20.20%.



Note. All KPI’s above were collected through pivot tables and manual calculations.

* Secondary KPI’s:

1. What are the total casualties with respect to vehicle type?

The most casualties belong to cars with 333,484 in total, following by motorcycles with 33,672, cargo vans 33,472, buses 12,798 and other vehicles 4,456. For the other vehicle category are included agricultural machines, bicycles, horses, and others not stated.

A screenshot of a phone

Description automatically generated

Note. All KPI’s above were collected through the total sum of the above categories in a pivot table.

1. We would like to see the monthly trend showing comparison of casualties for current year (2022) and previous year (2021).

As shown below, on the left side of our line chart we have the number of casualties and on the bottom the months. It is noticeable that there is a reduction of casualties in 2022(195,737) compering to 2021(222,145), where on both years the worst month regarding the number of casualties was in November.   
  
The lowest casualties for 2021 was in February with a total of 14,648 and for 2022 in January with 13,163 total. Is also important to mention that the casualties for February of both years are very similar with a difference of 156 casualties.

A graph of a trend

Description automatically generated with medium confidence

The reasons for why 2022 casualties are lower to 2021 are to speculations since we lack domain data like the origin of the dataset, country of origin (most probably UK), weather, changes on road structure, awareness campaigns, absence of stakeholders to talk to and others.

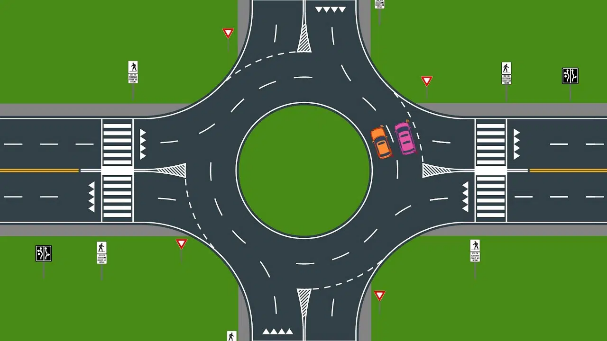
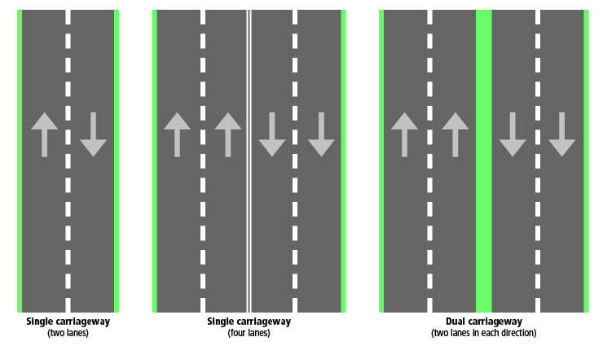
1. What are the maximum casualties by road type?

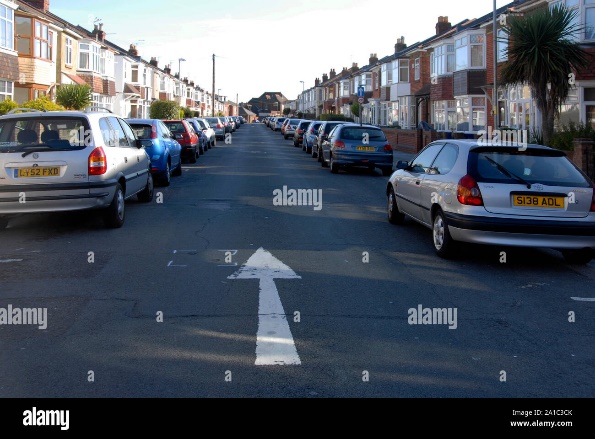
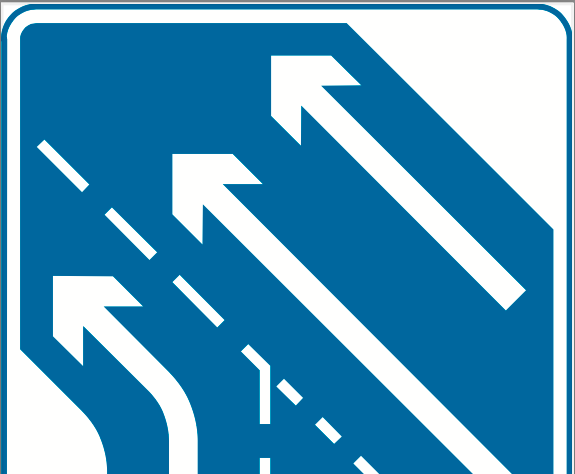
Single carriageway represents most casualties with 309.7 K in total, next is follow by dual carriageway with 67.4 K, roundabout 26.8 K, slip roads 4.7 K and a not specified category with 1.9 K.

A screenshot of a computer

Description automatically generated

Below I attached explanatory images of each category for a better understanding.



Note. From left to right, first image Dual and Single carriageway, second image Roundabout, third One-way Street, and the last one Slip Road.

1. We need the distribution of total casualties by road surface.

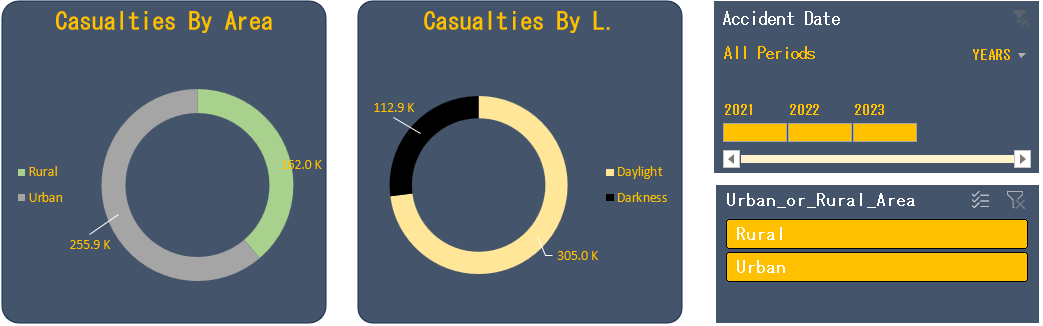
Most casualties occur on dry roads with 279,445 in total and it represents the 66.87% of all casualties. Wet is the second most common with 115, 261 and 27.58%, followed by Snow/Ice roads with 22,780 and 5.45%. There is also a fourth category for not specified roads that represent 0.09% and 396 casualties, which is almost imperceptible at the bottom right of our tree map chart.

A screenshot of a computer screen

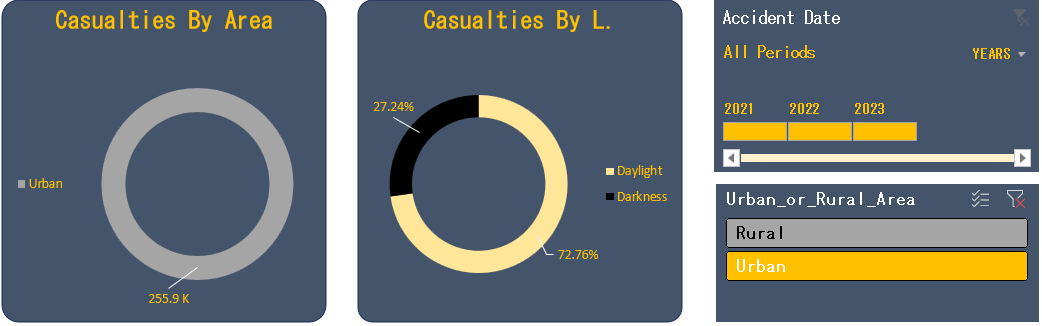
Description automatically generated

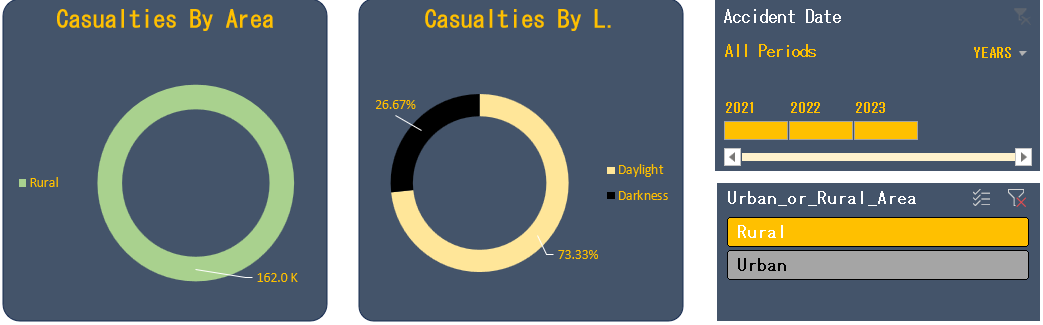
1. Is there a relation between casualties by area/location and by day/night?

Here are the original doughnut chart including the rural and urban areas.



And now if we filter it by urban and later by rural in percentages.





It is noticeable that the percentages remain similar, so we cannot conclude there is a relation between area and light conditions.

Note. All charts were generated using pivot tables.

## Findings

* During 2021 and 2022 almost half a million casualties happened (417 K), where 84.10% represented slight, 14.19% serious and 1.71% fatal casualties.
* Most casualties by vehicle were by car, in total 333,484 which represents the 79.80% of the total.
* Compering casualties from 2021 to 2022 there is a reduction of 26,408, being 2022 the one with less casualties. Monthly there was a difference from 156 to 5376.
* The top three casualties by road type belong to Single carriageway with 309.7 K and representing 74.11%, followed by Dual carriageway with 67 K and 16.42% and roundabout with 26 K and 6.12%. The rest of categories representing 3.35% only.
* Casualties by road surface are dominated by dry surface with 66.87%. Wet surface represents 27.58% and snow/ice 5.45%.
* There is not a relation between casualties by area/location and by day/night while compering the data, since percentages remained very similar for both cases.

There is a clear problematic regarding road safety, even though there is a reduction of casualties and most represent slight ones.

Note. Some percentages and calculations were added on some tables to improve the understanding of the data. These are represented by red cells on the title and are not included in the data analysis above.

## Recommendations

I would advise to:

1. Short Term:

* The Public Transportation Ministry should start a road safety and awareness campaign on multimedia platforms. Campaign should focus on general safe driving, distractions during driving, what to do after a traffic accident and obeying traffic laws.
* Reduce the speed limit on certain areas, especially high pedestrian areas and residential areas. A vehicle traveling 35 mph doubles the likelihood of killing the victim.
* Guidelines for later start times of parking enforcement to encourage impaired drivers to leave their car overnight.
* Installment of safety cameras and better position of the already existing ones will encourage drivers to behave better for the fear of tickets.

1. Long Term:

* Redesign streets to include landscaped medians, limited left turns, buffered bike lanes, enhanced bus transit, and additional crosswalks. Many of the road traffic systems are out of date or were never designed to reduce accidents. Implementing these changes can reduce accidents by more than 40 percent.
* Traffic Safety Education programs starting from early ages is a great idea to create conscience before even the legal age for driving.
* Strict law enforcer could be useful with a zero tolerance for the use of alcohol or drugs, this could mean the driver license suspension and heavy fines to discourage said behaviors.
* Legalization of driverless cars when fully operational and autonomous is great idea since most of the accidents are the result distractions.

## Conclusion

To conclude this report, based on our findings we know there are 417,882 casualties from 2021 to 2022, where slight casualties represent 84.10%, serious 14.19% and fatal 1.71%. Also, that in most of casualties, cars were involved with 79.80% of all vehicle types.

I highly recommend acting based in the result of our findings, since we should always strive to minimize traffic accidents and have a goal of zero fatalities and serious injuries. According to Illinois Department of Transportation just by improving and redesigning streets to include landscaped medians, limited left turns, buffered bike lanes, enhanced bus transit, and additional crosswalks and traffic signs can reduce by more than 40% accidents.

This percentage alone is a huge improvement alone, now imagine the difference if all the recommendations are followed on the short and long term. I urge all our collaborators and stakeholders to act to make our roads safer.

## Appendices

Here you will find essential supplementary material that supports the main report.

* **Road Accident Data:** composed of 23 columns and 307,976 rows. Columns are Accident Index, Accident Date, Day of Week, Month, Year, Junction Control, Junction Detail, Accident Severity, Latitude, Light Conditions, Local Authority (District), Carriageway Hazards, Longitude, Number of Casualties, Number of Vehicles, Police Force, Road Surface Conditions, Road Type, Speed Limit, time, Urban or Rural Areas, Weather Conditions and Vehicle Type.

A screen shot of a computer

Description automatically generated

* **Pivot table for total casualties:** Sum of all casualties. Yellow table to export data to dashboard.

****

* **Pivot table for accident severity:** Sum of all casualties by severity.

A screenshot of a computer

Description automatically generated

* **Pivot table for casualties by vehicle type:** Sum of all casualties by vehicle. Yellow table to export data to dashboard.

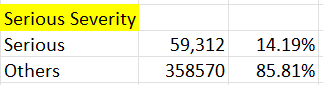
**A screenshot of a computer

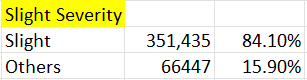
Description automatically generated**

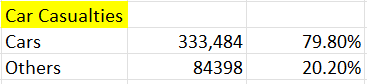
* **Manual tables:** To export data in percentages to dashboard for Fatal Severity, Serious Severity, Slight Severity and Car Casualties.

**A screenshot of a phone number

Description automatically generated**

****

****

****

* **Pivot table for Monthly trend:** Two pivot tables, one for each year with a sum of casualties by month. Yellow table to export data to dashboard. Red to display difference of casualties.

A screenshot of a computer

Description automatically generated

* **Pivot table for casualties by road type:** Sum of all casualties by road type. Red table to display percentages.

A screenshot of a computer

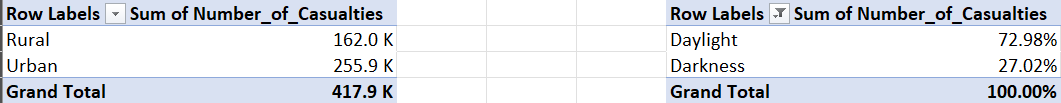
Description automatically generated

* **Pivot table for casualties by road surface:** Sum of all casualties by road surface. Yellow table to export data to dashboard. Red to display percentages.

A screenshot of a computer

Description automatically generated

* **Pivot tables for area and light condition casualties:** Sum of casualties by area and percentage of light conditions.



## References

* **Sources:** <https://www.youtube.com/watch?v=XeWfLNe3moM&list=WL&index=30>

<https://www.arrowheadclinic.com/category/blog/14-unbelievable-car-accident-facts>

<https://www.nsc.org/saferoads>

<https://natlawreview.com/article/improving-roadway-designs-traffic-safety>

<https://highways.dot.gov/public-roads/winter-2022/01>

<https://www.fmcsa.dot.gov/ourroads/all-our-roads-our-safety-campaigns>

<https://www.idot.illinois.gov/transportation-system/safety/roadway/index>

<https://medium.com/analytics-vidhya/ways-to-handle-categorical-column-missing-data-its-implementations-15dc4a56893>

* **Tool citations:** Excel.
* **Acknowledgments:** To Data Tutorials on his guided excel project.