Road Accident Analysis in Excel

This article is for educational purposes only. The numbers do not reflect real data.

- Find access to the video tutorial here
- Download the dataset here

INTRODUCTION

Data analysis in Excel is widely used in many industries, from finance and accounting to marketing and sales. In this project, we will use several data analysis tools in Excel to create a stunning interactive dashboard.

The dataset used in this project is about road accident casualties in UK in year 2021 and 2022. This project will helped you understand

some key insights about how various factors cause road accident casualties.

THE GOAL OF THIS PROJECT:

The first step is to understand the KPIs of the project. The primary KPIs of this project are:

- Total casualties and percentage of total with respect to accident severity
- Number of casualties with respect to type of vehicle

The secondary KPIs of this project are:

- Monthly trend of showing casualties for year 2021-2022
- Casualties with respect to road type
- Casualties with respect to road surface
- Relation between casualties with Urban/Rural area
- Relation between casualties with Day/Night

THE DATA

This dataset has approximately 31k rows and 23 variables. Each row holds information about a road accident. This information includes following features relevant to KPIs:

- Accident date
- Accident_Severity
- Number of casualties
- Road_surface_ condition

- Light_Conditions
- Speed_limit
- Vehicle_type
- Urban or Rural area

A significant part of working on a data analysis project is cleaning data to make it ready to analyze. Begin the project by cleaning the data in Excel.

Go over each column/field and primarily use filtering Excel feature to clean data through this following steps:

1. Searched for misspelling and typos and corrected them (e.g Fetal and Fatal in Acedent_Severity). You have to replace all Fetal with Fatal
2. Deleted duplicated entries
3. Made sure there were no blank entries
4. Made sure the units of measurement were consistent

Once the data looked clean, start with data processing and feature engineering. Data processing and feature engineering, helps you transform data into usable information and make it more insightful.

Follow various steps to make data more meaningful and ready to accomplish the KPIs mentioned above:

1. Grouped similar row categories to get a more general data trend
2. Added a new feature "month" by extracting month from the date column to get a better idea of monthly trend of accident casualties
3. Added a new feature "year" by extracting year from the date column to get a better idea of yearly trend of accident casualties.

KEY INSIGHTS

After analyzing this dataset, you will discover:

- Total casualties was 417883 in different locations across UK in year 2021-2022.
- Most of the casualties was caused by slight severity accidents as compared to serious and fatal severity accidents.
- Highest number of the casualties was caused by car accidents as compared to other vehicle types.

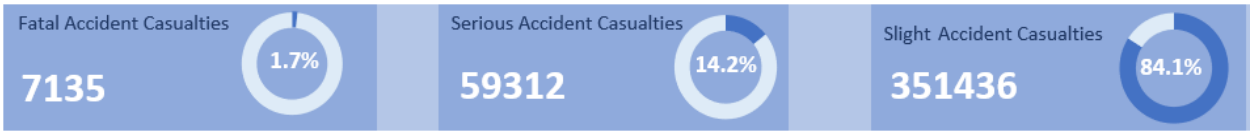
ANALYSIS:

Start the project by looking at the total number of casualties caused in road accidents by creating a pivot table to find the sum of all casualties:

Sum of Number_of_Casualties
417883

Move on to the other primary KPIs, create several pivot tables to find out the required information.

Explore which accident severity category caused most casualties and create a pivot table for all three categories and plot donut charts for percentages of each categories.



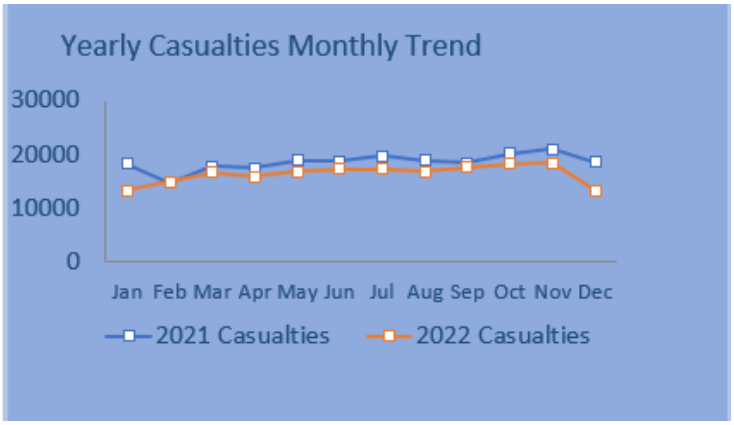
Results show that maximum number of casualties was caused by slight severity accidents which constitutes 84.1% of the total casualties.

Which vehicle type had most casualties?

Row Labels	Sum of Number_of_Casualties
Agricultural ve	1032
Cars	333485
Motorcycle	33672
Buses	12798
Goods Vehicle	33472
Others	3424
Grand Total	**417883**

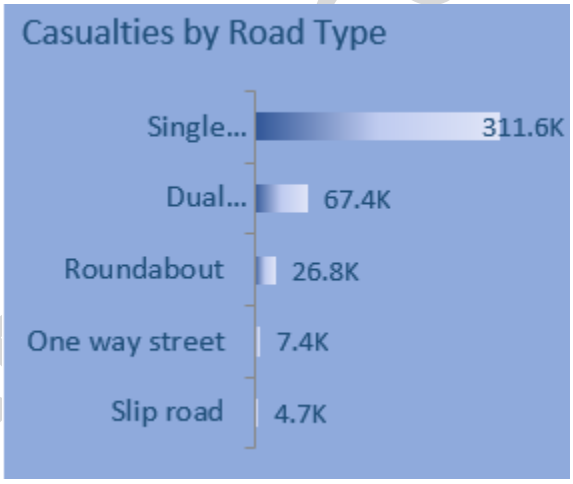
To create this pivot table, group some similar categories of vehicle types and finalize the ones shows above. Highest number of casualties was for car accidents. It is interesting to note that second highest number of casualties is by goods vehicles for examples trucks.

Which year had more casualties?



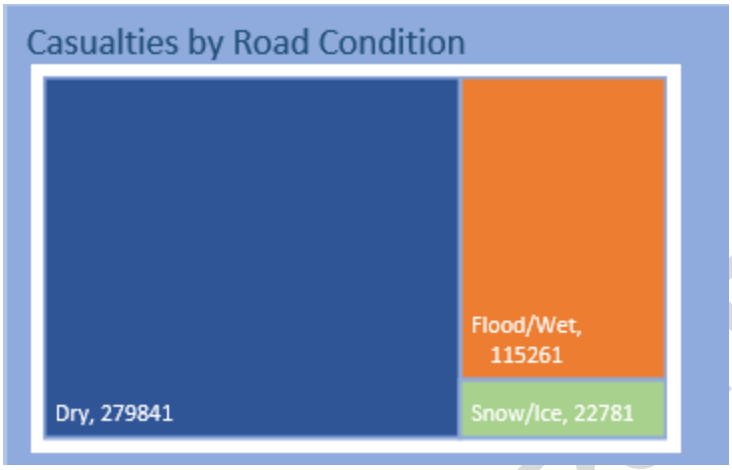
To plot this line chart, use filter to separate casualties in years 2021 and 2022 and create pivot tables for those years. Trend is almost same for both years except for the month of January. However it is interesting to note that year 2022 overall shows lesser number of casualties as compared to 2021. Further investigation needs to be done to find out what measures improve these numbers in year 2022.

Which road type had most casualties?



To answer this question, create a pivot table to get sum of casualties on all road type categories and plot the data on a bar chart. It shows that most of the casualties happened as a result of accidents on Single Carriageway roads.

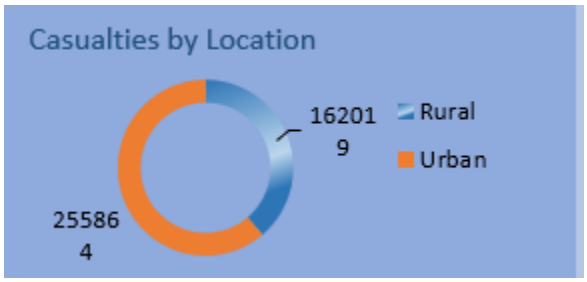
What road conditions caused most casualties?



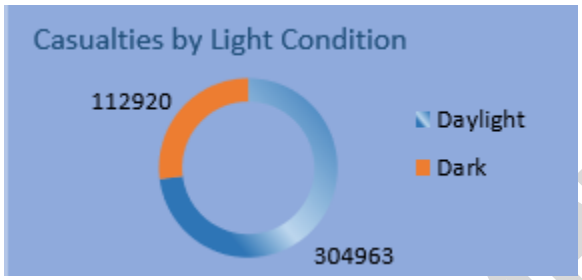
Create a tree map to show how different road conditions impacted number of casualties. The results show dry condition caused 279841 casualties which was the highest as compared to snow and flood conditions. This is probably because people tend to stay at home in bad weather conditions and leave only if it's necessary hence the frequency of accidents is less in bad weather as compared to good weather. However, relationship between accident severity and weather conditions would be an interesting analysis.

Did more casualties happen in rural or urban roads?

Create a donut chart, which from chat bellow shows more casualties happened in urban areas as compared to rural areas.

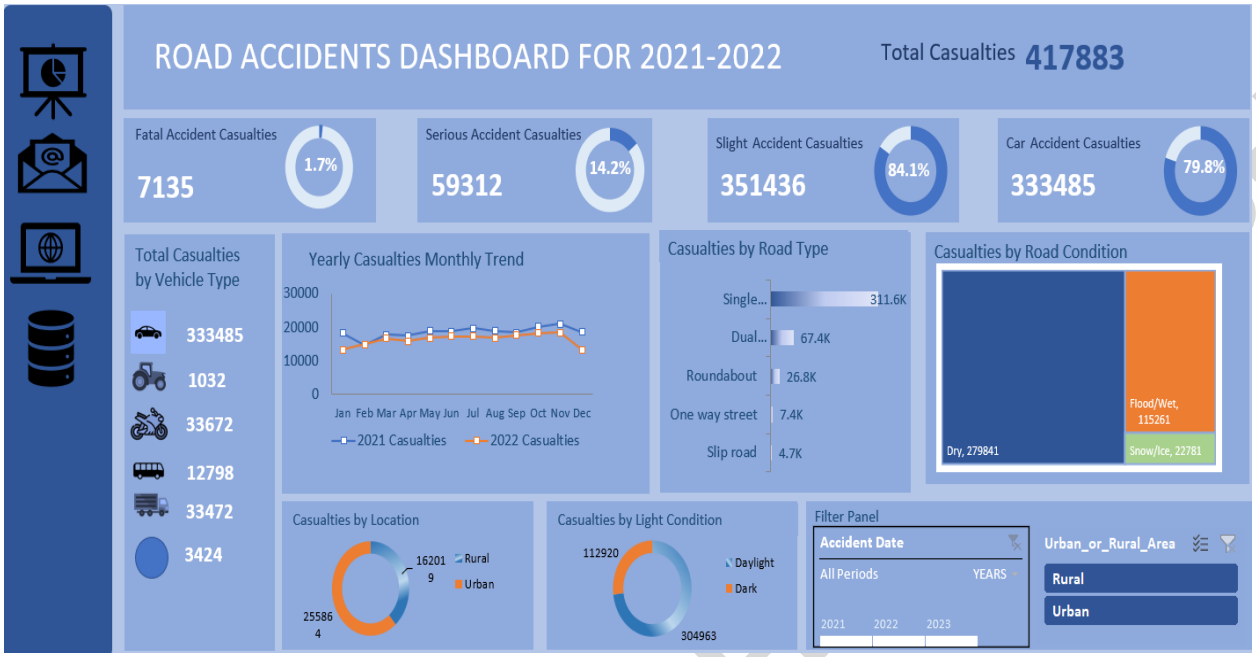


Did more casualties happen during the day or night?



Also, donut chat will help solve this. Surprisingly more casualties happened during day light hours; one of the reasons is there are more vehicles on roads during the day which increases frequency of accidents.

THE INTERACTIVE DASHBOARD:



Click here to watch a short video of the interactive dashboard

Important Excel features to be use in this dashboard are:

- Pivot Charts
- Filtering
- Sorting
- Donut charts, line charts, tree map, bar chart
- Slicer
- Timeline
- Linking timeline and slicer with the dashboard to make it interactive
- Linking dashboard sheet to the data sheet and vice versa

FINAL WORD:

Excel is a crucial tool for data analysis, and it offers a range of features that enable users to manipulate and analyze large amounts of data efficiently. In this project you will be able to create an interactive dashboard about in Excel to get insights about road accident casualties in UK. Along the way you can explore various data analysis features. The major target is to make the dashboard interactive.

TASK

After you are done with the dashboard successful, post it on social medial (LinkedIn, Instagram, Facebook) with # tag (#MyFirstExcelDashboard, #DataAnalysis, #DataVisualization #adefam). Also tag Adefam Computer Infotech on all the platform.

References:

- Full Project in Excel with Interactive Dashboard | Excel Project | Excel Project from Start to End by ***Data Tutorials***
- I can make your data tell a story! By ***Naima M.***