

ROAD ACCIDENT ANALYSIS

Hello, I'm Clifford and I embarked on a project to analyze road accident data, aiming to extract meaningful insights and visualize them effectively using Microsoft Excel. Here's how I went by the project, from data acquisition to developing an interactive dashboard.

Project Overview

The goal of this project was to develop a Road Accident Dashboard for the years 2023 and 2024, focusing on analyzing traffic accident data. The dashboard would give important information to stakeholders and shape strategic policies to reduce accident casualties.

Some insights to be drawn were:

1. Total casualties post-accident.
2. Total casualties by accident severity and their percentage.
3. Total casualties with respect to vehicle type.
4. Monthly trend comparing casualties for the current and previous years.
5. Maximum casualties by road type, top 3.
6. Relationship between casualties on the basis of Area/Location and Day/Night.

Important Stakeholders:

Ministry of Transport, Road Transport Department, Police Force, Emergency Services Department, Road Safety Corps, Transport Operators, Traffic Management Agencies, Public, Media.

Getting the Data

First, I needed to get the data. I downloaded this dataset from Kaggle: [Coffee Bean Sales Raw Dataset](#). The dataset had only one sheet which contained all the necessary records. The dataset has 307,974 rows and 21 columns. It contains accident details with fields like accident_index (unique for every data entry recorded), accident date, location details, accident severity, number of casualties, road type, vehicle type, and others.

After obtaining the dataset, I performed the following steps which are:

1. Data Cleaning
2. Data Analysis
3. Data Visualization

Methodology

1. Data Cleaning

a. Data Exploration:

A thorough inspection of the dataset in Microsoft Excel was made to enhance comprehension and identify areas requiring cleaning. Notable issues included blank entries and spelling errors.

Specifically, the 'Accident_Severity' and 'Junction_Control' columns revealed errors such as 'Fetal' instead of 'Fatal' and 'Auto traffic sigl' instead of 'Auto traffic signal'.

b. Data Cleaning:

The data cleaning process started with copying the entire workbook into another workbook. I always do this and save the original dataset so I can come back to it for any reference. The subsequent cleaning procedures were executed on this dedicated sheet, comprising:

- Columns containing typos were corrected.
- Similar data in the ***Light Conditions*** column were all changed to one. (eg. Darkness – no lit, Darkness – lit were all changed to Darkness). There were to light conditions afterwards; Daylight and Darkness
- The years were 2021 and 2022, I changed them to 2023 and 2024 respectively.
- All errors were corrected using the ***Find and Replace*** function in Excel.



2. Data Analysis

a. Data Modelling:

The dashboard was to be showcased in monthly and yearly insight, but from our data, we only have an ***Accident Date*** column with the full date for each data entry. New columns were created for the day, month and year of accident.

b. Data Analysis:

In this step, I used pivot tables to generate some useful insights. I had a total of six pivot tables which show the following insights:

1. Total number of casualties.
2. Total number of casualties for each Accident Severity type (fatal, severe, slight). The percentage of total casualties for each Accident Severity type was calculated.
3. Total number of casualties by Location, by Day/Night, by Road type.
4. Total number of casualties by Vehicle type. Similar vehicles were consolidated into one using **Calculated Item**. (E.g., Taxi and Private cars were now grouped as Cars)
5. Total number of casualties grouped by months for the two years.

6. Relationship Between Casualties based on Area/Location and Day/Night: Understanding the relationship between these variables contributes to targeted safety measures, especially in specific areas.

Totl Number of Casualties			Number of casualties by road type		
Sum of Number_of_Casualties			Road_Type	Sum of Number_of_Casualties	
	417883		Roundabout		6.64%
			Dual carriageway		16.68%
			Single carriageway		76.68%
Number of casualties by location			Grand Total		100.00%
Urban_or_Rural_Area	Sum of Number_of_Casualties				
Rural		38.8%			
Urban		61.2%			
Grand Total		100.00%			
			Number of casualties by Vehicle type		
Number of casualties by date			Vehicle_Type	Sum of Number_of_Casualties	
Sum of Number_of_Casualties	Year		Cars		333485
Month	2023	2024	Bus		34096
Jan		18173 13163	Motorcycle		33764
Feb		14648 14804	Goods Vehicle		13206
Mar		17815 16575	Other Transports		3332
Apr		17335 15767	Grand Total		417883
May		18852 16775			
Jun		18728 17230	Number of casualties by Day/Night		
Jul		19682 17201	Row Labels	Sum of Number_of_Casualties	
Aug		18797 16796	Daylight		73.0%
Sep		18456 17500	Darkness		27.0%
Oct		20109 18287	Grand Total		100.00%
Nov		20975 18439			
Dec		18576 13200	Number of casualties by Severity		
			Accident_Severity	Sum of Number_of_Casualties	
			Fatal		7135
			Serious		59312
			Slight		351436
			Grand Total		417883

Also, I created a section called ***Making The KPI*** where I calculated the totals which would be used for the KPIs.

I calculated the total casualties for each accident severity and its percentage.

I calculated the total casualties for each vehicle type.

MAKING THE KPI			
Fatal	7,135	1.7%	
Others	410,748	98.3%	
Serious	59,312	14.2%	
Others	358,571	85.8%	
Slight	351,436	84.1%	
Others	66,447	15.9%	
Total Casualties	417,883		
Car	333,485		
Bus	34,096		
Motorcycle	33,764		
Goods Vehicle	13,206		
Other Transport	3,332		

3. Data Visualization

From the pivot tables, I made relatable charts and integrated slicers and a timeline for enhanced dashboard navigation.

1. Slicers were created for the Location and Light Conditions.
2. A timeline was created for the Accident dates. It could be filtered by day, month, quarter and year.
3. Bar charts were created to show the Top three Road Types with the highest casualties.
4. A line chart was created for the Monthly Trend of Casualties comparing the two years.
5. I created doughnut charts to show the Casualties by Day/Night and Casualties by Location
6. Lastly, fields were created for the KPIs (Accident Severity and Total Number of Casualties) using shapes and text fields.
7. The colors used were:



Conclusion

This Road Accident Analysis project was a fascinating journey into the power of data analytics. By preparing the data and leveraging Excel, I created a comprehensive and interactive dashboard that serves as a valuable tool for the Major Stakeholders.

This project not only shows current business performance but also helps in making strategic decisions by highlighting key trends and areas for improvement to help reduce road accidents.