U.K. Accidents- Ten Years History.

Sheetal Munjewar

2023-02-12

Contents

Introduction

Road safety is the common concern around the world, As a part of this exercise we are going to explore U.K road safety data about the circumstances of personal injury road accidents in GB from 2005 to 2014,

Data Source link : https://www.kaggle.com/datasets/benoit72/uk-accidents-10-years-history-with-many-variables

Different data Sources files (cvs):

Accident file: main data set contains information about accident severity, weather, location, date, hour, day of week, road type. . . Vehicle file : contains information about vehicle type, vehicle model, engine size, driver sex, driver age, car age. . . Casualty file: contains information about casualty severity, age, sex social class, casualty type, pedestrian or car passenger. . . Lookup file : contains the text description of all variable code in the three files

License - http://www.nationalarchives.gov.uk/doc/open-government-licence/version/3/

Research Questions

- Does road types has an effect on road accident count ?
- Co-relation between weather with number or severity of an accident?
- Does driver age has an effect on the number of accident?
- What is the relation between hour, day, week, month with number of fatal accident?
- Are certain car models safer than others?
- Is the social class of a casualty dependent of the accident severity?

Approach

Data must be collected from legal source (Publicly available), Check for missing data, merge the different data sources/files into one data frame. In out case we have four data sources. Map column codes with text string for look up table, map and assign column names. map log/lat into the countries. Filter required columns to address research questions and use graphs for visualizations.

How your approach addresses (fully or partially) the problem.

Project approach is the address following future forcast :

Can you forecast the future daily/weekly/monthly accidents? Action that can prevent future accident based on variable relationship and predictions ? Fatal accidents can be predict or avoided ? Variables contributing rise in fatal accidents ?

Data

Four data Sources(cvs):

Accident file: main data set contains information about accident severity, weather, location, date, hour, day of week, road type. . . Vehicle file : contains information about vehicle type, vehicle model, engine size, driver sex, driver age, car age. . . Casualty file: contains information about casualty severity, age, sex social class, casualty type, pedestrian or car passenger. . . Lookup file : contains the text description of all variable code in the three files

Sources : https://www.kaggle.com/datasets/benoit72/uk-accidents-10-years-history-with-many-variables

Required Packages

Base packages plus "ggplot2", "dplyr", "magrittr", "tidyverse", "broom", "purrr", "GGally", "scales", "reshape", "moments", "ggpubr", "readxl" .. and more on need basis.

Plots and Table Needs

scatter plots, time-series plot and histograms to analyze and visualize the data patterns.

Questions for future steps

Wide data set, wrangling will be challenging to all together at one place and pick selective columns to address out research questions, In addition fear of unknown as we move forward.