Text

Description automatically generated with medium confidence

**John O’Brien**

Email:

[John.obrien@henkel.com](mailto:John.obrien@henkel.com)

[Jobrien2.irl@gmail.com](mailto:Jobrien2.irl@gmail.com)

Project Report GitHub URL (insert URL here)

**Abstract (Short overview of the entire project and features)**

The project looks at the road casualty statistics for the UK over the period of 2016 to 2020 inclusive. An analysis of 2020 data shows that the data is skewed due to the impact of Covid-19 and the national lockdown. Through much of the analysis, 2020 data was omitted.

The project combines datasets on road casualty statistics and vehicle statistics

**Introduction (Explain why you chose this project use case)**

Why did I use this data

**Dataset (Provide a description of your dataset and source. Also justify why you chose this source)**

The project uses data from gov.uk website <https://www.gov.uk/government/statistical-data-sets/> including the 2016, 2017, 2018, 2019 & 2020 road casualty statistics datasets.

Road Safety Data with vehicle statistics was sourced from [Road Safety Data - data.gov.uk](https://data.gov.uk/dataset/cb7ae6f0-4be6-4935-9277-47e5ce24a11f/road-safety-data) and was used to link road casualty statistics.

The total number of registered cars by make and model was obtained again from gov.uk. <https://www.gov.uk/government/statistical-data-sets/vehicle-licensing-statistics-data-files>

Only information on registered cars for end of quarter 2021 could be accessed from the website, and although the dataset is more recent than the casualty data, the dataset is used nonetheless to give an indication of the number of cars on the road and the comparative accidents by car make.

**Implementation Process (Describe your entire process in detail)**

Before bringing in datasets, the appropriate packages were imported to the shell. As the project developed, additional packages were imported and scripted in the same cell.

Pip install was used to install packages. Once a additional package was installed, the pip install statement was hashed - # - out

**Step one – Import data:**

The CSV files were imported with pandas function ‘pd.read\_csv( )’ to convert the CSV directly into a dataframe

BeautifulSoup package was used to scrape the uk.gov website for an additional CSV file

The first objective was to analyse car accidents as a proportion of the number of cars on the road by make. As I will be merging two datasets, to simply the analysis, all car makes will be joined based on the first name. Regex was used to extract the first car name from the Make column. the data frame was grouped by the regex output make.

2019 data was used instead of the 2020 data as there was an obvious impact of accident data due to the Covid-19 lockdown which would obscure the analysis

Chart, bar chart, histogram

Description automatically generated

The 2019 data set was used to compare the accidents by number of cars on the road as a proportion.

The make of the car was gleaned using the same regex code. The two datasets were then merged on the regex identified make name.

Chart

Description automatically generated

Chart, scatter chart

Description automatically generated