**Data Visualisation and Insight: Assignment**

**(Final Project)**

**Traffic Accident By Plotly Dash App**

By

Khubim Kumar Chhetri (d00251757@student.dkit.ie),

*Computing Science and Mathematics.*

*Dundalk Institute of Technology.*

Dundalk, Ireland.

1. **Project Summary:**

* **Introduction:**

Each year, 1.35 million people are killed on roadways around the world. Every day, almost 3,700 people are killed globally in crashes involving cars, buses, motorcycles, bicycles, trucks, or pedestrians. More than half of those killed are pedestrians, motorcyclists, or cyclists..Between 20 and 50 million more people suffer non-fatal injuries, with many incurring a disability as a result of their injury..So Plotly Dash in python use different Data Visualization techniques to interpret all the information from them. This is one of the best ways of Data representation.

Road traffic injuries cause considerable economic losses to individuals, their families, and to nations as a whole. These losses arise from the cost of treatment as well as lost productivity for those killed or disabled by their injuries, and for family members who need to take time off work or school to care for the injured. Road traffic crashes cost most countries 3% of their gross domestic product..

The main aim of this project is to create visualizations using Plotly dash Python .I have taken the accident dataset from the <https://www.gov.uk/government/collections/road-accidents-and-safety-statistics,Whichis> used to visualise using plotly.

* **Findings in the Project:**
* Traffic accident by Day and Time
* Accidents by speed limit
* County,City or Place ,District,where the Accident happened.

1. **Data Description and Wrangling:**

The Raw data consists of 12 fields Accident\_Index , Longitude, Latitude,Number\_of\_vehicles,Number-of\_casualities, Date ,Time ,Speed Limit,Day of week ,and Accident Severity

* **Programming language:** Python
* **Libraries:** Pandas, Plotly Dash
* **Data Cleaning Activities**:

1. Column names are changed to lower cases without spaces.
2. Change the time column to hours.
3. Days are sorted to in a order for simplicity

* **Data Description:**

The final dataset that was cleaned contains the following columns and comes under the labelled categories.

|  |  |  |
| --- | --- | --- |
| **Column Name** | **Category** | **Information** |
| Accident\_Index | Nominal Categorical |  |
| Longitude | Float | Location of Accident |
| Latitude | Date | Location of Accident |
| Number\_of\_Vehicles | Nominal Categorical |  |
| Number\_of\_Casualties | Nominal Categorical | Number of casualities |
| Date | Date | Accident date |
| Time | Time | Time of Accident |
| Speed\_limit | Continous Numerical | Region of Country belongs |
| Police\_Force | Nominal Categorical | District Police |
| Accident\_Severity | Nominal Categorical | Seriousness of accident |
| Day\_of\_Week | Nominal Categorical | Days of accident |
| Local\_Authority\_(District) | Nominal Categorical | County or district |

**Data Visualization Techniques:** Plotly dash Python visualisation is used to represent the following features and charateristics.so we found that the Accident by speed limit and the mapbox where the accident took place. The data used to power the visualization dashboard in this project has been gathered from various sources online such as open data archives and other services. A majority of the raw data was obtained fromthe open data portal by Transport for London, particularly the road safety which includes extracts for road

collision data. Transport for London is a local government body responsible for the transport system in Greater

London, England. The data they have collected and made available provides information for road traffic collisions that involve personal injury occurring on the public highway reported to the police. Damage only collision are not included and all data belongs to the Greater London area which comprises 32 London boroughs and the City of London. Data is collected by police at the scene of an accident or in some cases reported by a member of the public at a police station, then processed and passed by the police to Transport for London for checking and analysis.

Our approach to data engineering for the purposes of this project was to collect datasets from

various sources, each pertaining to one relevant aspect that we felt could pose as an underlying factor that

contributes to accidents in London, for example: weather, vehicle details etc. Some of these individual

datasets were joined to one another on some common dimension such as date, while others were used independently throughout the process.

**Git-Hub Repository for Codes and Dashboards:** [**https://github.com/khubim/Plotly\_dash\_app**](https://github.com/khubim/Plotly_dash_app)

**https://khubim-app.herokuapp.com/**

**Conclusion:**. We have utilized the principles of good visualization design that we learned to prepare a dashboard that can be help the relevant audience to analyse accidents in London and ask important questions such as :

“What are the factors involved in the origin of accidents that happen in London?”

Our dashboard will help individuals to gain insights about many more questions. It will allow them to access relevant and important information at a glance and perform desired analytical tasks. Hidden patterns, trends and relationships that are not inherently observable have a better chance of being highlighted through our visualization dashboard.

**References**

"Global Power City Index 2017". Institute for Urban Strategies – The Mori Memorial Foundation. Retrieved

23 November 2018.

https://www.london.ca/residents/Roads-Transportation/Pages

https://www.standard.co.uk/news/london/1220-killed-or-seriously-injured-in-london-road-accidents-overthree-

months-a4069436.html

https://en.wikipedia.org/wiki/Traffic\_collision#Economic\_costs