Data Mining Project - Spring 2022

UK Traffic Accidents Analysis

Introduction

You are given with the traffic accidents dataset of UK from 2000 to 2016. The dataset contains over 1.6 million accidents. All data is amassed from police reports and contains major accidents.

Dataset Description

Data column	Description
Year	Year in which accident occurred
CP	Count point. A unique reference for the road link
Estimation_method	Accident is counted or estimated
Estimation_method_detailed	How the accident was counted/estimated
Region	Region in which accident took place
Local Authority	Breaking down regions further
Road	Roads are either M for Motorway (the biggest roads), or A which are major roads.
RoadCategory	Roads are further categorized on the basis of class i.e. PM for Class A Principal Motorway, PR for Class A Principal Road in Rural Area, PU for Class A Principal Road in Urban
LinkLength_km	Length of road in kilometer
LinkLength_miles	Legth of road in miles
Lat	Latitude of the location where the incident took place
Lon	Longitude of the location where the incident took place
Non motorized vehicle	It includes PedalCycles
Motor vehicles	It includes Motorcycles, CarsTaxis, BusesCoaches
Motor LGV	It includes LightGoodsVehicles
All HGVs	It includes heavy good vehicles like (V2AxleRigidHGV, V3AxleRigidHGV, V4or5AxleRigidHGV,

	V3or4AxleArticHGV, V5AxleArticHGV, V6orMoreAxleArticHGV)
All Motor Vehicles	It includes motor vehicles + motor LGV + All HGVs

Deliverables

The project will have three deliverables. The details of each deliverable are as follows:

1. **Deliverable 1: EDA - Exploratory Data Analysis (40%)**
   This deliverable will be focused to get familiar with the data. It will include initial data exploration, visualizations etc. This could include but not limited to the following points:
   - Data preprocessing (identify duplicate data and remove it, find missing values, normalize the data)
   - Calculate summary statistics
   - Find correlation between attributes
   - Visualization of dataset using maps
   - Yearly analysis of accidents
   - Types of vehicle that spotted largely in accidents

   Think of as much as you can!

2. **Clustering & Frequent Pattern Mining (40%)**
   This deliverable will be more focused towards feature selection and engineering for data mining tasks. Based on selected set of features you will perform specific tasks that can extract useful insights from data. Some of the suggestions are as follows:
   - Perform frequent pattern mining to identify patterns between year and region in which accidents occurred.
   - Perform cluster analysis
   - Explain your findings
3. **Final Report (20%)**

   As a final deliverable, discuss your findings and give reasoning for the techniques which you have used at every step of the project. Explain with the figures where needed. You need to identify what meaningful information can be extracted from this data.

Dataset Link

https://drive.google.com/file/d/1JbFl6OB4QWZ7Y6Z8FEVXaoOeRBbSDP4k/view?usp=sharing