Project Description

## Name: Hamza Saleh

## Problem Statement:

Understanding the factors affecting road accidents helps a lot in avoiding the repetition of such problems and solving them in the future. In this project i will try to analyze the accidents thatsssss occurred between 2017-2020 and identify risk patterns such as road conditions, weather, and the demographic origin of individuals.

## Possible Impact of Your Analysis:

The importance of this project is to help lower the number of accidents caused by factors that can be studied and identified. This can save more lives and reduce the country’s financial costs. All of this will be possible through a clear understanding of the data.

## Dataset(s):

(Include information on where your dataset is from, the license for using the dataset, and a brief description of the variables included in your dataset.)

**Addis Ababa Road Traffic Accident Dataset (2017-2020**)

<https://www.kaggle.com/datasets/saurabhshahane/road-traffic-accidents?select=RTA+Dataset.csv>

**\* Note: The data contains two data files, the first is not cleaned and the other is cleaned. I will work with the uncleaned one.**

The dataset is derived from manual records of road traffic accidents (RTAs) that occurred between 2017 and 2020 , it is collected from the Addis Ababa Sub City police departments as part of a Master's research project.

The dataset was prepared by Bedane, Tarikwa Tesfa (2020) as part of the research titled “Road Traffic Accident Dataset of Addis Ababa City.” The dataset is available on Mendeley Data, V1, DOI: 10.17632/xytv86278f.1.

Project Scoping Document

## Business Problem

Understanding the factors affecting road accidents helps a lot in avoiding the repetition of such problems and solving them in the future. In this project i will try to analyze the accidents thatsssss occurred between 2017-2020 and identify risk patterns such as road conditions, weather, and the demographic origin of individuals.

## Business Impact

The importance of this project is to help lower the number of accidents caused by factors that can be studied and identified. This can save more lives and reduce the country’s financial costs. All of this will be possible through a clear understanding of the data.

## Dataset(s)

**Addis Ababa Road Traffic Accident Dataset (2017-2020)**

**Strengths:**

- Contains over 12,000 records of traffic accidents in Addis Ababa between 2017 and 2020.

- Covers 32 features, including key factors like road conditions, weather, driver demographics, and accident severity.

- Allows for comprehensive analysis of accident causes and identification of risk patterns.

**Weaknesses:**

- Limited to accidents in Addis Ababa, reducing its applicability to other regions.

- Based on manual records, which may introduce inconsistencies or errors.

- Sensitive information has been excluded (such as the exact personal history of drivers) , which may limit certain types of analysis .

- Some fields may contain missing values, affecting data completeness.

## Methods

**Variables and Comparisons**

* Age band of driver
* Gender of driver
* Driving experience
* Weather conditions
* Road surface conditions
* Type of vehicle
* Time
* Day of week
* Cause of accident

**Each of these will be compared with Accident severity.**

## Dashboard

We will display accident severity across variables like age, gender, weather conditions, time, and vehicle type, allowing for interactive filtering and analysis

## Milestones

* Perform exploratory data analysis (EDA)
* Clean and wrangle data
* Identify key factors and relationships between variables
* Create visualizations
* Build a dashboard with interactive features
* Write final report and conclusions

## Timeline

|  |  |
| --- | --- |
| Week | Tasks |
| Week 1 | Data cleaning and wrangling |
| Week 2 | Identify key patterns and relationships  Create initial visualizations |
| Week 3 | Build and refine the dashboard  Analyze results and adjust visualizations |
| Week 4 | Write final report and conclusions |