**PROJECT PLAN**

Project Name: An Analysis of the Relationship between Weather and Environment factors with Motor Vehicle Accidents

**Project Lead:** Jemima Joaquina Dias

# Start Date: 1st Jan 2023

**Introduction:**

The purpose of this report is to analyze the crash data from New York City's Open Data website. The data used is from July 2012 onward for the five boroughs of NYC: Brooklyn, Bronx, Manhattan, Queens, and Staten Island. The aim is to identify patterns and trends in the data, and draw insights that can help in improving road safety in the city.

# **Scope and complexity estimate:**

The scope of the project is extensive, as it involves analyzing crash data from New York City's Open Data website for the five boroughs of NYC: Brooklyn, Bronx, Manhattan, Queens, and Staten Island. The data was collected over a period of several years and covered a wide range of variables, such as land area, time of day, day of the week, type of vehicle, injuries, and deaths. The complexity of the project is also high, as it involves a detailed analysis of large volumes of data, using statistical and visualization techniques to identify patterns and trends. The project requires expertise in data analysis and visualization, as well as knowledge of road safety policies and regulations. Additionally, the project may involve collaborating with stakeholders such as the NYPD and city authorities to gain a deeper understanding of the data and its implications.

# Project time estimate:

2 month project scheduled over an eight-hours per week

# Resources estimate:

The resources needed to complete the project in a timely manner are listed.

* 1. High-performance computers with sufficient storage and memory capacity
  2. Visualization and graphics software, such as PowerBI or AMCharts
  3. Microsoft Office Suite for documentation and reporting
  4. Project management software.

**Milestone schedule**

|  |  |  |  |
| --- | --- | --- | --- |
| **Milestones** | **Responsibility** | **Start Date** | **End Date** |
| Extract Weather Data | Jemima, Nalin | 01/01/2023 | 07/01/2023 |
| Extract Crash Data | Nainesh, Nalin | 07/01/2023 | 14/01/2023 |
| EDA Weather Data | Jemima, Kajal | 14/01/2023 | 28/01/2023 |
| EDA Crash Data | Nalin,Kamal | 28/01/2023 | 04/02/2023 |
| Merge Data | Nainesh, Jemima | 04/02/2023 | 11/01/2023 |
| Create MongoDB Cluster | Nainesh, Kamal, Jemima,Kajal,Nalin | 11/02/2023 | 28/02/2023 |
| Push data into MongoDB | Nainesh, Kamal, Jemima,Kajal,Nalin | 18/02/2023 | 25/02/2023 |
| Pull data from MongoDB | Nainesh, Kamal,Jemima,Kajal,Nalin | 18/02/2023 | 25/02/2023 |
| Applying ML techniques | Kajal, Nainesh | 25/02/2023 | 06/01/2023 |
| Creating front-end | Nalin, Kamal | 25/02/2023 | 06/03/2023 |
| Documentation | Kamal | 25/02/2023 | 06/03/2023 |

# Roles and responsibilities

Nainesh: Pyspark, ML, MongoDB

Kajal: ML, MongoDB

Kamal: ML, MongoDB, Visualization, Documentation

Jemima: ML, MongoDB

Nalin: Visualization, Extraction, MongoDB

# Production plan

Creating a python virtual environment with libraries like pymongo, pyspark, jupyter-lab, sklearn.

Create MongoDB Atlas cluster for storing data.

Use Machine learning techniques for classification problem

For visualization create a frontend using html and powerBI.

# Revision process

Tweak and hyper tune ML techniques parameters to get best outcome.

# Testing plans:

* Data verification: Checking the integrity of the data used in the analysis, such as verifying that all data points are correctly labeled, and that there are no inconsistencies or errors in the data.
* Data validation: Checking the validity of the data used in the analysis, such as ensuring that the data comes from a reliable source and that it represents an accurate picture of the crash patterns and trends in New York City.
* Software testing: This would involve testing the visualization software used to create the reports, such as testing for bugs or glitches that could affect the accuracy or usability of the reports.
* User testing: This would involve testing the usability and functionality of the reports by having users interact with them and provide feedback on their effectiveness.
* Overall, the testing plans for this project would be focused on ensuring that the data and analysis are reliable, accurate, and accessible to stakeholders who may use the insights gained from the analysis to improve road safety in New York City.