**Traffic Incident Analysis in Calgary**

Project Overview

This project aims to perform an extensive analysis of traffic incidents in Calgary, leveraging a dataset obtained from the City of Calgary. The dataset contains details such as incident descriptions, timestamps, geographic coordinates, and quadrants within the city. The primary goal is to extract meaningful insights, identify patterns, and provide visualizations to better understand traffic trends and incidents.

Dataset Description

* Dataset Source: City of Calgary Open Data
* Total Rows: 51,756
* Columns: 10

Key Columns

1. INCIDENT INFO - Brief information about the incident's location.
2. DESCRIPTION - Detailed description of the incident.
3. START\_DT - Timestamp indicating when the incident started.
4. MODIFIED\_DT - Last updated timestamp.
5. QUADRANT - Quadrant (NE, NW, SE, SW) of Calgary where the incident occurred.
6. Longitude & Latitude - Geographic coordinates for incident location.
7. Count - Number of incidents recorded (always 1 per row).
8. Point - Geospatial representation of longitude and latitude.

Project Tasks and Goals

1. Data Cleaning and Preprocessing

* Handle missing values in the DESCRIPTION, MODIFIED\_DT, and QUADRANT columns.
* Convert START\_DT and MODIFIED\_DT into datetime format.
* Extract additional time-based features such as day, hour, and weekday.
* Verify and correct any inconsistencies in geographic coordinates.

2. Exploratory Data Analysis (EDA)

* Generate summary statistics for all numerical and categorical columns.
* Analyze incident frequency by:
  + Time of day
  + Day of the week
  + Quadrants
  + Incident type (based on descriptions)
* Visualize trends over time (daily, weekly, monthly patterns).
* Create bar plots and pie charts for categorical data distribution.

3. Geospatial Analysis

* Map incident locations using Folium or Plotly.
* Create heatmaps to identify high-density areas for incidents.
* Cluster incident locations using techniques like KMeans to group hotspots.
* Analyze incidents in proximity to major intersections and highways.

4. Pattern and Trend Identification

* Examine seasonality and recurring patterns in incident data.
* Identify relationships between incident types and locations.
* Investigate patterns related to time (e.g., peak hours, weekends vs. weekdays).

5. Statistical Analysis

* Test correlations between variables such as incident type, time, and location.
* Perform hypothesis testing to detect significant trends.
* Calculate probabilities for specific incidents based on location and time.

6. Predictive Modeling (Optional)

* Apply KMeans clustering for incident grouping.
* Develop a time-series forecasting model to predict future incidents based on historical trends.
* Test classification algorithms (e.g., Logistic Regression, Decision Trees) to predict incident severity.

7. Reporting and Visualization

* Generate interactive dashboards using Plotly Dash or Tableau Public.
* Summarize findings in a report highlighting:
  + Key patterns and trends.
  + High-risk areas and peak times for incidents.
  + Actionable insights for city planning and traffic management.

Long-Term Goals

* Continue improving the analysis by integrating additional datasets, such as weather data and road conditions.
* Implement machine learning algorithms for anomaly detection in traffic patterns.
* Explore integration with real-time data feeds to create dynamic visualizations.
* Document findings, challenges, and progress in the GitHub repository for continuous updates.

Timeline

* Week 1: Data cleaning, preprocessing, and initial visualizations.
* Week 2: Exploratory data analysis and pattern identification.
* Week 3: Geospatial analysis and statistical testing.
* Week 4: Advanced modeling and dashboard creation.
* Ongoing: Iterative improvements, documentation, and publishing insights.

This document will act as a reference for tracking progress and defining tasks throughout the project.