Vehicle Crash Analysis

1 Overview

The Vehicle Crash Analysis Application performs various analytics on vehicle crash data to derive insights related to accidents, vehicle makes, driver demographics, and contributing factors. The analysis is modular, leveraging Spark DataFrame APIs, and is configured to read input and output file paths from YAML configuration files.

2 Project Structure

2.1 Configuration

- **Input Config:** Contains paths to input CSV files.
- **Output Config:** Defines output paths for analysis results and visualizations.
- **Data Directory:** Contains the raw CSV files, such as `charges.csv`, `damages.csv`, etc.
- **Output Directory:** Organized into subfolders for each analysis and visualizations.

2.2 Source Code

- **Initialization:** `src/__init__.py`
- **Configuration Reader:** Reads configuration files.
- **Data Cleaning:** Functions to clean input data.
- **Crash Analysis:** Contains various analytical functions for processing the crash data.
- **Visualizations:** Functions for generating and saving visualizations.
- **Main Execution Script:** Orchestrates data loading, cleaning, analysis, and result saving.

3 Key Analytics Tasks

- Count of Crashes with Males Killed: Analyzes incidents where more than two males were killed.
- Two-Wheelers Involved: Counts the number of motorcycle crashes.
- Top Vehicle Makes: Identifies vehicle makes involved in fatalities where airbags did not deploy.
- Hit and Run Valid Licenses: Counts vehicles involved in hit-and-run incidents with valid driver licenses.
- State with Highest Accidents: Determines the state with the most accidents excluding females.
- Vehicle Makes for Injuries: Analyzes vehicle makes contributing to injuries and fatalities.
- Top Ethnic Group by Body Style: Identifies dominant ethnic groups for each vehicle body style involved in crashes.
- Top Zip Codes for Alcohol-Related Crashes: Analyzes zip codes with the highest incidents where alcohol was a factor.
- Distinct Crash IDs with High Damage: Counts crash IDs where no damage was observed but damage level is above 4.
- Vehicle Makes Charged with Speeding: Identifies top vehicle makes involved in speeding offenses with specific conditions.

4 How to Run

Run the main script using the following command:

```
python main.py
```

5 Setup

1. **Clone the repository:**

   ```
   git clone <repository-url>
   cd <repository-directory>
   ```

2. **Install required packages:**

   ```
   pip install -r requirements.txt
   ```

6 Installation Instructions

Ensure you have Python installed on your machine. You can download it from `https://www.python.org/downloads/`.

6.1 Required Packages

Make sure to install the required packages listed in `requirements.txt` for the project to function properly. The list includes libraries for data manipulation, analysis, and visualization, such as:

- Pandas
- NumPy
- Matplotlib
- Seaborn