Project Plan: Analysis and Visualization of Auto Accident Data

Determine Most Common Causes and any Correlation between Marijuana Legalization and State Alcoholism Reporting

1. Introduction

* Objective: Develop a visually appealing and technically sound project analyzing the most common causes of auto accidents. I will also compare state marijuana legalization data and alcoholism levels to review any correlation with the data
* Tools and Technologies: Pandas, SQLite, Tableau, Jupyter Notebook, Urllib, OS

1. Data Acquisition and Cleaning

* Loading Data
  + Read data file: Auto Accident Data 2016 – 2023 (.csv) and Alcoholism dataset (dataset still to be found)
  + Web Scraping: Scrape marijuana legalization timeframes from Wikipedia
* Data Cleaning and Merging
  + Clean datasets using Pandas
  + Convert data to databases and merge with SQL utilizing common attributes

1. Data Analysis and Visualization

* Develop a Tableau dashboard to present key insights

1. Best Practices Implementations

* Virtual Environment
  + Set up a virtual environment and provide setup instructions in the README
* Data Dictionary
  + Set up a dictionary to breakdown the various datapoints within the auto accident dataset

1. Data Interpretation and Documentation

* Code Annotation
  + Use markdown cells in Jupyter Notebooks to annotate code and explain the steps taken
  + Ensure clear and concise comments within code
* Project Documentation
  + Write a comprehensive README.md file covering:
    - Project Overview and Objectives
    - Data sources and cleaning steps
    - Analysis and visualization methods
    - Instructions for setting up the virtual environment and running the project
    - Summary of findings and interpretations

1. Review and Polishing

* Internal Review
  + Ensure accuracy and readability of visual components
  + Ensure project is fully functional and well-documented
* External Feedback
  + Present the project to peers or mentors for feedback
  + Incorporate feedback to improve project