|  |  |
| --- | --- |
| **NAME:** | Nathan Cardoso |
| **UID No.** | 2021600009 |
| **BRANCH:** | B.E CSE-AIML |
| **BATCH:** | D |
| **SUBJECT** | Advanced Data Visualization |
| **EXPERIMENT No.** | 7 |
| **DATE:** | 20/10/2024 |

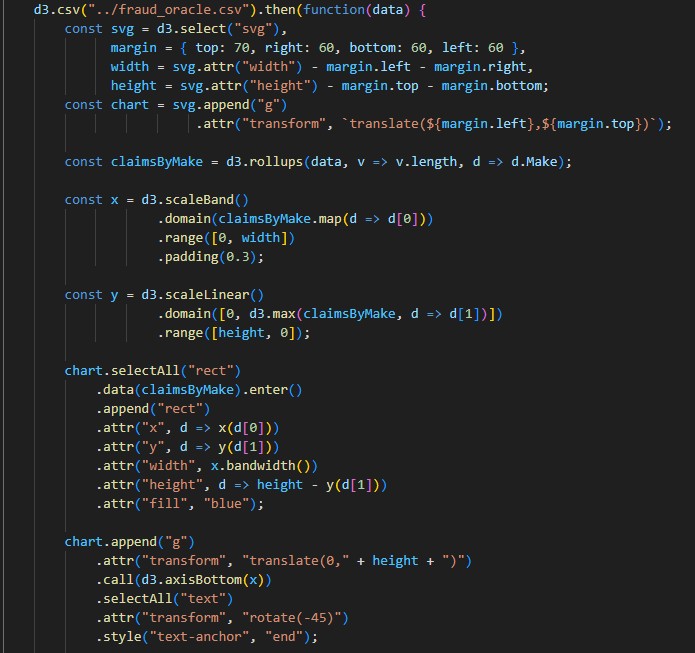
**AIM:** Experiment Design for Creating Visualizations using D3.js on a Finance Dataset

**DATASET:** <https://www.kaggle.com/datasets/shivamb/vehicle-claim-fraud-detection>

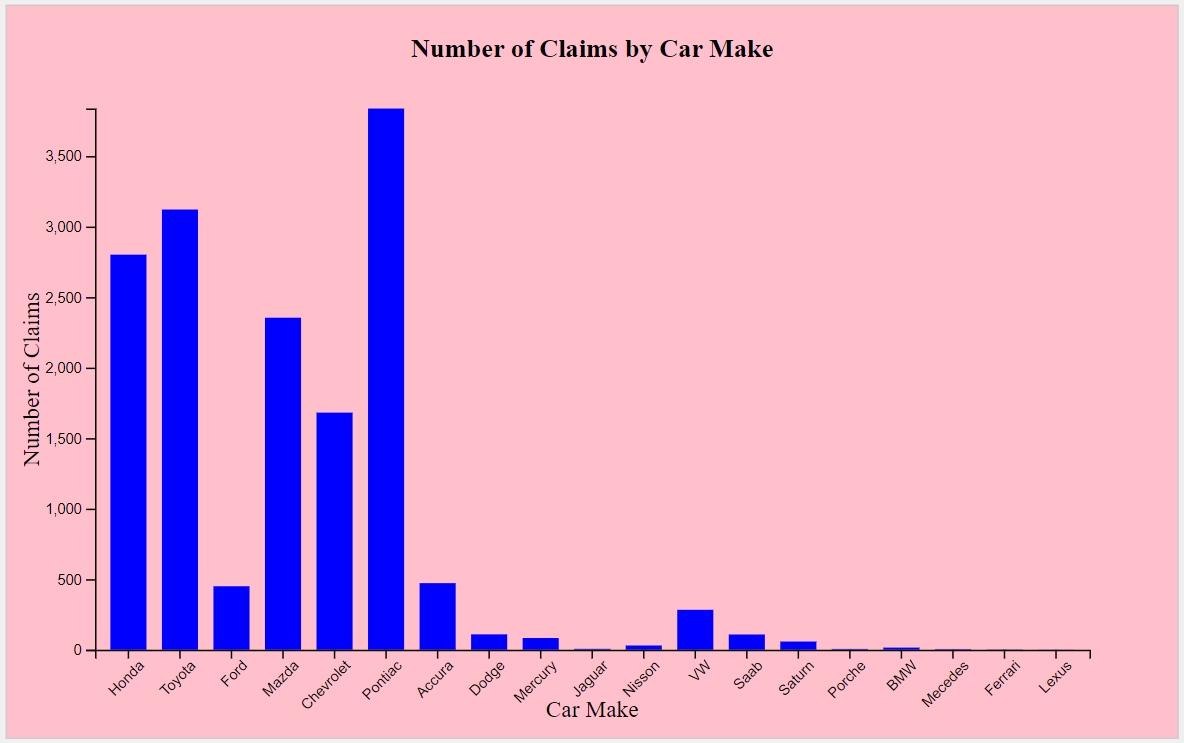
The dataset used here contains details about insurance claims, including columns like Age, Sex, VehiclePrice, and PolicyType, among others, with the aim of detecting potential fraud. It provides valuable insights into patterns of claims based on customer demographics, vehicle characteristics, and policy details.

# ANALYSIS:

1. Bar Graph

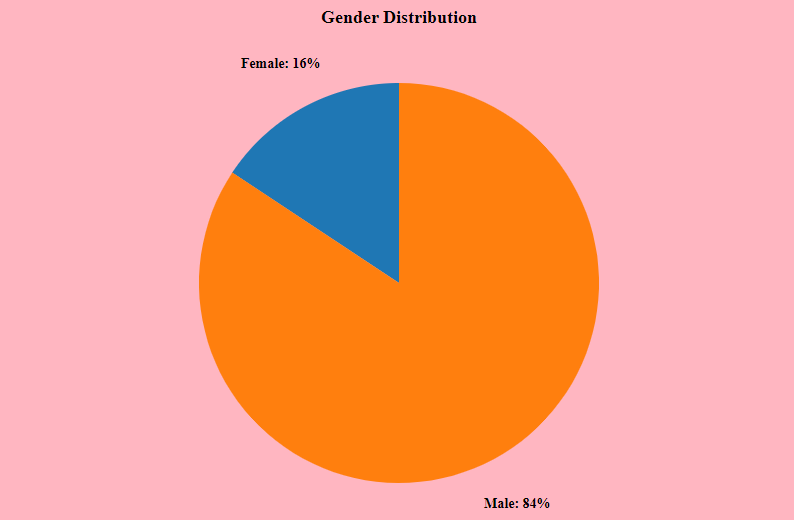






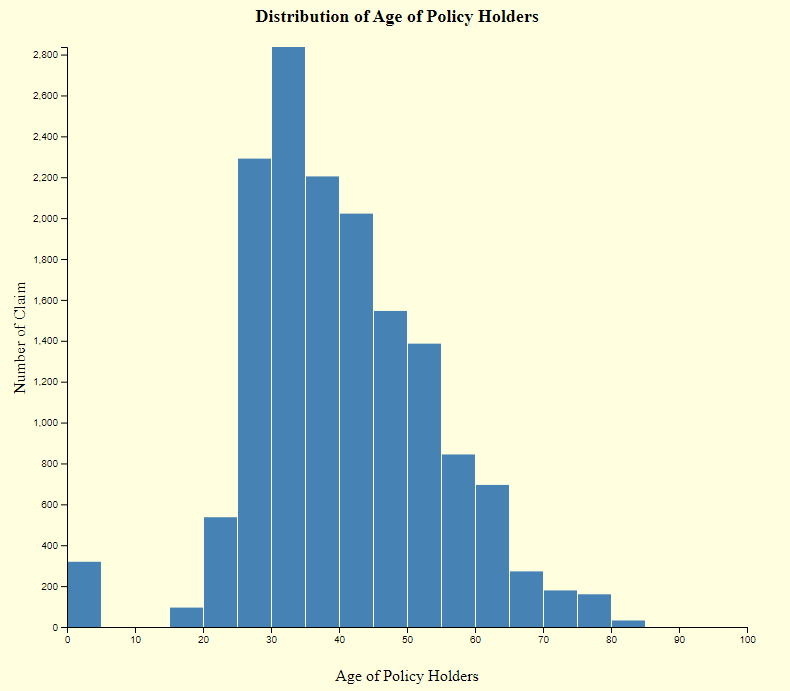
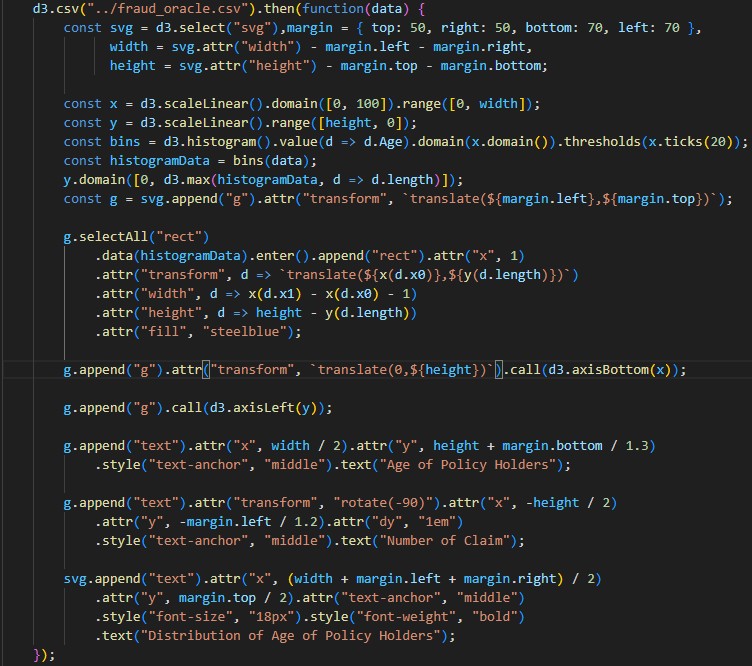
The above Bar Graph shows the Number of Policy Claims as per the Car Model(Make). The above graph shows that the car make Pontiac has resulted to higher number of policy claims.

1. Pie Graph



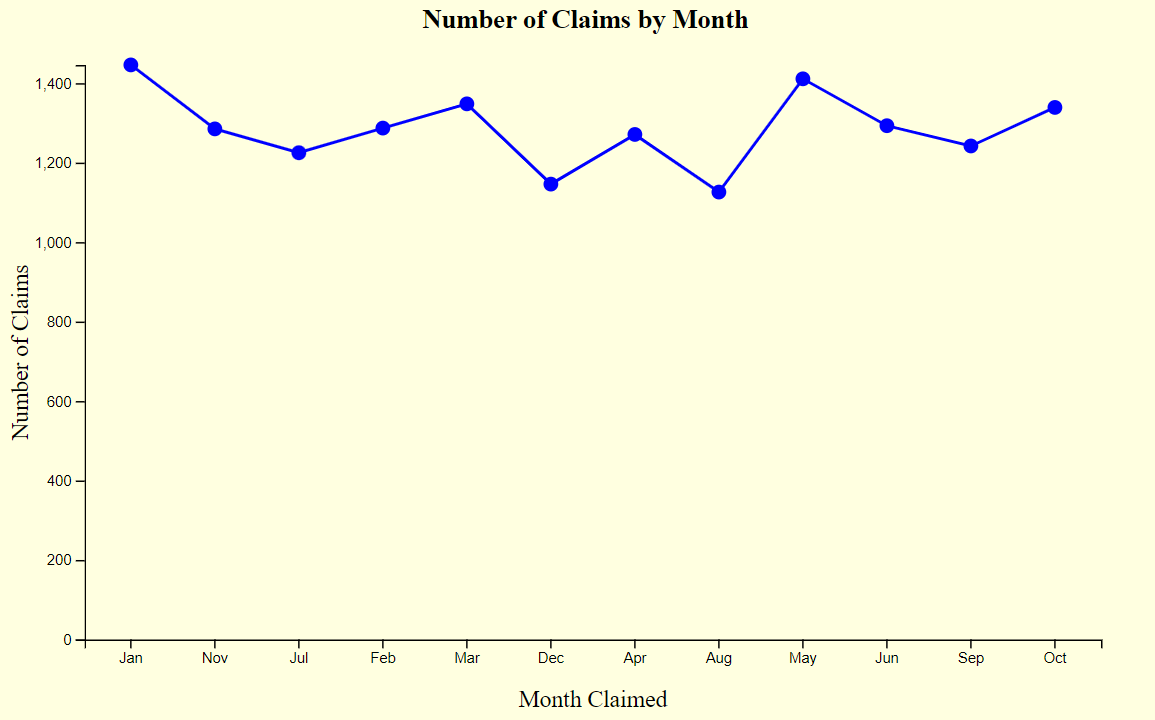
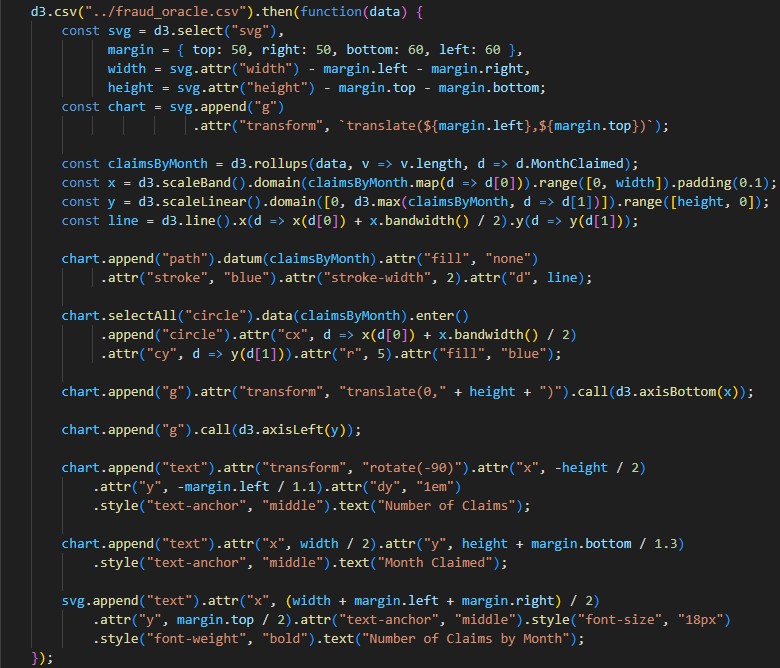
The above pie chart shows the distribution of policy claims according to genders. The chart states that males drivers have a higher number of claims than female.

1. Histogram



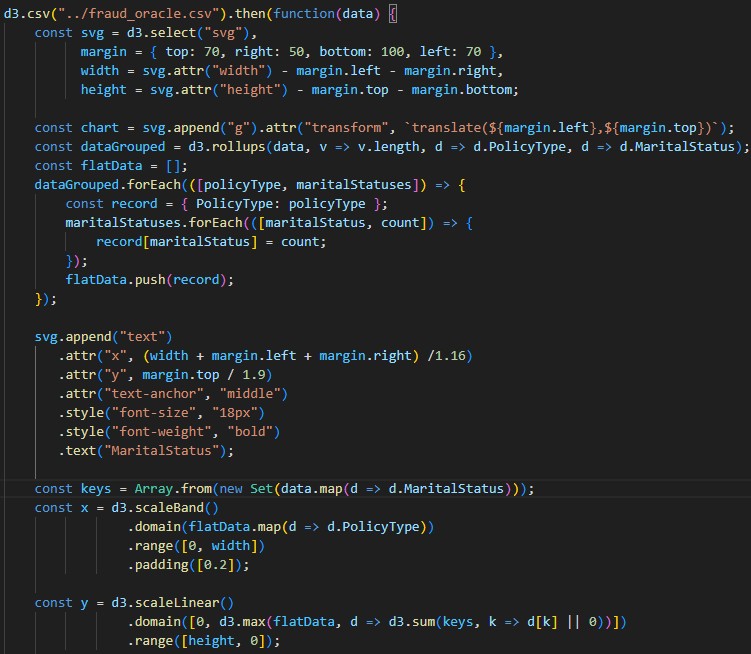
The above Histogram shows the distributions of policy claims as per age. The median comes to around 38 and it has the highest number of claims.

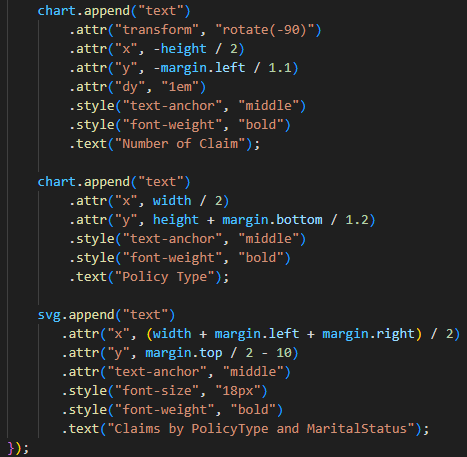
1. Line Graph

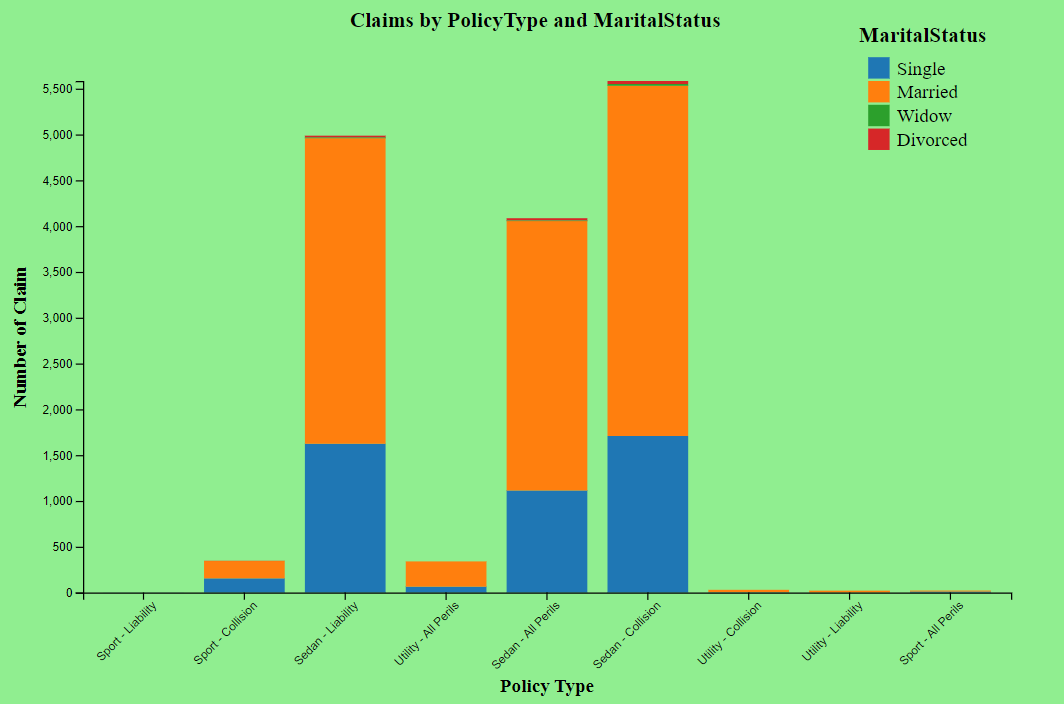


The above Timeline Graph shows the number of claims as per the months. The number is close for all the months but is seen to be higher in the month of January and May.

1. Stacked Bar Graph

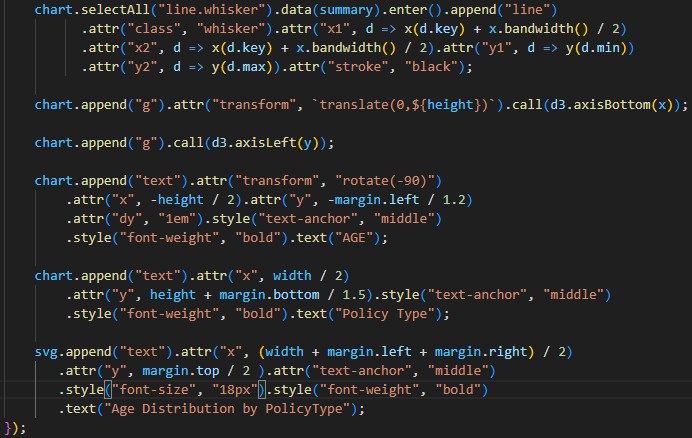
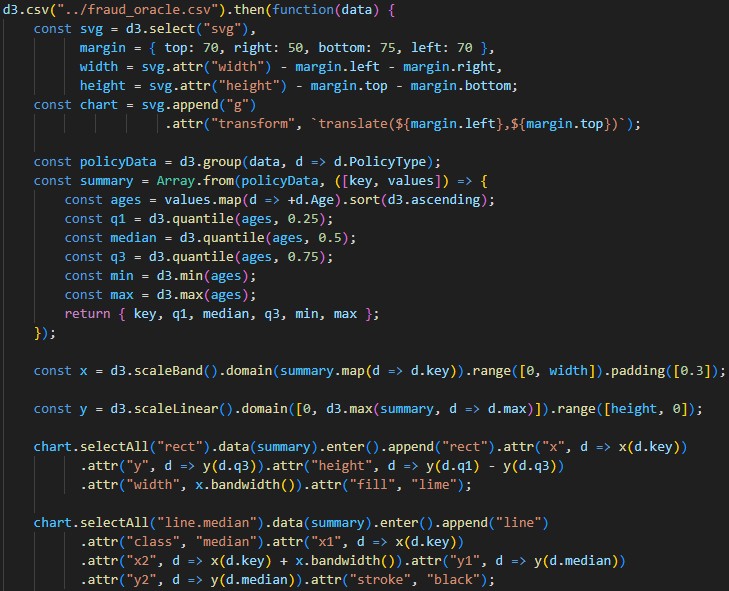


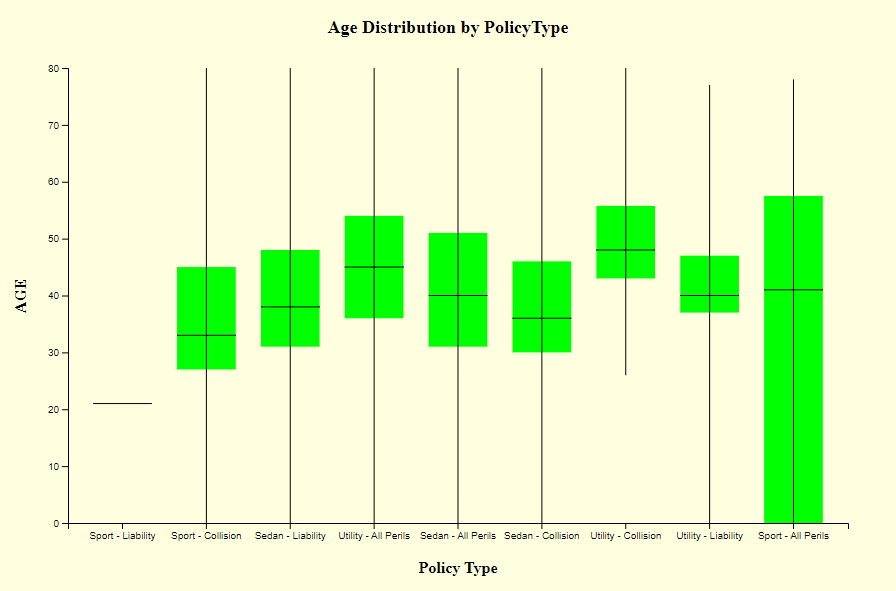




The above Staked Bar Graph is shows us the distribution of number of claims as per policy type and marital status. The above graph shows that the maximum number of claims are filed by married people and the least by widows. The most common claim policy types are Sedan – Liability and Sedan – Collision.

1. Box and Whisker Graph





The above Box and Wisker Graph shows distribution of policy claims as per policy type and age. It can be seen that Sport – All Perils has group of people from almost all the ages and Sport – Liability does not have that much crowd.

1. Word Cloud

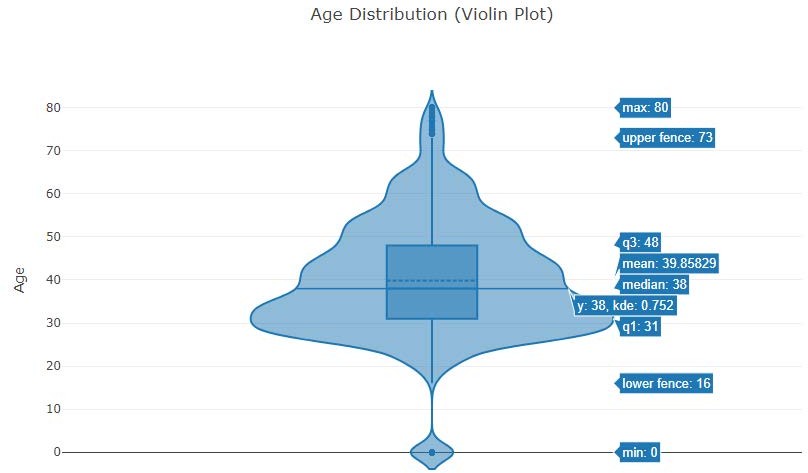




The above word cloud shows the frequency of policy claims as per the vehicle category. Sedan has the highest number of claims and Utility has the lowest number of claims.

1. Violin Plot

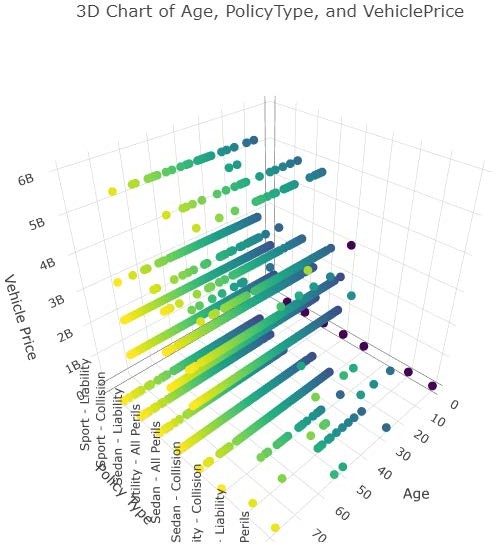




The above Violin Graph shows distribution as per age. It shows that the maximum number of claims is around the age of 30, 38 being the median, 0 being the lowest and 80 being the highest age of claim.

1. 3D Chart

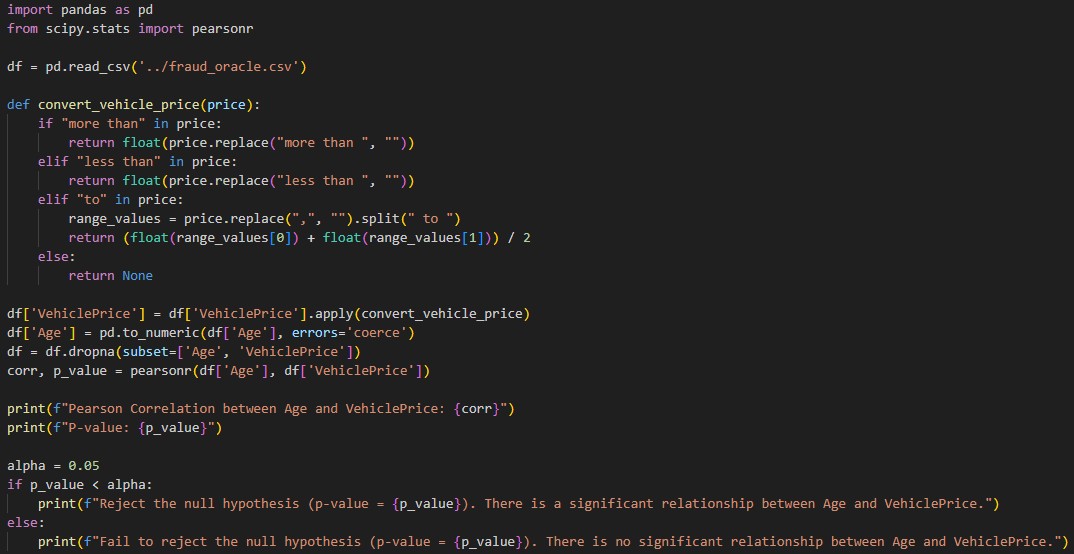




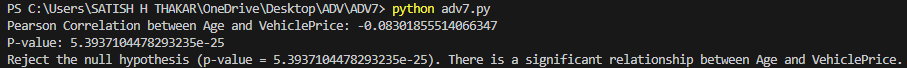
The above 3D Graph shows the claims as per age, policy and vehicle price. It can be seen that most cars are in the range from 0 to 3B. There are few entries for age 0 that states it is a part or outlier.

# HYPOTHESIS TESTING:

To perform hypothesis testing using the Pearson correlation coefficient to evaluate relationships between numerical variables in the dataset.



# OUTPUT:



The above relation states that there is a weak negative correlation between Age and Vehicle Price which is then supported using the hypothesis testing as we reject the null hypothesis thus stating that there is a significant relation between Age and Vehicle Price.