Pre-owned Cars Dashboard

Toshitt Ahuja
Rutgers University
Piscataway, NJ, USA
toshitt.ahuja@rutgers.edu

Abhinay Mannepalli
Rutgers University
Piscataway, NJ, USA
am2977@scarletmail.rutgers.edu

Pavankumar Prakash Savanur
Rutgers University
Piscataway, NJ, USA
ps1149@scarletmail.rutgers.edu

Abstract— We have worked on building a multiple featured, multilevel, highly interactive dashboard that allows users to demonstrate and get an idea of car prices in different states of the US. The data[1] was scraped from Craigslist and was open-sourced at Kaggle. We have visualized the features of data through various plots and maps provided by Plotly[2] and Dash[3], python libraries based on concept of visualization.

Keywords— Plotly, Dash, Choropleth Maps

I. PROJECT DESCRIPTION

Searching for the perfect car, is quite a task. To strike a cordial balance between the price and requirements is a herculean task and one feels lost in the vast sea of resources in which it is almost impossible to identify between dubious and the decisive resources. In the quest to search for a panacea, we are set to build a dashboard where users can search, interact and lookup statistics to make an informed decision to buy a used car.

The purpose of this project is establish a one spot shop to find the pertinent information on used cars for example, price, year, state, odometer, manufacturer etc. With the help of this dashboard, the intended users are able to get a better idea about which car would be the perfect fit for a given budget and desired requirements. Our efforts to cover all angles of visualizing the data has been done with help of different plots and maps ranging from Choropleth Maps to Bar Charts and Pie Charts.

A. Stage1 - The requirements gathering stage.

Our dataset is taken from vehicles listing from Craigslist.com which is in kaggle. Craigslist is the world's most extensive collection of used vehicles for sale, yet it's challenging to collect all of them in the same place. A scraper is built on it to create this dataset which includes every used vehicle entry within United States on Craiglist. This dataset contains all the relevant information that Craiglist provides on car sales. The dataset is 1.45GB, and it has over 400k records. In this project, we mainly focus on 11 columns out of 23 columns, namely region, manufacturer, cylinders, fuel, odometer, price, transmission, type, state, latitude, longitude. We cleaned the data by removing NaN value columns, and after cleaning the data, our dataset size turns to be around 400k records.

The knowledge about the dataset is provided in Kaggle. We learnt more about the columns of the dataset when we started the implementation phase. This application is designed for:

- **Users:** The intended users for the dashboard are Buyers and Sellers, Analysts in Car Manufacturing companies, and it also acts like Used cars seller platforms like Cargurus, Autograder. Buyers can interact with the specific feature based on their personal choice and buy the cars accordingly, comparing local listings based on used vehicles. Sellers can interact with the dashboard to know how much there can be sold based on the features on the dashboard(for example, region). Analysts of Car Manufacturing companies can look at the features that most customers are looking for.

Scenario 1: We render an interactive Choropleth Map to users where they can search for a car, and then they can hover on the state and find the features of the car. There are filters for the choropleth map where the user can the cars based on these filters. We have also created an extension of Choropleth, called Advanced, which instead of using filters will allow the users to insert SQL queries to generate choropleth maps of the required findings.

- This smart interactive way can save them a lot of time for the users. Another advantage is that users can select only required features which helps in the reducing time to find targeted cars. This enables the user to buy the right fit car given their constraints.
- The same goes with the advanced page which will allow the user to exploit the power of SQL to carry out the necessary visual analytics, making the process effective manifold.

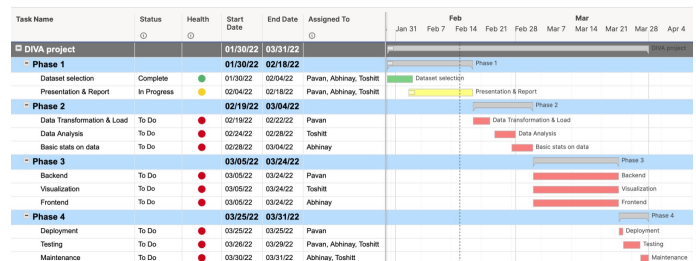- Project Timeline and Division of Labor.

Fig. 1. 8 weeks Gantt chart that was followed to complete the project

Toshitt Ahuja: Data Analysis,Visualization,Testing, Maintenance
Abhinay Mannepalli: Statistical Data Analysis, Front-end, Testing, Maintenance
Pavankumar Prakash Savanur: Data Transformation and load, Back-end, Deployment

B. Stage2 - The design stage.

- In our web application, we have 4 web pages, i.e., Statistics, Plots, Dashboard and Advanced Dashboard. The first page is the Statistics page, where the user can have a peek of the statistics and trends in the data, based on the car manufacturers, car prices, number of cars sold etc.
- The second page is plots, where we have different plots, i.e., Barchart, Line Chart, Pie Chart.The line chart helps visualizing 1. Distribution of car sales 2. Distribution of cars based on fuel type. Bar Chart helps visualizing among two variations 1. Manufacturer vs mean prize of cars 2. Number of cars sold in

a year. The Line Chart visualizes only one variation i.e.increase in yearly prices in a particular state.

- The third page is the Dashboard, followed by a fourth page named Dashboard Advanced (an extension of the Dashboard) briefly described in scenario 1 in stage 1.
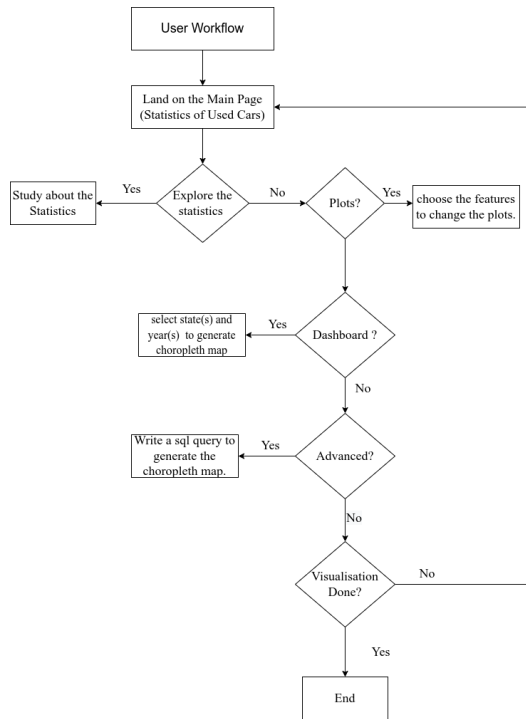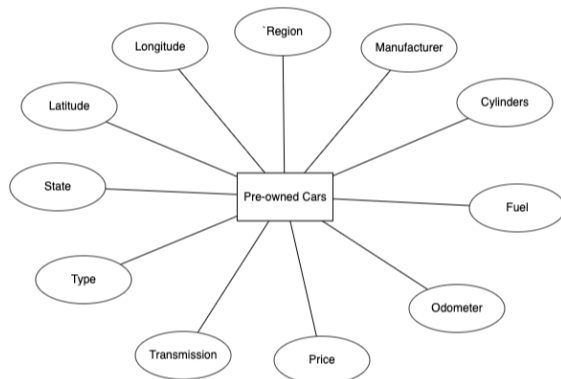- Data Flow Diagram:

Fig. 2. App flow diagram

Fig. 3. ER Diagram.

C. Stage3 - The implementation stage.

- This web application primarily uses client-server architecture with MVT (Model, View, Template) design pattern.
- Raw dataset had multiple csv files with over 400,000 records and all the data has been stored in local storage itself.
- View in the design pattern is implemented as Python functions that serve the calls from client. Templates are the web pages

that are viewed by the client. Dash Components are used for coding the template.
- Dash[3] and Plotly[2] with Python is used for the construction of the UI in this project.

The steps involved in the implementation are as follows:
- Downloaded the dataset from Kaggle.
- Cleaned the data to remove non-important columns and NaN values.
- Performed the exploratory data analysis on this data to understand the relationship between various features like Price, Odometer, Fuel, etc.
- Render choropleth maps, statewise and nationwide statistics on request.
- Implemented a text editor which allows the user to enter SQL queries which on pressing submit generates a Choropleth map displaying the results obtained from the query.

D. Stage4 - User interface.

The User Interface were developed with Dash and Plotly. We performed Exploratory Data Analysis to find the relationship between various features with respect to Price and are displaying interesting statistics along with interactive graphs on request.

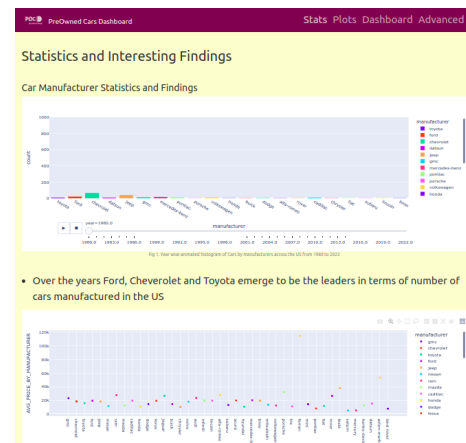The following are some screenshots from the different pages in our web application.

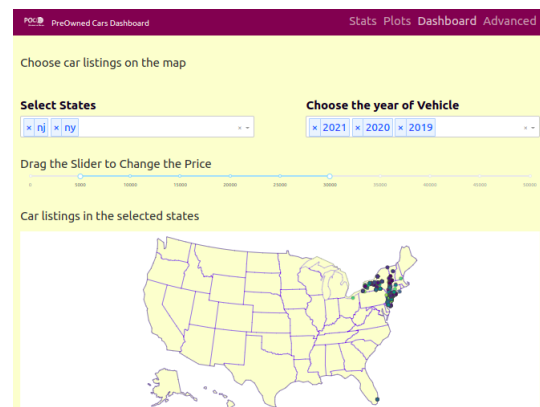Fig. 4. Home page with various statistics and interesting findings.
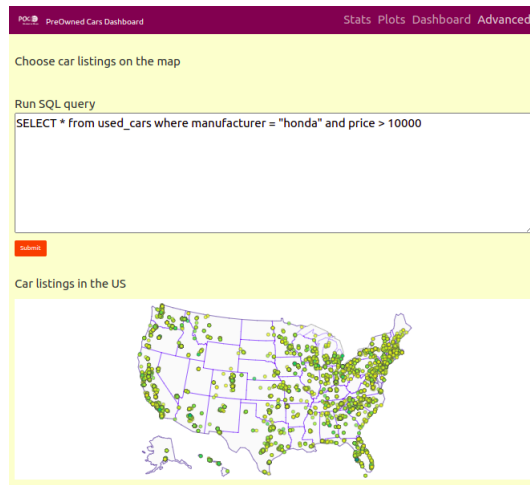

Fig. 5. Choropleth Map.

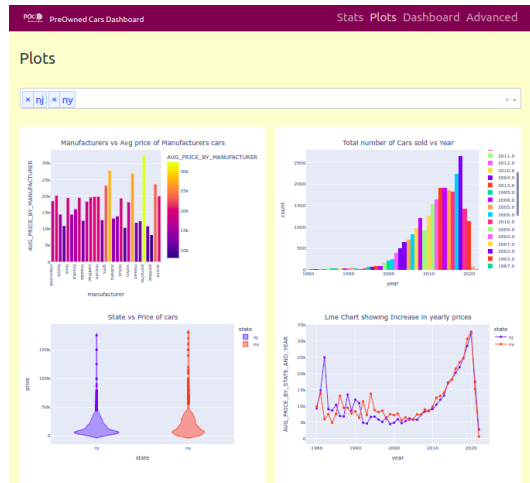Fig. 6. Advanced:- SQL query based Choropleth map.

Fig. 7. Various statewide and nationwide plots.

II. PROJECT HIGHLIGHTS.

Our dashboard has been able to churn out insightful facts and insights from the given dataset and are mentioned as follows.

1) Our dataset is based on the entries made by different users over the years on Craigslist. Our dataset is about 1.46 GB in size and has over 23 features.

2) We implemented a page based on statistics, which acts a platform to serve information on the go. The user can find the desired information on the feature they require without the hassle of visualizing the entire feature.

3) We implemented different pages in our web application so the user feels comfortable in visualizing the data in which ever way that suits them the most. For example, we have made use of choropleth maps to give out a visually aesthetic and easy to comprehend method to visualize and compare different features of certain cars all across the US.

4) We have sketched out the relationships between different features and how they affect one another. We have been able to deduce a price trend based on important features. We also have tried to crack the relation between how price and requirements vary from state to state.

5) We were able to bring out various insights and interesting findings from our data:

   a) Ford, Toyota and Chevrolet are leaders in number of cars sold in the US.

   b) In terms of average car price, Aston Martin, Kia and Tesla are the top 3 car manufacturers.

   c) We found out that gas driven cars are the most popular cars across the US, capturing 84.4 percent of the market share and also the carries the lowest average car price amongst all fuel types.

   d) Electric and Diesel driven cars hold the highest average car prices among all the fuel types.

   e) Looking at the year-wise trends the number of cars sold in a year have gone up from 1980s to 2020, with the exception of the year 2009.

   f) Considering year-wise trends, the average price of a car has gone up with exception of early 1990s to 2000, 2008-09 and 2021.

III. FUTURE WORK.

- An ML based recommender system, which will show suggestions based on the requirements and price entered by the user.
- A compare feature between the two recommended cars by the recommendation engine or manually selected cars.

REFERENCES

[1] A. Reese. (2019) Used car dataset. [Online]. Available: https://www.kaggle.com/austinreese/craigslist-carstrucks-data

[2] A. J. Jack Parmer Chris Parmer and M. Sundquist. (2012) Plotly. [Online]. Available: https://github.com/plotly

[3] Alex Johnson Jack Parmer Chris Parmer and M. Sundquist. (2015) Dash by plotly. [Online]. Available: https://github.com/plotly/dash