

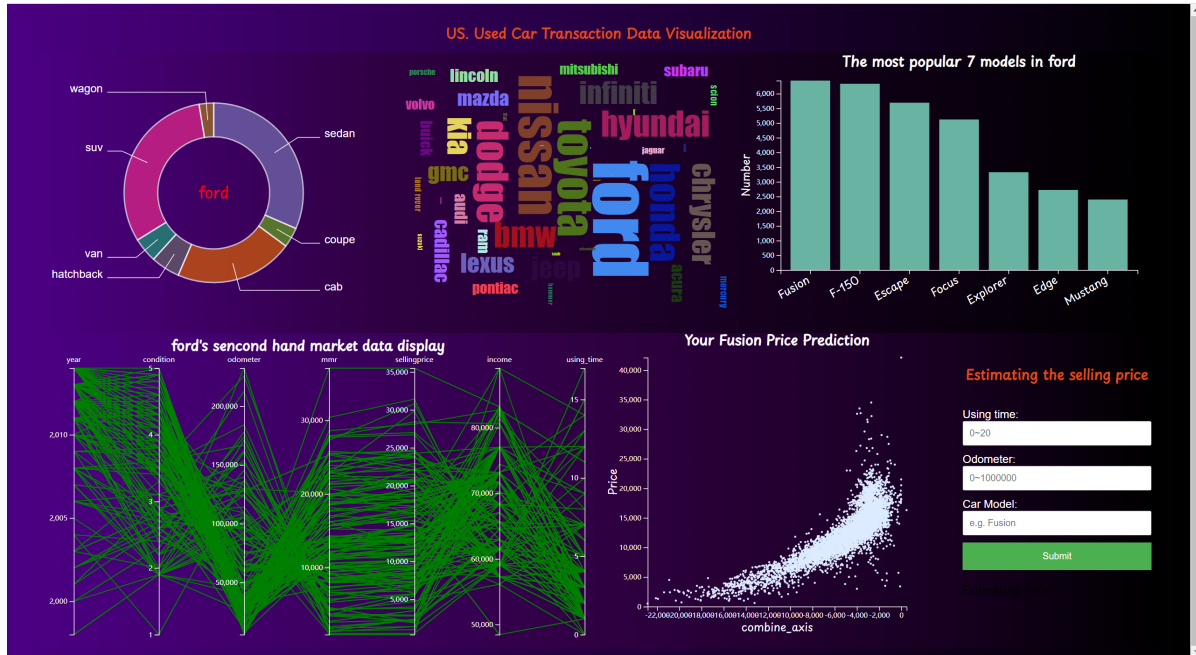
CSE 332 Lab 5: Project Report

Qijia He

116010713

01 Overview

1.1 The Dashboard



This dashboard comprises **5 main components**:

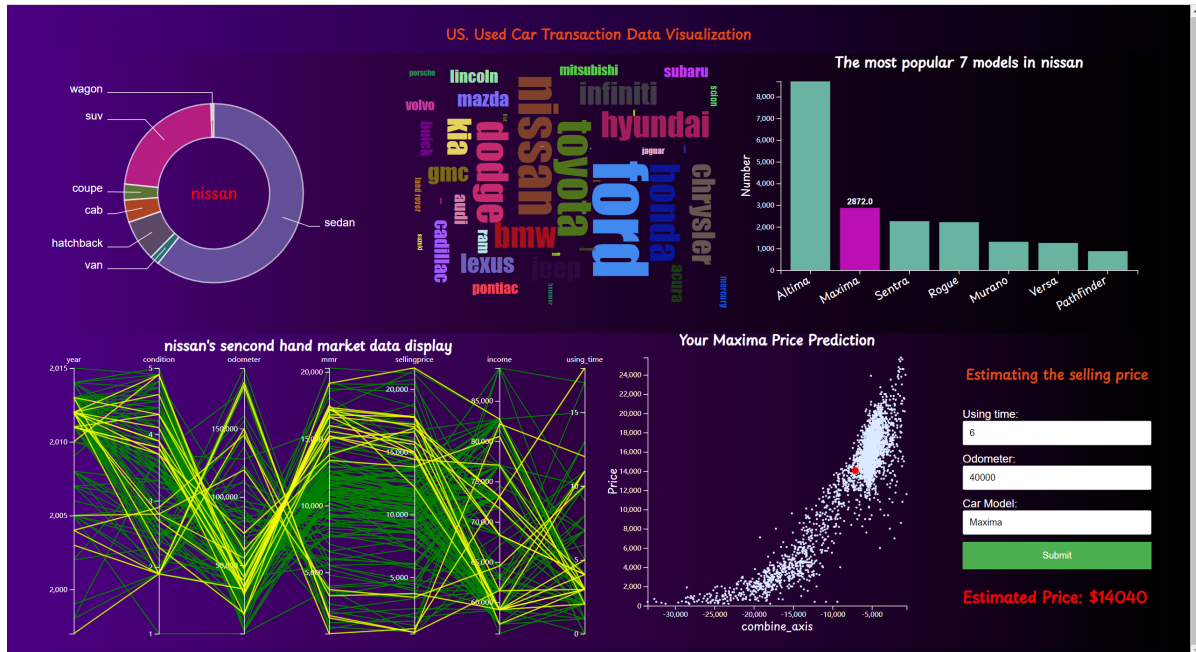
1. The **pie chart** positioned in the upper-left corner illustrates the distribution of car body types for the selected car brand (Ford is depicted in the example image).
2. In the middle, the **word cloud picture** shows the car brands present in the dataset. The size of each word corresponds to the number of cars within the dataset belonging to that brand.
3. In the upper-right corner, the **bar chart** identifies the most popular models within the selected car company (Ford is exemplified in the image).
4. The **parallel coordinate display** visually represents the interrelation among various features of the cars.
5. The **scatter plot** offers a 2-D representation of the data of the selected model (Ford's Fusion in this case). Additionally, the **final Table** located in the bottom-right utilizes big data technology to predict your car's price based on factors such as car model, usage time, and odometer reading. The resulting predictions are reflected on the scatter plot as a prominent red point.

1.2 Designed Interactions

- The pie chart will be highlighted if your mouse is on the pie.
- if you click the pie chart, the particular part of data will turn into yellow in the parallel coordinate display.
- If you click the word on the word-cloud, the bar chart and pie chart will visualize the data of the company you clicked.

- The bar chart will be highlighted if your mouse is on the bar, and a random color will be assigned.
- if the bar was clicked, the coordinate display will highlight the corresponding data in yellow.
- click the submit button on the table with the right input, your input will be visualized on the scatter plot as a red dot. and the predicted price will be at the bottom of the table

An interaction example:



Algorithms of the prediction function

Here we use **KNN algorithms**, filtered the dataset such that the data only preserve the data that only contains `model==input_model`, and the calculated the `combine_axis` based on `using_time` and `odometer`, and Finally find the 10 closest points on the scatter plot, and use the average selling price as the final price of the car

- How to compute `combie_axis` :

The combine axis is a result due to **Binomial Linear Regression**. I regressed the whole dataset using feature `odimeter` and `using_time`, and get its corresponding parameters, in that way the scatter plot will be more like a linear function, and patterns will be easier to find.

$$combine_axis = (-720.9678)using_time + (-0.06936549)odometer;$$

02 Two Interesting Stories

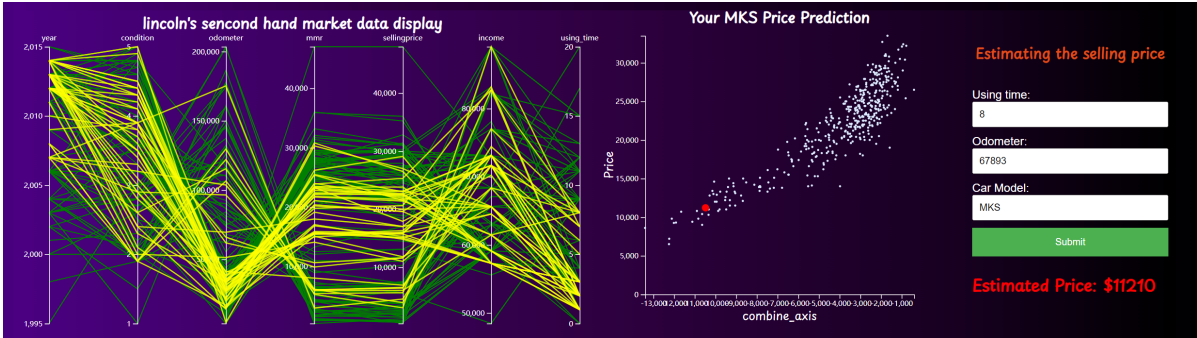
2.1 Mike wants to sold his old Lincoln ?

Mike is a data scientist working in New York, However, somehow his company was bankrupted few months ago, and Mike lost his job.

Unfortunately, at the same time, his wife was ill, so most of the savings went to his wife's medical treatment.

Mike has no choice but to sell his old Lincoln.

Then Mike opened my web page, his car was 8 years old, and Its mileage has exceeded sixty thousand miles...



It can only be sold like \$1, 1210...

But he still remember he spent more than \$3, 0000 to buy the car!

It doesn't worth it... he thought for a while.

Then he gave up his idea of selling a car, and try to find other way to make enough money for his wife, everything will be figured out!

2.2 Simon wants to buy a car...

Simon was very proud, because his daughter graduated from high school, and get an offer from Stony Brook University

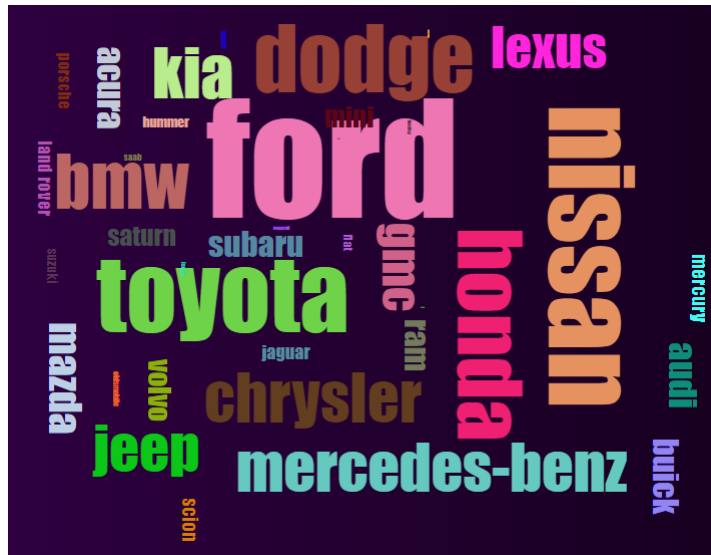
But Stony Brook University is far from city, and it is pretty hard to hang out with out a car

So Old Simon decided to buy a secondary sedan for his daughter as a gift

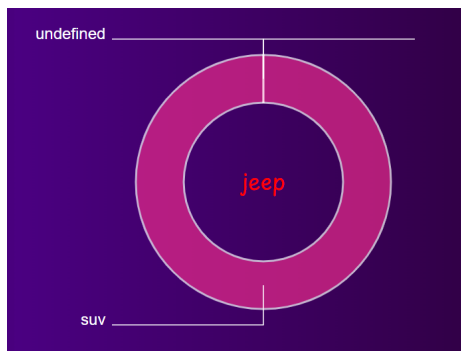
But he have no idea which car to buy...

Then he opened my web page

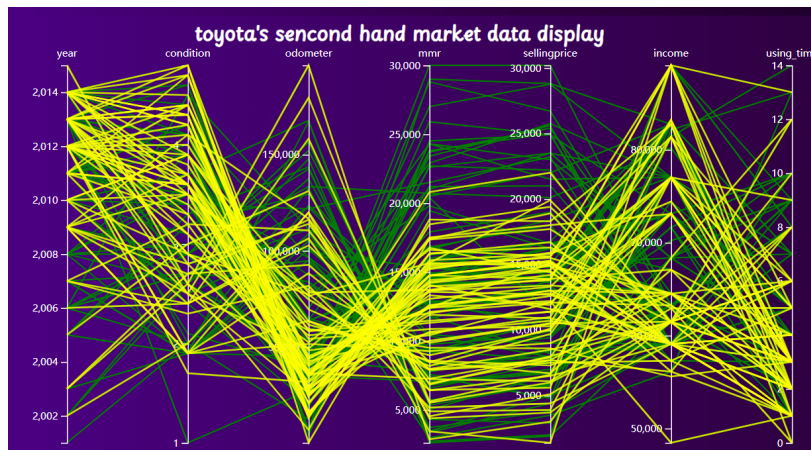
At first glance he saw jeep, is that a good choice? he wondered.



and then he clicked the button...Oops! to his surprise, `Jeep` is a car company that do not sell sedan!



He keeps searching... `toyota` might be a good choice, but is that very expensive?

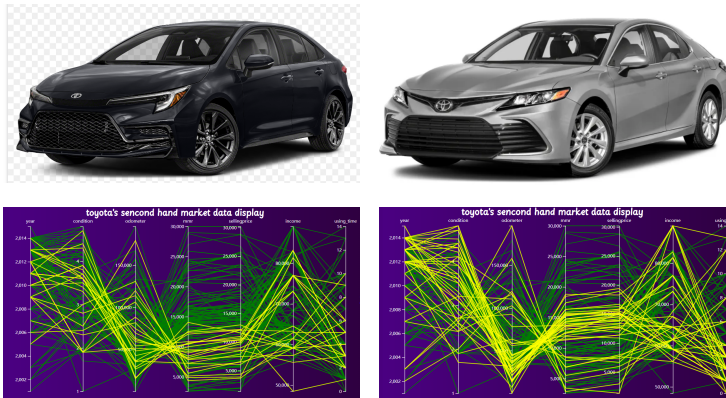
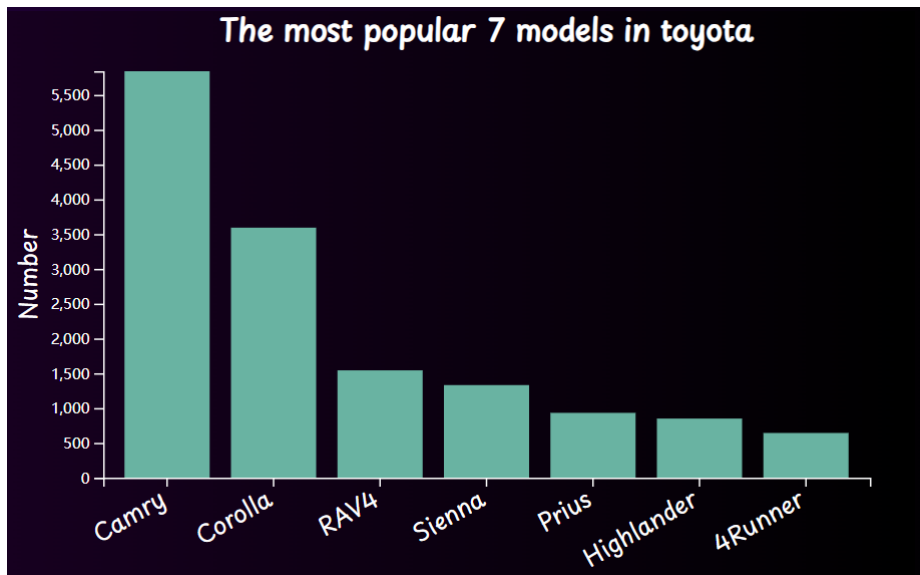


he clicked the button: some cars are indeed expensive, but some are not, and most sedans are at the bottom of the line! (cheers)

But there are many options as well. Camry and Corolla are 2 good choices.

Which one to buy? Their appearance looks similar...

Then he compared the 2 coordinate picture...



Uhhhhh, Corolla seems to be much cheaper than Camry in secondary market! Hence he finally buy a Corolla for his daughter.

Appendix

All the csv files and codes are available on [github Heqijia/CSE332_lab5\(github.com\)](https://github.com/Heqijia/CSE332_lab5)