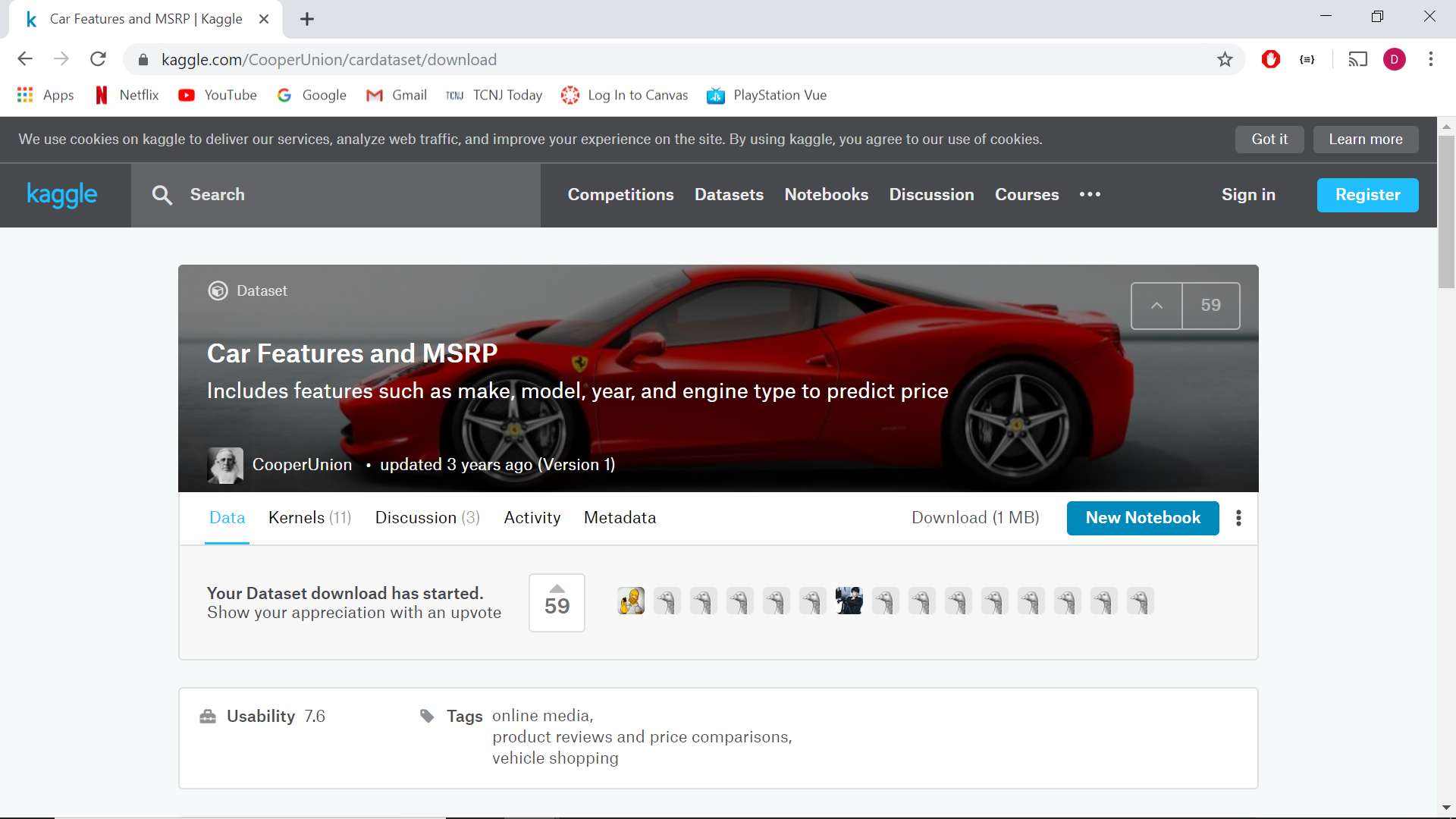
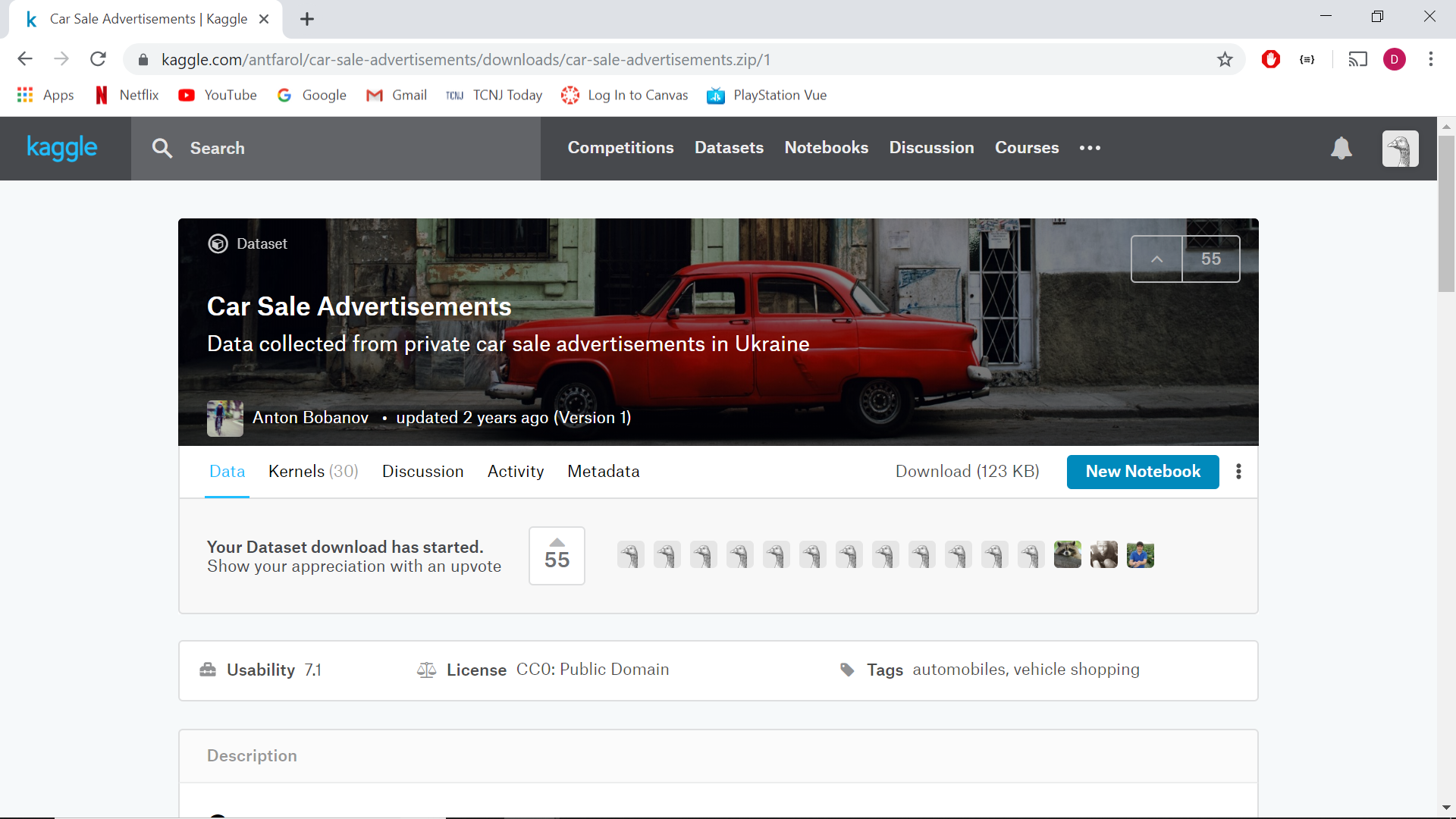
**Danyelle Tolud**

**ETL Project Final Report**

**Extract**

The datasets were found on Kaggle:

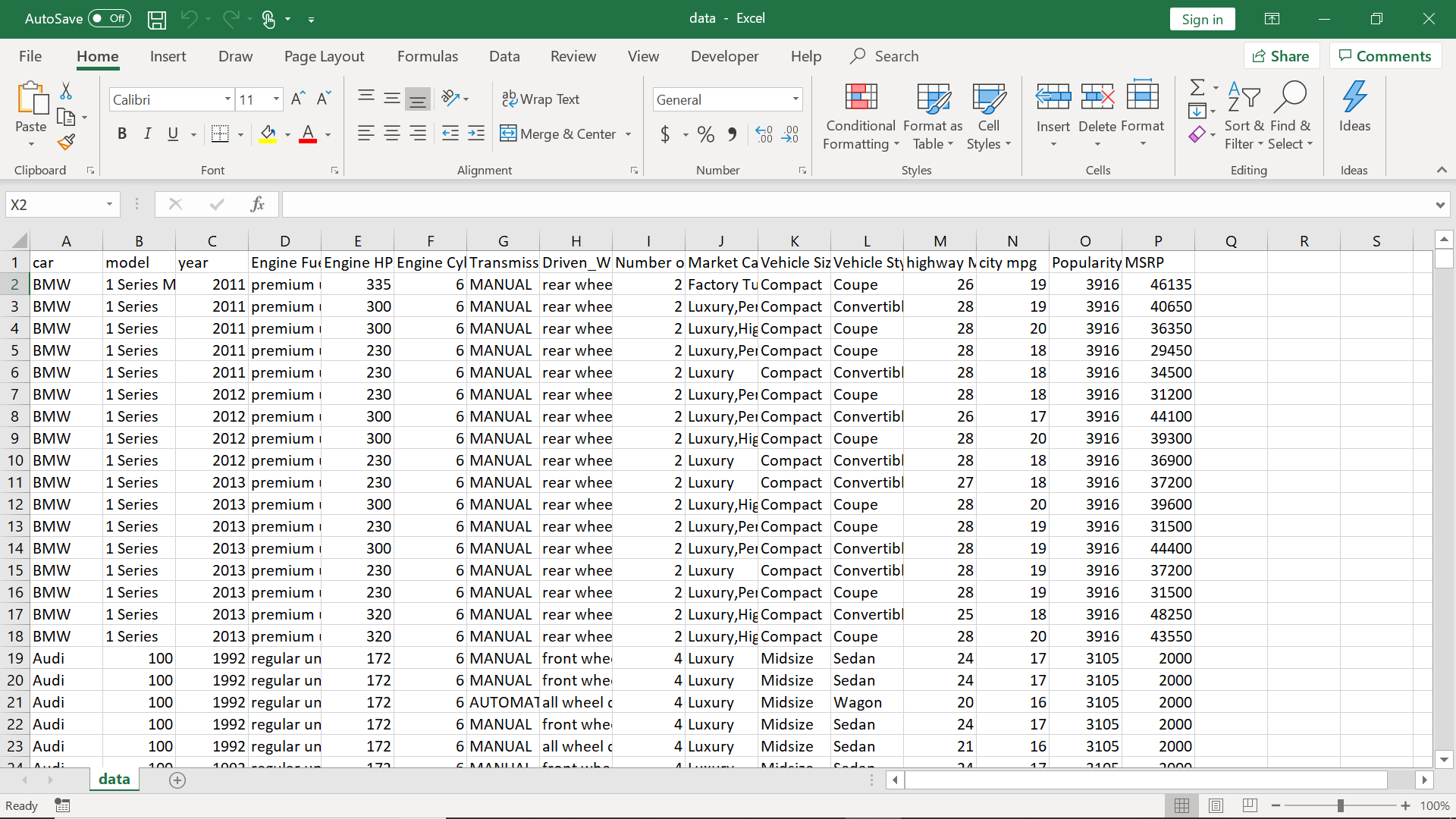




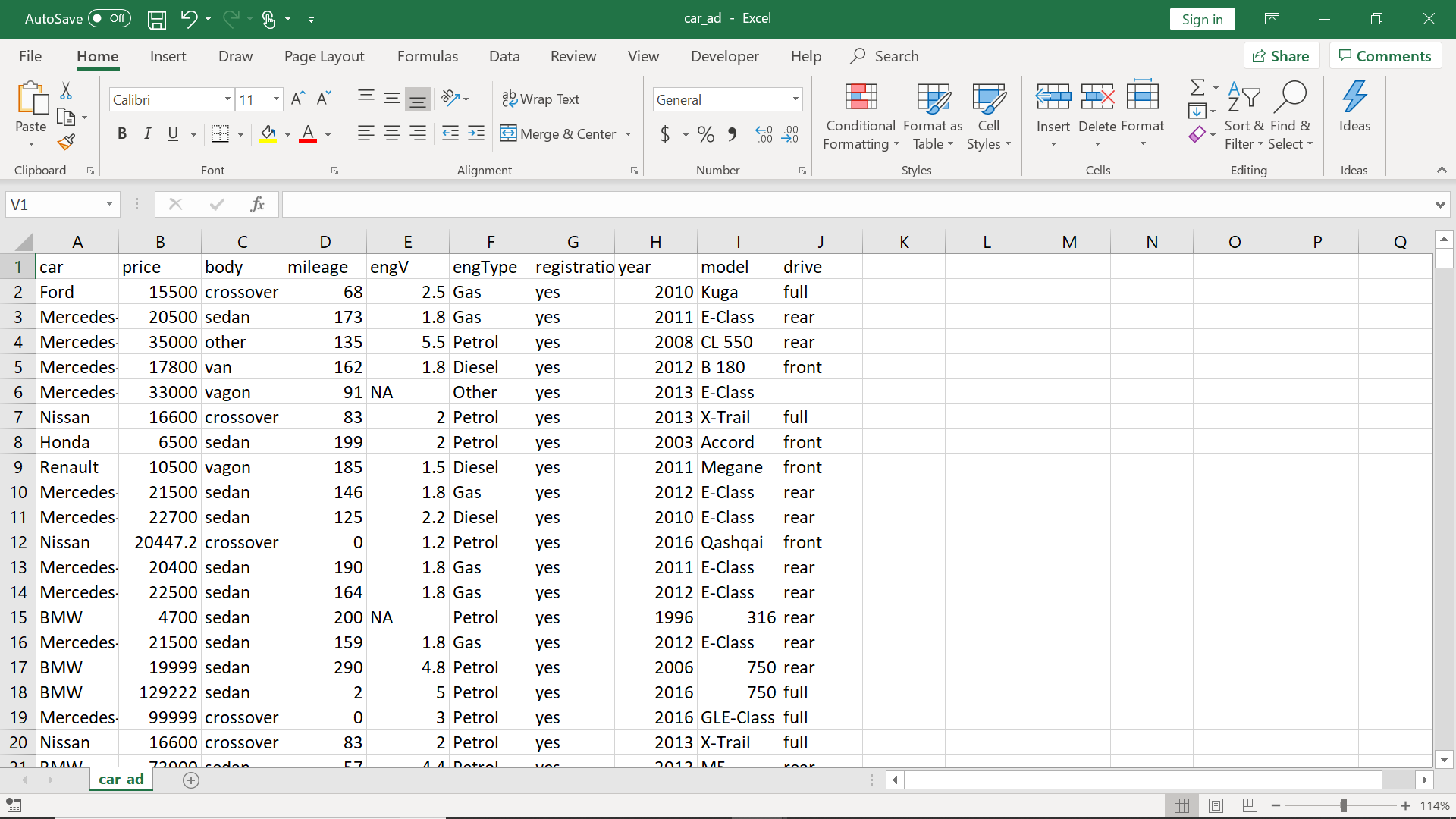
Downloaded the CSV Files:

(car data and popularity)

data.csv :

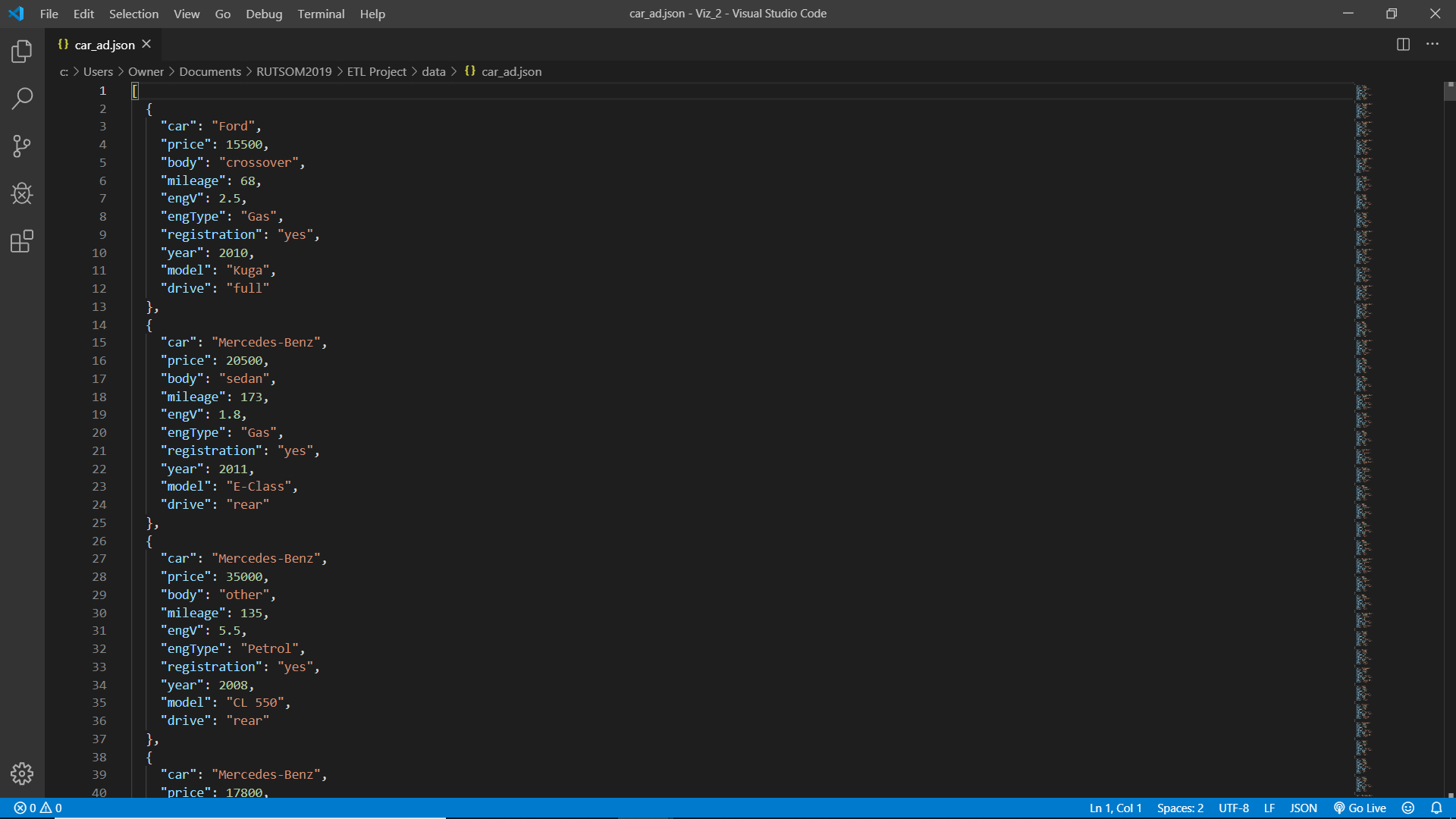


(car advertising and sales data)

car\_ad.csv :

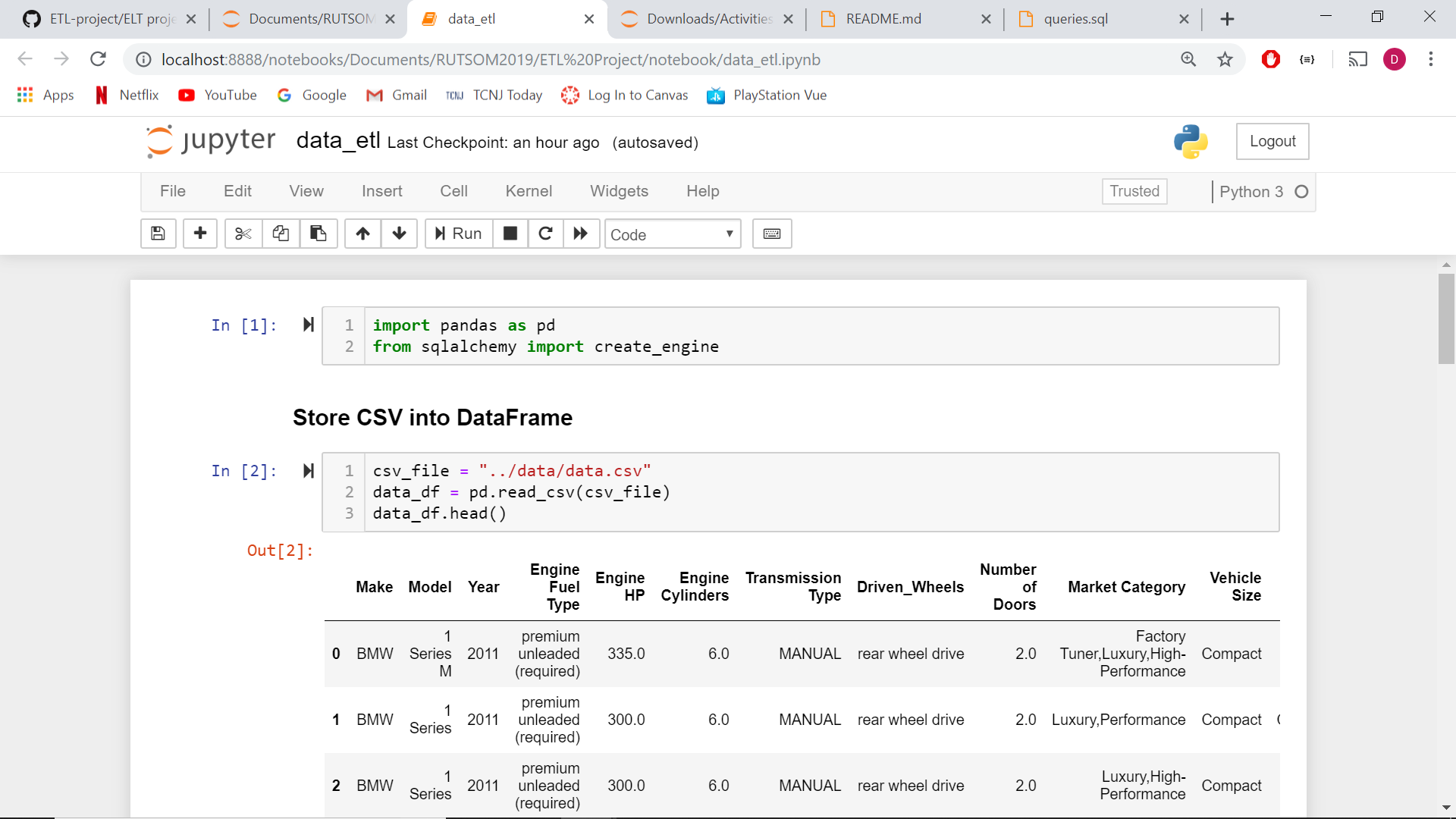
converted >>> CSV to JSON

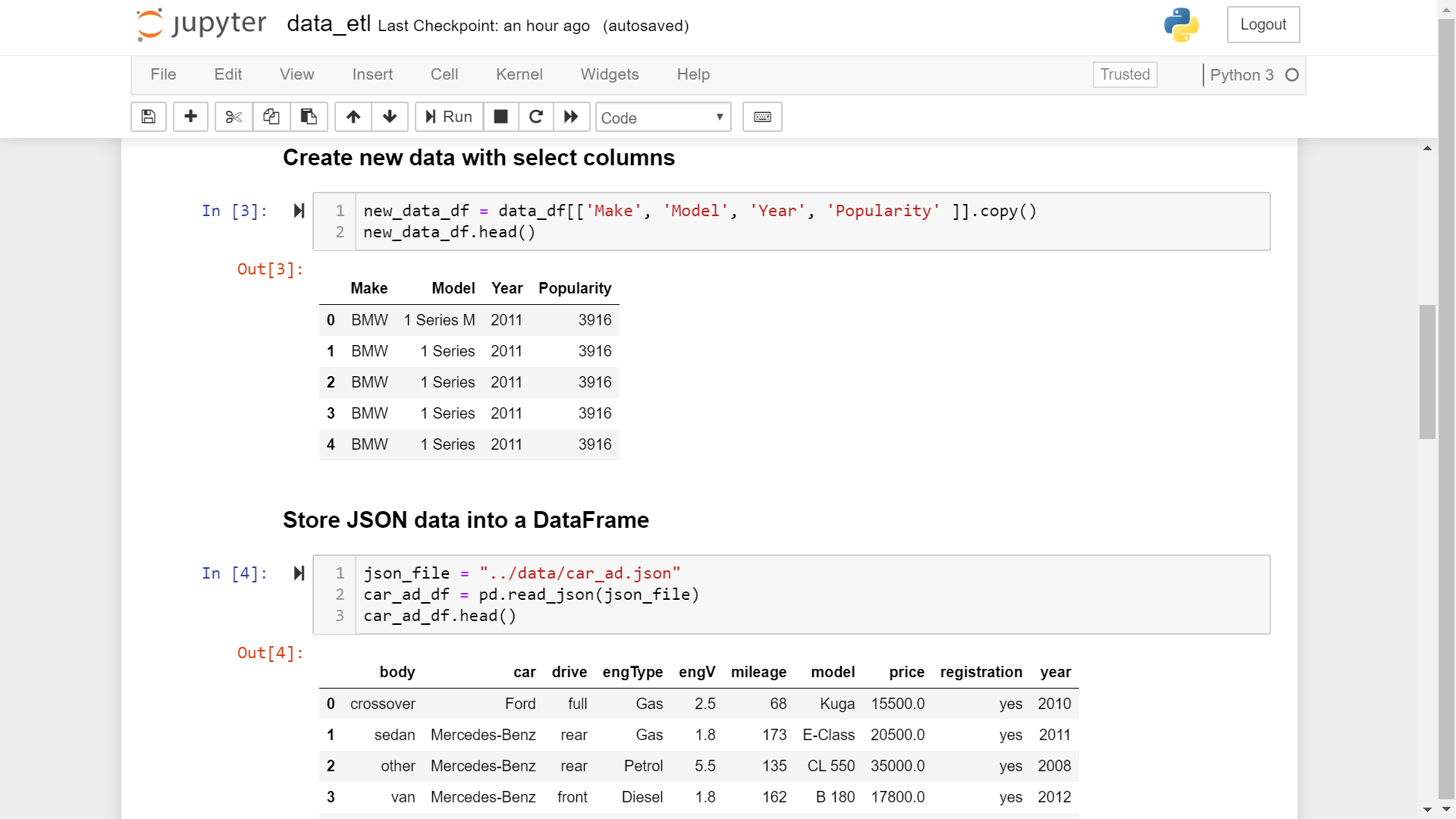
car\_ad.json :

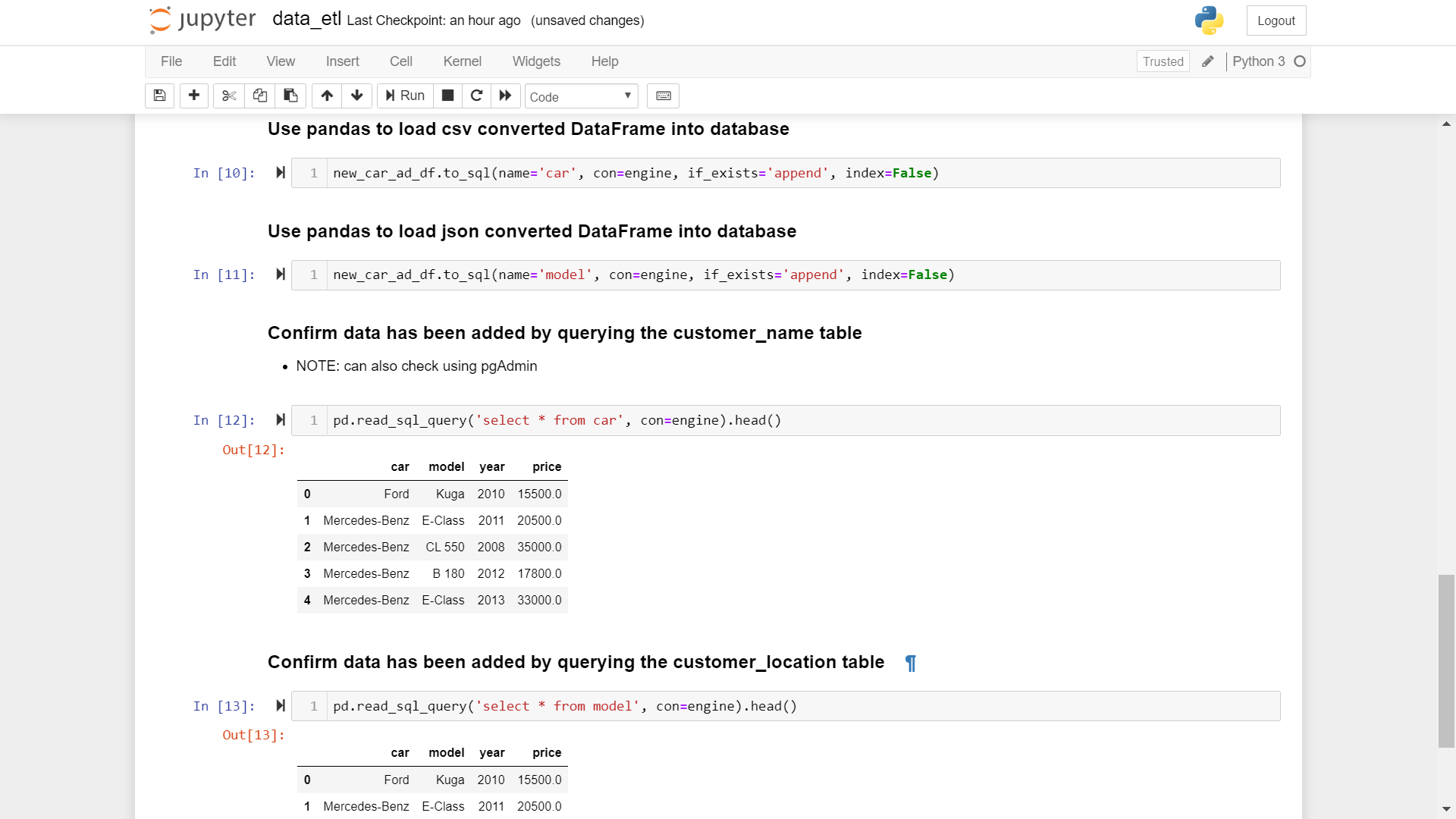
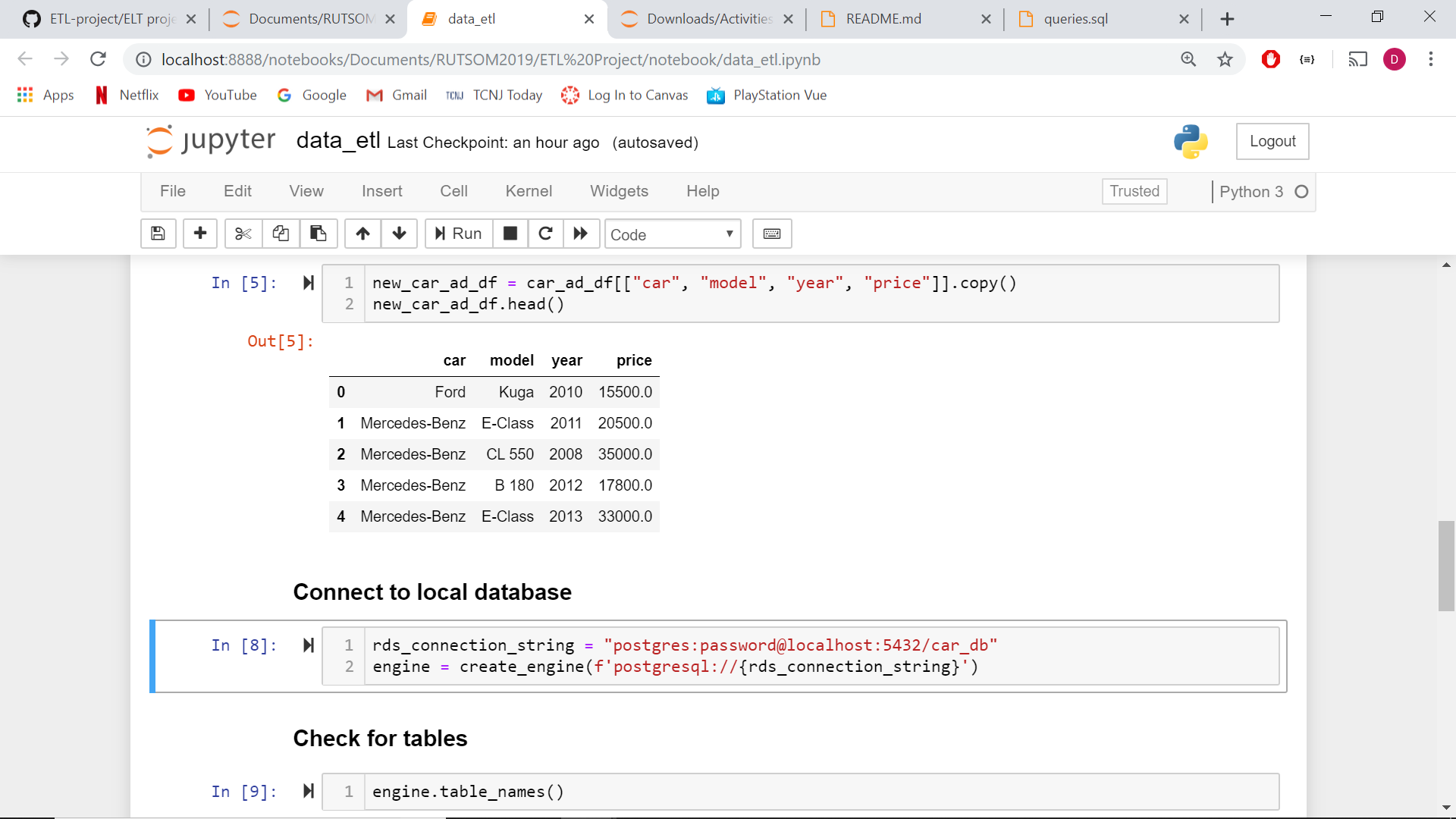


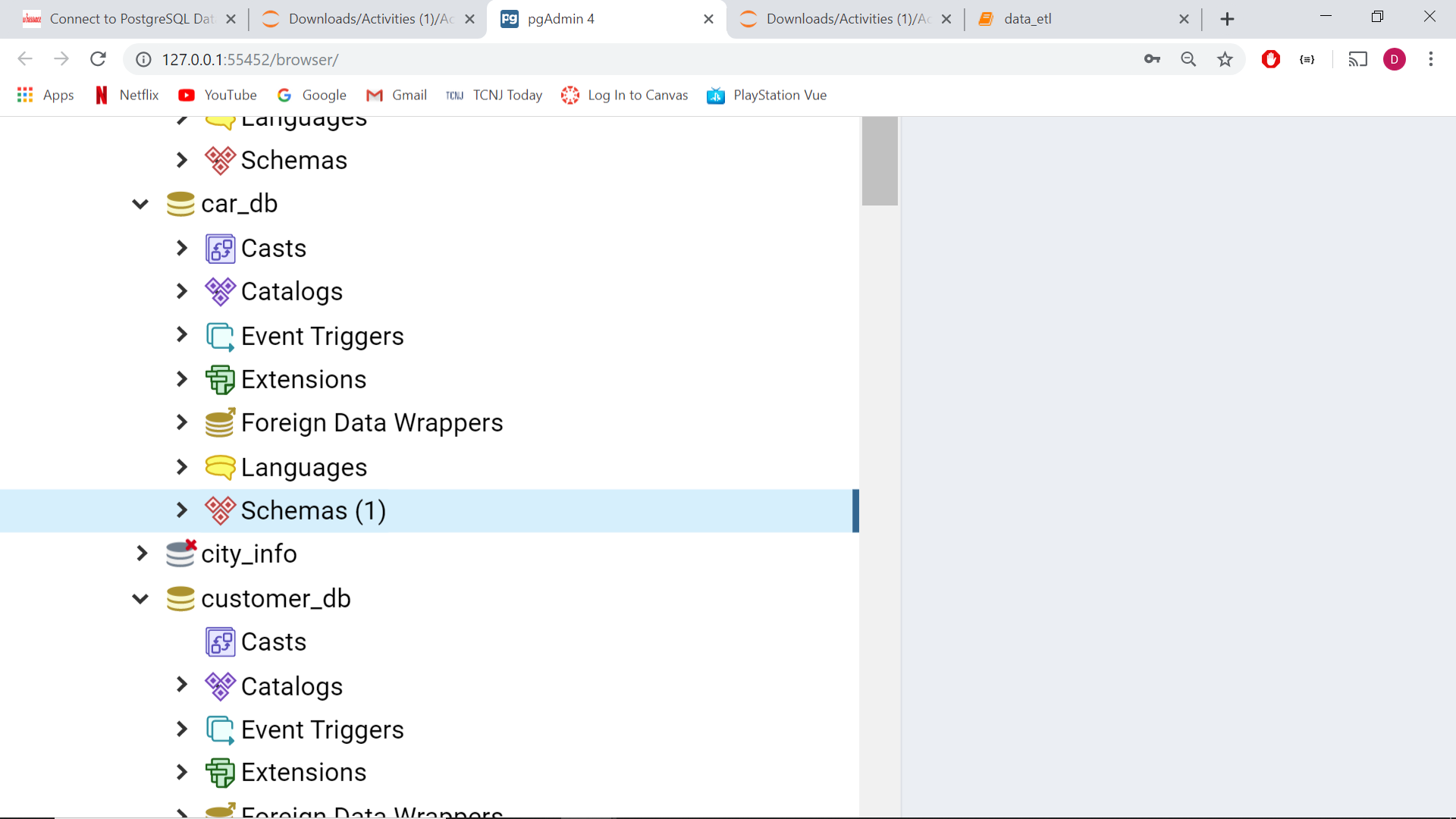
**Transform**

For the ETL process, the data was extracted from the original data.csv file and a converted csv to json car\_ad.json file. Then the data was cleaned, transformed, and joined into one dataframe that included data from both datasets, joined by car make, model, and year to include popularity and price from the two different datasets. This was done in order to analyze car popularity and car price together as they are not both included in either dataset but are in the combined dataframe.



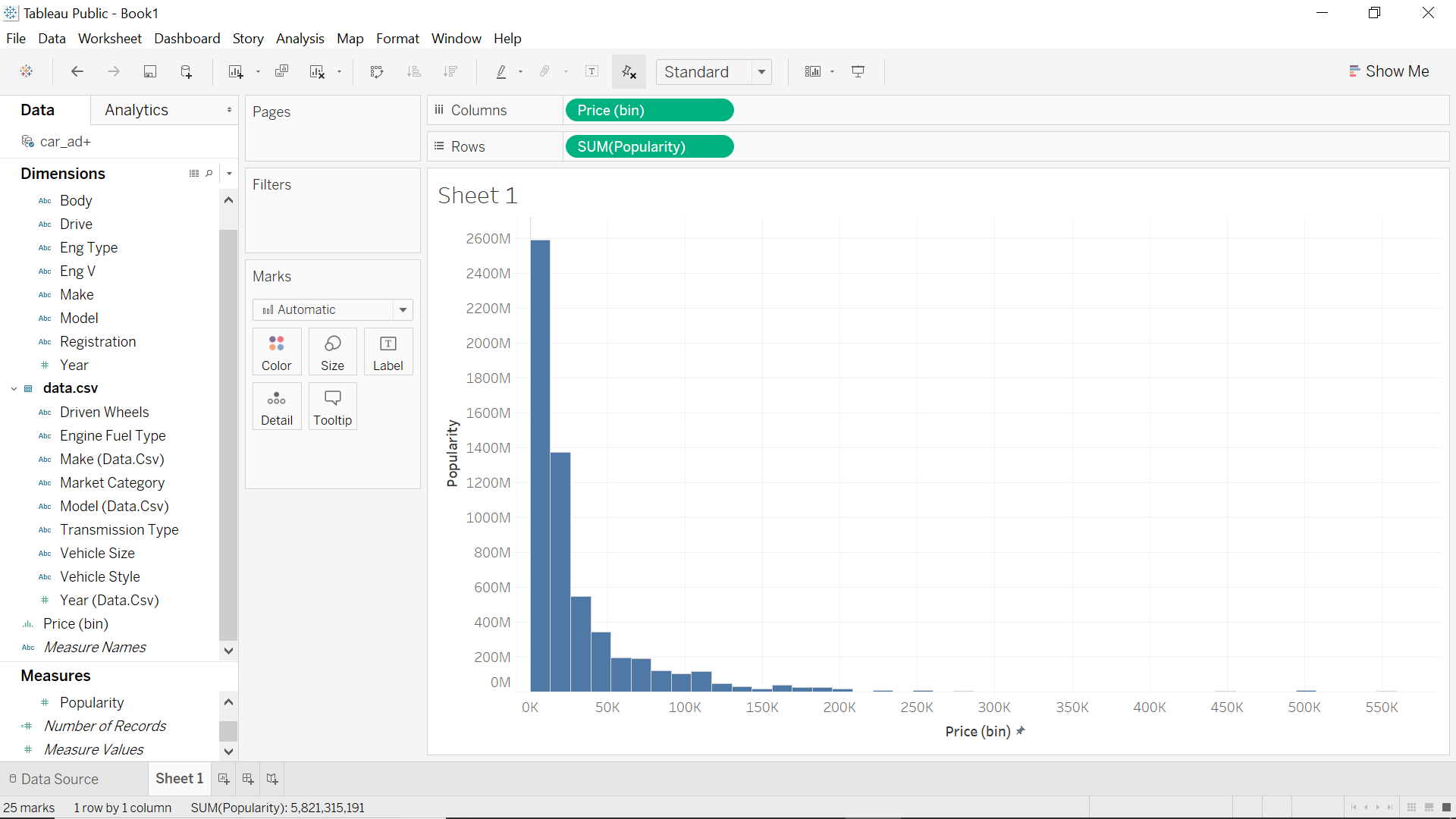






**Load**

Either dataset included some of the same data on cars, which made it possible to join the data together. From the cars.csv dataset we included car make, noted as simply as car, car model, year and popularity. From the car\_ad.json dataset we included car make, model, and year as well, but also price. Because either set included popularity or price but not both, we cleaned the data and loaded them together to analyze both. We can now observe car popularity by price.



Price

From the results, shown in the graph above, we can see a negative correlation—as price increases car popularity decreases. More expensive cars are less popular, and we can see that the most popular cars cost a lot less. This may not necessarily be because cheaper cars are more well-known, perform any better, or have any higher efficiency, etc. but maybe just because they are more affordable. A lot more people can afford cheaper cars, so this causes more expensive cars to be less “popular” and purchased less frequently. However, just because more expensive cars are not bought as often does not mean they have a poor reputation or need increased advertising. People may be well-aware of an expensive car and want to purchase it, more than any other car, but still choose to purchase a cheaper car because it is within their budget. Advertisers should be aware of this when marketing ads, so instead of increasing advertising, they could make it more efficient. Given that price is strongly correlated to popularity, advertisers need focus on promoting affordability if they wish to increase popularity. Future analysis could include looking at cars sales by different demographics. Because cost is indicative of popularity, we could see how income or location ties into that. Then in terms of marketing, companies directly target the appropriate demographic to have more efficient means of advertising.