**INITIAL EDA- 220512**

Calculating the Top 20 Most Popular Vehicles according to ‘Year’

1. Combine ‘manufacturer’ and ‘model’ to create new attribute ‘make and model’—> for ease of reference
2. Bar graph —> because discrete variables
3. Plot x axis= ‘make and model’
4. Plot y axis= ‘number of postings’
5. User input—> slider for ‘year’ (from xxx to xxx year)

Chart, bar chart

Description automatically generated

Timeline

Description automatically generated

Calculating the Price of Top 20 Most Popular Models according to ‘type’

1. Combine ‘manufacturer’ and ‘model’ to create new attribute ‘make and model’—> for ease of reference
2. Box and whisker diagram—> max/ min/ mean/ median/ outliers easily identified
3. Plot x axis= ‘make and model’
4. Plot y axis= ‘price’
5. User input—> dropdown menu of all available ‘type’ (truck, suv, pickup, coupe, mini-van, sedan, convertible, wagon, van, hatchback, **other**) *?Should ‘other’ be included?*

Chart

Description automatically generated

A picture containing text

Description automatically generated

Geographic representation of which states have the most used cars according to ‘make and model’ and ’type’

1. Combine ‘manufacturer’ and ‘model’ to create new attribute ‘make and model’—> for ease of reference
2. Select attributes ’states’, ‘number of postings’, ‘make and model’, ‘type’,
3. Heat map—> to show states with most number of available used cars
4. User input—> checkbox- users can filter according to ‘make and model’ and ‘type’

Map

Description automatically generated

Graphical user interface, application

Description automatically generated

Reference: <https://www.kaggle.com/code/alexandersylvester/used-cars-eda-with-depreciation-analysis>

Heat map ref:

<https://liuyanguu.github.io/post/2019/04/17/ggplot-heatmap-us-50-states-map-and-china-province-map/>

*Dataset:* [*https://www.kaggle.com/datasets/austinreese/craigslist-carstrucks-data*](https://www.kaggle.com/datasets/austinreese/craigslist-carstrucks-data)