# **Week 8 - Final Project**

Final Project

Kimberly Sultan 12/17/18

**Getting the Data**

Data scraping was completely new to me. With the basics, I was able to scrape from the cars.com website the main car data, like name, model, and price. R Studio was useful for cleaning and writing data to files. This process was the most time intensive. It was also iterative so having done some process for one car group, I’d apply the same steps to others to ensure consistency.

**Selecting Price Point**

Using boxplots allowed me to see the spread and average price of the cars for that manufacturer. Using the most middle of the road of the data, I chose one car category from the dataset of returned cars. Cars under 30k met this requirement and were mostly from the middle of the boxplot.

**Story**

The story is about the car buying process in the company. Though we have the corporate criteria, additional criteria (insurance, fuel economy, etc.) effects what employees may perceive to be the best company car. Importance was place on looking at cost over time with the expected ownership of 2 years. Safety was the most important criteria for the company with 3 of the 4 cars having a top crash test rating of 5. Having other criteria was helpful since it’s hard to eliminate a car if they all have 5 of 5 stars.

**Vehicle Choices**

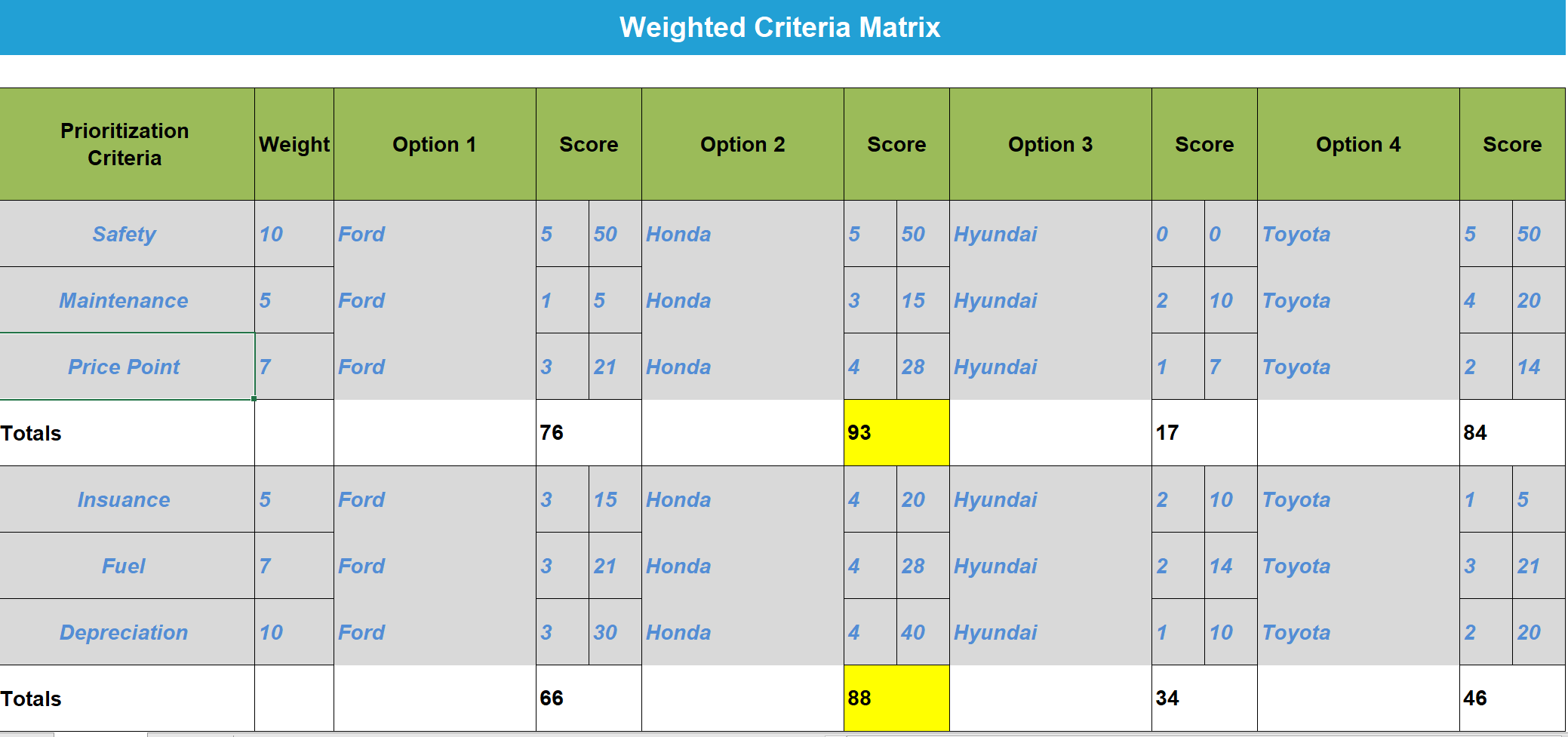
* 2017 Ford Escape
* 2017 Honda CRV
* 2017 Hyundai Santa Fe
* 2017 Toyota Rav 4

**Criteria 1 (Company Criteria)**

* Safety features – weighted at 10 (based on Kelly Blue Book Crash Test Rating)
* Maintenance cost – weighted at 5 (based on 2-year maintenance costs for Edmunds.com)
* Price point – weighted at 7 (analysis of box plots)

**Criteria 2 (Personal Criteria)**

* Insurance – weighted at 5 (based on 2-year insurance costs)
* Fuel Economy – weighted at 7 (based on 2-year fuel costs)
* Resale Value – weighted at 10 (based on 2-year depreciation costs)



**Data Scraping Methods**

Data was scraped to obtain vehicle make, model, year from website cars.com. I Used R code to pull the data from the URLs. In R, I used histograms and boxplots for each car’s raw data to see the distribution of price. The separate car data frames were merged together into a single dataset and written to the csv file. There were 183 car records. I used this data to narrow down the list to the 4 car types within each car’s group, the basic model and style for that make of car.

I was unable to scrape the costs items by the method used above. The html codes did not allow any kind of scaping of summarized data. I used Edmunds.com to compare the 4 cars’ yearly cost and total 5-year true cost of ownership. I created a spreadsheet since it was data for 4 car choices listed in matrix. I ranked them in an inverse order by the lowest cost and then multiplied by the weight for the category. For instance, the car with the lowest maintenance costs for 2 years had the highest score of 4. The next lowest had a score of 3 etc. The score could then be multiplied by the weight. The car with highest score would be chosen.

**Evaluation**  
Based on criteria 1 and with the matrix scoring the Honda CRV is the top choice. The score was 93.

Based on criteria 2 and with the matrix scoring the Honda CRA also out-scored the others. The score was 88.

The Honda CVR was highest score for both sets of criteria.

**Sources**  
Edmunds Costs Criteria – Cost to Own tables

<https://www.edmunds.com/ford/escape/2017/st-401628717/cost-to-own/>

<https://www.edmunds.com/honda/cr-v/2017/st-401696116/cost-to-own/>

<https://www.edmunds.com/hyundai/santa-fe/2017/st-401628708/cost-to-own/>

<https://www.edmunds.com/toyota/rav4/2017/st-401691480/cost-to-own/>

KBB Crash Test Rating URLs:

<https://www.kbb.com/ford/escape/2017/>

<https://www.kbb.com/honda/cr-v/2017/>

<https://www.kbb.com/hyundai/santa-fe/2017/> (ratings were not available)

<https://www.kbb.com/toyota/rav4/2017/>

Cars.com URLs:

url\_Ford <- "https://www.cars.com/for-sale/searchresults.action/?rd=99999&mkId=20015&mdId=21088&searchSource=ADVANCED\_SEARCH&yrId=30031936&prMx=15000&zc=80027&stkTypId=28880"

url\_Honda <- "https://www.cars.com/for-sale/searchresults.action/?mdId=20762&mkId=20017&page=1&perPage=50&rd=99999&searchSource=GN\_REFINEMENT&showMore=true&sort=relevance&stkTypId=28880&yrId=30031936&zc=80027"

url\_Hyundai <- "https://www.cars.com/for-sale/searchresults.action/?mdId=21899&mkId=20064&page=1&perPage=50&rd=99999&searchSource=GN\_REFINEMENT&showMore=true&sort=relevance&stkTypId=28880&yrId=30031936&zc=80027"

url\_Toyota <- "https://www.cars.com/for-sale/searchresults.action/?mdId=21780&mkId=20088&page=1&perPage=50&rd=99999&searchSource=GN\_REFINEMENT&showMore=true&sort=relevance&stkTypId=28880&yrId=30031936&zc=80027"

**Attachments:**

Data files (raw and cleaned) in spreadsheet format

"C:\Users\kimbe\Downloads\AllCarsPBI.xlsx"

"C:\Users\kimbe\Downloads\carsCriteria.xlsx"

Visualizations / PDFs

"C:\Users\kimbe\OneDrive\Desktop\Data Visualization\Week 8 Data Viz\Final Project\_ksultan.pdf"

"C:\Users\kimbe\OneDrive\Desktop\Data Visualization\Week 8 Data Viz\Final\_Project\_R Markdown Kim\_Sultan.pdf"