RECOMMENDATION FOR VEHICLE PURCHASE

RISHAB GUPTA
MSc IN COMPUTING AND DATA ANALYTICS
ST. MARY'S UNIVERSITY

Table of Contents

Executive Summary

The purpose of this report is to give a recommendation to Sue, who intends to buy a car and has expressed certain criteria that are important in her decision.

The dataset is comprised of specifications and dimensions of 10 potential cars. Based on the given priorities the data set was analyzed to identify the key variables crucial for the recommendation. It was observed that out of the 18 given features, only 8 features were vital to proceed with the analysis.

Safety rating, being the most important factor for Sue in her decision was given the highest priority during the analysis. Therefore, as the first step, the cars were sorted by Safety Rating and the cars with the highest safety rating were filtered out to be analyzed.

Tableau was chosen as a choice of visualization tool and exploratory analysis was performed to understand the range of values and distribution of each of the variables. It was observed that Honda CRV has the highest initial cost of $20,000 and has the largest cargo volume of 1000. Whereas, Toyota Prius was the most fuel-efficient with the second largest cargo volume.

Some new features were now created to provide a reasonable recommendation. A few assumptions, like, the fuel cost and the percentage of drive on the highway and in the city, were made. Using the given data, the total cost of owning and operating the vehicle for the duration of the ownership was calculated, which was used as a yardstick in the final recommendation.

After the feature engineering, each of the car was individually compared with the other cars and a thorough and rigorous analysis was conducted. Keeping in mind the safety rating, the fuel efficiency, cost of owning and operating the vehicle and a large cargo volume, it was finally concluded that Toyota Prius would be the best choice of car for Sue.

Problem Statement

Sue intends to buy a vehicle. She has a certain criterion that needs to be met. Using the data provided on potential vehicles, we are asked to come up with a recommendation for the vehicle that best suits her needs. Stated below are Sue's priorities:

1. Sue is a family person and has a dog
2. This will be her family's primary mode of transportation
3. Sue intends to own the vehicle until the odometer reads 250,000 km
4. Sue expects roughly 25,00 km of annual driving
5. Storage space is important to Sue, owing to her long-distance visits to family, camping trips, etc.
6. Sue considers the vehicle's safety rating of utmost importance, and the total cost of owning and operating the vehicle as her second most important factor in the decision.

Data Set

The data set is comprised of ten car models with their respective specifications:

- Make and Model of the Car
- Safety rating
- Price
- Kilometers driven
- Highway and City fuel efficiency in litres per 100 km
- Tire Size
- Car dimensions
- Front and Rear legroom
- Cargo volume
- Horsepower and Engine size

Visualization

Visualization was performed using Tableau and the dashboard can be accessed here. (https://public.tableau.com/profile/rishab.gupta#!/vizhome/RecommendationforVehiclePurchase/SCover).

Exploratory Analysis

Exploratory analysis was performed to identify the range of values and distribution of each variable. It was noted that only some of the variables contributed towards providing a recommendation.

1. Sue gives safety the highest priority in her decision to buy a car. From the data provided, cars fall into two categories of safety rating, 4 and 5

Safety Rating	
5	CRV | Focus | Prius | Rondo | Soul EX |
4	CX-5 | Edge SEL | Matrix | RAV4 LE | Rogue |

Fig. 1 Safety Rating of each car

2. The price, fuel efficiency and cargo volume of the car are another important factors that would drive Sue's decision. Shown below is the price, fuel efficiency and cargo volume of each car with a safety rating of 5.

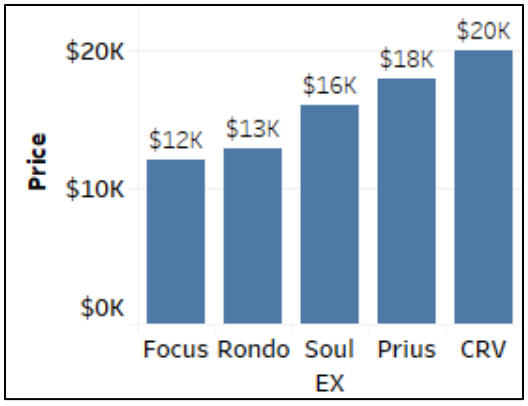

Fig. 2 Bar chart of car prices

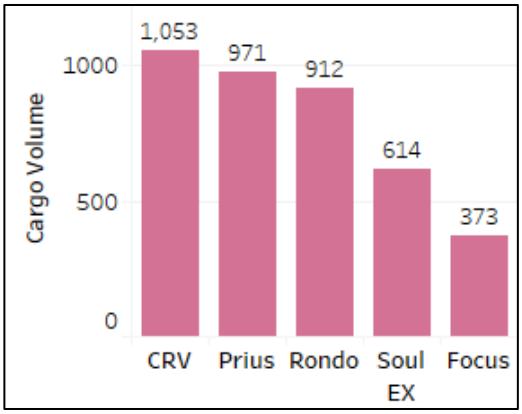

Fig. 3 Bar chart of cargo volume

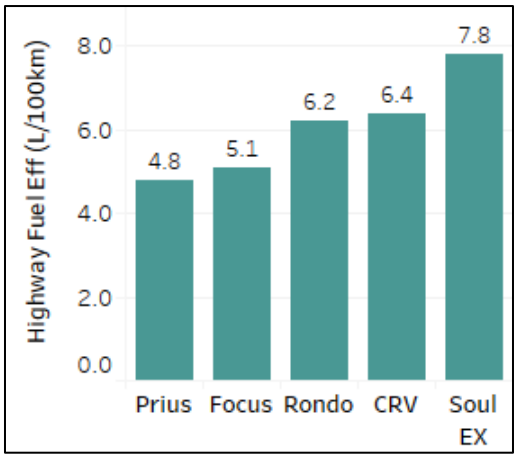

Fig. 4 Bar chart of highway fuel efficiency (the lower, the better)

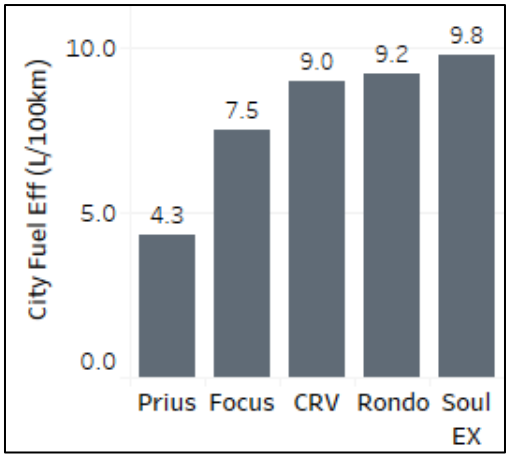

Fig. 5 Bar chart of city fuel efficiency (the lower, the better)

Assumptions

1. Sue lives in Nova Scotia and the average fuel price in Nova Scotia is $1.30 per litre for the duration of the car owned.
2. Out of 25,000 km of annual driving, Sue will drive 50% of the times on a highway and 50% of the times in a city.

Feature Engineering

Feature engineering was performed to better represent the recommendation system and achieve accurate results

1. **Number of years Sue can own the car** – given the current odometer reading of the cars, the average annual driving (i.e. 25,000 km) and the final odometer reading until Sue intends to own the car (i.e. 250,000 km); the total number of years Sue can own the car was calculated using the formula:

```
Total Years = (250000 – current odometer reading)/25000
```

2. **Highway and City mileage** – using the 'Highway Fuel Efficiency' and 'City Fuel Efficiency' fields given in litres per 100 kilometers, mileage on a highway and in a city were calculated respectively using the formula:

```
Highway Mileage = 100/Highway Fuel Efficiency
                    and
City Mileage = 100/City Fuel Efficiency
```

3. **Total cost of the fuel to drive for the duration of the car owned** – using the assumption that Sue will drive 50% of the times on a highway and 50% of the times in a city and considering the average fuel price in Nova Scotia, i.e. $1.30 per litre, the total fuel price was calculated using the formula:

```
Total Fuel Cost =
  (Fuel Price * 50% of highway drive * Total Years)/Highway Mileage +
     (Fuel Price * 50% of city drive * Total Years)/City Mileage
```

4. **Total cost of the car –** The total cost of the car is the summation of the Car Price and the Total Fuel Cost. This field was calculated with the aim to find out the total amount of money Sue will have to pay to own and operate the car until the odometer reads 250,000 km. The Total cost of the car was calculated using the formula:

```
Total Cost of the car = Total Fuel Cost + Car Price
```

Analysis

After performing the exploratory analysis, data cleaning and feature engineering, the following results were produced:

Make	Model	Safety Rating	Total Years	Cargo Volume	Price	Highway Fuel Eff (L/100km)	City Fuel Eff (L/100km)	Total Cost of Car
Ford	Focus	5	8.11	373	$11,999.00	5.1	7.5	$28,608.57
Toyota	Prius	5	7.76	971	$18,000.00	4.8	4.3	$29,471.26
Kia	Rondo	5	8.18	912	$12,884.00	6.2	9.2	$33,350.85
Honda	CRV	5	8.21	1053	$19,995.00	6.4	9	$40,549.03
Kia	Soul EX	5	8.76	614	$15,995.00	7.8	9.8	$41,060.04

Table 1. Feature Engineered dataset

From the table above, it is observed that:

- The initial price and the 'Total cost of car" is the lowest for Ford Focus. However, it compromises on the cargo volume to large extent which is almost half of the cargo volume of Kia Soul EX.

- Toyota Prius stands second lowest in 'Total cost of car' after Ford Focus. The initial cost of the car is $6000 higher than latter, however, there is an insignificant difference in the total price of both the cars. The cargo volume of Toyota Prius is also three folds greater than that of Ford Focus. Moreover, out of all the cars, Toyota Prius is the most fuel-efficient.

- Kia Rondo has a low initial price but $1000 higher than that of Ford Focus's. However, it stands in the top three of the highest "Total cost of car". It can be clearly interpreted that Kia Rondo is not fuel-efficient car.

- Honda CRV has the highest initial price of $19995 and has the largest cargo volume of 1053, crossing the 971 mark of Toyota Prius. However, the 'Total cost of car' is significantly high, touching $40,000, and is not as fuel efficient as Toyota Prius or Ford Focus.

- Kia Soul EX is a mid range car and has the third lowest initial price after Ford Focus and Kia Rondo but, has the highest 'Total cost of car' compared to all the other cars. This is owing to the large numbers of Highway Fuel Eff and City Fuel Eff which are almost the double of Toyota Prius.

From the above analysis, it is evident that Toyota Prius would be the best car that meets all of Sue's requirements. However, it is noteworthy that Toyota Prius has the least number of total years of ownership of the car. It is arguable that owing to the lesser total years of ownership, the total price of the Toyota Prius is lower compared to the other cars. Therefore, to address this problem, the 'Total Years' of all the cars was set to equal to the 'Total Years' of Toyota Prius, i.e. 7.76

Make	**Model**	**Safety Rating**	**Total Years**	**Cargo Volume**	**Price**	**Highway Fuel Eff (L/100km)**	**City Fuel Eff (L/100km)**	**Total cost of car**
Ford	Focus	5	7.76	373	$11,999.00	5.1	7.5	$27,887.60
Toyota	Prius	5	7.76	971	$18,000.00	4.8	4.3	$29,475.10
Kia	Rondo	5	7.76	912	$12,884.00	6.2	9.2	$32,303.40
Honda	CRV	5	7.76	1053	$19,995.00	6.4	9	$39,414.40
Kia	Soul EX	5	7.76	614	$15,995.00	7.8	9.8	$38,188.60

Table 2. Feature Engineered dataset with same value of Total Years

Once again, from the table above it is evident that there is an insignificant difference in the 'Total cost of car' and considering the long-term ownership, Toyota Prius is still the best suitable car for Sue.

Conclusion

The data was filtered, and feature engineered to meet Sue's priorities. The highest priority was given to the Safety Rating of the car, followed by the price of owning and operating the vehicle and the cargo volume. After making few assumptions and calculating the total price of each car, an analysis was performed to best suit Sue's priorities.

It was challenging to close on one car that would be best suitable to Sue's requirements. However, using logic and reasoning on the analysis, it is concluded that Sue can confidently buy Toyota Prius which meets all her requirements, and this will the best choice considering she will own the car for the next 7.76 years.