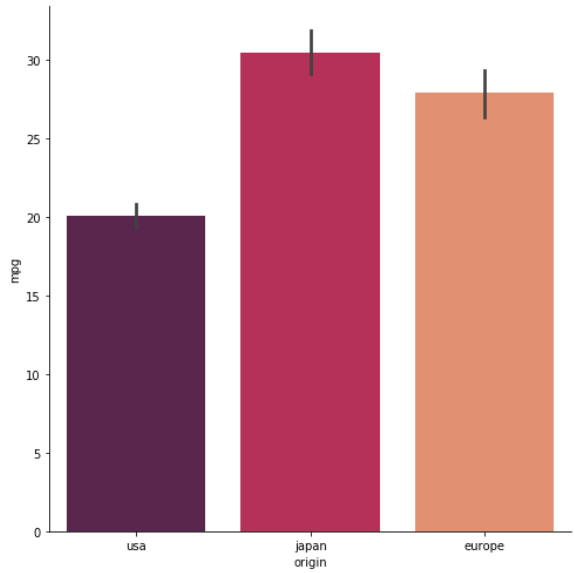
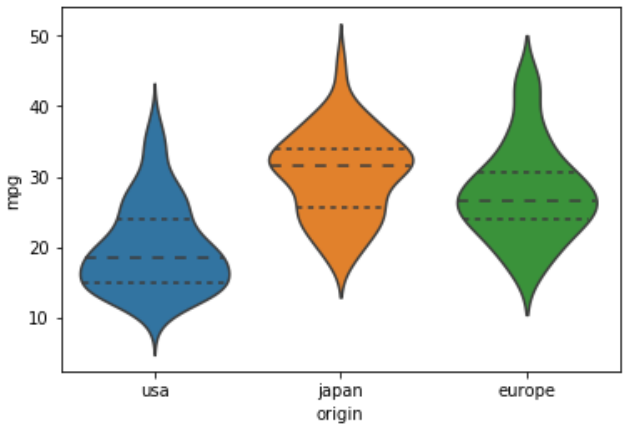
Does a car's origin impact its longevity?





The problem

Buying a car is a major milestone and major decision that should not be taken lightly. In the U.S. 14 million new cars on average are sold every year. It is often stated that American cars are not as well made as foreign cars but does that mean that the car is not as efficient, or more prone to repairs and car death? Insight into the average mpg dataset from seaborn suggests that Japanese and European cars are drastically more efficient when it comes to average gas consumption. This led me to question if differences in origin result in drastic differences in longevity. What is the impact of car origin to the longevity of a vehicle? Is car origin or maintenance more impactful in the longevity of a car? In this study, I define longevity as the number of miles the car can drive before drastic repairs (engine, transmission, motor) or car death.

Study Design

This experiment will follow new car owners of different car origin types and car classes as they record problems, maintenance, and mileage over the course of 6 years to further understand if origin corresponds with reliability. In examining distributions of car lifespans, car buyers can know which origin of cars generally have the best longevity and return on their investment.

Hypothesis

Japanese cars, as shown to be more efficient in gas consumption, will have greater longevity and reliability than European and American cars

Null Hypothesis

There will be no significant difference between longevity and reliability among the different origins of the cars.

The method of testing the solution

In a partnership with new car dealerships across the country, we will randomly select 900,000 non-luxury new car buyers from across the country of models 2016 forward. We will select from

- American dealerships (300,000)
- European dealerships (300,000)
- Japanese dealerships (300,000)

Additionally, the choose automobiles must span across these classes: (100,000 of each origin)

- cars
- SUVs
- crossovers

- While pickup trucks tend to be a popular purchase, they are excluded from the study, because there are few European and Japanese trucks of origin. When examining longevity, according to body type, pickup trucks tend to last the longest and tend to be abundantly American. This would create a form of bias in the data. Additionally, many European cars tend to be luxury which would also create a form of bias in the data, leading to the exclusion of luxury cars. Subgroups will be very important in this study to avoid a Simpson's paradox.

- Data collection will happen through an online portal with a simple and easy to use interface that we design for the study. Upon buying a new vehicle, owners will be

prompted to sign into our online portal, receive instructions of how to log the state of their car and be informed of the study and the benefits of participation (a yearly payout of $50 and discounts on gas)

- The portal will prompt users to maintain a car profile over the course of 6 years where on a monthly basis they will:
    - report car mileage
    - report car washes
    - report car repairs - a key metric
    - report maintenance (non-repair, checkups, personal work)
    - report additional problems

- Secondary metrics that will be collected in partnership with dealerships include:
    - income of participants
    - model year
    - sex of the owner
    - age of the owner

After the six-year period, the data will be examined and categorized by mileage ranges, number of repairs, and the amount of maintenance done. Under each division of the data, the differences in origin type of the car will be studied.

Evaluation

Evaluation of the success of this study will mostly be in the collection of data. If the proper amount of data is collected through the portals over the selected time span to yield significance, the study can be deemed successful and can move towards the final data analysis.

Statistical Test

Next, we will conduct a number of statistical tests to evaluate the validity of our data. To do this we will conduct a t-test. Comparing the mean number of repairs across varying mileage (high, medium, low) and origin (American, European, Japanese). We will conduct an additional t-test in comparing the means of cars categorized by maintenance. We divide the data according to how well kept the car is (high maintenance, medium maintenance, low maintenance) to see if there are significant differences in the number of repairs. Additionally, we will use to the p-value to ensure that our data is meaningful and not due to chance across each test.

Conclusion

If Japanese cars, regardless of mileage and amount of maintenance, resulting in fewer over the time span of the study, then we can reject our null hypothesis and conclude that Japanese cars are more reliable.