# A DSS for Eco-Friendly Car Purchase

**Debora Maia Silva, Prasad Gujela, Shan Lin, Taeyong Kim, Ke Yin**

Purdue University

Department of Management, 403 W. State Street

West Lafayette, IN 47907

[demaiasil@purdue.edu](mailto:demaiasil@purdue.edu); [pgujela@purdue.edu](mailto:pgujela@purdue.edu); [lin882@purdue.edu](mailto:lin882@purdue.edu); [kim2679@purdue.edu](mailto:kim2679@purdue.edu); [yin111@purdue.edu](mailto:yin111@purdue.edu)

**Abstract**

The United States is a country on the wheel. However, the large number of vehicles in this country has serious impacts on public health and the environment. According to reports, passenger vehicles are a major pollution contributor. Choosing the right car is not only significant for consumers, but also important for the well-being of the country. Our Decision-Support-System (DSS) will help American consumers to find the eco-friendliest car in the market based on what they are looking for in a vehicle. Users can input desired features such as their budgets, usage of the cars and desired performance to find the suitable options (automaker and model will be provided as response, along with a local dealer to purchase the vehicle).

Keywords: decision support; eco-friendly car;

## **Business Problem**

According to Union of Concerned Scientists, nearly one half of all Americans – an estimated 150 million – live in areas that don’t meet federal air quality standards. At the same time, the data provided by United States Environment Protection Agency (EPA) shows that, 27% of U.S. greenhouse gas (GHG) emissions is from transportation. Among all different means of transportation, passenger vehicles stand out as the major pollution contributor, producing significant nitrogen oxides, carbon monoxide, and other pollution.

Are there any solutions to solve this intensifying problem? Yes. The development of clean vehicles and fuel technologies can provide consumers with an affordable, available means of reducing transportation-related air pollution and climate change emissions. But how can consumers find out the eco-friendly options from tons of different models? With such question, we decided to create a DSS for consumers to identify the eco-friendliest model that matches their requirements.

The major stakeholders for this DSS are consumers with a specific concern on eco-friendly feature and manufacturers. When it comes to purchasing an eco-friendly car, consumers have their own personal preferences. By inputting their desirable features, such as budgets, car performances, consumers can find the most suitable options with automaker and model. While for manufacturers, this system can give them a brief idea about who’s selling the similar model. It is also possible for them to identify the potential consumers for their specific model.

In order to build our model, we first considered the variables which might affect a consumer’s purchasing decision. We identified several variables that could materially shape the optimal vehicle choice for each customer. Firstly, consumers might have two parts of budgets: their expected car price and their annual fuel cost. Secondly, in order to define car performance, several variables need to be clarified. For instance, cylinders, engine displacement and drive axle type. Besides, consumers might also consider whether they use cars for commuting or they will have a lot of road trips. City mpg and highway mpg will be two important parameters to help them find the more suitable types. After obtaining all the information, we will be able to put these parameters in our model and find out the result.

There are several constraints that we took into consideration when building our model. In order to make our DSS more practical for consumers to purchase new cars, we focus only on cars with the year of 2017 and 2018. The next constraint we take into consideration is whether we would like to include electric vehicles. In order to test the relationship with fuel type and the ecofriendly performance, we decide to only consider cars that burn fuel. We also limit fuel type to conventional fuel, which indicates that these are more likely to be passenger vehicles.

Our initial set of business profits is further elaborated below:

1. This system can help consumers to make better comparison with different car models by their eco-friendly performance and help them to better decide which one to purchase.
2. This system can help manufacturers to increase revenues. By providing a better shopping experience for consumers, manufacturers might attract more consumers. Besides, manufacturers can identify their potential competitors, and they can improve their products accordingly.
3. This system can also help manufacturers to reduce costs. By utilizing this system, consumers can find their ideal cars by themselves. This can help companies to reduce the number of sales people and then reduce their costs.

(~~Statement claiming that stakeholder agreement on the business problem statement has been determined~~)

## **Analytics Problem**

In order to find out the most suitable option, we have to analyze consumers’ needs. When consumers look for a new car, there are mainly three questions for them to consider: how much do I want to pay for this car? What is the main purpose for using my car? What performance do I want for my car? In order to solve these questions, we’ll have to take a look at the variables for different questions.

Firstly, the budget. In this project, consumers’ budget for a new car can be divided into two parts. The net car price and also the fuel cost for the car. In fact, one limitation for our project is that we are not able to grab the actual price for each model. But in order to test the feasibility of our system, we set random values between $15,000 and $70,000 for each car model. Other than the car price, we will take a look at the annual fuel cost for different models.

Secondly, the main purpose of the car. Will the consumer use the car more for commuting in a city or travel between different cities? City mpg and Highway mpg can help consumers to decide which one will be more cost-efficient. The higher the mpg number, the more cost-efficient the car.

Thirdly, how to define a good performance car? Consumers can set different value for cylinder, engine displacement in liters and drive axle type. Generally speaking, with more cylinders, engine displacement in liters and drive axle type, the car will have a better performance, however, it is also likely to have higher fuel consumption and less eco-friendly.

By taking consideration of budget, main purpose and performance, we should be able to find out the one that matches all these features. But instead of showing only one result, the model will instead show four car models that are compatible with the consumer’s desired features. Each model will also be shown with its eco-friendly performance measure - “Green Car”, “Kinda Green”, or “Monster”. A detailed logic of the criteria of eco-friendly performance will be given later in this report.

In this project, we assume that consumers are concerned with the environment and have some value for eco-friendly performance. We also assume that budgets, purpose of the car and performance of the car are the three most important things for consumers to consider when purchase a new car. These key set of assumptions is the basis to support us to solve the problem.

The key metrics of success for this project is whether our app can find out the most suitable option after inputting all different features.

## **Data**

For this project, we have 2 datasets on hand. The first one is vehicles that include all basic information such as vehicle type, fuel type, mpg and so on. The second one is specifically focus on emissions and provides the scores for different vehicles. We combine these two datasets by the column “id”. After loading the whole dataset in R, there are 33,455 observations of 90 variables.

In order to make our DSS more practical for consumers to purchase new cars, we clean our data to cars with the year of 2017 and 2018. In order to test the relationship with fuel type and the eco-friendly performance, we decide to exclude electric vehicles (columns with charge120 or charge240). We haven’t found out a reliable dataset of prices for all new cars, thus, we add random prices (between 15,000 and 70,000) in this dataset in order to test the feasibility of our DSS.

The detailed dataset description is attached below:

make: the automaker / manufacturer of the car.

model: model name.

price: a random price from 15,000 to 70,000.

city08: city MPG for fuelType1.

highway08: highway MPG for fuelType1.

cylinders: engine cylinders.

displ: engine displacement in liters.

drive: drive axle type.

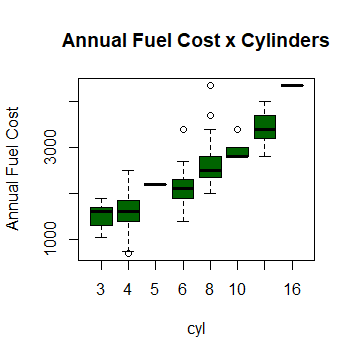
fuelCost08: annual fuel cost for fuelType1.

fuelType1: fuel type 1. For single fuel vehicles, this will be the only fuel. For dual fuel vehicles, this will be the conventional fuel.

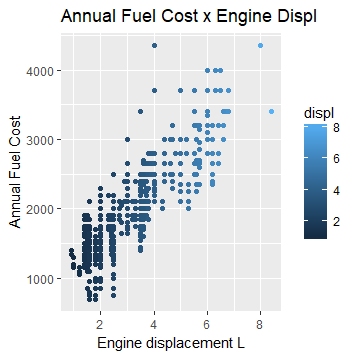
ecoFriendly: There are two types of score to test whether the car is eco-friendly: column “score” in dataset “emission” and column “ghgScore” in dataset “vehicle”. When both score is more than or equal to 8, the car will be marked as “Nature Lover”, which indicates an excellent eco-friendly performance. For ghgScore is more than or equal to 8 while score is more than 6, the car will be marked as “Green Car”, standing for relatively good eco-friendly performance. The medium performance is ghgScore more than or equal to 5 while score is more than 5. The rest represents bad performance in eco-friendly feature.

* Identify relationships in the data (with plots)

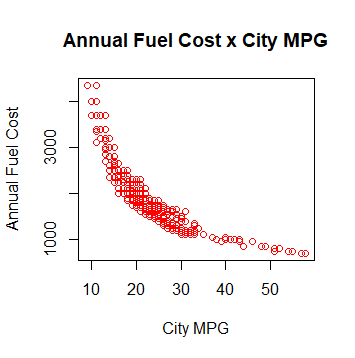
1. Cylinders and FuelCost

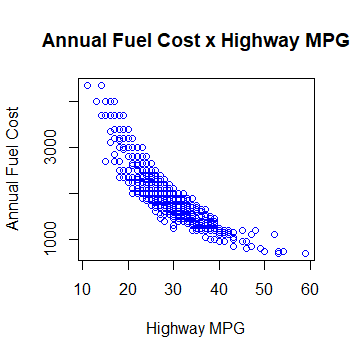


1. Displacement and FuelCost

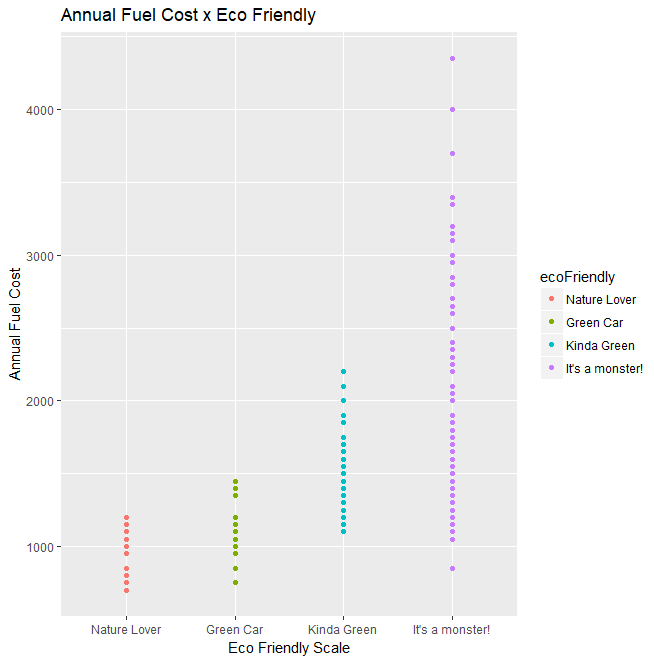


1. Cylinders and Engine Displacement (this one doesn’t make a lot of sense I guess)
2. Drive and FuelCost (this also doesn’t)
3. ?City08 and FuelCost + Highway08 and FuelCost





1. ghg Score, Score and ecofriendly



## **Methodology Selection**

In order to solve the problem, the first thing for our team to do is to clearly define the problem. Nowadays, there are so many websites that can help consumers to find out different type of cars. However, few of them will provide a reference for the eco-friendly performance. We believe that some consumers with an eco-friendly focus are facing problems of finding out the most suitable cars. Thus, this system can help them to identify the most suitable one.

After that our team gathered together and generate ideas to find out the best solution. We had several meetings and tried to brainstorm together. The first thing for us to do is to identify key variables that would help a consumer find out the best car selection. After this step, we would like to find out the best approach to build our model and present our system with our consumers. We went to several different websites, for instance, cars.com and fueleconomy.gov to find out their features. Then we found that the limitation on these websites is that they can’t provide consumers with a powerful recommendation with clear definitions of eco-friendly performance. Based on this finding, we plan to define eco-friendly performance with our own ways and give consumers suggestions based on our metrics.

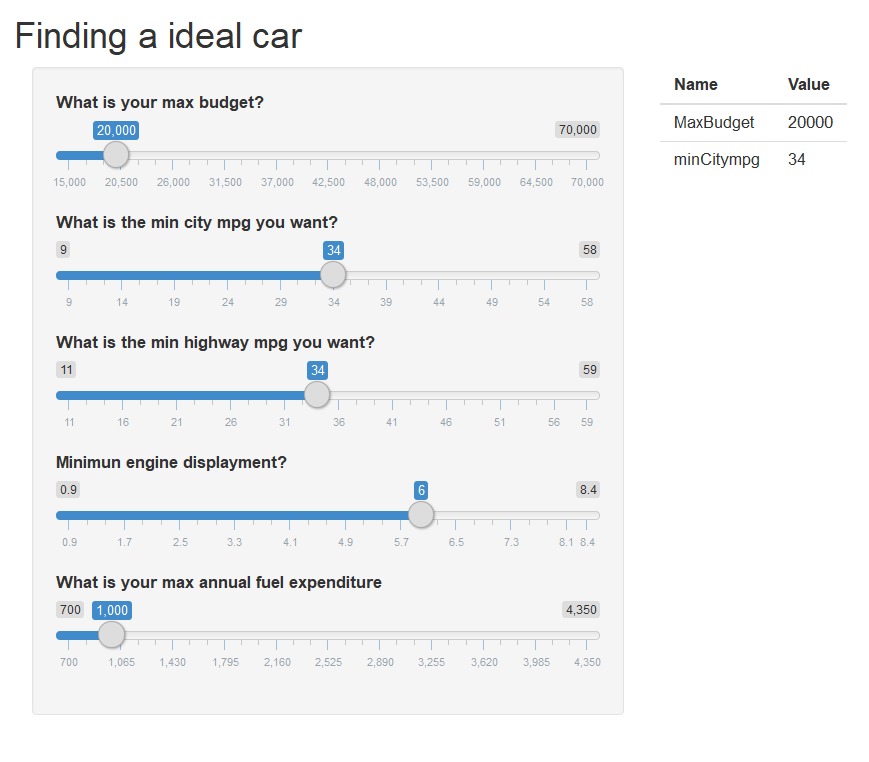
With the purpose of finding the best approach of analyzing data, we tried to compare both Excel and R. We have two datasets in .csv format, and we tried to use Pivot table to clean the data and generate some findings based on our data analysis. But due to the large size of dataset, it took long time to deal with all data. To this extent, R is actually a better tool that supporting larger datasets and has more powerful data manipulation capabilities with easier automation and faster computation. In addition, we would like to create an interactive app, so with R Shiny package, R might be a better solution for this problem. Apart from R Shiny package, there are also many available packages for us to clean the data, analyze the data and build our model. With R, it is also easier for us to test our system and to find and fix the errors.

* Test and select an approach or approaches you believe might work

## **Model Building**

* Run and evaluate the model(s)
* Calibrate model and data
* Discussion of integrating the model back to the problem
* Discussion of any findings (including assumptions, limitations, and constraints)

## **Functionality**



(need to update screenshot)

Consumers can find their ideal cars by setting different value in our Shiny App. Below is a detailed instruction on the functionality of our DSS.

Input:

Price: Consumers can set the max budget on the slider.

City mpg: Consumers can set their minimum value on the slider.

Highway mpg: the same as City mpg.

Cylinder: Consumers can have more than one option.

Engine displacement:

Drive:

Fuel Cost 08: Consumers can put their expected max annual fuel expenditure on the slider.

Fuel Type:

Eco-friendly:

Zip Code:

(screenshot 2 with output)

After inputting all their desired features, our DSS will provide them with four options that have different eco-friendly performances. Besides, by typing zip code in our app, we will actually provide the consumers with a map of different dealers. Here we use “West Lafayette” as sample in this app.

If we have more time or experience, we would like to add the real price in our dataset and make it more practical. Besides, we also hope to have more places and maps in our app.

* Discuss any R packages you found useful – R Shiny. dplyr.
* Did you have to write any conditional logic?

## **GUI Design and Functionality**

We have tested this app several times, with a success rate of \_\_\_\_\_%.

Several errors appear: 1. …

Compared with the student examples provided, we think our app is (advantages). But we do have some room for improvement. For example, (advice for improvement).

**Conclusions**

## **References (if any)**