# 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), based on R Shiny package available in RStudio, 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; optimization; utility surplus; eco-friendly car; R

## **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, manufacturers and government organizations. 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. For government organizations, as well as some non-profit organizations, they can identify the manufacturers that are promoting eco-friendly cars and they can provide those manufacturers with better policy guidelines and tax credits. This is a virtuous circle which helps in growing the market of eco-friendly vehicles.

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 the industry association and government to market and educate customers about eco-friendly vehicles.
3. This system can indirectly encourage manufacturers who are promoting eco-friendly vehicles.

After we determined our business problem, we talked to a representative of Department of Energy. The stakeholder agreed on our business problem.

## **Analytics Problem**

In order to work out the most suitable model to analyze consumers’ needs, we have to find out the key variables. 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. Given a car matches the minimum performance criteria of the customer, the customer will purchase a car that is eco-friendlier. 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 purchasing 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 and whether it can locate the nearby dealer where they can find that car model.

After our second meeting, we updated our meeting minutes with our stakeholder and notified them about the approach we are taking to solve our analytics problem. The stakeholder agreed on our approach.

## **Data**

We discussed about the various parameters that will be involved in solving the analytics problem. We agreed that we will require an authentic source of secondary data. We found out manufacturers’ car data on fueleconomy.gov.

For this project, we have 2 datasets from fueleconomy.gov, which is maintained by Department of Energy. The first one is “vehicles.csv” that include all basic information such as vehicle type, fuel type, mpg and so on. The second one is “emission.csv” with a specific focus on emissions and providing 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.

fuelCost08: annual fuel cost for fuelType1.

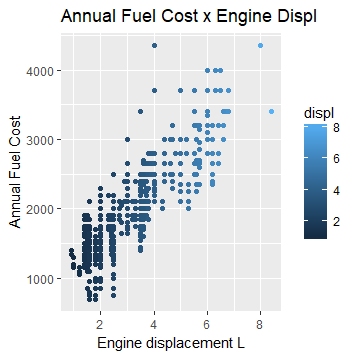
fuelType1: fuel type 1. For single fuel vehicles, this will be the only fuel. For dual fuel vehicles, this will be the conventional fuel.

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.

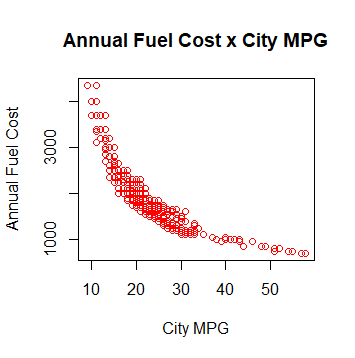
In order to work with data, we need to normalize the variables. This was required as we had to use it for mapping the performance of cars across categories. Thus a ‘z score’ was calculated for each observation. The z score is obtained by subtracting the observed value by the mean for that variable and dividing it by the standard deviation of that variable.

The relationships between different variables are tested below:

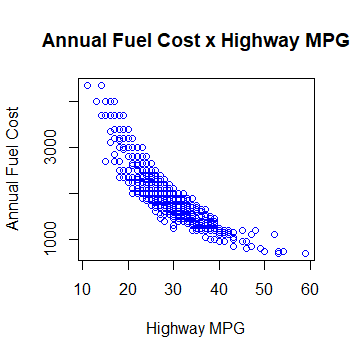
1. Displacement and FuelCost: The larger number of Engine Displacement, the higher the annual Fuel Cost.



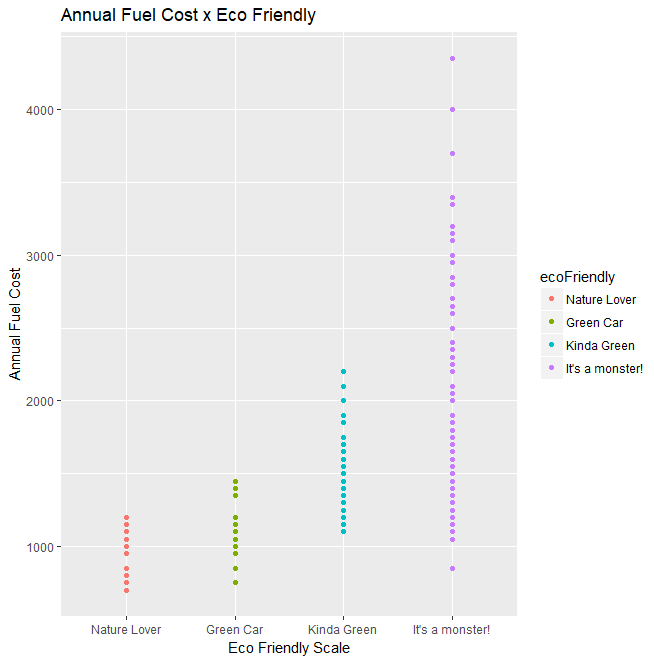
1. City08 and FuelCost: The higher the City MPG, the lower the annual Fuel Cost.



1. Highway08 and FuelCost: The higher the highway MPG, the lower the annual Fuel Cost.



1. Annual FuelCost and ghgScore (eco-friendly performance): “Nature lover” cars have a tendency for lower annual Fuel Cost. As a consequence, cars that are least eco-friendly will have a higher annual Fuel Cost.



From the insights obtained from the analysis of graph above, it is clear that we need to create a DSS for consumers to identify the eco-friendliest model that matches their requirements. There’s no need for us to refine the business and analytics problem statements.

## **Methodology Selection**

There are many approaches to solve this optimization the problem. We have to find the minimum ‘Euclidean distance’ between the input parameters and the already defined parameters.

Critical point (mathematics), Differential calculus, Gradient, Hessian matrix, Positive definite matrix, Lipschitz continuity, Rademacher's theorem, Convex function, and Convex analysis are all mathematical methods to solve the optimization problem.

The easiest method to find the Euclidean distance between two set of points on a normalized plane is the technique of Linear Programming. We will use the category that the user wants to purchase as a constraint to solve the problem.

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.

We believe that utility maximization function can be solved using a linear programming method, as it is easy to implement using Euclidean distance.

**Model Building**

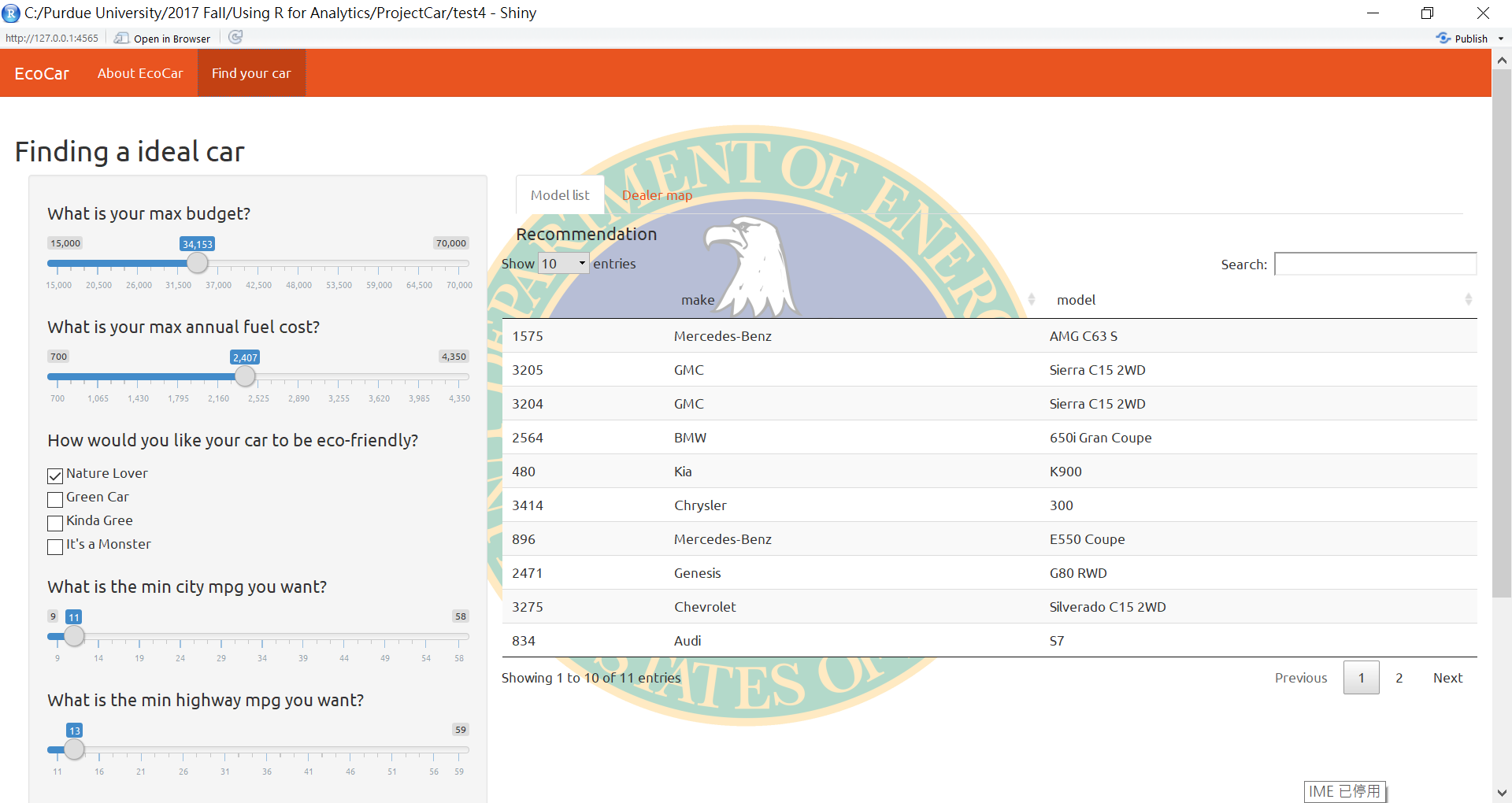
Our objective is to find the minimum Euclidean distance, it makes no sense to compare the distance of one observation with another observation. The optimization function is only supposed to calculate the distance between the input parameter with all other parameter in a normalized plane. This can be done by finding Euclidean distance and using some variables such as Price as constraints.

In the data, we have 2 variables, “ghgScore” and “score”, both indicating the eco-friendly features of the car, so we combined them into a single metric called “ecoFriendly”. Apart from calibrating the data, we don’t have to calibrate the model.

We added new variables to the problem which was normalized, the Euclidean distance between this new input that we obtained after entering the various parameters describing the car and all other variables in dataset was calculated. The nearest observed normalized variable, i.e. the one with shortest Euclidean distance gave us the car model which best suited user requirements.

The limitations of the model, given the nature of problem is that we cannot verify the accuracy of the model. The model works and gives us the closest possible match of car based on the parameters that the user has defined.

## **Functionality**



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. For both City mpg and Highway mpg, we highly recommend consumers to set a relatively higher value. This indicates more cost-efficient performance.

*Engine displacement*: Consumers can select the number of engines on the slider.

*Ecofriendly*: Consumers can decide their ecofriendly performance on the slider from scale 1 to 3. (1 is the eco-friendliest and 3 is the least eco-friendly)

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

*Zip Code*: Consumers can input their zip code to find the dealers nearby. So far, we will only provide dealers around West Lafayette.

After inputting all their desired features, our DSS will provide them with a recommendation list (with 10 car models). Besides, by default, we will actually provide the consumers with a recommended dealer in the map near Lafayette.

In our DSS, the useful R packages we found are Shinydashboard, Leaflet, and dplyr. By using Shinydashboard, we can have a nice dashboard in our Shiny App. It helps us with the theme on top of “Shiny”. Leaflet gave us opportunity to present a map in our app. While dplyr help us work with data frame like objects, both in memory and out of memory.

We have to use conditional statements to determine the optimal output based on how the customer describe his preference for the vehicle. We use “if…else” statements to get the desired result.

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 hope to help consumers in different places to find the dealers near them.

## **GUI Design and Functionality**

We have tested this app several times with a success rate of 100%. However, some errors or problems have been diagnosed during our tests. There are mainly three problems:

1. In Format and Layout, it’s hard for us to create a recommendation list. Calling the input from ui.R file failed several times, and our team members spent a long time in fixing this problem.
2. We faced a version problem when we tried to combine codes from two members.

Compared with the student examples provides, we think our GUI design is better than those examples.

**Conclusions**

Buying an eco-friendly car is a big deal, not only for consumers, but also for the whole industry and the government. With our DSS, we provide a solution to this problem.

In order to solve our established business and analytics problem, we utilized our business analytics skills as well as our data analysis knowledge. Based on our model, we created our DSS with RStudio and presented our solution with a user-friendly and interactive interface. With such visual interface, decision makers can have a better decision making and are able to find their car models by using our R Shiny app.

## **References**

EPA. (2017). *Carbon Pollution from Transportation*. Retrieved from https://www.epa.gov/air-pollution-transportation/carbon-pollution-transportation

Union of Concerned Scientists. (2017). *Vehicles, Air Pollution, and Human Health*. Retrieved from http://www.ucsusa.org/clean-vehicles/vehicles-air-pollution-and-humanhealth#.Wc6rCkyZPeQ