**Midterm Project Proposal**

Yujia Wang

**Personal Statement**:

My career goal is becoming a data analyst and my ideal work field is sales company. I am especially interested in [automobile](javascript:;) [industry](javascript:;).

The bigger the company, the more they care about their reputation and take on more social responsibilities. So in the midterm project, I focus on finding some patterns about the CO2 emission of cars in various conditions because CO2 quantity is an important index to identify whether a car is “environmental friendly” or not. Finding greener parts can guide companies in developing cleaner cars. Also, a good reputation can [stimulate](javascript:;) [sales](javascript:;). For example, Tesla is using new energy vehicles as their selling point.

**Question**:

My main dataset is about CO2 emissions of cars with different features. I want to figure out the discrepancy of CO2 emission in different vehicle class, fuel type and transmission type. The dataset also has other [variable](javascript:;)s such as engine size, cylinders and fuel consumption in a variety of roads.

Besides, I will merge fuel consumption ratings such as CO2 rating and smog rating into the main dataset if it is necessary. This data comes from another website.

In conclusion, I will fit multilevel models in order to find out the type of car that consumes the least carbon dioxide, while consumes the least gas for the same mileage. This is both an environmental concept and a selling point to attract consumers, which they can spend less at the pump.

**The data sources**:

I got the data from Kaggle and Government of Canada.

Link 1:

<https://www.kaggle.com/debajyotipodder/co2-emission-by-vehicles?select=CO2+Emssion_Canda.csv>

Link 2:

<https://open.canada.ca/data/en/dataset/98f1a129-f628-4ce4-b24d-6f16bf24dd64#wb-auto-6>

**Proposed Timeline of work**:

EDA: 11/5-11/11

Data Processing: 11/12-11/18

Modeling and Validation: 11/19-11/25

Write up: 11/26-12/2