**FTEC 6321 Final Project**

**Yuming Chang**

**Problem Identification:**

* For almost all the investment companies or fund managers, they all interested in getting the first hand of the stock information so that they can get ahead of the game. That's the motivation for this project.
* In this project, I will find the hidden factors or leading indicators from demography to predict the waste data in The United States, and then find the correlation between the companies' financial performance and the increment of waste in the USA to better predict the revenue and stock price.
* The demography indicators might not be intuitively connected to the final result, but with large data involved and processes being logical. I have confidence that the outcome will be effective in making a good investment decision.

**Data Gathering:**

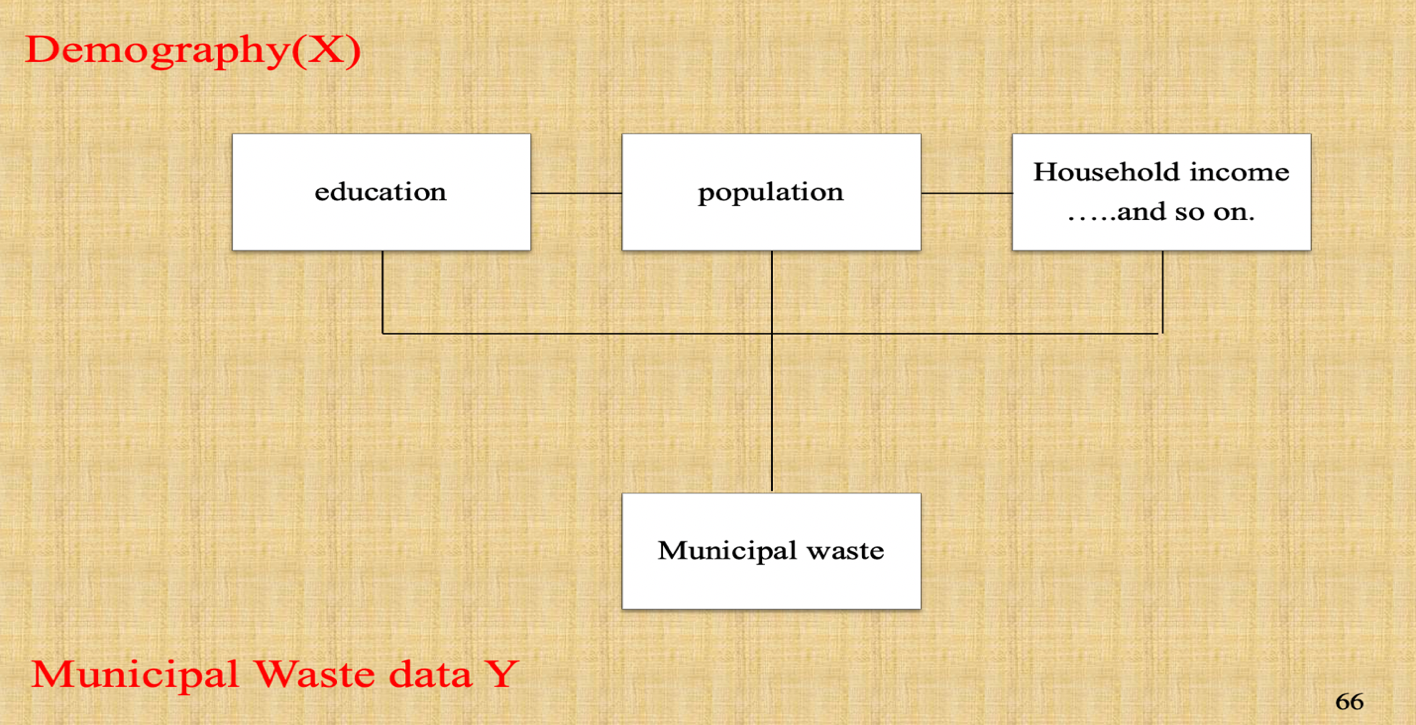
* All the data used in this project will be stored in Github:

<https://github.com/yxc190029/adv-stats-and-probability>

* The demography data I use for the features are respectively "population", "gender", "household income(median)", "education", "lifespan", "race". I also add 'oil usage in USA' as an independent variable.
* The data in this project is gathered from diverse sources, most of the data in the first layer are from [Census.gov](http://census.gov/)  and [epa.gov.](https://www.epa.gov/facts-and-figures-about-materials-waste-and-recycling/studies-summary-tables-and-data-related) (The United States Environmental Protection Agency)

In addition, the US oil consumption data is from [ycharts.com](https://ycharts.com/indicators/us_oil_consumption).

**First layer:**



* For the second layer, I will classify the municipal data into multiple genres, and see the correlation between these genres and the revenue performance. The classification will be "Combusted", "landfilled", "recycled".
* The final goal of this project will be predicting the upward or downward of the stock price, so to make a comprehensive project, other indicators for the companies' management will be considered. The financial factors that I choose as in independent variables would be "Revenue", "EPS", "Ebitda", "income", "capital change", "tax".
* The data for the financial data is from [Warton wrds.](https://wrds-www.wharton.upenn.edu/)

**Second Layer:**

