DS5220
Project Proposal
Evan Chan
001245773

Publicly Available Dataset:

The publicly available dataset that we intend to explore is "Supermarket store branches sales analysis." This extensive dataset includes various measurements relating to different stores of a supermarket company. The available measurements include:

- store ID (the ID of a particular store)
- store area (area of a particular store in yard squared)
- items available (number of different items available at a particular store)
- daily customer count (number of daily customers on average over a month)
- store sales (the US dollar amount that a particular store made)

With funding, we hope to perform an exploratory data analysis on this dataset that we are confident is a feasible undertaking. The data measurements are straightforward and will allow for various hypothesis that can be explored with some low-risk options and a few high-risk, high-reward teasers.

Exploratory Data Analysis (EDA) Goals:

The goals for our exploratory data analysis (EDA) involve gaining insight into the supermarket, a self-service shop that is ubiquitous in American culture. The supermarkets and grocery stores industry accounts for the largest food retail channel in the United States, making the applications of answers to any of the following questions posed in our EDA essential and possibly profitable insights. Questions that we believe can be answered using methods developed in this class include:

- What relationships do the variables have?
  - clean data frame
  - pairs plot
  - heat map
- Does the daily customer count have an influence on the store sales?
  - assign variables to model
  - fit model into simple regression
  - assess model accuracy in regression analysis
- Does the store area have an influence on the items available?
  - assign variables to model
  - fit model into simple regression
  - assess model accuracy in regression analysis
- Does the daily customer count and the items available have an influence on store sales?
  - assign variables to model
  - fit model into multiple regression
  - assess model accuracy in regression analysis
- Does the store area and the items available have an influence on daily customer count?
  - assign variables to model
  - fit model into multiple regression
  - assess model accuracy in regression analysis