Clay Heidrich

March 2020

Tuesday Project Class

**Sales Analysis Project Plan**

# **Project Overview**

In my project I will use Python, Pandas & Python Matplotlib to analyze and answer some basic business questions relating to a data set. The data will be gathered from 12 different .csv files (one per month) containing store purchases information relating to a year’s worth of sales data. First I will combine the separated files into one data set. Once concatenated, I will try to answer a few basic questions that would be helpful for a business to improve their business strategy for the coming year based on my analysis.

Question #1: What was the best month for sales?

Question #2: What city sold the most product?

Question #3: What time should we display advertisements to maximize the likelihood of purchases?

Question #4: What products are most often sold together?

Question #5: What product sold the most? Why?

I will use several different graphs to visualize and represent the answers to the questions above.

**Technical Summary**

* Pandas
* Python
* Jupyter notebook
* Matplotlib

# **Feature List**

- Concatenating multiple csvs together to create a new DataFrame (pd.concat)

- Adding columns

- Parsing cells as strings to make new columns (.str)

- Using groupby to perform aggregate analysis

- Plotting bar charts and lines graphs to visualize results