QTM2000   Name_____
Case Studies in Business Analytics
Professor Mathaisel
Fall 2021

<center>Final Project</center>

<center>**Individual Assignment**</center>

<center>The following pledge must be on your exam cover sheet and signed.</center>

<center>I pledge my honor that I have neither received nor provided unauthorized assistance during the completion of this work. Please Initial:_____</center>

The objective of this project is to gain insight into customer purchases during "Black Friday". The analytical method used is Association Rules (arules). For this project, you are being asked to write an "Executive Summary" to accomplish that objective using the dataset and its accompanying R script provided. Your summary should be brief, but it must include: Introduction; Analysis; and Conclusions. The analysis is divided into two parts: 1. Exploratory Data Analysis (EDA); and 2. Association Rules.

<center>

Part 1.   50 points
Part 2.   50 points
_____
        100 points

</center>

Introduction: (Provided by: Dean Abate, Analyst at Aon, Chicago,)

"The origins of 'Black Friday' stem not from a day filled with shopping, discounts, and a turn of the holiday season, but rather with a **financial crisis**! The first recorded use of the term "Black Friday" was recorded on September 24th, 1869 when two Wall Street businessmen, Jay Gould and Jim Fisk, decided to artificially inflate the price of gold and attempted to sell it for profit. As a result of their nefarious actions, on that specific Friday in 1869, the price of Gold dropped and the United States plunged into a state of financial devastation.
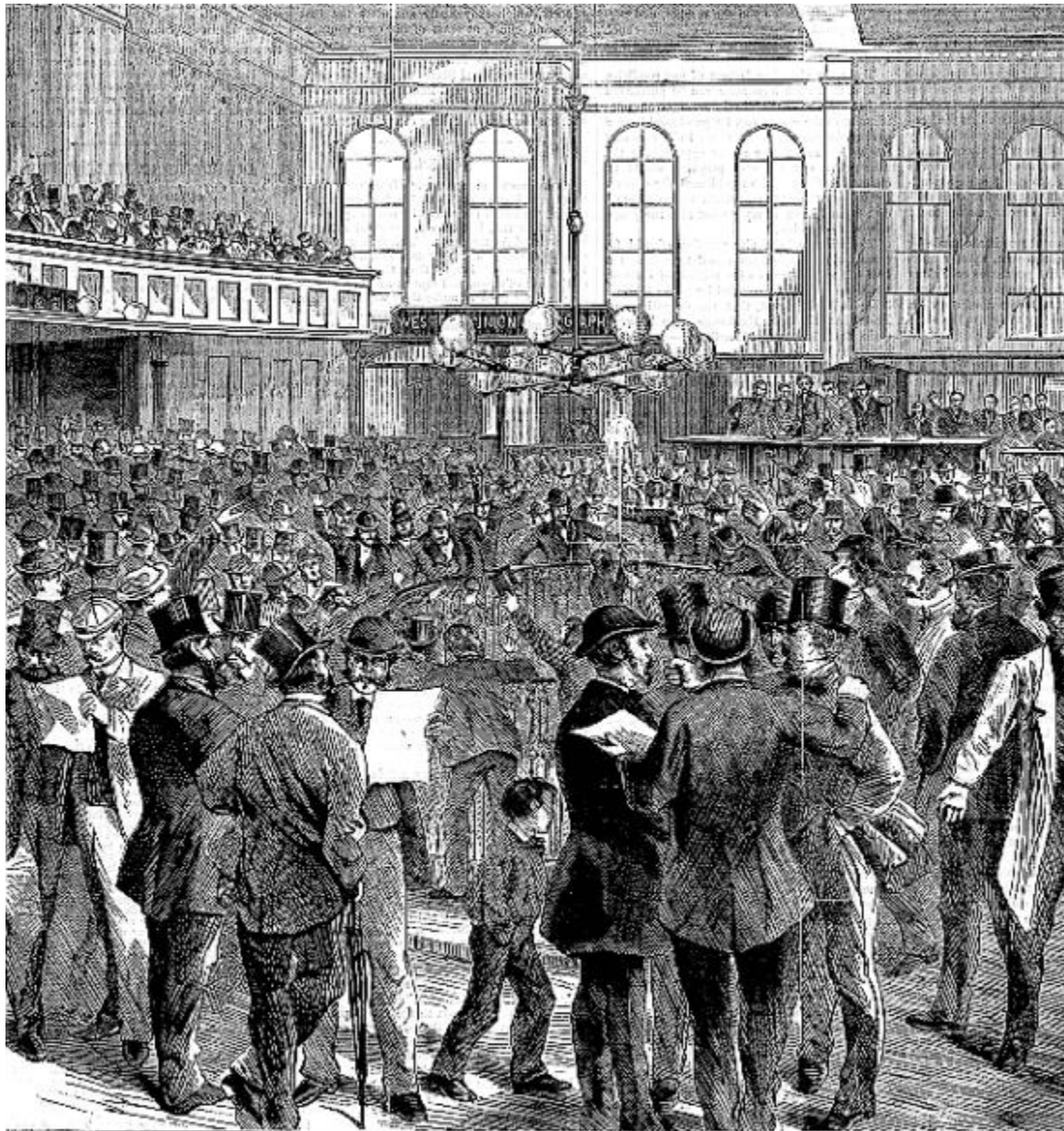
SCENE IN THE GOLD ROOM, NEW YORK CITY, DURING THE INTENSE EXCITEMENT OF FRIDAY, SEPTEMBER 24, 1869.

Various stories exist regarding the first recorded use of the term as it relates to holiday shopping, but its connotation continued to keep a negative stigma associated with it until the late 20th century. Black Friday's relation to consumerism was first derived from 1950s Philadelphia. Philadelphia suburbanites descended on the city after the Thanksgiving holiday to watch the traditional Army/Navy college football game and take advantage of sales and promotions brought about by the influx of spectators to the city. Philadelphia Police officers, who were assigned to work that weekend, coined the term due to their long grueling shifts and the massive amounts of people and shoppers. Philadelphia businesses also started to use the term to describe the long lines and shopping mayhem at their stores. Although the term Black Friday originally represented the pitfalls of two Wall Street businessmen and the mayhem in Philadelphia following Thanksgiving, it is now familiarly known as the busiest shopping day of the year. One possible explanation for the term as it relates to consumers and retailers is that "Black Friday" represents the first day of the year in which businesses were turning profits and accounting was done on a hand-written ledger. As described in the Oxford Dictionary, "The use of colors in accounting refers back to the bookkeeping practice of recording the credit side of an account in a ledger in *black* ink and the debit side in *red* ink." Hence the name Black Friday became associated with business debits overtaking credits. Although this idea might make sense, the claim hasn't been completely verified.

Black Friday as we know it today is an extravaganza of sales, promotions, and long lines outside stores. Retailers such as Target, Best Buy, Amazon, and many others look forward to this day every year with the hopes that consumers will take advantage of deals. Black Friday has also spawned other retail holidays, such as "Cyber Monday", "Small-Business Saturday", and "Giving Tuesday." "

Objective of the Project:

A retailers' goal is to attract customers for Black Friday deals, which has become a competition among retailers to get the customers to their online platforms or physical retail stores. Data science techniques help retailers with proper insight into this goal using historical data and Association Rules. The project is divided into two main parts: exploring the Black Friday data, and using Association Rules to find possible associations between purchases. The idea is to try to get consumers who purchase one product/service (Antecedent) to also purchase another (Consequent).

Data Dictionary:

The dataset is comprised of 550,000 observations about Black Friday shoppers in a retail store. It contains different kinds of data: numerical and categorical. It also contains missing values (source: Mehdi Dagdoug).

The "Black Friday" dataset contains 12 different columns:

- *User_ID*: Unique identifier of shopper.
- *Product_ID*: Unique identifier of product. (No key given)
- *Gender*: Sex of shopper.
- *Age*: Age of shopper split into bins.
- *Occupation*: Occupation of shopper. (No key given)
- *City_Category*: Residence location of shopper. (No key given)
- *Stay_In_Current_City_Years*: Number of years stay in current city.
- *Marital_Status*: Marital status of shopper. (No key given) (See the R script for coding.)
- *Product_Category_1*: Product category of purchase.
- *Product_Category_2*: Product may belong to other category.
- *Product_Category_3*: Product may belong to other category.
- *Purchase*: Purchase amount (No key given) (Currency is not certain).

When we look at the first few rows of our dataset, we will see that each row represents a transaction, or item purchased by a specific customer. This same customer may also purchase another product, resulting in a different transaction. This will come into play later on when we group all transactions by a specific *User_ID* to get a sum of all purchases made by a single consumer.

One caution or critique we can make regarding this dataset is that there isn't a key (Data Dictionary) for many of the attributes, especially the Product_IDs and the items they represent (i.e., we can't attribute the Product_ID P00265242 to an item that is easily recognizable). In reality, we would want to have another dataset that provides the name of an Item and its Product_ID, and then join it to our existing dataset. This won't affect our EDA, but it would be more useful explaining the results of the Association Rules algorithm if we could link it to an actual product, instead of referring it as P00265242. But, alas, those keys were not provided.

The Project:

There are two parts:

Part 1. Exploratory Data Analysis (EDA).

Briefly and concisely tell a story of what you visualize from the results and corresponding graphs generated by the R script. Use as few words as possible, but convey your story.

Part 2. Association Rules

Summarize the results of the arules model.

   a. A brief discussion of what the model is trying to accomplish and how the technique works.
   b. Discuss the resulting analysis.