CONSUMER PURCHASE BEHAVIOUR ANALYSIS

TABLE OF CONTENTS

ABSTRACT

Black Friday (BF), the day after Thanksgiving, is reported to be a day of both extreme bargain shopping as well as one of the busiest shopping days of the year. BF has become notorious in the media and general public for egregious consumer behaviour. Research regarding BF remains relatively new and a focus on consumer misbehaviour on BF is slowly growing. Utilizing two major theoretical perspectives (General Aggression Model and Stimulus-Organism-Response Model), the goal of the present research was intended to explore potentially influential personal and situational factors that affect consumer behaviour on BF. From a survey of BF shoppers, in which four hypothetical BF scenarios were manipulated by two levels (goal blockage or goal fulfilment), we found that public self-consciousness and self-control are negatively related to consumer misbehaviour and that the behaviours of fellow consumers and goal blockage evoke negative emotions. We also found that negative emotions are positively related to consumer misbehaviour. Our results support the General Aggression Model in explaining consumer misbehaviour on BF.

1. INTRODUCTION

1.1 Purpose

The main point of doing consumer behavior analysis is to know the trends of how people make purchase decisions with regard to a product, service, or organization. We can understand customers on basis of their buying habits and also understand purchasing habits according to age groups, occupation and city category. Black Friday is one of the most days in which a shopping fever spreads all over the world. The popularity of Black Friday witnessed a huge increase with the appearance of online shopping and the fact that individuals started to prefer online shopping over regular shopping. This study examined how aware the Jordanian consumers of black Friday and to what extent it managed to change their shopping behaviour.

1.2 Black Friday: History and Origins BF Black Friday (BF): It is traditionally regarded as the first day of the shopping season for the winter holiday (Fletcher, 2009). Such a peculiar name is hypothesized to originate from a dated method of recording business accounts. Profits were recorded in black ink thus many business owners considered BF as a turning point in sales for the year (Fletcher, 2009). Its origins date back as far as the late 19th century in Philadelphia and it has remained an unofficial holiday since (Fletcher, 2009).

1.3 Consumer Behaviour Theory: Consumer behaviour theories generally differentiate between rational (Ho, Tang & Bell, 1998) and irrational (Dholakia, 2000) behavioural models. Most commonly associated with irrational models is the assumption that consumers are most likely to exhibit varying degrees of negative behaviour (i.e., damaging other consumers, themselves, or employees). Consumer misbehaviour, within the context of this research, is defined as “the extent to which a customer deliberately behaves in a way that violates the norms and unwritten rules of an individual service setting in a negative fashion” (Reynolds & Harris, 2009, p 321). Definitive elements of BF make this day especially vulnerable to enabling consumer misbehaviour. The hours of operation for businesses (starting as early as the evening before BF), products that are offered at 2 severe discounts, the sheer volume of advertising, self-selection, and rowdy customers may each play a role in promoting consumer misbehaviour.

2. LITERATURE SURVEY

2.1 Promotions: Promotions are important to consider for an analysis of BF. Within the context of BF, promotions are a key element that set this day apart from other retail sales during the year. BF is known for having longer hours of operation in stores, significant discounts, and offering other special deals. The effect of promotions on consumer behaviour has been studied in several capacities, including the effects of promotions on the consumer as well as the influence of consumer intent on promotion effectiveness. Consumer behavioural trends vary depending on the purchasing and product intent of the shopping trip. Most relevant to BF, consumers who intended to shop specifically for price specials tended to be more sensitive to flyers and advertised promotions, as was reflected in their purchases. Consumers with the intent to do other types of shopping, such as for specific items or a major shopping trip for which they intended to purchase multiple items, were less sensitive to flyers and advertised promotions. In regard to consumer expectations and the importance of advertising on BF, we can assume that most shoppers intend to purchase specific items, or at the very least are browsing for items that are advertised at a desirable deal due to the nature of BF (i.e., heavily promotion based). If consumers have different reactions to promotions, based on the promotions themselves and pre-existing perceptions of fairness and the products, then perhaps outward behaviour may vary as well.

2.2 Consumer Expectations: Regardless of the occasion, when consumers visit retail sites with an expectation to purchase a product, they intend to trade an amount of money for a product in an exchange that they perceive as being fair. One can argue that due to BF's portrayal in the media as a day with extensive discounts, consumers have higher expectations for the exchange. One complication on BF arises when expectations for advertised products or interactions with fellow consumers are not met. The basic generalization is that individuals form expectations about events and how people should behave in the present and future. Huang, Lin and Wen (2009) studied other consumer behaviour and showed that consumers were most affected by the behaviour of other consumers when they perceived that management could and should control the deviant behaviour and that the behaviour violated some social standard of behaviour. Within the context of BF, outward displays of consumer misbehaviour are consistently reported in the media. Difficulty obtaining items (due to stock outs or long lines) and aggressive consumer behaviour are commonly depicted on BF In relation to Huang et al.'s (2009) research, consumer frustration on BF may result partly from the expectation that management should handle the retail scene more efficiently and appropriately than is typical. In organizations such as retail, front-line employees are crucial to the exchange as they are constantly and directly interacting with consumers.

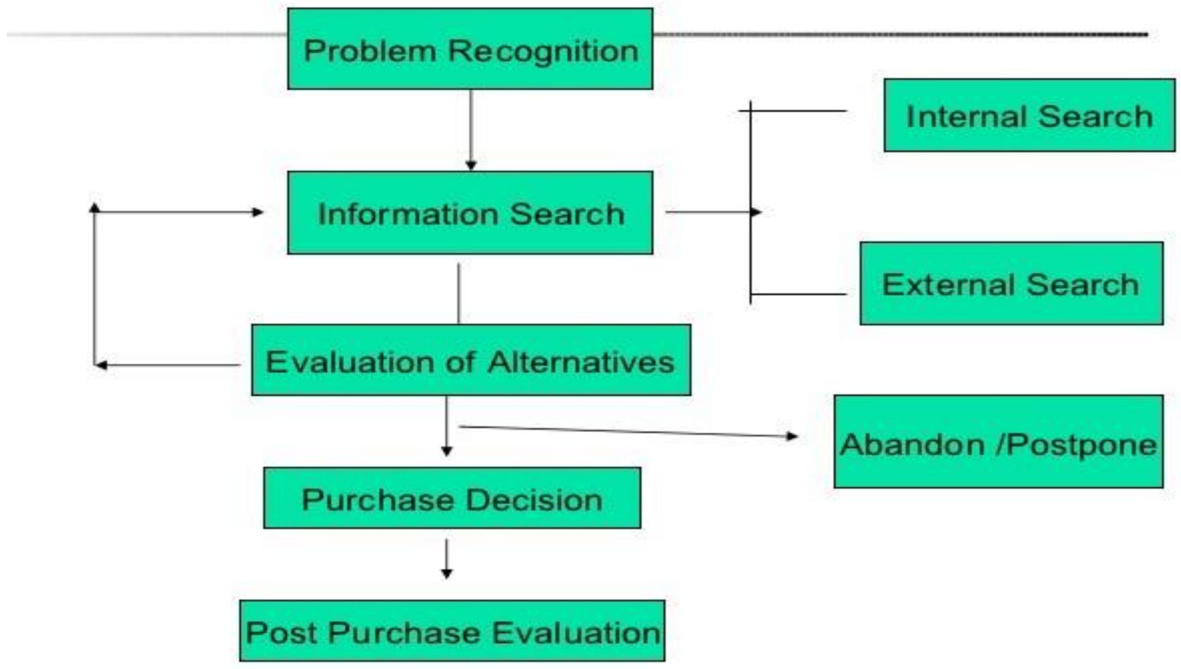
On BF these roles become crucial, as consumers are perceived by employees, other consumers and the media as being more likely to be aggressive, anxious, and determined to find promoted items. Research has shown that although employee functionality is important in complementing business success, services capes (defined as aspects of visual merchandising and dress variables) are also important in influencing consumer attitudes towards businesses in certain situations.

By administering questionnaires to student customers and frontline employees regarding negative employee/customer interactions, McCollKennedy et al. (2009) were able to empirically support that customer rage incorporates 8 negative emotions such as wrath, which is expressed through physical, verbal, or nonverbal actions that are generally displaced on others. These emotional stressors may be especially influential on BF when stores are packed with more consumers, employees, and products. Despite the hypothesized influence of environmental and employee factors on consumer expectations on BF, the role of the consumer is a crucial part of this analysis. By fully understanding what influences consumers to behave negatively or positively, employers and businesses may be able to create environments that are not conducive to egregious Black Friday misbehaviour.

2.3 Emotion: Emotion is especially prevalent in regards to current BF research. Emotion has been previously defined in many ways. A common definition of emotion is "a valenced affective reaction to perceptions of situations". Much like other natural conceptualized items or events, categories of emotion are also individually formed by experiences and pre-existing stereotypes. Shaver et al. (1987) theorized that emotions are organized in a hierarchy with emotions used on a daily basis (i.e., sadness, joy, anger, love etc.) being the most accessible. Emotions identified as being less accessible, within the hierarchy, were divided into two categories: positive and negative emotions. Joy was included within positive emotions and anger was included within negative emotion. Moreover, emotions related to anger were distinguished as the highest in potency whereas emotions related to joy were distinguished as an intermediate degree of potency. In the context of BF, this may help to explain the affinity of consumers to readily display frequent acts of aggression if they are innately more prone to be affected by anger-related emotions. This data, as well as other research literature, influenced the development of a large pool of emotion items used by Lennon et al. (2011) and the current research.

3. THEORETICAL ANALYSIS

3.1 Block Diagram



The above block diagram shows the flow of analysis

3.2 Hardware/Software Designing: The model is designed using Jupyter Notebook in Anaconda which is an open-source distribution of the python programming language for scientific computing that aims to simplify package management and deployment. The application building is done using flask which is a web framework that provides tools, libraries and technologies that allows the developer to create a web application. The web application can be a web page, blog, a commercial website.

3.3 Hypothesis

City Level Hypotheses:

1. City Type and Size: Urban or Tier 1 cities should have higher sales because of the higher income levels of people there.

2. Population Density: Cities with densely populated areas should have higher sales because of more demand.

3. Younger Population: Cities with younger populations might have higher tendency to spend more on Black Friday.

Customer Level Hypotheses:

1. Income: People with higher income should spend more on products.
2. Age and Gender: Men with ages ranging from 25 to 40 should spend more on techlogical products.
3. Family Size: Families should be more contained on spending's, just buying the best offers and only needed products.
4. Purchase History: Customer with a purchase history should be more willing to purchase more products on this day.

Store Level Hypotheses:

1. Location: Stores with a location in well moved streets should have better sales.
2. Size: Bigger stores with higher stores and variety of products should have better sales.
3. Competition: Stores with no competitors nearby must have the highest sales.
4. Marketing: Do stores which spend more on marketing should have the best sales results

Product Level Hypotheses:

1. Category: Most clients should be looking to buy technological products;
2. Price: Customer will spend more on products with higher discounts
3. Advertising: More advertised products should sell more
4. Visibility: More visible products should sell more
5. Brand: Clients will invest more on already known brands

4. EXPERIMENTAL INVESTIGATIONS

4.1 Available Data: This is the current data available:

Data

Variable	Definition
User_ID	User ID
Product_ID	Product ID
Gender	Sex of User
Age	Age in bins
Occupation	Occupation (Masked)
City_Category	Category of the City (A,B,C)
Stay_In_Current_City_Years	Number of years stay in current city
Marital_Status	Marital Status
Product_Category_1	Product Category (Masked)
Product_Category_2	Product may belongs to other category also (Masked)
Product_Category_3	Product may belongs to other category also (Masked)
Purchase	Purchase Amount (Target Variable)

There is some information related to the customer such as age group, sex, occupation and marital status. On the other hand, we have data on the city's size and how many years the customer has lived in it whereas on the product's side there is only information regarding the categories and the amount spent.

It is my belief that Gender, Age, City_Category, Product_Category_1 are the predictors that will influence more the amount spent by a customer on this day.

4.2 The target variable (Purchase):

Name	Type	Subtype	Description	Segment	Expectation
User_ID	Numeric	Discrete	User ID	Customer	Low_Impact
Product_ID	Numeric	Discrete	Product ID	Product	Low_Impact
Gender	Categorical	Nominal	Sex of User	Customer	High_Impact
Age	Categorical	Ordinal	Age in bins	Customer	High_Impact
Occupation	Categorical	Nominal	Occupation (Masked)	Customer	Medium_Impact
City_Category	Categorical	Ordinal	Category of the city (A,B,C)	City	High_Impact
Stay_In_Current_City_Years	Categorical	Ordinal	Number of years stay in current city	City	Low_Impact
Marital_Status	Categorical	Ordinal	Marital Status	Customer	Low_Impact
Product_Category_1	Categorical	Nominal	Product Category (Masked)	Product	High_Impact
Product_Category_2	Categorical	Nominal	Product may belongs to other category also (Masked)	Product	Low_Impact
Product_Category_3	Categorical	Nominal	Product may belongs to other category also (Masked)	Product	Low_Impact
Purchase	Numeric	Continuous	Purchase Amount (Target Variable)	Product	NAN

4.3 Analysis

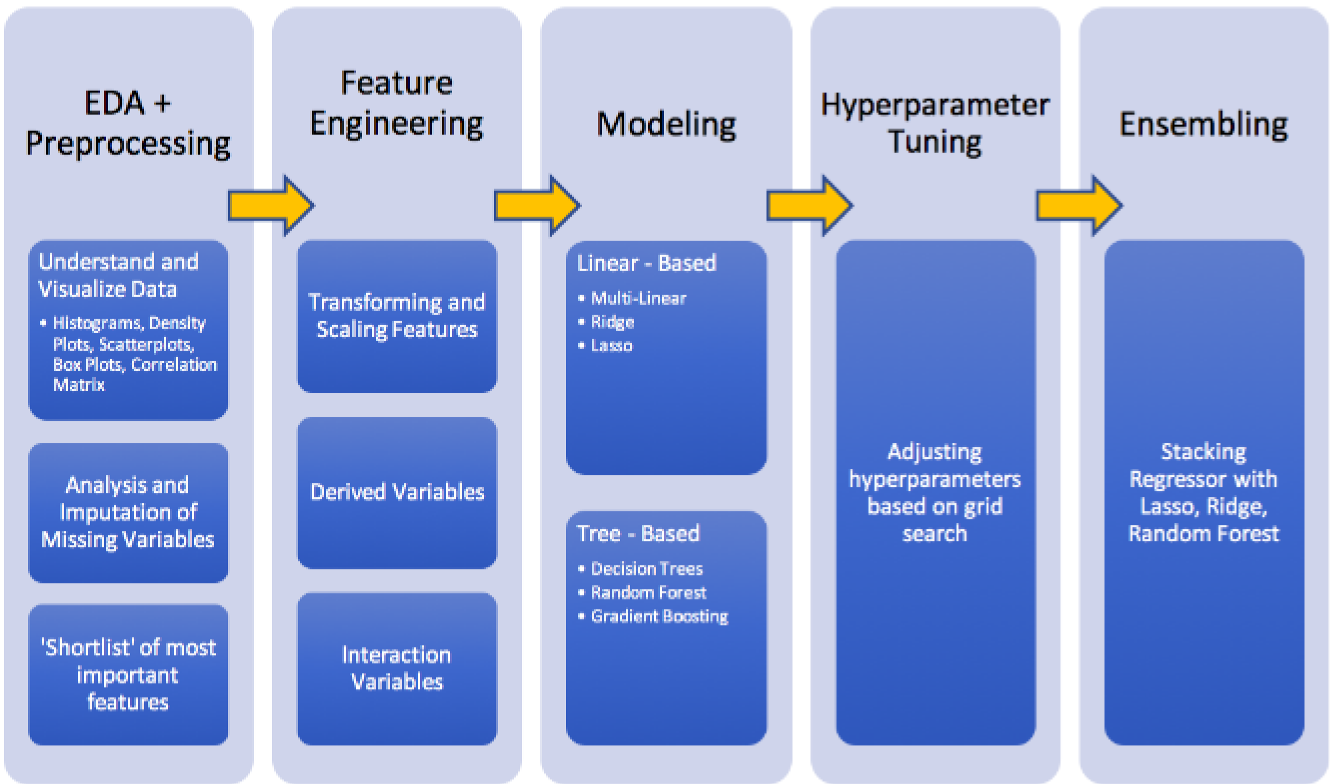
	Highest purchase	Lowest purchase
Amount	23961.0	12.0
Product ID	375436.0	142.0
Age group	Above 55 years	0-17 years
Gender	Male	Female
City category	City C	City A
Occupation	20^{th} category	0^{th} category

Above table represents the results of the model we have built using machine learning algorithms. It's been used to identify the highest purchase and lowest purchase made by the customer and also to know the highest and lowest amount paid, which product have the highest and lowest price, which age group is highest spender, which gender made the highest payment and the which gender made it.

5. FLOW CHART

Our goal as a Data Scientist is to identify the most important variables and to define the best regression model for predicting out target variable. Hence, this analysis will be divided into five stages:

1. Exploratory data analysis (EDA);
2. Data Pre-processing;
3. Feature engineering;
4. Feature Transformation;
5. Modeling;
6. Hyper parameter tuning
7. Ensembling.



6. RESULTS

1. The majority of the sample was students who had almost no financial responsibilities like bills, rent and household needs. This result was seen to be logical given that mostly students who invest their allowance in buying items that they need.

2. Females formed the majority of the sample and also that result was seen to be logical given that female are known for their interest in shopping more than males and they are most likely to go shopping for many reasons and in all occasions.

3. There appeared to be a good level of awareness from Jordanian consumers regarding the idea of Black Friday. Respondents seemed to know the meaning of the concept and were aware of what comes along with this kind of days.

4. All statements in the questionnaire were answered correctly and in a positive way which indicates that individuals understood each and every statement and answered them based on their true experience.

This section displays the accuracy of the model that we have built for the “BLACK FRIDAY” dataset using regression algorithms during this project. Regression algorithms accuracy is good enough to confidently predict the purchase amount a customer makes when given all the features required to predict. This model which is built by following all the necessary steps of machine learning algorithm is capable of predicting the purchase amount accurately.

7. ADVANTAGES AND DISADVANTAGES

Advantages

Consumer purchase behaviour analysis helps how people make purchase decisions with regard to a product, service, or organization. It helps **marketers** decide how to present their products in a way that generates maximum impact on consumers. **Consumers** have the ability to compare products and prices at various stores. Packaged products are uniform across multiple stores, differing only by brand name and style. Consumers have to use a degree of restraint in their shopping to compare the prices of products across multiple stores. And also, it helps companies to understand how the decision to buy was made and how they hunted for the product. These information help companies and business managers to know the reasons behind the purchase or rejection of a product or service by the customer.

Disadvantages

Understanding Customer Expectations: First and foremost, the biggest challenge is understanding the changing expectations of the customers. Most of the company's target the audience based on interests, age, gender, etc. But the desires of the customers change while purchasing a product or service.

8. APPLICATIONS

One of the popular applications of AI is Machine Learning (ML), in which computers, software, and devices perform via cognition (very similar to human brain). Herein, we share few examples of machine learning that we use everyday, and perhaps have no idea that they are driven by ML.

A. Virtual Personal Assistants

Machine learning is an important part of these personal assistants as they collect and refine the information on the basis of your previous involvement with them. Later, this set of data is utilized to render results that are tailored to your preferences.

Virtual Assistants are integrated to a variety of platforms. For example:

- Smart Speakers: Amazon Echo and Google Home
- Smartphones: Samsung Bixby on Samsung S8
- Mobile Apps: Google Allo

B. Predictions while Commuting

Traffic Predictions: We all have been using GPS navigation services. While we do that, our current locations and velocities are being saved at a central server for managing traffic. This data is then used to build a map of current traffic. While this helps in preventing the traffic and does congestion analysis, the underlying problem is that there are less number of cars that are equipped with GPS. Machine learning in such scenarios helps to estimate the regions where congestion can be found on the basis of daily experiences.

Online Transportation Networks: When booking a cab, the app estimates the price of the ride. When sharing these services, how do they minimize the detours? The answer is machine learning. Jeff Schneider, the engineering lead at Uber ATC reveals in an interview that they use ML to define price surge hours by predicting the rider demand. In the entire cycle of the services, ML is playing a major role.

C. Videos Surveillance

The video surveillance system nowadays are powered by AI that makes it possible to detect crime before they happen. They track unusual behaviour of people like standing motionless for a long time, stumbling, or napping on benches etc. The system can thus give an alert to human attendants, which can ultimately help to avoid mishaps. And when such activities are reported and counted to be true, they help to improve the surveillance services. This happens with machine learning doing its job at the backend.

D. Social Media Services

From personalizing your news feed to better ads targeting, social media platforms are utilizing machine learning for their own and user benefits. Here are a few examples that you must be noticing, using, and loving in your social media accounts, without realizing that these wonderful features are nothing but the applications of ML.

. *People You May Know*
. *Face Recognition*
. *Similar Pins*

E. Email Spam and Malware Filtering

There are a number of spam filtering approaches that email clients use. To ascertain that these spam filters are continuously updated, they are powered by machine learning. When rule-based spam filtering is done, it fails to track the latest tricks adopted by spammers. Multi-Layer Perceptron, C 4.5 Decision Tree Induction are some of the spam filtering techniques that are powered by ML.

F. Online Customer Support

A number of websites nowadays offer the option to chat with customer support representative while they are navigating within the site. However, not every website has a live executive to answer your queries. In most of the cases, you talk to a chatbot. These bots tend to extract information from the website and present it to the customers. Meanwhile, the chatbots advances with time. They tend to understand the user queries better and serve them with better answers, which is possible due to its machine learning algorithms.

G. Search Engine Result Refining

Google and other search engines use machine learning to improve the search results for you. Every time you execute a search, the algorithms at the backend keep a watch at how you respond to the results. If you open the top results and stay on the web page for long, the search engine assumes that the results it displayed were in accordance to the query. Similarly, if you reach the second or third page of the search results but do not open any of the results, the search engine estimates that the results served did not match requirement. This way, the algorithms working at the backend improve the search results.

H. Product Recommendations

You shopped for a product online few days back and then you keep receiving emails for shopping suggestions. If not this, then you might have noticed that the shopping website or the app recommends you some items that somehow matches with your taste. Certainly, this refines the shopping experience but did you know that it's machine learning doing the magic for you? On the basis of your behaviour with the website/app, past purchases, items liked or added to cart, brand preferences etc., the product recommendations are made.

I. Online Fraud Detection

Machine learning is proving its potential to make cyberspace a secure place and tracking monetary frauds online is one of its examples. For example: Paypal is using ML for protection against money laundering. The company uses a set of tools that helps them to compare millions of transactions taking place and distinguish between legitimate or illegitimate transactions taking place between the buyers and sellers.

9. CONCLUSION

We tried to explore different trends from any shopping dataset (black Friday). We extracted useful information that answered questions such as: what gender shops more. Do the occupations of the people have any impact on sales? Which age group is the highest spender? We **created a simple machine learning algorithm** that predicts the amount of money that a person is likely to spend on any occasion (ex: black Friday, amazon big billion days etc,) depending on features. An application is built which is integrated to the model built to display the predictions on UI.

10. FUTURE SCOPE

The results of this study were limited by constraints of the sampled demographic. If the current research were to be repeated perhaps other personality traits could be studied in relation to the likelihood to engage in consumer misbehaviour on BF. Longitudinal studies that track a wide-array of consumers that shop in the same retail locations for both BF and average retailing days may be of value. Interviews gathered employee opinions on differences and similarities of store environment on BF in comparison to average days but consumer reports of store differences may reveal other differences that could be explored in more detail in the future. Future researchers may benefit from developing gender studies where the behaviours of both men and women are analysed, keeping in mind that women and men may frequent different types of stores for different products on BF (i.e., men may be more likely to seek out electronic or hardware items whereas women may be prone to seek clothing, toys and other accessories). Aggression specifically has been studied between genders and extent literature as well as crime statistics suggest that men are more likely to engage in physical aggression than females. Research suggests that historically, gender differences in regards to aggression have remained stable. Furthermore, in observing homicide trends in the United States it is clear that most victims and offenders in homicides and violent crimes have been predominately male consistently since 1975 (U.S. Department of Justice, 2012). These behavioural differences would be interesting to explore within the context of BF and consumer misbehaviour. Another constraint of the current research was the age of participants. In the interviews, employees were in agreement that consumer behavioural trends differed 42 according to age; thus, an interesting study may try to isolate different age ranges of consumers. Within the context of this research only college undergraduates were used for pilot testing and college undergraduates and graduates were issued the main survey. Although we had a relatively large range of ages (18 to 63) it was not an evenly distributed sample. Future research may also look to study cultural differences of opinions and behaviours of consumers on BF. BF may be a culturally contrived phenomenon in the United States as BF attracts a large population of immigrants every year. According to Lyon and Trevisani (2011) Brazilians in particular will travel over 5,000 miles in order to partake in the great deals offered in the U.S. during the holiday season (i.e., BF) especially. BF is portrayed mostly in the United States as a retail holiday centred on egregious behaviours so it would be interesting to see what trends, if any, exist in other countries. When considering BF, one must consider scarcity in conjunction with consumer behaviour. Scarcity was defined as a “perception…sustained by a competitive pressure on the demand side, and the consumer infers from this competition that the scarce good should possess some inner tangible property”. Thus, products that are scarce are considered more desirable by consumers.

Research suggests that not only do consumers assume that scarce products are rated as more popular and of greater quality but they are more likely to choose scarcer products as well. A major complaint identified by consumers and employees alike regarding BF is the limited number of products and high tendency for stock outs. Research regarding the conjunction of 43 scarcity and consumer misbehaviour would be of high interest in future experiment design. Another interesting aspect of BF that should be explored further would be research regarding the physical store set-up, specific store-type and employee qualifications and appearance. Interviews with frontline retail employees raised issues that may play a role in promoting egregious consumer behaviour. Employees shared that the store layout often changes specifically for BF although the consumer demographic does not really undergo any significant noticeable changes. Participants also shared that it was impossible to compare behaviours of consumers who frequent Wal-Mart compared to establishments such as Nordstrom because the demographic of consumers were too contrasting. This would be an interesting concept to consider further. In terms of employee influence, interviews revealed that employees noticed that consumers tended to treat employees differently based on factors such as dress and age. Employee appropriateness of dress within the context of BF would also be of interest for further exploration. In a study by Shao Baker and Wagner (2004) the effects of appropriateness of employee dress on consumer expectations of service quality and purchase intent in the banking field was explored. The results of this study presented data that supports the importance of service employee dress in consumer evaluation of the employee and quality of the firm (Shao et al., 2004). Although this concept has been explored in other contexts, it may be of interest in regards to BF.

11. BIBILOGRAPHY

1. https://www.google.com/url?q=https://www.lsbf.org.uk/blog/news/analyse-consumer-behaviour/121862&sa=D&source=hangouts&ust=1591123763219000&usg=AFQjCNE5-OFvBaUgL9GAC5h_mIKJ5GMDEw

2. https://www.google.com/url?q=https://www.abacademies.org/articles/how-black-friday-can-mess-with-jordanian-customers-minds-8250.html&sa=D&source=hangouts&ust=1591123793890000&usg=AFQjCNGsPWTyJfaiYotzfJPW4k7HYsSwqw

APPENDIX

SOURCE CODE

```
import numpy as np
import pandas as pd
from sklearn.model_selection import train_test_split
from sklearn.preprocessing import StandardScaler
from sklearn.model_selection import train_test_split
from sklearn.linear_model import LinearRegression
from sklearn.tree import DecisionTreeRegressor
from sklearn.ensemble import RandomForestRegressor
from sklearn.metrics import r2_score
import matplotlib.pyplot as plt
%matplotlib inline
import seaborn as sns
import pickle
from sklearn import linear_model as lm


lr = LinearRegression()
lr.fit(x_train,y_train)
pred = lr.predict(x_test)
accuracy = r2_score(y_test,pred)



mlr  = LinearRegression()
mlr.fit(x_train,y_train)
y_pred = mlr.predict(x_test)
accuracy = r2_score(y_test,y_pred)


lr = LinearRegression()
dtr = DecisionTreeRegressor()
rfr = RandomForestRegressor()
fit1 = lr.fit(x_train,y_train)
fit2 = dtr.fit(x_train,y_train)
fit3 = rfr.fit(x_train,y_train)
```

decision tree

```
ydtr=dtr.predict(x_test)
ydtr
accuracydtr=r2_score(y_test,ydtr)
accuracydtr
```

random forest

```
yrfr=rfr.predict(x_test)
yrfr
accuracyrfr=r2_score(y_test,yrfr)
accuracyrfr
```