Walmart Sales Predictor
DATA 5000 Y: Walmart Sales Predictor

Muhammad Salman (100995527) Marziyeh Zamiri (101214196)

January 31st, 2021

Abstract:
Sales predictions can be used to allow businesses to develop a plan and order inventory better suited for their customers, driving profits higher, while cutting losses simultaneously. Given a Walmart dataset containing store/department information, we can predict future sales per department across a number of different locations and predict the effect of sales during different holiday seasons. Datasets are open access and we will download the required datasets and information from Kaggle (link provided in reference [1] below).

Introduction/Motivation:
As the world's population grows, customer preferences have become broader and more diverse, resulting in manufacturing of various products. Retail stores, serving as the joiners between industries and customers, must manage the availability of a significant number of items at different times, to ensure customer satisfaction as well as maximum profit. Therefore, it is crucial to understand the mentality of millions of customers around the globe. This is done through collecting/analyzing sales data.

Data science classifies structured and unstructured data using statistical methods and scientific algorithms [2,3]. Nowadays, it is playing a notable role in business analysis and has an increasing impact on business models in all industries, including retail. It has shown that the use of Big Data techniques lead to an average sales increase of 62% in major retail stores [4]. Walmart is one retailer that has applied data science methodology in various supply chain arenas. In Walmart Labs, some essential applications of data science are: offering a smarter customer experience, using social media to forecast trends, analyzing the path to purchase, optimizing prices, and maximizing inventory, etc. Utilizing data science in business improves customer experience remarkably, while also having the potential to boost retail sales.

The project will be undertaken by Muhammad Salman, who has a background in Python and C, and is eager to learn more about data analytics with different software such as R, and by Marziyeh Zamiri, who has an optoelectronic background, currently doing her M.Sc.in Electrical Engineering, with a Data Science focus.

Description of the Data:
The dataset that this group will be analyzing contains data for approximately 45 Walmart branches, subject to decrease during data clean-up. A brief description and size of each store is also provided, realizing that there are multiple departments within each store [1].

The provided training data ranges over a two year span, from February, 2010 to November, 2012. There is also data pertaining to price markdowns, especially before seasonal holidays, with a strong focus on Christmas, Thanksgiving, the Super Bowl, and finally Labour Day [1].

Some data fields found within the files are: number of departments/stores, week of sales, if the week was within a holiday season, and weekly revenue [1].

Problem Statement:
Walmart is one of the world's largest departmental retail stores with growing revenues on a yearly basis. Here, we define a very interesting Big Data project based on Walmart's open source datasets: *forecasting future sales across various departments within different Walmart branches, depending on the season*. This will be performed through technical analysis, and data training/machine learning techniques.

Objectives:
The challenging aspect of this project is to consider the different sales rates during the four major holidays, as mentioned above, within a year. By predicting the sales patterns, a sufficient product supply can be determined, in order to meet the demand and generate maximum profit.

Training such data could potentially allow business decisions that will drive sales and minimize losses. This will lead appropriate inventory to be stored and less stock shortages. Though this may be an obvious example, ordering more Father's Day cards before Father's Day would raise profits, while ordering Christmas Cards for the same holiday would lead to greater losses.

Conclusion:
In conclusion, we are looking forward to undertaking this project as part of the DATA 5000 seminar class and delivering an exceptional final analysis. We hope that Professor Genkin approves this proposal so that we may begin with our dataset.

References:
[1] Kaggle.com. 2014. *Walmart Recruiting - Store Sales Forecasting | Kaggle*. [online] Available at: <https://www.kaggle.com/c/walmart-recruiting-store-sales-forecasting/data> [Accessed 31 January 2021].
[2] Dhar, Vasant. "Data science and prediction." *Communications of the ACM* 56, no. 12 (2013): 64-73.
[3] Leek, Jeff. "The key word in 'data science' is not data, it is science." *Simply statistics* 12 (2013).
[4] Bennani, Y., 2018. *Data Science applied to the retail industry: 10 essential use cases - Margo*.[online] Margo. Available at: <https://www.margo-group.com/en/news/data-science-applied-to-the-retail-industry-10-essential-use-cases/> [Accessed 31 January 2021].