**Walmart Sales Analysis**

Vaishali Maheshwari

Regis University

MSDS 670: Data Visualization

Professor John Koenig

March 10, 2024

YouTube Presentation Link : https://youtu.be/18JQmaJWFTE

**Introduction**

This century has seen an outburst of data being generated as a result of the

continuous use of edge technologies. Retail giants like Walmart consider this data

as their biggest asset as this helps them predict future sales and customers and finally helps lay out plans to generate profits and compete with other organizations. The employment of modern technological approaches is crucial for any organization to survive in a cutting-edge global market and enable services that distinguish it from its competitors.

**Purpose**

The idea behind this project is to explore the Walmart data for 45 stores and draw the analysis based on available historical data and identify whether factors like temperature, unemployment, fuel prices, CPI etc affect the weekly sales of these stores under study. The aim is to understand whether sales are relatively higher during holidays like Christmas and Thanksgiving, Labor day and Super Bowl than normal days so that stores can work on creating promotional offers that increase sales and generate higher revenue.

Walmart runs several promotional markdown sales throughout the year on days immediately following the prominent holidays in the United States which are taken into consideration to determine the impact on weekly sales and to drive resources towards such key strategic initiatives.

**Tools Used**

The analysis for this data has been done using Python and Tableau on the dataset provided by Walmart at “*https://data.world/tommywilczek/walmart”*. For the initial EDA Python libraries like matplotlib, seaborn, etc have been implemented. Packages like numpy, pandas, etc. have been used for data wrangling and manipulation. For more advanced visualizations, the tableau dashboard has been used.

**Exploratory data analysis**

The extensive analysis of the Walmart dataset is performed to depict the following :-

* Identifying store as well as department-wide sales in Walmart
* Identifying sales based on store size and type
* Identifying how much sales increase during holidays
* Correlation between the different factors that affect sales
* Monthly Total and Average Sales Analysis
* Day-wise sales analysis across the years 2010 to 2012
* Impact on sales as per temperature, fuel price and unemployment.

A correlation matrix shows Department and IsHoliday flag has a maximum dependency on the Weekly Sales values. The size of the store also has an impact on the sales. Markdown fields have no direct impact on the Weekly Sales. Fuel Price, Temperature and CPI have the least correlation with the sales hence they are not important while predicting the sales.

**Data Cleaning**

Deletion of the unnamed columns that were not required for analysis. The records with weekly sales less than 0 were removed to avoid the error because of negative values. Temperature outliers were identified in the data but can be ignored. Markdown columns have the null values that were handled. All the 5 markdown missing fields are imputed with the mean of the available column values.

**Feature Engineering**

Considering four holiday weeks in the data, added the columns for super\_bowl, Thanksgiving, Labor\_Day and Christmas. The date field was split to get more insights about monthly, day-wise or yearly sales. The holiday\_month column was introduced to explore the sales. The holiday months marked are February, September, November and December. Days left for Thanksgiving and Christmas are created with the expectation that sales will be higher before the holidays than on the exact holiday.

The quarterly average sale details for 25 stores with the highest sales are as below :-



Quarter 4 has recorded the maximum sales probably due to the holiday season in this period. Quarter 2 has the next highest sales which can be due to the pleasant weather during that period of the year.

**Conclusions**

* Store size is a great factor that affects sales; the bigger the store, the higher the sales. Store A still has the highest sales, followed by stores B and C.
* Stores 20,4 and 14 have the highest sales for Type A, and stores 10, 23 and 18 have the highest sales for Type B.
* Departments 95, 92 and 38 generate the highest revenue for Walmart across all the sizes of stores. The stores with these departments will result in higher sales.
* Sales are the highest during the holiday season (in November and December month).
* Days around the Thanksgiving holidays have recorded higher sales than any other holiday.
* Extreme hot and cold weather has a profound impact on sales and pleasant weather encourages higher sales.
* Unemployment impacts the sales for bigger stores like type A and B.

**Resources**

* <https://seaborn.pydata.org/generated/seaborn.heatmap.html>
* <https://www.thedataschool.com.au/jason-hu/four-different-methods-to-calculate-the-percent-of-total-in-tableau/>
* <https://www.tableau.com/dashboard-examples>