Utkarsh Nigam, DATS 6401, Summer 2020

Final Technical Report

## **Impact of Holiday Season 2019 on US Online Retail Sales**

**Abstract**

As per the [Forbes Report](https://www.forbes.com/sites/bryanpearson/2016/12/22/holiday-spending-to-exceed-1-trillion-and-11-other-surprising-data-points-of-christmas/#7748f391247f), total holiday sales in the USA exceeded USD 1 Trillion in 2016, which is roughly 5% of the US GDP and more than 2/3rd US citizens shopped on these days. With the retail industry being the most holiday sensitive, this project will use different visualizations and analysis to uncover some interesting trends and insights from 2019 online sales data for a multi-category store. This project will also be including data-backed business recommendations for the upcoming holiday season 2020.

**Objectives**

The primary objective of this project is the application of complex visualizations to learn customer patterns & trends, from the 2019 online sales data, which can help derive actionable insights to create a truly unique and differentiated customer experience this year. The approach would be to analyze the performance of non-holiday months i.e. January till October and compare them with holiday months i.e. November-December. Including significant dimensions such as device type, geography, products category, time of the day will provide one level deeper insights on how the retail industry moves during the holiday season.

**Functional Requirements**

This project must be able to preprocess, aggregate, structure and finally, deliver appropriate visualization of the online sales data with respect to different custom views a user wants to look at in order to generate insights followed by recommendations. To maintain consistent experience throughout, the end-user will be able to interact with an HTML based web page. Each visualization will contain its respective level of detail specifications based on its type.

**System Architecture and Description**

The complete project was distributed into brief sequential milestones as listed below:

1. Primary Data Collection & Secondary Research
2. [Hypothesis Building](https://github.com/utkarsh-nigam/DATS-6401-Final/blob/master/Final%20Project_Hypothesis.xlsx)
3. Data Preprocessing
4. Data Visualization & Front End Development
5. Story Scaping

E-commerce sales data for the time range January-2019 to April-2020 was downloaded from Kaggle. Data preprocessing and consolidation were majorly done using [python](https://github.com/utkarsh-nigam/DATS-6401-Final/blob/master/Preprocessing_FinalProject.py). Zip code data was integrated later using [zipcodestogo.com](https://www.zipcodestogo.com/ZIP-Codes-by-State.htm) to analyze the state-wise behavior during the holiday season. Campaign level data was segregated separately to illustrate user behavior specific to the campaign. [Processed data](https://github.com/utkarsh-nigam/DATS-6401-Final/tree/master/Data) for each section of the project is stored separately in Excel files that will then be read to render visualizations using d3.

**Development Platforms**

The user-facing front end of this analysis will consist of one fixed side navigation web page with scrolling page anchors, that includes various tabs in a sequence to showcase a story through the graphics. Tabs will include sections for High-Level Key Performance Indicators, comparison of Holiday Season with Non-Holiday Season, Geographical Performance, and so on. Every section will also showcase key insights and recommendations based on data visualization. Web page will be stylized using Cascading Style Sheets (CSS). Visualizations will be developed using the D3 JavaScript library and Google Visualization API.

**Proposed Visualizations**

A combination of D3 JavaScript library and Google Visualization API will be used to create visualizations for this project. Vertical bar charts will be used to showcase the summary of the high-level numbers and their breakdown. Dual-axis combo charts will be used to indicate trends for metrics like order conversion rate and traffic throughout the holiday season. Interactive pie charts will be implemented to reflect and compare the performance of each product category and also highlight the holiday sensitivity. Geo plots will be rendered using Google Visualization API to display and analyze the state-wise behavior. Since the retail industry is highly volatile to seasonality, therefore, calendar charts will be used to illustrate how some metrics vary depending on the day of the week. Horizontally stacked bar charts will showcase the user purchase behavior like propensity to buy and conversion rate during the peak sale days i.e., Wednesday before Thanksgiving, Thanksgiving, Black Friday, and Cyber Monday.

**Experimental Analyses and Conclusions**

From high-level observation, holiday period i.e. November & December together contributed to almost 34% of the total yearly revenue for 2019, which clearly signifies its impact on businesses. The calendar chart showed that peak sale days generated 5X more revenue than the rest of the holiday season. Usage of the mobile device outnumbers desktop and tablet, which is the reason why marketers prioritize responsive mobile-friendly experience. In terms of product categories, Electronics and Home Furnishing are the most holiday sensitive categories, for which bulk of the buying happens during the holiday season. Therefore, this may help fine-tune the marketing mix by adding sensitive categories for customer acquisition campaigns. The geo plots displayed that California and Texas are the top revenue-generating states overall and for every campaign. Interestingly, US customers show different buying behavior during every sale day. Most of them order either early in the morning on Thanksgiving, or late night on Black Friday and Cyber Monday.

With this study, we conclude by highlighting the importance of Holiday Season for the retail industry and the reason why businesses want to plan well ahead of time to prepare for this shopping extravaganza. Now due to this ongoing pandemic, more of the holiday shopping can be expected to be done digitally this year, and that is why the power of data becomes even more essential to create a truly unique and differentiated customer experience, and outshine from their competitors by increasing the online sales, revenue and reducing the operational costs.