CS7DS4 Data Visualization

2019-2020

Student Name: Suhrid Chatterjee

Student Number: 18312290

Assignment 3

(Open the workbook and press F7 to run it in presentation mode)

INTRODUCTION

QVC is a retail company based in the United States of America that has television networks and home shopping channels through televised advertisements. The dataset was released as a part of the QVC Analytics Challenge in 2018. Following the lines of the original challenge, I attempted to perform an exploratory analysis to detect patterns in the data and gain insights about product delivery through their nationwide network, purchasing behaviors and customer geography, that are beneficial from a business standpoint.

DESCRIPTION

THE DATA

In the dataset provided, there are 6 major data files named "**QVC Data 1**" to "**QVC Data 6**" along with a randomly generated sample of smaller size. These are transactional data that describe business activities aggregated during the course of QVC's transactional activities. There are other reference data files containing data dictionaries and distribution center data.

Each of the six files contain data for all the relevant states in the USA. The type of data ranges from numerical values to highly diverse categorical attributes.

THE APPROACH: DATA HANDLING AND MANAGEMENT

The six data files were congregated to generate "**QVC Data Total**" using R:

```
> QVC_Data_Total <-rbind(QVCData1,QVCData2,QVCData3,QVCData4,QVCData5,QVCData6,deparse.level = 1)
  str(QVCDataTotal)
  write.csv(QVC_Data_Total,file="G:/DataVisualization/A3/QVC_DV/QVC DATA-20200415T171929Z-001/QVC DATA/QVC Data Total.csv",row.names =
          TRUE)
```

- QVCData1, QVCData2, QVCData3 and QVCData5 had one unnecessary boolean attribute which were removed from the datasets. After this, all the files had 36 attributes each.
- The **Source_Ship_Warehouse_Nbr** attribute was used to join the total data **QVC Data Total** with the data file containing the distribution center data, that is "**QVCdist_ctr**". The format of this attribute was different in the two files – text in **QVCdist_ctr** and number in **QVC Data Total**. Also, there were no leading zeroes in the attribute in **QVC Data Total** like there was in **QVCdist_ctr**. These issues were taken care of before joining the two files.
- There were also a few invalid state codes that did not correspond to any state in the USA such as LI, NL, S and WC.
- Attributes with large amount of missing data were omitted from the dataset such as **Shipped_Dt**, **Package_Scan_Dttm**, **Pickup_Dt**, **Rescheduled_Delivery_Dt**, **Scheduled_Delivery_Dt**, **Shipping_Priority_Ind** and **SHIP_TO_CITY**.

THE FINDINGS

The tableau storyboard attempts to understand the relationships between multiple factors influencing customer loyalty, such as the speed with which products are delivered to them.

From the visualization as a whole, the conclusions derived are outlined below:

- Warehouses in New Jersey, New York and California should increase their storage provisions for clothing and accessories (Apparels) since they account for the majority of orders in this category.
- North Carolina is responsible for the maximum number of health and beauty product orders and strategically placed warehouses here can result in mitigation of logistical expenses for these products.
- While the states in the eastern half of the country benefitted from many warehouses located in their close proximity thus leading to a delivery waiting time of a week maximum, the western half suffered from a lack thereof. Hence, the natural progression would be to add the number of warehouses in the northwestern states.
- Decisions to place warehouses in these states where the reasoning is backed by visual analysis would expedite the delivery process of products in demand specific to each state.
- This leads us to arguably the most important conclusion from this visualization - it was discovered that the customers who enjoyed shorter delivery times not only inclined to stick to the same retailer for future purchases, but also tend to buy more from the them. It is thus the most important takeaway that warehouses must be placed such that it is able to deliver products to its customers efficiently while also increasing their revenue, creating a win-win situation.

CITATIONS AND REFERENCES

Data source: https://drive.google.com/drive/folders/1dkaoFlk-hoogmTC0E9UCdShoWAYVXKYW