Data 608 – Final Project Proposal

Contents

[**Business Scenario:** 1](#_Toc23187852)

[**Dataset:** 1](#_Toc23187853)

[**Dataset Description:** 1](#_Toc23187854)

[**Project Goal:** 2](#_Toc23187855)

[**Outcome:** 3](#_Toc23187856)

[**Dataset Link:** 3](#_Toc23187857)

## Business Scenario:

My final project highlights a current business scenario. In today’s world more and more businesses are online and competition is fierce. Businesses are striving for better customer engagement and understanding of customer buying behaviors. This helps businesses to serve customers better and they can effectively retain existing customers while attracting future customers.

One of the key techniques used by large retailers for customer engagement and enrichment is called Market Basket Analysis (MBA). This is a technique which uncovers association between products by looking for combinations of products that frequently co-occure in transactions. In other word it helps retailers to identify products which customers buy together. For example a customer who buys flash light will likely to buy batteries.

Retailers can use insights gained from MBA in number of ways including:

1] Grouping products that co-occure in the design of store’s layout

2] Driving online recommendation engine

3] Targeting marketing campaigns by sending out promotional schemes to customers about related products

## Dataset:

My project is about performing market basket analysis for large retailer called Instacart. Instacart has made their dataset publicly available. The dataset is about relational set of files describing customers order over time. The goal of analysis to figure out which products are frequently bought together by customers. The dataset is anonymized and contains sample of over 3 million grocery orders from more than 200,000 Instacart users. For each user there are about 4 to 100 orders with the sequence of products purchased in each order

## Dataset Description:

Aisles.csv: Contains aisles information

aisle\_id,aisle

1,prepared soups salads

2,specialty cheeses

3,energy granola bars

Departments.csv: Contains department information within Instacart

department\_id,department

1,frozen

2,other

3,bakery

Order\_products.csv: These files specify which products were purchased in each order. order\_products\_\_prior.csv contains previous order contents for all customers. 'reordered' indicates that the customer has a previous order that contains the product.

order\_id,product\_id,add\_to\_cart\_order,reordered

1,49302,1,1

1,11109,2,1

1,10246,3,0

Orders.csv: Contains order details

order\_id,user\_id,eval\_set,order\_number,order\_dow,order\_hour\_of\_day,days\_since\_prior\_order

2539329,1,prior,1,2,08,

2398795,1,prior,2,3,07,15.0

473747,1,prior,3,3,12,21.0

Products.csv: Contains product details

product\_id,product\_name,aisle\_id,department\_id

1,Chocolate Sandwich Cookies,61,19

2,All-Seasons Salt,104,13

3,Robust Golden Unsweetened Oolong Tea,94,7

## Project Goal:

Goal of the project is to perform details MBA for Instacart by using effective data visualization techniques. Project will focus on following areas

1] Exploratory Analysis –

* Explaining data.
* Data Imputation
* Outlier Detection
* Variable encoding etc.

2] Descriptive Statistics – Quick description statistics showing following facts

* What time of the day people mostly order?
* What day of the week is most busy?
* When do people order again?
* Home many prior orders exists for customers?
* How many items people generally buy?
* What are the bestsellers
* How often people order same item again?

3] Market Basket Analysis – Finding association between products. We will focus on following areas

* Probabilistic estimation of reorder based on time
* Probabilistic estimation of reorder based on quantity
* Association between organic and in-organic products

4] Uber product portfolio analysis – Perform overall product portfolio analysis for Instacart. We will highlight following fact using data visualization techniques

* Visualizing product portfolio
* Unique products per aisle
* Selling frequency of products

## Outcome:

This project will be a very good showcase where we can see how effective data visualization techniques can be leveraged to solve complex business problem like Market Basket Analysis. MBA is an integral part of retailer’s strategy to retain and gain customers. This project will demonstrate how Data Scientist can leverage advance visualization techniques to add required clarity on massive datasets and enable leadership to take key decisions to impact business bottom line.

This project will have following highlights

* Dealing with real world complex datasets
* Data cleaning and data transformation
* Leveraging advance data visualization technique to surface out complex facts enabling key decision making
* Effective use of multiple charts to add required clarity to underlying data.
* Detailed explanation for each analysis explaining how it helps towards decision making

## Dataset Link:

<https://github.com/mlforsachid/MSDSQ4/tree/master/Data608/FinalProject/Dataset>