| **Batch details: Gurgaon Jan 21 Batch** |
| --- |
| **Team members: Akshay Goel, Harshika Prakash, Raghav Verma, Rishabh Gupta, Shivam Nagpal, Srishti Taneja** |
| **Domain of Project: Consumer Behaviour** |
| **Proposed project title: THE INFLUENCE OF CONSUMER BEHAVIOUR ON E-COMMERCE** |
| **Group Number: 3** |
| **Team Leader: Shivam Nagpal** |
| **Mentor Name: Vibha Santhanam** |



**Dataset Name:**

**Brazilian E-Commerce Public Dataset by Olist**

**Brazilian E-Commerce Public Dataset by Olist**

**INTRODUCTION**

Research in consumer behaviour shows that we have a consumer driven society where the ultimate motive of business products is to satisfy consumer expectations which makes them happy and remain loyal to a brand. Therefore, a perfect understanding of consumer behaviour is determining.

**PROBLEM STATEMENT**

To find out which products can be placed together.  This would give retailer good information about related sales on group of goods basis. (Market basket analysis)

To predict customer scores based on customer segmentation (Recency, frequency, monetary value analysis) for the company.

**ABSTRACT**

Consumer behaviour involves the psychological processes that consumers go through in recognizing needs, finding ways to solve these needs, making purchase decisions (e.g., whether or not to purchase a product and, if so, which brand and where), interpret information, make plans, and implement these plans. Consumers often buy products not because of their attributes per se but rather because of the ultimate benefits that these attributes provide, in turn leading to the satisfaction of ultimate values. The important thing in a means-end chain is to start with an attribute, a concrete characteristic of the product, and then logically progress to a series of consequences (which tend to become progressively more abstract) that end with a value being satisfied. Thus, each chain must start with an attribute and end with a value. An important implication of means-end chains is that it is usually most effective in advertising to focus on higher level items.  
  
A market comes into existence because it fulfils the needs of the consumer. Consumer behaviour is a complex, dynamic, multidimensional process, and all marketing decisions are based on assumptions about consumer behaviour. Models of consumer behaviour play a key role in modern empirical Industrial Organization.

**METHODOLOGY**

* Apriori algorithm: **Apriori algorithm** is used for finding frequent item sets in a dataset for Boolean association rule. Name of the algorithm is Apriori because it uses prior knowledge of frequent itemset properties. We apply an iterative approach or level-wise search where k-frequent item sets are used to find k+1 item sets.

To improve the efficiency of level-wise generation of frequent item sets, an important property is used called *Apriori property* which helps by reducing the search space.

* Churn analysis: Churn analysis is the evaluation of a company's customer loss rate in order to reduce it. Also referred to as customer attrition rate, churn can be minimized by assessing your product and how people use it.

**WEEKLY PLAN**

Week 1: Generating the problem statement

We define our problem statement and the expected outcomes from the project. This phase defines what exactly are our end goals through the project.

Week 2: Connecting with the mentor

After creating the problem statement, we connect with the mentor for expert guidance on how to proceed further with the project.

Week 3 and Week 4: EDA and Pre-processing

After we had our problem, we need to create at least one hypothesis that will help solve the problem. The hypothesis is our belief about how the data reacts to certain variables.

Our hypotheses need to have data that will allow us to prove or disprove them. This is where we need to look in the data set for variables that affect the problem.

As much as data scientists prefer to have clean, ready-to-go data, the reality is seldom neat or orderly. We may have outlier data that we can’t readily explain. In these cases, it’s up to us as a data scientist to remove those outliers and add missing data so that the data is more or less consistent. Without these changes, our results will become skewed and the outlier data will affect the results, sometimes drastically.

Week 5: Modelling the dataset

At some point, we’ll have to come up with models to support our hypotheses.

At this stage, we need to start assigning variables to our data. We need to factor in what will affect your data. We may have to modify certain variables we created in order to have a better prediction of sales.

Week 6: Finalizing and generating report

Create a data visualization or a presentation that explains our results to the layman and presenting the final project report.

**REFERENCES**

[**www.kaggle.com**](http://www.kaggle.com)**/olistbr/brazilian-ecommerce**

[**www.nairaproject.com**](http://www.nairaproject.com)

[**www.theconversation.com**](http://www.theconversation.com)

[**www.papers.ssrn.com**](http://www.papers.ssrn.com)

**Declaration: This is to declare that the dataset that we are using for our capstone project does not have any relevant legality associated to it** **and can be used to showcase the work we do on it as a presentation in Great Learning.**