

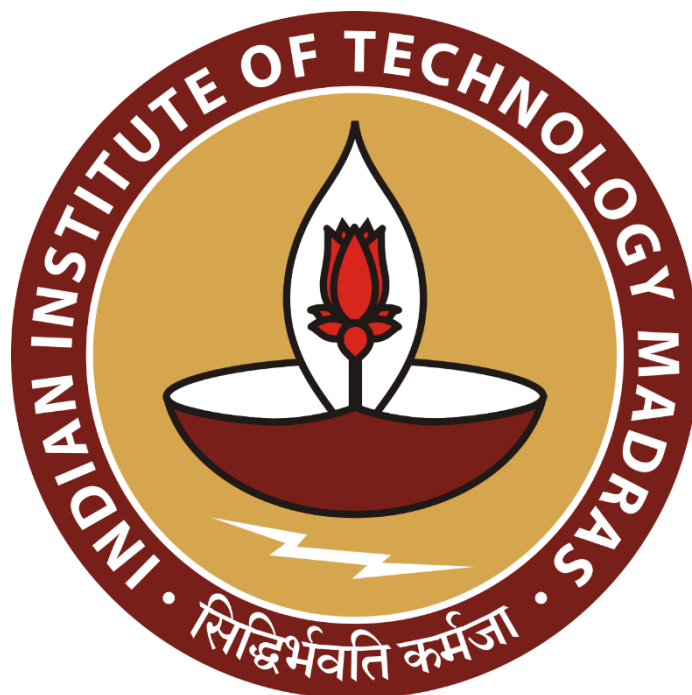
Brazilian E-Commerce [BDM - Project]
A Proposal report for the BDM capstone Project

Submitted by

Name: Vivek Singh Rao

Roll number: 21f3002861@ds.study.iitm.ac.in

Date of submission- 02-12-2024



IITM Online BS Degree Program,
Indian Institute of Technology, Madras, Chennai, Tamil Nadu, India, 600036

Contents

1	Executive Summary and Title	2
2	Organization Background	2
3	Problem Statement (Listed as objectives)	3
3.1	Problem statement 1: Explanation in 1-2 lines	3
3.2	Problem statement 2: Explanation in 1-2 lines	3
4	Background of the Problem	
5	Problem Solving Approach	3
6	Expected Timeline	4
7	Expected Outcome	5

Declaration Statement

I am working on a Project titled “Brazilian Ecommerce”. I extend my appreciation to Kaggle site, for providing the necessary resources that enabled me to conduct my project.

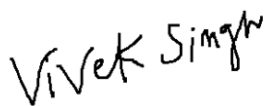
I hereby assert that the data presented and assessed in this project report is genuine and precise to the utmost extent of my knowledge and capabilities. The data has been gathered from primary sources and carefully analyzed to assure its reliability.

Additionally, I affirm that all procedures employed for the purpose of data collection and analysis have been duly explained in this report. The outcomes and inferences derived from the data are an accurate depiction of the findings acquired through thorough analytical procedures.

I am dedicated to adhering to the principles of academic honesty and integrity, and I am receptive to any additional examination or validation of the data contained in this project report.

I understand that the execution of this project is intended for individual completion and is not to be undertaken collectively. I thus affirm that I am not engaged in any form of collaboration with other individuals, and that all the work undertaken has been solely conducted by me. In the event that plagiarism is detected in the report at any stage of the project's completion, I am fully aware and prepared to accept disciplinary measures imposed by the relevant authority.

I understand that all recommendations made in this project report are within the context of the academic project taken up towards course fulfillment in the BS Degree Program offered by IIT Madras. The institution does not endorse any of the claims or comments.



Signature of Candidate: **(Digital Signature)**

Name: Vivek Singh Rao

Date: 2024/12/03

1 Executive Summary and Title

This project conducts an analysis of a Brazilian ecommerce dataset sourced from Olist. Olist facilitates connections for small businesses across Brazil, allowing them to access various sales channels seamlessly under a single contract. These merchants can market their products via the Olist Store and utilize Olist's logistics partners for direct shipping to customers. The dataset encompasses 100,000 orders from the years 2016 to 2018, providing valuable insights into order performance, logistics, customer demographics, product characteristics, and customer feedback.

The dataset sheds light on two main problems: poor customer loyalty, as 10% of buyers come back for more, and a big gap in sales across regions, with 80% of revenue coming from just two regions. These findings point to missed opportunities in regions that aren't doing as well and show we need to connect better with customers and tailor our approach to each area to grow the business long-term.

Doing customer segmentation analysis to tailor the strategies for different regions. Correlating variables like order status and freight cost features in these regions. Doing some sentimental analysis on the review comment feature and correlating it with the other features like review score and order status, customer city. By doing analysis like this we will get some insight into solving this problem.

Kaggle Dataset Ref and EDA Notebook: -

<https://www.kaggle.com/datasets/olistbr/brazilian-ecommerce>
<https://www.kaggle.com/code/viveksinghrao/2024-t3-bdm>

2 Problem Statement

1. Customer Retention:

The data set reveals low customer retention, with only 10% of customers returning repeat purchases. This illustrates the challenges of building customer loyalty, which can be due to limited personalization. Retention consistent post-purchase experiences, or unmet expectations Retention is critical to reducing costs on customer acquisition to achieve sustainable growth.

2. Non-Uniform Region Wise Sales

Around 80 percent of sales are concentrated in just two regions while the remaining 20 percent is distributed over the other 30-40 regions. Such an imbalance indicates the unexplored potential in the lagging regions that could be due to limited access or logistical constraints at the regional level or due to lack of localized product strategies.

3 Background of the Problem

The low customer retention rate might be because of a few things. Not personalizing marketing, giving different experiences after buying, or not keeping customers involved enough could all play a part. Also, tough competition and not knowing what customers want can make it hard to build loyalty.

As for sales being uneven across regions, this happens because marketing and shipping focus on just a few areas leaving others without much attention. Not getting what each region likes having trouble with shipping, and differences in local economies also limit sales in places that aren't doing well. To grow and keep customers, it's key to tackling these main issues.

4 Problem Solving Approach

Method: -

To tackle **customer retention** and **non-uniform sales distribution**, I will first create a unified dataset through SQL Joins and handle missing values with mean imputation. For retention, I'll perform **Sentiment Analysis** on reviews, correlate scores with order status or review score, and use **Cohort Analysis** and **RFM Analysis** to understand customer behaviors. For sales disparity, I'll conduct **Freight Cost Analysis**, **AOV Analysis**, and **Geo-Spatial Analysis** to identify regional inefficiencies and customer spending patterns

Data: -

The dataset has been taken from the Kaggle website. The dataset is highly popular in ecommerce analysis with an upvotes of 3393 and golden badge.

<https://www.kaggle.com/datasets/olistbr/brazilian-ecommerce>

Tools: -

Python Libraries: - For Performing EDA.

PostgreSQL Database: - For Data Cleaning and handling various data relations

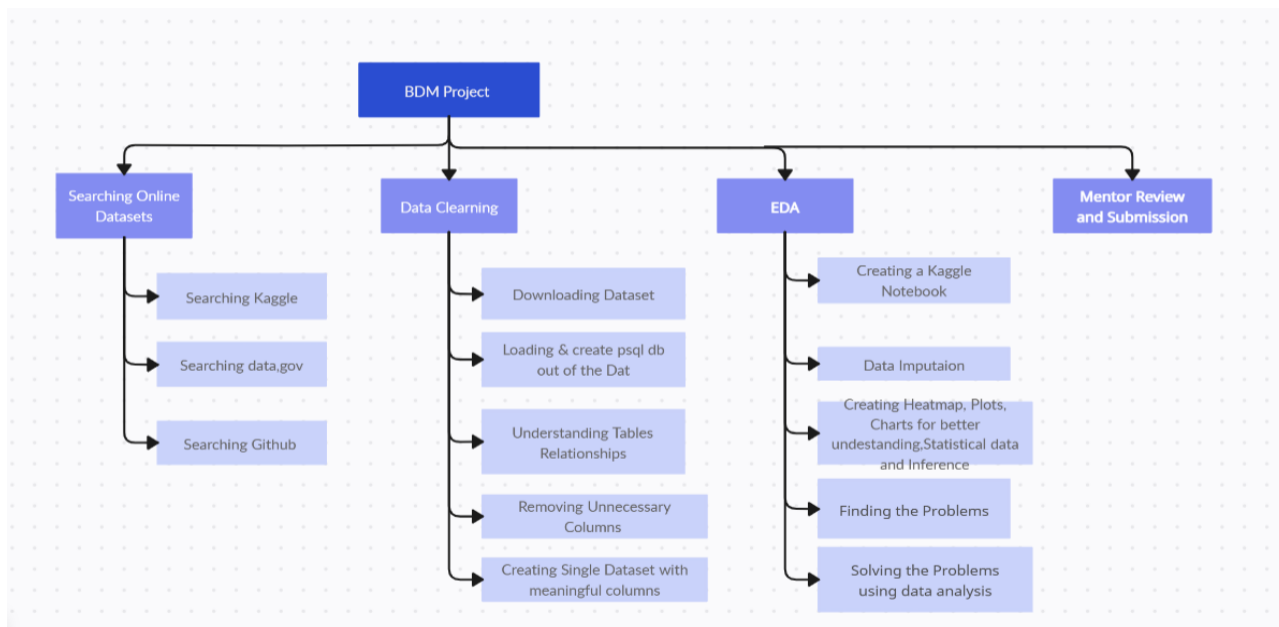
Power BI: - For making charts and analysis

5 Expected Timeline

Gantt Chart

ID	Task Name	2024-11					2024-12		
		01	03	10	17	24	01	08	15
1	Searching Secondary Dataset online								
2	Data Cleaning								
3	EDA -I Problems analysis								
4	EDA - II (Solution to Business Problems)								
5	Insights & Recommendations								

WBS Chart



6 Expected Outcome

With the help of the insights after the data analysis, Ecommerce business would be able to increase their customer retention and increase the sales in other cities as well. By putting data-driven insight to optimum utility for improving customer retention and the geographic sales, it is a safe bet that an organization will expect long-term sustainable growth along with gain-market competitiveness, and have a stronger, loyal customer base.