From Data to Decisions: Building a Smart E-Commerce Recommendation System

WHAT IS A RECOMMENDATION SYSTEM?

Introduction

Recommender systems are software tools and techniques that provide suggestions for items that are most likely to be of interest to a particular user - *Ricci, Rokach & Shapira (2011) – Recommender Systems Handbook*

These systems are widely used in e-commerce (Amazon, Alibaba), streaming platforms (Netflix, Spotify), and social media (YouTube, TikTok). The effectiveness of a recommendation system depends on its algorithm. The three main types are:

Problem Statement and Objective

- **Problem:** Traditional recommendation systems often lack accuracy and fail to provide truly personalized suggestions.

- **Intended Solution:** Implement a machine learning model to enhance recommendation quality and user experience.

PROJECT BODY

Building, evaluating and selecting the best machine learning model

Data Source and Overview

01 The dataset was obtained from Kaggle

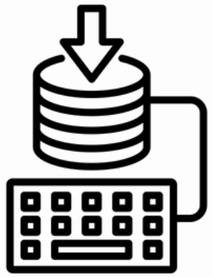
02 Dataset includes,
events_df: User interactions

03 **tem_properties_df**: Item attributes

04 **category_tree_df**: Item relationships

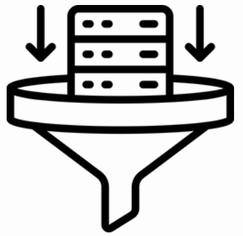
Model Building Workflow

DEFINING THE PROBLEM STATEMENT



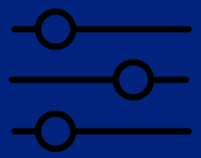
DATA IMPORTATION

EDA AND FEATURE ENGINEERING



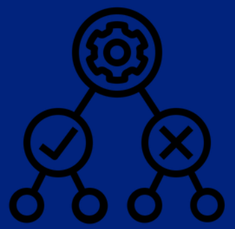
PRE PROCESSING FOR MODELLING

MODEL BUILDING



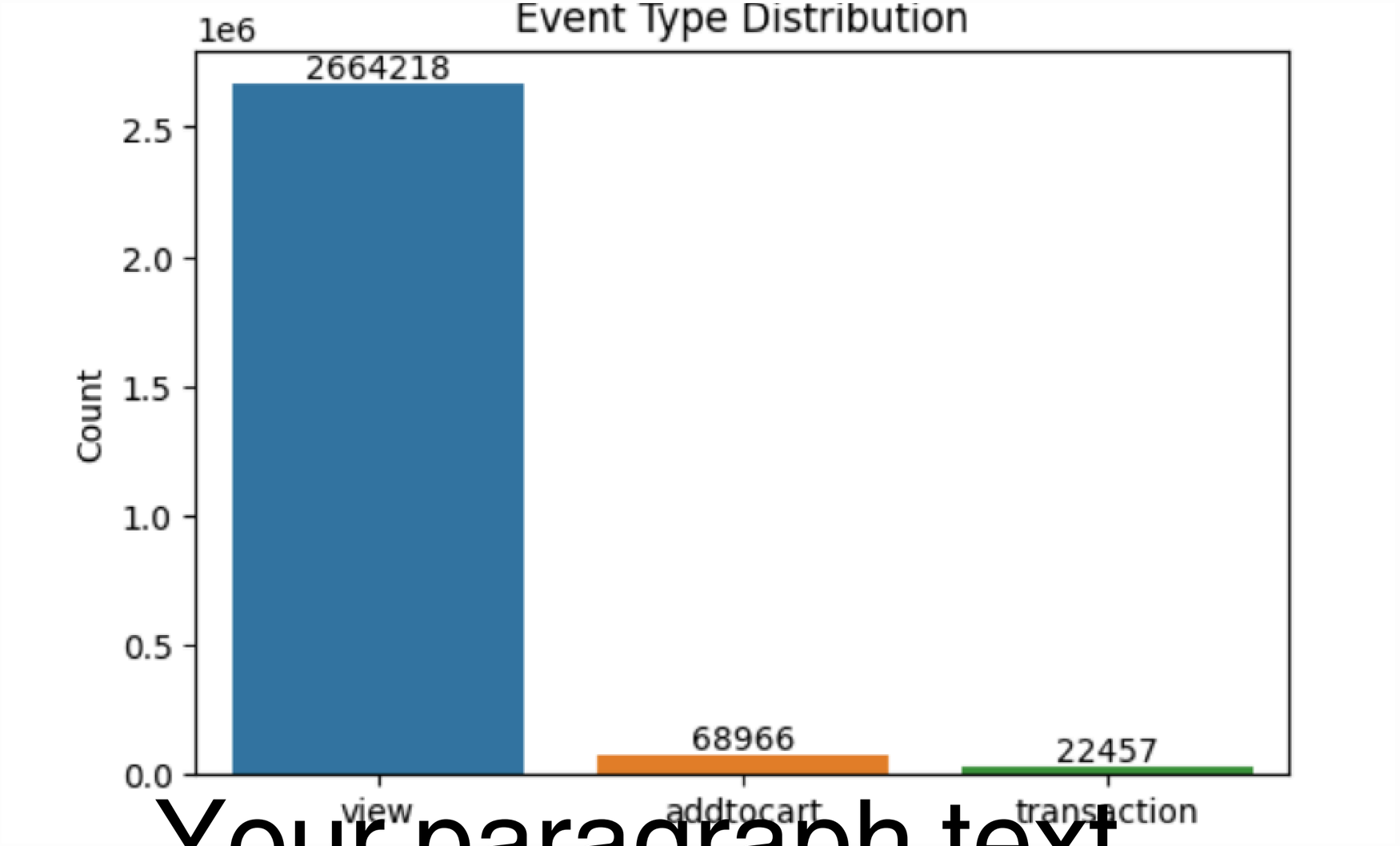
HYPERPARAMETER TUNING

MODEL VALIDATION AND EVALUATION



MODEL SELECTION

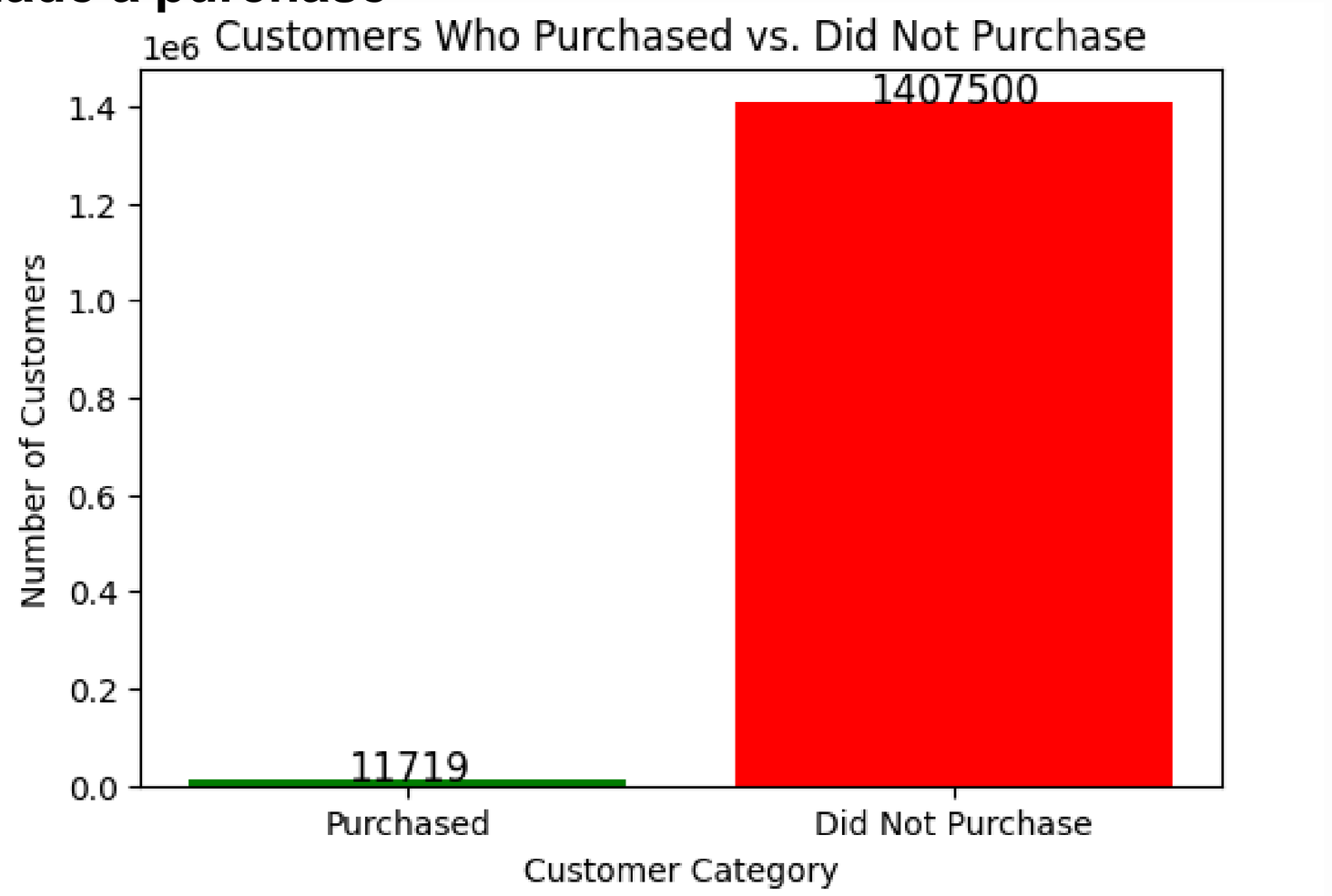
EDA: Event Type Distribution



Your paragraph text

Research Questions

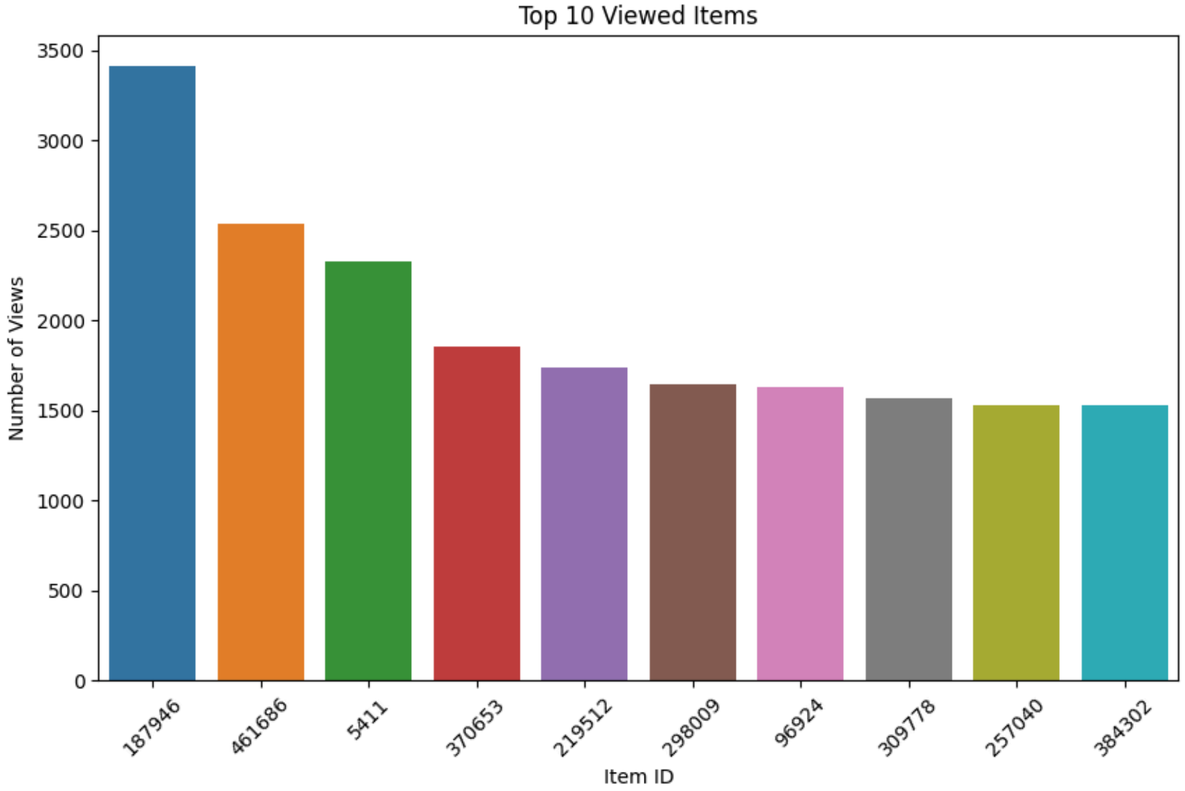
Customers who made a purchase

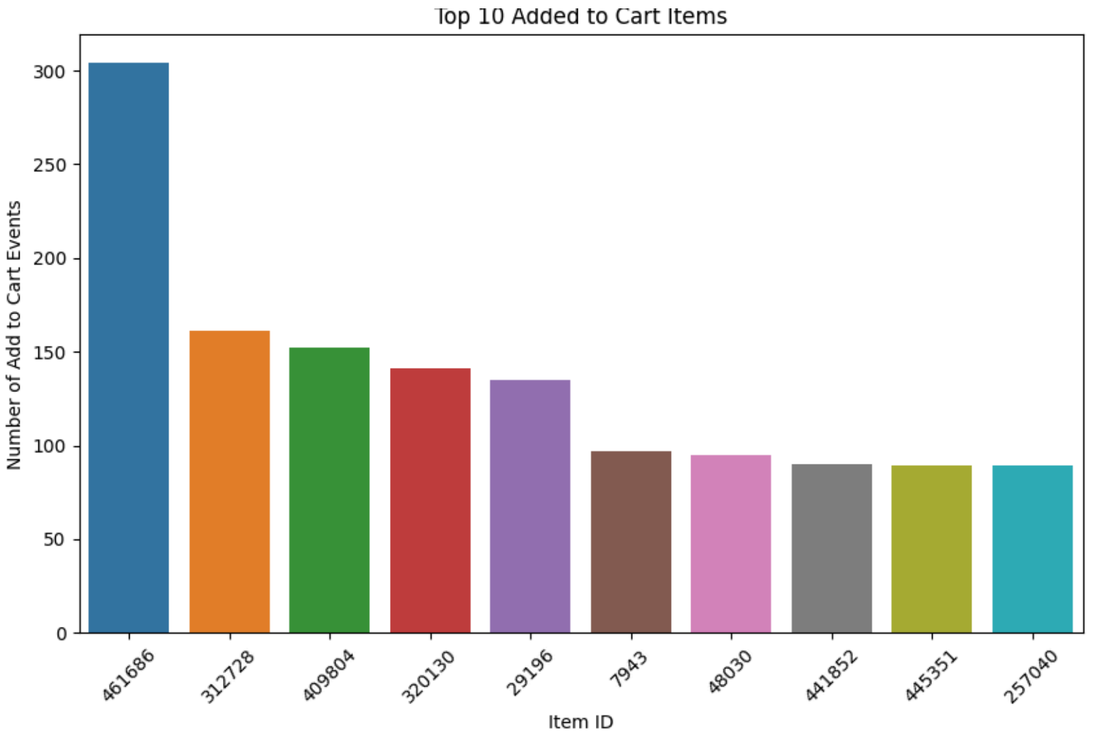


out of 1,419,219 visitors, only **1,1719 (0.83%)** customers made a purchase with the remaining **1,405,700 (99.17%)**either viwing or adding items to cart

Research Questions cont.

2. Top 10 most viewed items and added to cart items

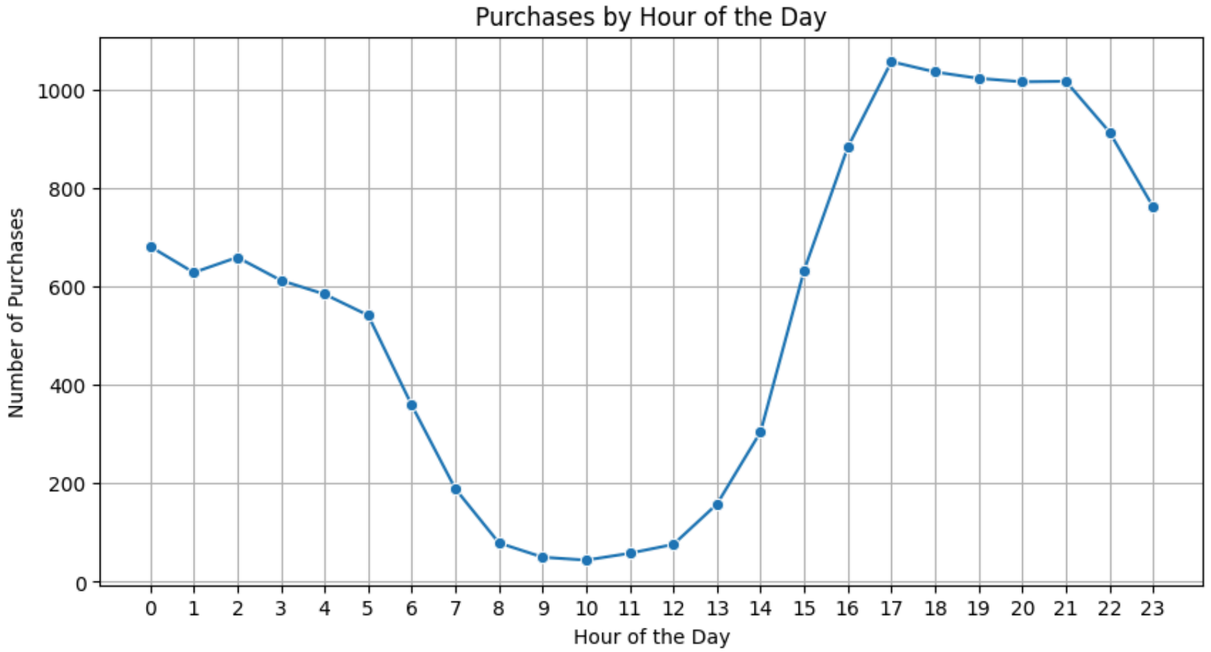


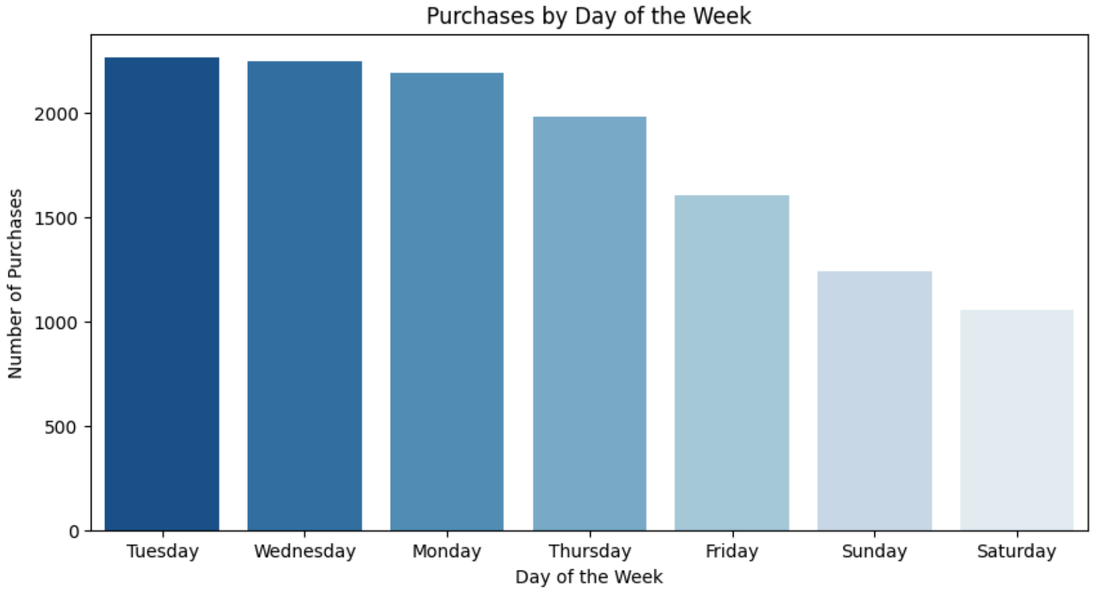


From the graph, the most viewed item is itemid187946, followed by itemid 461686. Although, itemid 187946 is the most viewed, it is not amongst the top 10 most added to cart items. Instead, itemid 461686 is the most added to cart item.

Research Questions cont.

3. Time of the day or day of the week which has the highest number of purchases



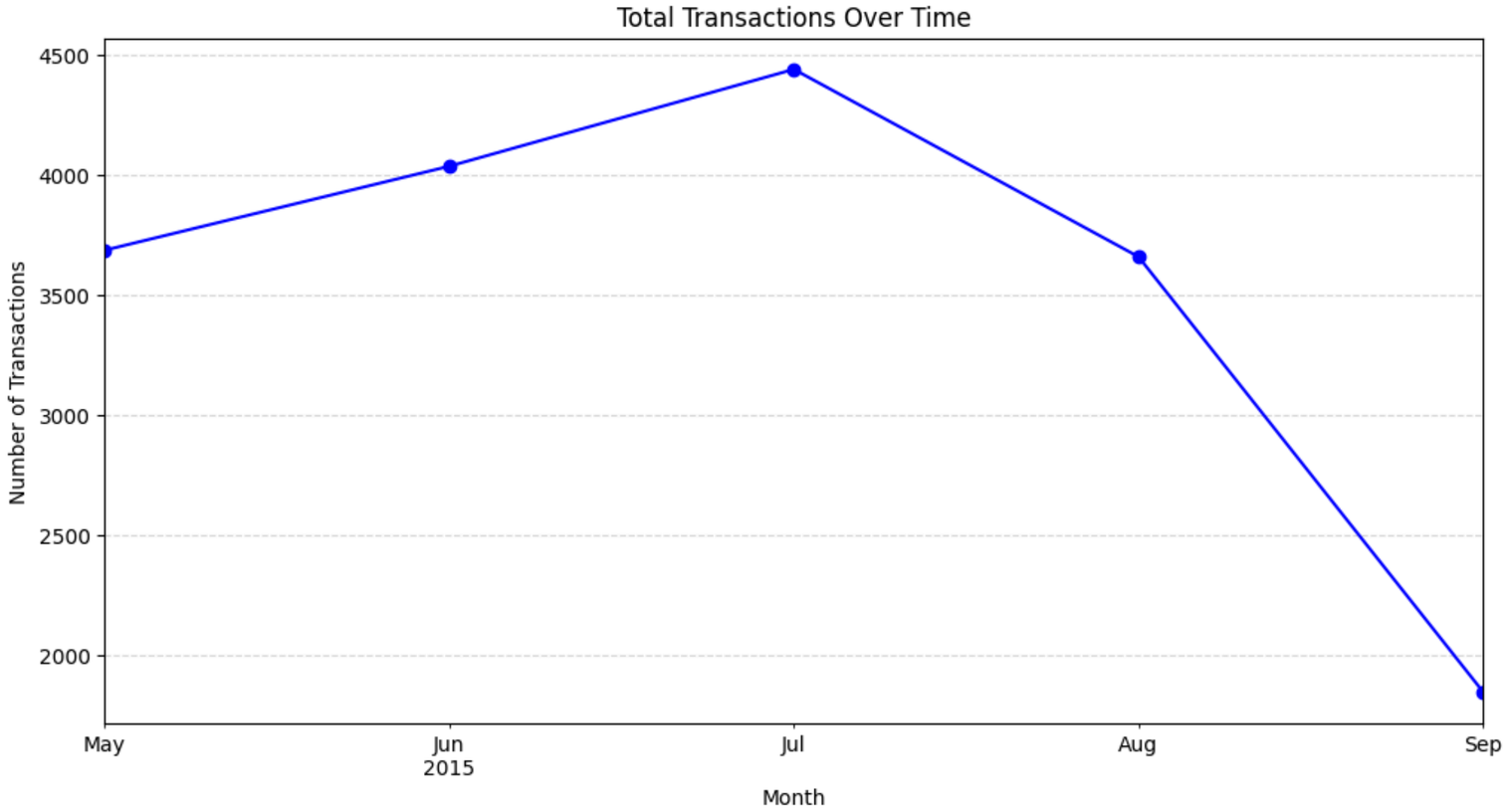


The highest purchase activity occurs between **16:00** and **21:00**, with a sharp peak around 17:00 whiles lowest purchase activity happens in the early morning hours, particularly between 06:00 and 12:00.

Also, purchases peak on **Tuesday** and **Wednesday**, with activity gradually decreasing towards the weekend, reaching the **lowest on Saturday.**

Research Questions cont.

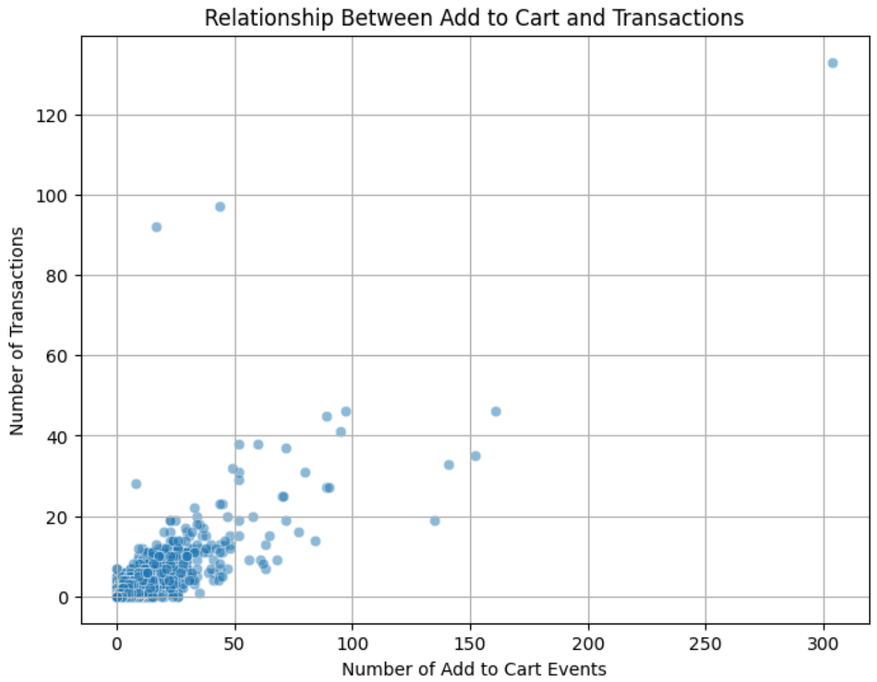
4. Total transactions made over time

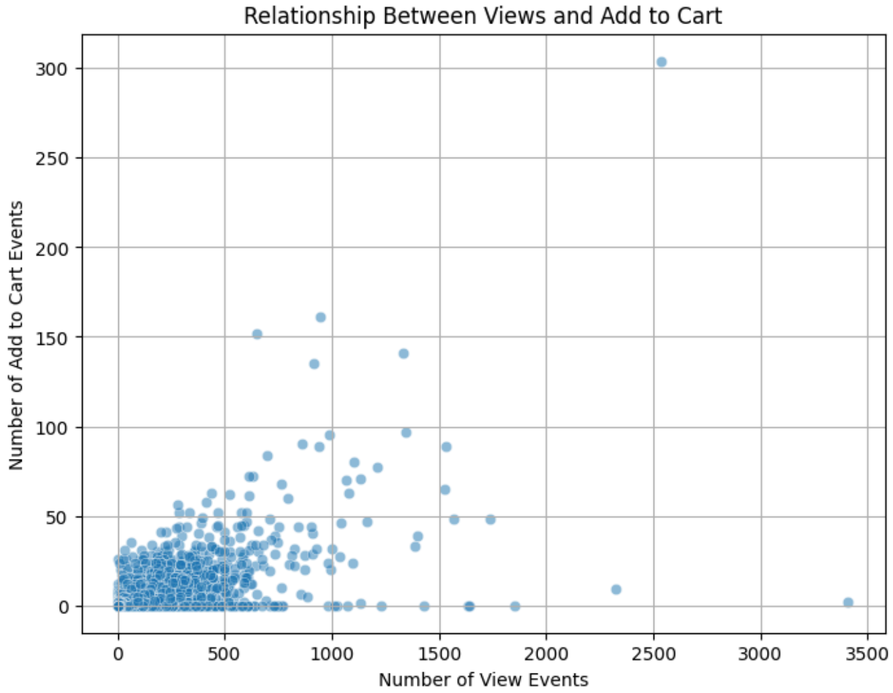


- The number of transactions increased from May to July, peaking in July.
- A sharp decline in transactions is observed from August to September, with September recording the lowest transactions.

Research Questions cont.

5. Relationship between addtocart and transaction & view and addtocart





- There is a positive correlation between add-to-cart events and transactions, but the conversion rate varies, with many products having low transaction counts despite high add-to-cart activity.
- There is a positive but weak correlation between view events and add-to-cart events, with many products experiencing high views but low conversions, indicating drop-off points, while a few outliers show unusually high add-to-cart rates despite lower views.

Preprocessing and Feature Engineering

1

Missing Values

After a final merge of the datasets, missing values in the property and value column were replaced with -1, clustering imputation.

2

User Features

Created user features like, Session length, views, cart count.

3

Item Features

Category, embeddings, popularity.

Anomaly Detection

Why?

Filter unusual user behaviors (bots).

Method

Isolation Forest to detect outliers.

14,059 abnormal users were identified and flagged.

Impact

Reduced noise for better predictions.

Model Selection and Training

1

Approach

Collaborative Filtering

2

Methods

ALS, SVD, DBSCAN considered.

3

Final Choice

Singular Value Decomposition (SVD).

Model Evaluation

MAE

Metric Used with a score of 0.34

RMSE

Metric Used with a score of 0.66

The evaluation results confirm that our model performs well, balancing accuracy and efficiency.

Recommendations

- Optimize marketing campaigns on **Tuesday** and **Wednesday**, since these days see the highest number of purchases.
- Schedule promotions and discounts during peak hours **(16:00–21:00)** to maximize conversions.
- Improve engagement during low activity hours (06:00–12:00) by offering time-limited morning discounts or email campaigns to drive early purchases.
- Implement targeted marketing campaigns or promotions in August and September to retain customer engagement and mitigate sales decline.
- Optimize product pages by improving descriptions, images, and customer reviews to encourage conversions.