GROUP 2 DATA REPORT ON ONLINE RECOMMENDATION SYSTEM

GROUP MEMBERS;

JULIA KARANJA

GRACE ANYANGO

DANIEL KIMUTAI

BRADLEY AZEGELE

PHELMA KANDIE

1. BUSINESS UNDERSTANDING

1.1Overview

Online shopping refers to a delivery and pick-up service. This particular service is offered by companies through a website or a mobile app. The service allows customers to order items from participating retailers with the shopping being done by a personal shopper or company.

Covid-19 has changed the way the world operates. The 'new normal' has also seen dramatic increases in online shopping trends for apparent reasons. Contactless transactions are safer, and lockdowns have given people no choice but to use online platforms to purchase goods. As you know, however, it doesn't last forever. Still, once a customer has experienced the convenience and safety of an initially obligated online shopping experience, they are very likely to use it again even when they don't have to.

Some of the advantages of online shopping include; there are a variety of products, it's convenient, there are a lot of online discounts, there is buyer/shopping secrecy, the customers can avoid crowds, you can find unique/rare products and there is detailed information about the products that you are buying.

To effect some of these advantages of online shopping a recommendation system or recommendation engine can be used for information filtering where it tries to predict the preferences of a user and provide suggestions based on these preferences.

Recommender systems automate personalization on the web enabling individual personalisation of each customer based on their past buying behavior or order as prediction of future buying behavior. The system takes in the effort of the user to personalize the site i.e. help the site adapt to each customer.

1.2 Problem Statement

One of the biggest challenges that new retailers trying to start a business face is that they have a variety of products to choose from in order to come up with their catalog. This process can be tedious as they have to go collect information on customer preferences, cost of items, quality of products, durability among other factors. In most cases these retailers are not able to get this information and therefore this can have a negative impact on their businesses.

We have therefore been appointed Data Scientist for an Online retail store so as to provide a recommendation system that will help both existing and new customers. Our main purpose of creating a recommendation system is to categorize the customers based on their purchase quantities and set this as our guide in recommending the products based on the size of a business. This will help create a better shopping environment for the retailers and cut costs on their initial research and development costs.

1.3 Business Objectives

1.3.1 Main Objective

To provide a recommendation system that will help both existing and new customers. Our main purpose of creating a recommendation system is to categorize the customers based on their purchase quantities and set this as our guide in recommending the products based on the size of a business.

1.3.2 Specific Objectives

1. To particularly provide product recommendations and to look for unique ways to personalize marketing to the customer base.

2. To determine the month with the highest sales so as to plan the optimal time to give discounts

3. To determine the peak hours of the day inorder to better strategize our marketing through online ads.

4. To determine if the unit price affects the quantity demanded

1.4 Business Questions

1.What is the month with the highest sales?

2.What are the peak hours of the day?

3.How does the unit price affect quantity?

1.5 Determining the Project Goals

1.To obtain a prediction for a specific user for a particular item.

2. Introduce a new user and make product recommendations for them.

3. Create a function that will return the top n recommendations for a user .

1.6 Determine the Project Success Criteri

RMSE of around 0.5

1.7 Assessing the situation

The resources to be used for this study include:

computers

Internet connectivity

Computer programming software

Jupyter notebook

Github

Jira

1.8 project plan

Business Understanding

Data Understanding

Data Preparation

Exploratory Data Analysis

Data Preparation

Modeling

Evaluation

Deployment

2.DATA UNDERSTANDING

Using the data we have, we will give insights on how to build recommendation systems based on their purchase quantities and set this as our guide in recommending the products based on the size of a business.

2.1 Data Description

Our data set is from the uci machine learning repository. It is from a non-store online shop selling unique gifts. The data set collected is from 2010 december to 2011 december.
It has eight features namely; invoice no, stock code, description, quantity, invoicedate, unit price, customerid, country.

2.2 Exploring the data

The data set includes fields that represent features that show the different dataset contains 26707 rows and 38 columns with no duplicates. The column definitions are displayed below. Some columns were found irrelevant for our analysis hence they were dropped.

2.3 Data Preparation

The following steps were followed in preparing the data;

v Importing the necessary libraries

v Loading the dataset from the CSV format it was stored in

v Creating a new data frame with the necessary columns for our research

v Cleaning the data

§ Checking for missing values and replacing them with mode

§ Checking for duplicates and renaming them

3 EXPLORATORY DATA ANALYSIS

3.1 UNIVARIATE DATA ANALYSIS

Representation of orders by description
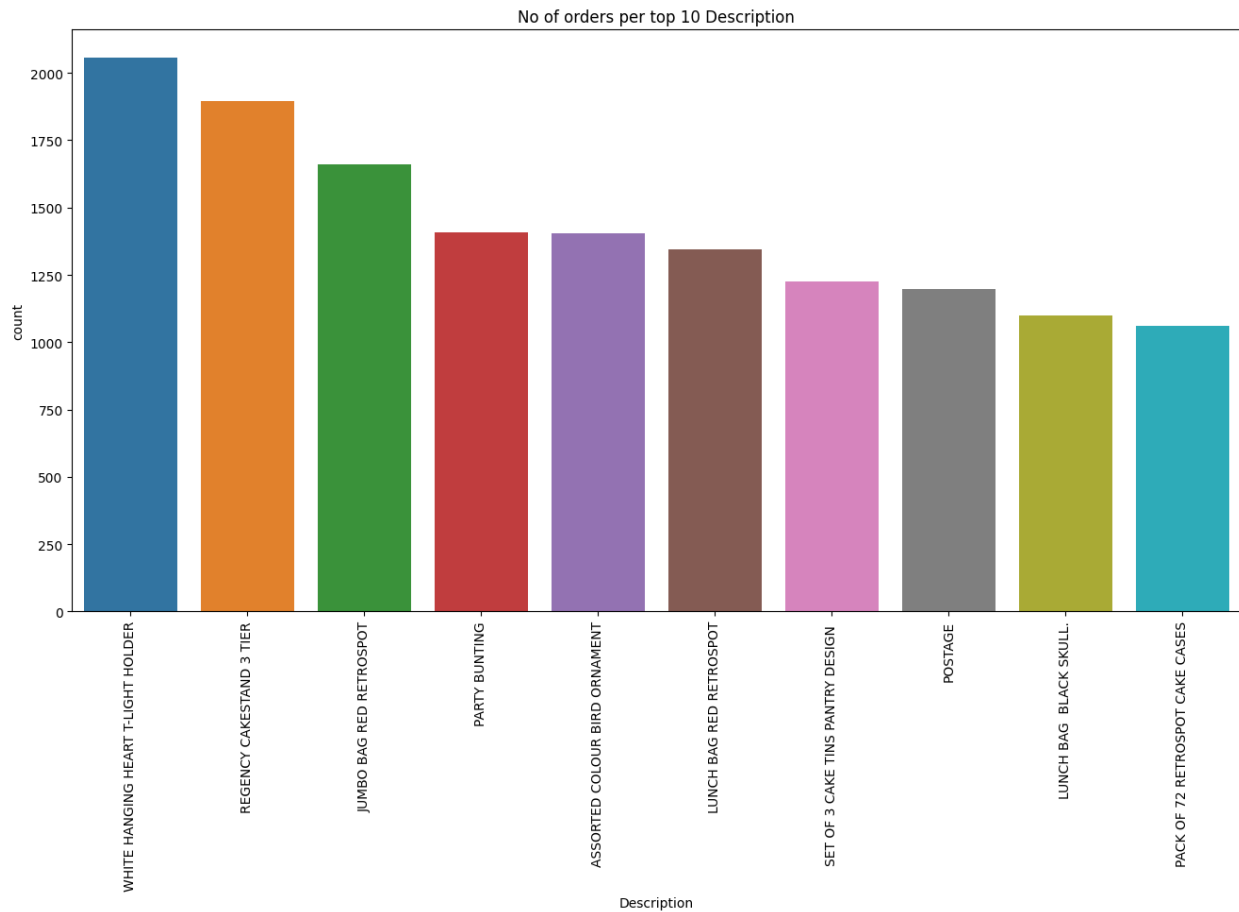
No of orders per top 10 Description

White hanging light holder has the highest number of orders followed by the regency casket. Pack of 72 retrosport cake has the least number of orders
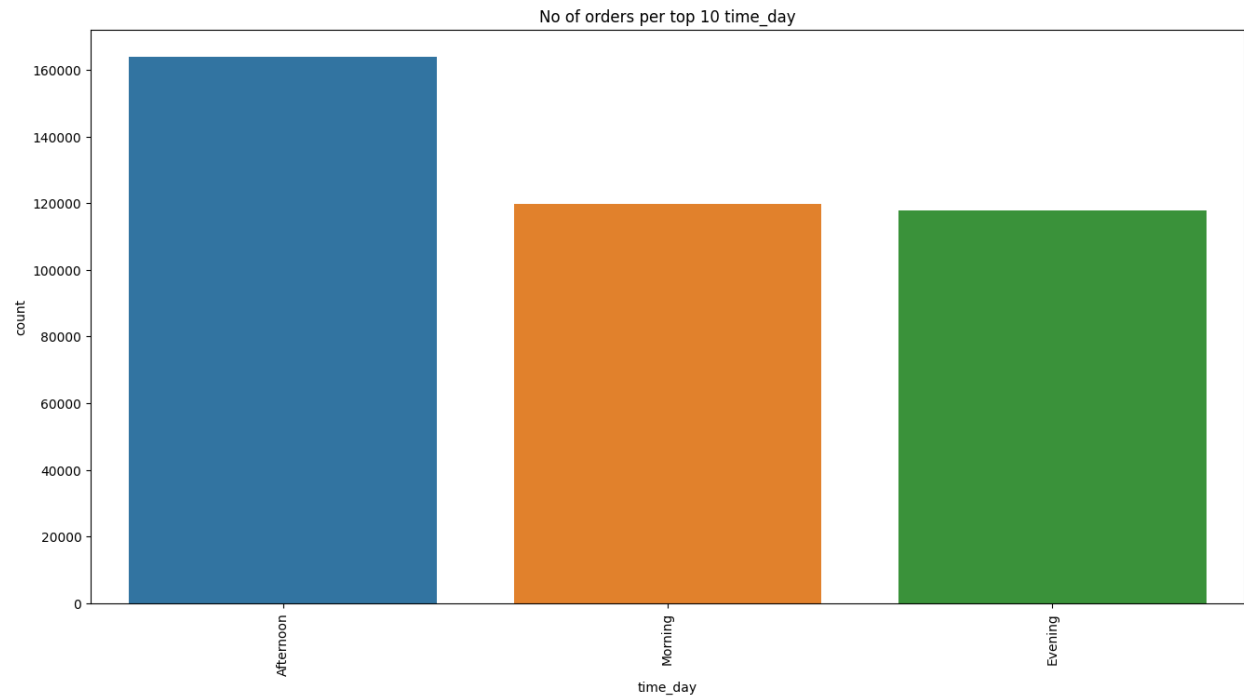
Representation of orders based on country

No of orders per top 10 Country
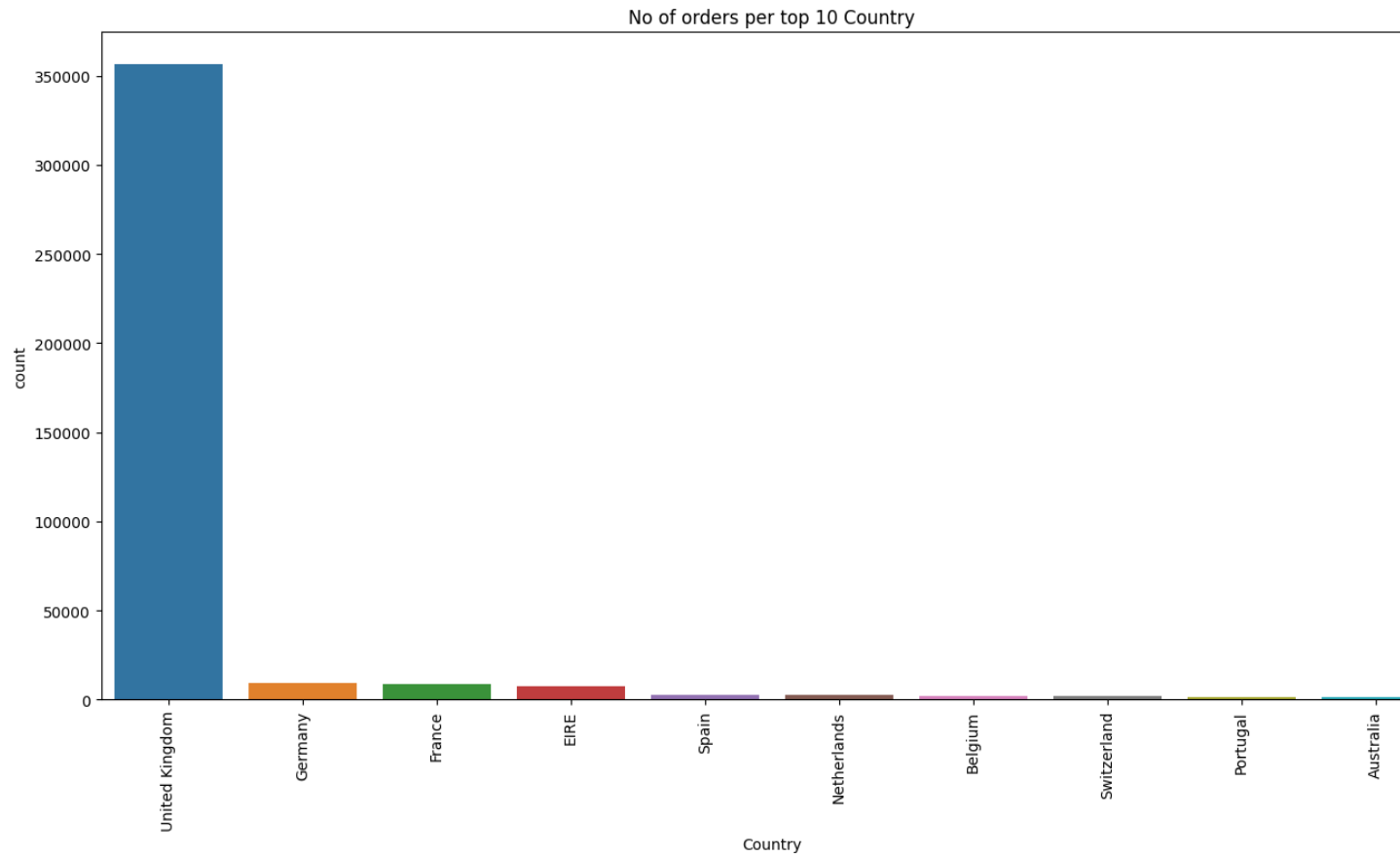
The United Kingdom has the highest number of Orders compared to Germany and other Countries.

**Representation of orders based on months**

No of orders per top 10 Month
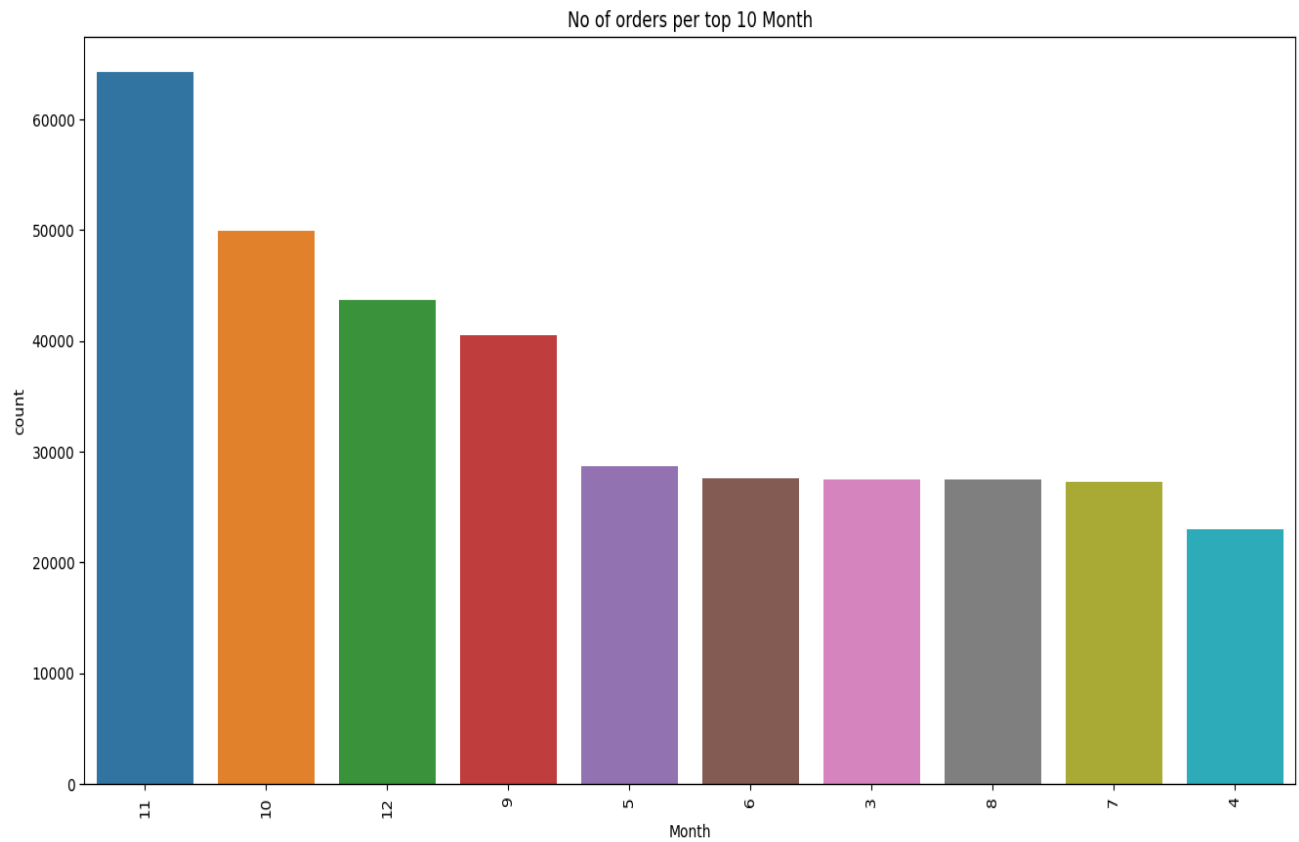
The Month with the highest count of orders is November followed by October then December. The one with the least count of orders are July and April

Representation of orders based on time of day

No of orders per top 10 time_day

Afternoon has the highest number of orders compared to morning and evening

Representation of orders based on days of the week
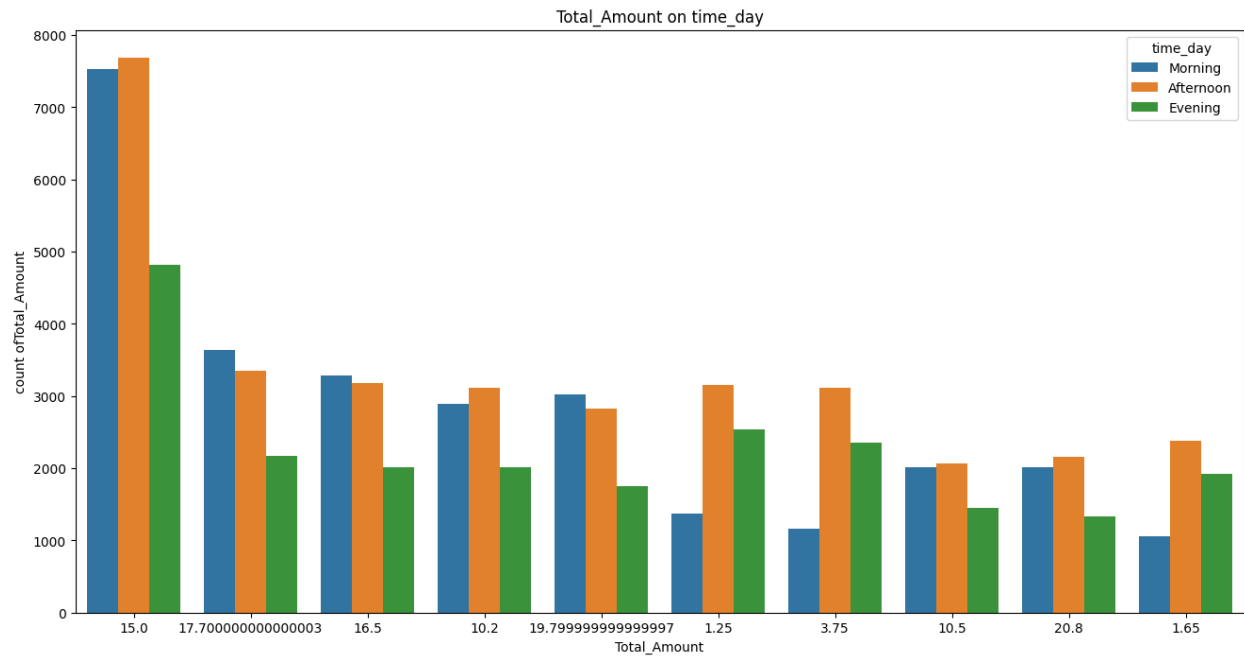

No of orders per top 10 Day_week
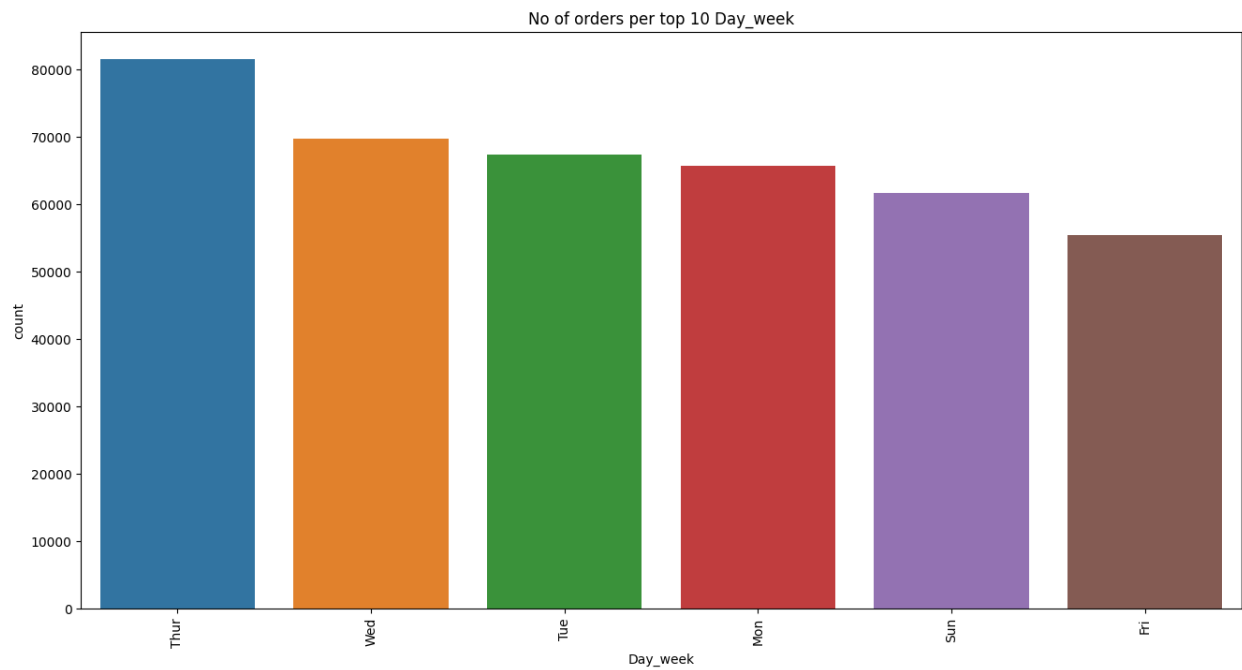
Many orders are made on Thursdays. The least orders are made on Friday.

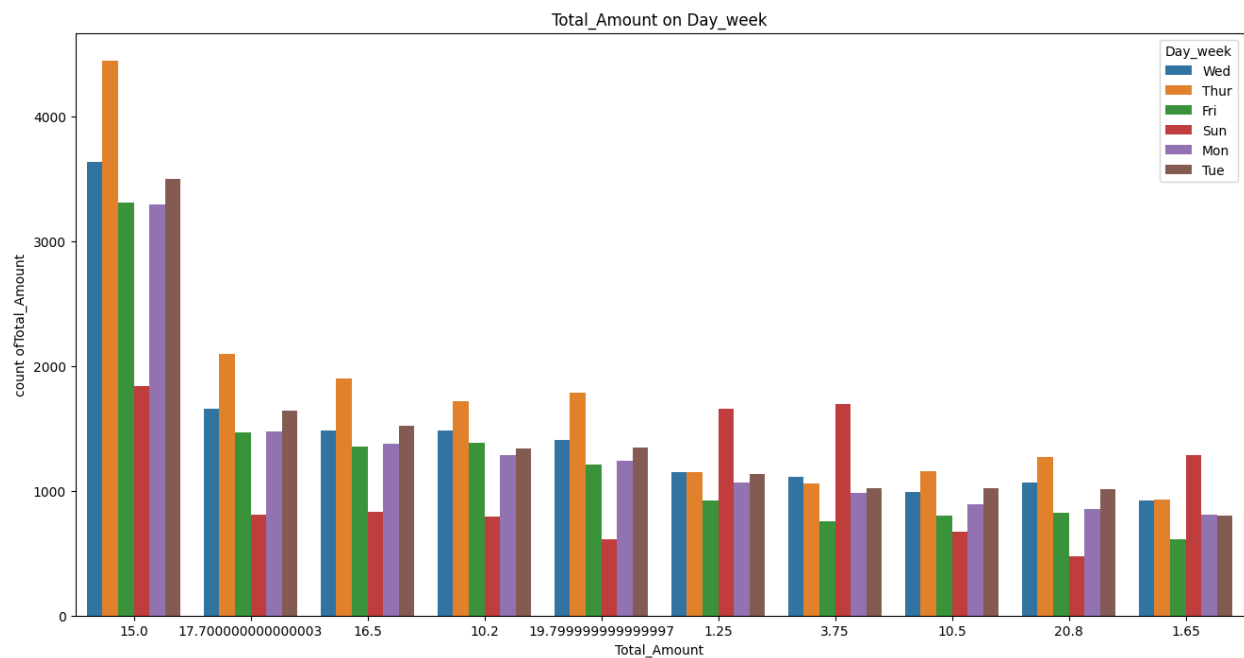## 3.2 BIVARIATE DATA ANALYSIS

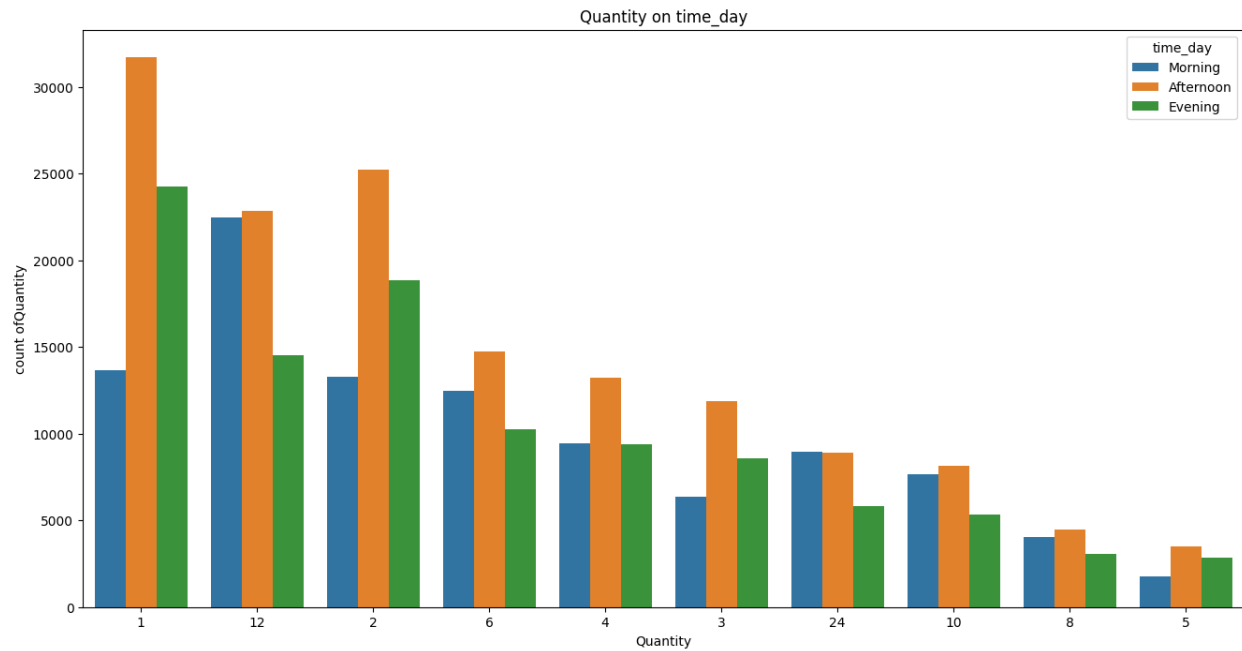## Relationship between total amount and Month

Afternoon has the highest number of sales compared to evening and morning

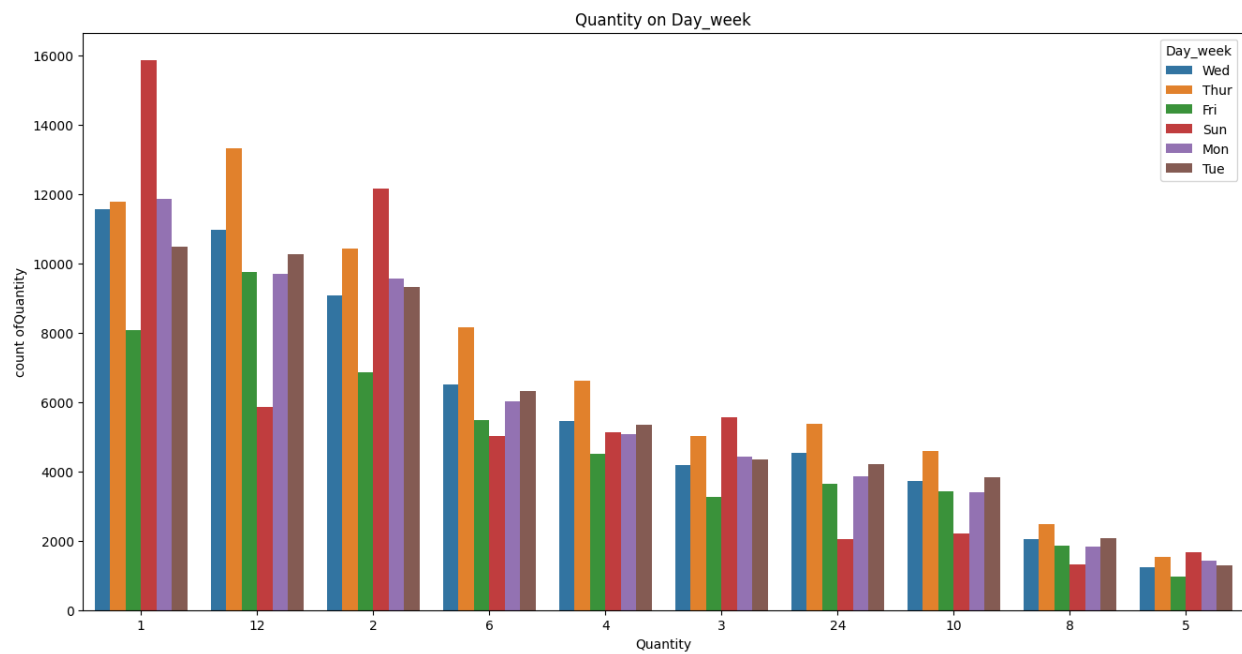## Relationship between total Amount Vs Day

Thursday has the highest number of sales and Friday has the least number of sales.

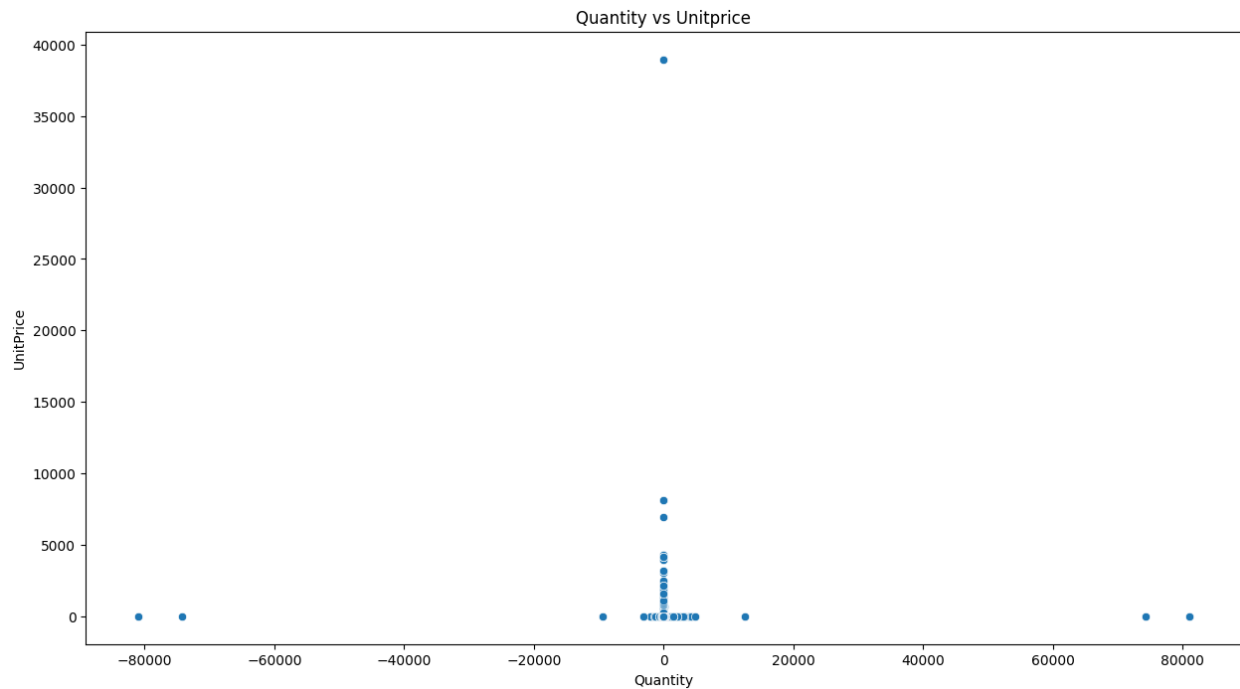**Relationship between quantity and time day**

Quantity on time_day

Afternoon has the highest number of sales compared to Morning and Evening

Relationship between quantity and day of the week

Quantity on Day_week

Thursday has the highest number of Quantity demands.

Relationship between quantity and unit price

There is no correlation between unit price and quantity..

4 DATA PREPROCESSING

This part of the analysis involves looking at trying to find solutions for:

❖ Missing values and all those values that are falsely labeled, but are missing values and some of the rather strangely looking labels

❖ Outliers and noise

❖ Encoding of categorical data

5 MODELING

5.1 Models Implemented

Collaborative Filtering

We used Cosine Similarity to recommend various Products to customers.

Memory_Based

We used the SVD model to create our model and test our data.

5.2 Findings

Memory based

We used Cosine Similarity to recommend various Products to customers. In our prediction we got an RMSE of 0.0655.

Model based
We used the SVD model to create our model and test our data. Here we got an RMSE of 0.3954

5.3 Evaluation
Our model meets the business success criteria which is less than 0.5 RMSE

6.CONCLUSION

The recommendation system we developed will be very helpful to the customers and e-commerce companies to recommend products based on experience with the products

7.LIMITATIONS

Collaborative Filtering gives solid recommendation systems, and at the same time requires fewer details than possible. However, it has a few limitations in some particular scenarios.

.It has a cold start like when a new item comes in, that cannot be recommended unless until it is rated or reviewed by the customer and also it doesn't have any extra

information of the product either to give any recommendations. Collaborative Filtering lacks transparency and explainability of this level of information.

To overcome these kinds of scenarios, I would recommend a Hybrid recommender which suggests content-based filtering and collaborative-based filtering recommendations which would be quite effective

8.OBSERVATIONS

- The month of November has the highest sales. That is the best time to give a discount.
- The peak hours of the day is the afternoon, that is the best time to strategize our marketing through online ads.
- White hanging light holder has the highest number of orders followed by the regency casket. Pack of 72 retrosport cake has the least number of orders
- Sundays and Thursdays have the highest quantity ordered compared to other days.
- There is no clear relationship between price and Quantity. Therefore price does not not determine quantity ordered.
- KNN (K-Nearest Neighbours) With Means model has an test RMSE value of 0.244 and cross validation RMSE value of 0.246.
- The SVD (Singular Value Decomposition) model has a test RMSE score of 0.364 and cross validation (CV) RMSE score of 0.363.

As for the recommendation, each client will have a variety of products suggested to them as they are gathered filling out missing entries in the matrix during matrix factorization using SVD.

.

.