Deriving Payment Insights Using Data Analytics

Enhancing insights using Big Query ,Open-refine And Power BI

ADTA 5240

Harvesting, storing and Retrieving Data

Nisarga Shivaprakash Aradhya (11818537)
Nitin Reddy Balaiahgari (11698724)
Himabindu Chunduri (11724307)
Jaya Prakash Reddy Munagala (11735187)
Harshini Vundavalli (11706868)

Company : ROCKSTART

- **Company Name : RockStart**

RockStart is U.S. Based clothing company started in 2017 with stores spread across the country within different states.

Employing over 2500 employes as of 2023 across the country in all stores.

The company covers a wide range of products in the clothing sector for males and females including other accessories like Watches, chains, shoes etc.

A leading fashion brand, RockStart offers a broad selection of clothing and lifestyle items for people of all ages by fusing comfort, style, and quality.

Note : This is a hypothetical company created only for data analysis and Class presentation requirement purposes

Problem Statement

- Purchase history analysis helps us increase seasonal shopping sales

Business Understanding

1) Who is asking for the data?

Audience: Clothing company trying to understand seasonal sales Based on gender

Purpose: to understand the differences between male and female customer behaviours to increase business sales.

2) What is the business trying to find out?

1)Spending Patterns: What are the different spending patterns among males and females

2)Seasonal Trends: Different spending patterns between males and females depending on the shopping season

3)Promo Code Usage: Which gender uses more promo codes male or females

4) Subscription usage: Who are more engages in loyalty programs males or females?

Data Collection

Customer Shopping Trends Dataset

Journey into Consumer Insights and Retail Evolution with Synthetic Data

Last Updated: a year ago (Version 2)

About this Dataset

Context

The Customer Shopping Preferences Dataset offers valuable insights into consumer behavior and purchasing patterns. Understanding customer preferences and trends is critical for businesses to tailor their products, marketing strategies, and overall customer experience. This dataset captures a wide range of customer attributes including age, gender, purchase history, preferred payment methods, frequency of purchases, and more. Analyzing this data can help businesses make informed decisions, optimize product offerings, and enhance customer satisfaction. The dataset stands as a valuable resource for businesses aiming to align their strategies with customer needs and preferences. It's important to note that this dataset is a Synthetic Dataset Created for Beginners to learn more about Data Analysis and Machine Learning.

Content

This dataset encompasses various features related to customer shopping preferences, gathering essential information for businesses seeking to enhance their understanding of their customer base. The features include customer age, gender, purchase amount, preferred payment methods, frequency of purchases, and feedback ratings. Additionally, data on the type of items purchased, shopping frequency, preferred shopping seasons, and interactions with promotional offers is included. With a collection of 3900 records, this dataset serves as a foundation for businesses looking to apply data-driven insights for better decision-making and customer-centric strategies.

shopping_trends.csv (453.25 kB)

Detail Compact Column 10 of 19 columns

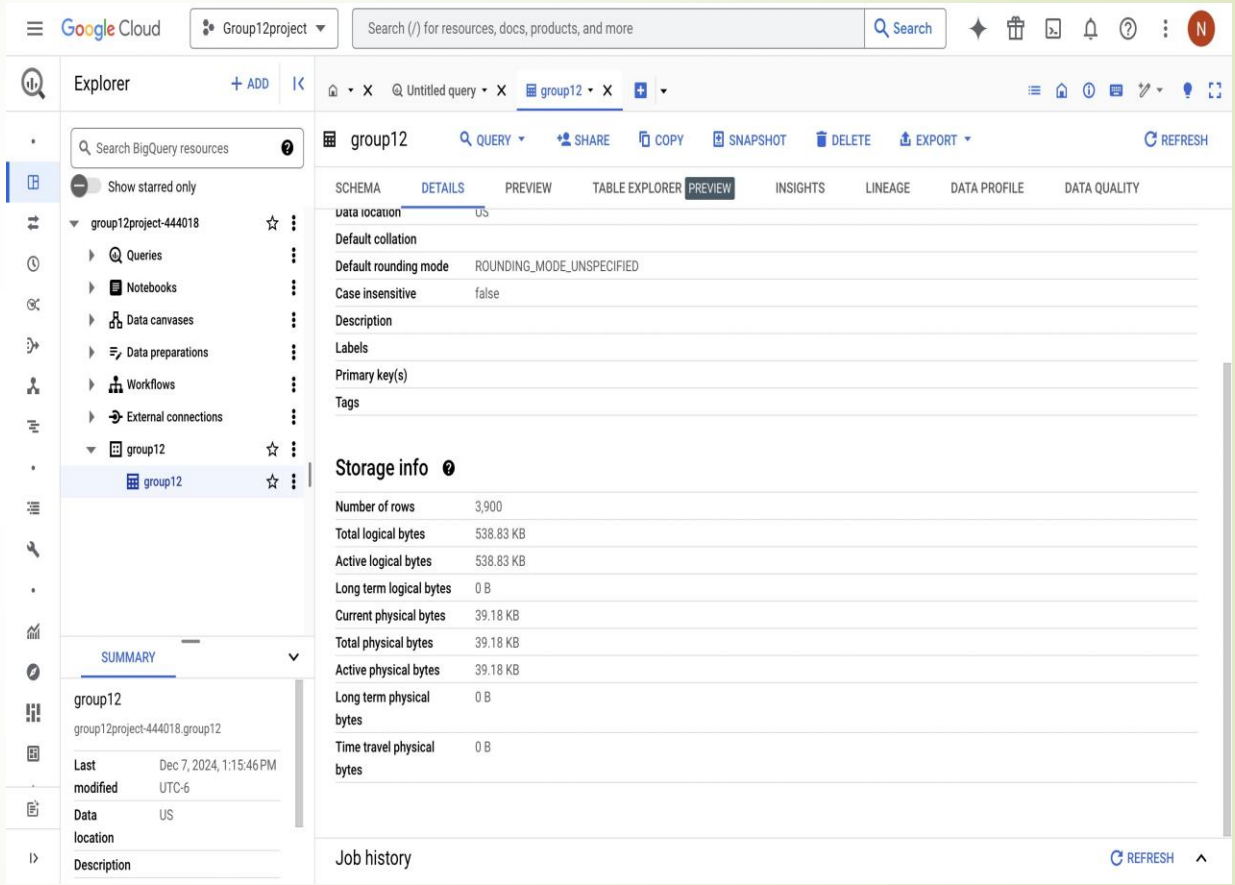
Customer Shopping Trends Dataset

Customer ID	Age	Gender	Item Purchased	Category
Customer ID	Age	Gender	Item Purchased	Category
1 3900	18 70	Male 68% Female 32%	Blouse 4% Pants 4% Other (3558) 91%	Clothing Accessories Other (923)
1	55	Male	Blouse	Clothing
2	19	Male	Sweater	Clothing
3	50	Male	Jeans	Clothing
4	21	Male	Sandals	Footwear
5	45	Male	Blouse	Clothing
6	46	Male	Sneakers	Footwear

Data set source :
https://www.kaggle.com/

Data storage

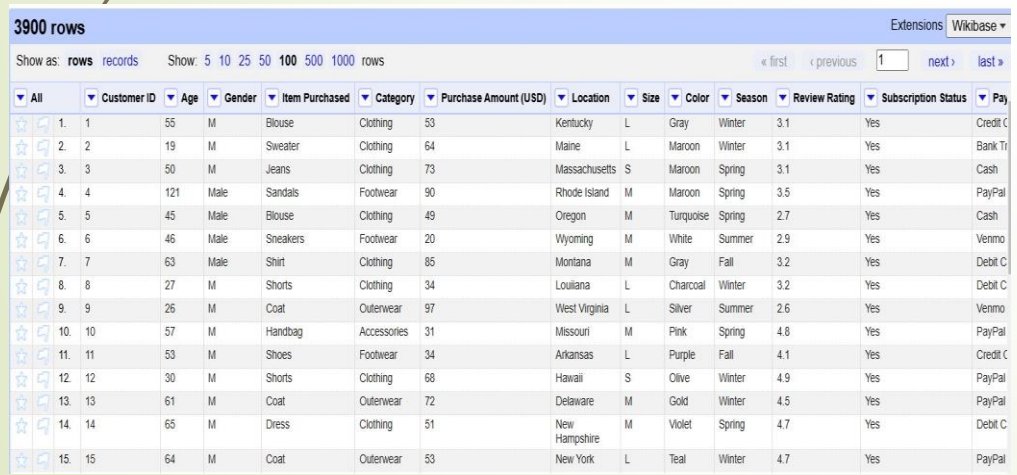
- We have used Big Query a data warehousing service offered by google cloud platform to store our data
- We have created a new project named group 12 and enabled BigQuery in the same project
- Next the dataset has been created named as group 12 and has been used for further configuration after uploading the data



Data Cleaning using Open Refine - Age

- **Cleaned the Age column:**
- Data set contained mispresenting values in age like 121.
- Replaced the misrepresenting data with average age 44 using the Edit cells and Replace values.
- Dataset had 20 missing values in age column.
- Replaced the missing values with average age 44 using Text facet edit values option.

Before Changes



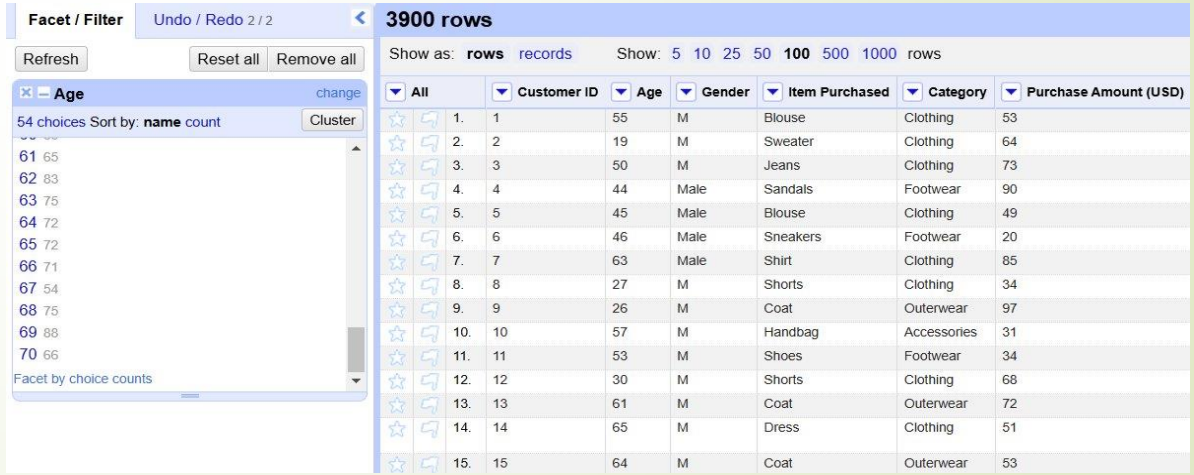
3900 rows

Extensions: Wikibase

Show as: **rows** records Show: 5 10 25 50 **100** 500 1000 rows « first ‹ previous 1 next › last »

All	Customer ID	Age	Gender	Item Purchased	Category	Purchase Amount (USD)	Location	Size	Color	Season	Review Rating	Subscription Status	Pay
1.	1	55	M	Blouse	Clothing	53	Kentucky	L	Gray	Winter	3.1	Yes	Credit C
2.	2	19	M	Sweater	Clothing	64	Maine	L	Maroon	Winter	3.1	Yes	Bank Tr
3.	3	50	M	Jeans	Clothing	73	Massachusetts	S	Maroon	Spring	3.1	Yes	Cash
4.	4	121	Male	Sandals	Footwear	90	Rhode Island	M	Maroon	Spring	3.5	Yes	PayPal
5.	5	45	Male	Blouse	Clothing	49	Oregon	M	Turquoise	Spring	2.7	Yes	Cash
6.	6	46	Male	Sneakers	Footwear	20	Wyoming	M	White	Summer	2.9	Yes	Venmo
7.	7	63	Male	Shirt	Clothing	85	Montana	M	Gray	Fall	3.2	Yes	Debit C
8.	8	27	M	Shorts	Clothing	34	Louiana	L	Charcoal	Winter	3.2	Yes	Debit C
9.	9	26	M	Coat	Outerwear	97	West Virginia	L	Silver	Summer	2.6	Yes	Venmo
10.	10	57	M	Handbag	Accessories	31	Missouri	M	Pink	Spring	4.8	Yes	PayPal
11.	11	53	M	Shoes	Footwear	34	Arkansas	L	Purple	Fall	4.1	Yes	Credit C
12.	12	30	M	Shorts	Clothing	68	Hawaii	S	Olive	Winter	4.9	Yes	PayPal
13.	13	61	M	Coat	Outerwear	72	Delaware	M	Gold	Winter	4.5	Yes	PayPal
14.	14	65	M	Dress	Clothing	51	New Hampshire	M	Violet	Spring	4.7	Yes	Debit C
15.	15	64	M	Coat	Outerwear	53	New York	L	Teal	Winter	4.7	Yes	PayPal

After changes



Facet / Filter | Undo / Redo 2 / 2

Refresh | Reset all | Remove all

Age change

54 choices Sort by: **name** count | Cluster

61 65
62 83
63 75
64 72
65 72
66 71
67 54
68 75
69 88
70 66

Facet by choice counts

3900 rows

Show as: **rows** records Show: 5 10 25 50 **100** 500 1000 rows

All	Customer ID	Age	Gender	Item Purchased	Category	Purchase Amount (USD)
1.	1	55	M	Blouse	Clothing	53
2.	2	19	M	Sweater	Clothing	64
3.	3	50	M	Jeans	Clothing	73
4.	4	44	Male	Sandals	Footwear	90
5.	5	45	Male	Blouse	Clothing	49
6.	6	46	Male	Sneakers	Footwear	20
7.	7	63	Male	Shirt	Clothing	85
8.	8	27	M	Shorts	Clothing	34
9.	9	26	M	Coat	Outerwear	97
10.	10	57	M	Handbag	Accessories	31
11.	11	53	M	Shoes	Footwear	34
12.	12	30	M	Shorts	Clothing	68
13.	13	61	M	Coat	Outerwear	72
14.	14	65	M	Dress	Clothing	51
15.	15	64	M	Coat	Outerwear	53

Data Cleaning using Open Refine - Gender

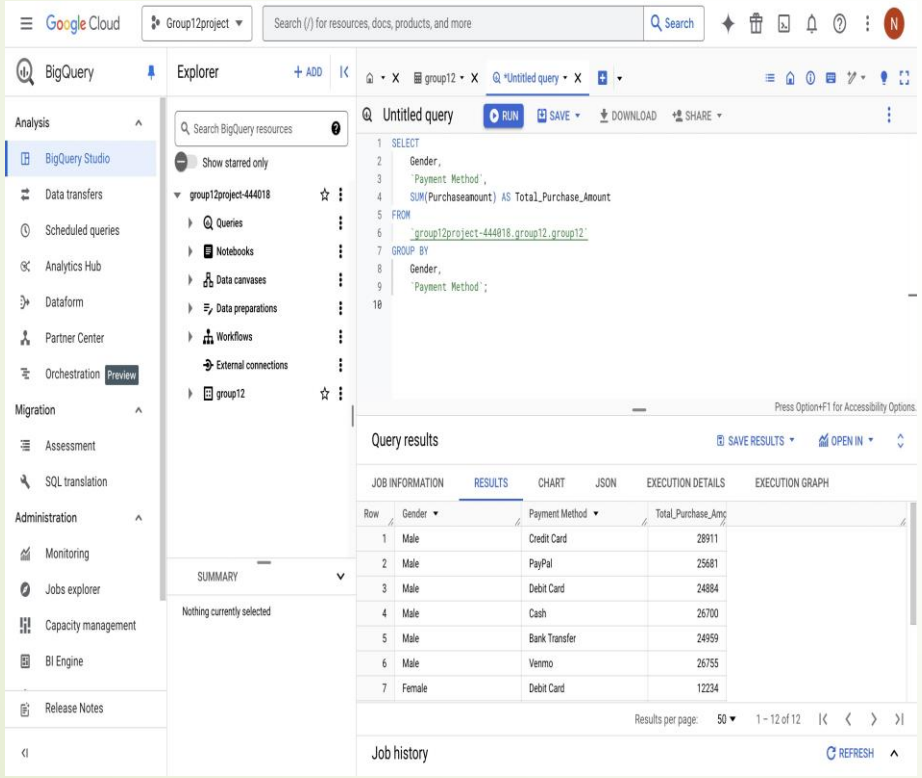
- **Cleaned the Gender column:**
- Gender columns contains various values like ‘M’, ‘Male’, ‘F’, ‘Female’
- Assumed M is Male and F is Female
- Replaced M as Male and F as Female using text facet and edit options.

Before changes

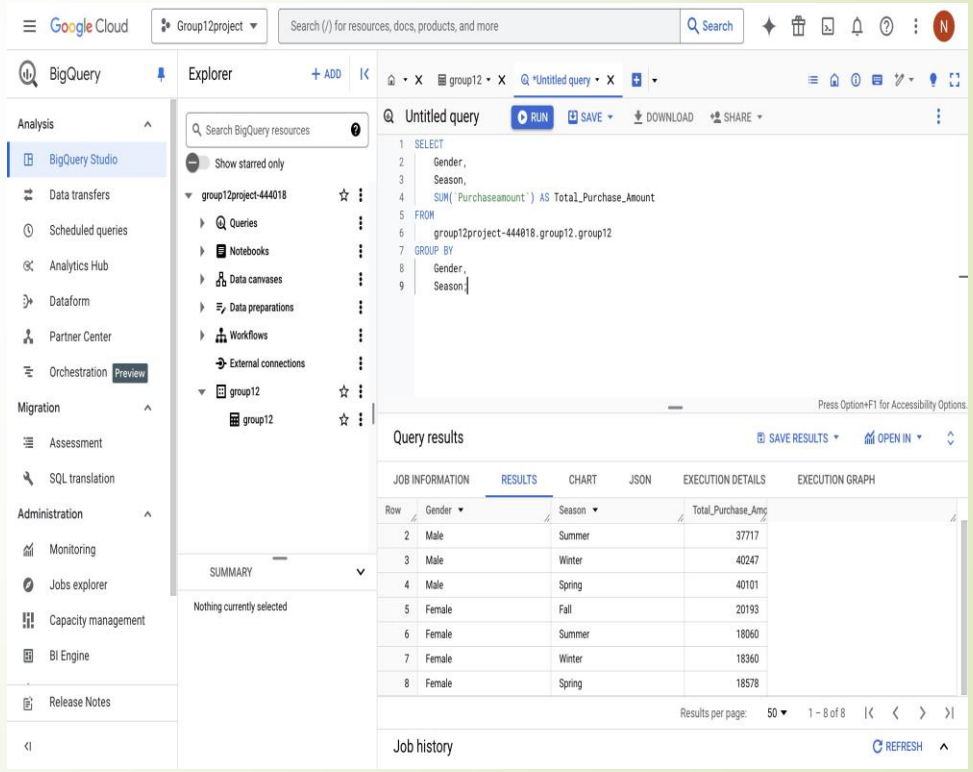
After changes

Data Analysis: Writing Queries to generate insights

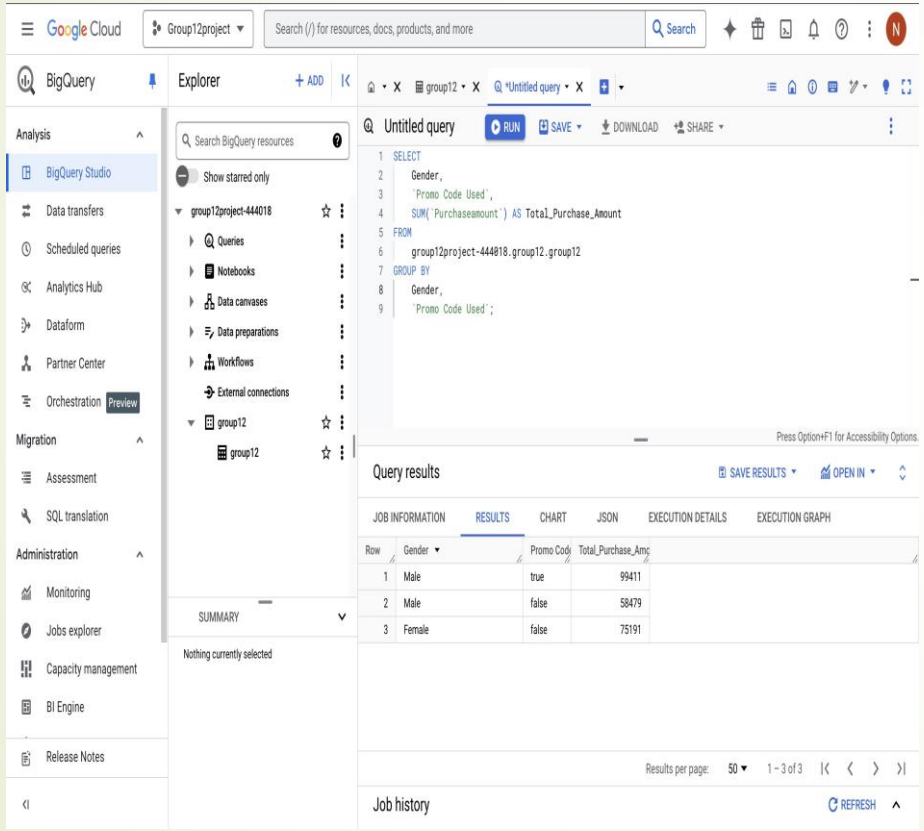
Total insights from payment methods(biased by gender)



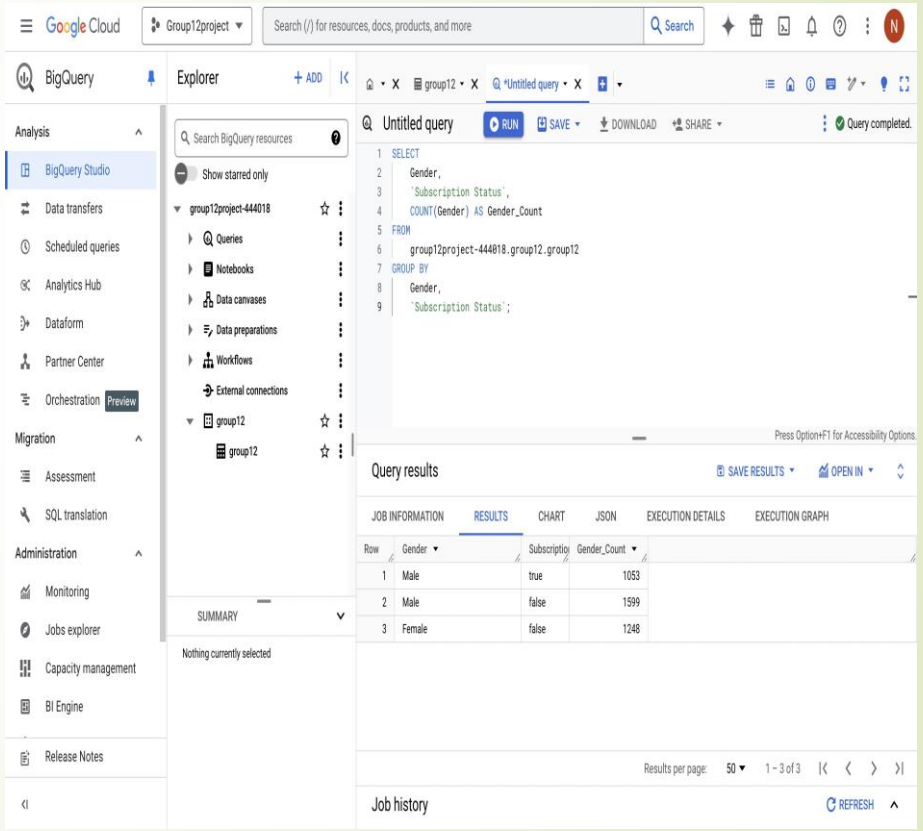
Overall Shopping spending in every season(biased by gender)



Usage of promo Codes (biased by gender)

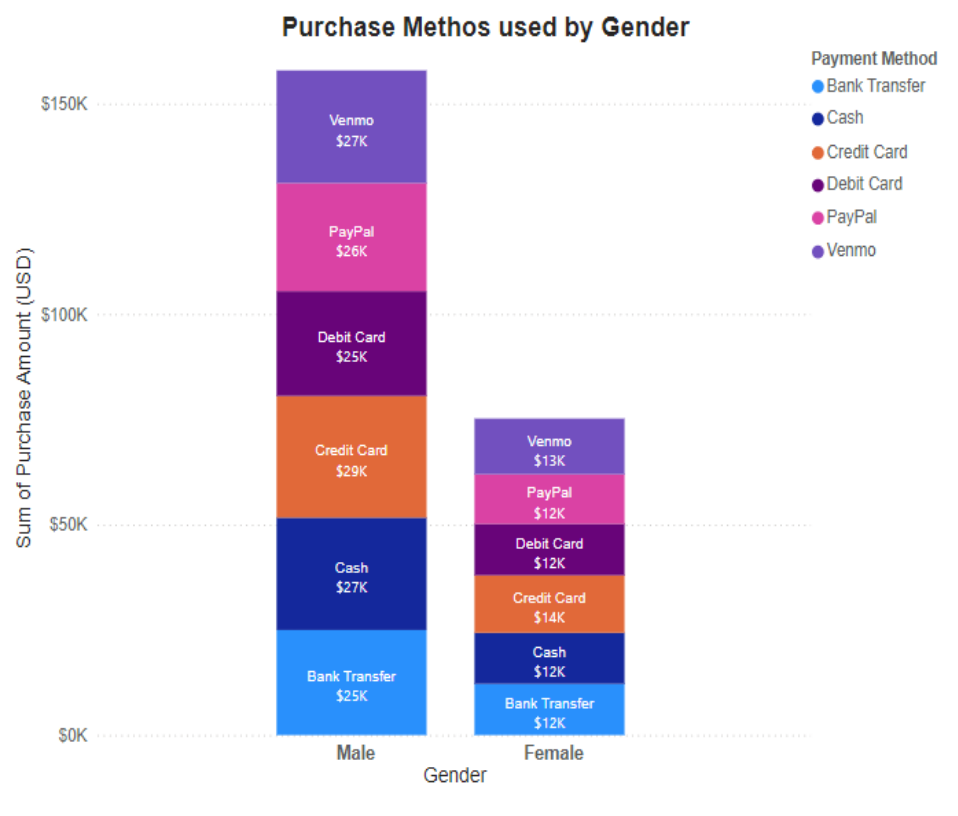


Subscription of Accounts (biased by gender)

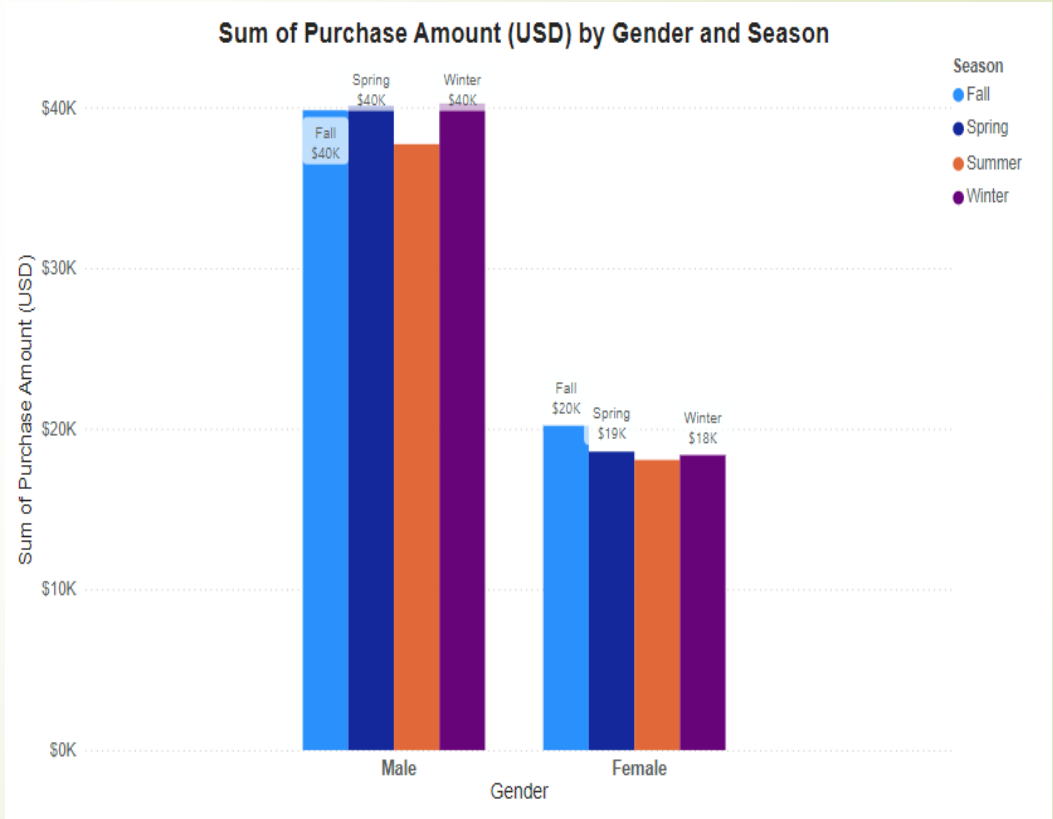


Data visualization (using power BI)

Total insights from payment methods(biased by gender)

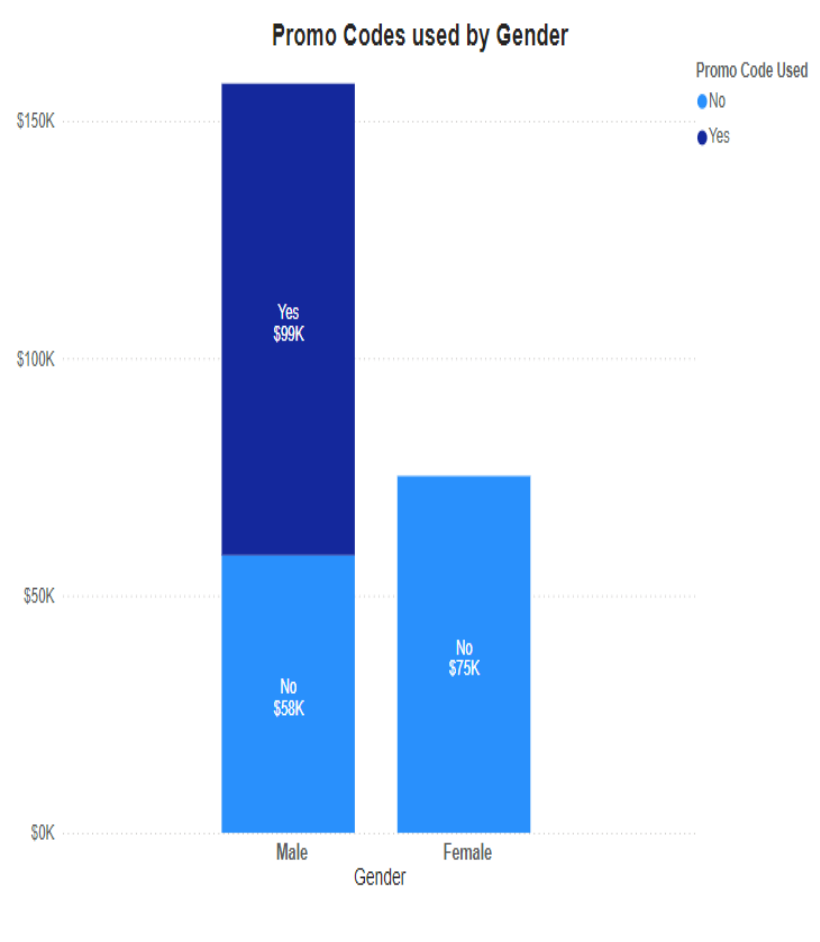


Overall Shopping spending in every season(biased by gender)

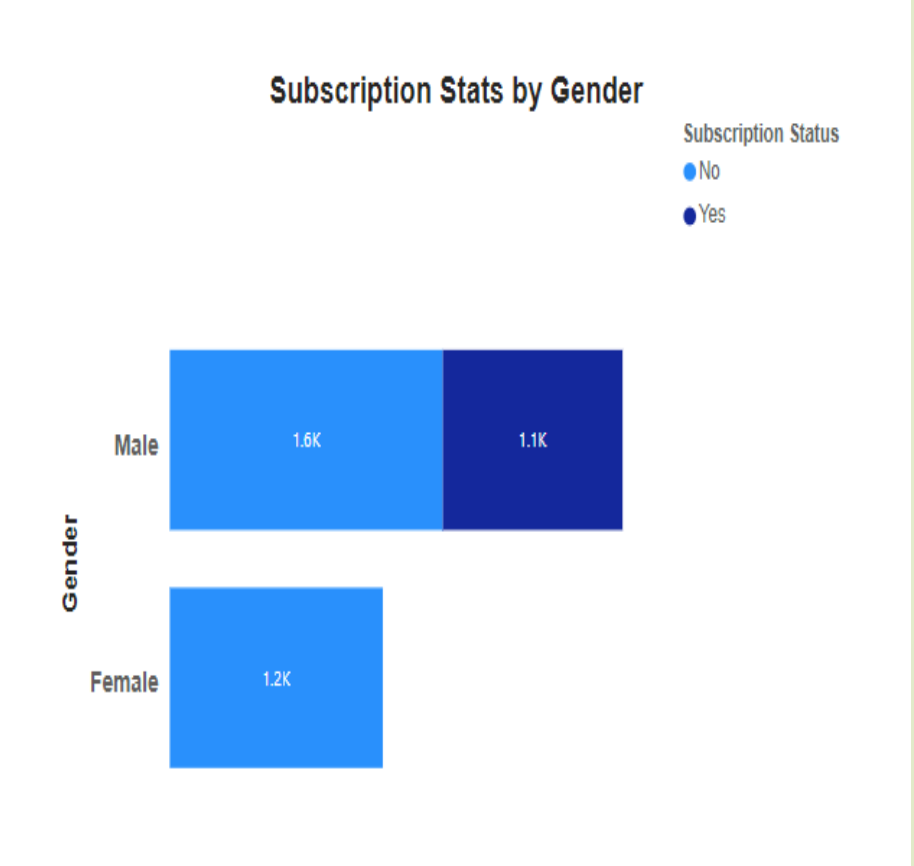


Munagala

Usage of promo Codes (biased by gender)



Subscription of Accounts (biased by gender)



Insights Gained

According to the charts above male customers tend to spend double of females using every payment gateways

According to the visualization, 63% (99K/157K) of the times males used a promo code when they were purchasing with the store

According to the visualization females tend to buy more clothes in the season of fall Whereas, males spend more money on clothes in spring and winter

Males subscribed to more store accounts than female did to earn points after their every purchase.

Recommendations to improve sales

Releasing more male garments and putting up male clearance sale during spring and winter , likewise female garments during fall improve clearance sale

Sending promo codes to male customer base during discount offer times increases the sales

Female are not using the Promo Codes OR subscriptions. So, We should advertise and promote among Females

Conclusion

We have Enhanced the sales by providing suggestions using PowerBi Studio,BigQuery and Open Refine

We have successfully enhanced the analysis using different tools and implemented the data lifecycle

Team Members and Roles

- Nisarga Shivaprakash Aradhya - Data Collection, Data Cleaning using Open Refine
- Nitin Reddy Balaiahgari - Data storage and Writing Queries to generate insights
- Himabindu Chunduri - Writing Queries to generate insights, Future Sales and Recommendations
- Jaya Prakash Reddy Munagala - Data visualization, Business Understanding
- Harshini Vundavalli - Data visualization, Insights Gained

Presentation Link

- https://drive.google.com/file/d/1hWw5JiBQG2xk9ZuFEbenpfYRLhUZNALp/view?usp=sharing

thanks