Post Graduate Certificate Programme in Data Science For Business Excellence And Innovation (DBI002)
Place: IIM Nagpur
Year: 2022-23
Capstone Project Presented by
Group 06

Sr.	Name of Group Members	Email ID
1	Bhagyashree Vyas	dbi002_bhagyashree@iimnagpur.ac.in
2	Chacko John Kuntharayil	dbi002_chacko@iimnagpur.ac.in
3	Rohan Wanzkhade	dbi002_rohan@iimnagpur.ac.in
4	Sagar Sambashiv Falke	dbi002_sagar@iimnagpur.ac.in
5	Saurabh Sheetlaprasad Dubey	dbi002_dubey@iimnagpur.ac.in

Black Friday Sales

What is Black Friday Sale?

Black Friday is the term for the Friday after the Thanksgiving in US.It is tradionally marks the start of the Christmas shopping season in the US.Many stores offer highly promoted sales at a discounted priceand often open early or sometimes as early as midnight or even on Thankgiving.

Approach:
Python Programming: Where We will Be Importing Libraries like **Pandas** And for EDA **Matplotlib** And **Seaborn.**

Data Visualization tool : **Power Bi**

Project Aim: Analysing the Black Friday sales to understand the customer behaviour based on their demographics and product details.

Project objectives:

1. Create a Buyers Demographics dashboard with charts showing the count of customers in different categories in each demographic's variable.
2. Create a Spending Dashboard to showcase the amount of money spent by different demography of buyers. This can include charts such as marital status and age groups vs total spending, occupation and gender vs total spending, etc.
3. On a Conclusion dashboard mentioning observation and findings from the analysis.

Problem Statement:

The problem to be addressed is to analyse and interpret Black Friday sales data to gain insights into customer behaviour and preferences. This includes identifying the most popular products, the purchasing patterns of different customer segments. Additionally, the analysis could focus on identifying any opportunities to increase sales or improve customer satisfaction during future Black Friday sales events.

The dataset consists of 537577 rows (transactions) and 12 columns (features) described below:

User_ID: This column represents the unique ID of each user. There are a total of 5891 users in the dataset.

Product_ID: This column represents the unique ID of each product. There are a total of 3623 products in the dataset.

Gender: This column indicates the gender of the person making the transaction.

Age: This column indicates the age group of the person making the transaction.

Occupation: This column shows the occupation of the user, already labelled with numbers 0 to 20.

City_Category: This column represents the user's living city category. The cities are categorized into three different categories: 'A', 'B', and 'C'.

Stay_In_Current_City_Years: This column indicates how long the user has lived in the current city.

Marital_Status: This column is 0 if the user is not married and 1 otherwise.

Product_Category_1 to Product_Category_3: These columns represent the categories of the product. All three categories are already labelled with numbers.

Purchase: This column represents the purchase amount.

Python Programming
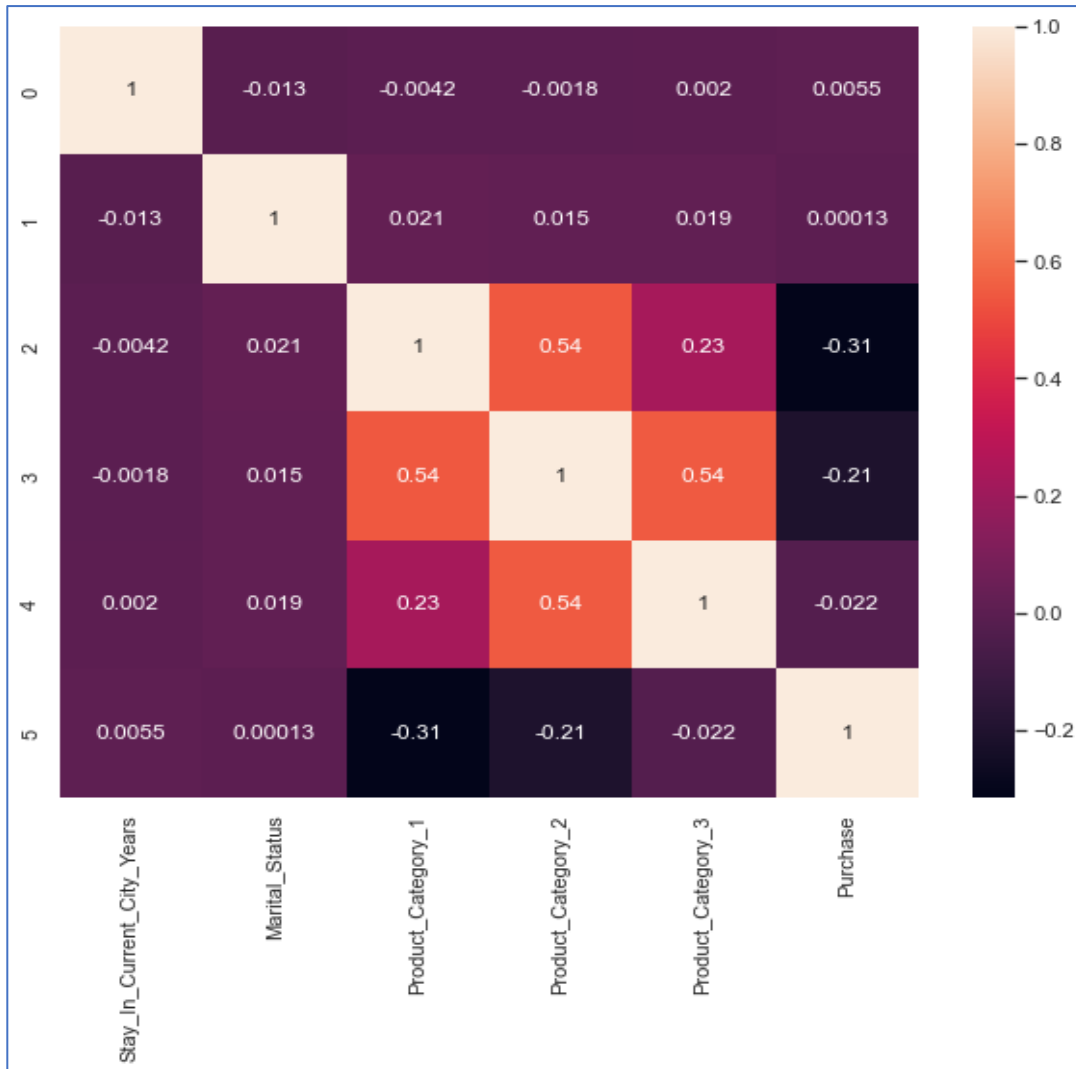

```python
# Import libraries
import pandas as pd
import seaborn as sns
import matplotlib.pyplot as plt

# Create a dataframe from the correlation data
data = {"Stay_In_Current_City_Years": [1, -0.012663171, -0.00418196, -
0.001780721, 0.002039186, 0.005469625],
    "Marital_Status": [-0.012663171, 1, 0.020545866, 0.015116324,
0.019452474, 0.000129018],
    "Product_Category_1": [-0.00418196, 0.020545866, 1, 0.5404235,
0.229490207, -0.314124735],
    "Product_Category_2": [-0.001780721, 0.015116324, 0.5404235,
1, 0.543543888, -0.209972881],
    "Product_Category_3": [0.002039186, 0.019452474, 0.229490207,
0.543543888, 1, -0.022256581],
    "Purchase": [0.005469625, 0.000129018, -0.314124735, -
0.209972881, -0.022256581, 1]}
df = pd.DataFrame(data=data)

# Create a heat map using seaborn
sns.set_theme()
plt.figure(figsize=(10,8))
sns.heatmap(df, annot=True)
plt.show()
```
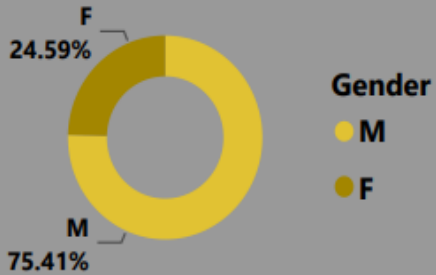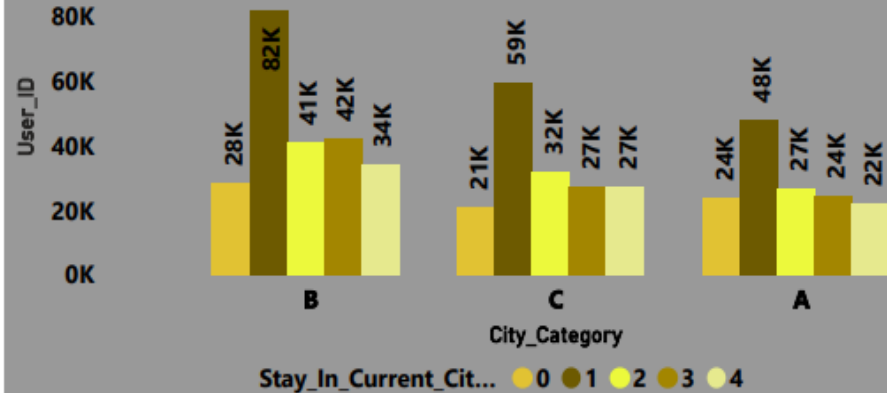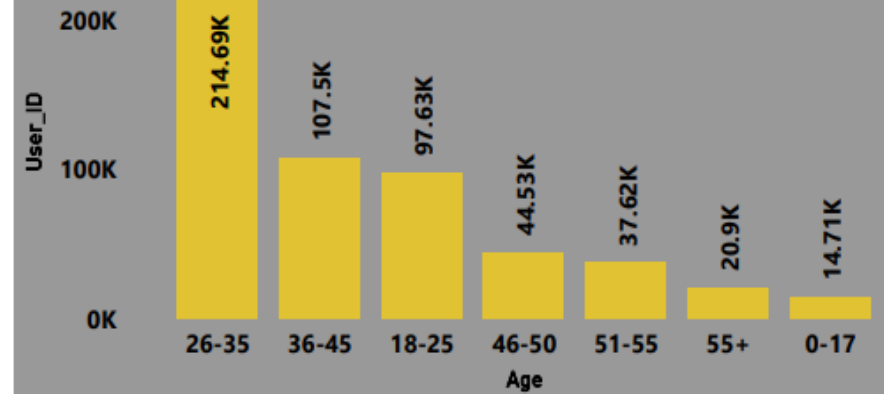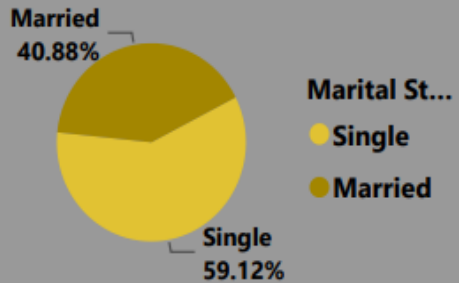
Buyers Demographic

Gender

F
24.59%

M
75.41%

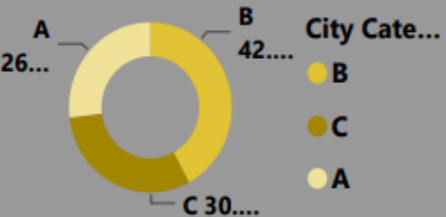**Gender**
- M
- F

Marital Status

Married
40.88%

Single
59.12%

Marital St...
- Single
- Married

City_Category Distribution

A
26...

B
42....

C 30....

City Cate...
- B
- C
- A

Year Spent In The City

City_Category	0	1	2	3	4
B	28K	82K	41K	42K	34K
C	21K	59K	32K	27K	27K
A	24K	48K	27K	24K	22K

Stay_In_Current_Cit... ● 0 ● 1 ● 2 ● 3 ● 4

Age Wise Distribution

Age	User_ID
26-35	214.69K
36-45	107.5K
18-25	97.63K
46-50	44.53K
51-55	37.62K
55+	20.9K
0-17	14.71K

User_ID by Occupation

Occupation	User_ID
4	71K
0	68K
7	58K
1	46K
17	39K
20	33K
12	30K
14	27K
2	26K
16	25K
6	20K
3	17K
10	13K
5	12K
15	12K
11	11K
19	8K
13	8K
18	7K
9	6K
8	2K
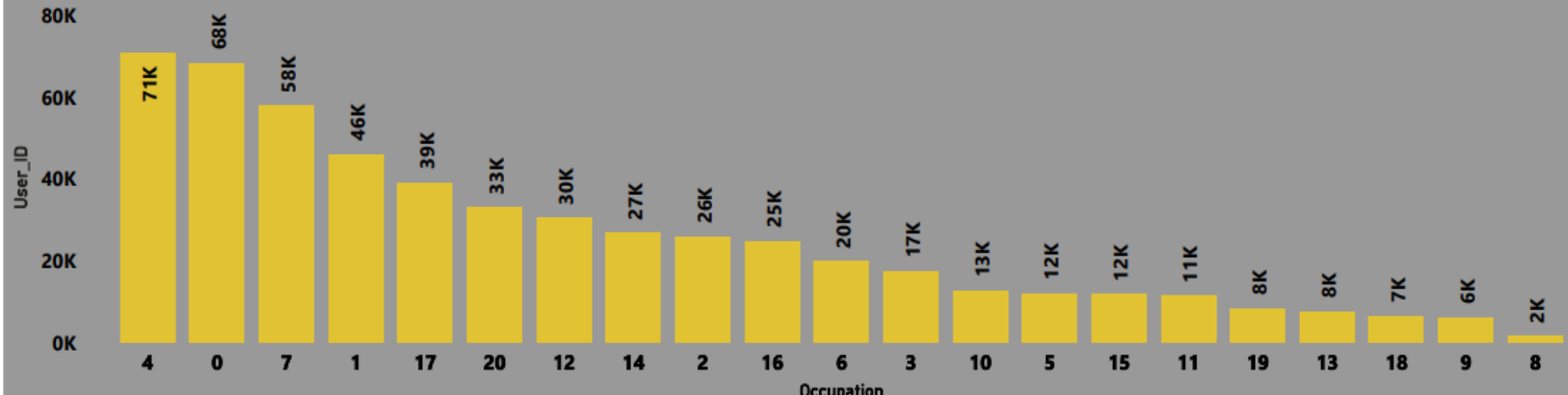
Spending Analysis

Total Purchase As Per Age And Marital Status

Marital Status ● Single ● Married

Age	Single	Married
0-17	133M	
18-25	715M	187M
26-35	1216M	784M
36-45	615M	395M
46-50		302M
51-55		259M
55+		
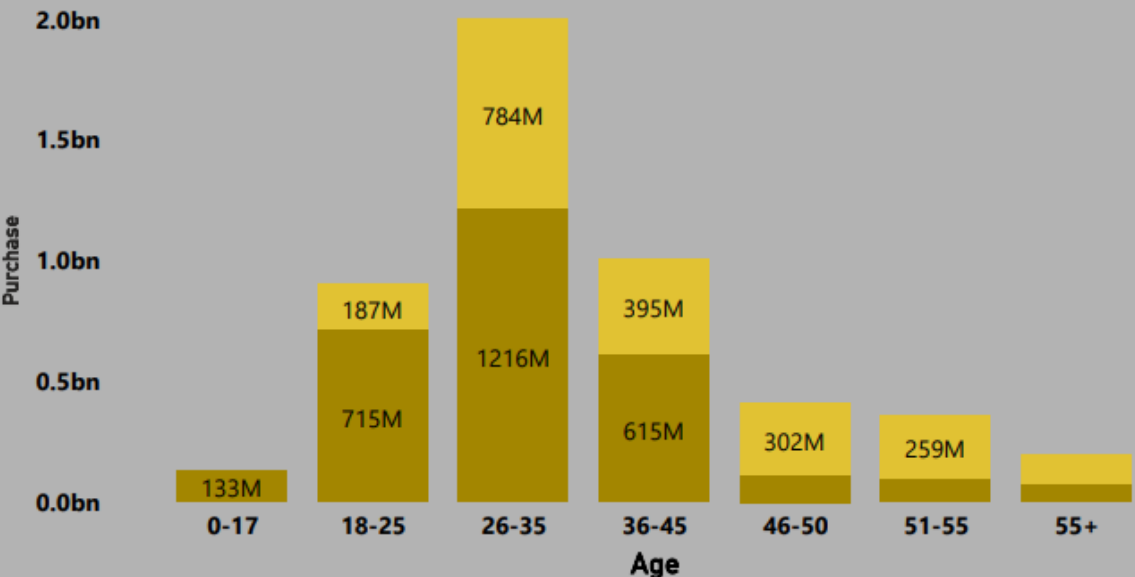
Purchase By Occupation And Gender

Gender ● F ● M

(Bar chart of Purchase by Occupation: 4, 0, 7, 1, 17, 12, 20, 14, 16, 2, 6, 3, 15, 10, 5, 11, 19, 13, 18, 9, 8)
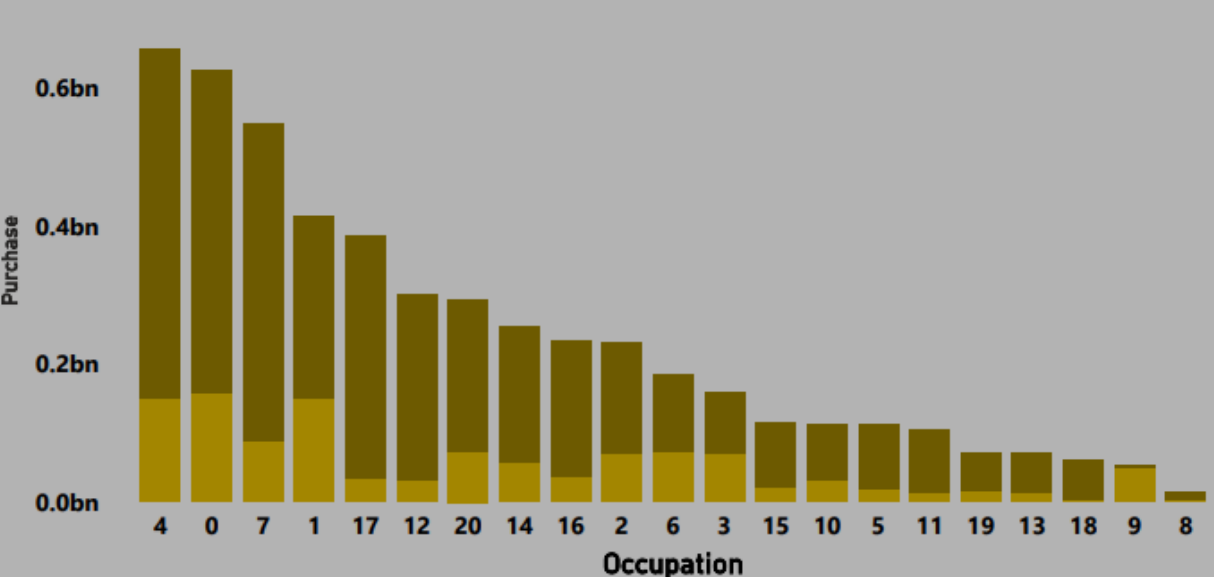
Total Purchase

5bn

City_Category
- A
- C
- B
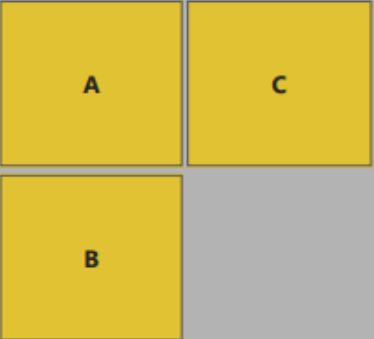
Gender
- F
- M

Stay_In_Current_City_Yea...
- 0
- 3
- 1
- 4
- 2

Custom_mari...
- Single
- Married

Conclusion

F

M

Age

☐ 0-17

☐ 18-25

☐ 26-35

Top 5 Product

Bar chart of Purchase vs Product_ID:
- P00025442: 28M
- P00110742: 26M
- P00255842: 25M
- P00184942: 24M
- P00059442: 24M

City Wise Sales

Pie chart:
- 1bn (25.82%)
- 2bn (41.52%) — B
- 2bn (32.66%)

Legend: B, C, A
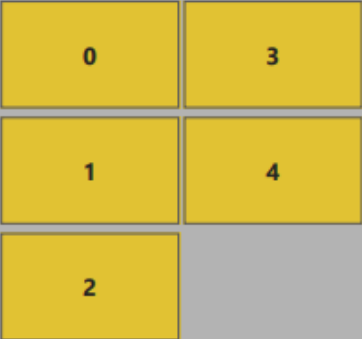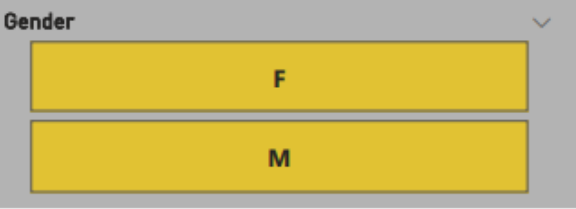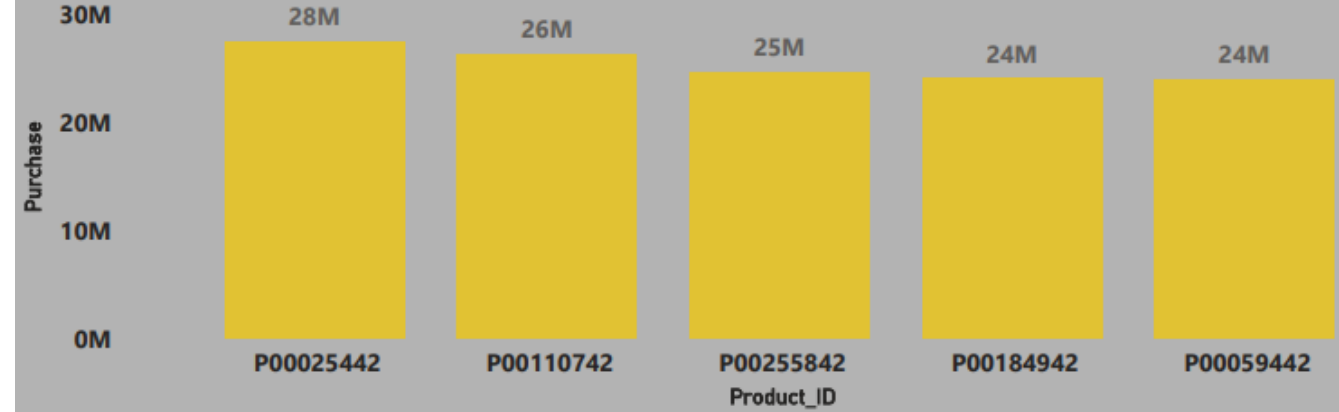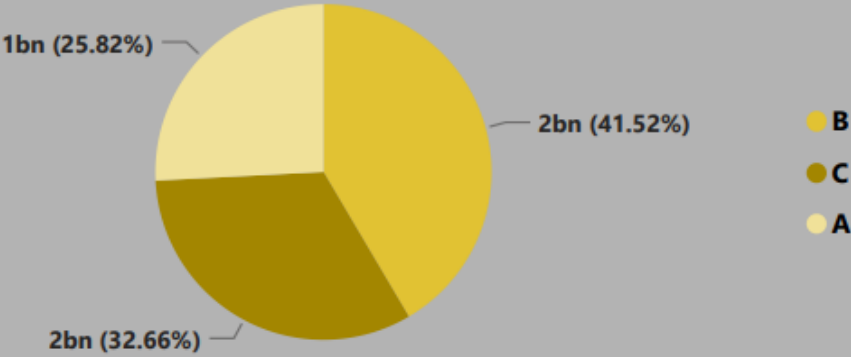
Purchase According to city Category And Gender

Bar chart of Purchase vs Gender:
- M: 3.9bn
- F: 1.2bn

Conclusion

1. Top Product = P00025442

2. Top City Purchase = B

3. Gender Wise Purchase = Male

Thank You