



# ONLINE STORE CUSTOMER DATA

By: Dao Phuoc Thinh

# AGENDA

01

OUR TEAM

02

DATA OVERVIEW

03

DATA PROCESSING

04

DATA INSIGHT

05

CONCLUSIONS

# Data Overview

1

## Dataset Name

Online Store Customer Data

(<https://www.kaggle.com/datasets/mountboy/online-store-customer-data/>)

2

## Dataset Detail

Dataset gồm 1 table (11 cột và 2513 dòng) chứa:

- Thông tin khách hàng
- Doanh thu theo đơn hàng

3

## Why do we choose this dataset?

Lĩnh vực thương mại điện tử/ mua bán trực tuyến ngày càng quan trọng và phát triển.

Phân tích bộ dữ liệu này sẽ giúp chúng tôi hiểu rõ hơn về các thách thức và cơ hội trong lĩnh vực này.

**Phân tích Sales Overview – Customer Demographic – Purchase Behaviour**

# Tools



## Clean Data & Check Value

Trong project này, nhóm sử dụng ngôn ngữ Python cùng một số thư viện để lọc và tiền xử lý dữ liệu thô.



## Report Visualization

Các bảng sau khi được làm sạch được nhập vào Power BI để tạo ra các biểu đồ nhằm rút ra các phân tích có ý nghĩa.



## Filter & Display Table Data

Xử lý dữ liệu để thể hiện dữ liệu theo dạng bảng



# Data Processing



# How to clean data?

Import Data lên  
Python

Xác định các cột/ dữ  
liệu cần xử lý

Xử lý dữ liệu



## 1

# Import Data lên Python

```
import pandas as pd  
df = pd.read_csv("Online store customer data.csv")  
print(df.head())
```

	Transaction_date	Transaction_ID	Gender	Age	Marital_status	State_names \
0	1/1/2019	151200	Female	19.0	Single	Kansas
1	1/1/2019	151201	Male	49.0	Single	Illinois
2	1/1/2019	151202	Male	63.0	Married	New Mexico
3	1/1/2019	151203	NaN	18.0	Single	Virginia
4	1/1/2019	151204	Male	27.0	Single	Connecticut

	Segment	Employees_status	Payment_method	Referral	Amount_spent
0	Basic	Unemployment	Other	1.0	2051.36
1	Basic	self-employed	Card	0.0	544.04
2	Basic	workers	PayPal	1.0	1572.60
3	Platinum	workers	Card	1.0	1199.79
4	Basic	self-employed	Card	0.0	NaN

# Data Definition

1. Transaction\_date - Ngày mua hàng
2. Transaction\_ID - ID đơn hàng
3. Gender - Giới tính
4. Age - Tuổi
5. Marital\_status - Trạng thái hôn nhân
6. State\_names - Địa lý
7. Segment - Hạng thành viên
8. Employees\_status - Trạng thái nghề nghiệp
9. Payment\_method - Hình thức thanh toán
10. Referral - Mua qua link giới thiệu
11. Amount\_spent - Doanh thu
12. year = `year(online_store_customer_data[Transaction_date])`: Năm mua hàng (xử lý tạo column mới trong PBI)
13. Month = `month(online_store_customer_data[Transaction_date])`: Tháng mua hàng (xử lý tạo column mới trong PBI)



## 2

## Xác định dữ liệu cần xử lý

```
[5] def check_values(df):  
    data = []  
    for column in df.columns:  
        data.append([  
            column, \  
            df[column].dtype, \  
            df[column].isna().sum(), \  
            round(100*(df[column].isna().sum()/len(df)),2), \  
            df[column].nunique()  
        ])  
    return pd.DataFrame(columns=['Data_Features','Data_Type','Null','Null_Percentage','Unique_Value'], data=data)
```

```
[6] check_values(df)
```

	Data_Features	Data_Type	Null	Null_Percentage	Unique_Value
0	Transaction_date	object	0	0.00	810
1	Transaction_ID	int64	0	0.00	2500
2	Gender	object	28	1.11	2
3	Age	float64	42	1.67	63
4	Marital_status	object	0	0.00	2
5	State_names	object	0	0.00	50
6	Segment	object	0	0.00	5
7	Employees_status	object	26	1.04	4
8	Payment_method	object	0	0.00	3
9	Referral	float64	155	6.17	2
10	Amount_spent	float64	242	9.63	2250

**Dataset có các cột sau có dữ liệu null cần được xử lý:**

- Gender
- Age
- Employees\_status
- Referral
- Amount\_spent

## Xử lý dữ liệu

```
[8] # Chuyển các giá trị null trong cột "Gender", "Employees_status", "Referral", "Amount_spent"
df['Gender'].fillna('Other', inplace=True)
df['Employees_status'].fillna('Undefined', inplace=True)
df['Referral'].fillna('0', inplace=True)
df['Amount_spent'].fillna('0', inplace=True)

# Chuyển các giá trị null trong cột "Age"
# Đầu tiên, tính độ tuổi trung bình của cột "Age" (bỏ qua các giá trị rỗng)
average_age = df['Age'].mean()
# Thay thế các giá trị rỗng trong cột "Age" bằng độ tuổi trung bình
df['Age'].fillna(average_age, inplace=True)

#Kiểm tra lại thông tin
check_values(df)
```

	Data_Features	Data_Type	Null	Null_Percentage	Unique_Value
0	Transaction_date	object	0	0.0	810
1	Transaction_ID	int64	0	0.0	2500
2	Gender	object	0	0.0	3
3	Age	float64	0	0.0	64
4	Marital_status	object	0	0.0	2
5	State_names	object	0	0.0	50
6	Segment	object	0	0.0	5
7	Employees_status	object	0	0.0	5
8	Payment_method	object	0	0.0	3
9	Referral	object	0	0.0	3
10	Amount_spent	object	0	0.0	2251

**Dataset có các cột sau có dữ liệu null cần được xử lý:**

- **Gender:** null -> Other
- **Age:** null -> AVG age (47)
- **Employees\_status:** null -> Undefined
- **Referral:** null -> 0
- **Amount\_spent:** null -> 0 (assumption: đơn hàng có promotion tặng sản phẩm (đơn hàng 0đ))

# Data Insight



# Directions for analysis

Sales Overview

Customer  
Demographic

Purchase  
behaviour



# Sales Overview

Compare Amount by month

year ● 2019 ● 2020 ● 2021

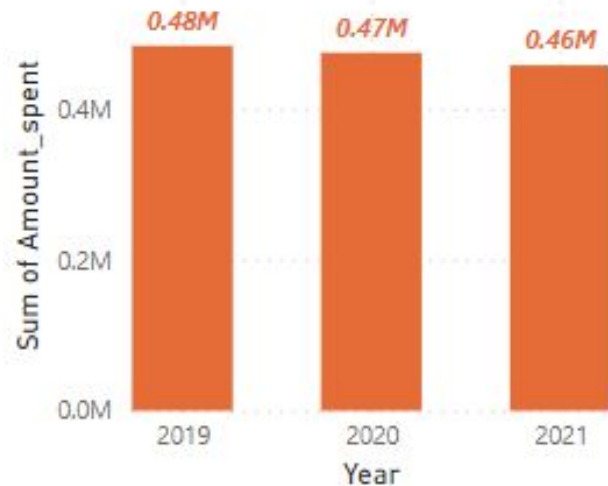


- Doanh số kinh doanh Online tại cửa hàng có xu hướng duy trì ổn định trong 3 năm, điều này là điểm đáng báo động trong tình hình kinh doanh hiện tại của công ty.
- Xét đến năm 2021 là năm hiện tại thì tháng 03 và 04/2023 tăng trưởng hơn 15% so với cùng kỳ năm ngoái
- Trung bình doanh thu theo tháng có xu hướng tăng nhẹ nhưng không đáng kể

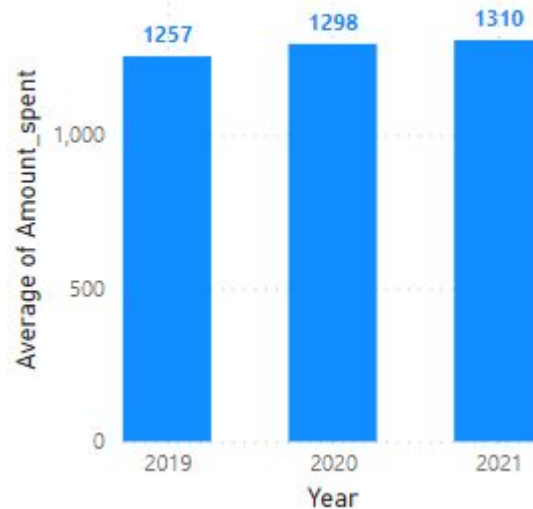
\*(Dữ liệu của Tháng 5/2021 không dùng để phân tích vì dữ liệu không đủ ngày)

# Sales Overview

Amount by year



Average order value



Lấy dữ liệu trong 4 tháng đầu năm thì doanh thu có xu hướng giảm tuy nhiên giá trị trung bình đơn tăng

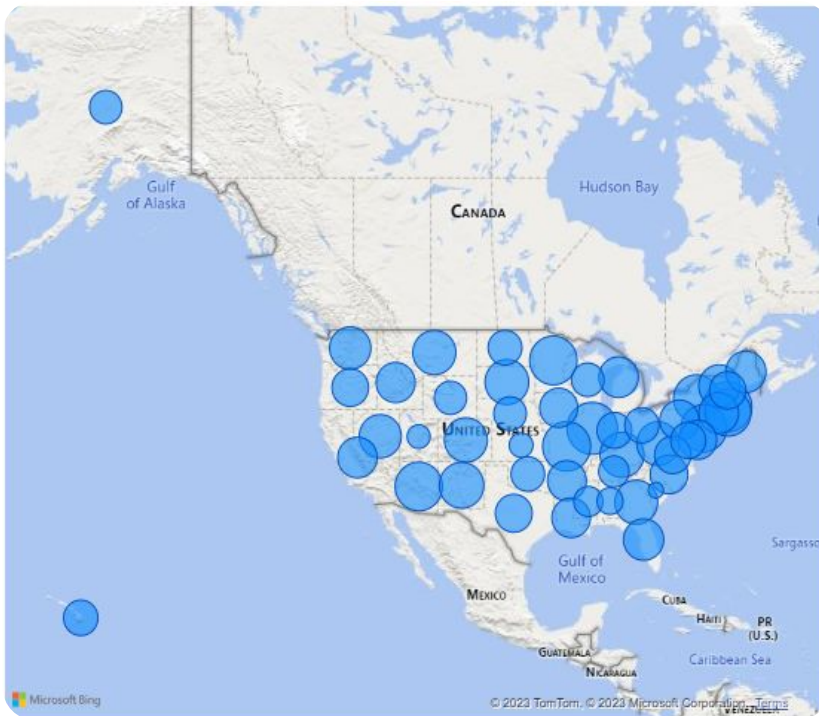
Qua các năm, người dùng ít mua sắm hơn nhưng chi tiêu nhiều tiền hơn cho 1 đơn hàng



## 1

# Sales Overview

**Phân bố doanh thu theo Khu vực:** Doanh số đến từ các khu vực không có sự chênh lệch quá lớn.



**TOP 5 Khu vực có doanh thu nổi bật:** Illinois, Arizona, Massachusetts, Missouri, Rhode Island

**TOP 2 Khu vực có trung bình Amount/Transaction lớn nhất là :** Idaho và Rhode Island

## Top 5 doanh thu cao nhất

	State_names ▾	Total_Amount_spent ▾
1	Illinois	8766065
2	Arizona	7722677
3	Massachusetts	7488362
4	Missouri	7404344
5	Rhode Island	7242385

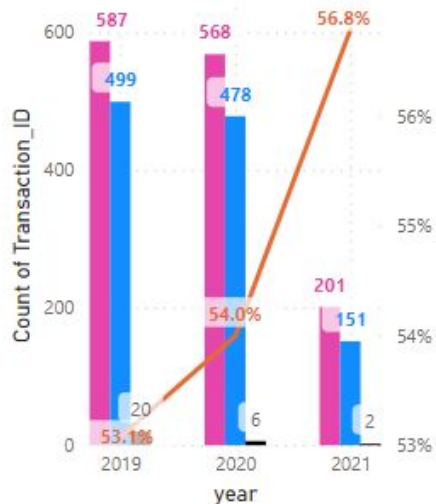
## Top 2 trung bình Amount spent/ Transaction

	State_names ▾	Avg_Amount_spent ▾
1	Idaho	155849,53658536586
2	Rhode Island	150883,02083333334

# Customer Demographic

Gender Ratio by Year

Gender Female Male Other FeMale\_r



Gender Ratio by year (amount)

Gender Female Male Other FeMale\_ratio



Người mua hàng là Nữ mua online nhiều hơn Nam chiếm tỷ trọng > 50% và có xu hướng tăng trong 2020 và 2021 cho cả số lượng đơn hàng và doanh thu. Cụ thể:

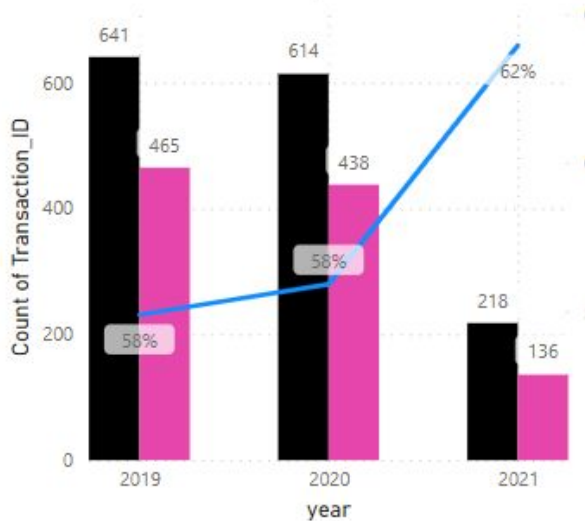
2019-2020 tăng từ 53.1% -> 54%  
2020-2021 tăng từ 54% -> 56.8%



## Customer Demographic

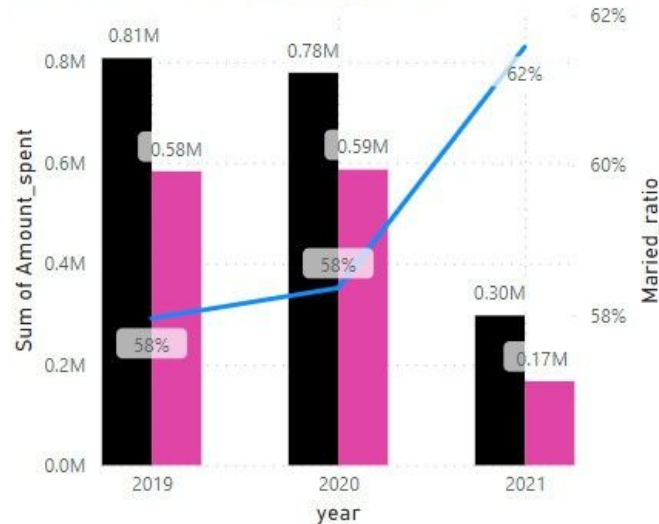
Marital Ratio by Year

Marital\_status ● Married ● Single ● Married\_ratio



Marital\_status by year (amount)

Marital\_status ● Married ● Single ● Married\_ratio

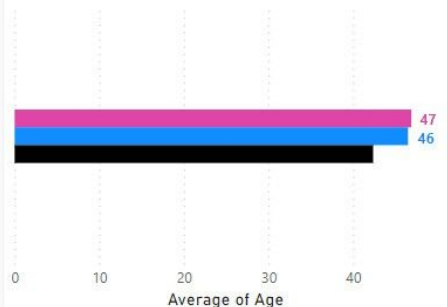


Đối tượng mua hàng qua kênh Online phần lớn là Married và có xu hướng tăng trưởng qua các năm. Cụ thể: 2020 qua 2021 tăng nổi bật từ 58% lên 62% trong khi 2019–2020 lại bằng nhau

## Customer Demographic

Average of Age by Gender

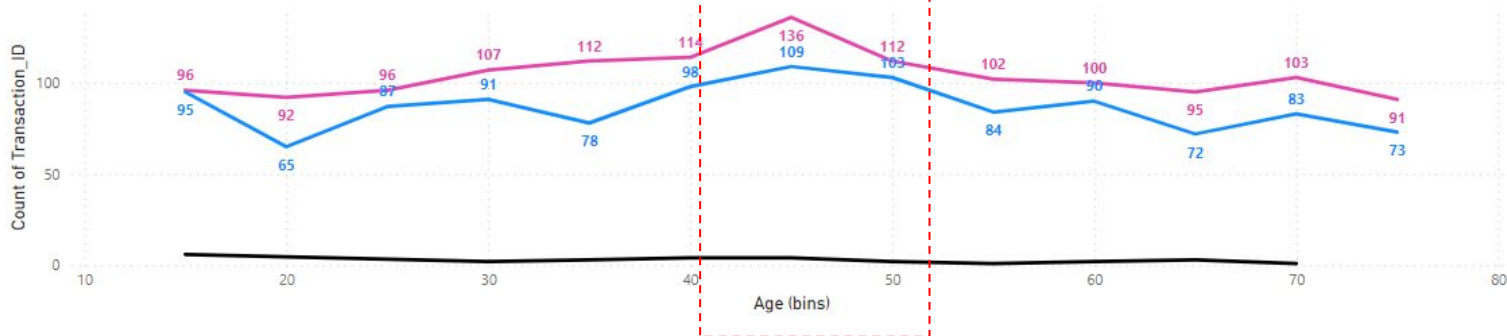
Gender ● Female ● Male ● Other



Độ tuổi trung bình của khách hàng Nam là 46 tuổi, nữ là 47 tuổi

Contribution of transaction by Age and Gender

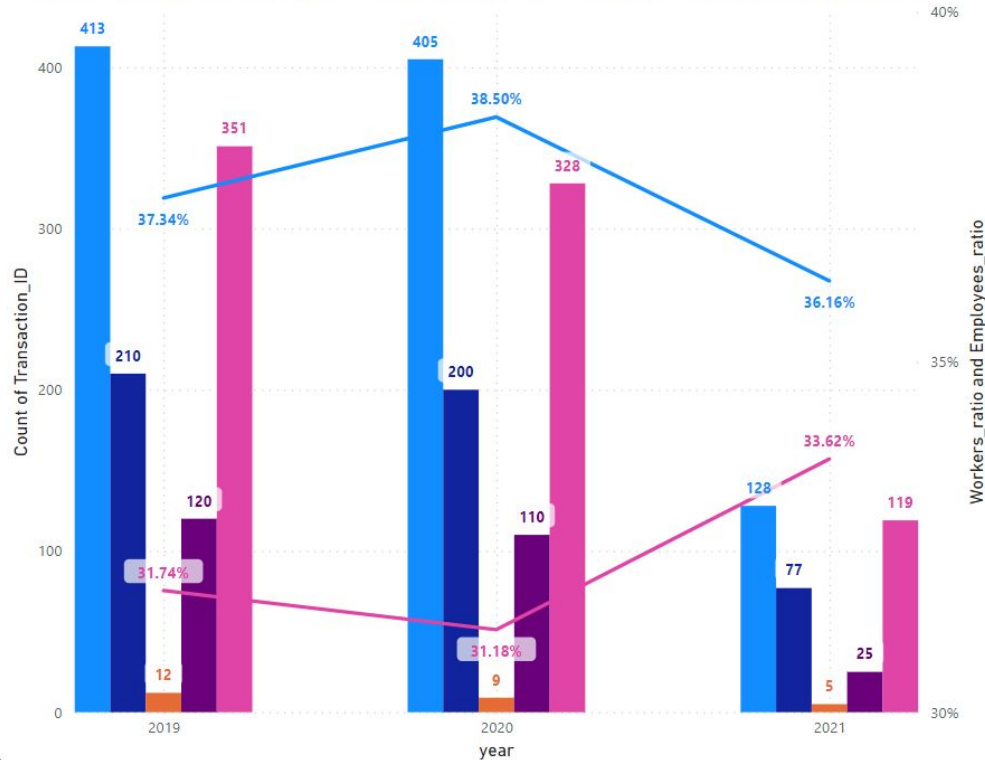
Gender ● Female ● Male ● Other



Phân khúc độ tuổi mua sắm Online trải rộng từ 15 - 78 tuổi. Tuy nhiên, Đối tượng có độ tuổi mua sắm nhiều nhất cho cả Nam và Nữ cho kênh Online khoản 40-50 tuổi.

contribution of transaction by Employees\_status

Employees\_status ● Employees ● self-employed ● Undefined ● Unemployment ● workers ● Workers\_ratio ● Employees\_ratio

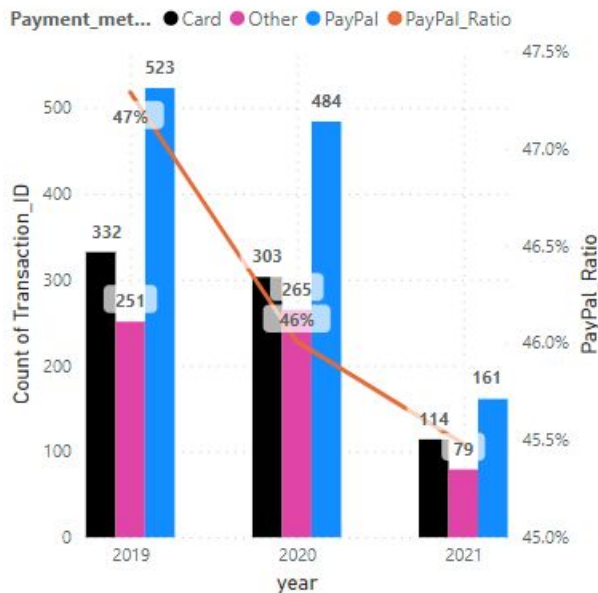


Kinh doanh Online người tiêu dùng phân lớn làm công công việc Employees or Workers. 2 ngành nghề này đang chiếm từ 68% -70%

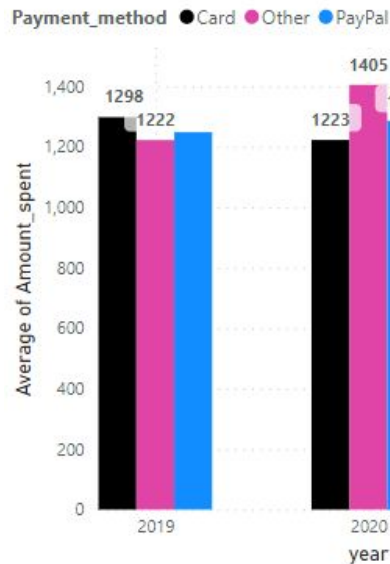
Khuyến nghị: Trong chương trình MKT online, Target vào 2 đối tượng này nhằm tiết kiệm chi phí MKT

# Customer Behaviour

## Trans by Payment method



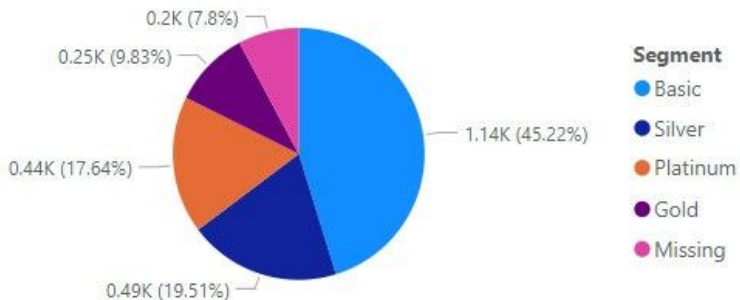
## AVG Amount by Payment method



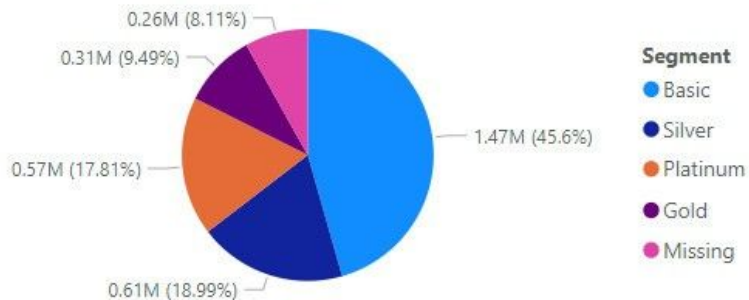
Đơn hàng được thanh toán qua Paypal chiếm số lượng đơn hàng cao nhất nhưng lại giảm qua các năm. Trong khi đó giá trị trung bình đơn hàng qua Paypal tăng trong giai đoạn 2019–2021.

# Customer Behaviour

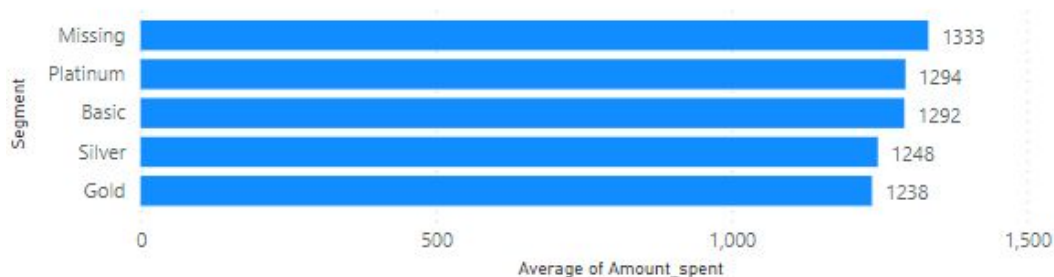
### Count of Transaction\_ID by Segment



### Sum of Amount\_spent by Segment



### Average of Amount\_spent by Segment



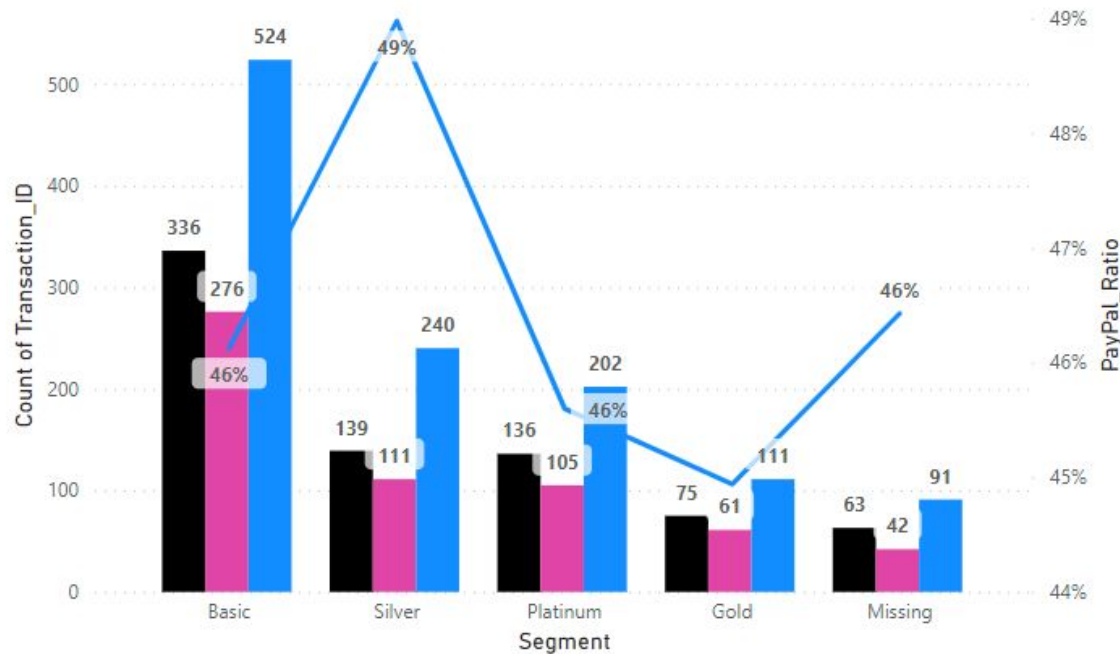
Segment Basic chiếm tỷ trọng lớn về số lượng đơn hàng và tổng doanh thu. Tuy nhiên giá trị trung bình đơn của nhóm này lại chỉ đứng thứ 3, trong khi đó nhóm missing lại chiếm tỷ trọng ít nhất nhưng có giá trị trung bình đơn hàng cao nhất.



## Customer Behaviour

Count of Transaction\_ID and PayPal\_Ratio by Segment and Payment\_method

Payment\_method ● Card ● Other ● PayPal ● PayPal\_Ratio



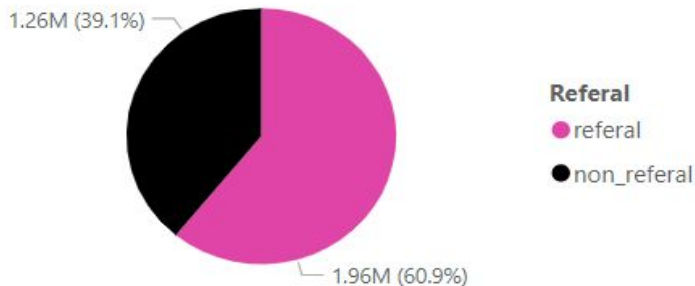
Hình thức thanh toán qua Paypal được sử dụng nhiều nhất trong tất cả các segment (chiếm tỷ trọng từ 46%–49%)

Đứng thứ 2 là thanh toán qua card và cuối cùng là các hình thức thanh toán khác

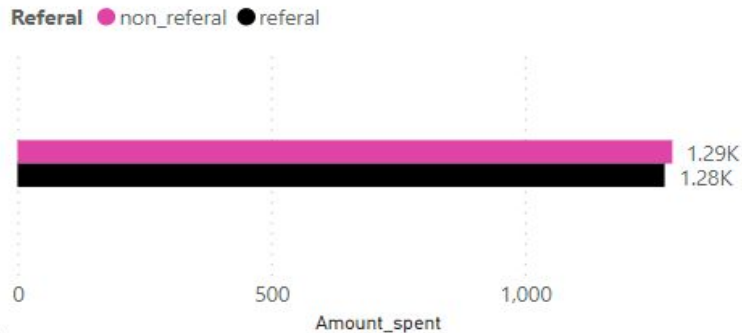


## Customer Behaviour

Amount\_spent by Referral



Amount\_spent by Referral

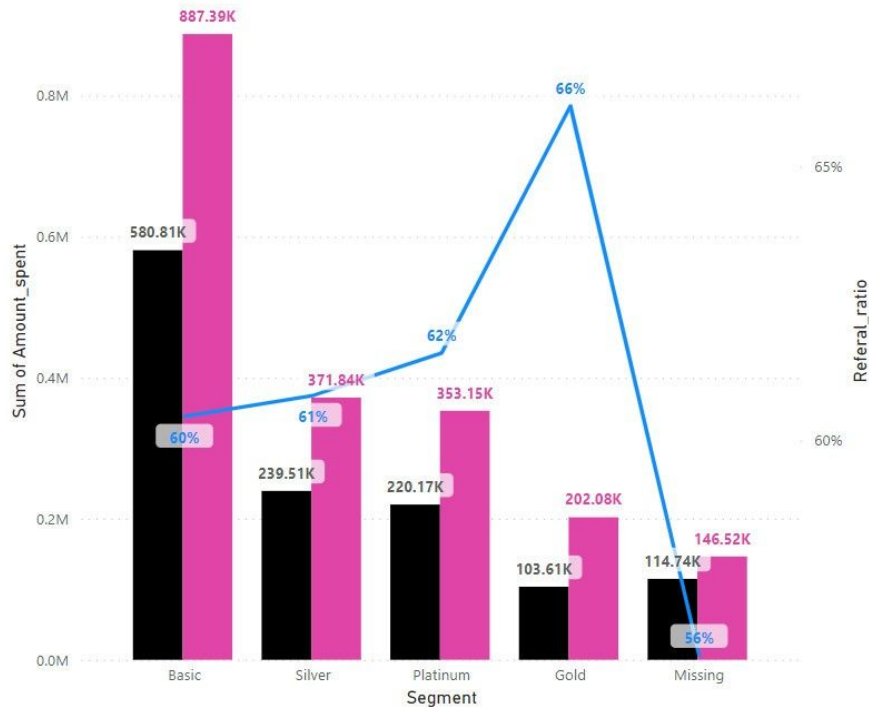


Doanh số đến từ kênh referral chiếm hơn 60% trong tổng doanh số nhưng giá trị trung bình đơn của kênh này thấp hơn (không đáng kể) so non-referral. Điều đó cho thấy bán hàng qua referral vẫn là một hướng tiềm năng cho việc kinh doanh Online

# Customer Behaviour

Segment by referral

Referral ● 0 ● 1 ● Referral\_ratio



Segment Basic có doanh số cao nhất đồng thời họ mua nhiều qua kênh referral nhiều nhất.

Với nhóm Gold doanh số chỉ đứng thứ 4, nhưng tỷ trọng mua hàng qua kênh referral của nhóm này chiếm tỷ trọng cao nhất (chiếm 66%)



# Summarize

Từ các dữ liệu trên chúng ta có thể đưa ra online target customer cho cửa hàng như sau

**Giới tính:** Nữ

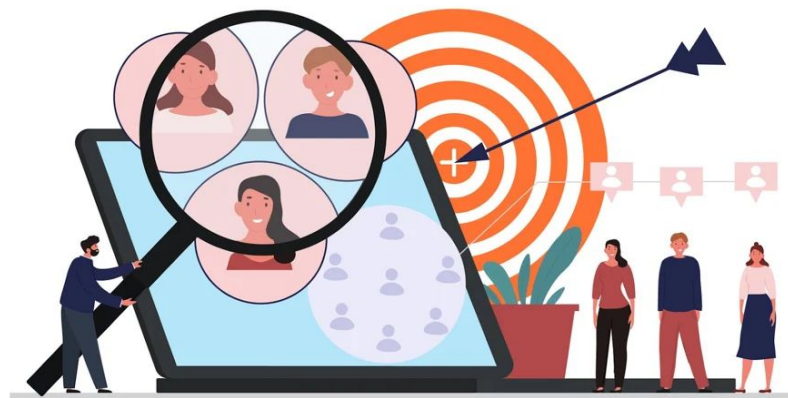
**Tuổi:** 40–50

**Trạng thái nghề nghiệp:** Employee, worker

**Trạng thái hôn nhân:** Đã kết hôn

**Địa lý:** Rhode Island và Montana

**Hành vi mua sắm:** Mua hàng qua giới thiệu (link referral) và thanh toán qua Paypal



# CONCLUSIONS



# OKRs suggested

## Key Objective: Tăng lợi nhuận ròng

Objective	Key Result	Tactic
Tăng doanh thu	Tăng số lượng đơn hàng	Tập trung quảng cáo và chăm sóc nhóm khách hàng mục tiêu
	Tăng giá trị đơn hàng	Triển khai các campaign tặng quà/ tăng chiết khấu cho kênh referral khi đơn hàng đạt giá trị X, tạo động lực mua hàng
Giảm chi phí	Giảm số chi phí logistics (assumption: hiện tại mỗi state đều có 1 kho trung chuyển)	Gộp kho trung chuyển cho các state gần nhau để giảm chi phí logistics. Chỉ giữ lại kho trung chuyển cho các state có doanh thu lớn
	Giảm chi phí marketing	Tập trung quảng cáo nhóm KH tiềm năng thay vì chạy QC mass
	Giảm tỷ lệ đơn hàng thất bại	Tiếp tục tập trung các hình thức thanh toán trả trước (Card/Paypal), giảm hình thức thanh toán bằng cash để giảm tỷ lệ bom hàng (giả định hình thức others là cash)

# Q&A



# THANKS

