Customer Shopping Behavior Analysis

1. Project Objective :

This project Analyses custom or shopping behavior using transactional data of purchases across the variety of category . The goal is to uncover the insights related to customer segmentation , product preferences , subscription behavior, spending patterns to boost business decision .

2. Dataset summary :

Rows : 3,900
Columns : 19
Data Structure :

- Customer demographics : customer_id,age,gender,location,subscription status
- Purchase details : item_purchased,category,purchase_amount,season,color,size
- Shopping details : review_rating ,shipping_type,discount_applied, previous_purchase,payment_method,frequency_of_purchase,age_group.

3. Exploratory Data Analysis :

- Data Loading : Imported the dataset using *Pandas.*
- Initial Exploration :
    - Used **df.head(10)** to get the preview of the data **.**
    - Used df.info() to study the structure & df.describe() for summary analysis
- Missing Data Handling :
  Checked the null values within each column and if found imputed with the best fitted measures of central tendency .
  So, found null values in review rating and replaced it with the median n value of each product category.
- Column Standarization :
  Rename column_name in snake_case for better readability.
- Feature Engineering :
  age_group : Created age_group column by binning customer age column.
- Data Consistency check :
  Discount_Applied & promo_code giving the same information , so being redundant removed the promo_code column.

4. Data Analysis Using SQL :

Performed structured analysis using MySQL

- **Total revenue gender wise :**

  Query :

```
select gender,sum(purchase_amount) total_revenue
from customer
group by gender;
```

gender	total_revenue
Male	157890
Female	75191

- **High Spending customers after using discounts**

  Query:

```
select customer_id
from customer
where discount_applied ="Yes" and purchase_amount>(select avg(purchase_amount)
from customer) ;
```

customer_id	purchase_amount
2	64
3	73
4	90
7	85
9	97
12	68

- **Top 5 products with highest average review rating**

  Query :

```
select item_purchased as product , round(avg(review_rating),2)
from customer
group by item_purchased
order by avg(review_rating) desc
limit 5;
```

product	round(avg(review_rating),2)
Sandals	3.82
Boots	3.79
Gloves	3.78
Hat	3.78
Handbag	3.78

3.78

- **The shipping courier which is used most ?**
  Query:
  ```
  select shipping_type , avg(purchase_amount)
  from customer
  where shipping_type in ('Express','Standard')
  group by shipping_type;
  ```

shipping_type	avg(purchase_amount)
Express	60.4752
Standard	58.4602

- **subscribers and non-subscribers :**
  Query:
  ```
  select subscription_status ,count(customer_id) total_customer ,avg(purchase_amount)
  average_spend ,sum(purchase_amount) total_revenue
  from customer
  group by subscription_status;
  ```

	subscription_status	total_customer	average_spend	total_revenue
▶	Yes	1053	59.4919	62645
	No	2847	59.8651	170436

- **Top products having the highest percentage of purchases with discounts applied.**
  Query :
  ```
  select item_purchased,sum(case when discount_applied='Yes' then 1 else 0 end
  )/count(*)*100 as discount_rate
  from customer
  group by item_purchased
  order by discount_rate desc
  limit 5;
  ```

	item_purchased	discount_rate
▶	Hat	50.0000
	Sneakers	49.655 50.0000
	Coat	49.0683
	Sweater	48.1707
	Pants	47.3684

- **Customer Segmentation :**
  Query :

```
select  t.customer_type ,count(t.customer_id)
from (select customer_id ,case when previous_purchases=1 then "new"
                               when previous_purchases between 2 and 10 then "returning"
          else "loyal"
          end as customer_type
      from customer ) as t
group by t.customer_type ;
```

customer_type	count(t.customer_id)
loyal	3116
returning	701
new	83

- **Top 3 most puchased products within each category**
  Query :

```
with view as (select category , item_purchased , row_number() over(partition by category
order by count(customer_id) desc) ranking
from customer
group by category ,item_purchased)

select category , item_purchased ,ranking
from view
where ranking <=3;
```

category	item_purchased	ranking
Accessories	Jewelry	1
Accessories	Sunglasses	2
Accessories	Belt	3
Clothing	Blouse	1
Clothing	Pants	2
Clothing	Shirt	3
Footwear	Sandals	1
Footwear	Shoes	2
Footwear	Sneakers	3
Outerwear	Jacket	1
Outerwear	Coat	2

- **Are Repeated buyers subscribers ?**
  Query :

```
select subscription_status,count(customer_id) as repeat_buyers
from customer
where previous_purchases >5
group by subscription_status;
```

subscription_status	repeat_buyers
Yes	958
No	2518

- **Revenue contribution of each age group**
  Query :

```
select age_grp ,sum(purchase_amount) as revenue
from customer
group by age_grp
order by revenue desc;
```

age_grp	revenue
Yound-adult	62143
Middle-aged	59197
adult	55978
Senior	55763

PowerBia Dashboard :

Business Recomendation :

- Increase Subscriber count by offering offer and benefit .
- Invest more in the most used Shipping type .
- Focus more on high revenue making age groups.
- Reward to repeated buyers so that they convert to loyal buyers.
- Use Top & best selling products for marketing campaign .