

PROJECT REPORT

Dataset: Customer Purchase Behaviour

SOFTWARE USED : -

- Database : My SQL Workbench
- Visualization: MS Power-BI
- Preprocessing & EDA: Jupyter Notebook

DATA DETAILS

This dataset containing customer purchase information for an online retail company. The dataset includes the following details:

- Transaction ID
- Customer ID
- Customer Name
- Product ID
- Product Name
- Product Category
- Purchase Quantity
- Purchase Price
- Purchase Date
- Country

1. Data Extraction and Transformation (SQL):

- **Setup a Database:** Create a database to store the customer purchase data.
 - create database myproject ;
- **Data Ingestion:** Write SQL scripts to import the provided purchase data into the database.

Steps :

- i. Format of data is csv
- ii. I used data ingestion manually by copying the script of the csv dataset using export method in sql.

Normalization of the data.

Data is normalized into 3 Tables :-

- i. Customers
- ii. Products
- iii. Purchase

Data Manipulation :

```
Create table customers1 as with cte as (select customername,  
country, purchasedate from raw_data)
```

```
Select row_number()over(order by purchasedate)+99 as  
customerid, customername, country from cte;
```

```
alter table raw_data add column customerid1 int;
```

- Update customer with the customerid1 values from customers1

```
update raw_data r
```

```
join customers1 c1 on r.customername = c1.customername and  
r.country = c1.country
```

```
set r.customerid1 = c1.customerid;
```

-- Replacing the Old customerid Values with new

```
alter table raw_data drop column customerid;
```

```
alter table raw_data change column customerid1 customerid int;
```

-- We can see Now customerid have unique data and we can delete our customer1 table which we used for mapping

```
drop table customers1;
```

Normalization of the data.

Data is normalized into 3 Tables :-

- i. Customers
- ii. Products
- iii. Purchase

```
create table purchases (transactionid int primary key,  
    customerid int,  
    productid int,  
    purchasequantity int,  
    purchaseprice double,  
    purchasedate date,  
    foreign key (customerid) references customers(customerid),  
    foreign key (productid) references products(productid)  
);  
  
insert into purchases (transactionid, customerid, productid,  
    purchasequantity, purchaseprice, purchasedate)  
  
select transactionid, customerid, productid, purchasequantity,  
    purchaseprice, purchasedate  
from raw_data;
```

Advanced queries to aggregate data

- *Finding total purchase by each customer*

-- Total spent per customer

```
select c.customerid, c.customername, count(*) as total_purchases,  
sum(p.purchasequantity * p.purchaseprice) as total_spent
```

```
from customers c
```

```
join purchases p on c.customerid = p.customerid
```

```
group by c.customerid, c.customername
```

```
order by total_spent ;
```

productid	productname	total_sales
861	Smartphone	4997.200000000001
302	Camera	4990.8499999999999
479	Headphones	4967.75
624	Microwave	4952.0999999999999
1155	Washing Machine	4950.65
327	Camera	4927.0999999999999
858	Smartphone	4890.65
1070	Toaster	4887.7999999999999
477	Headphones	4828.5
612	Microwave	4793.3

- *Total sales per product*

```
-- Total sales per product
```

```
select p.productid, p.productname, sum(pur.purchasequantity * pur.purchaseprice) as  
total_sales
```

```
from products p
```

```
join purchases pur on p.productid = pur.productid
```

```
group by p.productid, p.productname
```

```
order by total_sales desc;
```

productid	productname	total_sales
861	Smartphone	4997.2000000000001
302	Camera	4990.8499999999999
479	Headphones	4967.75
624	Microwave	4952.0999999999999
1155	Washing Machine	4950.65
327	Camera	4927.0999999999999
858	Smartphone	4890.65
1070	Toaster	4887.7999999999999
477	Headphones	4828.5
612	Microwave	4793.3

- *Number of customer per country*

select country, count(*) as num_of_customers

from customers

group by country

order by num_of_customers desc;

country	num_of_customers
Sudan	11
Palau	10
Romania	9
Botswana	9
Sri Lanka	9
Philippines	9
Eritrea	8
Oman	8
Luxembourg	8
Guyana	8



- *Top 5 product in each category*

select p.productname, sum(ps.purchasequantity) as total_sales

```

from products p
join purchases ps on p.productid = ps.productid
group by p.productname
order by total_sales desc
limit 5;

```

Result Grid |   Filter F

	productname	total_sa
	Toaster	191
	Heater	183
	Microwave	180
	Refrigerator	167
	Air Conditioner	165

- *Greatest purchase Quantity*

```

select customerid, productid, max(purchasequantity) as
greatest_purchase_quantity
from purchases
group by customerid, productid
order by greatest_purchase_quantity desc
limit 1;

```

customerid	productid	greatest_purchase_quantity
825	795	5

2. Data Analysis (Python):

Libraries used :-

```
import sqlalchemy
import pymysql
import pandas as pd
import matplotlib.pyplot as plt
import seaborn as sns
```

- **Data Extraction:**

```
engine =
sqlalchemy.create_engine('mysql+pymysql://root:12345@localhost:3306/
myproject')
```

```
#Loading Tables from 'myproject' database
```

```
customers = pd.read_sql_table("customers", engine)
```

```
products = pd.read_sql_table("products", engine)
```

```
purchases = pd.read_sql_table("purchases", engine)
```

- **Data Analysis:** Perform the following analysis using basic Python:
 - *Calculate total purchases, total revenue, and average purchase value.*

```
total_purchases total_purchases = combined_df['purchasequantity'].sum()
```

```
print(f"Total Purchase: {total_purchases}")
```

```
print(total_purchases)
```

```
total_revenue = total_revenue = (combined_df['purchasequantity'] *
combined_df['purchaseprice']).sum()
```

```
print(f"Total Revenue: {total_revenue}") :
```

```
Avg_purchase = #average_purchahse_value
```

```
Avg_purchase = round(combined_df['purchaseprice'].mean(),2)
```

```
print(f"Average Purchase Value: {Avg_purchase}")
```

Calculate total purchases, total revenue, and average purchase value.

```
#total_purchases
total_purchases = combined_df['purchasequantity'].sum()
print(f"Total Purchase: {total_purchases}")
```

Total Purchase: 3053

```
#total_revenue
total_revenue = (combined_df['purchasequantity'] * combined_df['purchaseprice']).sum()
print(f"Total Revenue: {total_revenue}")
```

Total Revenue: 1485760.5499999998

```
#average_purchse_value
Avg_purchase = round(combined_df['purchaseprice'].mean(),2)
print(f"Average Purchase Value: {Avg_purchase}")
```

Average Purchase Value: 489.27

- **Identify top customers and their purchasing behavior.**

```
top_customers = purchases.groupby('customerid').agg(
    total_purchases=('purchasequantity', 'sum'),
    total_spent=('purchaseprice', 'sum')
).nlargest(10, 'total_spent')
plt.savefig('top_customer.png')
print(top_customers)
```

customerid	total_purchases	total_spent
877	2	999.98
367	5	999.44
381	5	998.17
173	3	997.45
603	4	996.01
971	3	995.97
559	4	994.58
1026	5	993.55
915	2	992.33
358	4	991.08

<Figure size 640x480 with 0 Axes>

- **Analyze purchase trends over time (monthly, quarterly, yearly).**

MONTHLY TRENDS

```
: #Monthly
monthly_trends = combined_df.groupby(['year', 'month']).agg({
    'purchasequantity': 'sum',
    'purchaseprice': 'sum'
}).reset_index()
monthly_trends['average_purchase'] = monthly_trends['purchaseprice'] / monthly_trends['purchasequantity']
print(monthly_trends)
```

	year	month	purchasequantity	purchaseprice	average_purchase
0	2023	6	28	4012.44	143.301429
1	2023	7	269	47956.69	178.277658
2	2023	8	236	33493.46	141.921441
3	2023	9	287	46764.80	162.943554
4	2023	10	243	38368.97	157.896996
5	2023	11	293	39839.29	135.970273
6	2023	12	295	51910.42	175.967525
7	2024	1	292	43575.14	149.229932
8	2024	2	272	41393.77	152.182978
9	2024	3	208	37442.66	180.012788
10	2024	4	195	34165.64	175.208410
11	2024	5	268	42799.97	159.701381
12	2024	6	167	27551.47	164.978862



QUATERLY TREND

MONTH

```
] : #Quarterly
quarterly_trends = combined_df.groupby('quarter').agg({
    'purchasequantity': 'sum',
    'purchaseprice': 'sum'
}).reset_index()
quarterly_trends['average_purchase'] = quarterly_trends['purchaseprice'] / quarterly_trends['purchasequantity']
print(quarterly_trends)
```

	quarter	purchasequantity	purchaseprice	average_purchase
0	2023Q2	28	4012.44	143.301429
1	2023Q3	792	128214.95	161.887563
2	2023Q4	831	130118.68	156.580842
3	2024Q1	772	122411.57	158.564210
4	2024Q2	630	104517.08	165.900127

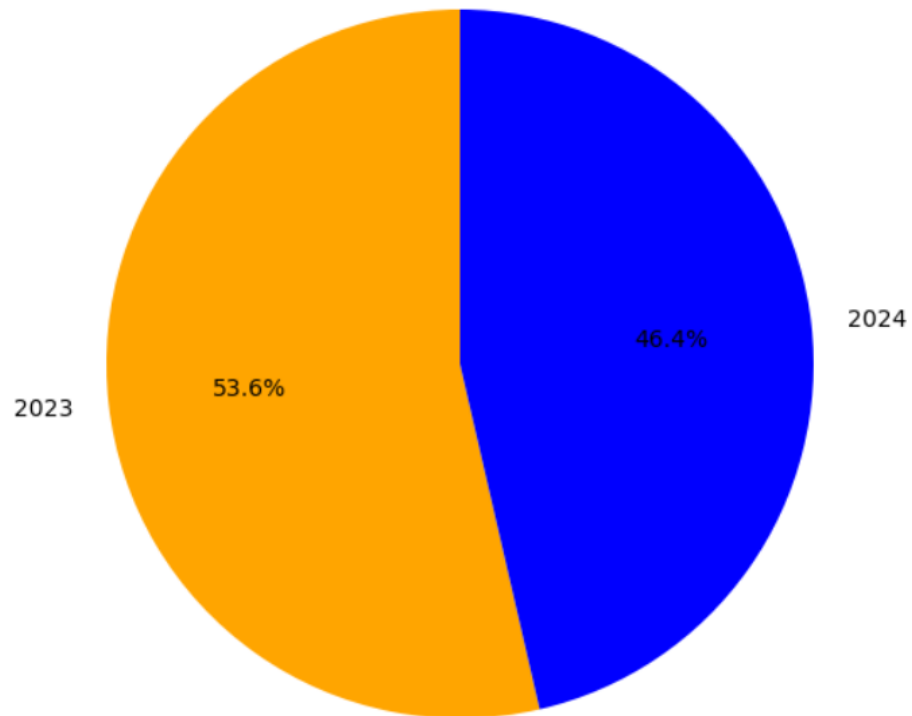


YEARLY TREND

```
yearly_trends = combined_df.groupby(combined_df['purchasedate'].dt.year)[['purchaseprice', 'purchasequantity']].sum().reset_index()
print(yearly_trends)
```

	purchasedate	purchaseprice	purchasequantity
0	2023	262346.07	1651
1	2024	226928.65	1402

Distribution of Purchase price by Year



- ***Identify the top-performing product categories.***

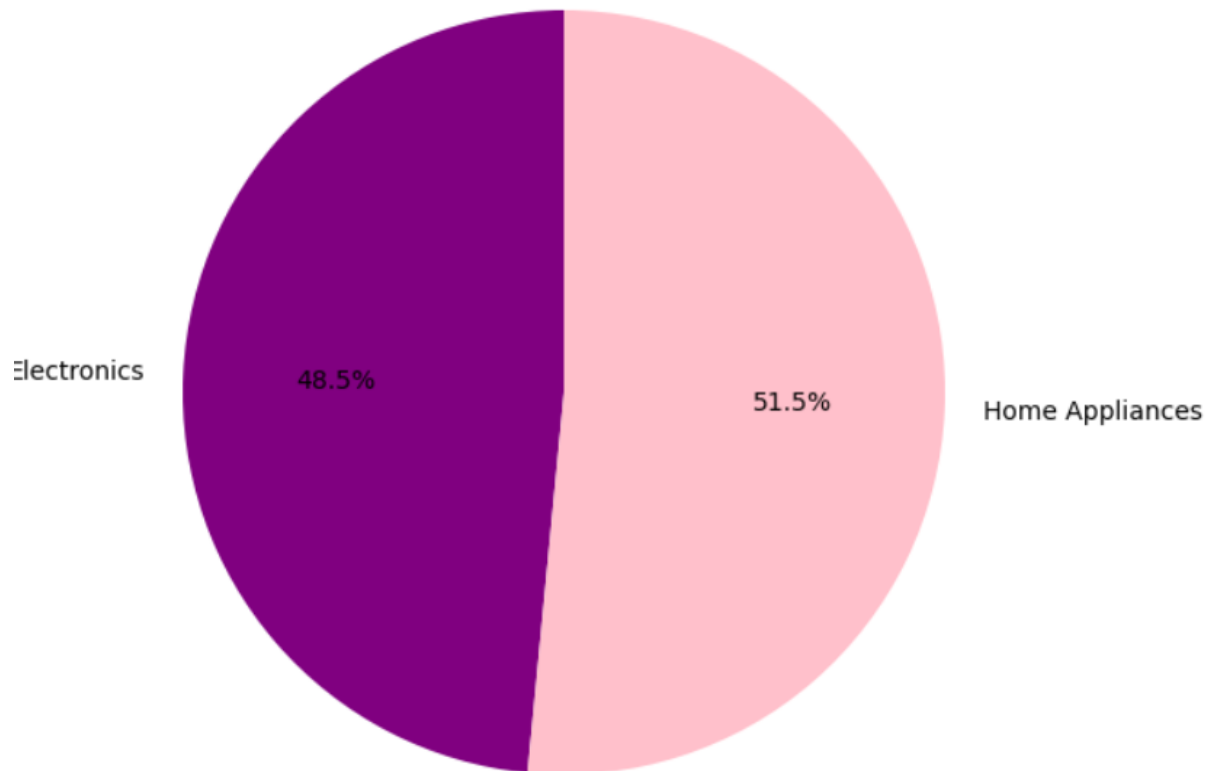
▼ Identify the top-performing product categories. ¶

```
] : top_product_categories = combined_df.groupby('productcategory').agg({
    'purchasequantity': 'sum',
    'purchaseprice': 'sum'
}).reset_index()

top_product_categories['average_purchase'] = top_product_categories['purchaseprice'] / top_product_categories['purchasequantity']
print(top_product_categories)
```

	productcategory	purchasequantity	purchaseprice	average_purchase
0	Electronics	1480	248194.56	167.699027
1	Home Appliances	1573	241080.16	153.261386

Top Product Categories: Purchase Quantity Distribution

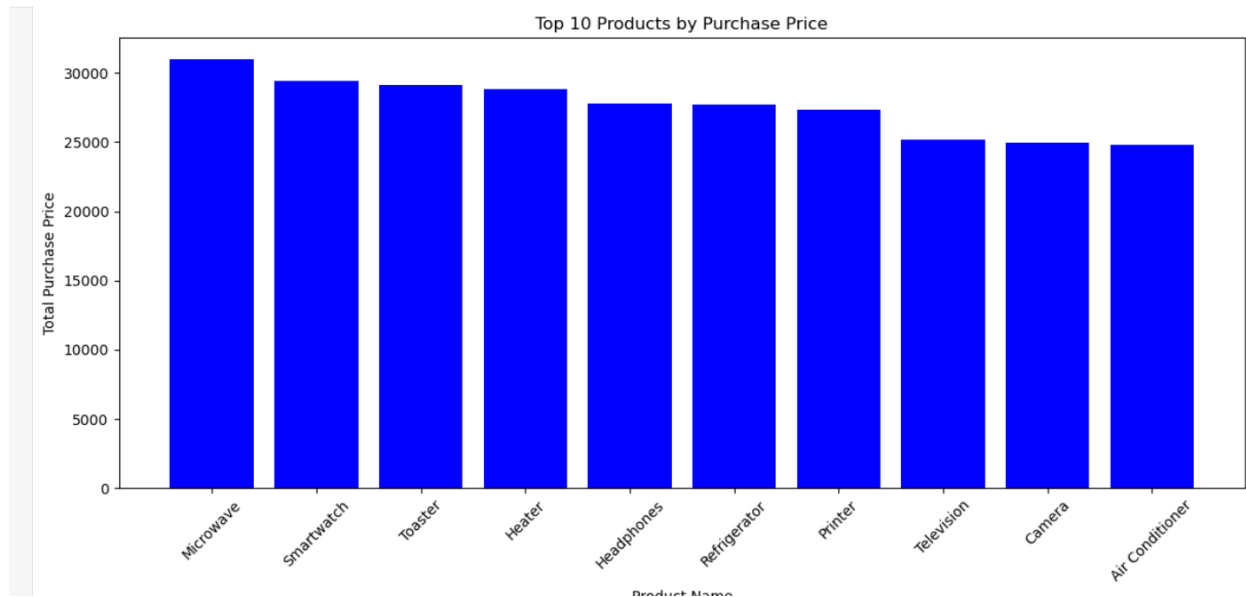


- **Identify the top-performing product categories.**

```
|: top_product = combined_df.groupby('productname')['purchaseprice'].sum().reset_index()
top_10 = top_product.sort_values(by='purchaseprice', ascending=False).head(10)
print("Top 10 product by purchase price:")
print(top_10)
```

Top 10 product by purchase price:

	productname	purchaseprice
8	Microwave	30970.75
14	Smartwatch	29407.68
17	Toaster	29116.03
6	Heater	28828.67
5	Headphones	27748.78
10	Refrigerator	27730.95
9	Printer	27360.29
16	Television	25188.60
2	Camera	24985.70
0	Air Conditioner	24812.05



Key Insight Report;

1. Total Purchases: 3053
2. Total Revenue: 1,485,760.54
3. Average Purchase Value: 489.29

Purchase_trend over time:

Monthly Trend in 2023:

December-2023 saw the highest purchase quantity (295 units) , purchase price of 51910.42, with an average purchase value of 175.96.

and lowest is the jun-2023 with the purchase quantity(28 unit) , purchase price of 4012.44 with an avg purchase value of 143.30

Purchases showed fluctuations across other months, indicating seasonal variations or promotional effects.

Quarterly Trend:

Q2 2023 had the lowest purchase quantity but a moderate average purchase value.

Q3 and Q4 2023 showed higher quantities and revenues

Q1 2024 maintained a steady purchase quantity and average purchase value.

Yearly Trend:

2023 saw the highest overall purchase quantity (1,651 units) and revenue (\$262,346.07), with a consistent average purchase value.

2024 started slightly lower but maintained a strong average purchase value.

Top performing product Performance:

Electronics: Customers made 1,480 purchases, generating 248,194.56 in revenue. On average, each purchase in this category was about 167.70.

Home Appliances: This category saw 1573 purchases, totaling 241,080.16 in revenue. The average purchase here was around 153.26.

Thank You.