Insights Report

Zeotap

Ishaan S

27th January, 2025

Introduction

1. Basic Descriptive Insights
2. Regions with the most sales
3. Regions with the most customers
4. Most popularly sold products
5. Customers who have spent the most
6. Input based insights on a specific product
7. Graphs

Basic Descriptive Insights

This is the code used to obtain the basic insights from each of these

```
customers = pd.read_csv("Customers.csv")
products = pd.read_csv("Products.csv")
transactions = pd.read_csv("Transactions.csv")
print("Customers:-\n", customers.describe())
print("Products:-\n", products.describe(include='all'))
print("Transactions:-\n", transactions.describe(include='all'))
```

The output is as shown below.

```
Customers:-
       CustomerID      CustomerName         Region  SignupDate
count         200               200            200         200
unique        200               200              4         179
top         C0001  Lawrence Carroll  South America  2024-11-11
freq            1                 1             59           3
Products:-
       ProductID            ProductName Category       Price
count        100                    100      100  100.000000
unique       100                     66        4         NaN
top         P001  ActiveWear Smartwatch    Books         NaN
freq           1                      4       26         NaN
mean         NaN                    NaN      NaN  267.551700
std          NaN                    NaN      NaN  143.219383
min          NaN                    NaN      NaN   16.080000
25%          NaN                    NaN      NaN  147.767500
50%          NaN                    NaN      NaN  292.875000
75%          NaN                    NaN      NaN  397.090000
max          NaN                    NaN      NaN  497.760000
Transactions:-
       TransactionID CustomerID ProductID      TransactionDate     Quantity   TotalValue       Price
count           1000       1000      1000                 1000  1000.000000  1000.000000  1000.00000
unique          1000        199       100                 1000          NaN          NaN         NaN
top           T00001      C0109      P059  2024-08-25 12:38:23          NaN          NaN         NaN
freq               1         11        19                    1          NaN          NaN         NaN
mean             NaN        NaN       NaN                  NaN     2.537000   689.995560   272.55407
std              NaN        NaN       NaN                  NaN     1.117981   493.144478   140.73639
min              NaN        NaN       NaN                  NaN     1.000000    16.080000    16.08000
25%              NaN        NaN       NaN                  NaN     2.000000   295.295000   147.95000
50%              NaN        NaN       NaN                  NaN     3.000000   588.880000   299.93000
75%              NaN        NaN       NaN                  NaN     4.000000  1011.660000   404.40000
max              NaN        NaN       NaN                  NaN     4.000000  1991.040000   497.76000
```

Regions with the most sales

This insight is obtained through merging the three datasets based on each of their respective IDs.

```
merged_data = transactions.merge(customers,
on="CustomerID").merge(products, on="ProductID")
```

The code specific to obtaining this data is as follows.

```
most_selling_regions=merged_data['Region'].value_counts()
```

The output of the code is shared below

```
The regions with the most sales :-
 Region
South America    304
North America    244
Europe           234
Asia             218
```

This insight can help a company devote more product stock or production in itself towards the regions that already have an established and thriving market for the specific product, and hence will lead to more sales. Whereas in terms of the regions that do not have as many sales can have more capital towards marketing and sales.

Regions with the most customers

The code specific to obtaining this data is as follows.

```
most_customers_region=customers['Region'].value_counts()
```

The output of the code is shared below

```
The regions with the most customers:-
 Region
South America    59
Europe           50
North America    46
Asia             45
```

When compared to the previous insight it becomes evident that in this specific example, the European customers are consuming less of the products when compared to the North American customers. This indicates that the North American customers are more loyal to the brand. This further indicates that they do not find any better deals for the products at any other store. Whereas in Europe, the case is not the same. Thus, a good course of action to improve sales in the European region can be to implement market price optimisation.

Most popularly sold products

This insight is obtained through merging the three datasets based on each of their respective IDs.

```
merged_data = transactions.merge(customers,
on="CustomerID").merge(products, on="ProductID")
```

The code specific to obtaining this data is as follows.

```
most_selling_products=merged_data.groupby('ProductName')['TotalValue'].sum
().sort_values(ascending=False)
```

The output of the code is shared below

```
The most popularly sold products :-
 ProductName
ActiveWear Smartwatch    39096.97
SoundWave Headphones     25211.64
SoundWave Novel          24507.90
ActiveWear Jacket        22712.56
ActiveWear Rug           22314.43
```

This is an insight that describes which specific products are being sold at a higher volume than the rest. This is an important insight as it can help the company dedicate a larger stock of these items to these specific regions and also help vastly in inventory management.

Customers who have spent the most

This insight is obtained through merging the three datasets based on each of their respective IDs.

```
merged_data = transactions.merge(customers,
on="CustomerID").merge(products, on="ProductID")
```

The code specific to obtaining this data is as follows.

```
top_customers_spending =
merged_data.groupby('CustomerID')['TotalValue'].sum().sort_values(ascendin
g=False).head(10)
```

The output of the code is shared below

```
The customers that have spent the most:-
 CustomerID
C0141    10673.87
C0054     8040.39
C0065     7663.70
C0156     7634.45
C0082     7572.91
C0188     7111.32
C0059     7073.28
C0028     6819.57
C0099     6715.72
C0165     6708.10
```

This is a list of this company's most loyal customers. These customers consistently purchase items from this store and hence are satisfied with the store's service. This means these customers can be recommended items based on their previous catalogue of various purchases.

Input based insights on a specific product

This insight is obtained through merging the three datasets based on each of their respective IDs.

```
merged_data = transactions.merge(customers,
on="CustomerID").merge(products, on="ProductID")
```

The code specific to obtaining this data is as follows.

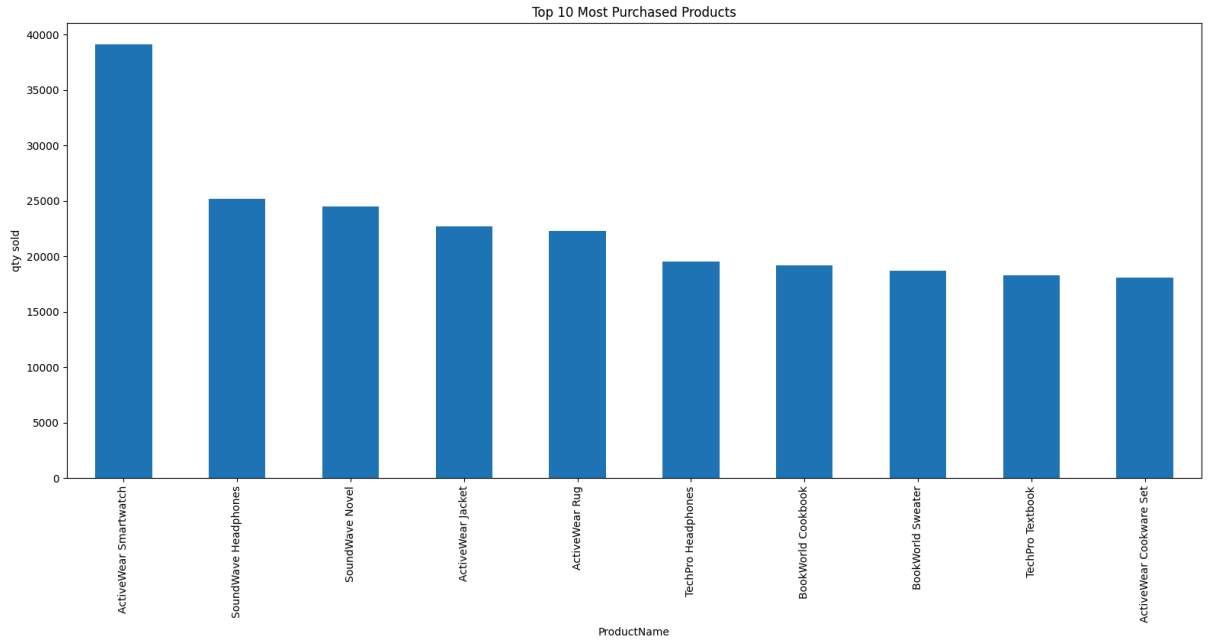
```
product_of_interest = input("Enter your desired product : ")

product_data = merged_data[merged_data['ProductName'] ==
product_of_interest]

purchase_counts =
product_data.groupby(['Region'])['ProductName'].count().reset_index(name='
PurchaseCount')
```

The output of the code is shared below

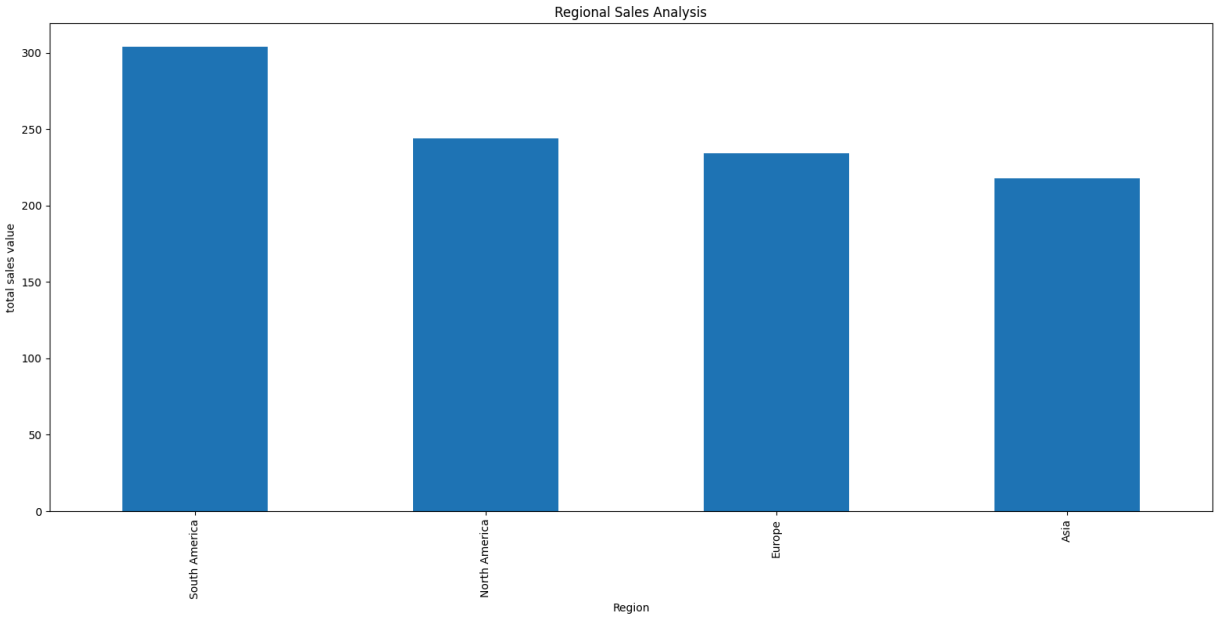
```
Enter your desired product : ActiveWear Biography
Purchase counts for 'ActiveWear Biography' by region and customer:
           Region  PurchaseCount
0            Asia              1
1          Europe              3
2   North America              2
3   South America              2
```

This is a more comprehensive list of which regions have a higher demand for a specific product when compared to the other. This can assist the inventory management with that product by dedicating the necessary amount to each region based on this insight.

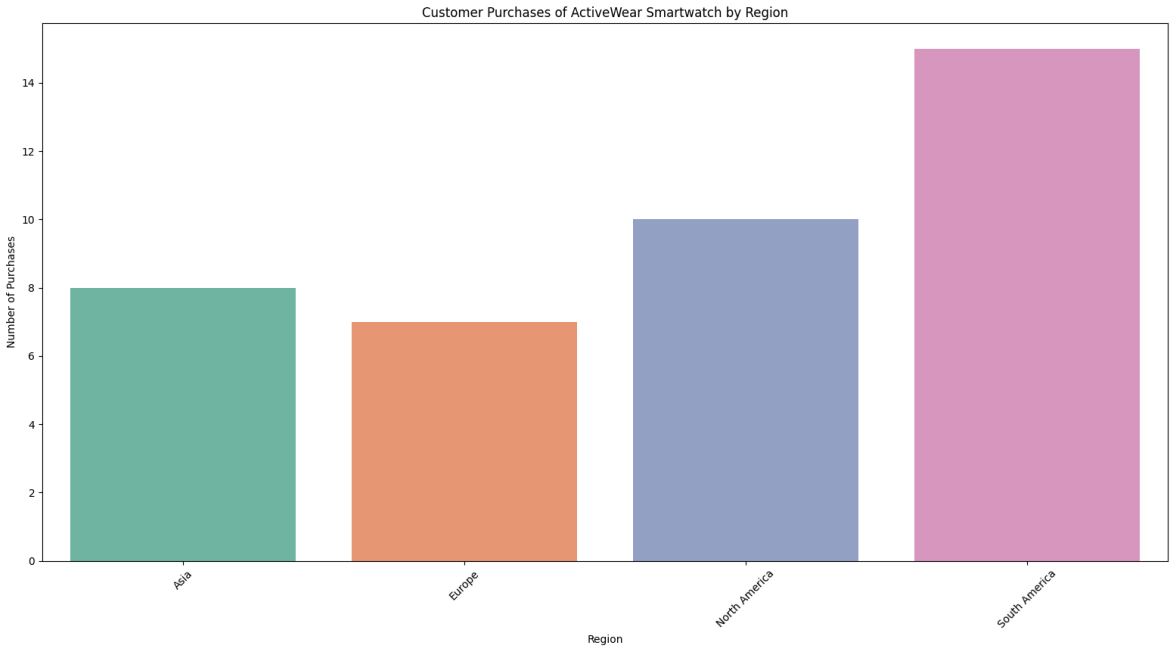
Graphs



Top 10 Most Purchased Products



Regional Sales Analysis



Customer Purchases of a Specific Product

Conclusion

These insights obtained through the datasets can be used to positively affect sales in the company.