27/01/2025, 22:09 Exploratory Data Analysis (EDA).ipynb - Colab

import pandas as pd

```
customers = pd.read_csv("Customers.csv")
products = pd.read_csv("Products.csv")
transactions = pd.read_csv("Transactions.csv")
print(customers.head())
print(products.head())
print(transactions.head())
print(customers.isnull().sum())
print(products.isnull().sum())
print(transactions.isnull().sum())
print(customers.info())
print(products.info())
print(transactions.info())
→ 4
          902.04 300.68
    CustomerID
    CustomerName
    Region
    SignupDate
    dtype: int64
    ProductID
    ProductName
                  0
    Category
    Price
    dtype: int64
    TransactionID
    CustomerID
    ProductID
    TransactionDate
    Quantity
    TotalValue
                      0
    Price
    dtype: int64
    <class 'pandas.core.frame.DataFrame'>
    RangeIndex: 200 entries, 0 to 199
    Data columns (total 4 columns):
     # Column
                     Non-Null Count Dtype
     0 CustomerID 200 non-null object
        CustomerName 200 non-null
                                     object
     2 Region
                      200 non-null
                                     object
     3 SignupDate 200 non-null
                                     object
    dtypes: object(4)
    memory usage: 6.4+ KB
    None
    <class 'pandas.core.frame.DataFrame'>
    RangeIndex: 100 entries, 0 to 99
    Data columns (total 4 columns):
                     Non-Null Count Dtype
     # Column
     0 ProductID 100 non-null
                                    object
        ProductName 100 non-null
                                    object
     2 Category 100 non-null
                                    object
     3 Price
                     100 non-null
                                    float64
    dtypes: float64(1), object(3)
    memory usage: 3.3+ KB
    None
    <class 'pandas.core.frame.DataFrame'>
    RangeIndex: 1000 entries, 0 to 999
    Data columns (total 7 columns):
                        Non-Null Count Dtype
     # Column
    ---
                         _____
        TransactionID 1000 non-null object
                        1000 non-null object
        CustomerID
                        1000 non-null
     2 ProductID
                                       object
        TransactionDate 1000 non-null
                                       object
     4 Quantity
                        1000 non-null int64
     5 TotalValue
                        1000 non-null float64
                        1000 non-null float64
     6 Price
    dtypes: float64(2), int64(1), object(4)
    memory usage: 54.8+ KB
    None
print(f"Total Customers: {customers.shape[0]}")
print(f"Total Products: {products.shape[0]}")
print(f"Total Transactions: {transactions.shape[0]}")
total_revenue = transactions["TotalValue"].sum()
print(f"Total Revenue Generated: ${total_revenue}")
revenue_by_region = transactions.merge(customers, on="CustomerID").groupby("Region")["TotalValue"].sum()
print(revenue_by_region)
best_selling = transactions.merge(products, on="ProductID").groupby("ProductName")["Quantity"].sum().sort_values(ascending=False)
print(best_selling.head())
Total Customers: 200
Total Products: 100
    Total Transactions: 1000
    Total Revenue Generated: $689995.56
    Region
                    152074.97
    Asia
                    166254.63
    Europe
                   152313.40
    North America
    South America 219352.56
    Name: TotalValue, dtype: float64
    ProductName
    ActiveWear Smartwatch 100
    SoundWave Headphones
                            97
    HomeSense Desk Lamp
                            81
    ActiveWear Rug
                            79
    SoundWave Cookbook
                            78
    Name: Quantity, dtype: int64
import matplotlib.pyplot as plt
import seaborn as sns
# Revenue by Region
revenue_by_region.plot(kind='bar', title='Revenue by Region', xlabel='Region', ylabel='Revenue', color='skyblue')
plt.show()
# Best-Selling Products
best_selling.head(10).plot(kind='bar', title='Top 10 Best-Selling Products', xlabel='Products', ylabel='Quantity Sold', color='orange')
```

 $https://colab.research.google.com/drive/1CqlW89xKE8Lkg-aAx5rxcXwYLQSDAG4G? authuser=1\#scrollTo=x_YtNhiQUPuT\&printMode=true$

1/2

plt.show()



