nkd4yopke

January 23, 2025

[1]:
```
import numpy as np
import pandas as pd
import matplotlib.pyplot as plt
import seaborn as sns
```

[2]:
```
data = pd.read_csv(r"C:\Users\91703\Downloads\customer_purchase_data.csv")
```

[3]:
```
data.head()
```

[3]:

	Age	Gender	AnnualIncome	NumberOfPurchases	ProductCategory	TimeSpentOnWebsite	LoyaltyProgram	DiscountsAvailed	PurchaseStatus
0	40	1	66120.26794	8	0	30.568601	0	5	1
1	20	1	23579.77358	4	2	38.240097	0	5	0
2	27	1	127821.30640	11	2	31.633212	1	0	1
3	24	1	137798.62310	19	3	46.167059	0	4	1
4	31	1	99300.96422	19	1	19.823592	0	0	1

[5]:
```
#How many missing values are there in the dataset?
missing_values = data.isnull().sum().sum()
print("Total Missing Values in dataset:", missing_values)
```

```
Total Missing Values in dataset: 0
```

[4]:
```
#What is the average Age of customers?
average_age = data['Age'].mean()
print("Average Age of customers:", average_age)
```

```
Average Age of customers: 44.29866666666667
```

[9]:
```
# What is the distribution of Gender among customers?*
gender_distribution = data['Gender'].value_counts()
print(gender_distribution)
```

```
Gender
1    757
0    743
Name: count, dtype: int64
```

[10]:
```
# What is the average AnnualIncome of customers?
average_annual_income = data['AnnualIncome'].mean()
print("Average Annual Income:", average_annual_income)
```

```
Average Annual Income: 84249.16433862
```

[11]:
```
# What is the total NumberOfPurchases made by all customers?
total_purchases = data['NumberOfPurchases'].sum()
print("Total NumberOfPurchases:", total_purchases)
```

```
Total NumberOfPurchases: 15630
```

[12]:
```
#What are the unique ProductCategory values?
unique_product_categories = data['ProductCategory'].unique()
print("Unique Product Categories:")
print(unique_product_categories)
```

```
Unique Product Categories:
[0 2 3 1 4]
```

[20]:
```
#What is the average TimeSpentOnWebsite by customers?
average_time_spent = data['TimeSpentOnWebsite'].mean()
print("Average TimeSpentOnWebsite:", average_time_spent)
```

```
Average TimeSpentOnWebsite: 30.469040226028
```

[15]:
```
#What percentage of customers are enrolled in the LoyaltyProgram?
loyalty_program_enrollment = data['LoyaltyProgram'].
 ↪value_counts(normalize=True) * 100
print("LoyaltyProgram Enrollment Percentage:")
print(loyalty_program_enrollment)
```

```
LoyaltyProgram Enrollment Percentage:
LoyaltyProgram
0    67.333333
1    32.666667
Name: proportion, dtype: float64
```

[16]:
```
#What is the average number of DiscountsAvailed by customers?
average_discounts_availed = data['DiscountsAvailed'].mean()
print("Average DiscountsAvailed:", average_discounts_availed)
```

```
Average DiscountsAvailed: 2.5553333333333335
```

[17]:
```
#What percentage of customers have PurchaseStatus as "True" or equivalent␣
 ↪(assuming "True" means successful purchase)
purchase_status_percentage = data['PurchaseStatus'].
 ↪value_counts(normalize=True) * 100
print("PurchaseStatus Percentage:")
print(purchase_status_percentage)
```

```
PurchaseStatus Percentage:
PurchaseStatus
0    56.8
1    43.2
Name: proportion, dtype: float64
```

[18]:
```
#What is the correlation between AnnualIncome and NumberOfPurchases?
correlation = data['AnnualIncome'].corr(data['NumberOfPurchases'])
print("Correlation between AnnualIncome and NumberOfPurchases:", correlation)
```

```
Correlation between AnnualIncome and NumberOfPurchases: 0.00027645500852801217
```