Defination

Customer Personality Analysis is a data-driven process used by businesses to understand their customers deeply by analyzing their demographics, behavior, lifestyle, and purchasing patterns. The main idea is to group customers based on similarities so companies can design targeted marketing strategies instead of treating all customers the same . This analysis involves collecting various types of information such as age, gender, income level, family details, education, occupation, and location. It also includes behavioral data—like what products customers buy, how often they shop, how much money they spend, and which marketing channels they respond to (email, social media, in-store, etc.).

Objectives

1. Understand customer demographics:

Identify age, gender, marital status, income, and family size to know who the customers are.

2. Segment customers into meaningful groups:

Create clusters/groups based on buying habits so the company can target each group differently.

3. Analyze customer spending patterns:

Study how much customers spend and on which product categories to understand their preferences.

4. Evaluate customer interaction with marketing channels:

Check which channels customers use most (email, store, web, social media) to improve communication.

5. Measure marketing campaign performance:

Understand which campaigns worked, how many people responded, and which ones failed.

6. Identify high-value customers:

Find customers who spend more, purchase frequently, and contribute maximum profit.

7. Detect low-engagement or inactive customers:

Identify customers who are not responding or have stopped buying so the company can re-engage them.

8. Personalize marketing and offers:

Use customer behavior insights to provide targeted discounts, recommendations, and messages.

9. Improve customer satisfaction:

Discover what customers like or dislike to help the company serve them better.

10. Optimize marketing costs:

Avoid spending money on people who don't respond and focus on profitable segments.

Data Incpection

import pandas as pd

s1 = pd.read_csv("C:\Customer_Personality_Analysis.csv")

s1

	CustomerID	Gender	Age	Education	MaritalStatus	Income	Country	SpendingScore	TotalSpent	ProductPreference	DateJoined
0	94260	Female	49	Basic	Single	116913.09	Canada	93	9091.24	Fish	11-10-2022
1	81359	Male	45	Graduation	Together	108466.47	France	2	7231.04	Gold	14-04-2018
2	77297	Male	27	Master	Divorced	100335.54	France	44	3333.32	Wine	13-12-2019
3	12701	Male	54	PhD	Single	84681.17	UK	84	3527.18	Gold	19-04-2019
4	92399	Male	60	PhD	Together	91396.81	France	52	7087.10	Meat	04-04-2019
...	...	...	...	...	...	...	...	...	...	...	...
995	99984	Male	23	PhD	Married	21004.37	USA	39	7772.67	Fruits	15-11-2020
996	32952	Female	68	PhD	Single	37221.37	France	91	8954.68	Meat	17-07-2019
997	38915	Male	52	Master	Single	92823.12	France	88	9653.71	Fruits	14-05-2020
998	56034	Female	18	Basic	Single	36589.32	Germany	28	9893.41	Meat	29-07-2018
999	53023	Male	24	Graduation	Together	106153.98	France	30	1988.48	Clothes	25-12-2019

1000 rows × 11 columns

s1.shape

(1000, 11)

s1.info

<bound method DataFrame.info of   CustomerID Gender Age   Education MaritalStatus   Income   Country \

0   94260 Female 49   Basic   Single 116913.09 Canada

1   81359   Male 45 Graduation   Together 108466.47 France

	CustomerID	Gender	Age	Education	MaritalStatus	Income	Country
2	77297	Male	27	Master	Divorced	100335.54	France
3	12701	Male	54	PhD	Single	84681.17	UK
4	92399	Male	60	PhD	Together	91396.81	France
..	...	...	...	...	...	...	...
995	99984	Male	23	PhD	Married	21004.37	USA
996	32952	Female	68	PhD	Single	37221.37	France
997	38915	Male	52	Master	Single	92823.12	France
998	56034	Female	18	Basic	Single	36589.32	Germany
999	53023	Male	24	Graduation	Together	106153.98	France

	SpendingScore	TotalSpent	ProductPreference	DateJoined
0	93	9091.24	Fish	11-10-2022
1	2	7231.04	Gold	14-04-2018
2	44	3333.32	Wine	13-12-2019
3	84	3527.18	Gold	19-04-2019
4	52	7087.10	Meat	04-04-2019
..	...	...	...	...
995	39	7772.67	Fruits	15-11-2020
996	91	8954.68	Meat	17-07-2019
997	88	9653.71	Fruits	14-05-2020
998	28	9893.41	Meat	29-07-2018
999	30	1988.48	Clothes	25-12-2019

[1000 rows x 11 columns]>

s1.tail

<bound method NDFrame.tail of   CustomerID Gender Age Education MaritalStatus   Income Country \

	CustomerID	Gender	Age	Education	MaritalStatus	Income	Country
0	94260	Female	49	Basic	Single	116913.09	Canada
1	81359	Male	45	Graduation	Together	108466.47	France
2	77297	Male	27	Master	Divorced	100335.54	France
3	12701	Male	54	PhD	Single	84681.17	UK

4	92399	Male	60	PhD	Together	91396.81	France
..	...	...	...	...	...	...	...
995	99984	Male	23	PhD	Married	21004.37	USA
996	32952	Female	68	PhD	Single	37221.37	France
997	38915	Male	52	Master	Single	92823.12	France
998	56034	Female	18	Basic	Single	36589.32	Germany
999	53023	Male	24	Graduation	Together	106153.98	France

	SpendingScore	TotalSpent	ProductPreference	DateJoined
0	93	9091.24	Fish	11-10-2022
1	2	7231.04	Gold	14-04-2018
2	44	3333.32	Wine	13-12-2019
3	84	3527.18	Gold	19-04-2019
4	52	7087.10	Meat	04-04-2019
..	...	...	...	...
995	39	7772.67	Fruits	15-11-2020
996	91	8954.68	Meat	17-07-2019
997	88	9653.71	Fruits	14-05-2020
998	28	9893.41	Meat	29-07-2018
999	30	1988.48	Clothes	25-12-2019

[1000 rows x 11 columns]>

s1.columns

Index(['CustomerID', 'Gender', 'Age', 'Education', 'MaritalStatus', 'Income',

'Country', 'SpendingScore', 'TotalSpent', 'ProductPreference',

'DateJoined'],

dtype='object')

Data Cleaning

s1= pd.read_csv(r"C:\Customer_Personality_Analysis.csv")

s1

	CustomerID	Gender	Age	Education	MaritalStatus	Income	Country	SpendingScore	TotalSpent	ProductPreference	DateJoined
0	94260	Female	49	Basic	Single	116913.09	Canada	93	9091.24	Fish	11-10-2022
1	81359	Male	45	Graduation	Together	108466.47	France	2	7231.04	Gold	14-04-2018
2	77297	Male	27	Master	Divorced	100335.54	France	44	3333.32	Wine	13-12-2019
3	12701	Male	54	PhD	Single	84681.17	UK	84	3527.18	Gold	19-04-2019
4	92399	Male	60	PhD	Together	91396.81	France	52	7087.10	Meat	04-04-2019
...	...	...	...	...	...	...	...	...	...	...	...
995	99984	Male	23	PhD	Married	21004.37	USA	39	7772.67	Fruits	15-11-2020
996	32952	Female	68	PhD	Single	37221.37	France	91	8954.68	Meat	17-07-2019
997	38915	Male	52	Master	Single	92823.12	France	88	9653.71	Fruits	14-05-2020
998	56034	Female	18	Basic	Single	36589.32	Germany	28	9893.41	Meat	29-07-2018
999	53023	Male	24	Graduation	Together	106153.98	France	30	1988.48	Clothes	25-12-2019

1000 rows × 11 columns

s1 = s1.drop_duplicates()

s1

	CustomerID	Gender	Age	Education	MaritalStatus	Income	Country	SpendingScore	TotalSpent	ProductPreference	DateJoined
0	94260	Female	49	Basic	Single	116913.09	Canada	93	9091.24	Fish	11-10-2022
1	81359	Male	45	Graduation	Together	108466.47	France	2	7231.04	Gold	14-04-2018

	CustomerID	Gender	Age	Education	MaritalStatus	Income	Country	SpendingScore	TotalSpent	ProductPreference	DateJoined
2	77297	Male	27	Master	Divorced	100335.54	France	44	3333.32	Wine	13-12-2019
3	12701	Male	54	PhD	Single	84681.17	UK	84	3527.18	Gold	19-04-2019
4	92399	Male	60	PhD	Together	91396.81	France	52	7087.10	Meat	04-04-2019
...	...	...	...	...	...	...	...	...	...	...	...
995	99984	Male	23	PhD	Married	21004.37	USA	39	7772.67	Fruits	15-11-2020
996	32952	Female	68	PhD	Single	37221.37	France	91	8954.68	Meat	17-07-2019
997	38915	Male	52	Master	Single	92823.12	France	88	9653.71	Fruits	14-05-2020
998	56034	Female	18	Basic	Single	36589.32	Germany	28	9893.41	Meat	29-07-2018
999	53023	Male	24	Graduation	Together	106153.98	France	30	1988.48	Clothes	25-12-2019

1000 rows × 11 columns

s1.dropna(inplace=True)

s1

	CustomerID	Gender	Age	Education	MaritalStatus	Income	Country	SpendingScore	TotalSpent	ProductPreference	DateJoined
0	94260	Female	49	Basic	Single	116913.09	Canada	93	9091.24	Fish	11-10-2022
1	81359	Male	45	Graduation	Together	108466.47	France	2	7231.04	Gold	14-04-2018
2	77297	Male	27	Master	Divorced	100335.54	France	44	3333.32	Wine	13-12-2019
3	12701	Male	54	PhD	Single	84681.17	UK	84	3527.18	Gold	19-04-2019

	Custom erID	Gen der	Ag e	Educati on	MaritalSt atus	Incom e	Count ry	Spending Score	TotalSp ent	ProductPref erence	DateJoi ned
4	92399	Male	60	PhD	Together	91396. 81	Franc e	52	7087.1 0	Meat	04-04-2019
...	...	...	...	...	...	...	...	...	...	...	...
99 5	99984	Male	23	PhD	Married	21004. 37	USA	39	7772.6 7	Fruits	15-11-2020
99 6	32952	Fem ale	68	PhD	Single	37221. 37	Franc e	91	8954.6 8	Meat	17-07-2019
99 7	38915	Male	52	Master	Single	92823. 12	Franc e	88	9653.7 1	Fruits	14-05-2020
99 8	56034	Fem ale	18	Basic	Single	36589. 32	Germ any	28	9893.4 1	Meat	29-07-2018
99 9	53023	Male	24	Gradua tion	Together	10615 3.98	Franc e	30	1988.4 8	Clothes	25-12-2019

1000 rows × 11 columns

s1.isnull().sum()

CustomerID   0

Gender   0

Age   0

Education   0

MaritalStatus   0

Income   0

Country   0

SpendingScore   0

TotalSpent   0

ProductPreference   0

DateJoined   0

dtype: int64

s1['CustomerID'].nunique()

994

s1['Gender'].value_counts()

Gender

Male   501

Female   499

Name: count, dtype: int64

check the datatype
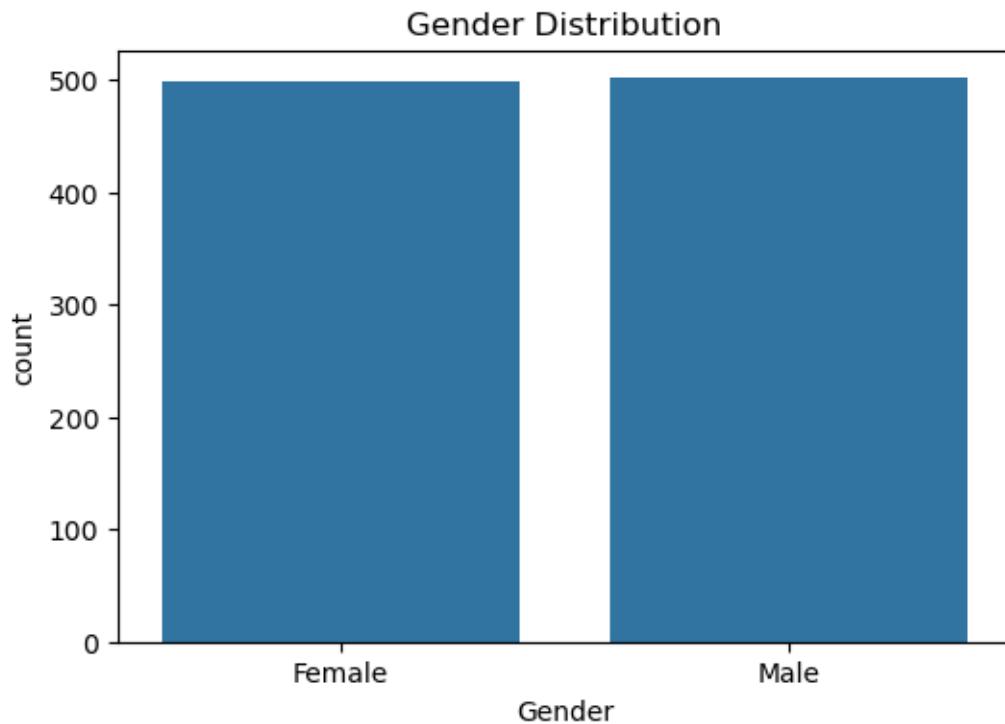
s1.dtypes

CustomerID   int64

Gender   object

Age   int64

Education   object

MaritalStatus   object

Income   float64

Country   object

SpendingScore   int64

TotalSpent   float64

ProductPreference   object

DateJoined   object
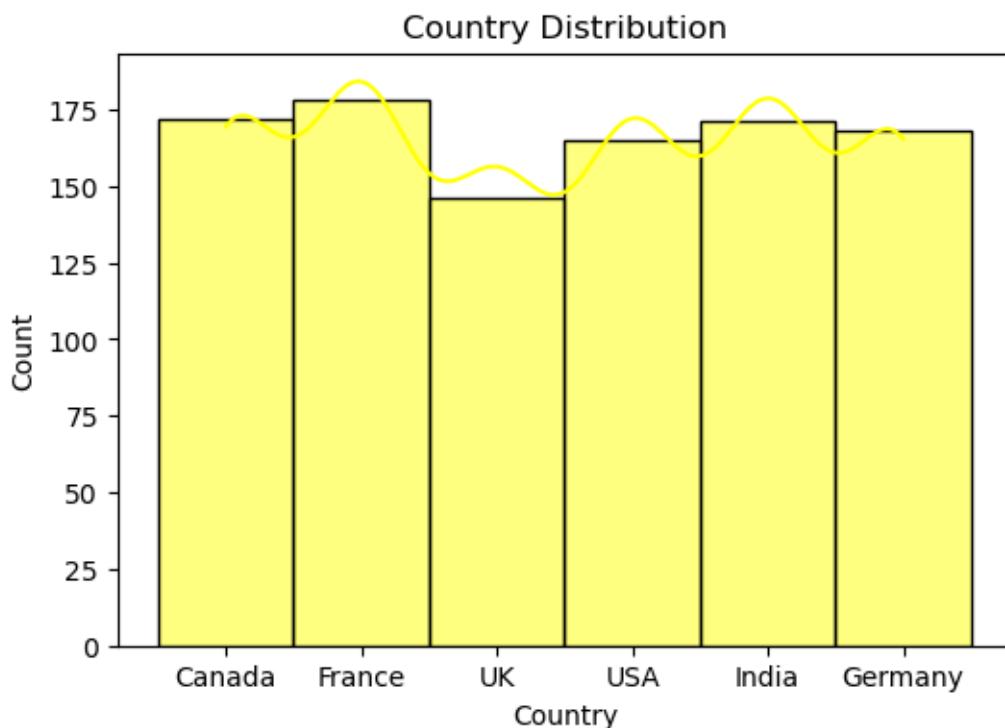
dtype: object

Exploratory Data Analysis (EDA)

```python
import pandas as pd
import numpy as np
import matplotlib.pyplot as plt
import seaborn as sns
plt.figure(figsize=(6,4))
sns.countplot(x='Gender', data=s1)
plt.title("Gender Distribution")
plt.show()
```

Gender Distribution
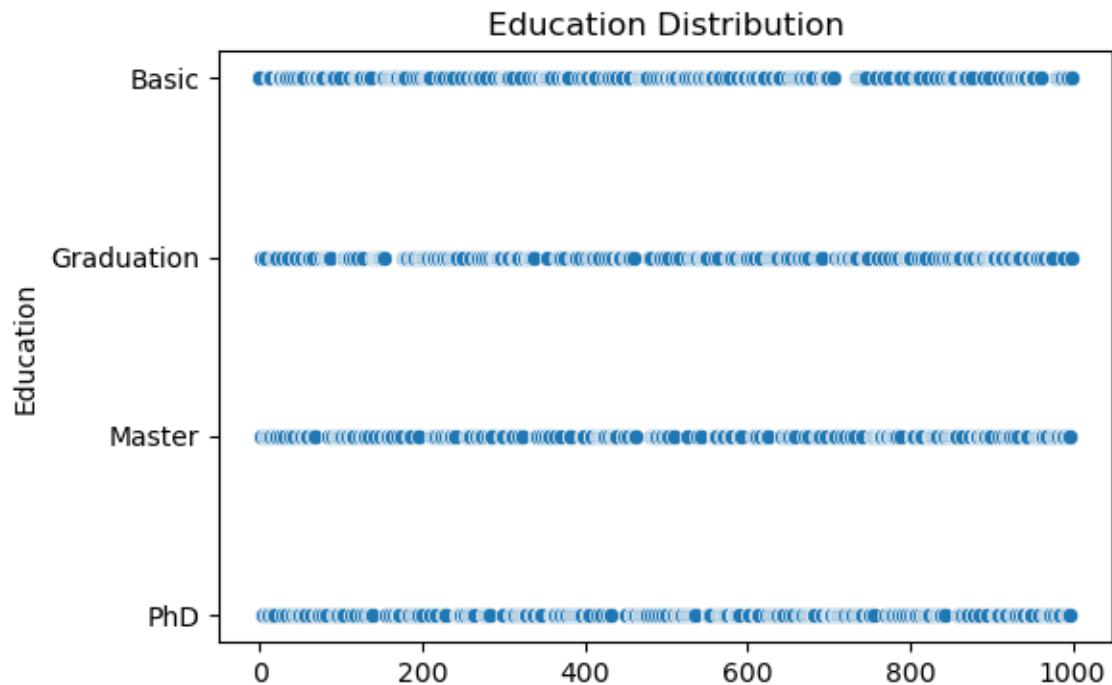
```
plt.figure(figsize=(6,4))

sns.histplot(s1['Country'], kde=True, color='yellow')

plt.title("Country Distribution")

plt.show()
```


Country Distribution
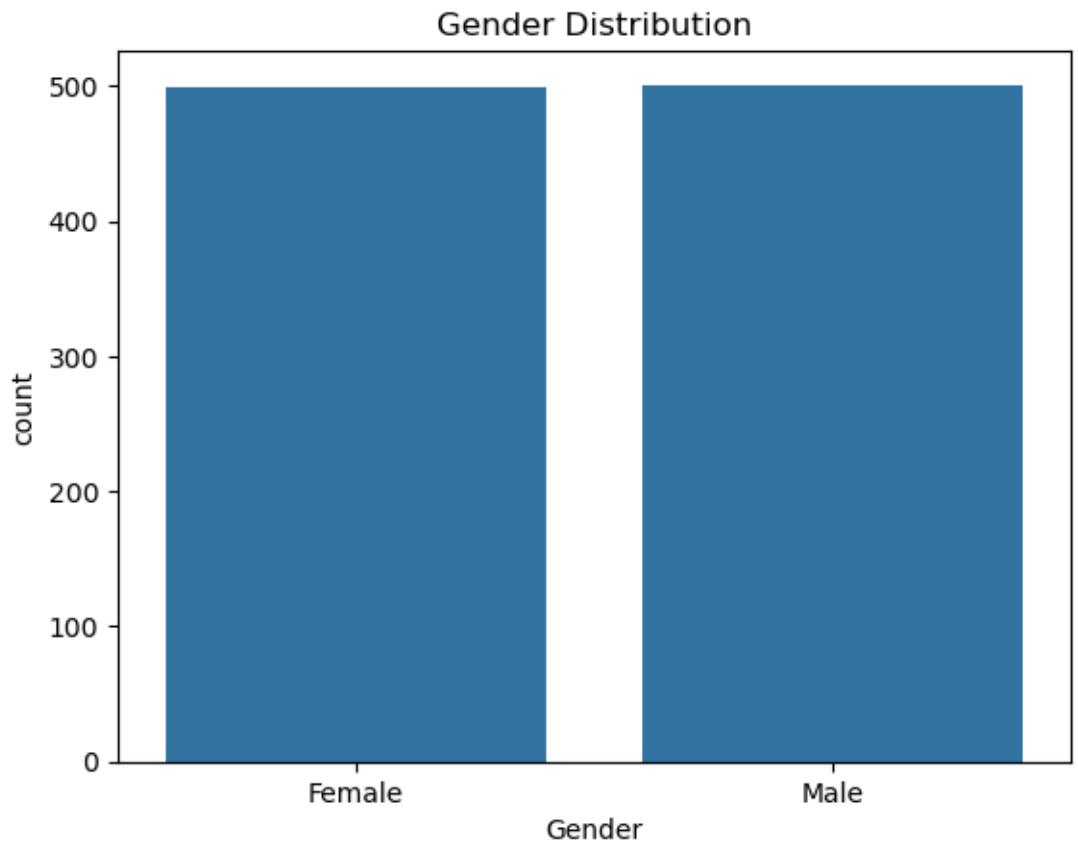
```
plt.figure(figsize=(6,4))
```

```
sns.scatterplot(s1['Education'])

plt.title("Education Distribution")

plt.show()
```


Education Distribution
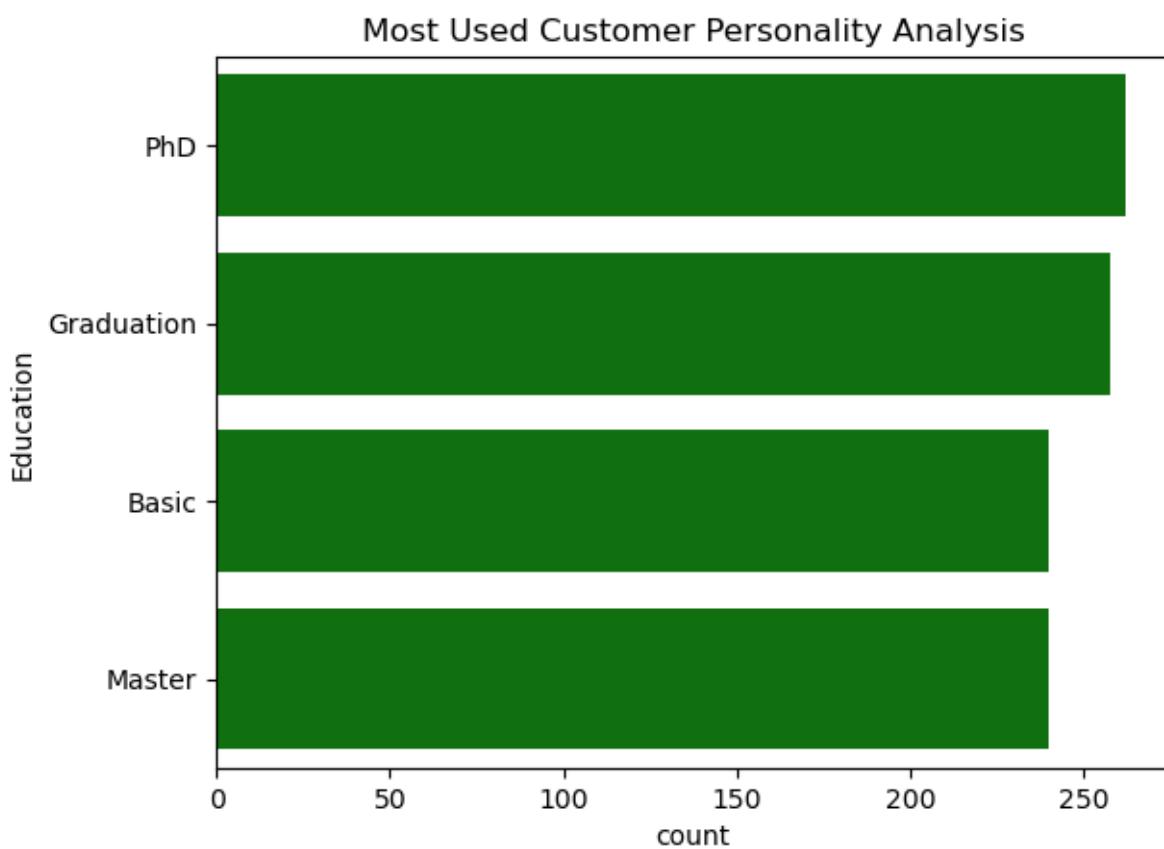
Visulaization

```
sns.countplot(x='Gender', data=s1)

plt.title("Gender Distribution")

plt.show()
```

Gender Distribution
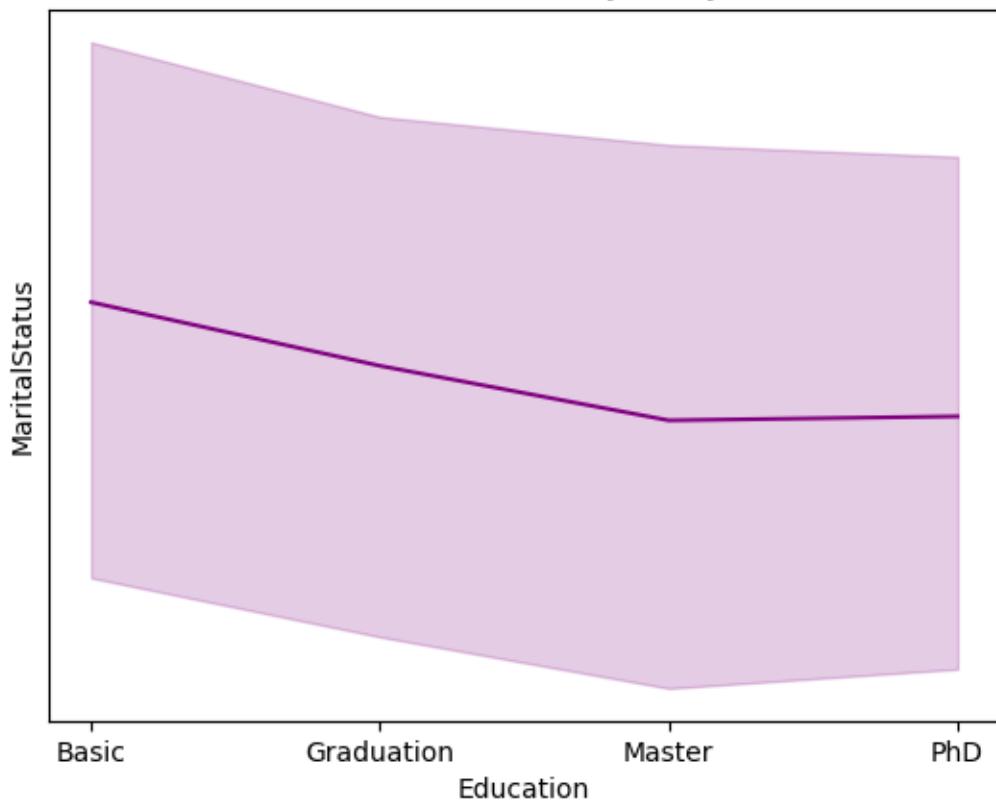
```
sns.countplot(y='Education',data=s1,order=s1['Education'].value_counts().index,color='green')
plt.title("Most Used Customer Personality Analysis")
plt.show()
```

Most Used Customer Personality Analysis

```
sns.lineplot(x='Education', y='MaritalStatus', data=s1, color='purple')

plt.title("Customer Personality Analysis")

plt.show()
```

Customer Personality Analysis

```
sns.heatmap(s1.corr(numeric_only=True), annot=True, cmap='coolwarm')

plt.title("Matrix")

plt.show()
```

Matrix