```python
import pandas as pd
import matplotlib.pyplot as plt
import seaborn as sns
```

```python
df = pd.read_csv('/cleaned_marketing_data.csv')
df.head()
```

|   | id | year_birth | education | marital_status | income | kidhome | teenhome | dt_customer | recency | mntwines | ... | acceptedcmp5 | acceptedcmp |
|---|----|-----------|-----------|----------------|--------|---------|----------|-------------|---------|----------|-----|--------------|-------------|
| 0 | 5524 | 1957 | graduation | single | 58138.0 | 0 | 0 | 2012-04-09 | 58 | 635 | ... | 0 | |
| 1 | 2174 | 1954 | graduation | single | 46344.0 | 1 | 1 | 2014-08-03 | 38 | 11 | ... | 0 | |
| 2 | 4141 | 1965 | graduation | married | 71613.0 | 0 | 0 | 2000-01-01 | 26 | 426 | ... | 0 | |
| 3 | 6182 | 1984 | graduation | married | 26646.0 | 1 | 0 | 2014-10-02 | 26 | 11 | ... | 0 | |
| 4 | 5324 | 1981 | phd | married | 58293.0 | 1 | 0 | 2000-01-01 | 94 | 173 | ... | 0 | |

5 rows × 32 columns

```python
print("Shape:", df.shape)
print("Columns:", df.columns)
```

```
Shape: (2216, 32)
Columns: Index(['id', 'year_birth', 'education', 'marital_status', 'income', 'kidhome',
       'teenhome', 'dt_customer', 'recency', 'mntwines', 'mntfruits',
       'mntmeatproducts', 'mntfishproducts', 'mntsweetproducts',
       'mntgoldprods', 'numdealspurchases', 'numwebpurchases',
       'numcatalogpurchases', 'numstorepurchases', 'numwebvisitsmonth',
       'acceptedcmp3', 'acceptedcmp4', 'acceptedcmp5', 'acceptedcmp1',
       'acceptedcmp2', 'complain', 'z_costcontact', 'z_revenue', 'response',
       'age', 'total_children', 'total_spent'],
      dtype='object')
```

```python
df.dtypes
df.isnull().sum()
```

|  | 0 |
| --- | --- |
| id | 0 |
| year_birth | 0 |
| education | 0 |
| marital_status | 0 |
| income | 0 |
| kidhome | 0 |
| teenhome | 0 |
| dt_customer | 0 |
| recency | 0 |
| mntwines | 0 |
| mntfruits | 0 |
| mntmeatproducts | 0 |
| mntfishproducts | 0 |
| mntsweetproducts | 0 |
| mntgoldprods | 0 |
| numdealspurchases | 0 |
| numwebpurchases | 0 |
| numcatalogpurchases | 0 |
| numstorepurchases | 0 |
| numwebvisitsmonth | 0 |
| acceptedcmp3 | 0 |
| acceptedcmp4 | 0 |
| acceptedcmp5 | 0 |
| acceptedcmp1 | 0 |
| acceptedcmp2 | 0 |
| complain | 0 |
| z_costcontact | 0 |
| z_revenue | 0 |
| response | 0 |
| age | 0 |
| total_children | 0 |
| total_spent | 0 |

**dtype:** int64

```
df.describe()
```

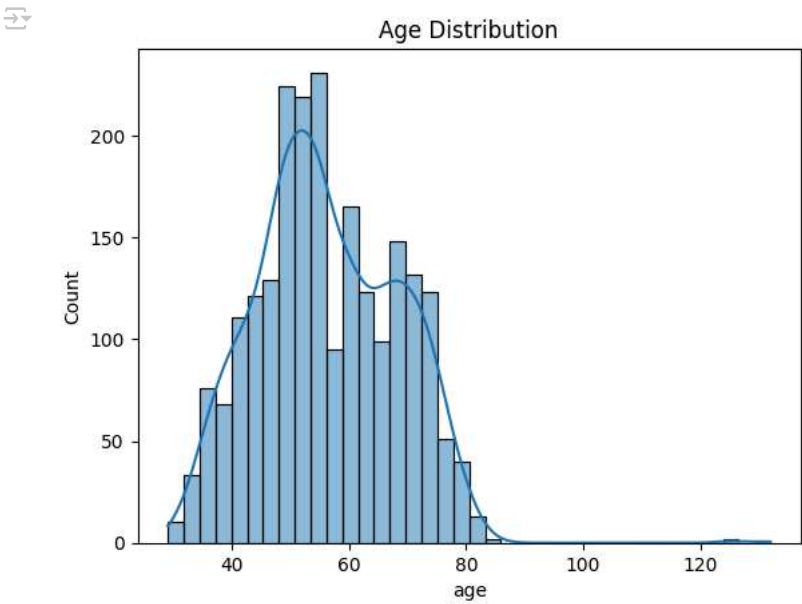|       | id | year_birth | income | kidhome | teenhome | recency | mntwines | mntfruits | mntmeatproducts | mntfishpr |
|-------|-----------|------------|--------------|----------|----------|-------------|-------------|------------|-----------------|-----------|
| count | 2216.000000 | 2216.000000 | 2216.000000 | 2216.000000 | 2216.000000 | 2216.000000 | 2216.000000 | 2216.000000 | 2216.000000 | 2216. |
| mean | 5588.353339 | 1968.820397 | 52247.251354 | 0.441787 | 0.505415 | 49.012635 | 305.091606 | 26.356047 | 166.995939 | 37. |
| std | 3249.376275 | 11.985554 | 25173.076661 | 0.536896 | 0.544181 | 28.948352 | 337.327920 | 39.793917 | 224.283273 | 54. |
| min | 0.000000 | 1893.000000 | 1730.000000 | 0.000000 | 0.000000 | 0.000000 | 0.000000 | 0.000000 | 0.000000 | 0. |
| 25% | 2814.750000 | 1959.000000 | 35303.000000 | 0.000000 | 0.000000 | 24.000000 | 24.000000 | 2.000000 | 16.000000 | 3. |
| 50% | 5458.500000 | 1970.000000 | 51381.500000 | 0.000000 | 0.000000 | 49.000000 | 174.500000 | 8.000000 | 68.000000 | 12. |
| 75% | 8421.750000 | 1977.000000 | 68522.000000 | 1.000000 | 1.000000 | 74.000000 | 505.000000 | 33.000000 | 232.250000 | 50. |
| max | 11191.000000 | 1996.000000 | 666666.000000 | 2.000000 | 2.000000 | 99.000000 | 1493.000000 | 199.000000 | 1725.000000 | 259. |

8 rows × 29 columns

```
df.duplicated().sum()
```

```
np.int64(0)
```

```
sns.histplot(df['age'], kde=True)
plt.title('Age Distribution')
plt.show()
```



```
df.columns.tolist()
```

```
['id',
 'year_birth',
 'education',
 'marital_status',
 'income',
 'kidhome',
 'teenhome',
 'dt_customer',
 'recency',
 'mntwines',
 'mntfruits',
 'mntmeatproducts',
 'mntfishproducts',
 'mntsweetproducts',
 'mntgoldprods',
 'numdealspurchases',
 'numwebpurchases',
 'numcatalogpurchases',
 'numstorepurchases',
 'numwebvisitsmonth',
```

```
        'acceptedcmp3',
        'acceptedcmp4',
        'acceptedcmp5',
        'acceptedcmp1',
        'acceptedcmp2',
        'complain',
        'z_costcontact',
        'z_revenue',
        'response',
        'age',
        'total_children',
        'total_spent']
```

```
df.columns = df.columns.str.strip().str.lower()
```

```
sns.scatterplot(x='age', y='income', data=df)
plt.title("Age vs Income")
plt.show()
```
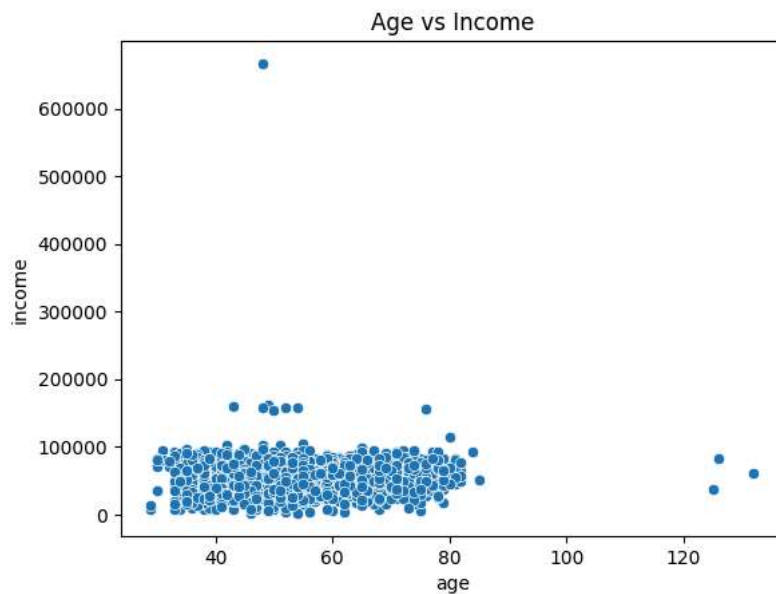


```
numeric_df = df.select_dtypes(include=['number'])
```

```
plt.figure(figsize=(10,6))
sns.heatmap(numeric_df.corr(), annot=True, cmap='coolwarm')
plt.title("Correlation Matrix")
plt.show()
```

Correlation Matrix

```
sns.boxplot(x=df['income'])
```