Project Title - Customer Personality Analysis

The Customer Personality data Analysis is one of the best analysis to collect a information from the customer and in which data maximum information is given. We analysis some informations to get important data like customer in which product to money investing. This data set collecting from kaggle.com, This dataset contains 29 variables and 2240 observations about different customers.

Here's a brief version of the data description file.

People

ID: Customer's unique identifier

Year_Birth: Customer's birth year

Education: Customer's education level

Marital_Status: Customer's marital status

Income: Customer's yearly household income

Kidhome: Number of children in customer's household

Teenhome: Number of teenagers in customer's household

Dt_Customer: Date of customer's enrollment with the company

Recency: Number of days since customer's last purchase

Complain: 1 if customer complained in the last 2 years, 0 otherwise

Products

MntWines: Amount spent on wine in last 2 years

MntFruits: Amount spent on fruits in last 2 years

MntMeatProducts: Amount spent on meat in last 2 years

MntFishProducts: Amount spent on fish in last 2 years

MntSweetProducts: Amount spent on sweets in last 2 years

MntGoldProds: Amount spent on gold in last 2 years

Promotion

NumDealsPurchases: Number of purchases made with a discount

AcceptedCmp1: 1 if customer accepted the offer in the 1st campaign, 0 otherwise

AcceptedCmp2: 1 if customer accepted the offer in the 2nd campaign, 0 otherwise

AcceptedCmp3: 1 if customer accepted the offer in the 3rd campaign, 0 otherwise

AcceptedCmp4: 1 if customer accepted the offer in the 4th campaign, 0 otherwise

AcceptedCmp5: 1 if customer accepted the offer in the 5th campaign, 0 otherwise

Response: 1 if customer accepted the offer in the last campaign, 0 otherwise

Place

NumWebPurchases: Number of purchases made through the company’s web site

NumCatalogPurchases: Number of purchases made using a catalogue

NumStorePurchases: Number of purchases made directly in stores

NumWebVisitsMonth: Number of visits to company’s web site in the last month

In [2]:
```
import pandas as pd
import numpy as np
import seaborn as sns
import matplotlib.pyplot as plt
```

In [4]:
```
cust_per_df = pd.read_csv("marketing_campaign_new.csv")
cust_per_df.head()
```

Out[4]:

	ID	Year_Birth	Education	Marital_Status	Income	Kidhome	Teenhome	Dt_Customer	Recency	MntWines	...	NumWebVisitsMonth	AcceptedCmp3	Acce
0	5524	1957	Graduation	Single	58138.0	0	0	04-09-2012	58	635	...	7	0	
1	2174	1954	Graduation	Single	46344.0	1	1	08-03-2014	38	11	...	5	0	
2	4141	1965	Graduation	Together	71613.0	0	0	21-08-2013	26	426	...	4	0	
3	6182	1984	Graduation	Together	26646.0	1	0	10-02-2014	26	11	...	6	0	
4	5324	1981	PhD	Married	58293.0	1	0	19-01-2014	94	173	...	5	0	

5 rows × 29 columns

In [5]:
```
cust_per_df.shape
```

Out[5]: (2240, 29)

In [6]:
```
cust_per_df1 = cust_per_df.copy()
cust_per_df1
```

Out[6]:

	ID	Year_Birth	Education	Marital_Status	Income	Kidhome	Teenhome	Dt_Customer	Recency	MntWines	...	NumWebVisitsMonth	AcceptedCmp3	A
0	5524	1957	Graduation	Single	58138.0	0	0	04-09-2012	58	635	...	7	0	
1	2174	1954	Graduation	Single	46344.0	1	1	08-03-2014	38	11	...	5	0	
2	4141	1965	Graduation	Together	71613.0	0	0	21-08-2013	26	426	...	4	0	
3	6182	1984	Graduation	Together	26646.0	1	0	10-02-2014	26	11	...	6	0	
4	5324	1981	PhD	Married	58293.0	1	0	19-01-2014	94	173	...	5	0	
...	...	...	...	...	...	...	...	...	...	...	...	...	...	...
2235	10870	1967	Graduation	Married	61223.0	0	1	13-06-2013	46	709	...	5	0	
2236	4001	1946	PhD	Together	64014.0	2	1	10-06-2014	56	406	...	7	0	
2237	7270	1981	Graduation	Divorced	56981.0	0	0	25-01-2014	91	908	...	6	0	
2238	8235	1956	Master	Together	69245.0	0	1	24-01-2014	8	428	...	3	0	
2239	9405	1954	PhD	Married	52869.0	1	1	15-10-2012	40	84	...	7	0	

2240 rows × 29 columns

In [7]:
```
# check the null values
cust_per_df1.isna().sum()
```

Out[7]:
```
ID                     0
Year_Birth             0
Education              0
Marital_Status         0
Income                24
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
dtype: int64
```

In [8]:
```
cust_per_df1["child"] = cust_per_df1["Kidhome"] + cust_per_df1["Teenhome"]
```

In [9]:
```
cust_per_df1
```

Out[9]:

Recency	MntWines	...	AcceptedCmp3	AcceptedCmp4	AcceptedCmp5	AcceptedCmp1	AcceptedCmp2	Complain	Z_CostContact	Z_Revenue	Response	child
58	635	...	0	0	0	0	0	0	3	11	1	0
38	11	...	0	0	0	0	0	0	3	11	0	2
26	426	...	0	0	0	0	0	0	3	11	0	0
26	11	...	0	0	0	0	0	0	3	11	0	1
94	173	...	0	0	0	0	0	0	3	11	0	1
...	...	...	...	...	...	...	...	...	...	...	...	...
46	709	...	0	0	0	0	0	0	3	11	0	1
56	406	...	0	0	0	1	0	0	3	11	0	3
91	908	...	0	1	0	0	0	0	3	11	0	0
8	428	...	0	0	0	0	0	0	3	11	0	1
40	84	...	0	0	0	0	0	0	3	11	1	2

In [10]:
```
cust_per_df1.columns
```

Out[10]:
```
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',
       'child'],
      dtype='object')
```

In [11]:
```
len(cust_per_df1.columns)
```

Out[11]:
```
30
```

In [12]:
```
cust_per_df1["total_amount_spent"] = cust_per_df1.MntWines + cust_per_df1.MntFruits + cust_per_df1.MntMeatProducts + \
                                     cust_per_df1.MntFishProducts + cust_per_df1.MntSweetProducts + cust_per_df1.MntGoldProds
```

In [14]:
```
cust_per_df1.head()
```

Out[14]:

	ID	Year_Birth	Education	Marital_Status	Income	Kidhome	Teenhome	Dt_Customer	Recency	MntWines	...	AcceptedCmp4	AcceptedCmp5	AcceptedC
0	5524	1957	Graduation	Single	58138.0	0	0	04-09-2012	58	635	...	0	0	
1	2174	1954	Graduation	Single	46344.0	1	1	08-03-2014	38	11	...	0	0	
2	4141	1965	Graduation	Together	71613.0	0	0	21-08-2013	26	426	...	0	0	
3	6182	1984	Graduation	Together	26646.0	1	0	10-02-2014	26	11	...	0	0	
4	5324	1981	PhD	Married	58293.0	1	0	19-01-2014	94	173	...	0	0	

5 rows × 31 columns

In [15]:
```
cust_per_df1["Dt_Customer"] = pd.to_datetime(cust_per_df1.Dt_Customer)
```

```
vide format or specify infer_datetime_format=True for consistent parsing.
  cust_per_df1["Dt_Customer"] = pd.to_datetime(cust_per_df1.Dt_Customer)
C:\Users\DHRUV\AppData\Local\Temp\ipykernel_5944\3321778344.py:1: UserWarning: Parsing '30-11-2013' in DD/MM/YYYY format. Pro
vide format or specify infer_datetime_format=True for consistent parsing.
  cust_per_df1["Dt_Customer"] = pd.to_datetime(cust_per_df1.Dt_Customer)
C:\Users\DHRUV\AppData\Local\Temp\ipykernel_5944\3321778344.py:1: UserWarning: Parsing '27-03-2014' in DD/MM/YYYY format. Pro
vide format or specify infer_datetime_format=True for consistent parsing.
  cust_per_df1["Dt_Customer"] = pd.to_datetime(cust_per_df1.Dt_Customer)
C:\Users\DHRUV\AppData\Local\Temp\ipykernel_5944\3321778344.py:1: UserWarning: Parsing '15-12-2012' in DD/MM/YYYY format. Pro
vide format or specify infer_datetime_format=True for consistent parsing.
  cust_per_df1["Dt_Customer"] = pd.to_datetime(cust_per_df1.Dt_Customer)
C:\Users\DHRUV\AppData\Local\Temp\ipykernel_5944\3321778344.py:1: UserWarning: Parsing '17-09-2012' in DD/MM/YYYY format. Pro
vide format or specify infer_datetime_format=True for consistent parsing.
  cust_per_df1["Dt_Customer"] = pd.to_datetime(cust_per_df1.Dt_Customer)
C:\Users\DHRUV\AppData\Local\Temp\ipykernel_5944\3321778344.py:1: UserWarning: Parsing '21-12-2012' in DD/MM/YYYY format. Pro
vide format or specify infer_datetime_format=True for consistent parsing.
  cust_per_df1["Dt_Customer"] = pd.to_datetime(cust_per_df1.Dt_Customer)
C:\Users\DHRUV\AppData\Local\Temp\ipykernel_5944\3321778344.py:1: UserWarning: Parsing '20-12-2012' in DD/MM/YYYY format. Pro
vide format or specify infer_datetime_format=True for consistent parsing.
  cust_per_df1["Dt_Customer"] = pd.to_datetime(cust_per_df1.Dt_Customer)
```

In [16]:
```
cust_per_df1
```

Out[16]:

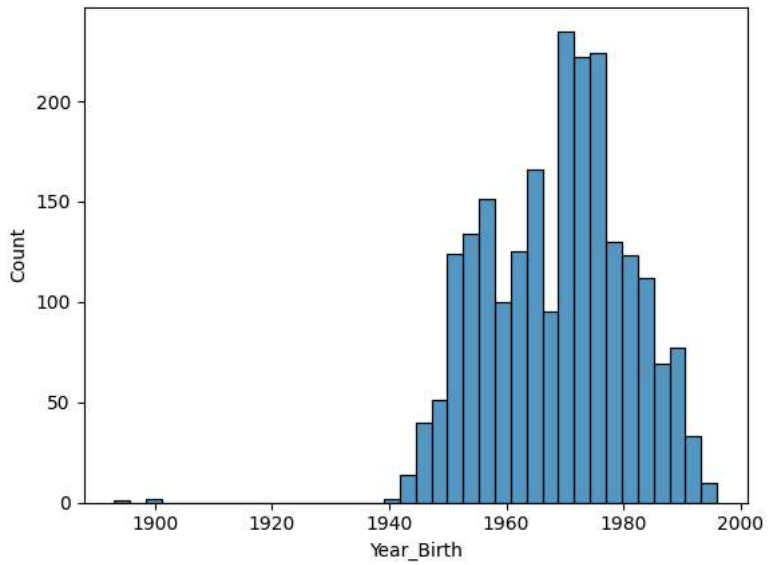
	ID	Year_Birth	Education	Marital_Status	Income	Kidhome	Teenhome	Dt_Customer	Recency	MntWines	...	AcceptedCmp4	AcceptedCmp5	Accep
0	5524	1957	Graduation	Single	58138.0	0	0	2012-04-09	58	635	...	0	0	
1	2174	1954	Graduation	Single	46344.0	1	1	2014-08-03	38	11	...	0	0	
2	4141	1965	Graduation	Together	71613.0	0	0	2013-08-21	26	426	...	0	0	
3	6182	1984	Graduation	Together	26646.0	1	0	2014-10-02	26	11	...	0	0	
4	5324	1981	PhD	Married	58293.0	1	0	2014-01-19	94	173	...	0	0	
...	...	...	...	...	...	...	...	...	...	...	...	...	...	
2235	10870	1967	Graduation	Married	61223.0	0	1	2013-06-13	46	709	...	0	0	
2236	4001	1946	PhD	Together	64014.0	2	1	2014-10-06	56	406	...	0	0	
2237	7270	1981	Graduation	Divorced	56981.0	0	0	2014-01-25	91	908	...	1	0	
2238	8235	1956	Master	Together	69245.0	0	1	2014-01-24	8	428	...	0	0	
2239	9405	1954	PhD	Married	52869.0	1	1	2012-10-15	40	84	...	0	0	

2240 rows × 31 columns

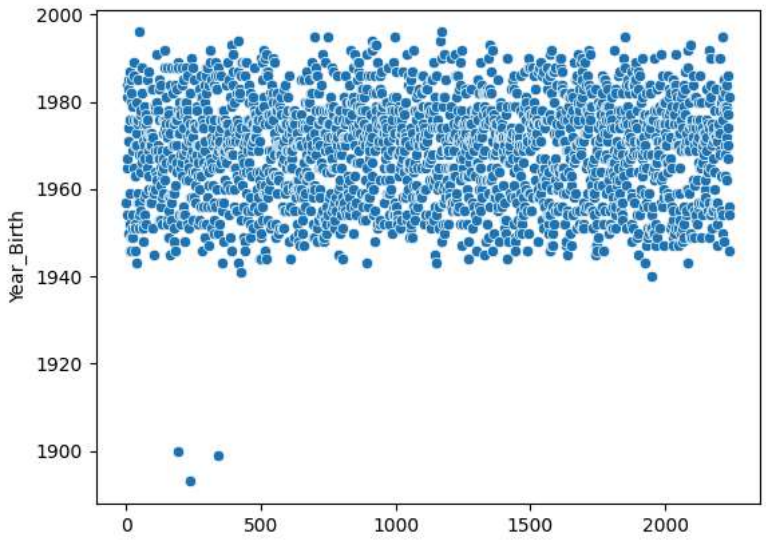
In [17]:
```
cust_per_df1.Year_Birth.unique()
```

Out[17]:
```
array([1957, 1954, 1965, 1984, 1981, 1967, 1971, 1985, 1974, 1950, 1983,
       1976, 1959, 1952, 1987, 1946, 1980, 1949, 1982, 1979, 1951, 1969,
       1986, 1989, 1963, 1970, 1973, 1943, 1975, 1996, 1968, 1964, 1977,
       1978, 1955, 1966, 1988, 1948, 1958, 1972, 1960, 1945, 1991, 1962,
       1953, 1961, 1956, 1992, 1900, 1893, 1990, 1947, 1899, 1993, 1994,
       1941, 1944, 1995, 1940], dtype=int64)
```

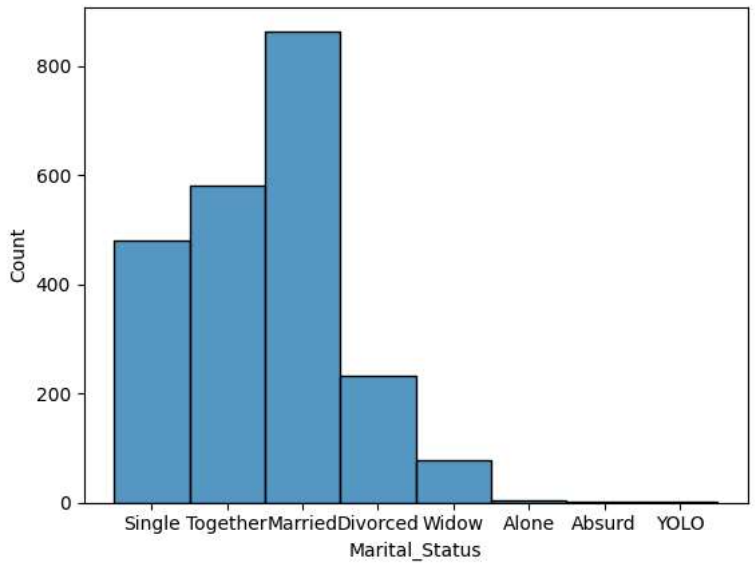
In [21]:
```
sns.histplot(cust_per_df1.Year_Birth)
plt.show()
```



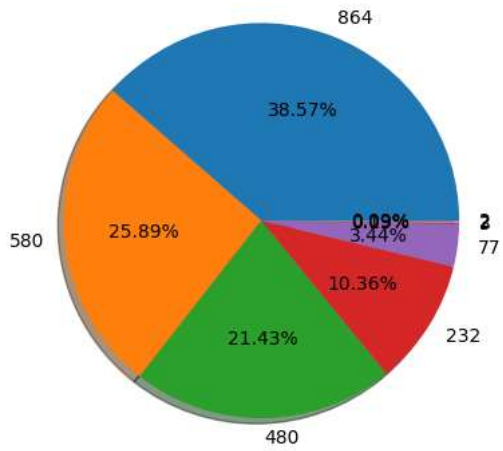
In [22]:
```
sns.scatterplot(cust_per_df1.Year_Birth)
plt.show()
```



In [23]:
```
sns.histplot(cust_per_df1.Marital_Status)
plt.show()
```



In [39]:
```
t.pie(cust_per_df1.Marital_Status.value_counts(),labels=cust_per_df1.Marital_Status.value_counts(),autopct='%1.2f%%',shadow=True)
t.show()
```



In [ ]:
```
# 1
```