

```
import numpy as np
import pandas as pd
import statsmodels.api as sm
import matplotlib.pyplot as plt
import seaborn as sb
from sklearn.cluster import KMeans
from sklearn.preprocessing import StandardScaler
%matplotlib inline
```

```
df = pd.read_csv('/content/ifood_df.csv')
```

```
df.head(50)
```

	Income	Kidhome	Teenhome	Recency	MntWines	MntFruits	MntMeatProducts	MntFishPro
0	58138.0	0	0	58	635	88	546	
1	46344.0	1	1	38	11	1	6	
2	71613.0	0	0	26	426	49	127	
3	26646.0	1	0	26	11	4	20	
4	58293.0	1	0	94	173	43	118	
5	62513.0	0	1	16	520	42	98	
6	55635.0	0	1	34	235	65	164	
7	33454.0	1	0	32	76	10	56	
8	30351.0	1	0	19	14	0	24	
9	5648.0	1	1	68	28	0	6	
10	7500.0	0	0	59	6	16	11	
11	63033.0	0	0	82	194	61	480	
12	59354.0	1	1	53	233	2	53	
13	17323.0	0	0	38	3	14	17	
14	82800.0	0	0	23	1006	22	115	
15	41850.0	1	1	51	53	5	19	
16	37760.0	0	0	20	84	5	38	
17	76995.0	0	1	91	1012	80	498	
18	33812.0	1	0	86	4	17	19	
19	37040.0	0	0	41	86	2	73	
20	2447.0	1	0	42	1	1	1725	
21	58607.0	0	1	63	867	0	86	
22	65324.0	0	1	0	384	0	102	
23	40689.0	0	1	69	270	3	27	
24	18589.0	0	0	89	6	4	25	
25	53359.0	1	1	4	173	4	30	
26	38360.0	1	0	26	36	2	42	
27	84618.0	0	0	96	684	100	801	
28	10979.0	0	0	34	8	4	10	
29	29670.0	0	0	56	112	17	11	

```
df.head()
```

	Income	Kidhome	Teenhome	Recency	MntWines	MntFruits	MntMeatProducts	MntFishProo
0	58138.0	0	0	58	635	88		546
1	46344.0	1	1	38	11	1		6
2	71613.0	0	0	26	426	49		127
3	26646.0	1	0	26	11	4		20
4	58293.0	1	0	94	173	43		118

5 rows × 39 columns

40	21994.0	0	1	4	9	0		6
----	---------	---	---	---	---	---	--	---

```
def basic_info(df):
    print("This dataset has ", df.shape[1], " columns and ", df.shape[0], " rows.")
    print("This dataset has ", df[df.duplicated()].shape[0], " duplicated rows.")
    print(" ")
    print("Descriptive statistics of the numeric features in the dataset: ")
    print(" ")
    print(df.describe())
    print(" ")
    print("Information about this dataset: ")
    print(" ")
    print(df.info())
```

```
basic_info(df)
```

This dataset has 39 columns and 2205 rows.

This dataset has 184 duplicated rows.

Descriptive statistics of the numeric features in the dataset:

	Income	Kidhome	Teenhome	Recency	MntWines \
count	2205.000000	2205.000000	2205.000000	2205.000000	2205.000000
mean	51622.094785	0.442177	0.506576	49.009070	306.164626
std	20713.063826	0.537132	0.544380	28.932111	337.493839
min	1730.000000	0.000000	0.000000	0.000000	0.000000
25%	35196.000000	0.000000	0.000000	24.000000	24.000000
50%	51287.000000	0.000000	0.000000	49.000000	178.000000
75%	68281.000000	1.000000	1.000000	74.000000	507.000000
max	113734.000000	2.000000	2.000000	99.000000	1493.000000

  

	MntFruits	MntMeatProducts	MntFishProducts	MntSweetProducts \
count	2205.000000	2205.000000	2205.000000	2205.000000
mean	26.403175	165.312018	37.756463	27.128345
std	39.784484	217.784507	54.824635	41.130468

min	0.000000	0.000000	0.000000	0.000000
25%	2.000000	16.000000	3.000000	1.000000
50%	8.000000	68.000000	12.000000	8.000000
75%	33.000000	232.000000	50.000000	34.000000
max	199.000000	1725.000000	259.000000	262.000000

	MntGoldProds	...	marital_Together	marital_Widow	education_2n Cycle	\
count	2205.000000	...	2205.000000	2205.000000	2205.000000	
mean	44.057143	...	0.257596	0.034467	0.089796	
std	51.736211	...	0.437410	0.182467	0.285954	
min	0.000000	...	0.000000	0.000000	0.000000	
25%	9.000000	...	0.000000	0.000000	0.000000	
50%	25.000000	...	0.000000	0.000000	0.000000	
75%	56.000000	...	1.000000	0.000000	0.000000	
max	321.000000	...	1.000000	1.000000	1.000000	

	education_Basic	education_Graduation	education_Master	education_PhD	\
count	2205.000000	2205.000000	2205.000000	2205.000000	
mean	0.024490	0.504762	0.165079	0.215873	
std	0.154599	0.500091	0.371336	0.411520	
min	0.000000	0.000000	0.000000	0.000000	
25%	0.000000	0.000000	0.000000	0.000000	
50%	0.000000	1.000000	0.000000	0.000000	
75%	0.000000	1.000000	0.000000	0.000000	
max	1.000000	1.000000	1.000000	1.000000	

	MntTotal	MntRegularProds	AcceptedCmpOverall
count	2205.000000	2205.000000	2205.000000
mean	562.764626	518.707483	0.29932
std	575.936911	553.847248	0.68044
min	4.000000	-283.000000	0.00000
25%	56.000000	42.000000	0.00000
50%	343.000000	288.000000	0.00000
75%	964.000000	884.000000	0.00000
max	2491.000000	2458.000000	4.00000

[8 rows x 39 columns]

Information about this dataset:

standardizing features

```
scaler = StandardScaler()
```

```
scaled_features = scaler.fit_transform(df[['Income', 'Kidhome', 'Teenhome', 'Recentyrscol']])
```

```
kmeans = KMeans(n_clusters=7, random_state=42)
```

```
kmeans.fit(scaled_features)
```

```

/usr/local/lib/python3.10/dist-packages/sklearn/cluster/_kmeans.py:870: FutureWarning: 1
from sklearn.metrics import silhouette_score

n_clusters = list(range(2, 21))

silhouette_scores = []

for n in n_clusters:
    kmeans = KMeans(n_clusters=n, random_state=42)
    cluster_labels = kmeans.fit_predict(scaled_features)
    silhouette_avg = silhouette_score(scaled_features, cluster_labels)
    silhouette_scores.append(silhouette_avg)

```



```

/usr/local/lib/python3.10/dist-packages/sklearn/cluster/_kmeans.py:870: FutureWarning: 1
    warnings.warn(
/usr/local/lib/python3.10/dist-packages/sklearn/cluster/_kmeans.py:870: FutureWarning: 1
    warnings.warn(
/usr/local/lib/python3.10/dist-packages/sklearn/cluster/_kmeans.py:870: FutureWarning: 1
    warnings.warn(
/usr/local/lib/python3.10/dist-packages/sklearn/cluster/_kmeans.py:870: FutureWarning: 1
    warnings.warn(
/usr/local/lib/python3.10/dist-packages/sklearn/cluster/_kmeans.py:870: FutureWarning: 1
    warnings.warn(
/usr/local/lib/python3.10/dist-packages/sklearn/cluster/_kmeans.py:870: FutureWarning: 1
    warnings.warn(
/usr/local/lib/python3.10/dist-packages/sklearn/cluster/_kmeans.py:870: FutureWarning: 1
    warnings.warn(
/usr/local/lib/python3.10/dist-packages/sklearn/cluster/_kmeans.py:870: FutureWarning: 1
    warnings.warn(
/usr/local/lib/python3.10/dist-packages/sklearn/cluster/_kmeans.py:870: FutureWarning: 1
    warnings.warn(
/usr/local/lib/python3.10/dist-packages/sklearn/cluster/_kmeans.py:870: FutureWarning: 1
    warnings.warn(
/usr/local/lib/python3.10/dist-packages/sklearn/cluster/_kmeans.py:870: FutureWarning: 1
    warnings.warn(
/usr/local/lib/python3.10/dist-packages/sklearn/cluster/_kmeans.py:870: FutureWarning: 1
    warnings.warn(
/usr/local/lib/python3.10/dist-packages/sklearn/cluster/_kmeans.py:870: FutureWarning: 1
    warnings.warn(
/usr/local/lib/python3.10/dist-packages/sklearn/cluster/_kmeans.py:870: FutureWarning: 1
    warnings.warn(
/usr/local/lib/python3.10/dist-packages/sklearn/cluster/_kmeans.py:870: FutureWarning: 1
    warnings.warn(
/usr/local/lib/python3.10/dist-packages/sklearn/cluster/_kmeans.py:870: FutureWarning: 1
    warnings.warn(
/usr/local/lib/python3.10/dist-packages/sklearn/cluster/_kmeans.py:870: FutureWarning: 1
    warnings.warn(
/usr/local/lib/python3.10/dist-packages/sklearn/cluster/_kmeans.py:870: FutureWarning: 1
    warnings.warn(
/usr/local/lib/python3.10/dist-packages/sklearn/cluster/_kmeans.py:870: FutureWarning: 1
    warnings.warn(

```

```
plt.figure(figsize=(10, 6))
plt.plot(n_clusters, silhouette_scores, marker='o', linestyle='-', color='b')
plt.title('Silhouette Analysis')
plt.xlabel('Number of Clusters')
plt.ylabel('Silhouette Score')
plt.grid(True)
plt.show()
```



