```
In [1]: import pandas as pd
import numpy as np

In [2]: df = pd.read_csv('sales_data_sample.csv', encoding='unicode_escape')

In [3]: df.head
```

Out[3]:	 bound	d method ND ER SALES		ead of	0	RDERNU	JMBER (OYTITMAUÇ	RDERED) PRIC	CEEACH	ORDERLIN
	0	1010			30	95	.70		2	2871.6	10	
	1	1012			34		. 35			2765.9		
	2	1013			41		.74			3884.3		
	3	1014	! 5		45	83.	. 26		6	3746.7	70	
	4	1015	59		49	100	.00		14	5205.2	27	
			•	•	• • •				• • •	• •	•	
	2818	1035	50		20	100.	.00		15	2244.4	10	
	2819	1037	' 3		29	100	.00		1	3978.5	51	
	2820	1038			43	100				5417.5		
	2821	1039			34	62				2116.1		
	2822				47		.52			3079.4		
	2022	1041	.4		47	05.	. 52		9	30/9.4	14	
		ORDE	RDATE	STATUS	QTR_	ID MO	ONTH_ID	YEAR_ID		\		
	0	2/24/2003	0:00	Shipped	-	1	_ 2	2 <u>0</u> 03				
	1	5/7/2003		Shipped		2	5	2003				
	2	7/1/2003				3	7	2003				
				Shipped								
	3	8/25/2003		Shipped		3	8	2003				
	4	10/10/2003	0:00	Shipped		4	10	2003				
	• • •		• • •		•	• •	• • •	• • •	• • •			
	2818	12/2/2004		Shipped		4	12	2004				
	2819	1/31/2005		Shipped		1	1	2005				
	2820	3/1/2005	0:00	Resolved		1	3	2005				
	2821	3/28/2005	0:00	Shipped		1	3	2005				
	2822	5/6/2005	0:00	On Hold		2	5	2005				
				ADDRESSL1		ADDRES	SSLINE2			STATE		
	0	897	_	rport Ave			NaN		NYC			
	1		59 rue	de l'Abb	paye		NaN		Reims	s NaN	1	
	2	27 rue du	Colonel	Pierre A	Avia		NaN		Paris	. NaN	J	
	3		78934	Hillside	Dr.		NaN	Pa	sadena	a CA	4	
	4			4 Strong			NaN					
	• • •		,,,	1 3010116			• • •	Sun Tru	•••			
	2818		C/ Mon	alzarzal			NaN	1	Madrid			
	2819		C/ MOI	Torikatı				'	Oulu			
			C / M				NaN	Í				
	2820	_		alzarzal,			NaN		Madrid			
	2821	1		ace-Lorra			NaN		ulouse			
	2822		8616 S	pinnaker	Dr.		NaN		Boston	n MA	4	
	1	POSTALCODE	COLINTA	V TEDDITA	NDV CO	NITACTI	ΛςΤΝΛΜ	E CONTACT	ETDCTN	IVWE DE	. A I C T 7 E	:
		10022	US		NaN	MIACIL	ASTNAM Y			wai	Small	
	0											
	1	51100	Franc		ΛΕΑ	_	Henrio			aul	Small	
	2	75508	Franc		1EA	[Da Cunh			niel	Medium	
	3	90003	US	A 1	NaN		Youn	g	Ju	ılie	Medium	1
	4	NaN	US	A 1	NaN		Brow	n	Ju	ılie	Medium	1
	2010	20024					· ·			• • •	· · ·	
	2818	28034	Spai		1EA		Freyr			ego	Small	
	2819	90110	Finlan		1EA	Ko	oskital			rkko	Medium	
	2820	28034	Spai		1EA		Freyr			.ego	Medium	
	2821	31000	Franc		1EA		Roule		Anne		Small	
	2822	51003	US	A 1	NaN		Yoshid	0	J	luri	Medium	1
			_									
	F 2 0 2 2		7	1.5								

[2823 rows x 25 columns]>

```
<bound method DataFrame.info of</pre>
                                                  ORDERNUMBER QUANTITYORDERED PRICEEACH ORDERL
Out[4]:
         INENUMBER
                       SALES \
         0
                      10107
                                            30
                                                    95.70
                                                                           2
                                                                               2871.00
         1
                      10121
                                            34
                                                    81.35
                                                                           5
                                                                               2765.90
         2
                                            41
                                                    94.74
                                                                           2
                      10134
                                                                               3884.34
         3
                      10145
                                            45
                                                    83.26
                                                                              3746.70
                                                                           6
         4
                      10159
                                            49
                                                   100.00
                                                                          14
                                                                               5205.27
                                           . . .
                                                                          . . .
         . . .
         2818
                      10350
                                            20
                                                   100.00
                                                                          15
                                                                               2244.40
                                            29
         2819
                                                                           1
                                                                               3978.51
                      10373
                                                   100.00
         2820
                      10386
                                            43
                                                   100.00
                                                                           4
                                                                               5417.57
         2821
                      10397
                                            34
                                                    62.24
                                                                           1
                                                                               2116.16
                                            47
                                                                               3079.44
         2822
                      10414
                                                    65.52
                      ORDERDATE
                                    STATUS QTR ID
                                                     MONTH ID
                                                                YEAR ID
         0
                2/24/2003 0:00
                                   Shipped
                                                  1
                                                             2
                                                                    2003
                                                  2
         1
                  5/7/2003 0:00
                                   Shipped
                                                             5
                                                                    2003
                                                                          . . .
         2
                                                  3
                                                             7
                                                                    2003
                 7/1/2003 0:00
                                   Shipped
         3
                                                  3
                8/25/2003 0:00
                                   Shipped
                                                             8
                                                                    2003
         4
                                   Shipped
                                                  4
               10/10/2003 0:00
                                                            10
                                                                    2003
                                       . . .
                                                           . . .
                                                                     . . .
                                                . . .
         2818
                12/2/2004 0:00
                                   Shipped
                                                  4
                                                            12
                                                                    2004
         2819
                1/31/2005 0:00
                                   Shipped
                                                  1
                                                             1
                                                                    2005
         2820
                 3/1/2005 0:00
                                  Resolved
                                                  1
                                                             3
                                                                    2005
                                                             3
         2821
                3/28/2005 0:00
                                   Shipped
                                                  1
                                                                    2005
                                                                    2005
         2822
                                                  2
                                                             5
                 5/6/2005 0:00
                                   On Hold
                                  ADDRESSLINE1 ADDRESSLINE2
                                                                          CITY STATE
         0
                      897 Long Airport Avenue
                                                                           NYC
                                                                                   NY
                                                           NaN
         1
                            59 rue de l'Abbaye
                                                           NaN
                                                                         Reims
                                                                                  NaN
         2
               27 rue du Colonel Pierre Avia
                                                           NaN
                                                                         Paris
                                                                                  NaN
         3
                           78934 Hillside Dr.
                                                           NaN
                                                                      Pasadena
                                                                                   CA
         4
                               7734 Strong St.
                                                           NaN
                                                               San Francisco
                                                                                   CA
         . . .
                                                           . . .
                           C/ Moralzarzal, 86
         2818
                                                           NaN
                                                                        Madrid
                                                                                  NaN
                                   Torikatu 38
         2819
                                                           NaN
                                                                          Oulu
                                                                                  NaN
         2820
                           C/ Moralzarzal, 86
                                                                        Madrid
                                                                                  NaN
                                                           NaN
         2821
                        1 rue Alsace-Lorraine
                                                           NaN
                                                                      Toulouse
                                                                                  NaN
         2822
                           8616 Spinnaker Dr.
                                                           NaN
                                                                        Boston
                                                                                   MΑ
                           COUNTRY TERRITORY CONTACTLASTNAME CONTACTFIRSTNAME DEALSIZE
              POSTALCODE
         0
                    10022
                                USA
                                           NaN
                                                             Yu
                                                                                      Small
                                                                              Kwai
         1
                    51100
                            France
                                          EMEA
                                                       Henriot
                                                                             Paul
                                                                                      Small
         2
                    75508
                            France
                                          EMEA
                                                       Da Cunha
                                                                           Daniel
                                                                                     Medium
         3
                    90003
                                USA
                                          NaN
                                                                             Julie
                                                                                     Medium
                                                          Young
         4
                      NaN
                                USA
                                           NaN
                                                          Brown
                                                                             Julie
                                                                                     Medium
                                . . .
                      . . .
                                           . . .
                                                            . . .
                                                                               . . .
         . . .
         2818
                    28034
                              Spain
                                          EMEA
                                                         Freyre
                                                                            Diego
                                                                                      Small
         2819
                    90110
                           Finland
                                          EMEA
                                                      Koskitalo
                                                                           Pirkko
                                                                                     Medium
         2820
                                                                                     Medium
                    28034
                                          EMEA
                                                         Freyre
                                                                            Diego
                              Spain
         2821
                    31000
                            France
                                          EMEA
                                                         Roulet
                                                                          Annette
                                                                                      Small
                    51003
                                                                                     Medium
         2822
                                USA
                                          NaN
                                                        Yoshido
                                                                              Juri
         [2823 rows x 25 columns]>
         to_drop = ['ADDRESSLINE1', 'ADDRESSLINE2', 'STATE', 'POSTALCODE', 'PHONE']
In [5]:
         df = df.drop(to drop, axis=1)
         df.isnull().sum()
```

```
ORDERNUMBER
                                  0
Out[6]:
         QUANTITYORDERED
                                  0
         PRICEEACH
                                  0
         ORDERLINENUMBER
                                  0
         SALES
                                  0
         ORDERDATE
                                  0
         STATUS
                                  0
         QTR_ID
                                  0
         MONTH ID
                                  0
         YEAR_ID
                                  0
         PRODUCTLINE
                                  0
                                  0
         MSRP
         PRODUCTCODE
                                 0
         CUSTOMERNAME
                                  0
         CITY
                                 0
         COUNTRY
                                 0
         TERRITORY
                              1074
         CONTACTLASTNAME
                                 0
                                 0
         CONTACTFIRSTNAME
         DEALSIZE
                                  0
         dtype: int64
In [7]: df.dtypes
         ORDERNUMBER
                                int64
Out[7]:
         QUANTITYORDERED
                                int64
         PRICEEACH
                              float64
         ORDERLINENUMBER
                                int64
         SALES
                              float64
         ORDERDATE
                               object
         STATUS
                               object
         QTR ID
                                int64
         MONTH_ID
                                int64
         YEAR ID
                                int64
         PRODUCTLINE
                               object
         MSRP
                                int64
         PRODUCTCODE
                               object
         CUSTOMERNAME
                               object
         CITY
                               object
         COUNTRY
                               object
         TERRITORY
                               object
         CONTACTLASTNAME
                               object
         CONTACTFIRSTNAME
                               object
         DEALSIZE
                               object
         dtype: object
In [8]:
         df['ORDERDATE'] = pd.to datetime(df['ORDERDATE'])
In [13]:
         import datetime as dt
          snapshot date = df['ORDERDATE'].max() + dt.timedelta(days = 1)
         df_RFM = df.groupby(['CUSTOMERNAME']).agg({
              'ORDERDATE' : lambda x : (snapshot_date - x.max()).days,
              'ORDERNUMBER' : 'count',
              'SALES' : 'sum'
         })
          df_RFM.rename(columns = {
              'ORDERDATE' : 'Recency',
              'ORDERNUMBER' : 'Frequency',
```

```
'SALES' : 'MonetaryValue'
          }, inplace=True)
In [14]: | df_RFM.head()
                                Recency Frequency Monetary Value
Out[14]:
               CUSTOMERNAME
                  AV Stores, Co.
                                    196
                                               51
                                                        157807.81
                  Alpha Cognac
                                    65
                                               20
                                                         70488.44
             Amica Models & Co.
                                    265
                                               26
                                                         94117.26
          Anna's Decorations, Ltd
                                    84
                                               46
                                                        153996.13
                                    188
                                                7
                                                         24179.96
               Atelier graphique
In [16]: | df_RFM['M'] = pd.qcut(df_RFM['MonetaryValue'], q = 4, labels = range(1,5))
          df_RFM['R'] = pd.qcut(df_RFM['Recency'], q = 4, labels = list(range(4,0,-1)))
          df_RFM['F'] = pd.qcut(df_RFM['Frequency'], q = 4, labels = range(1,5))
          df_RFM.head()
Out[16]:
                                Recency Frequency Monetary Value M R F
               CUSTOMERNAME
                  AV Stores, Co.
                                    196
                                               51
                                                        157807.81 4 2 4
                                                                  2 4 2
                  Alpha Cognac
                                     65
                                               20
                                                         70488.44
             Amica Models & Co.
                                    265
                                                         94117.26
                                                                 3 1 2
                                               26
          Anna's Decorations, Ltd
                                     84
                                               46
                                                        153996.13
                                                                  4 3 4
                                    188
                                                7
                                                         24179.96
                                                                 1 2 1
               Atelier graphique
          df_RFM['RFM_Score'] = df_RFM[['R', 'M', 'F']].sum(axis=1)
In [17]:
          df_RFM.head()
Out[17]:
                                Recency Frequency Monetary Value M R F RFM_Score
               CUSTOMERNAME
                  AV Stores, Co.
                                    196
                                               51
                                                        157807.81 4 2 4
                                                                                  10
                                                                  2 4 2
                  Alpha Cognac
                                     65
                                               20
                                                         70488.44
                                                                                   8
             Amica Models & Co.
                                                         94117.26
                                                                  3 1 2
                                                                                   6
                                    265
                                               26
                                                                  4 3 4
          Anna's Decorations, Ltd
                                    84
                                               46
                                                        153996.13
                                                                                  11
                                                7
                                                                                   4
               Atelier graphique
                                    188
                                                         24179.96
                                                                 1 2 1
          def rfm_level(df):
In [20]:
              if bool(df['RFM_Score'] >= 10):
                   return 'High Value Customer'
              elif bool(df['RFM_Score'] < 10) and bool(df['RFM_Score'] >= 6):
```

```
return 'Mid Value Customer'
else:
    return 'Low Value Customer'

df_RFM['RFM_Level'] = df_RFM.apply(rfm_level, axis = 1)

df_RFM.head()
```

Out[20]:		Recency	Frequency	MonetaryValue	М	R	F	RFM_Score	RFM_Level
	CUSTOMERNAME								
	AV Stores, Co.	196	51	157807.81	4	2	4	10	High Value Customer
	Alpha Cognac	65	20	70488.44	2	4	2	8	Mid Value Customer
	Amica Models & Co.	265	26	94117.26	3	1	2	6	Mid Value Customer
	Anna's Decorations, Ltd	84	46	153996.13	4	3	4	11	High Value Customer
	Atelier graphique	188	7	24179.96	1	2	1	4	Low Value Customer
In [21]:	<pre>data = df_RFM[['Recency', 'Frequency', 'MonetaryValue']] data.head()</pre>								

Out[21]: Recency Frequency MonetaryValue

CUSTOMERNAME

AV Stores, Co.	196	51	157807.81
Alpha Cognac	65	20	70488.44
Amica Models & Co.	265	26	94117.26
Anna's Decorations, Ltd	84	46	153996.13
Atelier graphique	188	7	24179.96

In [22]: data_log = np.log(data)
 data_log.head()

Out[22]: Recency Frequency MonetaryValue

CUSTOMERNAME

AV Stores, Co.	5.278115	3.931826	11.969133
Alpha Cognac	4.174387	2.995732	11.163204
Amica Models & Co.	5.579730	3.258097	11.452297
Anna's Decorations, Ltd	4.430817	3.828641	11.944683
Atelier graphique	5.236442	1.945910	10.093279

```
In [25]: from sklearn.preprocessing import StandardScaler
    scaler = StandardScaler()
```

```
scaler.fit(data_log)
data_normalized = scaler.transform(data_log)
data_normalized = pd.DataFrame(data_normalized, index = data_log.index, columns=data_l
data_normalized.describe().round(2)
```

Out[25]: Recency Frequency MonetaryValue

count	92.00	92.00	92.00
mean	0.00	-0.00	0.00
std	1.01	1.01	1.01
min	-3.51	-3.67	-3.82
25%	-0.24	-0.41	-0.39
50%	0.37	0.06	-0.04
75%	0.53	0.45	0.52
max	1.12	4.03	3.92

```
import matplotlib.pyplot as plt
import seaborn as sns
from sklearn.cluster import KMeans

sse = {}

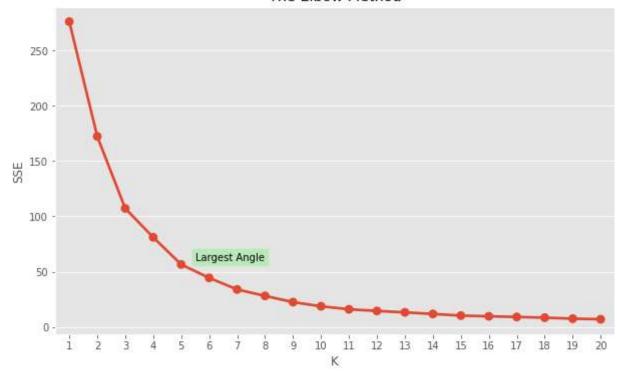
for k in range(1, 21):
    kmeans = KMeans(n_clusters = k, random_state = 1)
    kmeans.fit(data_normalized)
    sse[k] = kmeans.inertia_
```

```
In [31]: plt.figure(figsize=(10,6))
    plt.title('The Elbow Method')

plt.xlabel('K')
    plt.ylabel('SSE')
    plt.style.use('ggplot')

sns.pointplot(x=list(sse.keys()), y = list(sse.values()))
    plt.text(4.5, 60, "Largest Angle", bbox = dict(facecolor = 'lightgreen', alpha = 0.5))
    plt.show()
```

The Elbow Method



```
In [32]: kmeans = KMeans(n_clusters=5, random_state=1)
    kmeans.fit(data_normalized)
    cluster_labels = kmeans.labels_

data_rfm = data.assign(Cluster = cluster_labels)
    data_rfm.head()
```

Out[32]:

Recency Frequency Monetary Value Cluster

CUSTOMERNAME

AV Stores, Co.	196	51	157807.81	3
Alpha Cognac	65	20	70488.44	0
Amica Models & Co.	265	26	94117.26	0
Anna's Decorations, Ltd	84	46	153996.13	3
Atelier graphique	188	7	24179.96	2