```
!ls '/content/Mall_Customers.csv'
```

```
/content/Mall_Customers.csv
```

Installing the Libraries

```
import numpy as np
import pandas as pd
import matplotlib.pyplot as plt
import seaborn as sns
```

Importing the Dataset

```
%time data = pd.read_csv('/content/Mall_Customers.csv')
print(data.shape)
```

```
CPU times: user 5.52 ms, sys: 55 µs, total: 5.58 ms
Wall time: 5.48 ms
(200, 5)
```

```
data.head()
```

	CustomerID	Genre	Age	Annual Income (k$)	Spending Score (1-100)
0	1	Male	19	15	39
1	2	Male	21	15	81
2	3	Female	20	16	6
3	4	Female	23	16	77
4	5	Female	31	17	40

Saving...

```
data.tail()
```

	CustomerID	Genre	Age	Annual Income (k$)	Spending Score (1-100)
195	196	Female	35	120	79
196	197	Female	45	126	28
197	198	Male	32	126	74
198	199	Male	32	137	18
199	200	Male	30	137	83

```
data.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 200 entries, 0 to 199
Data columns (total 5 columns):
 #   Column                  Non-Null Count  Dtype
---  ------                  --------------  -----
 0   CustomerID              200 non-null    int64
 1   Genre                   200 non-null    object
 2   Age                     200 non-null    int64
 3   Annual Income (k$)      200 non-null    int64
 4   Spending Score (1-100)  200 non-null    int64
dtypes: int64(4), object(1)
memory usage: 7.9+ KB
```

```
data.describe()
```

	CustomerID	Age	Annual Income (k$)	Spending Score (1-100)
count	200.000000	200.000000	200.000000	200.000000
mean	100.500000	38.850000	60.560000	50.200000
std	57.879185	13.969007	26.264721	25.823522
min	1.000000	18.000000	15.000000	1.000000
25%	50.750000	28.750000	41.500000	34.750000
50%	100.500000	36.000000	61.500000	50.000000
75%	150.250000	49.000000	78.000000	73.000000
max	200.000000	70.000000	137.000000	99.000000

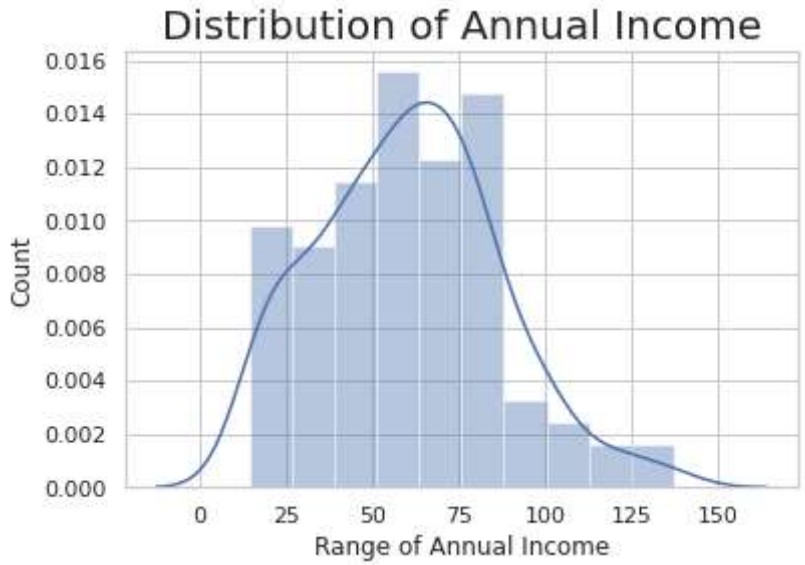
```
data.isnull().any()
```



```
Age                     False
Annual Income (k$)      False
Spending Score (1-100)  False
dtype: bool
```

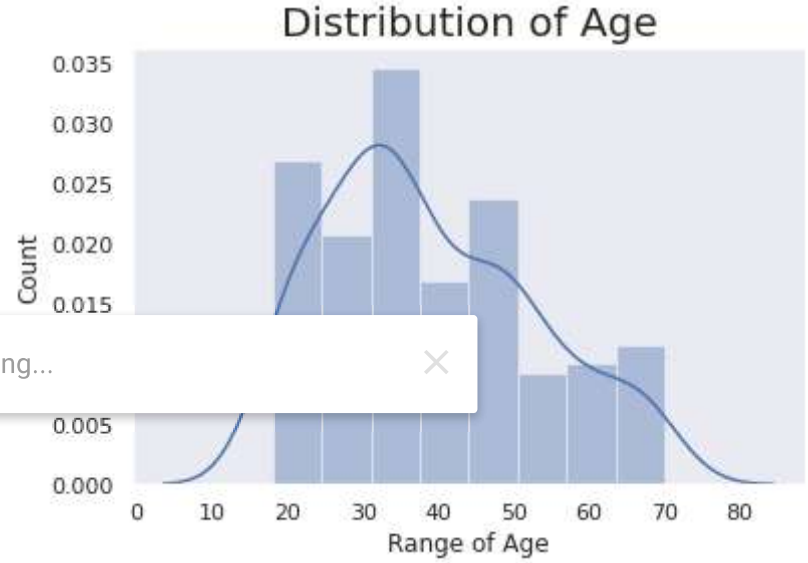
```
sns.set(style = 'whitegrid')
sns.distplot(data['Annual Income (k$)'])
plt.title('Distribution of Annual Income', fontsize = 20)
plt.xlabel('Range of Annual Income')
plt.ylabel('Count')
plt.show()
```

```
/usr/local/lib/python3.7/dist-packages/seaborn/distributions.py:2619: FutureWarning: `di
  warnings.warn(msg, FutureWarning)
```




```
sns.set(style = 'dark')
sns.distplot(data['Age'])
plt.title('Distribution of Age', fontsize = 20)
plt.xlabel('Range of Age')
plt.ylabel('Count')
plt.show()
```

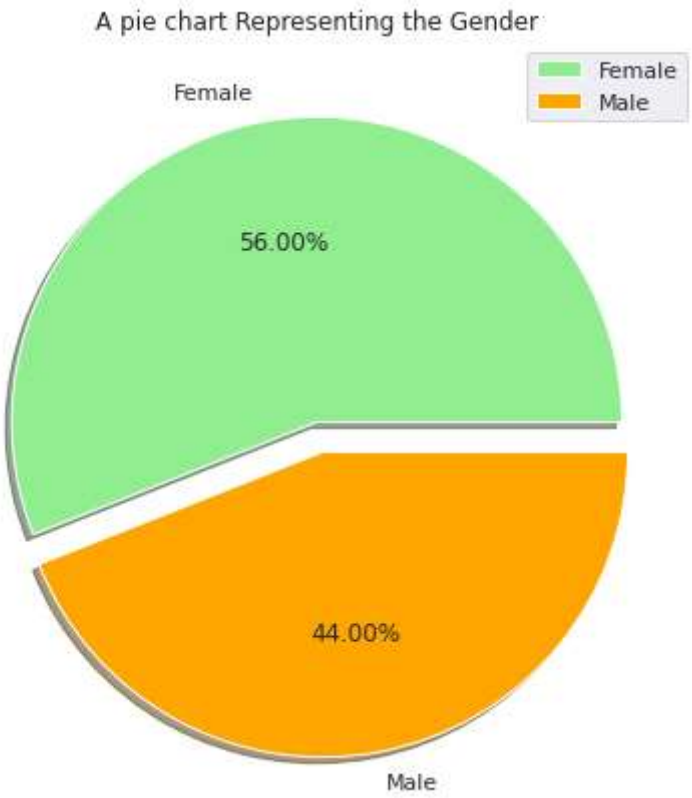
```
/usr/local/lib/python3.7/dist-packages/seaborn/distributions.py:2619: FutureWarning: `di
  warnings.warn(msg, FutureWarning)
```




```
data['Genre'].value_counts()
```

```
Female    112
Male       88
Name: Genre, dtype: int64
```

```
labels = ['Female', 'Male']
size = [112, 88]
colors = ['lightgreen', 'orange']
explode = [0, 0.1]
plt.rcParams['figure.figsize'] = (7, 7)
plt.pie(size, colors = colors, explode = explode, labels = labels, shadow = True, autopct = '
plt.title('A pie chart Representing the Gender')
plt.axis('off')
plt.legend()
plt.show()
```

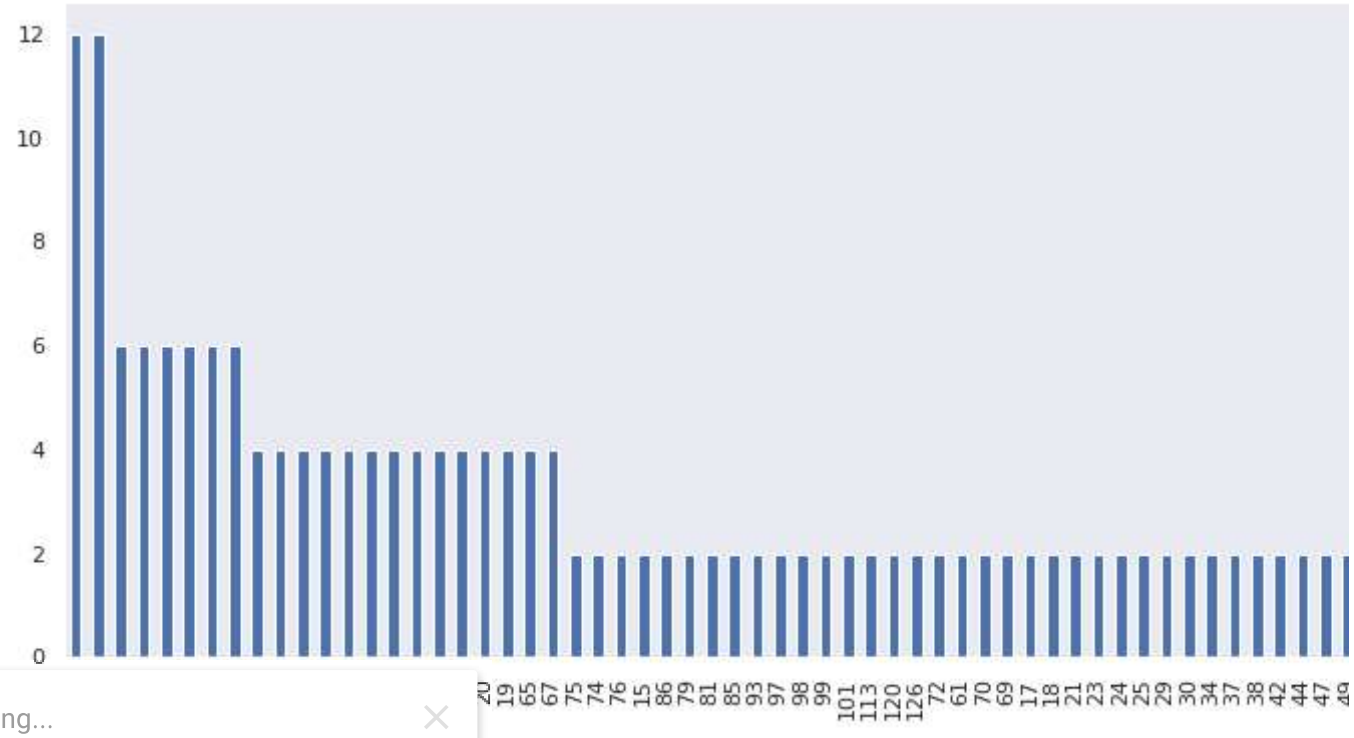



```
data['Age'].value_counts().plot.bar(figsize = (9, 9))
```

```
<matplotlib.axes._subplots.AxesSubplot at 0x7fcd37f51cd0>
```

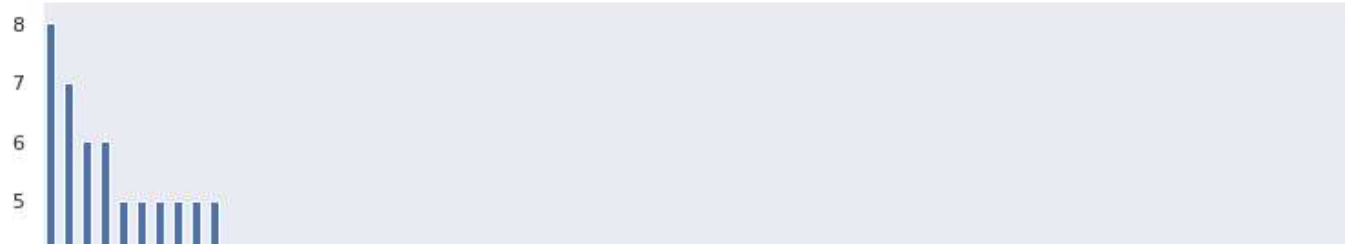

```
data['Annual Income (k$)'].value_counts().plot.bar(figsize = (13, 6))
```

```
<matplotlib.axes._subplots.AxesSubplot at 0x7fcd37d7d850>
```



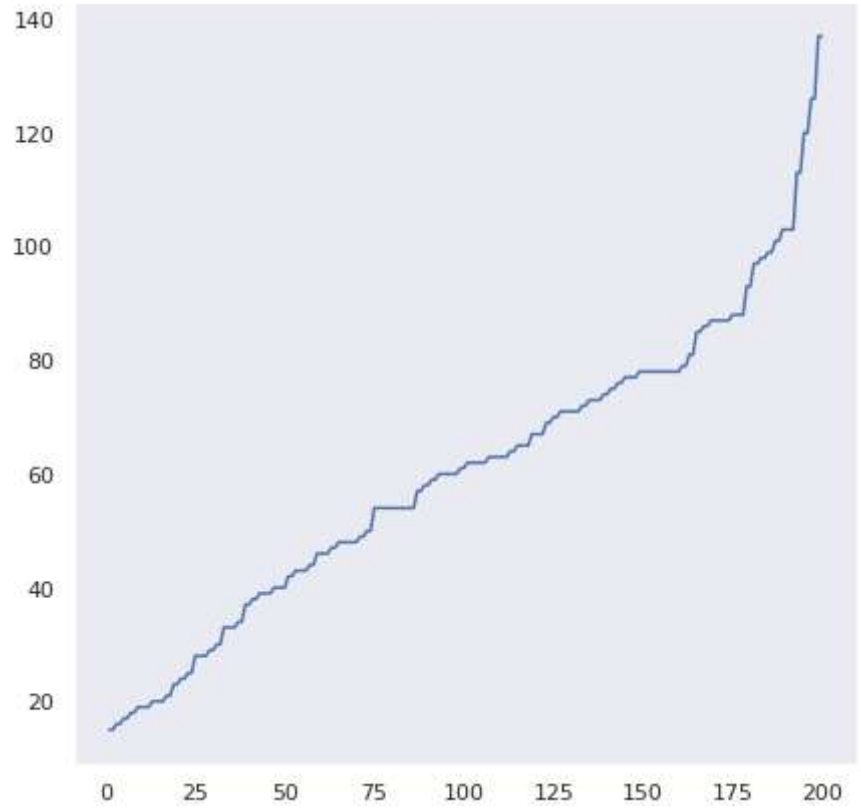

```
data['Spending Score (1-100)'].value_counts().plot.bar(figsize = (15, 5))
```

```
<matplotlib.axes._subplots.AxesSubplot at 0x7fcd37eab890>
```




```
x = data['CustomerID']
y = data['Annual Income (k$)']
plt.plot(x, y)
```

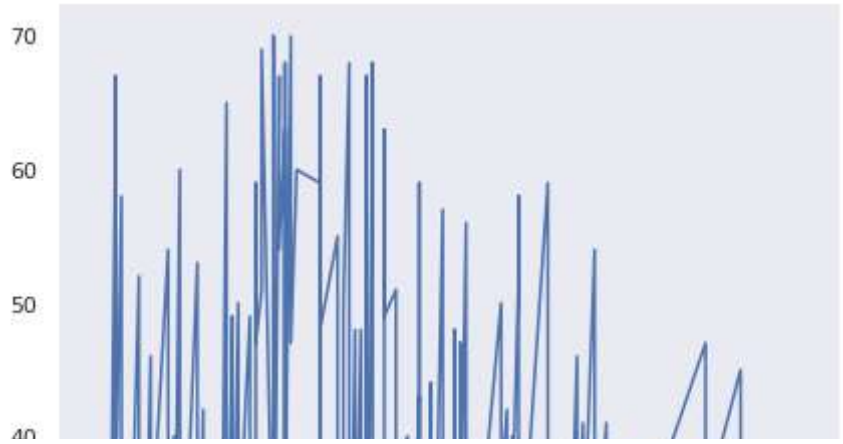
```
[<matplotlib.lines.Line2D at 0x7fcd32310150>]
```



Saving...

```
plt.plot(x, y)
```

```
[<matplotlib.lines.Line2D at 0x7fcd32272b90>]
```




```
x = data.iloc[:, [3, 4]].values
print(x.shape)
```

```
(200, 2)
```