```python
import pandas as pd
import numpy as np
import matplotlib.pyplot as plt


data = pd.read_csv('/content/drive/MyDrive/DataSet/Mall_Customers.csv')


df = data.copy()


df.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

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

```python
data.dtypes
```

```
CustomerID                int64
Genre                    object
Age                       int64
Annual Income (k$)        int64
Spending Score (1-100)    int64
dtype: object
```

```python
data.isnull().sum()
```

```
CustomerID                0
Genre                     0
```

```
Age                     0
Annual Income (k$)      0
Spending Score (1-100)  0
dtype: int64
```

```python
X = df.iloc[:, [3, 4]].values
```

```python
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

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


%matplotlib inline
sns.set_style('whitegrid')


plt.figure(figsize=(12,9))
sns.heatmap(data.corr(), annot=True)
```
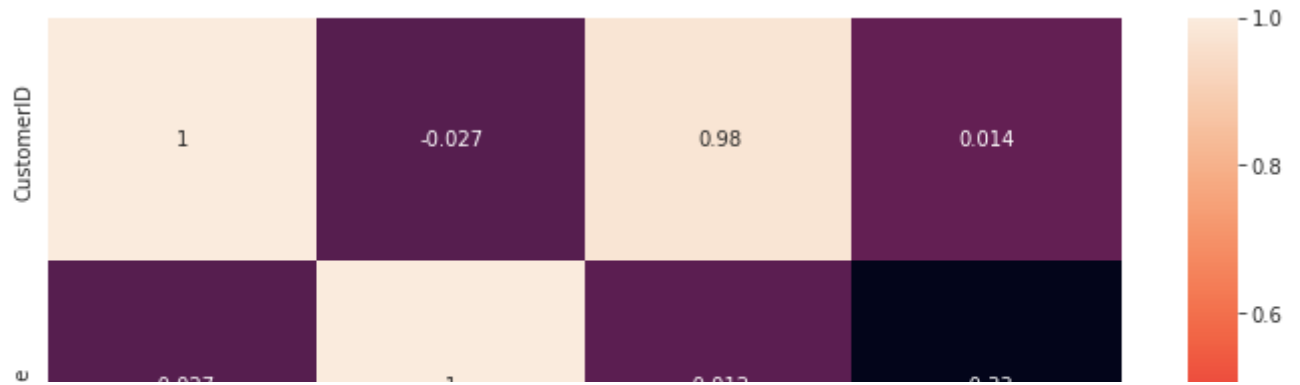
```
<matplotlib.axes._subplots.AxesSubplot at 0x7fdb96fec110>
```



```
import pandas as pd
data = {'Name': ['Jai', 'Princi', 'Gaurav',
               'Anuj', 'Ravi', 'Natasha', 'Riya'],
      'Age': [17, 17, 18, 17, 18, 17, 17],
      'Gender': ['M', 'F', 'M', 'M', 'M', 'F', 'F'],
      'Marks': [90, 76, 'NaN', 74, 65, 'NaN', 71]}
df = pd.DataFrame(data)
df
```

	Name	Age	Gender	Marks
0	Jai	17	M	90
1	Princi	17	F	76
2	Gaurav	18	M	NaN
3	Anuj	17	M	74
4	Ravi	18	M	65
5	Natasha	17	F	NaN
6	Riya	17	F	71

```
df['Gender'] = df['Gender'].map({'M': 0,
                                'F': 1, }).astype(float)
df
```

	Name	Age	Gender	Marks
0	Jai	17	0.0	90
1	Princi	17	1.0	76
2	Gaurav	18	0.0	NaN
3	Anuj	17	0.0	74
4	Ravi	18	0.0	65
5	Natasha	17	1.0	NaN
6	Riya	17	1.0	71

```
import pandas as pd
details = pd.DataFrame({
    'ID': [101, 102, 103, 104, 105, 106,
        107, 108, 109, 110],
    'NAME': ['Jagroop', 'Praveen', 'Harjot',
            'Pooja', 'Rahul', 'Nikita',
            'Saurabh', 'Ayush', 'Dolly', "Mohit"],
    'BRANCH': ['CSE', 'CSE', 'CSE', 'CSE', 'CSE',
            'CSE', 'CSE', 'CSE', 'CSE', 'CSE']})
print(details)
```

```
     ID     NAME BRANCH
0   101  Jagroop    CSE
1   102  Praveen    CSE
2   103   Harjot    CSE
3   104    Pooja    CSE
4   105    Rahul    CSE
5   106   Nikita    CSE
6   107  Saurabh    CSE
7   108    Ayush    CSE
8   109    Dolly    CSE
9   110    Mohit    CSE
```

```
import pandas as pd
fees_status = pd.DataFrame(
    {'ID': [101, 102, 103, 104, 105,
            106, 107, 108, 109, 110],
    'PENDING': ['5000', '250', 'NIL',
                '9000', '15000', 'NIL',
                '4500', '1800', '250', 'NIL']})
print(fees_status)
```

```
     ID PENDING
0   101    5000
1   102     250
2   103     NIL
3   104    9000
4   105   15000
5   106     NIL
6   107    4500
7   108    1800
8   109     250
9   110     NIL
```

DETAILS STUDENTS DATA WHO WANT TO PARTICIPATE IN THE EVENT*

```
import pandas as pd
student_data = {'Name': ['Amit', 'Praveen', 'Jagroop',
```

```
                'Rahul', 'Vishal', 'Suraj',
                'Rishab', 'Satyapal', 'Amit',
                'Rahul', 'Praveen', 'Amit'],

        'Roll_no': [23, 54, 29, 36, 59, 38,
                12, 45, 34, 36, 54, 23],

        'Email': ['xxxx@gmail.com', 'xxxxxx@gmail.com',
                'xxxxxx@gmail.com', 'xx@gmail.com',
                'xxxx@gmail.com', 'xxxxx@gmail.com',
                'xxxxx@gmail.com', 'xxxxx@gmail.com',
                'xxxxx@gmail.com', 'xxxxxx@gmail.com',
                'xxxxxxxxx@gmail.com', 'xxxxxxxxx@gmail.com']}
df = pd.DataFrame(student_data)
print(df)
```

	Name	Roll_no	Email
0	Amit	23	xxxx@gmail.com
1	Praveen	54	xxxxxx@gmail.com
2	Jagroop	29	xxxxxx@gmail.com
3	Rahul	36	xx@gmail.com
4	Vishal	59	xxxx@gmail.com
5	Suraj	38	xxxxx@gmail.com
6	Rishab	12	xxxxx@gmail.com
7	Satyapal	45	xxxxx@gmail.com
8	Amit	34	xxxxx@gmail.com
9	Rahul	36	xxxxxx@gmail.com
10	Praveen	54	xxxxxxxxx@gmail.com
11	Amit	23	xxxxxxxxx@gmail.com

```
import pandas as pd
student_data = {'Name': ['Amit', 'Praveen', 'Jagroop',
                'Rahul', 'Vishal', 'Suraj',
                'Rishab', 'Satyapal', 'Amit',
                'Rahul', 'Praveen', 'Amit'],

        'Roll_no': [23, 54, 29, 36, 59, 38,
                12, 45, 34, 36, 54, 23],
        'Email': ['xxxx@gmail.com', 'xxxxxx@gmail.com',
                'xxxxxx@gmail.com', 'xx@gmail.com',
                'xxxx@gmail.com', 'xxxxx@gmail.com',
                'xxxxx@gmail.com', 'xxxxx@gmail.com',
                'xxxxx@gmail.com', 'xxxxxx@gmail.com',
                'xxxxxxxxx@gmail.com', 'xxxxxxxxx@gmail.com']}
df = pd.DataFrame(student_data)
non_duplicate = df[~df.duplicated('Roll_no')]
print(non_duplicate)
```

	Name	Roll_no	Email
0	Amit	23	xxxx@gmail.com
1	Praveen	54	xxxxxx@gmail.com
2	Jagroop	29	xxxxxx@gmail.com
3	Rahul	36	xx@gmail.com
4	Vishal	59	xxxx@gmail.com
5	Suraj	38	xxxxx@gmail.com

```
6     Rishab    12    xxxxx@gmail.com
7   Satyapal    45    xxxxx@gmail.com
8       Amit    34    xxxxx@gmail.com
```

✓ 0s   completed at 4:43 PM   ● ✕