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

In [2]:
```
import matplotlib.pyplot as plt
```

In [3]:
```
import seaborn as sns
```

In [4]:
```
from sklearn.preprocessing import LabelEncoder
```

In [6]:
```
from sklearn.cluster import KMeans
```

In [8]:
```
df = pd.read_csv("C:\\Users\\dell\\mall customer.csv")
```

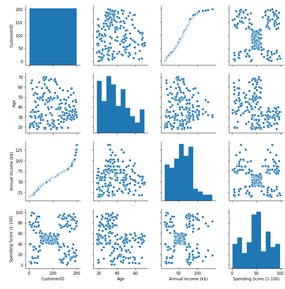
In [9]:
```
df.head()
```

Out[9]:

	CustomerID	Gender	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

In [10]:
```
sns.pairplot(df)
```

Out[10]:
```
<seaborn.axisgrid.PairGrid at 0x1d8f67e9f48>
```

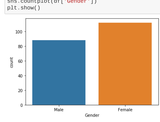


In [11]:
```
df.info()
```

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

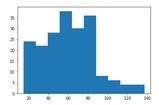
In [13]:
```
import warnings
```

In [14]:
```
warnings.filterwarnings('ignore')
sns.countplot(df['Gender'])
plt.show()
```

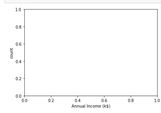


In [15]:
```
plt.hist(df['Annual Income (k$)'])
```

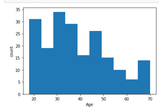
Out[15]:
```
(array([24., 22., 28., 38., 30., 36.,  8.,  6.,  4.,  4.]),
 array([ 15.  ,  27.2 ,  39.4 ,  51.6 ,  63.8 ,  76.  ,  88.2 , 100.4 , 112.6 ,
        124.8 , 137.  ]),
 <a list of 10 Patch objects>)
```



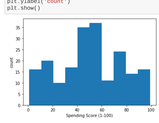
In [17]:
```
plt.xlabel('Annual Income (k$)')
plt.ylabel('count')
plt.show()
```



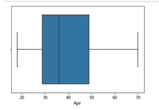
In [18]:
```
plt.hist(df['Age'])
plt.xlabel('Age')
plt.ylabel('count')
plt.show()
```



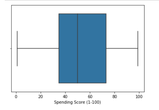
In [24]:
```
plt.hist(df['Spending Score (1-100)'])
plt.xlabel('Spending Score (1-100)')
plt.ylabel('count')
plt.show()
```



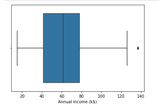
In [25]:
```
sns.boxplot(df['Age'])
plt.show()
```



In [26]:
```
sns.boxplot(df['Spending Score (1-100)'])
plt.show()
```



In [27]:
```
sns.boxplot(df['Annual Income (k$)'])
plt.show()
```

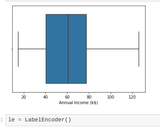


In [28]:
```
df1 = df.copy()
```

In [29]:
```
indices= df1[df1['Annual Income (k$)']>= df1['Annual Income (k$)'].max()].index
```

In [30]:
```
df1.drop(indices, inplace=True)
```

In [31]:
```
sns.boxplot(df1['Annual Income (k$)'])
plt.show()
```



In [32]:
```
le = LabelEncoder()
```

In [33]:
```
df1['Gender']=le.fit_transform(df1['Gender'])
```

In [35]:
```
df1.head()
```

Out[35]:

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

km_new = KMeans(n_clusters=5)

In [40]:
```
km_new.fit(df1.drop('CustomerID',axis=1))
```

Out[40]:
```
KMeans(algorithm='auto', copy_x=True, init='k-means++', max_iter=300,
       n_clusters=5, n_init=10, n_jobs=None, precompute_distances='auto',
       random_state=None, tol=0.0001, verbose=0)
```

In [41]:
```
segment = km_new.predict(df1.drop('CustomerID',axis=1))
```

In [42]:
```
df1['segment']=segment
```

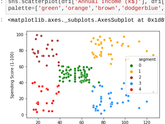
In [43]:
```
df1.head()
```

Out[43]:

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

In [44]:
```
sns.scatterplot(df1['Annual Income (k$)'], df1['Spending Score (1-100)'],hue=df1['segment'],
palette=['green','orange','brown','dodgerblue','red'])
```

Out[44]:
```
<matplotlib.axes._subplots.AxesSubplot at 0x1d8f772e608>
```



In [ ]:
```
```