In [1]:
```
import pandas as pd
from sklearn.cluster import KMeans as KM
import matplotlib.pyplot as plt
import plotly.express as px
from sklearn.preprocessing import StandardScaler as SS
```

In [2]:
```
df=pd.read_csv(r'D:\Mall_Customers.csv')
```

In [3]:
```
df
```

Out[3]:

	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
...	...	...	...	...	...
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

200 rows × 5 columns

In [4]:
```
df.describe()
```

Out[4]:

	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

In [5]:
```
df.isnull().sum()
```

Out[5]:
```
CustomerID                0
Gender                    0
Age                       0
Annual Income (k$)        0
Spending Score (1-100)    0
dtype: int64
```

In [6]:
```
fea=['Age','Annual Income (k$)','Spending Score (1-100)']
```

In [7]:
```
sc=SS()
sc_df=sc.fit_transform(df[fea])
```

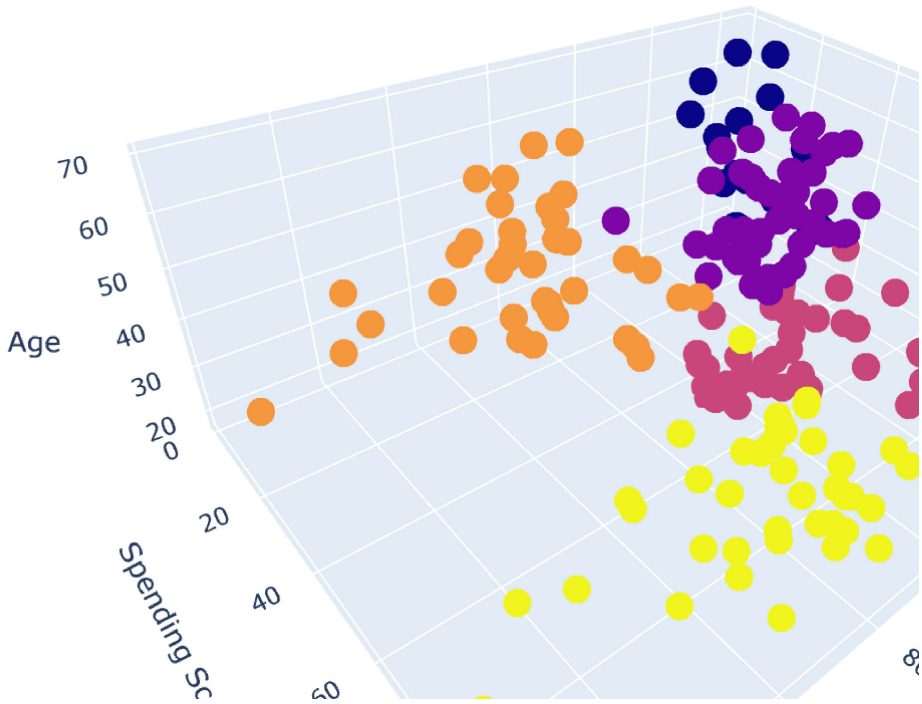
In [8]:
```
km=KM(n_clusters=5)
km.fit(sc_df)
```

```
C:\Users\nikhil\anaconda3\Lib\site-packages\sklearn\cluster\_kmeans.py:1412:
FutureWarning: The default value of `n_init` will change from 10 to 'auto' in
1.4. Set the value of `n_init` explicitly to suppress the warning
  super()._check_params_vs_input(X, default_n_init=10)
C:\Users\nikhil\anaconda3\Lib\site-packages\sklearn\cluster\_kmeans.py:1436:
UserWarning: KMeans is known to have a memory leak on Windows with MKL, when
there are less chunks than available threads. You can avoid it by setting the
environment variable OMP_NUM_THREADS=1.
  warnings.warn(
```

Out[8]:
```
KMeans
KMeans(n_clusters=5)
```

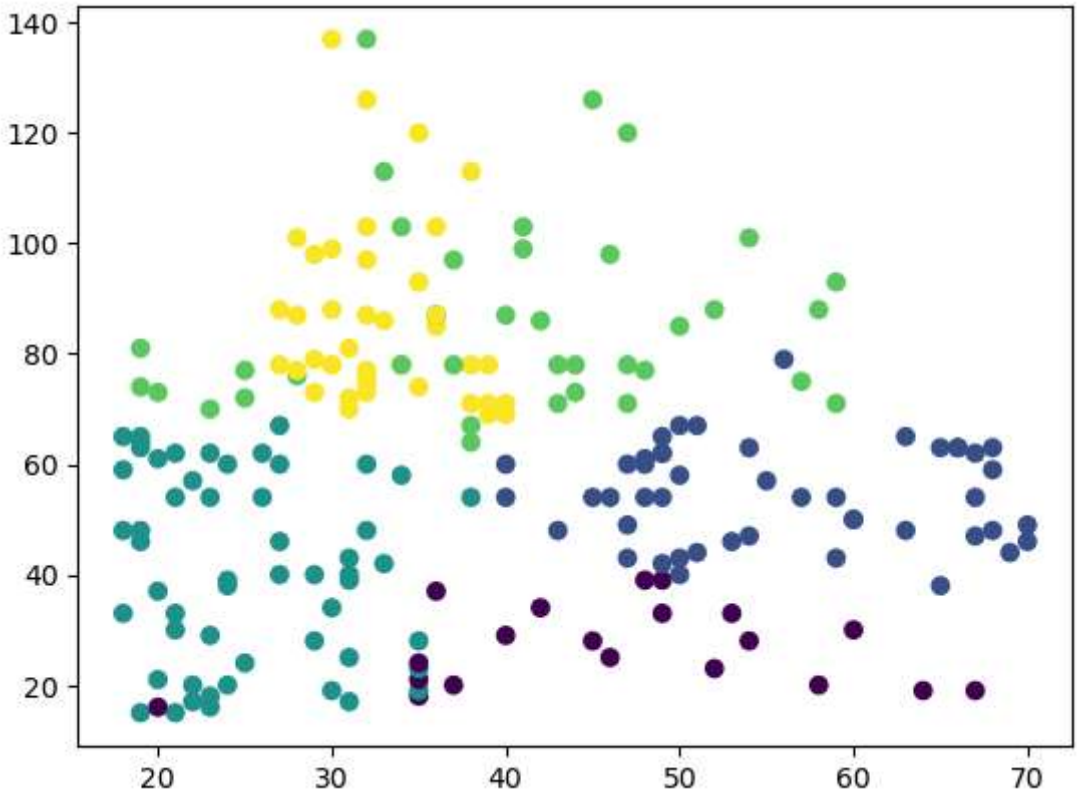
In [9]:
```
df['cluster']=km.labels_
```

In [10]:
```
# 3D Plot For KMeans Clustering
px.scatter_3d(df,x='Age',y='Annual Income (k$)',z='Spending Score (1-100)',col
```

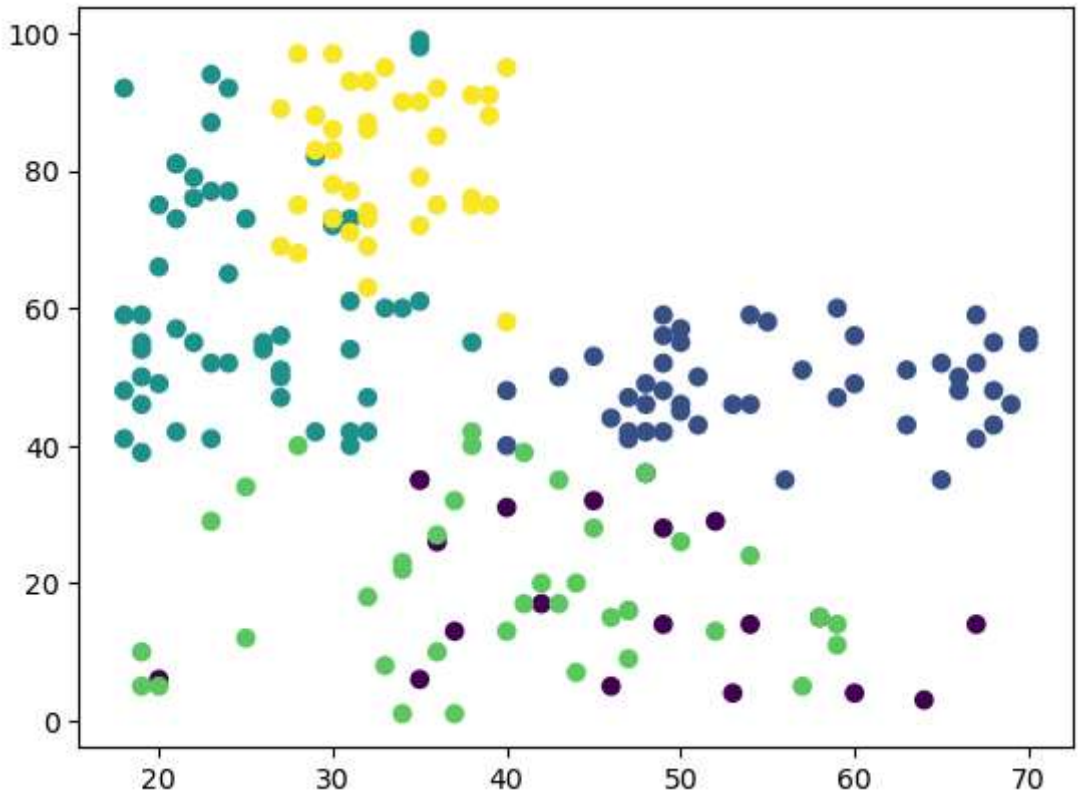


In [11]:
```
# 2D Plot For Kmeans Clustering
plt.scatter(x=df['Age'],y=df['Annual Income (k$)'],c=df['cluster'])
```

Out[11]:
```
<matplotlib.collections.PathCollection at 0x2467f74d290>
```



In [12]:
```
plt.scatter(x=df['Age'],y=df['Spending Score (1-100)'],c=df['cluster'])
plt.show()
```



In [13]:
```
plt.scatter(x=df['Annual Income (k$)'],y=df['Spending Score (1-100)'],c=df['cl
plt.show()
```