K-Means Clustering for "Mall_Customers.csv"

In [56]:
```
import pandas as pd
import matplotlib.pyplot as plt
from sklearn.cluster import KMeans
from sklearn.metrics import silhouette_score
```

In [57]:
```
# Load the dataset
data = pd.read_csv('Mall_Customers.csv')
```

In [58]:
```
# features for clustering ('Annual Income' and 'Spending Score')
X = data[['Annual Income (k$)', 'Spending Score (1-100)']]
```

In [59]:
```
# the number of clusters (K)
k = 5
```

In [60]:
```
# Create the K-means model
kmeans = KMeans(n_clusters=k, random_state=42)
```

In [61]:
```
# Fit the model to the dataset
kmeans.fit(X)
```

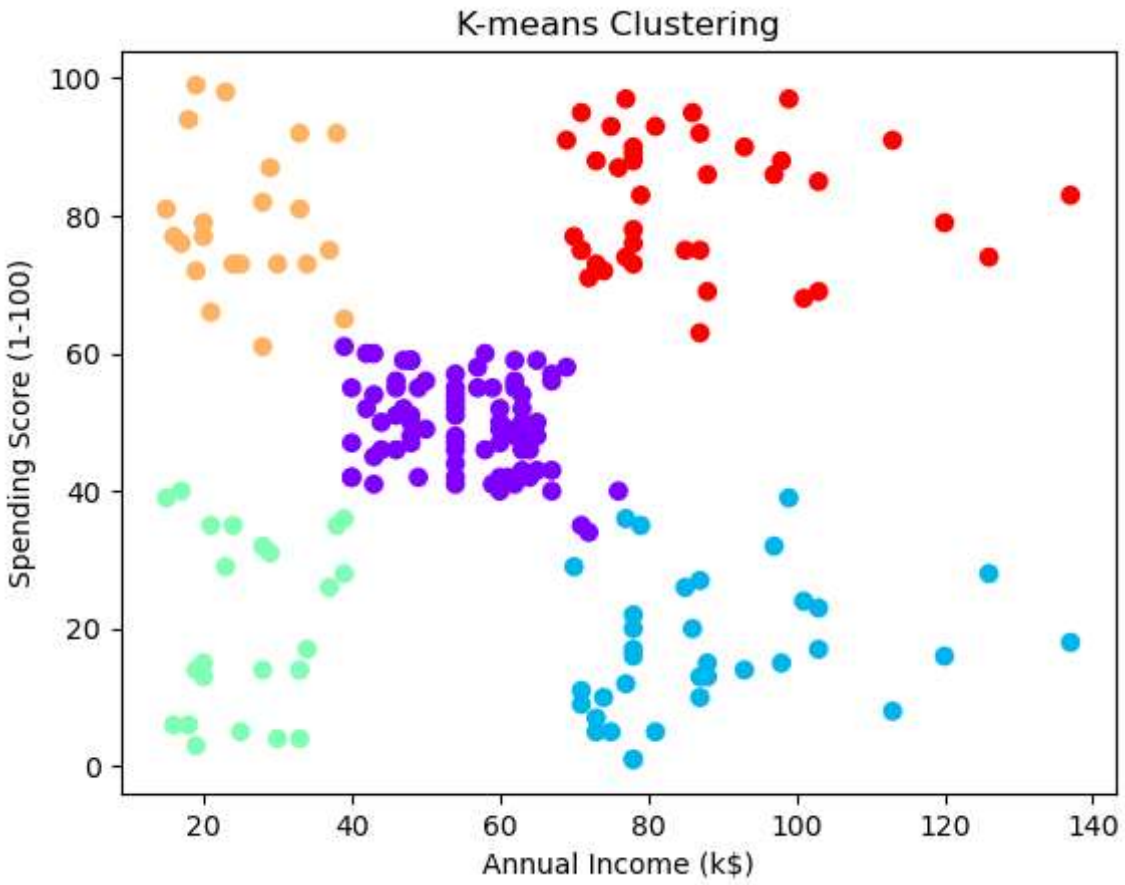
Out[61]:
```
KMeans(n_clusters=5, random_state=42)
```

In [62]:
```
# Get cluster assignments for each data point
labels = kmeans.labels_
```

In [63]:
```
# Calculate the silhouette score (higher is better)
silhouette_avg = silhouette_score(X, labels)
print(f"Silhouette Score: {silhouette_avg}")
```

```
Silhouette Score: 0.553931997444648
```

In [64]:
```
# Visualize the clustering results
plt.scatter(X['Annual Income (k$)'], X['Spending Score (1-100)'], c=labels, cm
plt.xlabel('Annual Income (k$)')
plt.ylabel('Spending Score (1-100)')
plt.title('K-means Clustering')
plt.show()
```



K-Means Clustering for "basic1.csv"

In [65]:
```
import pandas as pd
import matplotlib.pyplot as plt
from sklearn.cluster import KMeans
from sklearn.preprocessing import StandardScaler
from sklearn.metrics import silhouette_score
```

In [66]:
```
# Load the dataset
data = pd.read_csv('basic1.csv')
```

In [67]:
```
# Handle missing values (choose an appropriate strategy)
data.fillna(method='ffill', inplace=True)
```

In [68]:
```
# Standardize the data
scaler = StandardScaler()
data_scaled = scaler.fit_transform(data)
```

In [69]:
```
# the number of clusters (k)
k = 5
```

In [70]:
```
# Create the K-means model
kmeans = KMeans(n_clusters=k, random_state=42)
```

In [71]:
```
# Fit the model to the dataset
kmeans.fit(data_scaled)
```

Out[71]:
```
KMeans(n_clusters=5, random_state=42)
```

In [72]:
```
# Get cluster assignments for each data point
labels = kmeans.labels_
```

In [73]:
```
# Calculate the silhouette score
silhouette_avg = silhouette_score(data_scaled, labels)
print(f"Silhouette Score: {silhouette_avg:.2f}")
```

```
Silhouette Score: 0.51
```

In [74]:
```
plt.scatter(data_scaled[:, 0], data_scaled[:, 1], c=labels, cmap='rainbow')
plt.xlabel('Feature 1')
plt.ylabel('Feature 2')
plt.title('K-means Clustering')
plt.show()
```