ICP 6

1. Apply K means clustering in this data set provided below:

https://umkc.box.com/s/s15r7m0gnxu7b1s2kaobvc5w7da2nc1c

•Remove any null values by the mean.

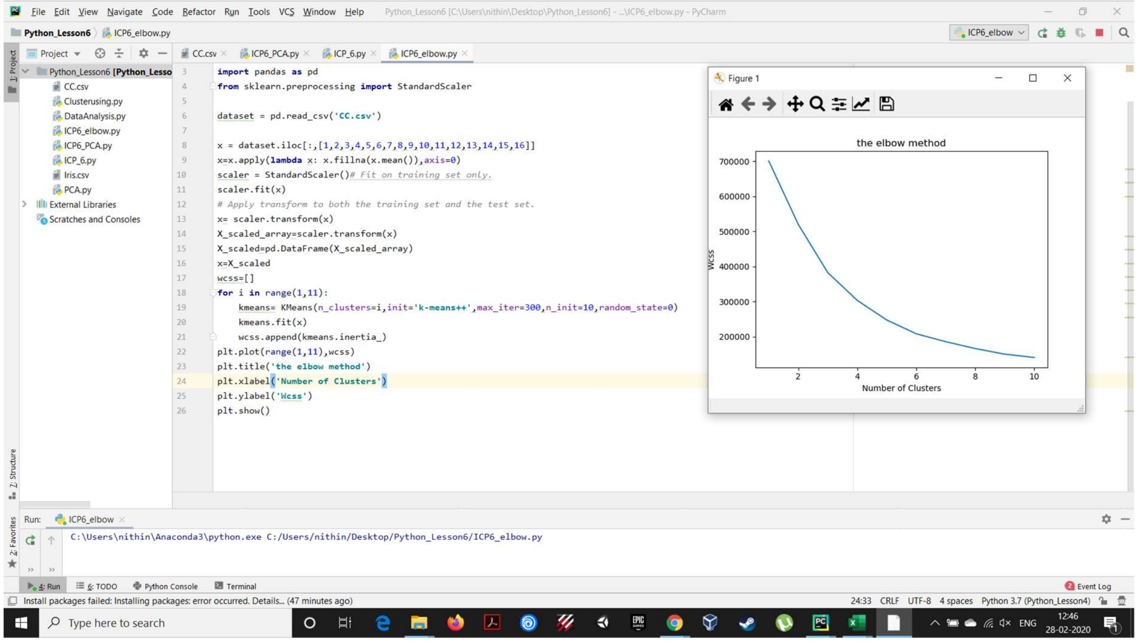
•Use the elbow method to find a good number of clusters with the KMeans algorithm

2.Calculate the silhouette score for the above clustering

3.Try feature scaling to see if it will improve the Silhouette score

4.Apply PCA on the same dataset. Data Description can be found in

https://umkc.box.com/s/cjeenva7pyj6s0vz8s8zlpbwxtmg6d17***



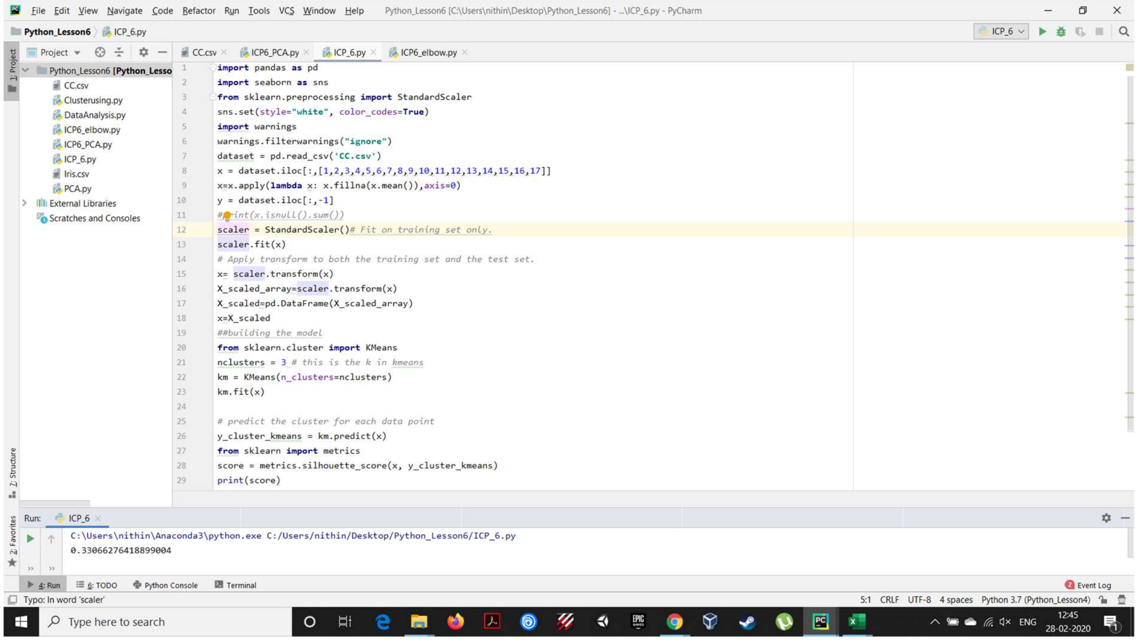
```python
import pandas as pd
from sklearn.preprocessing import StandardScaler

dataset = pd.read_csv('CC.csv')

x = dataset.iloc[:,[1,2,3,4,5,6,7,8,9,10,11,12,13,14,15,16]]
x=x.apply(lambda x: x.fillna(x.mean()),axis=0)
scaler = StandardScaler()# Fit on training set only.
scaler.fit(x)
# Apply transform to both the training set and the test set.
x= scaler.transform(x)
X_scaled_array=scaler.transform(x)
X_scaled=pd.DataFrame(X_scaled_array)
x=X_scaled
wcss=[]
for i in range(1,11):
    kmeans= KMeans(n_clusters=i,init='k-means++',max_iter=300,n_init=10,random_state=0)
    kmeans.fit(x)
    wcss.append(kmeans.inertia_)
plt.plot(range(1,11),wcss)
plt.title('the elbow method')
plt.xlabel('Number of Clusters')
plt.ylabel('Wcss')
plt.show()
```

```
C:\Users\nithin\Anaconda3\python.exe C:/Users/nithin/Desktop/Python_Lesson6/ICP6_elbow.py
```

```
Install packages failed: Installing packages: error occurred. Details... (47 minutes ago)
```


File Edit View Navigate Code Refactor Run Tools VCS Window Help
Python_Lesson6 [C:\Users\nithin\Desktop\Python_Lesson6] - ...\ICP_6.py - PyCharm
Python_Lesson6 ICP_6.py
Project
CC.csv ICP6_PCA.py ICP_6.py ICP6_elbow.py
Python_Lesson6 [Python_Lesso
CC.csv
Clusterusing.py
DataAnalysis.py
ICP6_elbow.py
ICP6_PCA.py
ICP_6.py
Iris.csv
PCA.py
External Libraries
Scratches and Consoles
Run: ICP_6
4: Run 6: TODO Python Console Terminal Event Log
5:1 CRLF UTF-8 4 spaces Python 3.7 (Python_Lesson4)
Type here to search
12:45
28-02-2020

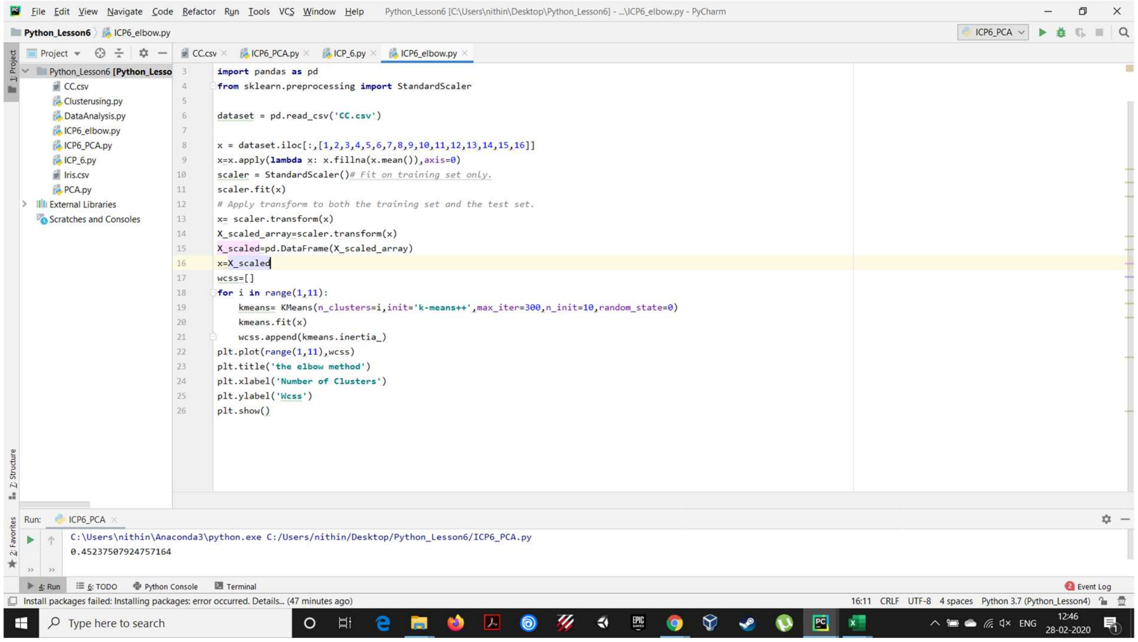
```
import pandas as pd
import seaborn as sns
from sklearn.preprocessing import StandardScaler
sns.set(style="white", color_codes=True)
import warnings
warnings.filterwarnings("ignore")
dataset = pd.read_csv('CC.csv')
x = dataset.iloc[:,[1,2,3,4,5,6,7,8,9,10,11,12,13,14,15,16,17]]
x=x.apply(lambda x: x.fillna(x.mean()),axis=0)
y = dataset.iloc[:,-1]
#print(x.isnull().sum())
scaler = StandardScaler()# Fit on training set only.
scaler.fit(x)
# Apply transform to both the training set and the test set.
x= scaler.transform(x)
X_scaled_array=scaler.transform(x)
X_scaled=pd.DataFrame(X_scaled_array)
x=X_scaled
##building the model
from sklearn.cluster import KMeans
nclusters = 3 # this is the k in kmeans
km = KMeans(n_clusters=nclusters)
km.fit(x)

# predict the cluster for each data point
y_cluster_kmeans = km.predict(x)
from sklearn import metrics
score = metrics.silhouette_score(x, y_cluster_kmeans)
print(score)
```

```
C:\Users\nithin\Anaconda3\python.exe C:/Users/nithin/Desktop/Python_Lesson6/ICP_6.py
0.33066276418899004
```

Typo: In word 'scaler'

2. Bonus points Apply kmeans algorithm on the PCA result and report your observation if the score improved or not?


File Edit View Navigate Code Refactor Run Tools VCS Window Help
Python_Lesson6 [C:\Users\nithin\Desktop\Python_Lesson6] - ...\ICP6_elbow.py - PyCharm
Python_Lesson6 ICP6_elbow.py
ICP6_PCA
Project
CC.csv ICP6_PCA.py ICP_6.py ICP6_elbow.py
Python_Lesson6 [Python_Lesso
CC.csv
Clusterusing.py
DataAnalysis.py
ICP6_elbow.py
ICP6_PCA.py
ICP_6.py
Iris.csv
PCA.py
External Libraries
Scratches and Consoles
Run: ICP6_PCA
4: Run 6: TODO Python Console Terminal Event Log
16:11 CRLF UTF-8 4 spaces Python 3.7 (Python_Lesson4)
Type here to search
12:46
28-02-2020

```
import pandas as pd
from sklearn.preprocessing import StandardScaler

dataset = pd.read_csv('CC.csv')

x = dataset.iloc[:,[1,2,3,4,5,6,7,8,9,10,11,12,13,14,15,16]]
x=x.apply(lambda x: x.fillna(x.mean()),axis=0)
scaler = StandardScaler()# Fit on training set only.
scaler.fit(x)
# Apply transform to both the training set and the test set.
x= scaler.transform(x)
X_scaled_array=scaler.transform(x)
X_scaled=pd.DataFrame(X_scaled_array)
x=X_scaled
wcss=[]
for i in range(1,11):
    kmeans= KMeans(n_clusters=i,init='k-means++',max_iter=300,n_init=10,random_state=0)
    kmeans.fit(x)
    wcss.append(kmeans.inertia_)
plt.plot(range(1,11),wcss)
plt.title('the elbow method')
plt.xlabel('Number of Clusters')
plt.ylabel('Wcss')
plt.show()
```

```
C:\Users\nithin\Anaconda3\python.exe C:/Users/nithin/Desktop/Python_Lesson6/ICP6_PCA.py
0.45237507924757164
```

Install packages failed: Installing packages: error occurred. Details... (47 minutes ago)

BY

DUKKIPATI SRI SAI NITHIN CHOWDARY

CLASS ID: 4