Web Mining Lab Assigment 6 (clustering)

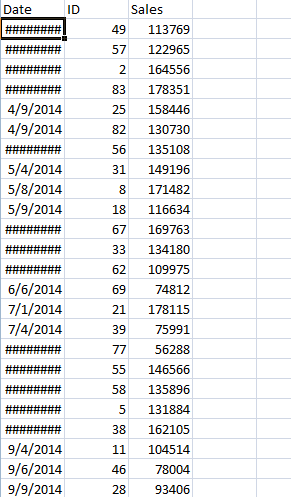
## Name: Om Ashish Mishra

## Registration Number: 16BCE0789

## Slot: B2

# The K-means

# The dataset:



Note: the date is coming encrypted but it still exits.

# The Code:

import pandas

import pylab as pl

from sklearn.cluster import KMeans

from sklearn.decomposition import PCA

variables = pandas.read\_csv('sales.csv')

Y = variables[['Sales']]

X = variables[['ID']]

X\_norm = (X - X.mean()) / (X.max() - X.min())

Y\_norm = (Y - Y.mean()) / (Y.max() - Y.min())

Nc = range(1, 20)

kmeans = [KMeans(n\_clusters=i) for i in Nc]

kmeans

score = [kmeans[i].fit(Y).score(Y) for i in range(len(kmeans))]

score

pl.plot(Nc,score)

pl.xlabel('Number of Clusters')

pl.ylabel('Score')

pl.title('Elbow Curve')

pl.show()

pca = PCA(n\_components=1).fit(Y)

pca\_d = pca.transform(Y)

pca\_c = pca.transform(X)

kmeans=KMeans(n\_clusters=3)

kmeansoutput=kmeans.fit(Y)

kmeansoutput

pl.figure('3 Cluster K-Means')

pl.scatter(pca\_c[:, 0], pca\_d[:, 0], c=kmeansoutput.labels\_)

pl.xlabel('Sales')

pl.ylabel('ID')

pl.title('3 Cluster K-Means')

pl.show()

# The Output:

