LAB REPORT-2

Name: Akansha Maloo

SBU ID: 113262290

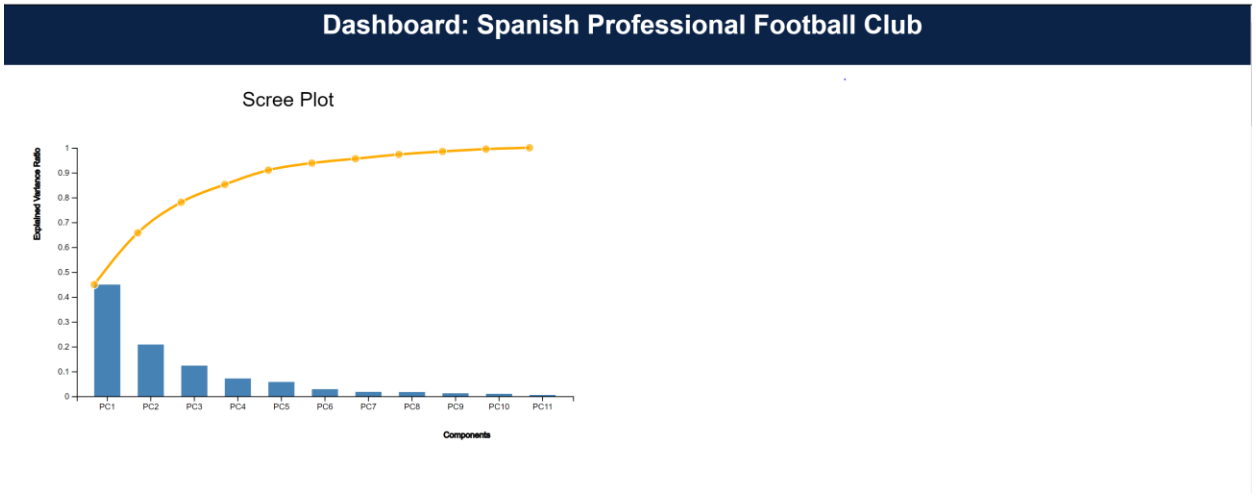
Chosen Data set: FIFA 2019 Players – Spanish Clubs

Link: https://www.kaggle.com/karangadiya/fifa19

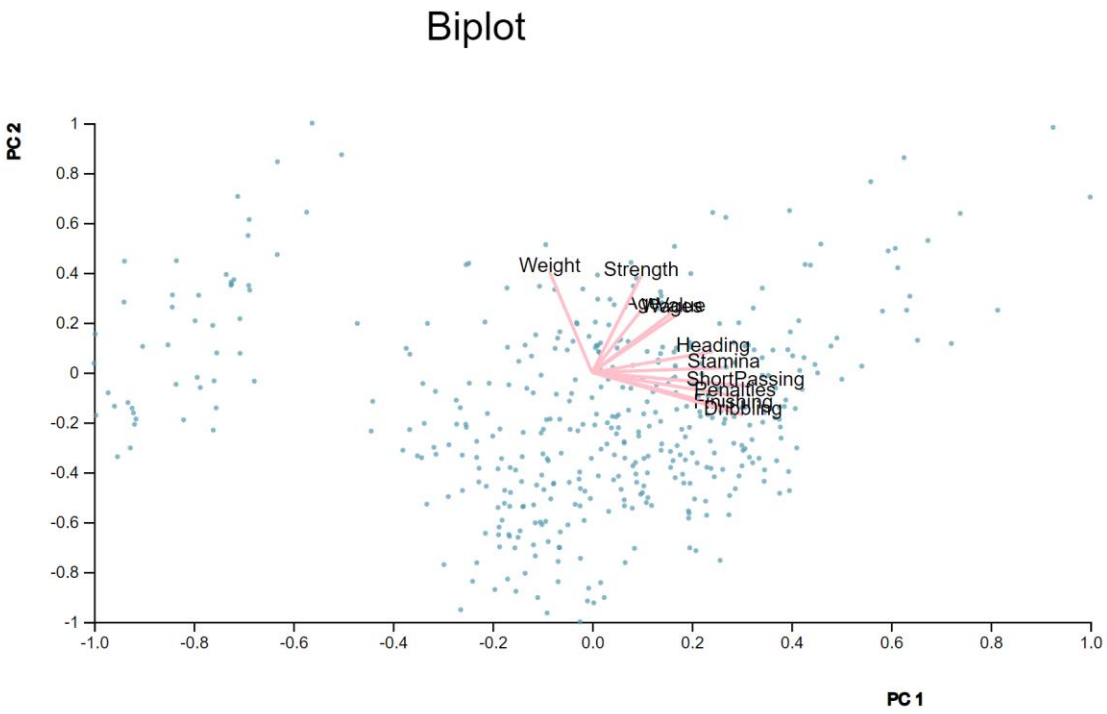
Aim: To implement a client-server system using flask. Implement Principal component Algorithm, K-means clustering and multi dimensional scaling on python and show them in plots using D3 in Javascript.

Task1 :

i. Implemented Scree plot for PCA components and variance ratio calculated python. Used tool tip functionality of d3 library to show value of cumulative covariance and component selected.

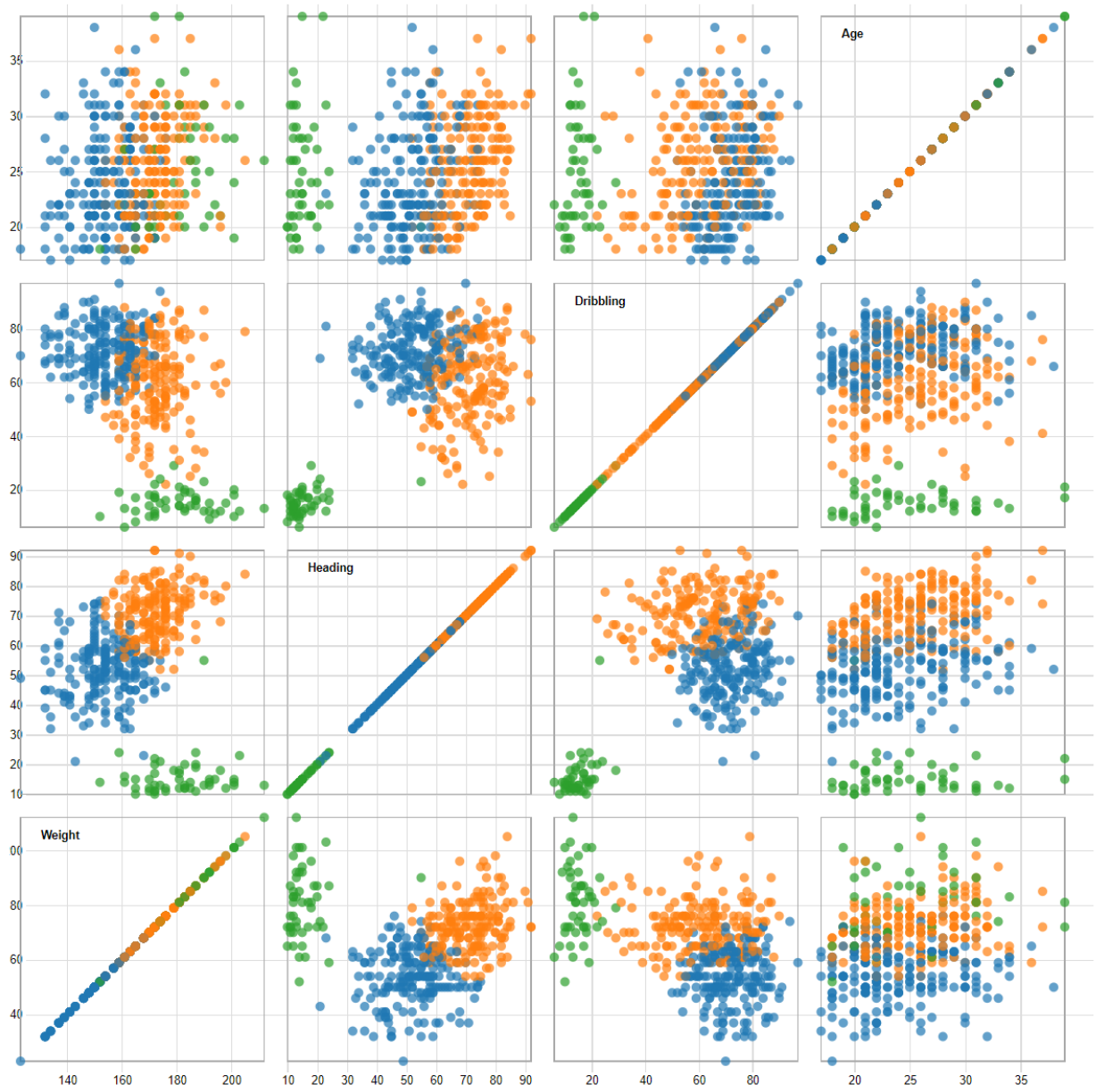


ii. Biplot of first two Principal components.TaTsk



Task 2: Create a Scatter matrix. Use k-means clustering to cluster data for the top 4 attributes . Top four attributes are selected by the user. It is found by summation of square of PC values of the components and the attributes with top 4 maximum values will be plotted. Suppose user selected PC6, then a square of sum of PC1, PC2,PC3,PC4,PC5,PC6 will be done.

Scatter Matrix

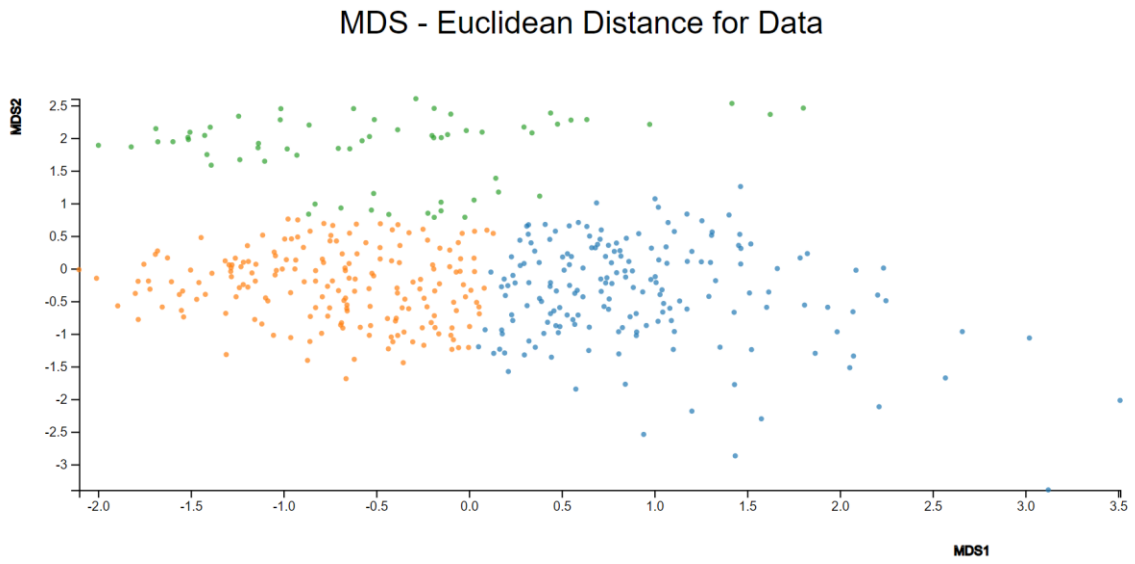


iii. Loading table of top 4 attributes.

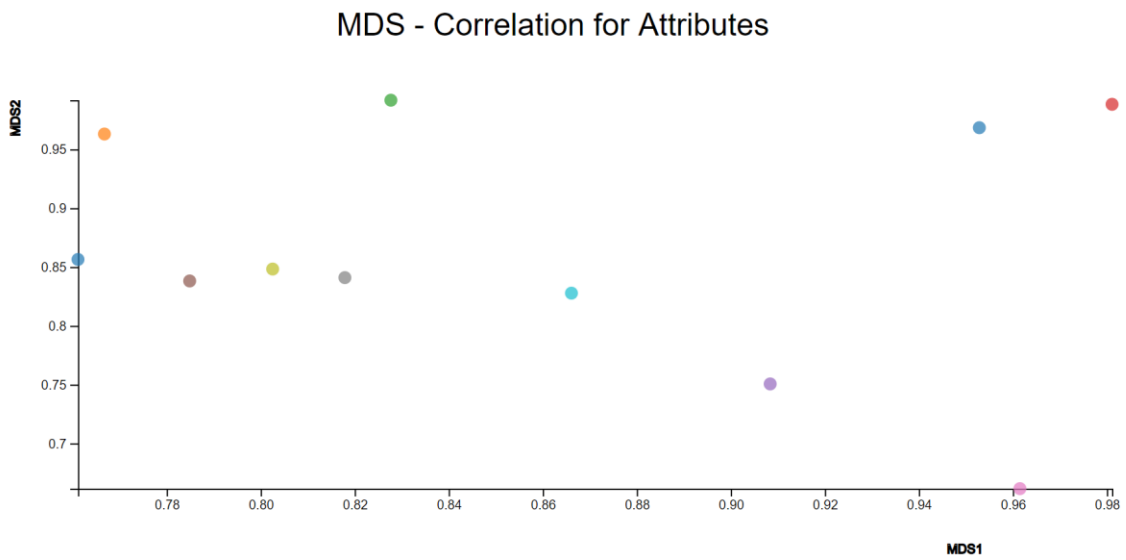
Loadings Table of top 4 Attributes

Attribute	PC1	PC2	PC3	PC4	PC5	PC6
Age	0.13406254224039174	0.33865630803370317	-0.13476827960146237	0.8328102151209049	-0.3295868155021333	-0.11432483606142257
Weight	-0.11361864789403163	0.5383944253652427	-0.15351115069892468	-0.08937714936232836	0.46446949916205	0.3060215562854305
Heading	0.32381569833332435	0.11174925549221608	-0.3851330516661118	-0.3210152679861661	0.009610906154967978	-0.6845445537383147
Stamina	0.35234424266237524	0.025226726865651073	-0.1542386630795917	-0.23729765961560673	-0.5349336464935476	0.507803972720401

Task 3:



ii. Each circle is a numerical attribute and when we hover over the dot, its attribute name appears. This is implemented using tool tip functionality in D3 library.



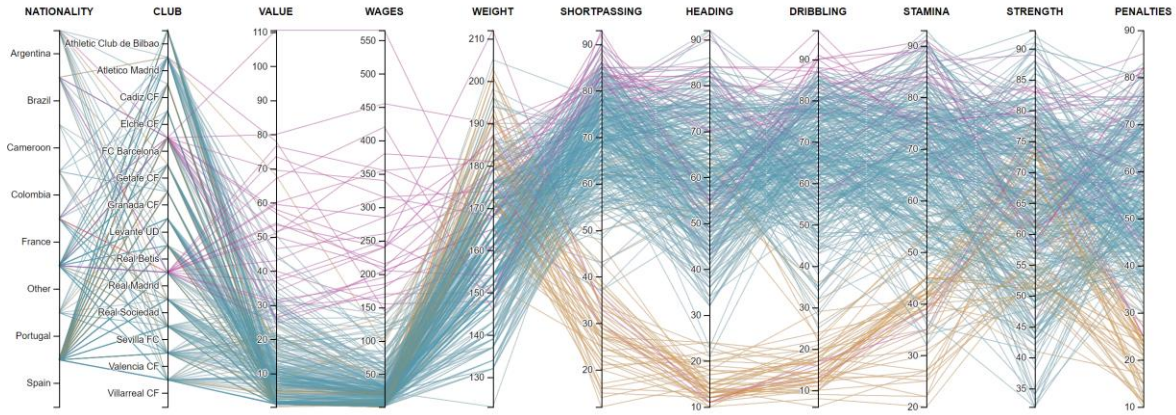
Task 4:

Implemented brushing using brush functionality in D3. Used drag functionality in D3 to implement reordering of axis feature. Used k-means clustering in python to cluster and colored to different colors.

Findings: From the plot below we can say that Values and Wages are positively correlated . We can also say that Weight and Shortpassing is negatively correlated.

Using brusing feature, we also found that the players with value more than 40 million Euros belong to FC Barcelona , Real Madrid and Atletico Madrid

Parallel Coordinates Plot of Players' Features



Parallel Coordinates Plot of Players' Features