**Week 8 Tutorial Answer Sheet: Clustering Analysis**

Suppose that you are hired by a newly established soccer team as a hiring manager. Your team is interested to hire a new soccer player at whatever cost. You have all soccer players’ performance data in English Premier League from 2017-2018. Download the file called “SoccerData.xlsx” from Canvas. Use SAS Enterprise Miner to perform cluster analysis. You should remove text variables such as Club, Name, Weekly Wage, and Position, in the data file.

Please address the following questions in *red and italic fonts*.

1. Please describe the steps to generate the clusters in SAS Enterprise Miner in bullet point formats to describe each step (after you have created a new diagram). For example, what components should you include in the diagram? Is there any property that you should change?
2. *Add nodes (file import & cluster) and connect them*
3. *Drop nominal variable*
4. *Run path and get result*
5. How many clusters do you identify?

*4*

1. Which cluster represents frequent “fouls” and the best scorers, respectively?
2. *Id 3 cluster*
3. *Id 4 cluster*
4. Which cluster represents the best defensive players? (Hint: the group with highest mean in Inter. Inter means interceptions).

*ID 1 cluster*

1. If your team wants to hire a good scorer, which player would you like to recruit and why?
   1. Hint 1: Go back to the path model in SAS Enterprise Miner. Click “Cluster” and on the left panel, click “…” next to “Exported Data” under “General”. Click the row of “Train” in the pop-up menu. You may click “browse” to view the data.
   2. Hint 2: You may right click the result table and choose “Export to Excel”. Then you may copy and paste the last column (i.e., Segment Description) to your original file of SoccerData.xlsx. Which player will you hire? You may pay attention to the “Weekly Wage” apart from “Segment Description”.

*In view of the performance, Gareth Barry has an outstanding one because he has the greatest win to loss ratio which is probably has once to win the shot out of two games. Moreover, his shooting accuracy is around the average of the scorer, so can reflect his steadiness in performance.*

*In terms of cost, he has no proposed weekly wage which we still have the bargaining area to minimise our cost on scorer.*