PROFILE - FREELANCER

EduManta (https://www.edumanta.com/) is an Online Educational Platform that provides live courses to professionals and the other wing of EduManta helps college students across the world by providing them doubts clarification and assignment solutions.

So, we are connecting with you to help those students in their academic doubts and assignments. Also, please note you will be hired as a Freelancer.

Job Profile

From the August; the session will start in our targeted countries and you will be required to provide the solution for assignments as and when needed. As assignments from these developed countries are real-time, and based on real situations, it will provide you with ample opportunity to learn. Not only that, you will get real-time feedback from the USA's university professors and all this experience will not only help you to earn a good amount per assignment or doubt clearing session but also an enriching experience that will help you to crack your future jobs.

Further on connecting with us will provide you a lifetime opportunity to earn part-time income from what you have learned throughout your life. The payment will be made on per assignment basis.

If you are interested in connecting with us, do solve this assignment and send it via intershala chat or via email.

SELECTION PROCESS – PYTHON ASSIGNMENT

Assessment Requirements / Tasks (include all guidance notes)

This assignment will use employment data of Wales from the StatsWales data source.

1. Data processing

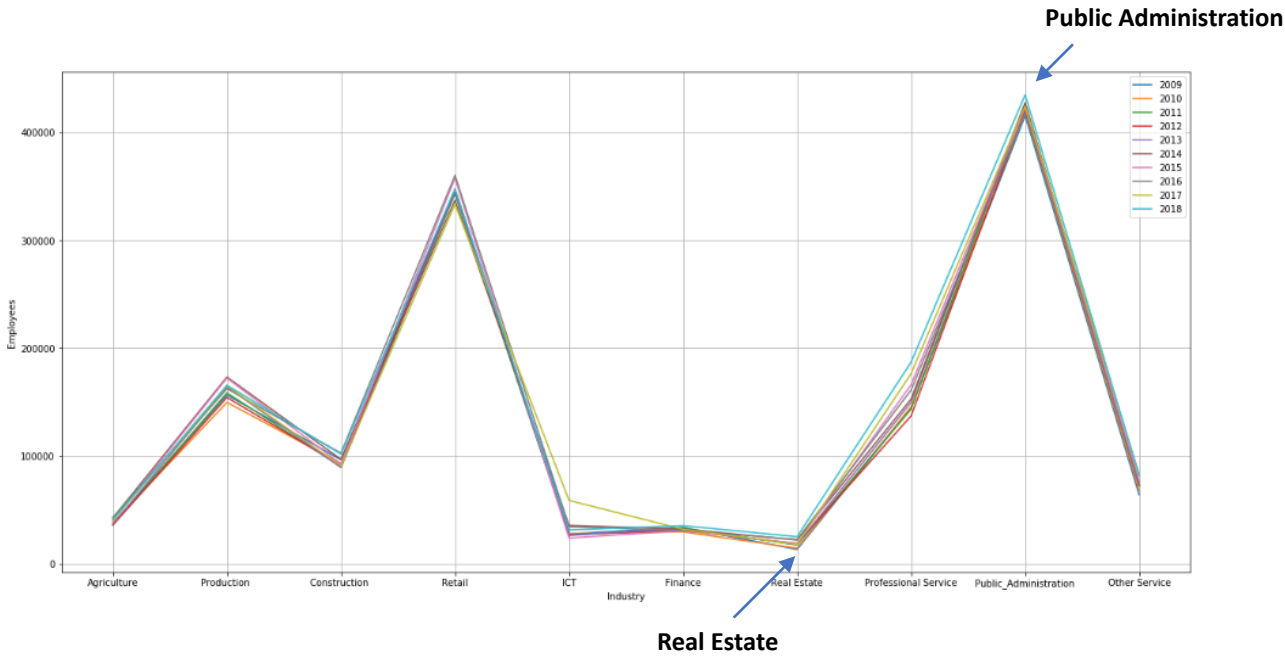
The final data frame :-

Industry	2009	2010	2011	2012	2013	2014	2015	2016	2017	2018
Agriculture	37700	38200	36100	36100	36800	42700	40700	43200	40200	41100
Production	156700	149800	158600	154400	164200	173300	172300	162500	165100	165700
Construction	96600	93200	90000	91300	89300	97000	92600	102700	90800	101800
Retail	345400	344500	343100	347300	345100	337300	357700	360200	333500	347600
ICT	27800	27900	26400	27200	26900	35700	24000	34400	58900	31500
Finance	33800	29800	33200	31100	32400	32400	30800	31000	32100	35500
Real Estate	13500	14600	17600	18800	18000	22200	19100	22700	18200	25200
Professional Service	144800	145800	143600	137300	149900	152900	166200	161200	176400	187100
Public Adminstration	415600	418600	425600	421000	427000	427600	423200	418500	424500	434900
Other Service	64200	68000	72400	72800	75500	73300	77200	72400	83200	81800

2. Data analysis

For each question provide graph/chart along with your own interpretation (~ 50 words)

2.1. Which industry employed highest and lowest workers over the period?

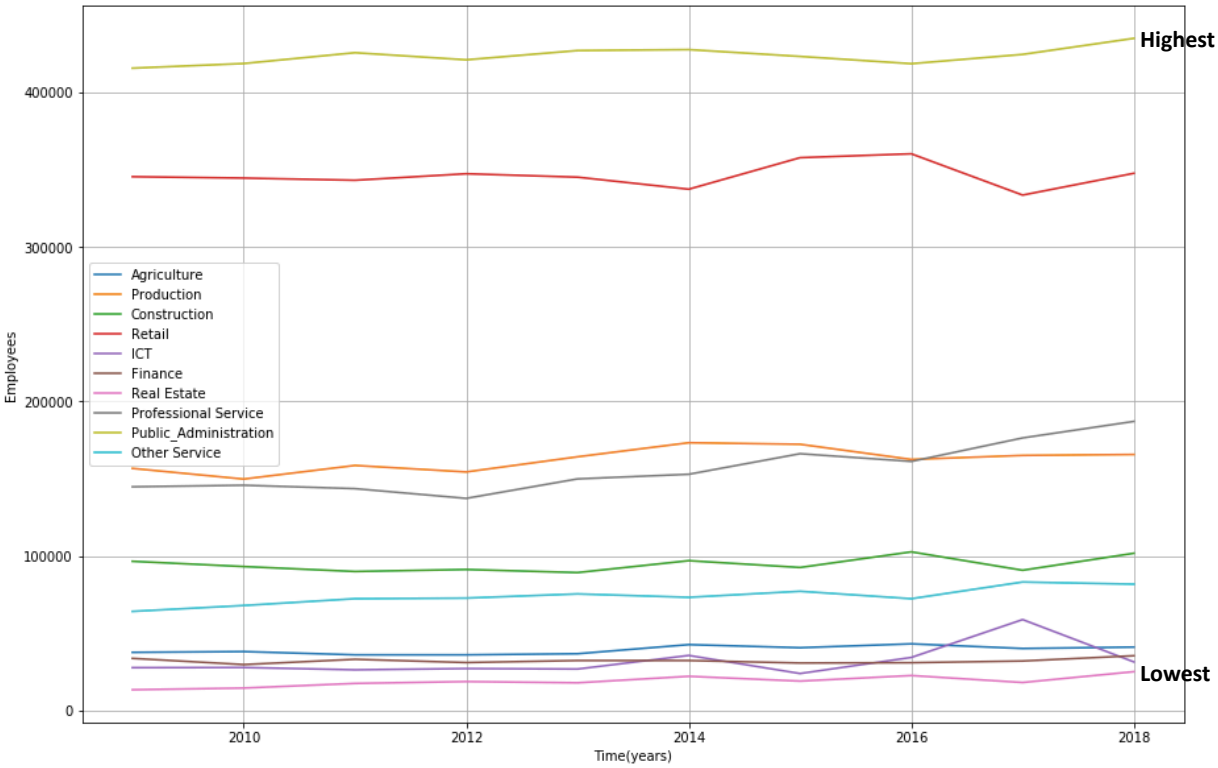


After plotting the graph btw Employees vs Industry where different colour lines represent different years. It is clear from the graph i.e.

Public Administration has the **highest** workers over the period

Real Estate has the **lowest** workers over the period

2.2. Which industry has the highest and lowest overall growth over the period?



After plotting the graph btw Employees vs Years where different colour lines represent different Industries. It is clear from the graph i.e.

Public Administration has the **highest** growth over the period

Real Estate has the **lowest** growth over the period

2.3. Which years are the best and worst performing year in relation to number of employments? (highest and lowest employment)

Best performing year in relation to number of lowest employment is 2018

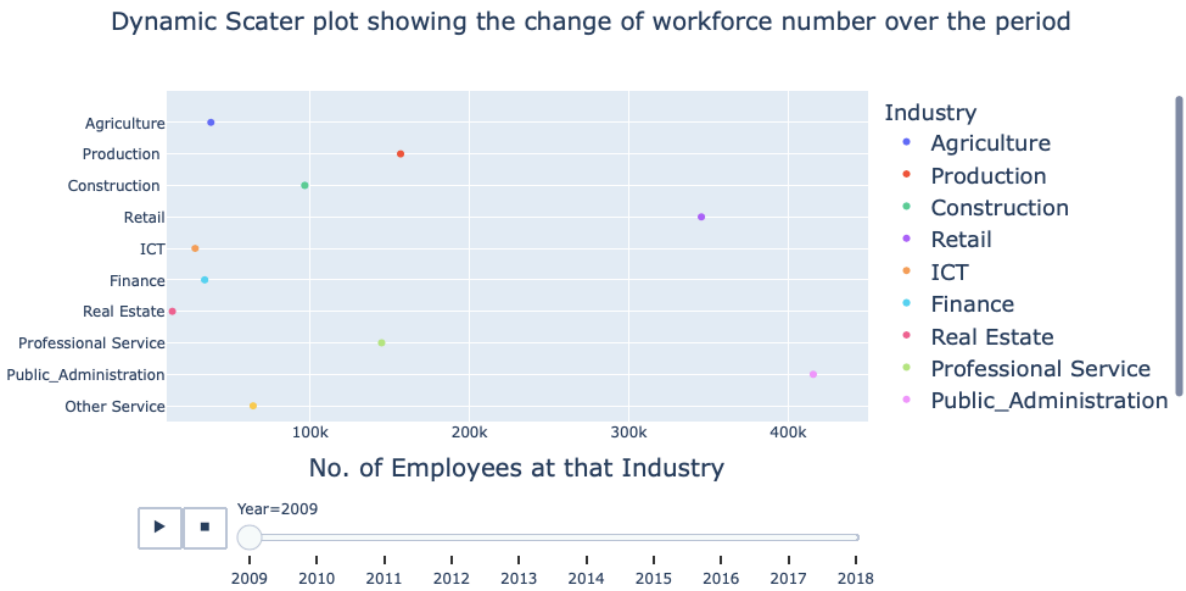
Best performing year in relation to number of highest employment is 2018

Worst performing year in relation to number of lowest employment is 2009

Worst performing year in relation to number of highest employment is 2009

3. Visual analysis

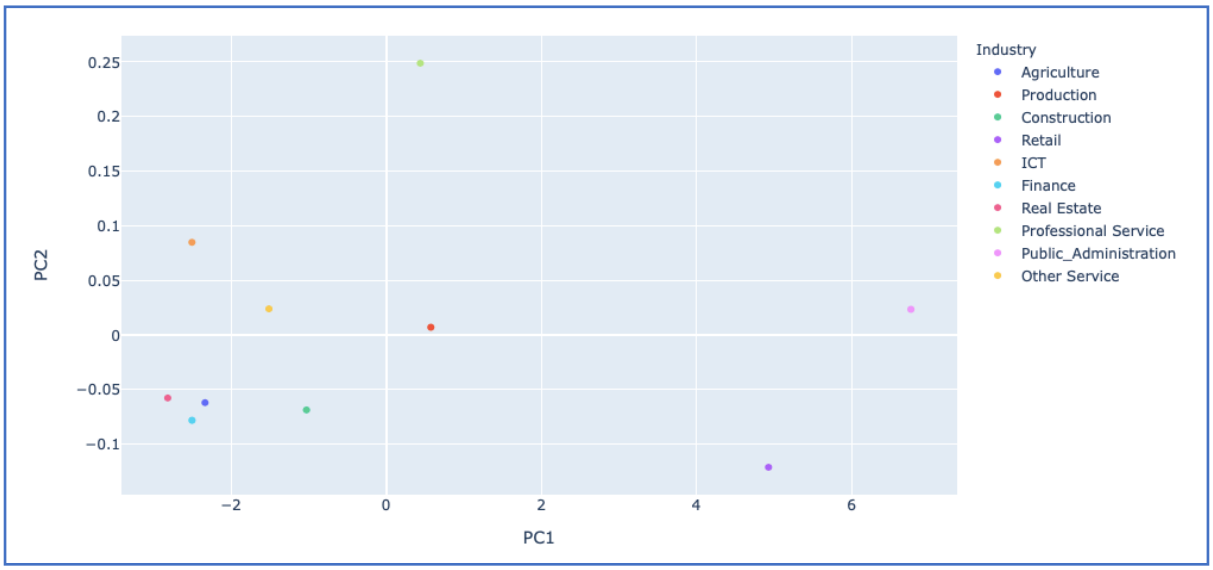
Create a dynamic scatter/bubble plot showing the change of workforce number over the period using Plotly express.



Animated scatter plot creating using Plotly Express. We can move the slider to change the year as per our convenience.

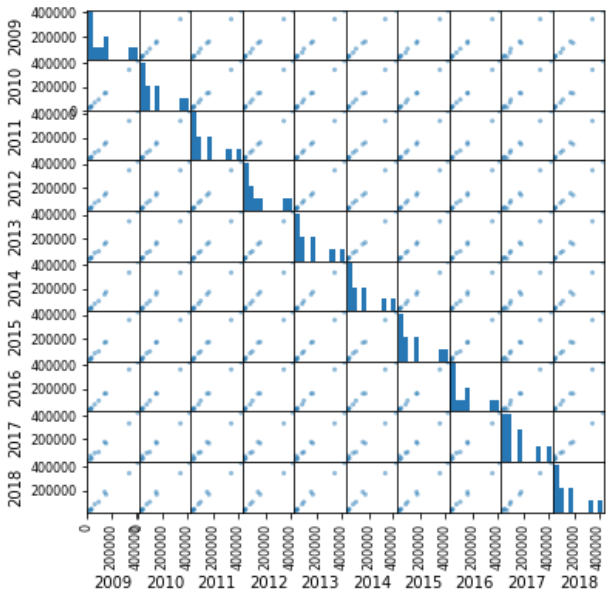
4. PCA/Correlation

4.1. Undertake a PCA (PC=2; columns should be like PC1, PC2, Industry) and produce a scatter plot. Write your interpretation about the plot and in relation to the analysis of section 2 & 3 (for example which industries are correlated over the years as well as in PCA etc.)



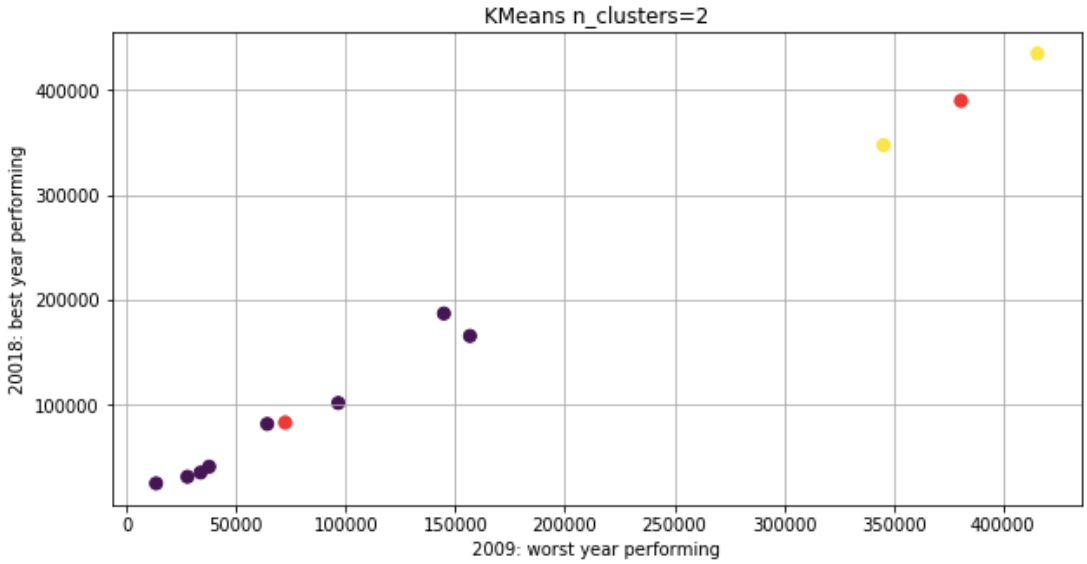
A PCA plot shows clusters of samples based on their similarity and shows how strongly each characteristic influences a principal component. We can see that Real Estate, Finance and Agriculture are strongly correlated. Public Administration follows the behaviour of Retail Industry closely. further we can relate ICT other services and production form a group which can be interrelated to each other

4.2. Make a year wise correlation for each industry. Does the aforementioned industries are also correlated over the years? Explain your answer.



5. Clustering (k means & hierarchical)

5.1. Using the best and worst performing year column's employment data (2.3) undertake a K means clustering analysis (K=2 & 3) and identify industries cluster together. Write your own interpretation (~100 words).

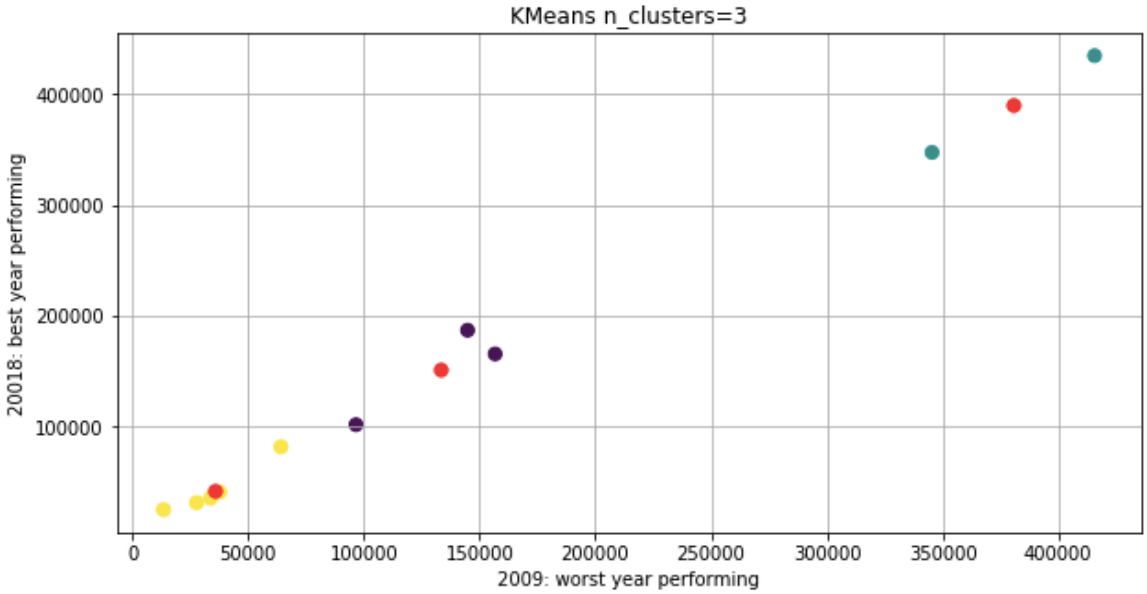


K-MEANS || No. of clusters : 2

Group-1 : Public Administration and Retail Industry

Group-2: Agriculture, Production, Construction, ICT, Finance, Real Estate, Professional Service, Other Service

When we keep number of clusters to be two. We can see there is two groups from the production and the retail form a group and the rest other from a group so many there are two groups and their relation can be interrelated during the worst and best year performance



K-MEANS || No. of clusters : 3

We can clearly see two partition and RED dot is the mean of each group

Group-1 : Public Administration and Retail Industry

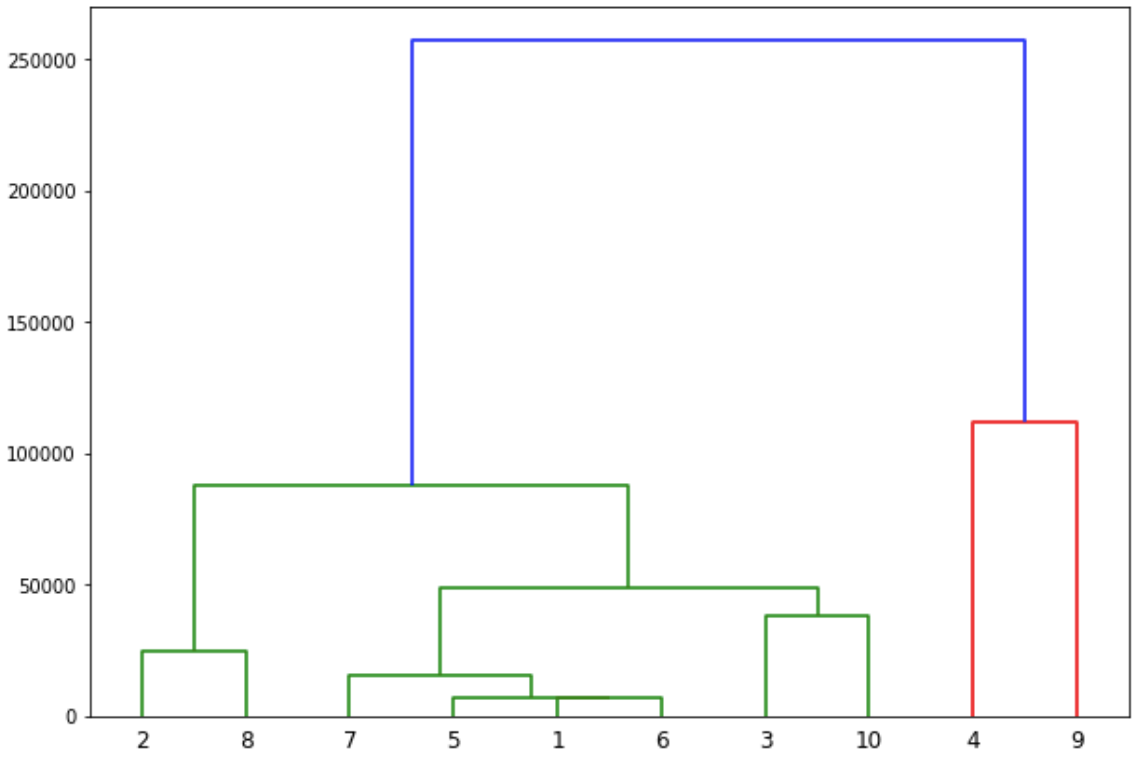
Group-2: ICT, Finance, Real Estate, Professional Service

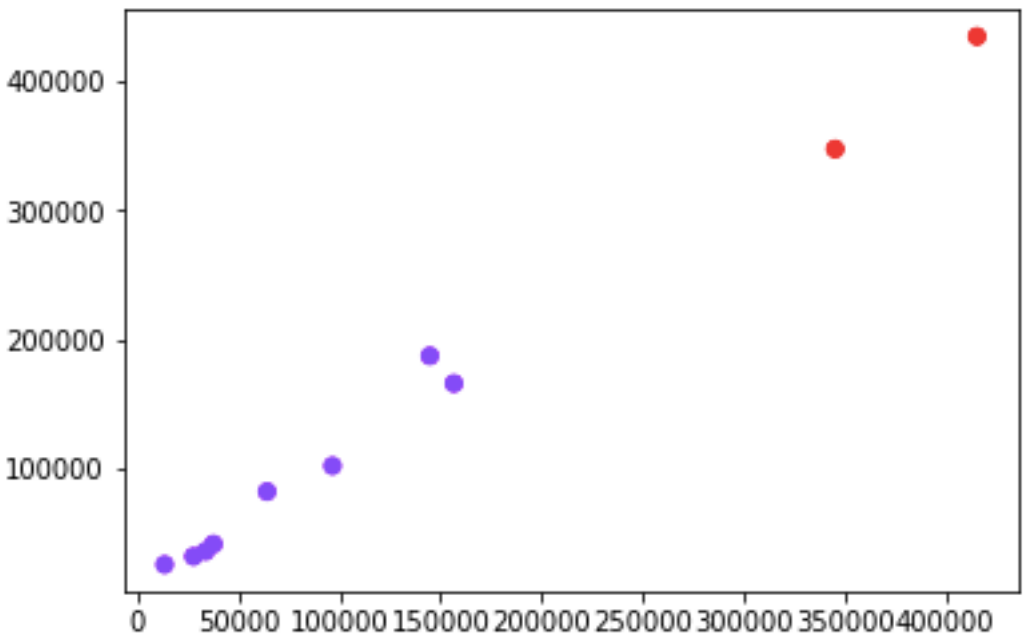
Group 3: Agriculture, Production, Construction, Other Service

the number of clusters are 3 and this is a better K means clustering then when the clusters was 2. As we can see a clear 3 group partition is found which are for which everyone has equal partition.

5.2. Using the same dataset (best & worst performing) create a hierarchical cluster. Compare the cluster with k means clusters. In K Means clustering, since we start with random choice of clusters, the results produced by running the algorithm multiple times might differ. While results are reproducible in Hierarchical clustering.

K Means clustering requires prior knowledge of K i.e. no. of clusters you want to divide your data into. But, you can stop at whatever number of clusters you find appropriate in hierarchical clustering by interpreting the dendrogram





6. Discussion

Provide a brief discussion (~ 300 words) on employment landscape of Wales based on the employment data analysis results.

There is no doubt that important demographic, economic, technological and environmental shifts are underway that will lead to structural changes to the world of work. This section concentrates on the existing evidence about how these drivers of change will affect the world of work in Wales. Much of the data and analysis is at the UK level, and it is important to note that these trends will be mediated by regional and local factors. These include social and community characteristics, such as a pre-existing industrial strategy or economic culture, the resultant skills base, the density of the population and its cultures, values and norms and as a result their impact will vary between sectors and in different localities.

From the employment data of Wales we can see that that the public administration has the highest employment and the real estate has the lowest number of workforce. 2018 has been the highest the best year for the employment whereas 2009 has been the worst year for the employment. We can also say that the public administration almost stays constant for the whole 10 years where is there is and drift change in the ICT sector in the year 2017 there is a high increase in the number of employment the graph of retail also varies in the York 2014 to 2000 17 the real estate lowest number of employment and almost remain same. Moreover after doing principal component analysis we can see that agriculture production and Finance are almost related to each other and their graph vary in the same way. Plus there is a positive correlation between industries in the year of 2009 and 2018

Assessment Criteria

1.1 Data preparation	**05**
1.2 Data preparation	**05**
1.3 Data preparation	**05**
2.1 Data analysis	**05**
2.2 Data analysis	**05**
2.3 Data analysis	**05**
3 Visual analysis	**20**
4.1 PCA	**10**
4.1 Correlation	**10**
5.1 Clustering	**10**
5.2 Clustering	**10**
6 Discussion	**10**

Note:

- **Download the Document and Rename it as (YourName_Python_EduManta)**
- **Deadline- 29th May, 2020**
- **You can provide part solution or complete solution as per your knowledge.**