CLUSTERING ANALYSIS OF COLLEGE MAJOR THAT PAY YOU MORE

Introduction:

It is big leap for any person to decide what college major to take after they complete their high school, depending on their interest and the amount of salary they will get after the completion of their college. Choosing a college major is a complex decision taking the personal interest, difficulty, and the career potential into consideration.

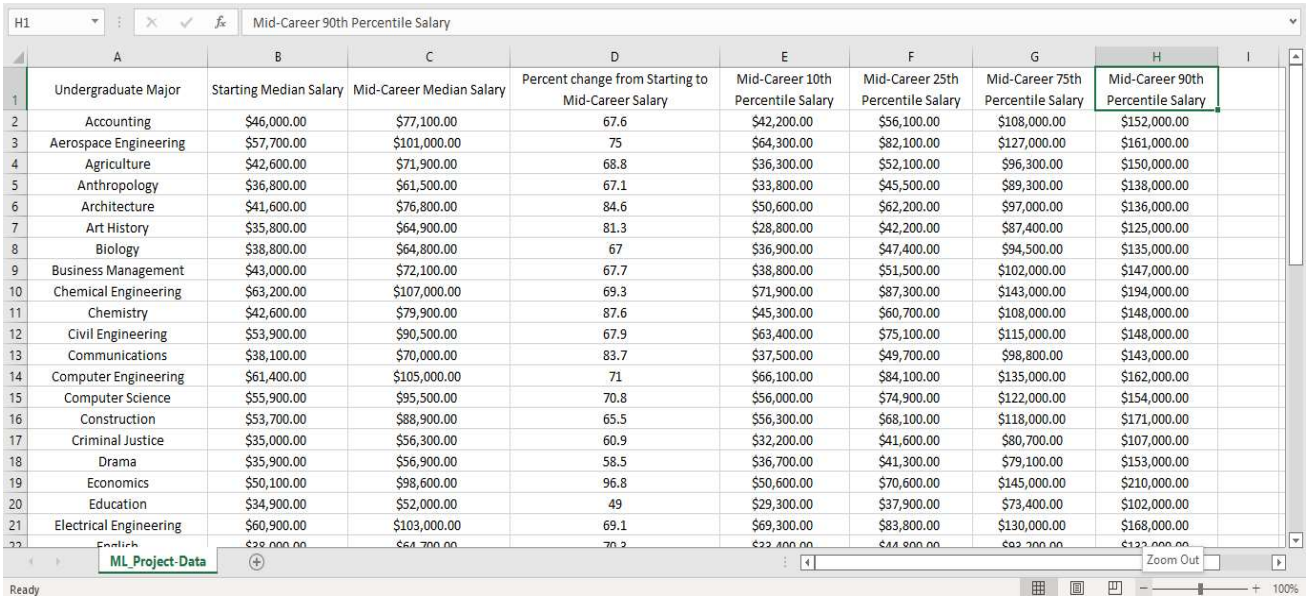
Problem Statement:

It would be a good help to know what type of major pays you more when you graduate. The report is based on The College Majors That Will Pay You More referenced in the Wall Street Journal Article. By applying the K-means Clustering to the data that is provided in the journal we can divide the majors into clusters depending on the starting salary, mid-career salary, career percentage growth the major is going to offer.

Data Set:

The data set contains the information about college majors along with the starting salary, mid-career salary, career growth percentage and the salaries in percentages. The data is taken from the wall street journal article from the PayScale Inc source. The dataset has 8 variables and 50 records.

Reference: http://online.wsj.com/public/resources/documents/info-Degrees_that_Pay_you_Back-sort.html?mod=article_inline



Undergraduate Major	Starting Median Salary	Mid-Career Median Salary	Percent change from Starting to Mid-Career Salary	Mid-Career 10th Percentile Salary	Mid-Career 25th Percentile Salary	Mid-Career 75th Percentile Salary	Mid-Career 90th Percentile Salary
Accounting	$46,000.00	$77,100.00	67.6	$42,200.00	$56,100.00	$108,000.00	$152,000.00
Aerospace Engineering	$57,700.00	$101,000.00	75	$64,300.00	$82,100.00	$127,000.00	$161,000.00
Agriculture	$42,600.00	$71,900.00	68.8	$36,300.00	$52,100.00	$96,300.00	$150,000.00
Anthropology	$36,800.00	$61,500.00	67.1	$33,800.00	$45,500.00	$89,300.00	$138,000.00
Architecture	$41,600.00	$76,800.00	84.6	$50,600.00	$62,200.00	$97,000.00	$136,000.00
Art History	$35,800.00	$64,900.00	81.3	$28,800.00	$42,200.00	$87,400.00	$125,000.00
Biology	$38,800.00	$64,800.00	67	$36,900.00	$47,400.00	$94,500.00	$135,000.00
Business Management	$43,000.00	$72,100.00	67.7	$38,800.00	$51,500.00	$102,000.00	$147,000.00
Chemical Engineering	$63,200.00	$107,000.00	69.3	$71,900.00	$87,300.00	$143,000.00	$194,000.00
Chemistry	$42,600.00	$79,900.00	87.6	$45,300.00	$60,700.00	$108,000.00	$148,000.00
Civil Engineering	$53,900.00	$90,500.00	67.9	$63,400.00	$75,100.00	$115,000.00	$148,000.00
Communications	$38,100.00	$70,000.00	83.7	$37,500.00	$49,700.00	$98,800.00	$143,000.00
Computer Engineering	$61,400.00	$105,000.00	71	$66,100.00	$84,100.00	$135,000.00	$162,000.00
Computer Science	$55,900.00	$95,500.00	70.8	$56,000.00	$74,900.00	$122,000.00	$154,000.00
Construction	$53,700.00	$88,900.00	65.5	$56,300.00	$68,100.00	$118,000.00	$171,000.00
Criminal Justice	$35,000.00	$56,300.00	60.9	$32,200.00	$41,600.00	$80,700.00	$107,000.00
Drama	$35,900.00	$56,900.00	58.5	$36,700.00	$41,300.00	$79,100.00	$153,000.00
Economics	$50,100.00	$98,600.00	96.8	$50,600.00	$70,600.00	$145,000.00	$210,000.00
Education	$34,900.00	$52,000.00	49	$29,300.00	$37,900.00	$73,400.00	$102,000.00
Electrical Engineering	$60,900.00	$103,000.00	69.1	$69,300.00	$83,800.00	$130,000.00	$168,000.00

Approach:

K-means Clustering is an unsupervised learning method having an iterative process in which the dataset is grouped into K number of predefined non-overlapping clusters. It is a process that uses the given data to divide into clusters in which each observation belongs to the cluster with the nearest mean. K-means clustering approach is one of the most often used clustering methods, the technique divides the data into "K" number of clusters based on the similarities and anomalies with the neighbouring observations. The similarities between the clusters can be related to the distances between the clusters.

The variables that are taken into consideration are: –

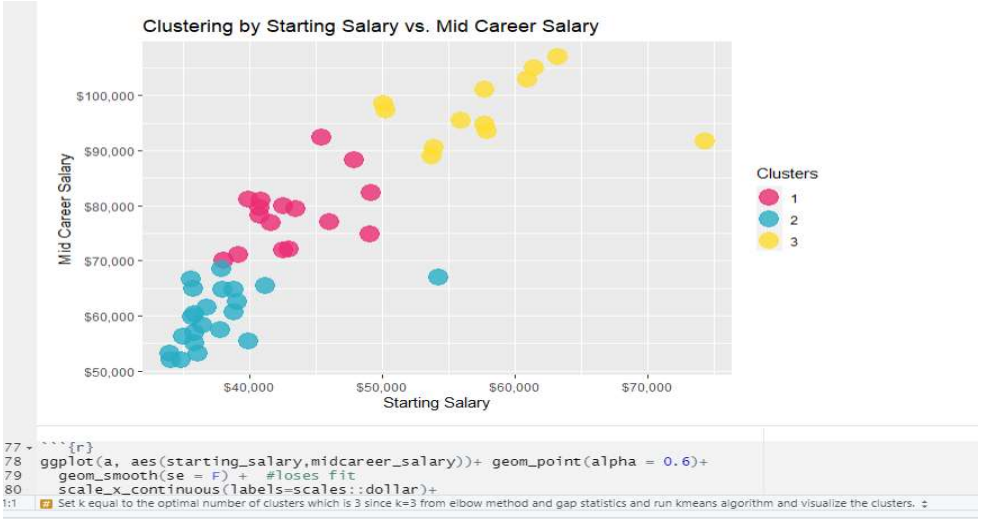
1. The starting salary that a major can be offered.
2. The mid-career salary of a major
3. The percentile salary (10,25,75,90) of the majors.

Analysis:

First determine the optimal number of clusters the data should be divided into using

1. Elbow Method
2. Silhouette Method
3. Gap Statistic Method

By using the above-mentioned methods, we get 3 clusters based on starting salary, mid-career salary.

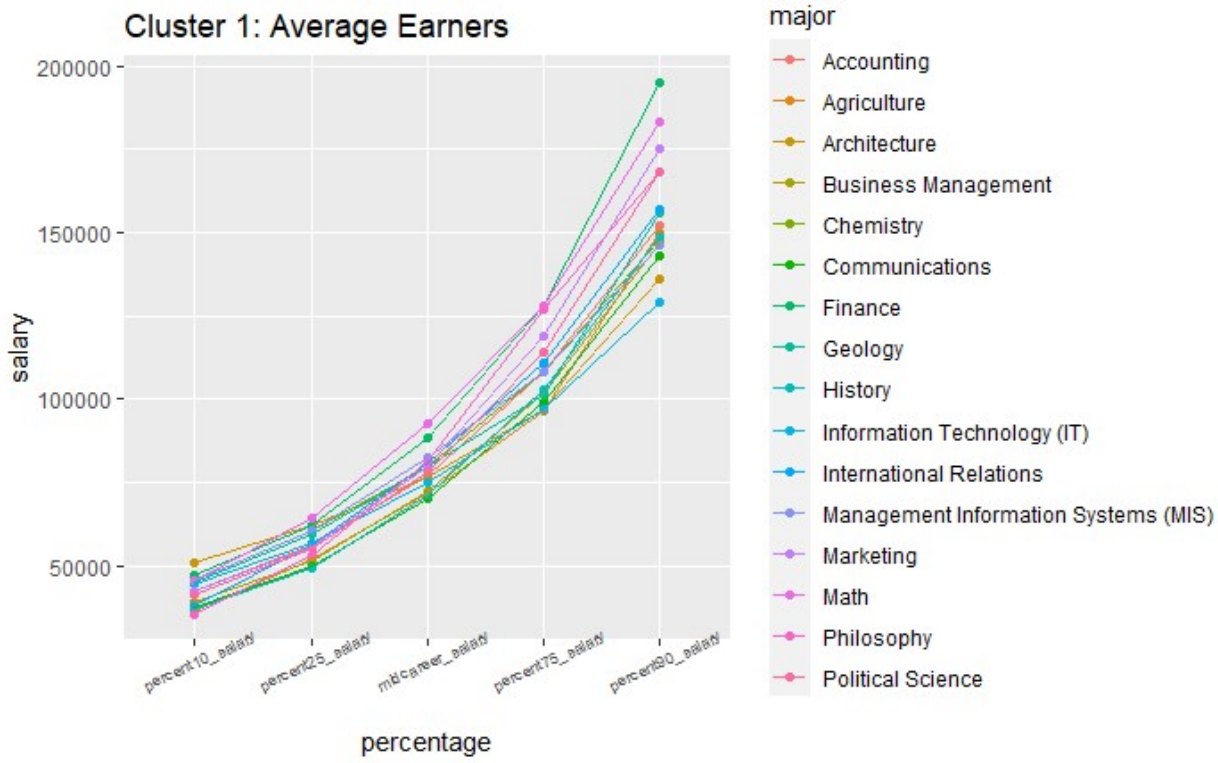


Cluster 1: The cluster 1 consists of majors that have the starting salary between $40000 to $50000 and mid-career salary ranging from $70000 to $90000.

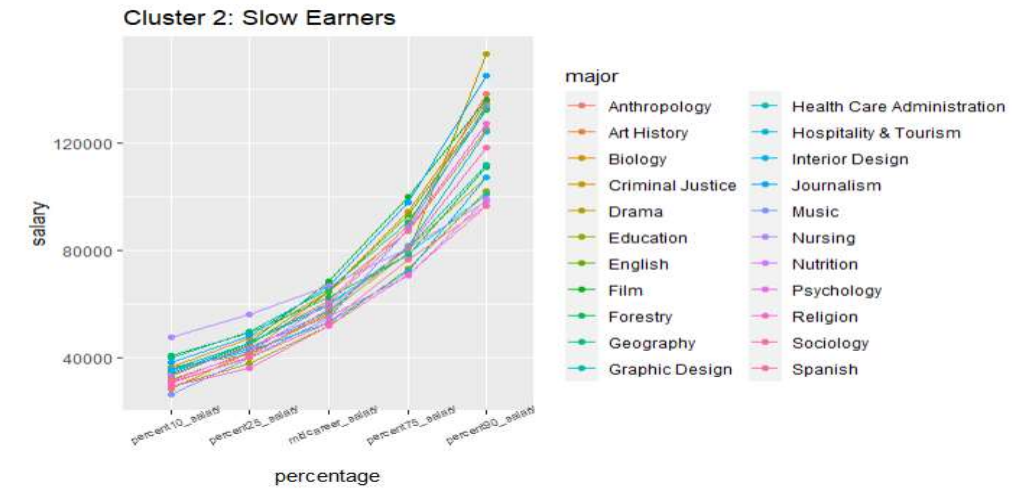
Cluster 2: The cluster 2 consists of majors that have the starting salary under $45000 and mid-career salary ranging from $50000 to $70000.

Cluster 3: The cluster 3 consists of majors that have the starting salary between $50000 to $65000 and mid-career salary ranging from $90000 to a little above $100000.

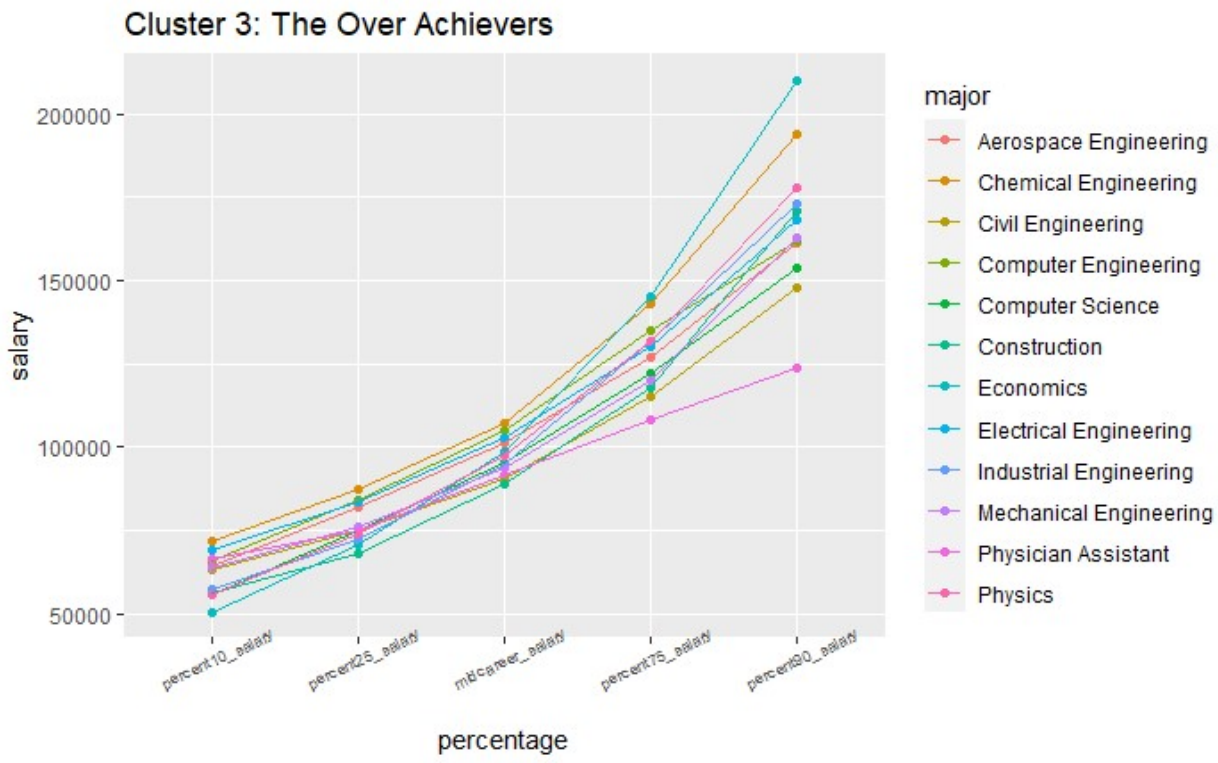
Plotting the cluster 1 majors based on the salary vs the salary percentage



Plotting the cluster 2 majors based on the salary vs the salary percentage.



Plotting the cluster 3 majors based on the salary vs the salary percentage



Conclusion:

Based on the above results, we can see that the resulting clusters show the patterns in starting salary vs mid-career salary along with growth by percentiles.

Depending on what type of salary you want like getting a high starting salary or high mid-career salary, you can select the major from the clusters.

For example, if one wants a high starting and mid-career salary, he/she wants to select a major from cluster 3.But remember that whatever the cluster a major is in, there are some other things that has to be taken into consideration like the college that you going to study, the city or village the college is present etc. As we can see that what ever the major you can join, the salary will start to increase as we can see them in the percentile growth.

Moreover, whatever the salary you may get, “Nothing Beats The Enjoyment One Will Get When He/She Likes The Job They Are Doing”.

LINK TO MY GITHUB ACCOUNT:

https://github.com/rmullapu/rmullapu_64060.git