BSMM8750-2-R-2023F Pred. Modeling & Decision Making

Section 2

Group - 6

Master of Management

Fall 2023

Submitted to

Dr. Brent Furneaux

Members	**Student ID**
Binz Baby Joseph	110094887
Sparsh Jindal	110093924
Jasleen Kaur Ahluwalia	110105482
Rifah Tasfia Amin	110099797

Table of Contents

1.Introduction

The primary focus of our project was how Odette School of Business can be successful in being ranked in the Financial Times (FT) top 100 ranking for master's in management. We refined our problem statement to what factors are driving the changes in the FT 100 ranking over the years. Specifically, identifying factors that cause ranking to either improve or decline. Understanding these factors will help us achieve our main objective: making Odette School of Business successful.

We used three approaches to accomplish this, the last two being particularly significant. Using multiple approaches enhanced the validity of our findings, thereby solidifying our insights. Our approaches are as follows:

1. Initially, we set out with the normal regression model (usually used to predict the rank based on the predictors), to understand the co-efficient of the attributes to supplement the Causal approach analysis.
2. Secondly, a fixed effect model was built focusing on changes in ranking occurring within a single year among universities, which is our dependent variable(Y), and the predictors (X) are also the changes in the factors within a single year. In other words, a causation analysis that tells us what is causing a change in ranking in a shorter term.
3. Finally, we constructed another fixed effect model, which has the same dependent variable: the changes in ranking taking place in 1 year. Meanwhile, the independent variables (X) were the changes in attributes occurring within two years. In other words, a broader window is used to consider the temporal order of events.

We completed our project implementation using two different tools for better learning purposes. In the first half, we used Microsoft Excel for the descriptive statistics and chart creation. In the second half, Google Colab Notebooks was used for model building, evaluation, and dataset manipulation.

2. Data Collection (2018-2023)

Initially we had a dataset ranging from 2018 to 2022, with different attributes. For data collection, we collected the data from the FT 100 ranking site for the year 2023 (*Business School Rankings* 2023). We based our analysis on the 2023 ranking methodology, as it was the most current information at the time of our study. Then we collected available information for the years 2018-2022 based on the ranking methodology (columns or attributes) of 2023, which was not available in the initial dataset; this was done from the same FT 100 ranking site (*Business School Rankings* 2023).

3. Data Preprocessing

3.1 Data Integration

Going forward all data was merged into one single sheet using Excel. During integration we found out that the data for the year 2018 had many missing features; entire columns were not available since most of those features were not used during the 2018 ranking. Hence, we excluded the 2018 data from the integrated sheet. Conversely, the 2019 data had only one missing feature which was 'Aims achieved (%)'. It is discussed in the 'Missing Value Handling' section. As mentioned earlier, the base of our columns considered was 2023. Some columns like 'Alumni network rank', 'Carbon footprint rank', and 'ESG and net zero teaching rank' were removed due to their presence in only one year: 2023 (*Business School Rankings* 2023). The final common columns in our integrated sheet are as shown in the Appendix. (figure 1)

3.2 Data Cleaning

The data cleaning process done using Excel was instrumental in improving our dataset. This phase was marked by a preliminary summary analysis aimed at identifying errors, such as presence of non-numeric values in numeric columns and missing value detection.

3.2.1 Removing or correcting erroneous data

In our dataset we found some alpha numeric values like the University Carlos III de Madrid had 'Female Faculty (%)' as "46 ††" in the year 2020 this was corrected. For alpha numeric values, whose meaning could not be interpreted were removed. There were only 2 rows that were non interpretable, so it did not impact our analysis that much.

3.2.2 Missing value handling

As mentioned earlier, 'Aims achieved (%)' column was missing in the year 2019. However, considering our specific scenario, which aimed to understand the factors influencing changes in university rankings, we only considered the universities which were ranked in all 5 years consistently because we were more concerned about the change in ranking and what is driving that change, this will be explained in the 'Final Dataset Used' section. So, for the 'Aims achieved (%)' column in 2019 we imputed it with the mean grouped by each university, indicating, we exclusively considered common universities. For example, let us take 'University A.' In the 'Aims achieved (%)' column, University A had recorded values for the years 2020 to 2023 as 86, 87, 88, and 87, respectively. In this case, the mean value of 87 was imputed for the year 2019 for that university and the same method was applied to each University. This was executed using pivot tables in Excel and VLOOKUP function taking the school's name and year as index. Additionally, we verified that no school was ranked twice in the same year by conducting duplicate detection checks for each year. For the missing value handling part, we discovered a few other missing values, like 6 of the rows in our integrated sheet had missing value in columns 'Internships (%)' and 'Avg_Course_Length(Months)' which accounted for 1% of the entire dataset. These universities were not consistent in being ranked in all 5 years so anyway these universities were not considered in our final dataset. So, these were also removed. The pivot table used for aims achieved is as shown in Appendix (figure 2)

The formula for the VLOOKUP used was "=VLOOKUP(lookupvalue=SchoolNamecolumn, table_array = pivottablearea, col_index_num = the column number of aims achieved, FALSE)"

3.2.3 Outlier Treatment

In our specific scenario, we are concerned with the change in ranking and what is causing that change, particularly those that lead to notable change in rank. Therefore, it is essential for us to keep the outliers in our analysis.

3.3 Data Transformation

In our data we only had to transform one column; 'Employed at three months (%)' highlighted in Figure 1. The data initially was in the following format: 100(92).

For example, for the HEC Paris university the value for 'Employed at three months (%)' is 100(92) This represents the percentage of the most recent graduating class that found employment within three months. The number inside the bracket indicates the data the school provided information for (Business School Rankings 2023). To better utilize this in our analysis we converted the column for all universities as follows:

New employed at 3 months = Number outside the bracket * (Number inside the bracket / 100)

In our case the calculation would be 100 * (92/100) = 92% This approach ensures fairness in our analysis for example, a different university which provided complete data and out of that 100% only 92% got a job so our value would be 92(100) This standardization method allows for a more equitable comparison between universities.

3.4 Scaling and Normalization

In causation analysis, since we are more interested in the causal relationship between the independent variables and the dependent variable, we did not scale or normalize the variables. Even in our normal prediction approach we were only concerned with the directions of the coefficients of the model so scaling was not necessary there as well.

4. Data Exploration and Analysis

This section involves looking into the data to identify patterns and relationships to gain valuable insights. This includes the measure of central tendency, measure of spread, the distribution, and the correlation matrix. The entire data exploration of the original dataset was done using excel. Let us look at some of the major insights found out:

4.1 Measure of Central Tendency

- When looking at the overall average in the 5 years most of the factors that were increasing over the years were in direct control of the universities (*Business School Rankings* 2023). Refer (figure 3) and (figure 4) in the Appendix.

4.2 Measure of Spread

4.2.1 Range

- Wide variation was seen in international board representation showing that some universities have limited international representation, while others have more.
- The range of Faculty of Doctorates (%) is around '38' with the minimum being 62% showing that most of the universities in this 5-year timespan is committed to this factor.

4.2.2 Standard Deviation and Variance

- Factors like international board, international work mobility rank, career progress rank, salary percentage increase, international faculty and international students exhibit high standard deviations. This indicates significant variation among universities concerning these factors (*Business School Rankings* 2023).
- However, factors like female faculty %, female student %, faculty with doctorates (%) and average course length have low variations when compared with universities in all 5 years (*Business School Rankings* 2023). Refer to the appendix (figure 5) for a sample of 'International board(%)' and 'Faculty with doctorates (%)'.
- The descriptive statistics of all attributes were generated using Microsoft Excel Data Analysis toolpak. This was done to learn the tools, as similar analyses can also be achieved using Python. While the charts and graphs were made using Microsoft Excel Pivot Charts and normal charts.

4.2.3 Distribution

To understand the distribution of the dataset, we plotted histograms for each of the 18 attributes, excluding the 'Location by primary campuses from our integrated dataset.

Some of the major ones are shown in the Appendix. One unique distribution found was that of Average Course Length which has a bimodal distribution. Refer to the Appendix (figure 6), (figure 7) and (figure 8)

4.3 Correlation

The correlation of all the 18 variables with the rank was generated, refer to figure 11 in appendix. This was like what was shown in the FT 100 ranking site. But the coefficients generated from our first approach were particularly useful in our analysis since it validated our findings.

5 Final Dataset Used

We used three different approaches using the same dataset in distinct ways. The data dictionary is provided in the appendix for better understanding.

1. **First Approach – Normal Prediction Model:**
   For our initial approach, which aimed to build a standard prediction model, we used the original cleaned dataset. This dataset included information on universities, ranks and the 18 attributes under consideration. However, to focus solely on identifying the factors affecting the ranks, we narrowed down our dataset, we considered only those universities which consistently received rank in all five years (2019-2023), excluding the year 2018, as mentioned earlier. So, the final dataset is a subset of the cleaned dataset ranging from the years 2019 to 2023. It consisted of 275 rows and 22 columns which includes Rank, School Name, Year, Location by primary campus, and the 18 independent variables. The name of the file is 'First_Approach_Data.csv.'
2. **Second Approach – Short term Causal Analysis:**
   In our second approach, focused on short term causal analysis, the dataset was generated using our first approach's cleaned dataset ('First_Approach_Data.csv' the one with 275 rows). Here, we calculated the change in rank (dependent variable 'y') for one year. Simultaneously, we calculated the changes in all 18 attributes (excluding 'Location by primary campus') for the same one-year period. It was done using a python script which is provided in the appendix. For instance, if we take 'HEC Paris' as an example, it had a rank of 2 in 2022, which improved to 1 in 2023. Therefore, the change in rank for this university for the year 2023-2022 is -1 (indicating improvement), while a positive value in the 'Rank Diff' column signifies a decline in rank. The attribute difference was also taken for the one-year difference; 2023-2022. We applied this calculation to each yearly difference: 2023-2022, 2022-2021, 2021-2020, and 2020-2019, respectively. The dataset generated is named as 'Second_Approach_Data.csv'
3. **Third Approach - Long-Term Causal Analysis (Broader Window):**
   Our third approach, focused on long term causal analysis using a broader window, also started with the original cleaned dataset used in the first approach ('First_Approach_Data.csv' the one with 275 rows). The rank difference part remains the same as mentioned in the second approach considering one year difference. However, the attributes or the 18 independent variables are considered over a two-year difference. For example, 'y' would remain same as above 2023-2022 while the independent variable change will be 2023-2021 that is a two-year difference. This will consider the temporal order of events since certain factors take some time to influence a change in rank and as we know the cause happens before the effect. This was also achieved through a python script provided in the appendix. The dataset generated is named as 'Third_Approach_Data.csv'.

6. Model Building and Evaluation

Since our analysis consisted of three different approaches let's go through all the approaches.

6.1 First Approach Normal Prediction Model

Considering our specific scenario, based on our data we built multiple regression models to predict the rank based on the 18 attributes, the dataset used was the 'First_Approach_Data.csv' we loaded the required libraries like 'sklearn' and 'pandas'. The dataset was divided into training and testing sets with an 80:20 split to validate the model's performance. Out of our analysis the three main models and their metrics are as follows:

Model Name	Accuracy
Random Forest Regression	86.9%
Bayesian Linear Regression	84.4%
ElasticNet Regression	84.4%

We conducted a multicollinearity assessment using the Variance Inflation Factor (VIF) method, and the results had no significant concerns, as none of the VIF values exceeded the threshold of 4. Details of this analysis and the corresponding code can be found in the appendix (9.3.1. Code for the First Approach - Normal Prediction Model #Multicollinearity Check)

The models provided some valuable insights. First, the Random Forest Regressor gave us the feature importance of the factors, but since our second and third approach are more reliable to find the important features, we took that for our reference. And for the other two models we checked the coefficients of Bayesian Linear Regression and ElasticNet Regression, here we are not much concerned about the value of the coefficients but rather the direction ('+' or '-') because a '-' direction indicates that those features improve rank. Since, dropping of our dependent variable 'y' means an improvement in our ranking. The code snippet for the coefficient is shown in the appendix (8.3.1. Code for the First Approach - Normal Prediction Model #Coefficients). Additionally, we cross-validated the direction of these coefficients (+ve or -ve) with our causal analysis findings. This can be referred in the Appendix (figure 14).

6.2 Causal Analysis

We employed a robust causal analysis to find which variables have a statistically significant impact and can be considered as a potential that causes a change in rank. Our data is panel data which contains a collection of data points for various universities across different years. These universities can exhibit variation over time, which can be analyzed. Hence, we used **fixed effects regression** which also considers for the unobserved time invariant variable like 'University Prestige'. It also eliminates bias from time-invariant confounders. Thus, we chose fixed effects regression in our causal analysis.

6.2.1 Second Approach Short-term Causal Analysis

In this approach, as mentioned earlier our dependent variable 'y' is the change in rank for one-year span and the predictors 'x' are also a change in the values in one-year span. This approach enables us to understand the immediate effects of various attributes on ranking shifts within a one-year period.

The code for implementing the second approach is provided in the appendix (9.3.2 Code for the Second Approach - Short Term Causal Analysis).

Initially, we ran the model for the dataset generated but found that the explanatory power of the model was very less. Thus, we filtered out the model for only taking the significant changes in ranking were we only considered the universities where a change in **rank is at least five or more,** this increased our explanatory power of the model as shown in the figure:

PanelOLS Estimation Summary

Dep. Variable:	Rank Diff	R-squared:	0.7277
Estimator:	PanelOLS	R-squared (Between):	0.2681
No. Observations:	129	R-squared (Within):	0.7277
Date:	Wed, Nov 29 2023	R-squared (Overall):	0.6198
Time:	20:49:02	Log-likelihood	-395.28
Cov. Estimator:	Clustered		
		F-statistic:	9.2070
Entities:	49	P-value	0.0000
Avg Obs:	2.6327	Distribution:	F(18,62)
Min Obs:	1.0000		
Max Obs:	4.0000	F-statistic (robust):	14.378
		P-value	0.0000
Time periods:	4	Distribution:	F(18,62)
Avg Obs:	32.250		
Min Obs:	31.000		
Max Obs:	35.000		

The R-square is 0.7277 which is a good accuracy, while the p-value is highly significant. The result of the model is as shown below:

Parameter Estimates

	Parameter	Std. Err.	T-stat	P-value	Lower CI	Upper CI
1yr Change in Female faculty (%)	-0.9188	0.3364	-2.7315	0.0082	-1.5912	-0.2464
1yr Change in Value for money rank	0.2324	0.1291	1.7997	0.0768	-0.0257	0.4904
1yr Change in International course experience rank	-0.0451	0.0904	-0.4992	0.6194	-0.2258	0.1356
1yr Change in International board (%)	0.0904	0.0761	1.1878	0.2395	-0.0618	0.2426
1yr Change in Faculty with doctorates (%)	-0.4898	0.3504	-1.3978	0.1672	-1.1903	0.2107
1yr Change in Nwemployedat3months	-0.1965	0.0700	-2.8061	0.0067	-0.3365	-0.0565
1yr Change in Women on board (%)	0.1319	0.1397	0.9443	0.3487	-0.1473	0.4111
1yr Change in Weighted salary (US$)	-0.0008	0.0002	-4.2901	0.0001	-0.0012	-0.0005
1yr Change in Career progress rank	0.1350	0.0459	2.9412	0.0046	0.0433	0.2268
1yr Change in Female students (%)	0.3272	0.1118	2.9255	0.0048	0.1036	0.5508
1yr Change in Careers service rank	0.1325	0.0566	2.3410	0.0225	0.0194	0.2457
1yr Change in Salary percentage increase	-0.2124	0.1129	-1.8818	0.0646	-0.4380	0.0132
1yr Change in International students (%)	0.1978	0.0842	2.3494	0.0220	0.0295	0.3662
1yr Change in International work mobility rank	0.1728	0.1194	1.4481	0.1526	-0.0657	0.4114
1yr Change in International faculty (%)	-0.8062	0.3245	-2.4845	0.0157	-1.4548	-0.1575
1yr Change in Aims achieved (%)	-0.9958	0.4663	-2.1353	0.0367	-1.9280	-0.0636
1yr Change in Internships(%)	-0.1030	0.0399	-2.5814	0.0122	-0.1828	-0.0232
1yr Change in Avg_Course_Length(Months)	1.1097	0.4425	2.5077	0.0148	0.2251	1.9943

Note that green indicates -ve sign which indicates that the rank difference is decreasing which is a good thing. Example: 'University A' had rank 5 in 2023 and in 2022 it had 7th rank. The difference is y=2023-2022=-2. Now if female faculty increases during this period of 2022 to 2023 by 1% then as shown the ranking difference drops further by -0.91 so instead of -2 it is -2.91 close to -3. This will result 'University A' to be ranked 4th in 2023 instead of 5th.

The red indicates +ve sign which is opposite of the previous scenario. Thus, it is associated with the worsening of the rank. However, rank variables like career progress rank and careers service rank though

+ve are still associated with improvement in rank because our lower rank in these variables would cause a lower rank in FT 100. Thus, an improvement in rank. All these causations show an average association with the rank difference, hence, should be looked at with caution and should only be applied with thorough domain knowledge. Thus, with the right domain knowledge we can filter out the useful insights and take them as an inspiration to our applications in real life.

We also filtered the data further to include a rank change of at least 10 to better understand the significant factors impacting significant changes. This increased the accuracy of the model to **95%**. The model results are shown in the Appendix (figure 9).

6.2.2 Third Approach - Long-Term Causal Analysis (Broader Window):

In this approach, like the second approach the dependent variable 'y' is the change in rank in one year span. While the independent variables 'x' is the change in their **respective values in two years**, which considers a **broader window** thus considering the temporal order of events. This approach enables us to understand that some of the factors do not show immediate effect on the rank. It also gives us an insight into what the universities can focus on in the longer run.

The code for implementing the third approach is shown in the appendix (9.3.3 Code for the Third Approach - Long Term Causal Analysis).

In this approach like the previous one, we filtered the dataset to consider a rank change of at least five to improve the explanatory power of the model. The fixed effect regression model results are as shown below:

```
                          PanelOLS Estimation Summary
================================================================================
Dep. Variable:              Rank Diff   R-squared:                        0.5313
Estimator:                   PanelOLS   R-squared (Between):             -0.2011
No. Observations:                  98   R-squared (Within):               0.5313
Date:                Wed, Nov 29 2023   R-squared (Overall):              0.2351
Time:                        20:51:57   Log-likelihood                   -315.61
Cov. Estimator:             Clustered
                                        F-statistic:                      1.9525
Entities:                          49   P-value                           0.0494
Avg Obs:                       2.0000   Distribution:                    F(18,31)
Min Obs:                       1.0000
Max Obs:                       3.0000   F-statistic (robust):             13.767
                                        P-value                           0.0000
Time periods:                       3   Distribution:                    F(18,31)
Avg Obs:                       32.667
Min Obs:                       31.000
Max Obs:                       35.000
```

Here even though the R-squared value is 53% and the p-value is 0.04 it is below the threshold of 0.05. Apart from that in Causation Analysis more than the accuracy of the model the p-value which shows the explanatory power of the model is more important. Hence, we are not much concerned with the accuracy of the model but rather how significant the results are:

Parameter Estimates

	Parameter	Std. Err.	T-stat	P-value	Lower CI	Upper CI
2yr Change in Female faculty (%)	-0.4472	0.5617	-0.7961	0.4320	-1.5927	0.6983
2yr Change in Value for money rank	0.2443	0.1592	1.5345	0.1351	-0.0804	0.5691
2yr Change in International course experience rank	0.2654	0.1580	1.6801	0.1030	-0.0568	0.5876
2yr Change in International board (%)	-0.0647	0.0766	-0.8442	0.4050	-0.2210	0.0916
2yr Change in Faculty with doctorates (%)	-0.0715	0.4291	-0.1667	0.8687	-0.9466	0.8036
2yr Change in Nwemployedat3months	-0.3672	0.1244	-2.9516	0.0060	-0.6209	-0.1135
2yr Change in Women on board (%)	0.0769	0.2064	0.3723	0.7122	-0.3441	0.4978
2yr Change in Weighted salary (US$)	-0.0006	0.0003	-2.4612	0.0196	-0.0012	-0.0001
2yr Change in Career progress rank	0.1875	0.0666	2.8150	0.0084	0.0517	0.3233
2yr Change in Female students (%)	-0.4820	0.2010	-2.3978	0.0227	-0.8920	-0.0720
2yr Change in Careers service rank	0.0381	0.1600	0.2383	0.8132	-0.2883	0.3645
2yr Change in Salary percentage increase	0.2166	0.1577	1.3739	0.1793	-0.1049	0.5381
2yr Change in International students (%)	-0.2594	0.2619	-0.9903	0.3297	-0.7936	0.2748
2yr Change in International work mobility rank	0.1181	0.2189	0.5396	0.5933	-0.3283	0.5644
2yr Change in International faculty (%)	0.0449	0.3213	0.1399	0.8897	-0.6103	0.7002
2yr Change in Aims achieved (%)	-0.5006	1.0054	-0.4979	0.6221	-2.5512	1.5500
2yr Change in Internships(%)	0.2132	0.0913	2.3349	0.0262	0.0270	0.3995
2yr Change in Avg_Course_Length(Months)	2.1469	0.8828	2.4320	0.0210	0.3465	3.9473

When we consider the long-term approach, it has some common factors with the short-term approach. While some key points here are that universities that focus on gender diversity in the long run are associated with an improvement in the ranking regardless of what other universities are doing.

Employed at 3 months has the strongest claim with an association towards improvement of ranking. Which brings up that we should also consider the p value, the smaller the p value the stronger the claim.

Average course length when increasing still shows an adverse effect on rank. The reason could be that even if the course length keeps on increasing eventually the course may be inefficient.

We also performed sensitivity analysis to validate the model explanatory power. Refer Appendix for code (9.3.3 Code for the Third Approach - Long Term Causal Analysis #Sensitivity Analysis)

7. Final Insight and Conclusion

- The results from the Short-Term Causal Analysis: The results are sorted from the strongest claim to the weakest claim but all of them are statistically significant. If we investigate the figure, it suggests that if Weighted salary (US$) increases by 1$, it improves the rank by 0.0008, Similarly if Weighted salary (US$) increases by 1000(US$) then the Rank improves by 0.8 which can be interpreted as an improvement by 1 rank position.

Short Term Change Effects (1yr change)	
Attributes	**Parameter**
Weighted salary (US$)	-0.0008
Career progress rank	0.135
Employed at 3 months	-0.1965
Female faculty (%)	-0.9188
Avg_Course_Length(Months)	1.1097
International faculty (%)	-0.8062
Careers service rank	0.1325
Aims achieved (%)	-0.9958

- The results from the Long-Term Causal Analysis: The results from the Long-Term Causal Analysis can also be interpreted in the same way as the short term one. This is also sorted from the strongest claim to the weakest claim. Employed at 3 months being the strongest claim in the Long Term Analysis.

Long Term Change Effects (2yr change)	
Attributes	**Parameter**
Employed at 3 months	-0.3672
Weighted salary (US$)	-0.0006
Career progress rank	0.1875
Female students (%)	-0.482
Avg_Course_Length(Months)	2.1469

- The highlighted parameters are common in all three approaches except for Avg_Course_Length(Months), which is linked with an improvement in rank in our First Approach. Conversely, it is linked with the worsening of rank in our other two approaches. The reason being that as shown in the appendix (figure 12) and (figure 13). This indicates that a course duration of 17-20 months appears to be optimal, so keeping on increasing after 20 could also suggest inefficiencies in the program's structure. This means that we can look into the courses of these universities to understand what additional elements are they offering or how their curriculum is structured over a 17-20 month period.

Conclusion:

Through our analysis, we aimed to unravel the driving forces that are influencing the Financial Times Top 100 business school ranking. Our significant insights delved deeper into our investigation that steamed from robust data collection, and data processing offer a clearer understanding of the factors that form this prestigious ranking.

To conclude, our findings not only highlight the criterias that navigate the rankings of top business schools but also provide a data-driven overview. Ultimately, this analysis steers us back to our core goal of how Odette School of Business can be successful in being ranked in the Financial Times (FT) top 100 ranking for master’s in management.

8. References

Business School Rankings. FT.com. (2023). https://rankings.ft.com/rankings/2948/masters-in-management-2023

9. Appendix

9.1 Data Dictionary

1. **Aims achieved (%):** The extent to which an alumni achieved their goals for pursuing a master's degree. (Business School Rankings 2023)

2. **Avg_Course_Length(Months):** The average length of the master's program in months. (*Business School Rankings* 2023)

3. **Career progress rank:** A rank determined based on changes in the level of seniority and the size of the organization alumni are working for, measured from the time of course completion to the present. (*Business School Rankings* 2023)

4. **Careers service rank:** A rank reflecting the effectiveness of the service in supporting student recruitment, as rated by alumni. (*Business School Rankings* 2023)

5. **Faculty with doctorates (%):** The percentage of faculty members holding doctoral degrees. (*Business School Rankings* 2023)

6. **Female faculty (%):** The percentage of females in comparison to the whole faculty board. (*Business School Rankings* 2023)

7. **Female students (%):** The percentage of female students enrolled in the program. (*Business School Rankings* 2023)

8. **International board (%):** The percentage of the school advisory board whose citizenship differs from the school's location. (*Business School Rankings* 2023)

9. **International course experience rank:** A rank reflecting the international experience offered by the course, including exchanges and internships. (*Business School Rankings* 2023)

10. **International faculty (%):** The percentage of international faculty members. (*Business School Rankings* 2023)

11. **International students (%):** The percentage of international students in the program. (*Business School Rankings* 2023)

12. **International work mobility rank:** A rank calculated based on changes in the location of employment of alumni between course completion and today, considering alumni citizenship. (*Business School Rankings* 2023)

13. **Internships (%):** The percentage of the last completed class who had internships as part of the course. (*Business School Rankings* 2023)

14. **Nwemployedat3months:** new employed at 3 months = number outside bracket * (numberinsidebracket/100) The percentage of the most recent graduating class that found employment within three months of completing their course. (*Business School Rankings* 2023)

15. **Rank diff:** The difference in rank compared to the previous year, indicating whether the university has moved up or down. Negative means the rank has improved while positive means the rank has declined. (*Business School Rankings* 2023)

16. **Salary percentage increase:** The average difference in alumni salary between course completion and the present, considering both absolute and relative percentage increases. (*Business School Rankings* 2023)

17. **School Name:** The name of the university or school participating in the ranking. (*Business School Rankings* 2023)

18. **Value for money rank:** A calculated rank based on alumni salaries, tuition, and other costs, indicating the perceived value for money. (*Business School Rankings* 2023)

19. **Weighted salary (US$):** The average graduate salary three years after course completion, adjusted for salary variations between sectors, in US dollars. (*Business School Rankings* 2023)

20. **Women on board (%):** The percentage of women on the school advisory board. (*Business School Rankings* 2023)

21. **Year:** The year for which the data is recorded. (*Business School Rankings* 2023)

22. **In the second and third approaches we mentioned the attributes as "1yr change in" + factor and "2yr change in" + factor respectively the meaning of the attributes remains the same.**

9.2 Figures

Figure 1 – The basic columns in the dataset

Final columns
Rank
School Name
Female faculty (%)
Value for money rank
International course experience rank
International board (%)
Faculty with doctorates (%)
Nwemployedat3months
Women on board (%)
Location by primary campus
Weighted salary (US$)
Career progress rank
Female students (%)
Careers service rank
Salary percentage increase
International students (%)
International work mobility rank
International faculty (%)
Aims achieved (%)
Internships(%)
Avg_Course_Length(Months)
Year

Figure 2 – The pivot table used for imputing missing values.

Row Labels	Average of Aims achieved (%)
Aalto University	86.18
Alliance Manchester Business School	82.46
Antwerp Management School	84.84
Audencia	84.18
Bayes Business School (formerly Cass)	84.83
BI Norwegian Business School	87.36
Burgundy School of Business	83.25
Católica Lisbon School of Business and Economics	81.45
Cems Global Alliance	87.00
Copenhagen Business School	86.75
Corvinus University of Budapest	81.27
Cranfield School of Management	84.44

Figure 3 – Descriptive statistics Average over the years

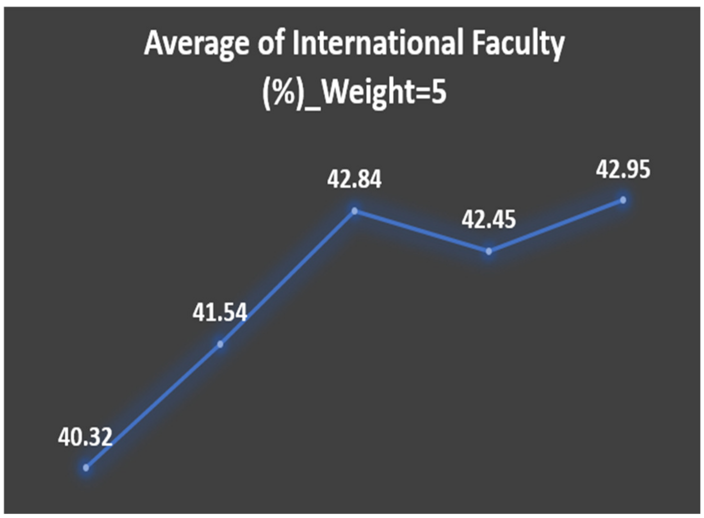

Figure 4 - Descriptive statistics Average over the years

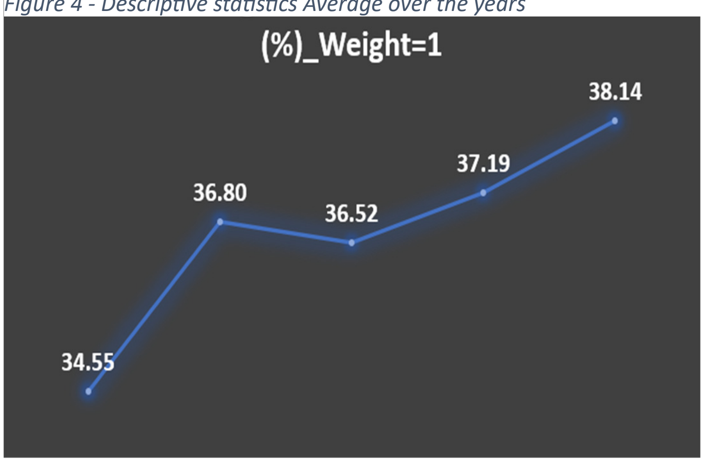

Figure 5 – The Descriptive Statistics

International board (%)		***Faculty with doctorates (%)***	
Mean	**42.70552147**	**Mean**	**93.50920245**
Standard Error	**1.254502761**	**Standard Error**	**0.343960116**
Median	**39**	**Median**	**96**
Mode	**0**	**Mode**	**100**
Standard Deviation	**27.74125158**	**Standard Deviation**	**7.606108496**
Sample Variance	**769.5770391**	**Sample Variance**	**57.85288645**
Kurtosis	**-0.912236435**	**Kurtosis**	**2.544230208**
Skewness	**0.300837059**	**Skewness**	**-1.593299038**
Range	**100**	**Range**	**38**
Minimum	**0**	**Minimum**	**62**
Maximum	**100**	**Maximum**	**100**

Figure 6 – Female Faculty(%) Distribution

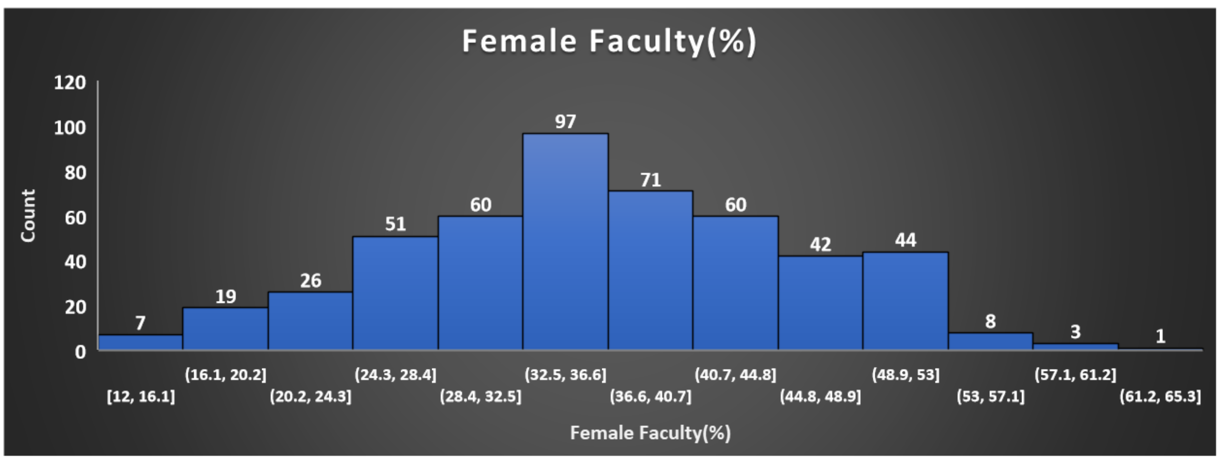

Figure 7 – Faculty with Doctorates(%) Distribution

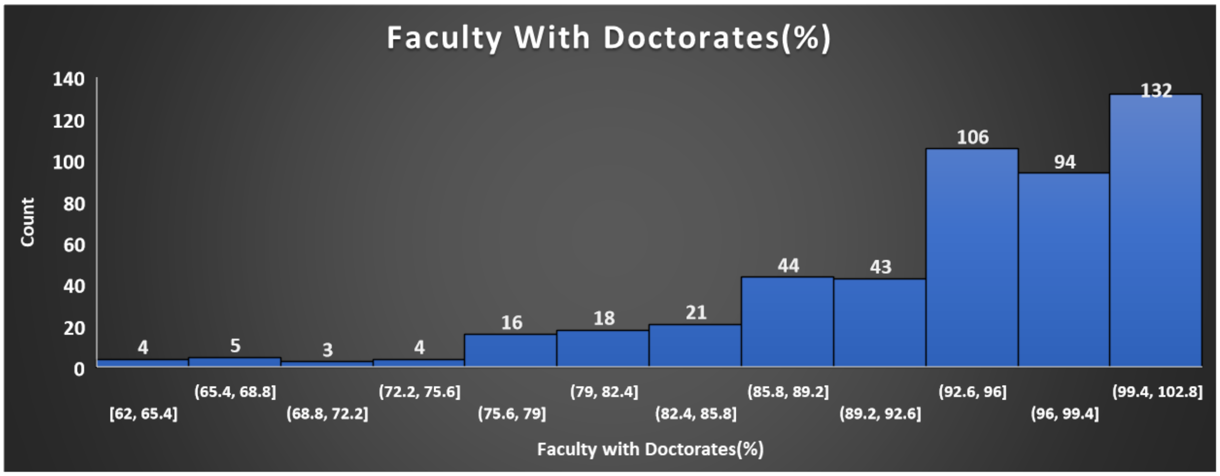

Figure 8 – Average of Course Length(Months) Distribution

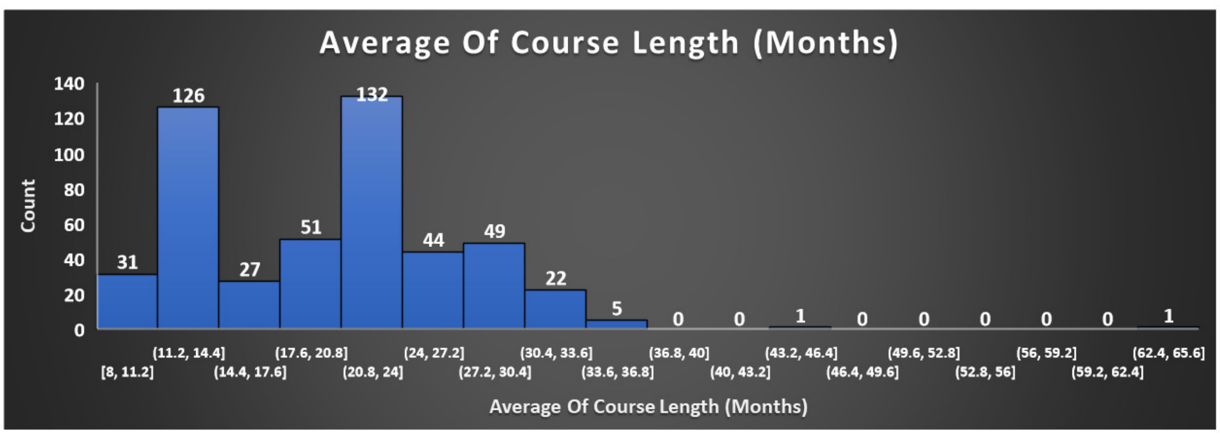

Figure 9 – For Second Approach Short Term considering the ranking changes of at least 10 or more

PanelOLS Estimation Summary

Dep. Variable:	Rank Diff	R-squared:	0.9598
Estimator:	PanelOLS	R-squared (Between):	-0.8414
No. Observations:	57	R-squared (Within):	0.9598
Date:	Wed, Nov 29 2023	R-squared (Overall):	-0.1700
Time:	20:49:25	Log-likelihood	-121.15
Cov. Estimator:	Clustered		
		F-statistic:	5.3100
Entities:	35	P-value	0.0586
Avg Obs:	1.6286	Distribution:	F(18,4)
Min Obs:	1.0000		
Max Obs:	3.0000	F-statistic (robust):	5.094e+15
		P-value	0.0000
Time periods:	4	Distribution:	F(18,4)
Avg Obs:	14.250		
Min Obs:	11.000		
Max Obs:	17.000		

Figure 10 – Results from the Second Approach when dataset was filtered with ranking changes of at least 10

Parameter Estimates

	Parameter	Std. Err.	T-stat	P-value	Lower CI	Upper CI
1yr Change in Female faculty (%)	-2.7772	0.8961	-3.0992	0.0362	-5.2652	-0.2892
1yr Change in Value for money rank	0.6982	0.4128	1.6916	0.1660	-0.4478	1.8443
1yr Change in International course experience rank	-0.0022	0.1237	-0.0176	0.9868	-0.3457	0.3414
1yr Change in International board (%)	-0.1125	0.1279	-0.8796	0.4287	-0.4676	0.2426
1yr Change in Faculty with doctorates (%)	0.3873	0.6648	0.5826	0.5914	-1.4584	2.2330
1yr Change in Nwemployedat3months	0.3568	0.1119	3.1895	0.0332	0.0462	0.6673
1yr Change in Women on board (%)	-1.0378	0.3329	-3.1178	0.0356	-1.9621	-0.1136
1yr Change in Weighted salary (US$)	-0.0012	0.0005	-2.4484	0.0706	-0.0025	0.0002
1yr Change in Career progress rank	0.1213	0.0729	1.6632	0.1716	-0.0812	0.3237
1yr Change in Female students (%)	0.4149	0.1825	2.2729	0.0855	-0.0919	0.9217
1yr Change in Careers service rank	0.2574	0.0724	3.5567	0.0237	0.0565	0.4584
1yr Change in Salary percentage increase	-0.0204	0.2266	-0.0898	0.9327	-0.6496	0.6089
1yr Change in International students (%)	0.2845	0.0734	3.8743	0.0179	0.0806	0.4884
1yr Change in International work mobility rank	-0.4205	0.2313	-1.8180	0.1432	-1.0627	0.2217
1yr Change in International faculty (%)	-1.7108	0.7422	-2.3052	0.0825	-3.7714	0.3497
1yr Change in Aims achieved (%)	-4.6571	0.8232	-5.6571	0.0048	-6.9428	-2.3714
1yr Change in Internships(%)	-0.5427	0.1304	-4.1624	0.0141	-0.9048	-0.1807
1yr Change in Avg_Course_Length(Months)	2.9649	1.9738	1.5021	0.2075	-2.5153	8.4451

Figure 11 – Correlation of the initial dataset

Female faculty (%)	0.194162
Value for money rank	0.071791
International course experience rank	0.386104
International board (%)	-0.107934
Faculty with doctorates (%)	-0.509386
Nwemployedat3months	-0.393009
Women on board (%)	-0.075627
Weighted salary (US$)	-0.758205
Career progress rank	0.063898
Female students (%)	0.074531
Careers service rank	0.438567
Salary percentage increase	-0.234883
International students (%)	-0.411235
International work mobility rank	0.552135
International faculty (%)	-0.362707
Aims achieved (%)	-0.709838
Internships(%)	-0.248931
Avg_Course_Length(Months)	0.029661

Figure 12 – Average of Rank is showing the lowest in the range of 17-20 Months of course.

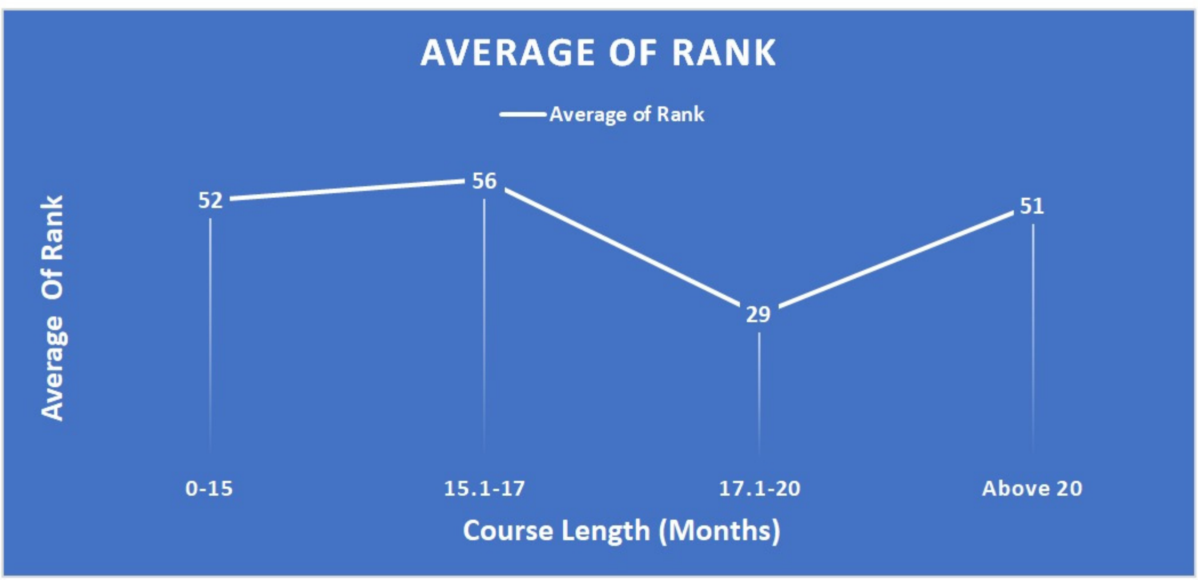

Figure 13 – Average course length distribution overall and in top 10 universities

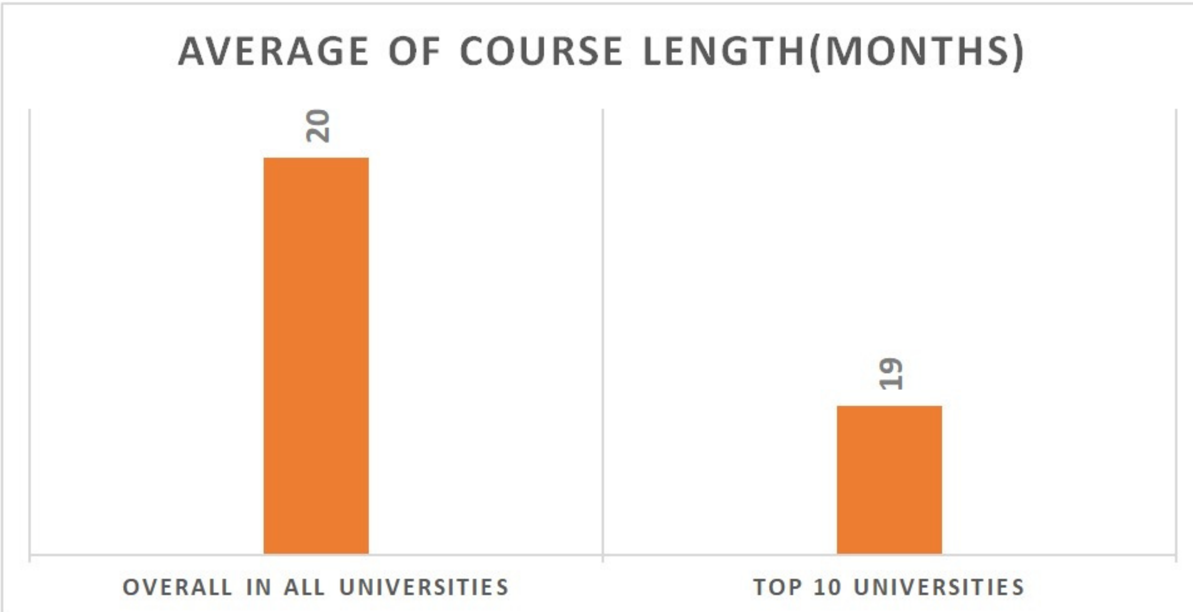

Figure 14 – Results of the Bayesian Linear Regression Coefficients

Column Name	Coefficient
Female faculty (%)	-0.674438
Value for money rank	0.082619
International course experience rank	0.197222
International board (%)	0.073485
Faculty with doctorates (%)	-1.042876
Nwemployedat3months	-0.121668
Women on board (%)	-0.107670
Weighted salary (US$)	-0.000546
Career progress rank	0.142825
Female students (%)	0.104736
Careers service rank	0.148620
Salary percentage increase	-0.393493
International students (%)	-0.063349
International work mobility rank	0.205374
International faculty (%)	-0.189960
Aims achieved (%)	-0.854718
Internships(%)	0.024590
Avg_Course_Length(Months)	-0.142564

9.3 The Python Script (Codes)

9.3.1. Code for the First Approach - Normal Prediction Model

 # Loading the libraries the first five are for Google Colab specifically

```
!pip install -U -q PyDrive
```

```
from pydrive.auth import GoogleAuth
from pydrive.drive import GoogleDrive
from google.colab import auth
from oauth2client.client import GoogleCredentials
import pandas as pd
from sklearn.model_selection import train_test_split
from sklearn.linear_model import LinearRegression
from sklearn.metrics import mean_squared_error, r2_score
from sklearn.model_selection import train_test_split
from sklearn import linear_model
from sklearn.preprocessing import PolynomialFeatures
from sklearn.linear_model import LinearRegression, Ridge, Lasso, ElasticNet, BayesianRidge
from sklearn.preprocessing import PolynomialFeatures
from sklearn.feature_selection import RFE
from sklearn.tree import DecisionTreeRegressor
from sklearn.ensemble import RandomForestRegressor
from sklearn.svm import SVR
from sklearn.metrics import mean_squared_error, r2_score
```

#For reading the csv first 5 is for google colab

```
auth.authenticate_user()
gauth = GoogleAuth()
gauth.credentials = GoogleCredentials.get_application_default()
drive = GoogleDrive(gauth)
file_id = 'First_Approach_data.csv'
df = pd.read_csv(file_id)
df
```

 # Defining the dependent variable (y) and independent variables (X)

y = df['Rank']

y

X = df[['Female faculty (%)','Value for money rank','International course experience rank','International board (%)','Faculty with doctorates (%)','Nwemployedat3months','Women on board (%)','Weighted salary (US$)','Career progress rank','Female students (%)','Careers service rank','Salary percentage increase','International students (%)','International work mobility rank','International faculty (%)','Aims achieved (%)','Internships(%)','Avg_Course_Length(Months)']]

X

 # Calculating the correlation matrix

correlation_matrix = X.corrwith(y)

print(correlation_matrix)

#Train test Split

X_train, X_test, y_train, y_test = train_test_split(X, y, test_size=0.2, random_state=42)

#Multicollinearity Check

#VIF multicollinearity

from statsmodels.stats.outliers_influence import variance_inflation_factor

#VIF for each feature

vif = pd.DataFrame()

vif["VIF Factor"] = [variance_inflation_factor(X.values, i) for i in range(X.shape[1])]

vif["features"] = X.columns

#VIF values

print(vif.round(1))

 # Models for the First Approach only the main ones are included

#Linear Regression

model = linear_model.LinearRegression()

model.fit(X_train, y_train)

y_pred = model.predict(X_test)

```
mse = mean_squared_error(y_test, y_pred)
r2 = r2_score(y_test, y_pred)
print(f"Mean Squared Error: {mse}")
print(f"R-squared (R2): {r2}")

# Random Forest Regression
rf_model = RandomForestRegressor()
rf_model.fit(X_train, y_train)
y_pred_rf = rf_model.predict(X_test)
mse = mean_squared_error(y_test, y_pred_rf)
r2 = r2_score(y_test, y_pred_rf)
print(f"Mean Squared Error: {mse}")
print(f"R-squared (R2): {r2}")
#for feature importance in Random Forest
rf_coeffs = dict(zip(X.columns, rf_model.feature_importances_))
print(rf_coeffs)

# ElasticNet Regression
elasticnet_model = ElasticNet(alpha=1.0, l1_ratio=0.5)
elasticnet_model.fit(X_train, y_train)
y_pred_elasticnet = elasticnet_model.predict(X_test)

mse = mean_squared_error(y_test, y_pred_elasticnet)
r2 = r2_score(y_test, y_pred_elasticnet)
print(f"Mean Squared Error: {mse}")
print(f"R-squared (R2): {r2}")

# Bayesian Linear Regression
bayesian_model = BayesianRidge()
```

```
bayesian_model.fit(X_train, y_train)

y_pred_bayesian = bayesian_model.predict(X_test)

mse = mean_squared_error(y_test, y_pred_bayesian)

r2 = r2_score(y_test, y_pred_bayesian)

print(f"Mean Squared Error: {mse}")

print(f"R-squared (R2): {r2}")


#Coefficients

bayesian_coeffs = dict(zip(X.columns, bayesian_model.coef_))

print(bayesian_coeffs)

# Convert the dictionary to a DataFrame

bayesian_coeffs_df = pd.DataFrame(bayesian_coeffs.items(), columns=['Column Name', 'Coefficient'])

# Display the DataFrame

print(bayesian_coeffs_df)
```

9.3.2 Code for the Second Approach - Short Term Causal Analysis

```
# Data Generation of second Approach

#First loading and reading the csv as done in the previous sections

!pip install -U -q PyDrive

from pydrive.auth import GoogleAuth

from pydrive.drive import GoogleDrive

from google.colab import auth

from oauth2client.client import GoogleCredentials

import pandas as pd

import pandas as pd

auth.authenticate_user()

gauth = GoogleAuth()

gauth.credentials = GoogleCredentials.get_application_default()

drive = GoogleDrive(gauth)
```

```
file_id = 'First_Approach_data.csv'
df = pd.read_csv(file_id)
df
```

 # Logic for the Data generation

```
# Defining the years for which we want to calculate differences
years = [2019, 2020, 2021, 2022, 2023]
# Creating an empty DataFrame to store the differences
diff_df = pd.DataFrame()

# Calculating the differences for each attribute between consecutive years
attribute_columns = [
    'Rank',
    'Female faculty (%)',
    'Value for money rank',
    'International course experience rank',
    'International board (%)',
    'Faculty with doctorates (%)',
    'Nwemployedat3months',
    'Women on board (%)',
    'Weighted salary (US$)',
    'Career progress rank',
    'Female students (%)',
    'Careers service rank',
    'Salary percentage increase',
    'International students (%)',
    'International work mobility rank',
    'International faculty (%)',
    'Aims achieved (%)',
```

```
    'Internships(%)',
    'Avg_Course_Length(Months)'
]
for i in range(1, len(years)):
    current_year = years[i]
    previous_year = years[i - 1]
    # Filtering the data for the current and previous years
    current_data = df[df['Year'] == current_year]
    previous_data = df[df['Year'] == previous_year]

    # Calculating the differences for each attribute by subtracting the previous year's value
    year_difference = f'{current_year}-{previous_year}'
    diff = current_data.set_index('School Name')[attribute_columns] - previous_data.set_index('School Name')[attribute_columns]
    diff.reset_index(inplace=True)
    diff['Year'] = year_difference
    # Appending the results to the diff_df DataFrame
    diff_df = pd.concat([diff_df, diff])
print(diff_df)
```

#Saving the dataframe to excel format and downloading the excel in google colab

```
# Saving the results to an Excel file
diff_df.to_excel('Second_Approach_data.xlsx', index=False)
#Downloading
from google.colab import files
# Saving the excel file
files.download('attribute_differences.xlsx')
```

The Actual Short-Term Causal Approach using the data generated above the earlier file was converted to csv before reading

#Importing Libraries and installing linear models first five is for google colab

```
!pip install -U -q PyDrive

from pydrive.auth import GoogleAuth

from pydrive.drive import GoogleDrive

from google.colab import auth

from oauth2client.client import GoogleCredentials

!pip install linearmodels

import pandas as pd

import matplotlib.pyplot as plt

import seaborn as sns

from linearmodels.panel import PanelOLS

import statsmodels.api as sm
```

 # Reading the csv file first five for google colab specific

```
auth.authenticate_user()

gauth = GoogleAuth()

gauth.credentials = GoogleCredentials.get_application_default()

drive = GoogleDrive(gauth)

file_id = ‘Second_Approach_data.csv’

imm_df_n = pd.read_csv(file_id)

imm_df_n
```

#Actual Logic

```
# Renaming 'Rank diff' to 'Rank Diff'

imm_df_n.rename(columns={'Rank diff': 'Rank Diff'}, inplace=True)

# Renaming 'Year' to 'Year Diff'
```

```
imm_df_n.rename(columns={'Year': 'Year Diff'}, inplace=True)

imm_df_n
```

#Filtering for difference greater than or equal to 5

```
imm_df = imm_df_n[imm_df_n['Rank Diff'].abs() >= 5]

imm_df
```

 # The Fixed Effect Regression Model

```
# Creating a single year identifier by taking the first year in the 'Year Diff' string

imm_df['Year'] = imm_df['Year Diff'].str.split('-').str[0].astype(int)

# Setting the index to be the entity identifier (School Name) and the new time identifier (Year)

imm_df = imm_df.set_index(['School Name', 'Year'])

# The dependent variable is 'Rank Diff' and the rest are independent variables

imm_df_y = imm_df['Rank Diff']

imm_df_X = imm_df.drop(columns=['Rank Diff', 'Year Diff'])  # Drop the original 'Year Diff' column

# Renaming the independent variables with '1yr Change in' prefix

imm_df_X = imm_df_X.rename(columns=lambda x: '1yr Change in ' + x)

# Creating the fixed effects model

model = PanelOLS(imm_df_y, imm_df_X, entity_effects=True)

# Fitting the model

fitted_model = model.fit(cov_type='clustered', cluster_entity=True)

# Output of the summary of the model

print(fitted_model.summary)
```

#Multicollinearity Check

```
#VIF multicollinearity

from statsmodels.stats.outliers_influence import variance_inflation_factor

# Calculating VIF for each feature
```

```
vif = pd.DataFrame()

vif["VIF Factor"] = [variance_inflation_factor(imm_df_x.values, i) for i in range(imm_df_x.shape[1])]

vif["features"] = imm_df_x.columns

# Inspecting VIF values

print(vif.round(1))
```

 # Converting the Significant Results into a DataFrame

```
table = fitted_model.summary.tables[1]

# Converting the table to a DataFrame

df = pd.DataFrame(table.data[1:], columns=table.data[0])

# Converting the numerical columns from string to float

df[['Parameter', 'Std. Err.', 'T-stat', 'P-value', 'Lower CI', 'Upper CI']] = df[['Parameter', 'Std. Err.', 'T-stat', 'P-value', 'Lower CI', 'Upper CI']].astype(float)

# Setting a threshold for significance, usually 0.05 or 0.01

significance_level = 0.05

# Filtering for significant factors only

significant_factors_df_short = df[df['P-value'] < significance_level]

significant_factors_df_short
```

#Exporting to Excel

```
# Exporting to Excel

significant_factors_df_short.to_excel('shorttermeffects.xlsx', index=False)

from google.colab import files

# Saving the excel file

files.download('shorttermeffects.xlsx')
```

#Filtering for difference greater than or equal to 10

```
imm_df_10 = imm_df[imm_df['Rank Diff'].abs() >= 10]
```

imm_df_10

#Model Building

```
# The dependent variable is 'Rank Diff' and the rest are independent variables
imm_df_10_y = imm_df_10['Rank Diff']
imm_df_10_X = imm_df_10.drop(columns=['Rank Diff', 'Year Diff'])  # Drop the original 'Year Diff' column
# Renaming the independent variables with '1yr Change in' prefix
imm_df_10_X = imm_df_10_X.rename(columns=lambda x: '1yr Change in ' + x)
# Creating the fixed effects model
model = PanelOLS(imm_df_10_y, imm_df_10_X, entity_effects=True)
# Fitting the model
fitted_model = model.fit(cov_type='clustered', cluster_entity=True)
# Output of the summary of the model
print(fitted_model.summary)
```

9.3.3 Code for the Third Approach- Long Term Causal Analysis

 # Libraries and reading the csv

```
!pip install -U -q PyDrive
!pip install linearmodels
from pydrive.auth import GoogleAuth
from pydrive.drive import GoogleDrive
from google.colab import auth
from oauth2client.client import GoogleCredentials
import pandas as pd
import matplotlib.pyplot as plt
import seaborn as sns
from linearmodels.panel import PanelOLS
```

```
import statsmodels.api as sm

#THIRD APPROACH STARTS HERE LONG-TERM ONE
broaderwindow = 'Third_Approach_data.csv'
broad_df_normal = pd.read_csv(broaderwindow)
broad_df_normal

# Filtering the dataset for significant changes in rank
broad_df = broad_df_normal[abs(broad_df_normal['Rank Diff']) >= 5]

#Actual logic, the order the columns is matched with the second approach
# Defining the correct order of the independent variables as specified
ordered_columns_diff = [
    'Female faculty (%) Diff',
    'Value for money rank Diff',
    'International course experience rank Diff',
    'International board (%) Diff',
    'Faculty with doctorates (%) Diff',
    'Nwemployedat3months Diff',
    'Women on board (%) Diff',
    'Weighted salary (US$) Diff',
    'Career progress rank Diff',
    'Female students (%) Diff',
    'Careers service rank Diff',
    'Salary percentage increase Diff',
    'International students (%) Diff',
    'International work mobility rank Diff',
    'International faculty (%) Diff',
```

```
    'Aims achieved (%) Diff',
    'Internships(%) Diff',
    'Avg_Course_Length(Months) Diff'
]
ordered_columns = [col.replace(' Diff', '') for col in ordered_columns_diff]
# Creating a single year identifier by taking the first year in the 'Year Diff' string
broad_df['Year'] = broad_df['Year Diff'].str.split('-').str[0].astype(int)
# Setting the index to be the entity identifier (School Name) and the new time identifier (Year)
broad_df = broad_df.set_index(['School Name', 'Year'])
broad_y = broad_df['Rank Diff']

# Dropping the 'Year Diff' column as we won't be using it in the model
broad_X = broad_df.drop(columns=['Rank Diff', 'Year Diff'])
# Reordering the columns based on the specified order with 'Diff'
broad_X = broad_X[ordered_columns_diff]
# Renaming the independent variables by removing ' Diff' and adding '2yr Change in' prefix
rename_mapping = {old: '2yr Change in ' + new for old, new in zip(ordered_columns_diff,
ordered_columns)}
broad_X = broad_X.rename(columns=rename_mapping)
# Creating the fixed effects model for the significant changes dataset
model_significant = PanelOLS(broad_y, broad_X, entity_effects=True)
# Fitting the model
fitted_model_significant = model_significant.fit(cov_type='clustered', cluster_entity=True)
# Output of the summary of the model for significant changes
print(fitted_model_significant.summary)
```

 # Sensitivity Analysis

```
# Defining a list of the predictors
all_predictors = [
```

```
    'Aims achieved (%) Diff', 'Avg_Course_Length(Months) Diff', 'Career progress rank Diff',
    'Careers service rank Diff', 'Faculty with doctorates (%) Diff', 'Female faculty (%) Diff',
    'Female students (%) Diff', 'International board (%) Diff', 'International course experience rank Diff',
    'International faculty (%) Diff', 'International students (%) Diff', 'International work mobility rank Diff',
    'Internships(%) Diff', 'Nwemployedat3months Diff', 'Salary percentage increase Diff',
    'Value for money rank Diff', 'Weighted salary (US$) Diff', 'Women on board (%) Diff'
]

# Defining the function to run the fixed effects model
def run_fixed_effects(broad_y, broad_X, entity_effects=True, cov_type='clustered', cluster_entity=True):
    model = PanelOLS(broad_y, broad_X, entity_effects=entity_effects)
    fitted_model = model.fit(cov_type=cov_type, cluster_entity=cluster_entity)
    return fitted_model

# Running the sensitivity analysis by excluding predictors one at a time based on p-values
for predictor in all_predictors:
    print(f"Running model without {predictor}")
    X_sensitivity = broad_df.drop(columns=[predictor, 'Rank Diff'])
    y_sensitivity = broad_df['Rank Diff']
    model_sensitivity = run_fixed_effects(y_sensitivity, X_sensitivity)
    print(model_sensitivity.summary)
```

 # Converting the significant results into a dataframe

```
# Here we are using for saving the data frame
significant_factors = fitted_model_significant.summary.tables[1]  # Table 1 usually contains the results
significant_factors_df = pd.DataFrame(significant_factors.data[1:], columns=significant_factors.data[0])
# Converting numerical columns from string to float
significant_factors_df[['Parameter', 'Std. Err.', 'T-stat', 'P-value', 'Lower CI', 'Upper CI']] =
significant_factors_df[['Parameter', 'Std. Err.', 'T-stat', 'P-value', 'Lower CI', 'Upper CI']].astype(float)
```

```
# Filtering only significant factors
significant_threshold = 0.05
significant_factors_df = significant_factors_df[significant_factors_df['P-value'] < significant_threshold]
significant_factors_df
```

 # Saving the DataFrame to Excel

```
# Saving the DataFrame to an Excel file
significant_factors_df.to_excel('significant_factors.xlsx', index=False)
# Downloading the excel file
from google.colab import files

# Saving the excel file
files.download('significant_factors.xlsx')
```