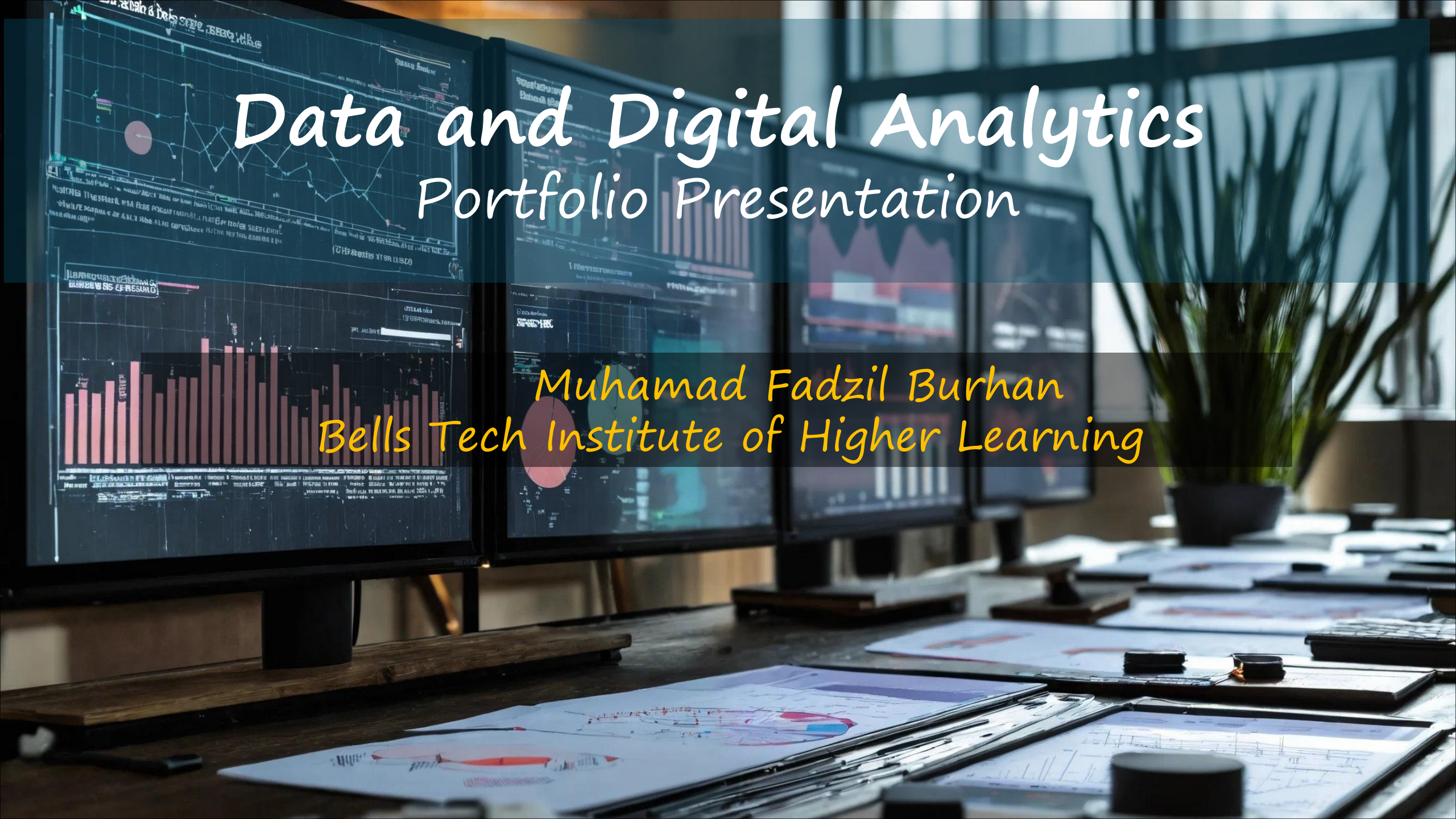


# Data and Digital Analytics

## Portfolio Presentation

Muhamad Fadzil Burhan
Bells Tech Institute of Higher Learning

**Portfolio Assignment and Presentation**

In this portfolio assignment, we will explore the college dataset containing information about various universities, with a focus on how a school counsellor can use statistical summaries, statistical analyses and regression analysis to provide recommendations. This portfolio assignment comprises 10 activities, each designed to cover different aspects of descriptive, diagnostic, predictive, and prescriptive analytics.

The candidate is able to use tools and techniques taught in Excel to answer the following activities. Include cross tables, graphs and charts to support your answer. A clear explanation must be provided to justify your answer.

**Deliverable:**

The candidate deliverable will be the successful completion of all the 10 activities along with their documentation in PowerPoint as well as a presentation of the same with the instructor.

**Activities for Portfolio Presentation:**

1) Introduce the college data set and calculate summary statistics (mean, median, standard deviation) for variables like SAT scores, admission rates, and tuition fees.
2) Use Pivot table to create a table and pivot chart (pie chart) of average (summation) of faculty salary_avg, across different regions.
3) Use PowerBI to visualize the average SAT score across different regions and break down by Control (private/public) school.
4) Use the PowerBI report for Q3, create a dashboard consisting of 4 visuals (for each visual add a filter for each region). Add cross filtering.
5) Create an SQL query to select only the columns (faculty_salary, region, control), filter for only “Private” control and aggregate the average faculty_salary for each region.
6) Connect the MySQL workbench using PowerQuery to PowerBI. You will need to create a new table and export the results of Q5 into the new table. Create a pie chart using this new table.
7) Using Orange data mining tool, you will create a regression model to predict sat_avg using the following input columns (admission_rate, under grads, tuition, faculty_salary_avg, loan_default_rate, median_debt). Display the regression metrics Use prescriptive analytics to recommend how an institution should allocate financial aid based on factors like admission rates and loan default rates.
8) Based on the previous regression model in Q7, use the Orange Rank widget and identify the 3 most important features that can be used to predict SAT_avg.

Data and Digital Analytics
Portfolio Presentation Instructional Booklet

BELLS INSTITUTE OF HIGHER LEARNING

9) Continue Based on the previous regression model in Q7, use the Data Table widget and view the coefficients for the regression model. Identify the features which must be increased and those that must be decreased in order to achieve better sat_avg scores.
10) Using Orange determine if there is a significant different between faculty_avg_salary for control (private/public), using hypothesis testing in the Boxplot widget.

1) Introduce the colleae data set and calculate summaru statistics (mean. median, standard deviation) for variables like SAT scores, admission rates, and tuition fees.

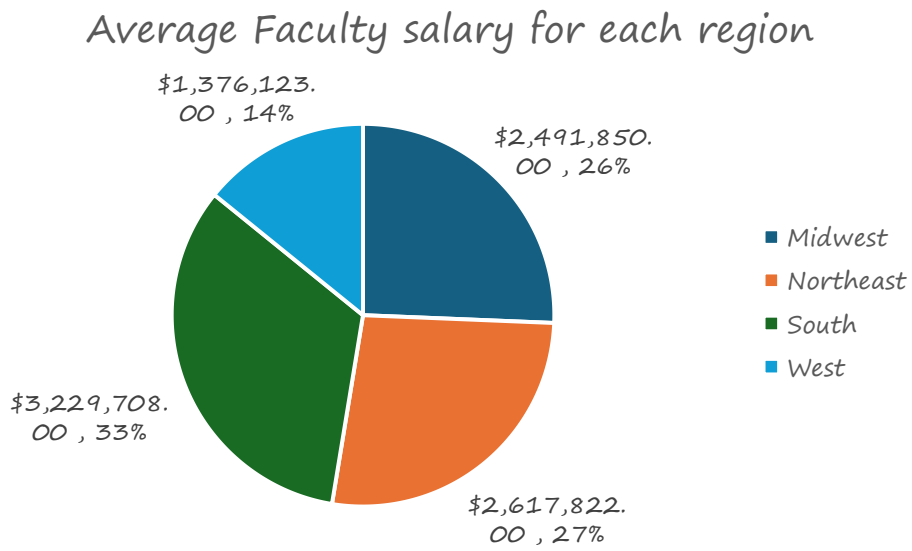
| *SAT score* | |
|---|---|
| Mean | 1059.512618 |
| Standard Error | 3.73744676 |
| Median | 1040 |
| Mode | 1050 |
| Standard Deviation | 133.0866955 |
| Sample Variance | 17712.06851 |
| Kurtosis | 1.139584563 |
| Skewness | 0.844072643 |
| Range | 825 |
| Minimum | 720 |
| Maximum | 1545 |
| Sum | 1343462 |
| Count | 1268 |

| *admission_rate* | |
|---|---|
| Mean | 0.650142396 |
| Standard Error | 0.005395029 |
| Median | 0.6687 |
| Mode | 1 |
| Standard Deviation | 0.192187246 |
| Sample Variance | 0.036935937 |
| Kurtosis | 0.158923733 |
| Skewness | -0.588255985 |
| Range | 0.9491 |
| Minimum | 0.0509 |
| Maximum | 1 |
| Sum | 825.0307 |
| Count | 1269 |

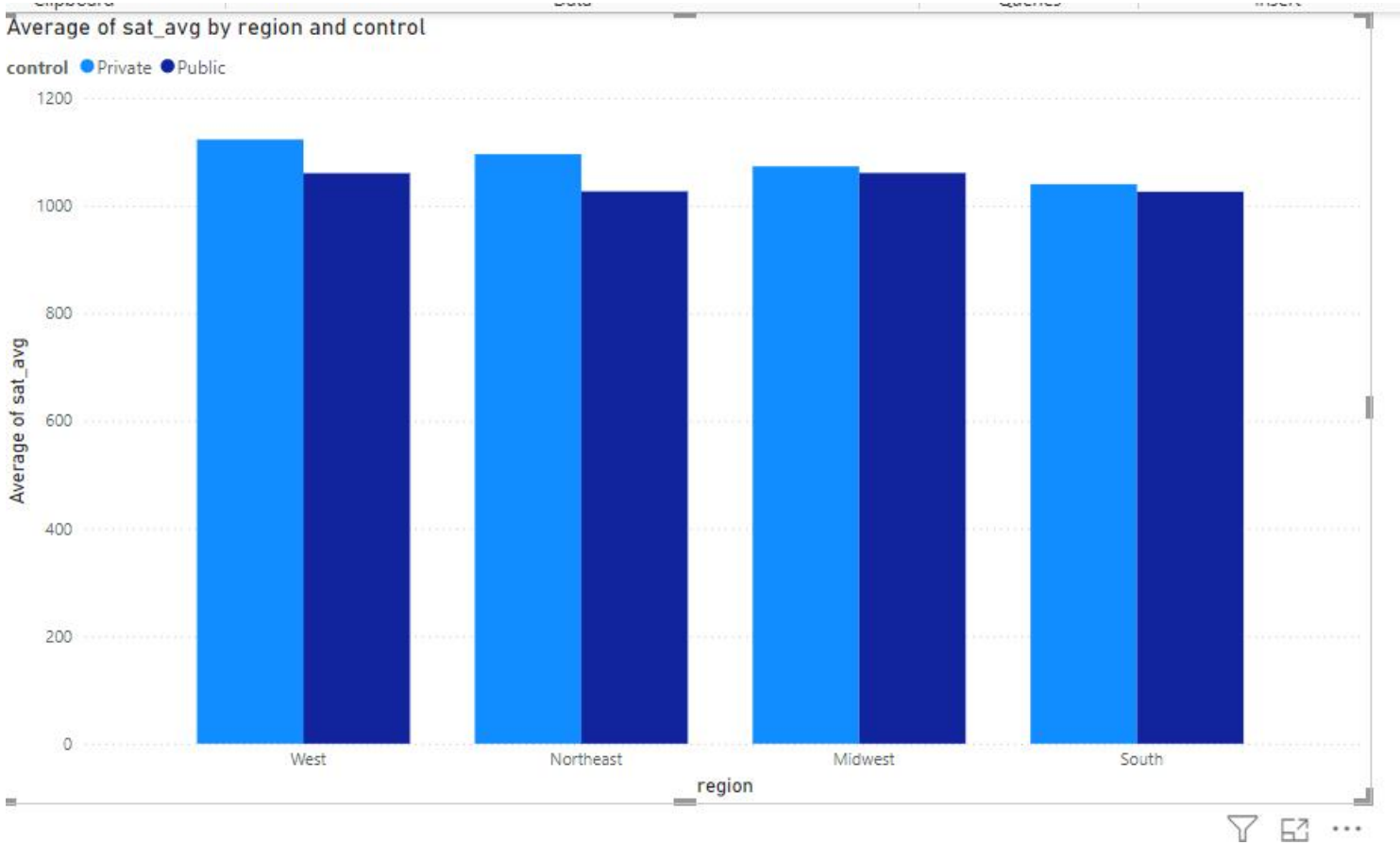
| *tuition* | |
|---|---|
| Mean | 21024.81166 |
| Standard Error | 352.1878712 |
| Median | 20000 |
| Mode | 26850 |
| Standard Deviation | 12545.99778 |
| Sample Variance | 157402060.4 |
| Kurtosis | -0.944504133 |
| Skewness | 0.452235898 |
| Range | 48276 |
| Minimum | 2732 |
| Maximum | 51008 |
| Sum | 26680486 |
| Count | 1269 |

2) Use Pivot table to create a table and pivot chart (pie chart) of average (summation) of faculty salary average, across different regions.

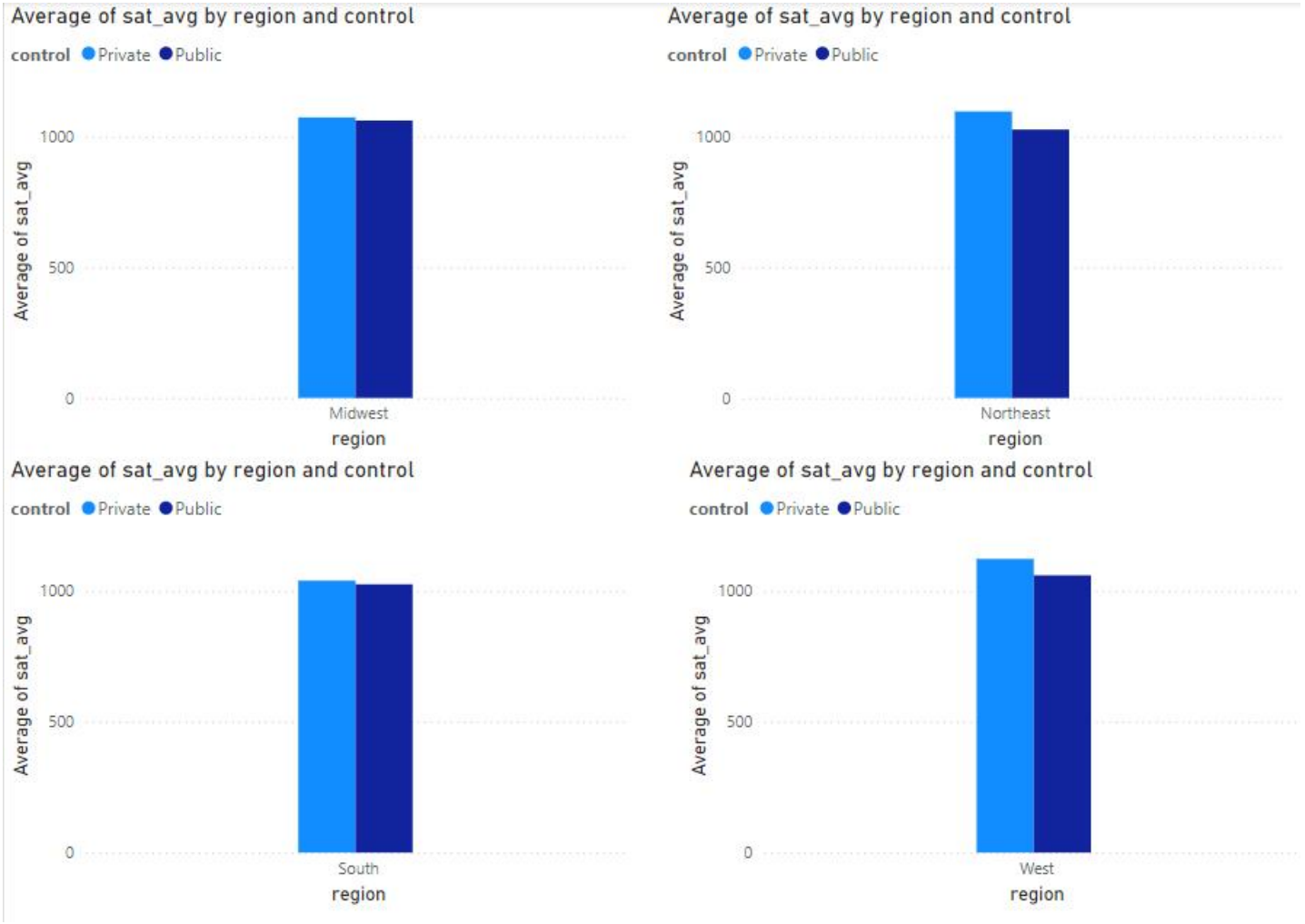
| Row Labels | Sum of faculty_salary_avg | |
|---|---|---|
| Midwest | $ | 2,491,850.00 |
| Northeast | $ | 2,617,822.00 |
| South | $ | 3,229,708.00 |
| West | $ | 1,376,123.00 |
| **Grand Total** | **$** | **9,715,503.00** |



3) Use PowerBI to visualize the average SAT score across different regions and breakdown by Control (private/public) school.



4) Use the PowerBI report for Q3. create a dashboard consisting of 4 visuals (for each visual add a filter for each region). Add cross filtering.



5) Create an SQL query to select only the columns (faculty salary, region, control), filter for only "Private" control and aggregate the average faculty salary for each region.

Query 1 | college - Schema | college - Schema

Limit to 1000 rows

```
1  select avg (faculty_salary_avg) as avg_salary, region, control
2  from college
3  where control = "Private"
4  group by region;
```

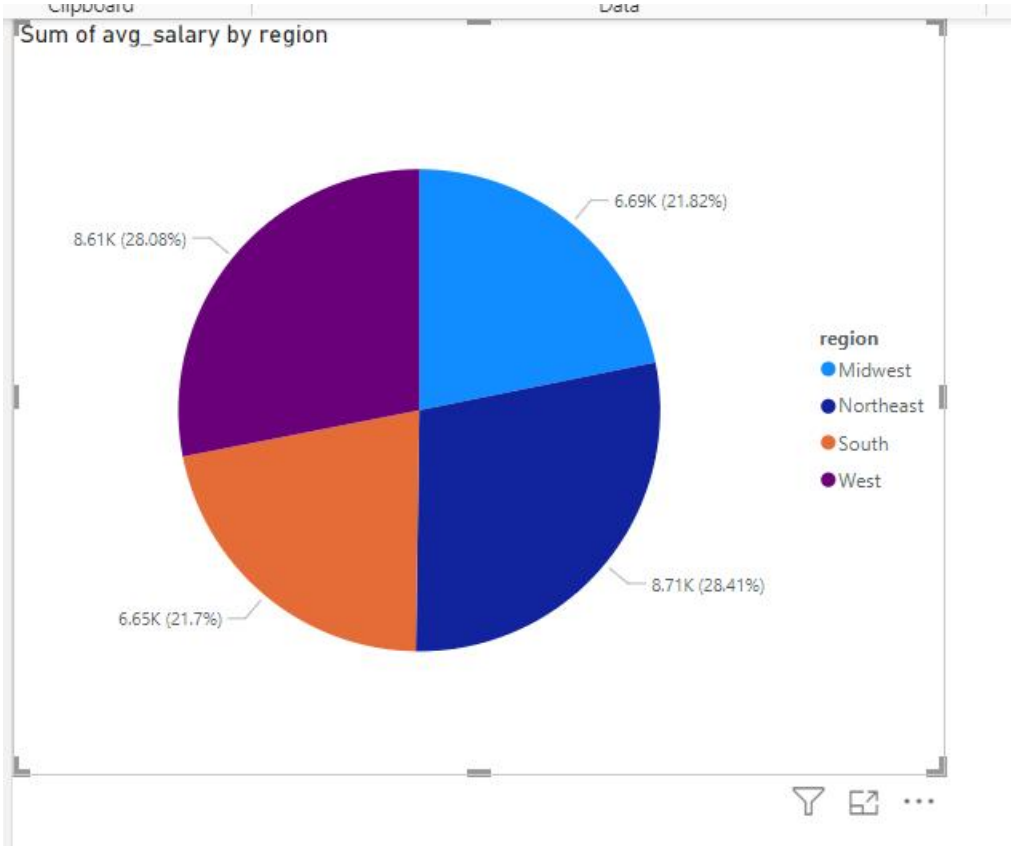
Result Grid | Filter Rows: | Export: | Wrap Cell Content:

| avg_salary | region | control |
|---|---|---|
| 8606.5000 | West | Private |
| 6651.8960 | South | Private |
| 8708.1730 | Northeast | Private |
| 6687.4268 | Midwest | Private |

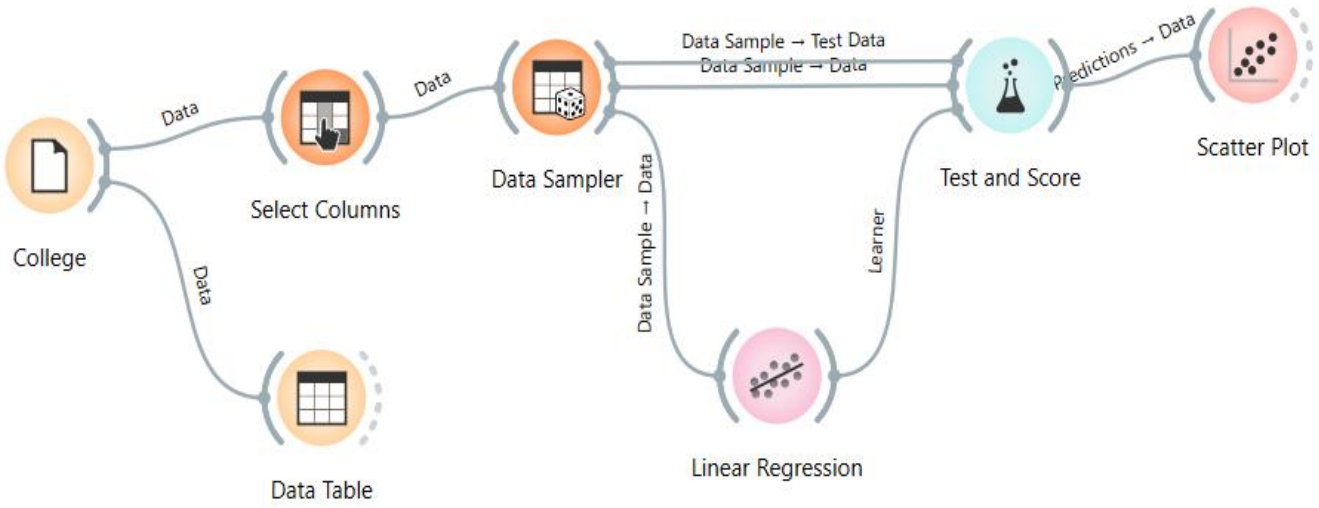
Result Grid | Form Editor | Read Only

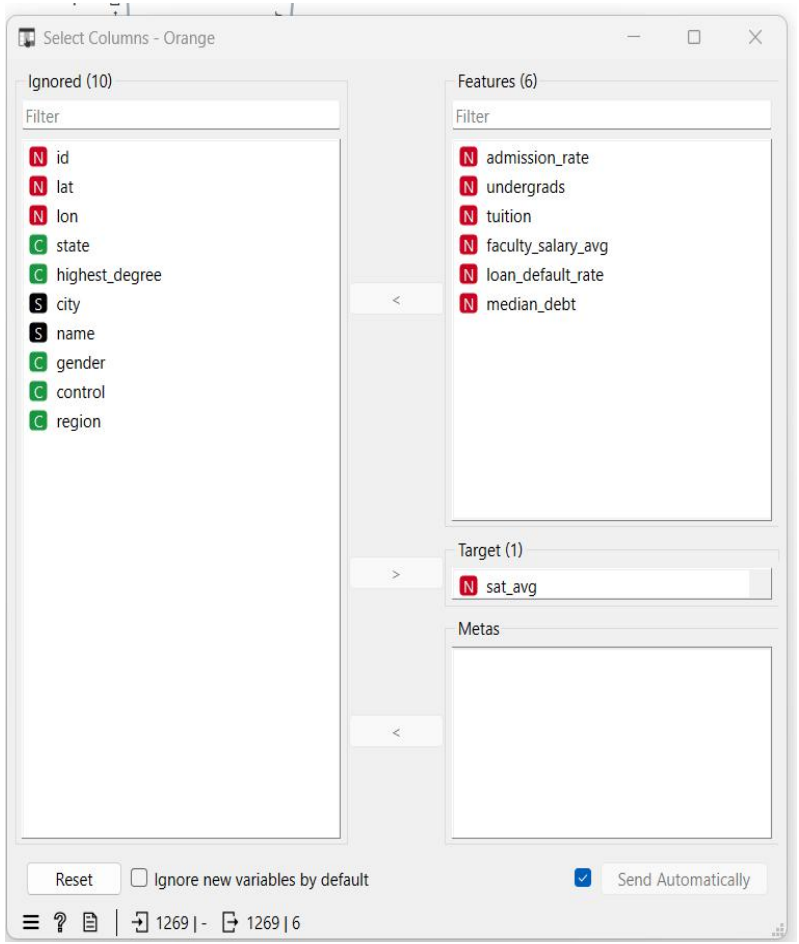
Result 2

6) Connect the MuSQL workbench usina PowerQueru to PowerBI. You will need to create a new table and export the results of Q5 into the new table. Create a pie chart using this new table.

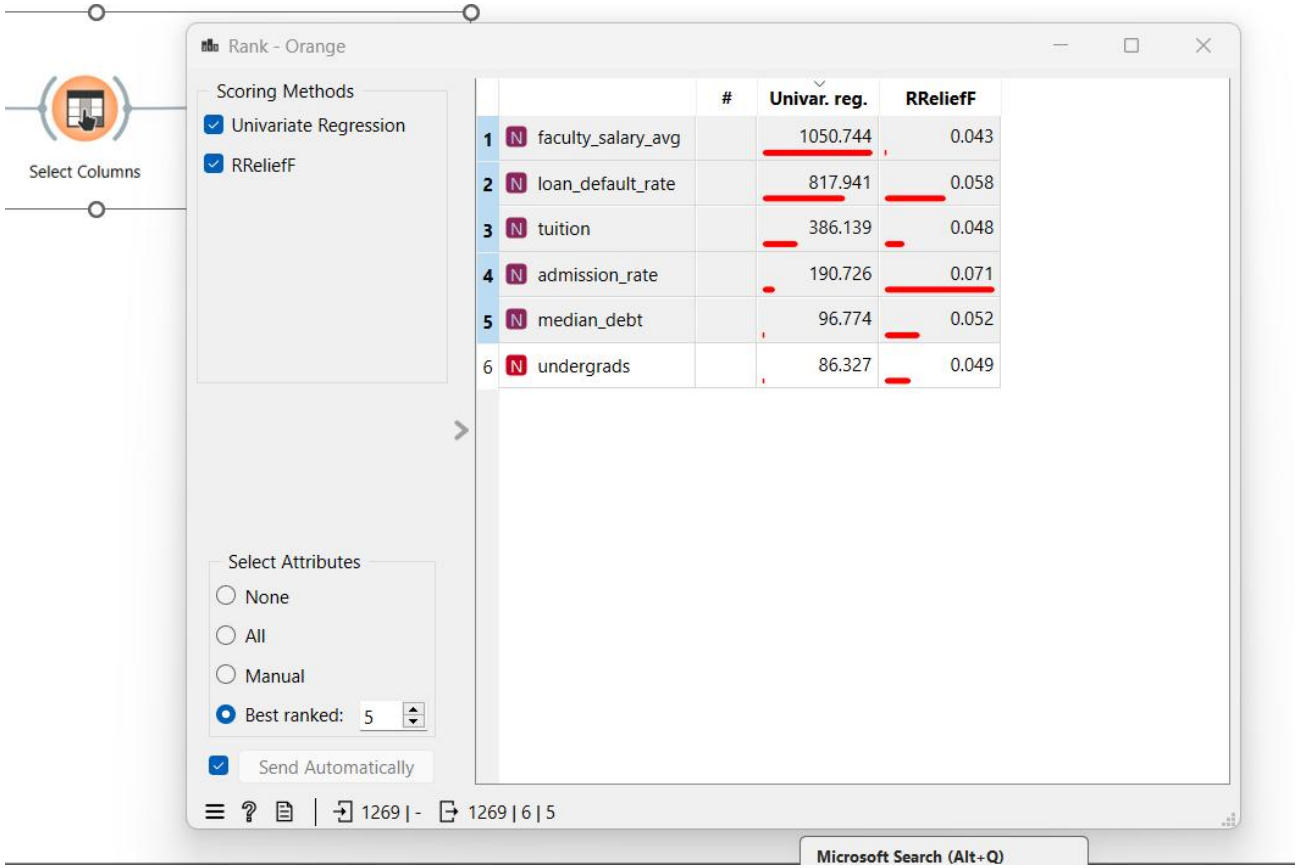


7) Connect Using Orange data mining tool, you will create a regression model to predict sat_avg using the following input columns (admissionrate, under grads, tuition, faculty_salary_avg, loan_default_rate, median_debt). Display the regression metrics Use prescriptive analytics to recommend how an institution should allocate financial aid based on factors like admission rates and loan default rates.

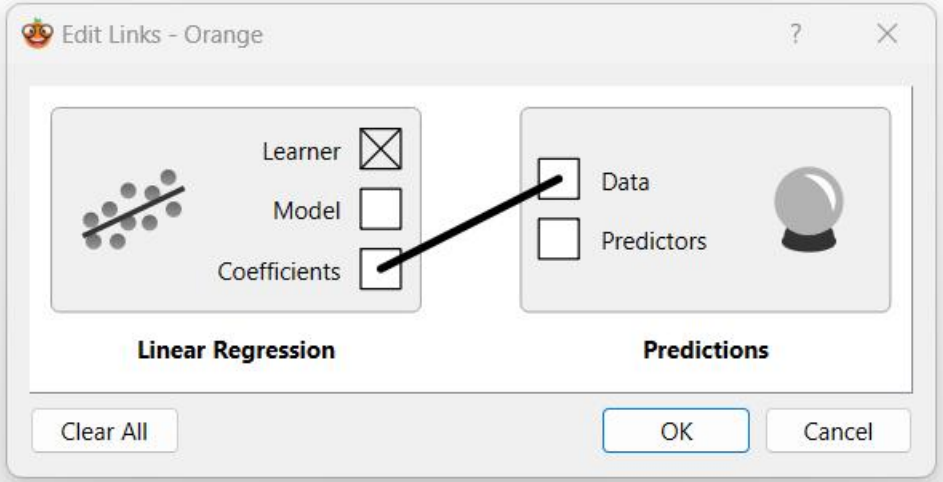


8) Based on the previous rearession model in Q7. use the Oranae Rank widget and identify the 3 most important features that can be used to predict SAT_avg.

9) Continue Based on the previous reagression model in Q7. use the Data Table widaet and view the coefficients for the reagression model. Identifu the features which must be increased and those that must be decreased in order to achieve beer sat ava scores. Based on the previous reagression model in Q7. use the Oranae Rank widget and identify the 3 most important features that can be used to predict SAT_avg.

10) Usina Oranae determine if there is a sianificant different between faculty_avg_salary for control (private/public), using hypothesis testing in the Boxplot widget.