

REPORT ON

Guided Learning Hours in Further Education and Sixth-Form Colleges

at

Faculty of Computing, Engineering and Media

**Name: Aathira Puthenpurayil**

**PNO: P2736418**

**Course: MSc Data Analytics**

|  |  |
| --- | --- |
| **Module name:** | **Analytics Programming** |
| **Module code:** | **IMAT5168** |
| **Title of the Assignment:** | **Report on Guided Learning Hours in Further Education and Sixth Form Colleges.** |

### 

Table of Contents

[1. Summary 3](#_Toc134954930)

[1.1 Overview of GLH in the UK 3](#_Toc134954931)

[2. Methods 3](#_Toc134954932)

[**2.1 Importing 3**](#_Toc134954933)

[**2.2 Identification and Handling of data anomalies 3**](#_Toc134954934)

[**2.3 Exploratory Data Analysis 3**](#_Toc134954935)

[**2.4 The choice of statistical model(s) and relevant validation 4**](#_Toc134954936)

[3. Results 5](#_Toc134954937)

[**3.1 Univariate Procedure for FE College 5**](#_Toc134954938)

[**3.2 Univariate Procedure for Sixth-Form College 6**](#_Toc134954939)

[**3.3 GLM Procedure 7**](#_Toc134954940)

[**3.3.1 Institution type 7**](#_Toc134954941)

[**3.3.2 Institution size 9**](#_Toc134954942)

[3.3.3 Region 9](#_Toc134954943)

[4. Conclusion 9](#_Toc134954944)

[5. References 10](#_Toc134954945)

[APPENDIX A: METHODS 11](#_Toc134954946)

[APPENDIX B: RESULTS 18](#_Toc134954947)

### 1. Summary

**1.1 Overview of GLH in the UK**

Guided learning hours is a term used in educational sectors of UK to find the total hours a learner spends engaged in planned learning activities. There are two dataset IMAT5168-6FORM.csv and IMAT5168-FE.csv contains data about guided learning hours for Sixth Form and Further Education colleges. In both the dataset there are columns containing Region, Institution type, Institution size, Total GLH, and learners for year 1, year 2, and year 3.

The aim of the project is to identify the effect of region, institution type, institution size, and year on GLH per learner. As we don’t have an institution size, it is calculated by taking the Total GLH for all years. For identifying various relationships average GLH and total GLH and Learners for each year are also calculated.

### 2. Methods

## **2.1 Importing**

Two datasets, IMAT5168-6FORM.csv and IMAT5168-FE.csv have been imported using INFILE, one with FE institutions and the other with sixth-form institutions. Both contain the total for individual regions for GLH and Learners which needs to be moved to a separate table as the region total. Moreover, both these institution types can be merged into a single table where all the necessary columns required for analysis need to be included such as institution size, Total GLH, Total Learners, AvgGLHPerLearners, and AvgGLHPerLearners for each year. (Refer to Appendix A for tables)

## **2.2 Identification and Handling of data anomalies**

For numeric data, there are missing values found for small institution sizes. The resulting dataset has been provided in the Appendices for reference. (Refer to Appendix A for tables)

There is one missing value for east midlands in FE College for Character data type. (Refer to Appendix A for tables)

For checking the outliers box plot has been plotted, and noticed there are a total of 7 outliers found in the plot. As there are only very few missing values, it’s not necessary to clean those, and if cleaned the institution size with small gets disappeared from the table and will not get the full context of the data. These outliers do not make much impact on the data analysis. (Refer to Appendix A for tables)

## **2.3 Exploratory Data Analysis**

Exploratory data analysis has been conducted by asking various questions to the dataset and analysing the results accordingly to get better insight Guided learning hours.

**Q1. How does the GLH (Guided Learning Hours) per learner in each year vary across different regions, institution types, and institution sizes?**

Proc means has been used to identify GLH Per Learner for three different years based on region, institution types, and institution sizes. As per the analysis of Regionthe East Midlands region has a lower average GLHPerLearner compared to the East of England, Greater London, and Northwest regions. On Institution Type Sixth Form Colleges generally have higher GLHPerLearner values than FE Colleges. On Institution size, large institutions have higher average GLHPerLearner values compared to medium and small institutions. (Refer to Appendix A for tables)

**Q2. Is there any correlation between the total GLH and the total number of learners in each institution?**

The Correlation between Total GLH and Total Learners is 0.981 which indicated it is positively correlated and P<0.0001 shows they are statistically significant. Whereas the Correlation between Total GLH and Avg GLH Per learners is -0.08330 which indicates it is negatively correlated and P>0.0001 shows they are not statistically significant. Moreover, the correlation between the Total Learners and Avg GLH Per learner is -0.2163 which indicates it is negatively correlated, and P<0.001 shows they are statistically significant. (Refer to Appendix A for tables)

**Q3. What is the overall trend in GLH per learner across all institution types?**

FE Colleges has a total of 254 records and a mean of 189.7158326 whereas the sixth Form colleges have a total of 100 records but interestingly the mean is much higher compared to FE as 559.0477275. From this, the sixth form has more GLH than FE colleges. (Refer to Appendix A for tables)

**Q4. Which region has the highest and lowest GLH per learner on average?**

From the means procedure table and provided scatter plot the Northwest region has the highest GLH per learner on average and the Southwest region has the lowest GLH per learner on average. (Refer to Appendix A for tables)

**Q5. Which institution type and size have the highest and lowest GLH per learner on average?**

Sixth Form Colleges have higher GLH per learner compared to Further education colleges. There is variation in the average GLH per learner across different institution sizes, with Large and Large-medium institutions having higher values. (Refer to Appendix A for tables)

**Q6.** **what is the total number of learners each year for all FE in each region?**

It is notable that Learner for each region follows a trend that for year 1 there is a high number of learners and gradually decreases in year 2 and 3. (Refer to Appendix A for tables)

**Q7. what is the total number of learners each year for all sixth forms in each region?**

Sixth Form colleges have the highest learners for all regions in year 1 itself but there is no gradual decrease in learners as it is seen regions including east midlands, northeast, and southwest have more in year 3 whereas remaining regions have more learners in year 2 compared to year 3. (Refer to Appendix A for tables)

**Q8. what is the total GLH each year for all FE in each region?**

Total GLH for the FE region is higher in Year 2 for all regions compared with the other two years. (Refer to Appendix A for tables)

**Q9. what is the total GLH in each year for all sixth forms in each region?**

Total GLH in year 3 is higher for regions including east midlands, east of England, northeast, northwest, southwest, west midlands, and Yorkshire. Whereas the remaining regions have a Total GLH higher in year 1. (Refer to Appendix A for tables)

## **2.4 The choice of statistical model(s) and relevant validation**

1. **PROC MEANS:** Proc means is used for GLH analysis to find the missing values of numeric variables as well as used in exploratory data analysis to find How does the GLH (Guided Learning Hours) per learner in each year vary across different regions, institution types, and institution sizes?, Compute means of Avg\_GLHPerLearner by institution type ,region and for calculating total number of learners and Total GLH in each year for all sixth form and FE in each regions. It helped in calculation of summary statistics, data validation, pre-processing, and comparison between different categories helped in understanding and exploring the dataset.
2. **PROC FREQ:** Proc freq is used in GLH to analyze missing values for character variables by Institution Type and region, calculate frequency distribution, and percentages, and help in analyzing relationships between different categories.
3. **Proc Univariate:** Proc univariate is used to analyze the distribution of average learners per year in each institution type and summary statistics. It provided a detailed analysis of distribution along with the normality assumptions as well as graphical representations.
4. **Proc sgplot:** Proc sgplot is used in GLH to visualize outliers for numeric variables and explore the dataset.
5. **Proc Scatterplot**: Proc Scatterplot is also used for visualizing the relationship between two continuous variables and identifying any patterns, trends, or outliers in the data. In GLH analysis it is used to visualize the mean GLH per learner by region.
6. **Proc GLM:** Proc glm is used in GLH to get a framework containing the relationship between GLH in each year with other variables such as institution type, institution size, and region as well as checked group differences, regression modeling, model comparison, the inclusion of covariates, and posthoc.
7. **Proc Sort:** Proc sort helped in organizing and arranging institution size, institution type, and region for the merging and analysis of GLH data.
8. **Proc corr**: Proc corr helped in understanding the correlation between Total\_GLH Total\_Learners Avg\_GLHPerLearner variables. It provides a quantitative measure which helped in getting an insight from the dataset.

### 3. Results

## **3.1 Univariate Procedure for FE College**

A screenshot of a computer screen

Description automatically generated with low confidence

A picture containing text, diagram, line, plot

Description automatically generated

The mean is 189.71 and the standard deviation is 60.098 for FE college. The goodness of fit test for normal distribution in Kolmogorov, Cramer, and Anderson test statistic (D) is 0.07069840, and the p-value is less than 0.005. This indicates that there is evidence to reject the null hypothesis of the data being normally distributed and significantly different. The graph looks bell-shaped and can be said as normally distributed. Both QQ Plot and PP Plot have almost all points lying approximately along a straight line which indicated data follows a normal distribution (Refer to Appendix B for plots) .We can also assume as sample size>=30 then it is close to normal.

## **3.2 Univariate Procedure for Sixth-Form College**

A screenshot of a computer

Description automatically generated with low confidence

A picture containing text, diagram, plot, line

Description automatically generated

The mean is 559.047 and the standard deviation is 173.87 for sixth form college. The goodness of fit test for normal distribution in Kolmogorov, Cramer, and Anderson test statistic (D) is 0.010, and the p-value is less than 0.005. This indicates that there is evidence to reject the null hypothesis of the data being normally distributed and significantly different. The graph looks bell-shaped and can be said as normally distributed. Both QQ Plot and PP Plot have almost all points lying approximately along a straight line which indicated data follows a normal distribution (Refer to Appendix B for plots).We can also assume as sample size>=30 then it is close to normal.

## **3.3 GLM Procedure**

## **3.3.1 Institution type**

A screenshot of a computer

Description automatically generated with medium confidence

A screenshot of a computer

Description automatically generated with medium confidenceA screenshot of a graph

Description automatically generated with low confidence

Checking the P value for typeIII table show p <0.05, so we reject H0, null hypothesis and combination of factors (Institution Type, Institution Size, and Region) has a significant impact on the response variable. R square with 0.075 suggests that there are 75% total variations can be explained. For institution type, p<0.0001 has a large impact on the response variable whereas P=0.0002 for institution size and p=0.0030 for region has a small impact on response variables. Tukey grouping of means also shows there is a huge difference in the mean and the model is statistically different.

## **3.3.2 Institution size**

A picture containing text, screenshot, diagram, parallel

Description automatically generated

From the box plot it is clear that large medium has highest average learners per year.

### 3.3.3 Region

A screenshot of a graph

Description automatically generated with medium confidence

Northwest regions have higher Avg GLH per learners and southwest has the least number of average GLH per learners.

### 4. Conclusion

GLH Analysis has been conducted with different procedures to analyse the effect of region, institution size, institution type and GLH per learner. As per the analysis, the sixth-form college has more GLH compared to further education even though the size of further education is more. Moreover, large-medium institutions have higher Avg GLH learners. Analysing the region, it is observed that North-West has a higher GLH per hour and the least for the southwest region. One interesting factor that has been noticed is that the FE colleges for all regions have more GLH learners in year 3 and gradually decrease in year 2 and year 1.

On further analysis, it is found that P<0.005, so we can reject H0, and the model is statistically different. The plotted points for QQ and PP Plot do appear to fall reasonably well on the straight line, so normality looks okay. We can also assume as the sample size>=30 for both FE and sixth-form institution types then it is close to normal.

Overall, the data has been analyzed using various statistical models and drawn conclusions from the analysis based on the institution type, size, region, and GLH per learner.

### 5. References

**1.** [PROC IMPORT: PROC IMPORT Statement (sas.com)](https://support.sas.com/documentation/cdl/en/acpcref/63184/HTML/default/viewer.htm#a003102096.htm)

**2**. Smithers, A., & Robinson, P. (2006). General learning hours and their role in determining funding for further education colleges. Journal of Further and Higher Education, 30(2), 113-126.

**3.** Textbook-Step-By-Step Programming with Base SAS 9.4, 2nd edn.

### APPENDIX A: METHODS

**Import Data set**

A screenshot of a computer

Description automatically generated with medium confidence

A screenshot of a computer

Description automatically generated with medium confidence

A screenshot of a computer

Description automatically generated with medium confidence

A screenshot of a computer

Description automatically generated with medium confidence

A screenshot of a data

Description automatically generated with medium confidence

**Identification and Handling of data anomalies**

|  |  |  |  |
| --- | --- | --- | --- |
| Region | Institution Type | Institution Size | Missing Numbers |
| East Midlands | FE | Small | 2 |
| East of England | FE | Small | 1 |
| Northwest | FE | Small | 5 |
| Northwest | Sixth Form | Small | 1 |
| Southeast | FE | Small | 2 |
| Southeast | Sixth Form | Small | 2 |
| West Midlands | Sixth Form | Small | 1 |
| Yorkshire and the Humber | FE | Small | 3 |

A screenshot of a computer

Description automatically generated with medium confidence

A picture containing text, screenshot, diagram, plot

Description automatically generated **Exploratory Data Analysis**

**Q1. How does the GLH (Guided Learning Hours) per learner in each year vary across different regions, institution types, and institution sizes?**

A picture containing text, screenshot, number, font

Description automatically generated

A screenshot of a computer screen

Description automatically generated with low confidence

**Q2. Is there any correlation between the total GLH and the total number of learners in each institution?**

A screenshot of a computer

Description automatically generated with low confidence

**Q3. What is the overall trend in GLH per learner across all institution types?**

A screenshot of a table

Description automatically generated with low confidence

**Q4. Which region has the highest and lowest GLH per learner on average?**

A picture containing text, screenshot, number, font

Description automatically generated

A screen shot of a graph

Description automatically generated with low confidence

**Q5. Which institution type and size have the highest and lowest GLH per learner on average?**

A picture containing text, screenshot, number, font

Description automatically generated

A picture containing text, screenshot, number, font

Description automatically generated

**Q6.** **what is the total number of learners each year for all FE in each region?**

A picture containing text, screenshot, number, font

Description automatically generated

**Q7. what is the total number of learners each year for all sixth forms in each region?**

A screenshot of a table

Description automatically generated with low confidence

**Q8. what is the total GLH each year for all FE in each region?**

A picture containing text, screenshot, number, font

Description automatically generated

**Q9. what is the total GLH in each year for all sixth forms in each region?**

A screenshot of a table

Description automatically generated with low confidence

### APPENDIX B: RESULTS

**Plot of Avg\_GLHPerLearner with InstitutionType: FE College**

A picture containing text, screenshot, line, diagram

Description automatically generated

A picture containing text, screenshot, line, plot

Description automatically generated

**Plot of Avg\_GLHPerLearner with InstitutionType: Sixth Form College**

A picture containing text, screenshot, line, diagram

Description automatically generated

A picture containing text, screenshot, line, plot

Description automatically generated