BIG DATA AND BUSINESS INTELLIGENCE

MODULE : CIS4008-N

UNIVERSITY STUDENTS DATA ANALYSIS

SECTION 1: BUSINESS INTELLIGENCE DESIGN

NAME: YUVRAJ CHANDRAKANT HARYAN

STUDENT ID: D3581054

SUBMISSION DATE : 09/01/2024

TABLE OF CONTENTS

1. Data Source Description and BI Requirements

1.1 Introduction

Using the University Students Dataset as a starting point for a Business Intelligence (BI) analysis is a calculated step into the world of academic data. The aim of this analysis is to gather insightful data that will help decision-makers, optimize educational strategies and improve the educational experience for all students. This analysis focuses on extracting actionable knowledge from the dataset by closely examining assessments, courses, student data, and virtual learning activities. This will help make well-informed decisions in the ever-changing world of higher education.

1.2 Data Set Source and Description

1.2.1 Source

The Anonymised University Students Dataset downloaded from the analyse.kmi.open.ac.uk : https://analyse.kmi.open.ac.uk/open_dataset#:~:text=This%20page%20introduces%20the%20anonymised,selected%20courses%20(called%20modules).

The dataset includes the "studentAssessment", "studentInfo", "studentVle", "vle", "assessments" and "courses" tables. It is a comprehensive collection of data pertaining to students in a university setting. The dataset is structured to provide a holistic view of student academic journeys, enabling thorough analysis for insights into factors influencing student success, demographic patterns, and the impact of virtual learning activities.

1.2.2 Description

Table 1 : assessments

Index	Column Name	Description
1	code_module	Name of the Course
2	code_presentation	Sub-category of the course
3	id_assessment	Unique ID for assessments
4	assessment_type	Type of the assessment conducted
5	weight	Weight of the assessment conducted

Table 2 : courses

Index	Column Name	Description
1	code_module	Name of the Course
2	code_presentation	Sub-category of the course
3	module_presentation_length	How long the module will be taught for

Table 3 : studentAssessment

Index	Column Name	Description
1	id_assessment	Unique ID for assessments
2	id_student	Unique ID for students
3	score	Score of the student in that assessment

Table 4 : studentInfo

Index	Column Name	Description
1	code_module	Name of the Course
2	code_presentation	Sub-category of the course
3	id_student	Unique ID for students
4	gender	Gender of the student
5	region	Home region of the student
6	highest_education	Highest level of education that the student has received
7	imd_band	Index of multiple deprivation band (this is a relative measure of deprivation i.e. it can tell you if one area is more deprived than another but not by how much)
8	age_band	The age group in which the student falls
9	num_of_prev_attempts	The number of attempts he required to clear his highest level of qualification
10	studied_credits	Student's total studied credits till now

11	disability	A flag to specify if the student has a disability
12	final_result	Result of the student.
13	attendance	Attendance of the student

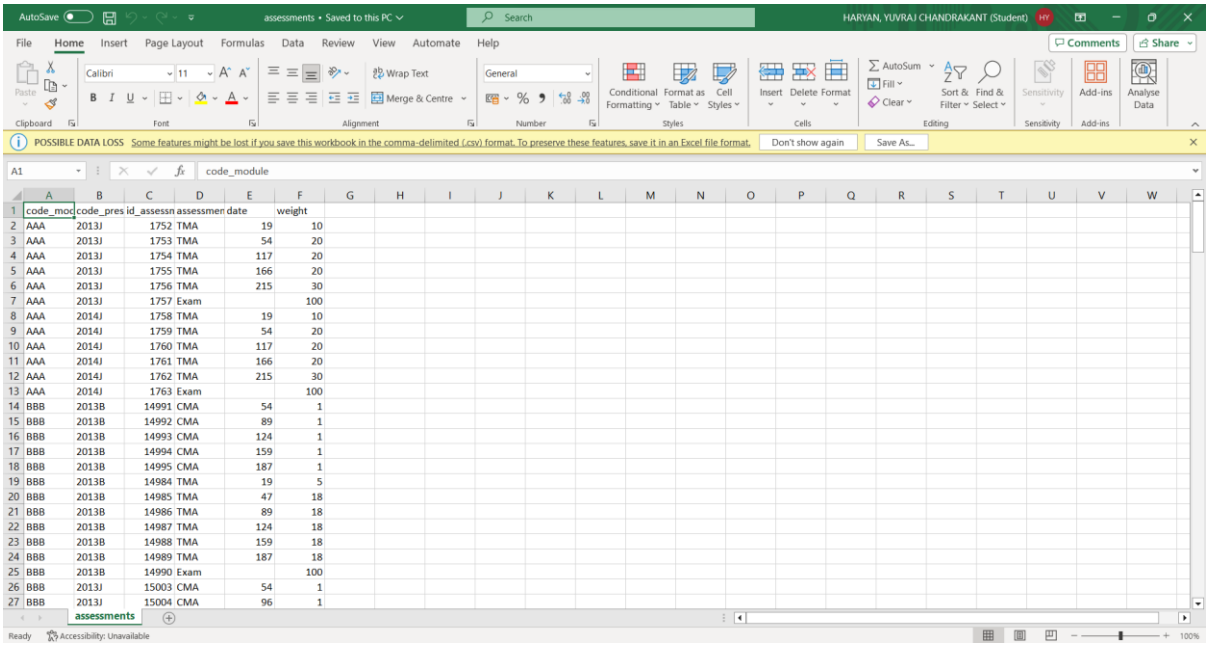
Table 5 : studentVle

Index	**Column Name**	**Description**
1	code_module	Name of the Course
2	code_presentation	Sub-category of the course
3	id_student	Unique ID for students
4	id_site	Unique ID for the location of the Virtual learning activity
5	sum_click	Total number of clicks on that Virtual learning activity by students

Table 6 : vle

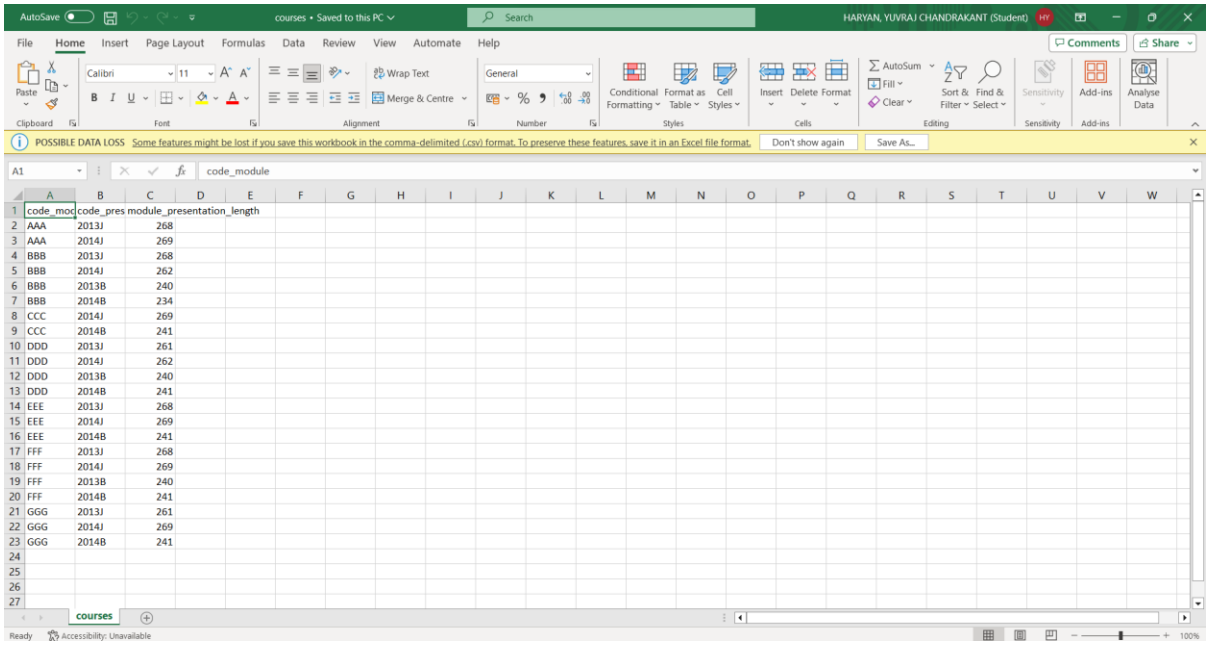
Index	**Column Name**	**Description**
1	id_site	Unique ID for the location of the Virtual learning activity
2	code_module	Name of the Course
3	code_presentation	Sub-category of the course
4	activity_type	Type of activity used for Virtual learning

Snapshots of the dataset are presented below :



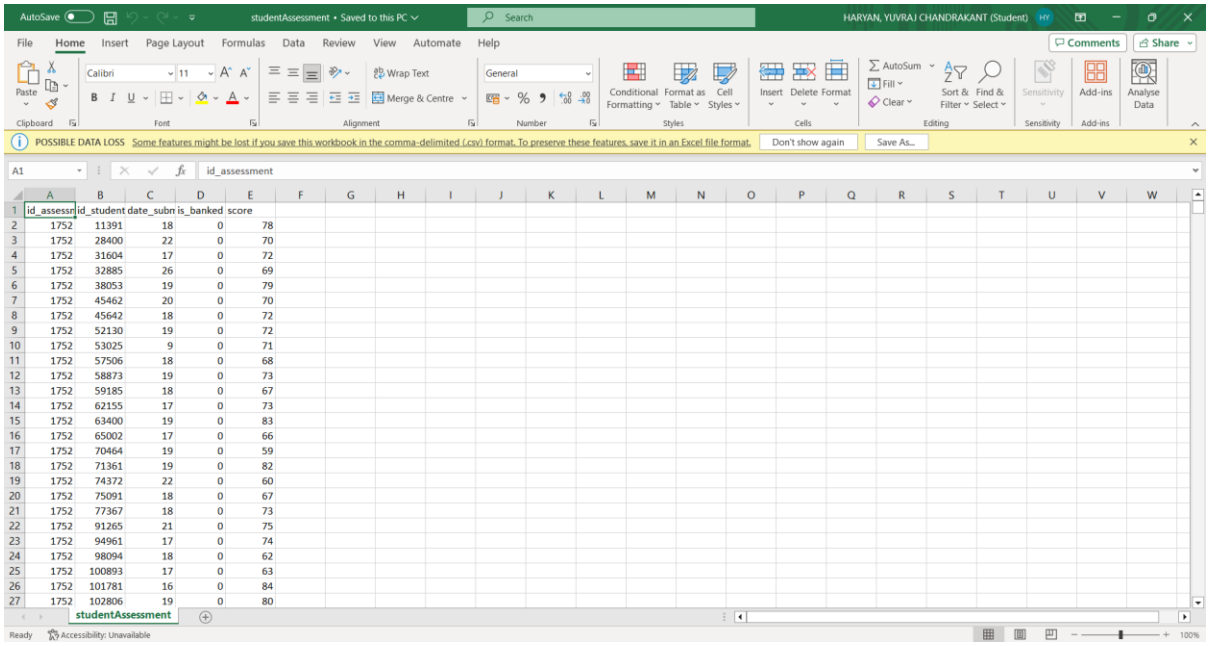
code_mod	code_pres	id_assessn	assessmen	date	weight
AAA	2013J	1752	TMA	19	10
AAA	2013J	1753	TMA	54	20
AAA	2013J	1754	TMA	117	20
AAA	2013J	1755	TMA	166	20
AAA	2013J	1756	TMA	215	30
AAA	2013J	1757	Exam		100
AAA	2014J	1758	TMA	19	10
AAA	2014J	1759	TMA	54	20
AAA	2014J	1760	TMA	117	20
AAA	2014J	1761	TMA	166	20
AAA	2014J	1762	TMA	215	30
AAA	2014J	1763	Exam		100
BBB	2013B	14991	CMA	54	1
BBB	2013B	14992	CMA	89	1
BBB	2013B	14993	CMA	124	1
BBB	2013B	14994	CMA	159	1
BBB	2013B	14995	CMA	187	1
BBB	2013B	14984	TMA	19	5
BBB	2013B	14985	TMA	47	18
BBB	2013B	14986	TMA	89	18
BBB	2013B	14987	TMA	124	18
BBB	2013B	14988	TMA	159	18
BBB	2013B	14989	TMA	187	18
BBB	2013B	14990	Exam		100
BBB	2013J	15003	CMA	54	1
BBB	2013J	15004	CMA	96	1

Fig 1 : assessments table data



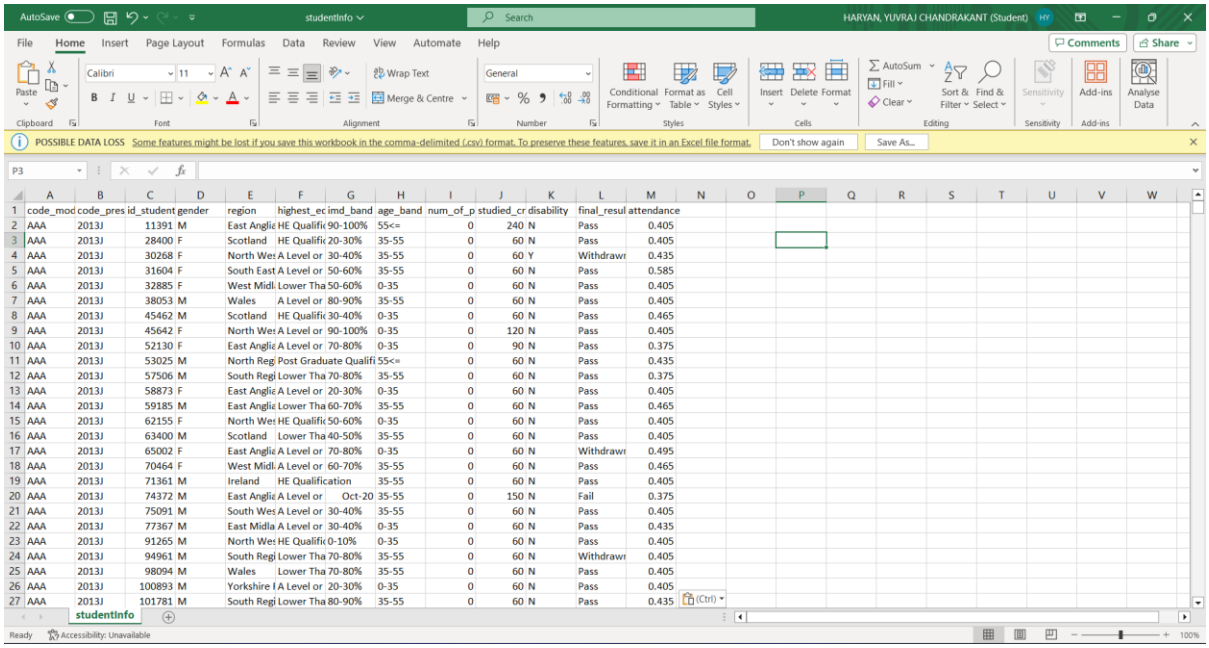
code_mod	code_pres	module_presentation_length
AAA	2013J	268
AAA	2014J	269
BBB	2013J	268
BBB	2014J	262
BBB	2013B	240
BBB	2014B	234
CCC	2014J	269
CCC	2014B	241
DDD	2013J	261
DDD	2014J	262
DDD	2013B	240
DDD	2014B	241
EEE	2013J	268
EEE	2014J	269
EEE	2014B	241
FFF	2013J	268
FFF	2014J	269
FFF	2013B	240
FFF	2014B	241
GGG	2013J	261
GGG	2014J	269
GGG	2014B	241

Fig 2 : courses table data



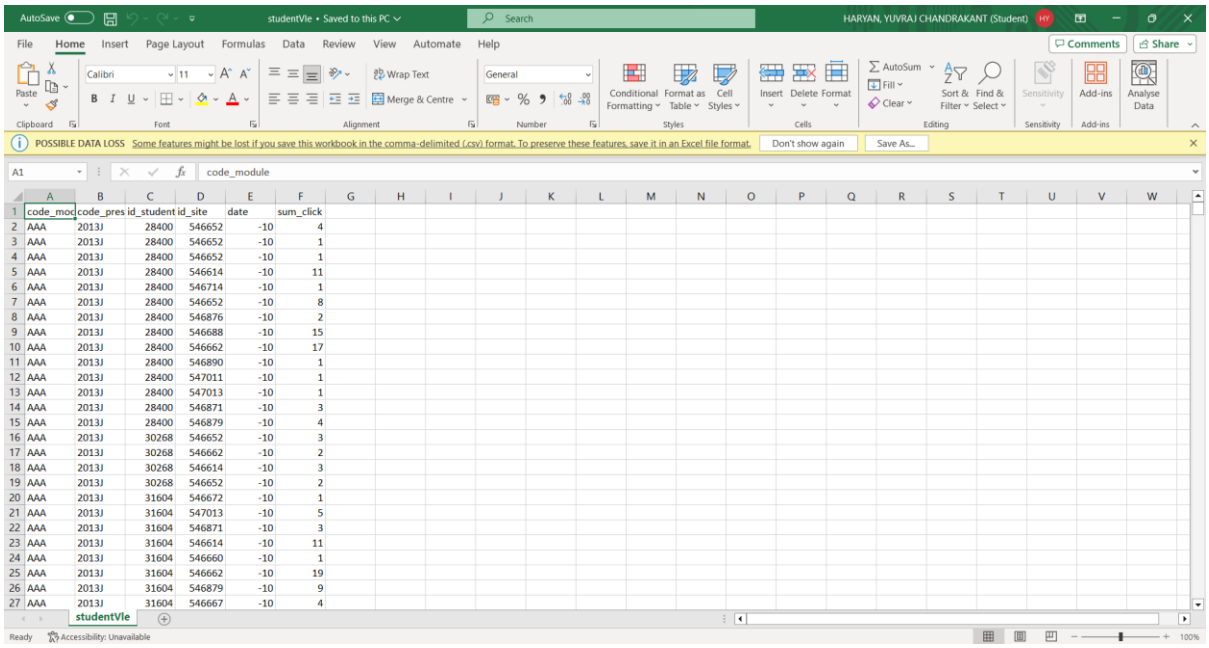
id_assessment	id_student	date_submitted	is_banked	score
1752	11391	18	0	78
1752	28400	22	0	70
1752	31604	17	0	72
1752	32885	26	0	69
1752	38053	19	0	79
1752	45462	20	0	70
1752	45642	18	0	72
1752	52130	19	0	72
1752	53025	9	0	71
1752	57506	18	0	68
1752	58873	19	0	73
1752	59185	18	0	67
1752	62155	17	0	73
1752	63400	19	0	83
1752	65002	17	0	66
1752	70464	19	0	59
1752	71361	19	0	82
1752	74372	22	0	60
1752	75091	18	0	67
1752	77367	18	0	73
1752	91265	21	0	75
1752	94961	17	0	74
1752	98094	18	0	62
1752	100893	17	0	63
1752	101781	16	0	84
1752	102806	19	0	80

Fig 3 : studentAssessment table data



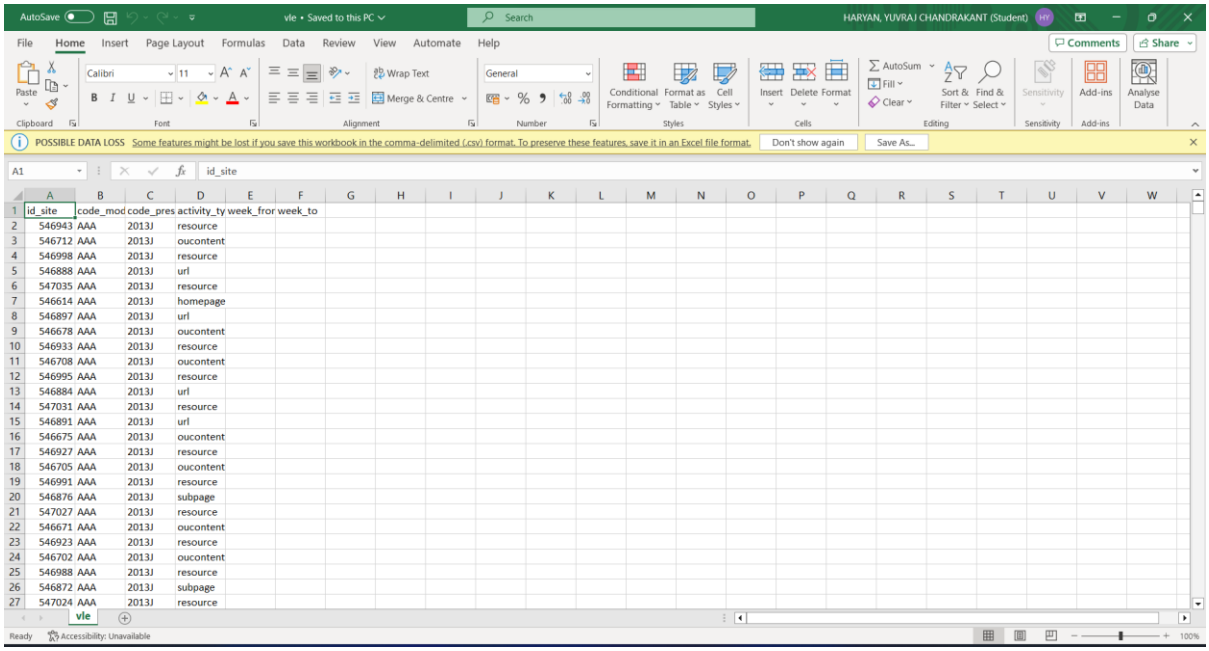
code_module	code_presentation	id_student	gender	region	highest_education	imd_band	age_band	num_of_prev_attempts	studied_credits	disability	final_result	attendance
AAA	2013J	11391	M	East Anglia	HE Qualific	90-100%	55<=	0	240	N	Pass	0.405
AAA	2013J	28400	F	Scotland	HE Qualific	20-30%	35-55	0	60	N	Pass	0.405
AAA	2013J	30268	F	North Wes	A Level or	30-40%	35-55	0	60	Y	Withdrawn	0.435
AAA	2013J	31604	F	South East	A Level or	50-60%	35-55	0	60	N	Pass	0.585
AAA	2013J	32885	F	West Midl	Lower Tha	50-60%	0-35	0	60	N	Pass	0.405
AAA	2013J	38053	M	Wales	A Level or	80-90%	35-55	0	60	N	Pass	0.405
AAA	2013J	45462	M	Scotland	HE Qualific	30-40%	0-35	0	60	N	Pass	0.465
AAA	2013J	45642	F	North Wes	A Level or	90-100%	0-35	0	120	N	Pass	0.405
AAA	2013J	52130	F	East Anglia	A Level or	70-80%	0-35	0	90	N	Pass	0.375
AAA	2013J	53025	M	North Regi	Post Graduate Qualifi	55<=		0	60	N	Pass	0.435
AAA	2013J	57506	M	South Regi	Lower Tha	70-80%	35-55	0	60	N	Pass	0.375
AAA	2013J	58873	F	East Anglia	A Level or	20-30%	0-35	0	60	N	Pass	0.405
AAA	2013J	59185	M	East Anglia	Lower Tha	60-70%	35-55	0	60	N	Pass	0.465
AAA	2013J	62155	F	North Wes	HE Qualific	50-60%	0-35	0	60	N	Pass	0.405
AAA	2013J	63400	M	Scotland	Lower Tha	40-50%	35-55	0	60	N	Pass	0.405
AAA	2013J	65002	F	East Anglia	A Level or	70-80%	0-35	0	60	N	Withdrawn	0.495
AAA	2013J	70464	F	West Midl	A Level or	60-70%	35-55	0	60	N	Pass	0.465
AAA	2013J	71361	M	Ireland	HE Qualification		35-55	0	60	N	Pass	0.405
AAA	2013J	74372	M	East Anglia	A Level or	Oct-20	35-55	0	150	N	Fail	0.375
AAA	2013J	75091	M	South Wes	A Level or	30-40%	35-55	0	60	N	Pass	0.405
AAA	2013J	77367	M	East Midla	A Level or	30-40%	0-35	0	60	N	Pass	0.435
AAA	2013J	91265	M	North Wes	HE Qualific	0-10%	0-35	0	60	N	Pass	0.405
AAA	2013J	94961	M	South Regi	Lower Tha	70-80%	35-55	0	60	N	Withdrawn	0.405
AAA	2013J	98094	M	Wales	Lower Tha	70-80%	35-55	0	60	N	Pass	0.405
AAA	2013J	100893	M	Yorkshire I	A Level or	20-30%	0-35	0	60	N	Pass	0.405
AAA	2013J	101781	M	South Regi	Lower Tha	80-90%	35-55	0	60	N	Pass	0.435

Fig 4 : studentInfo table data



code_mod	code_pres	id_student	id_site	date	sum_click
AAA	2013J	28400	546652	-10	4
AAA	2013J	28400	546652	-10	1
AAA	2013J	28400	546652	-10	1
AAA	2013J	28400	546614	-10	11
AAA	2013J	28400	546714	-10	1
AAA	2013J	28400	546652	-10	8
AAA	2013J	28400	546876	-10	2
AAA	2013J	28400	546688	-10	15
AAA	2013J	28400	546662	-10	17
AAA	2013J	28400	546890	-10	1
AAA	2013J	28400	547011	-10	1
AAA	2013J	28400	547013	-10	1
AAA	2013J	28400	546871	-10	3
AAA	2013J	28400	546879	-10	4
AAA	2013J	30268	546652	-10	3
AAA	2013J	30268	546662	-10	2
AAA	2013J	30268	546614	-10	3
AAA	2013J	30268	546652	-10	2
AAA	2013J	31604	546672	-10	1
AAA	2013J	31604	547013	-10	5
AAA	2013J	31604	546871	-10	3
AAA	2013J	31604	546614	-10	11
AAA	2013J	31604	546660	-10	1
AAA	2013J	31604	546662	-10	19
AAA	2013J	31604	546879	-10	9
AAA	2013J	31604	546667	-10	4

Fig 5 : studentVle table data



id_site	code_mod	code_pres	activity_ty	week_fror	week_to
546943	AAA	2013J	resource		
546712	AAA	2013J	oucontent		
546998	AAA	2013J	resource		
546888	AAA	2013J	url		
547035	AAA	2013J	resource		
546614	AAA	2013J	homepage		
546897	AAA	2013J	url		
546678	AAA	2013J	oucontent		
546933	AAA	2013J	resource		
546708	AAA	2013J	oucontent		
546995	AAA	2013J	resource		
546884	AAA	2013J	url		
547031	AAA	2013J	resource		
546891	AAA	2013J	url		
546675	AAA	2013J	oucontent		
546927	AAA	2013J	resource		
546705	AAA	2013J	oucontent		
546991	AAA	2013J	resource		
546876	AAA	2013J	subpage		
547027	AAA	2013J	resource		
546671	AAA	2013J	oucontent		
546923	AAA	2013J	resource		
546702	AAA	2013J	oucontent		
546988	AAA	2013J	resource		
546872	AAA	2013J	subpage		
547024	AAA	2013J	resource		

Fig 6 : vle table data

1.3 Rationale for Choosing Dataset

- With information on tests, courses, student assessments, student information, virtual learning activities, and learning environment features, the dataset provides an all-encompassing perspective of university student data. Due to the abundance of data, a comprehensive study and a thorough grasp of the many facets pertaining to student performance and engagement are made possible.
- The dataset is organised into several tables, each offering a unique viewpoint. The tables labelled "assessments" and "courses" provide details about the types, weights of assessments and duration of modules respectively. The "studentAssessment" table displays student assessment scores; the "studentInfo" table provides demographic and academic background information; and the "studentVle" and "vle" tables offer information about virtual learning activities. This diversity allows for a more thorough investigation.
- Given that it includes data on student performance, engagement, and demographics, the dataset is very pertinent for research in education. Examining assessment trends, comprehending what influences student achievement, and examining engagement in virtual learning environments can all provide insightful information about how to improve teaching methods.
- Due to the complexity of the information, there are possibilities to learn and practise a variety of business skills, such as relationship building, data modelling, data cleaning and preprocessing, and using different data manipulation languages. Professionals in data-driven fields need these abilities, which may be developed via real-world use on a dataset this size and complexity.
- The dataset has big data-related issues, including managing multiple data sources, handling different kinds of information, and combining data from multiple tables. This dataset is an invaluable tool for tackling big data issues and improving problem-solving abilities since it presents challenges that are consistent with real-world situations involving vast and diverse information.

1.4 Objectives

Gaining knowledge and actionable insights that can guide strategic decision-making and enhance educational outcomes is the main objective of this data analysis on the University Students Dataset. This will provide educators, administrators, and policymakers with useful information that will improve the overall educational experience and results for students.

1.5 Scope of Work and BI Questions

The BI analysis seeks to explain the following questions by carefully examining the various dimensions present in the dataset, such as assessments, courses, student assessments, student information, and virtual learning activities:

1. How does the student's previous education level impact their performance? (KPI: Average score based on highest education level)

2. How well do students are making use of the Virtual Learning Environment (VLe) ?

3. Are there any trends in the way students use the VLe and are there any resource types that are popular among students?

4. Is there any trend in the way of students choosing the courses?

5. From what region are most of the students at the University?

6. Can a portal-like page where all the student data is viewed at a glance can be made using power BI?

2. Data Pre-Processing and Cleaning

2.1 Data Loading

Loading the dataset is the first step in this analysis. The 'Get Data' button on the home tab is used to accomplish this. As seen in Fig 7, this button displays a drop-down menu with the various methods for loading data into Power BI.

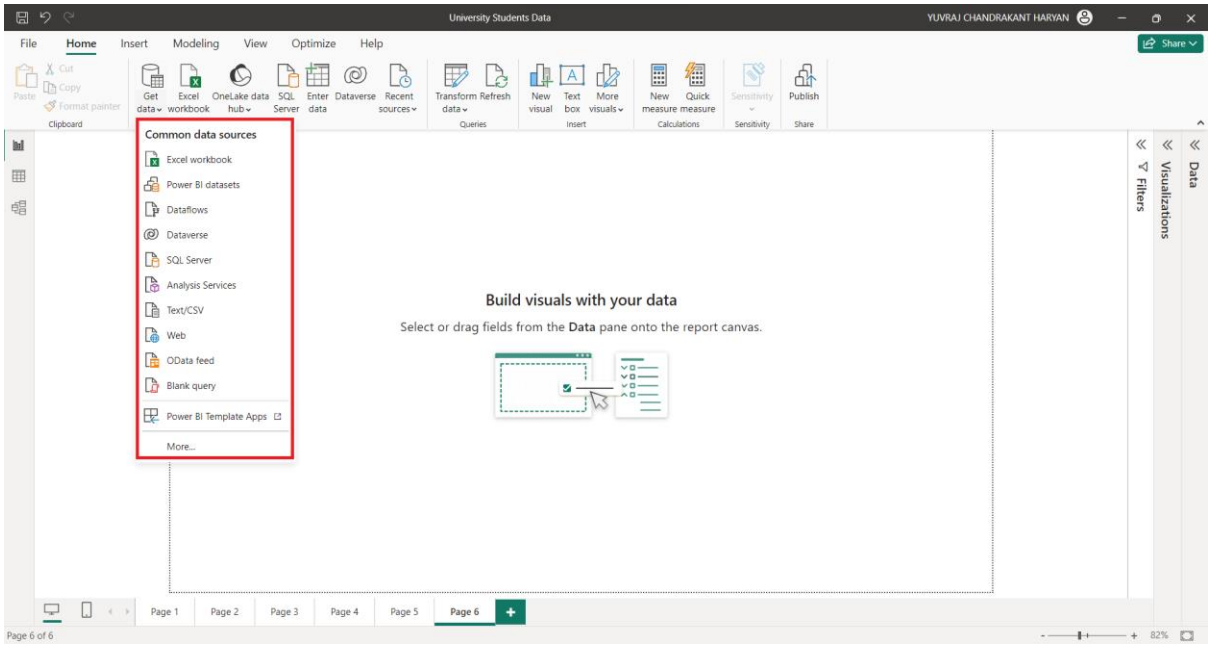


Fig 7 : Options for loading data

The Text/CSV option was chosen because the project's data is in CSV format. The dataset for assessments table was selected. This resulted in the display of a dialogue box with the option to load or modify data, as indicated in Figure 8

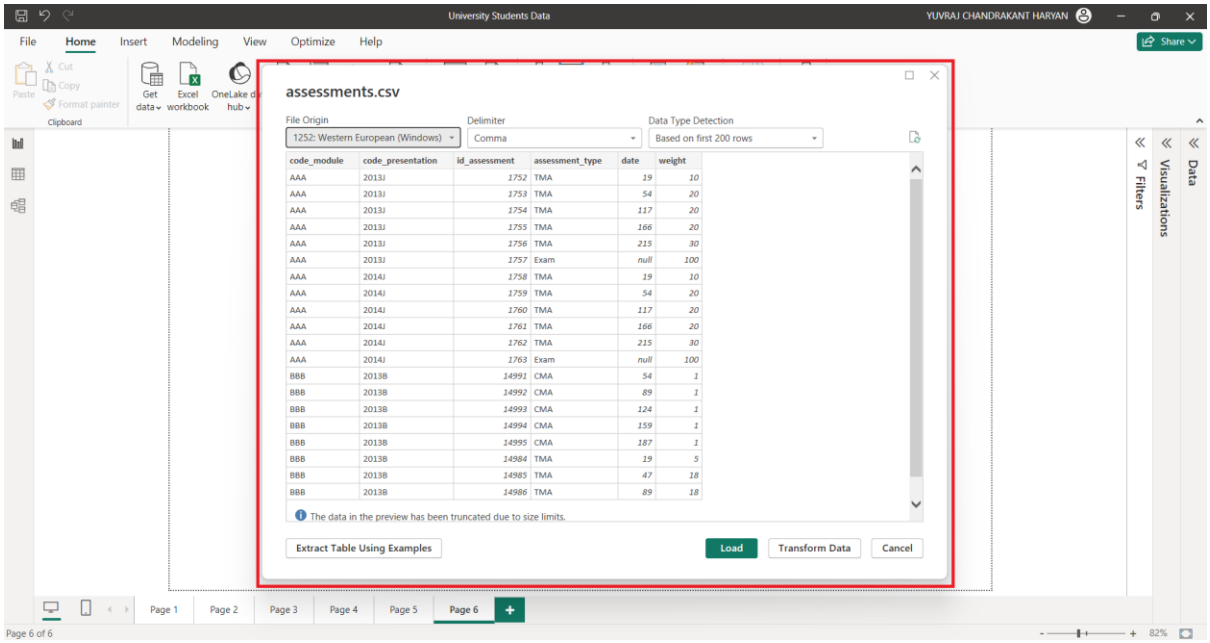


Fig 8 : Dialog box showcasing the data and options to edit the data before loading

After selecting the load option, the assessments data table was loaded successfully. The other tables were loaded using the same procedure. The outcome of loading all the tables into Power BI is displayed in Fig 9. The drop-down headers are the table names while the expanded list comprises of its column names.

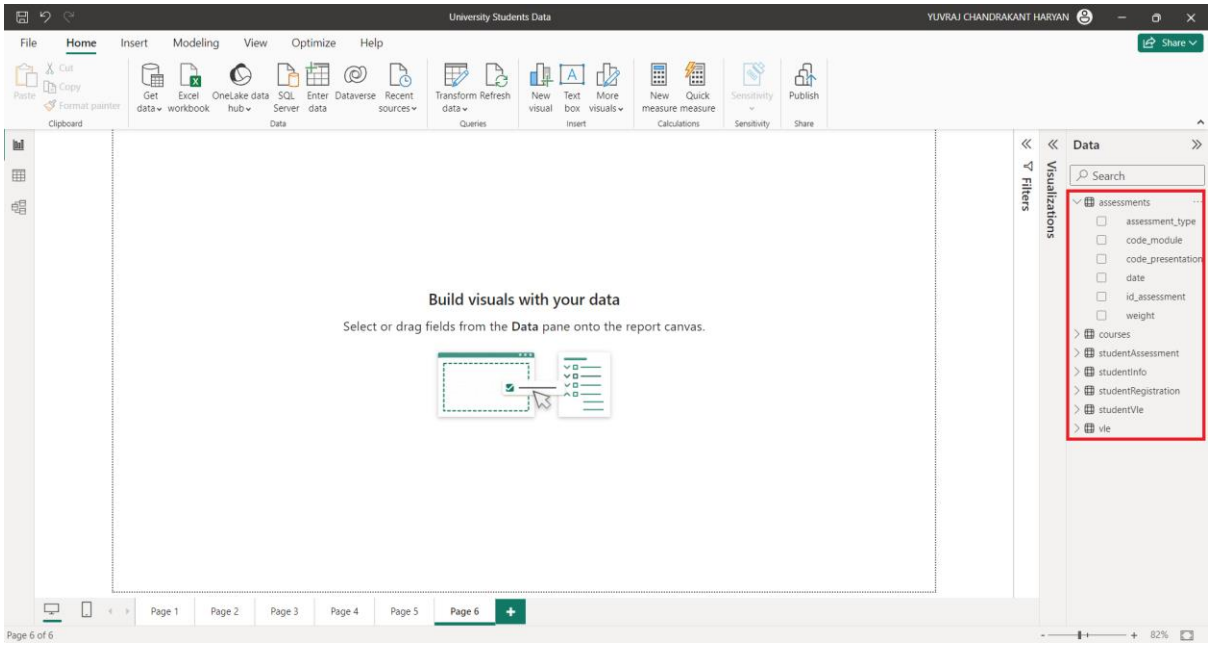


Fig 9 : Successfully loaded tables

2.2 Data Cleaning

The pre-processing and data cleansing will be completed in Power Query as the following stage. As seen in Fig 10, to accomplish this, the Transform data drop down on the home tab was selected. When Transform data was selected under this, the Power Query Editor appeared, as seen in Fig 11.

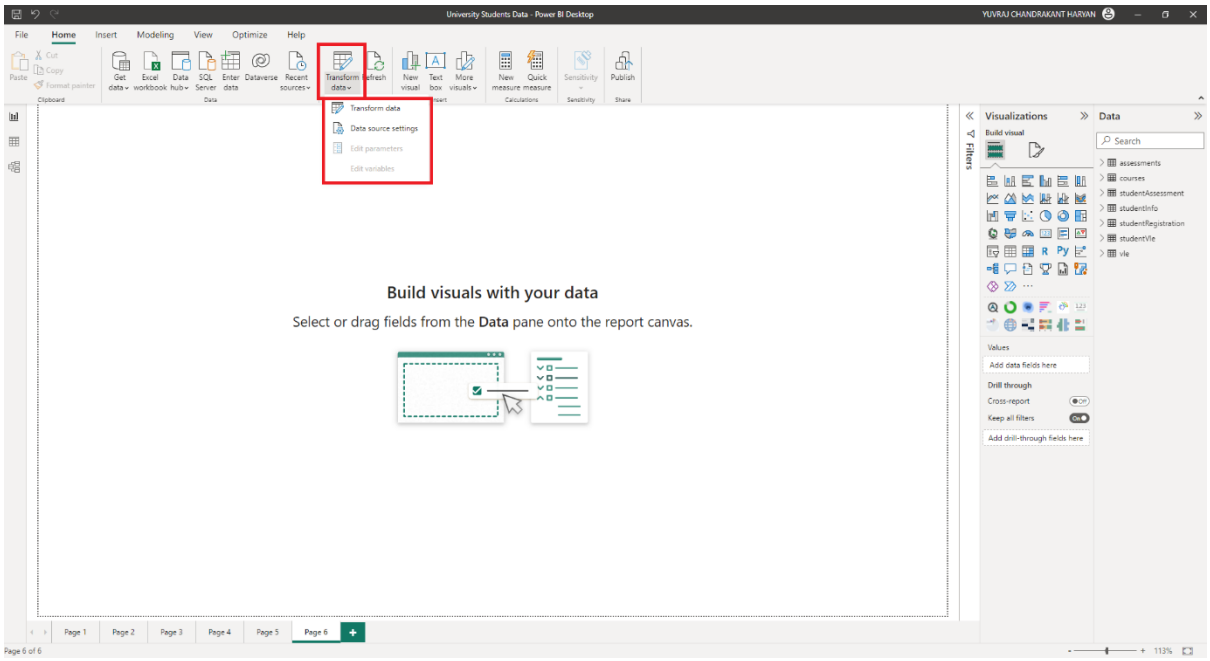


Fig 10 : Transform data option

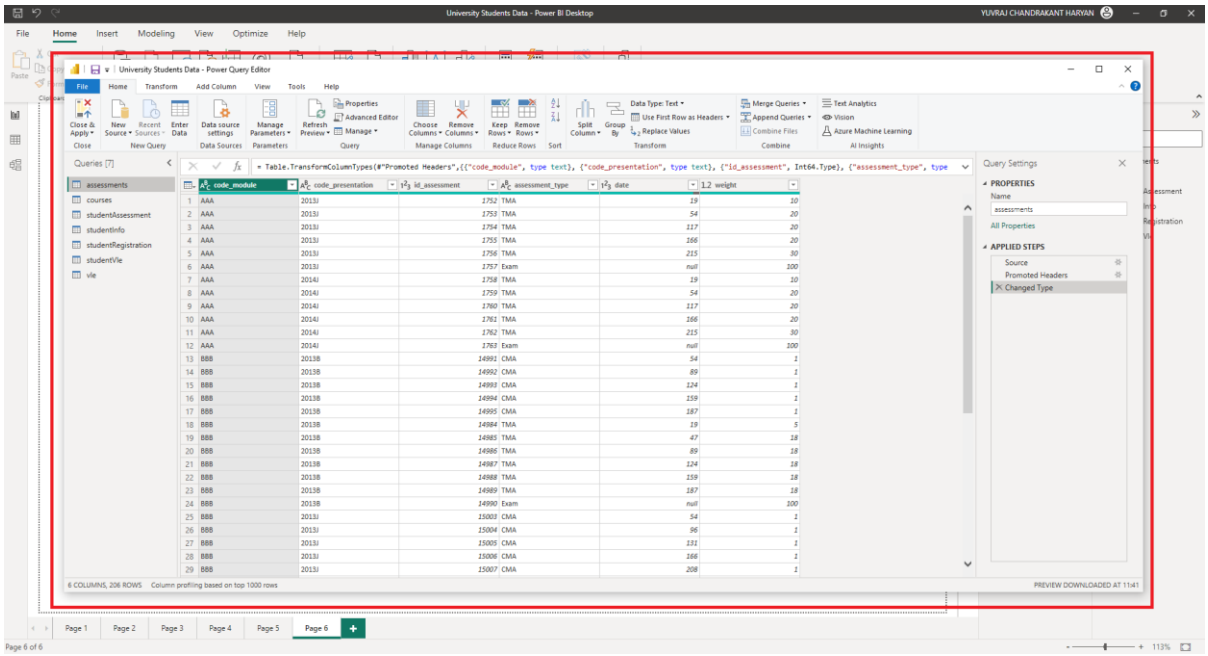


Fig 11 : Power Query Editor Dialog box

Promoting Headers :

It can be seen in Fig 12 that first row of my data table has its headers.

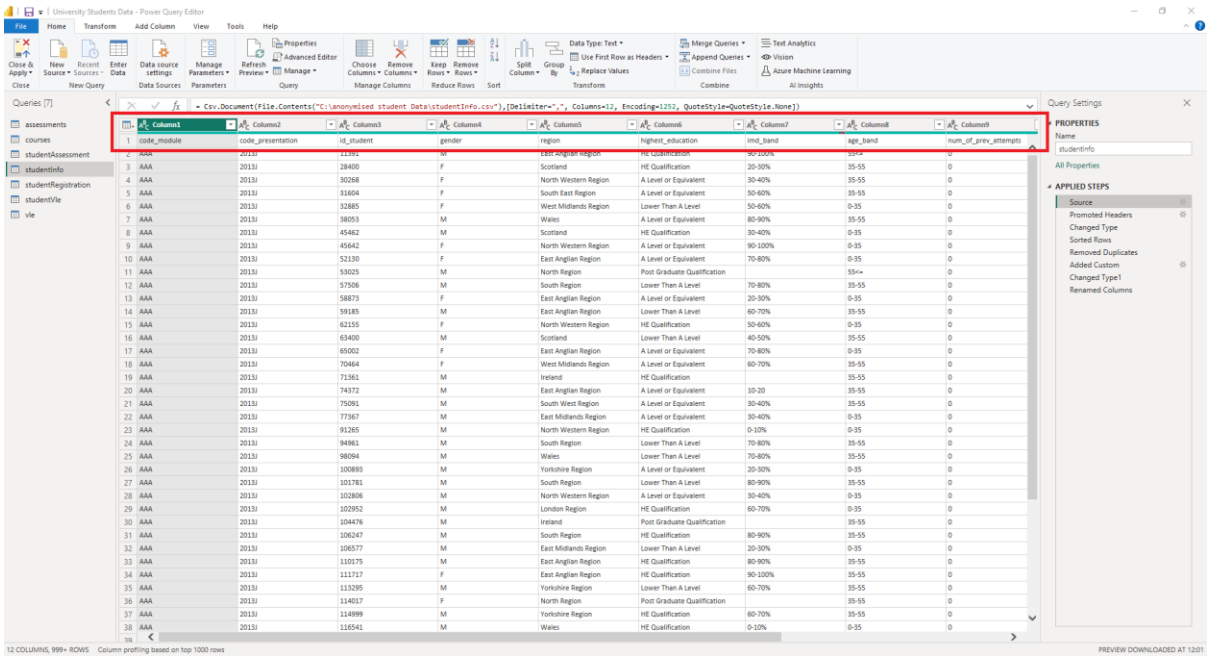


Fig 12 : First row of data table has the headers

Power BI can be instructed to extract the headers of a data table from its first row. This is called promoted headers which is displayed in Fig 13 and Fig 14

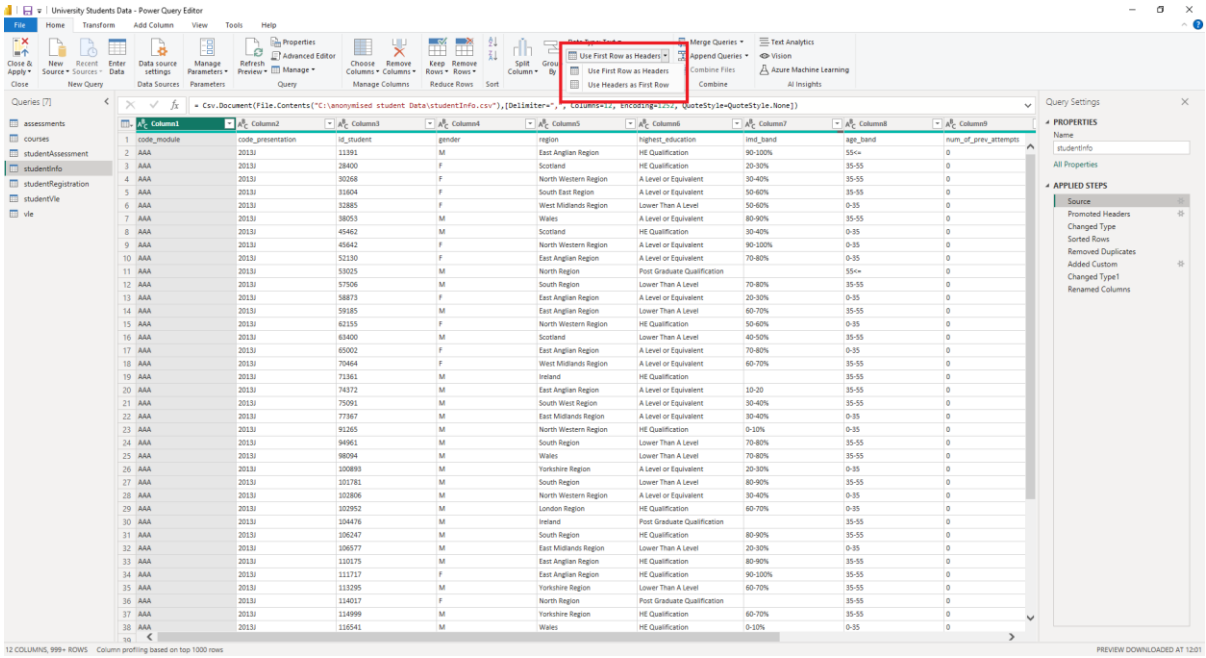


Fig 13 : Promoted Headers option

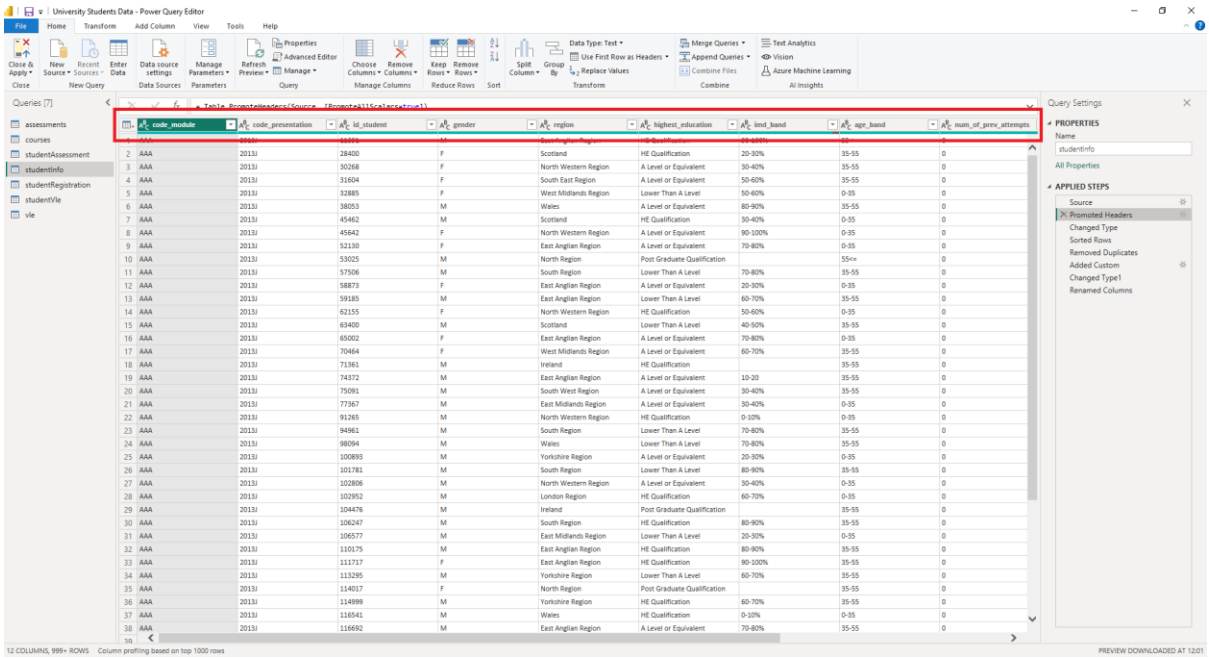


Fig 14 : Promoted Headers

This step has been carried out on all the other tables as well.

Changing Data Type :

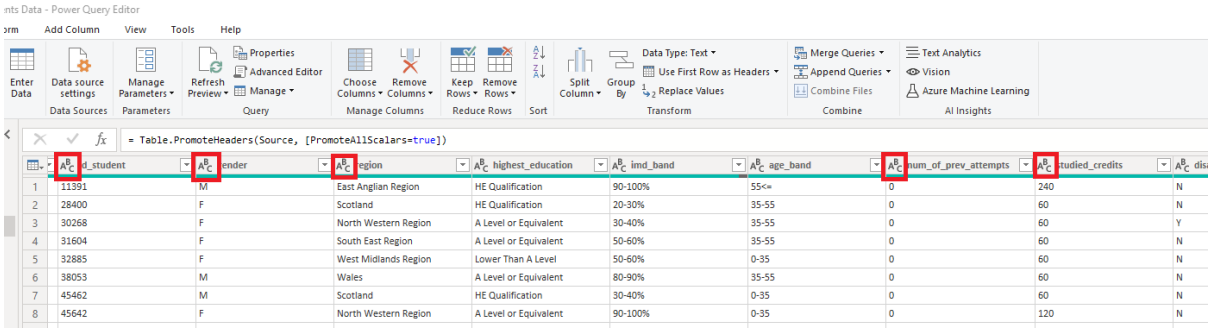


Fig 15 : Data Type of Columns

The data types of the column can be seen by the small icon on the left of the column header as shown in the Fig 15. As all the columns in this dataset are set to text Type, they need to be altered according to the requirement.

The small icon indicating the data type of the column can be clicked to open a drop-down of different available data types as shown in the Fig 16. As this is the ‘studied_credits’ column which should contain numeric values only. The whole number data type was selected.

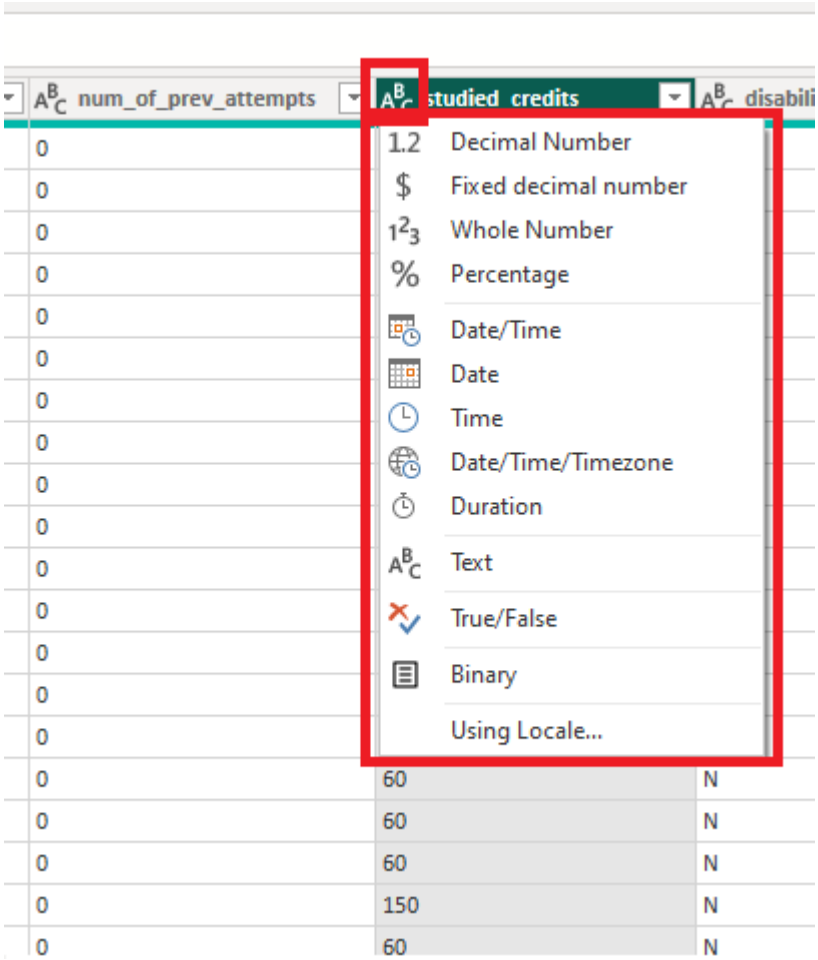


Fig 16 : Changing Data type

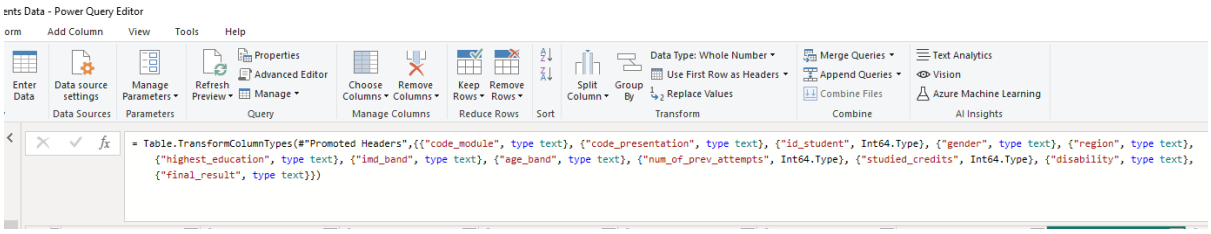


Fig 17 : Changing Data type using M language

Similarly, the data types of the columns can even be changed using the M language query as shown in the Fig 17.

Further, the ‘id_student’ and ‘num_of_prev_attempts’ columns in the studentInfo table needs to be changed to the numeric type following the same steps. The Altered types are displayed in the Fig 18.

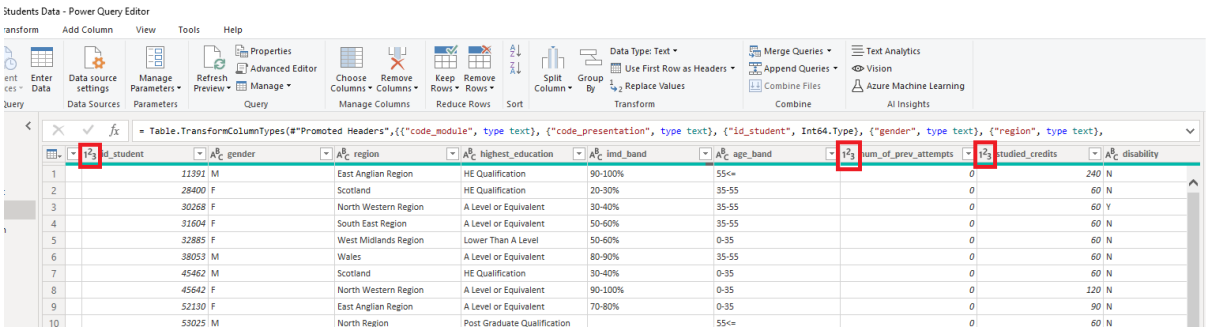


Fig 18 : Changed Data type

The same procedure was followed to altered the data types of the desired columns in rest of the tables of the dataset.

Sorting Rows of the table :

As the ‘id_student’ column is the primary key in the studentInfo table it makes sense sorting the table according to the student id.

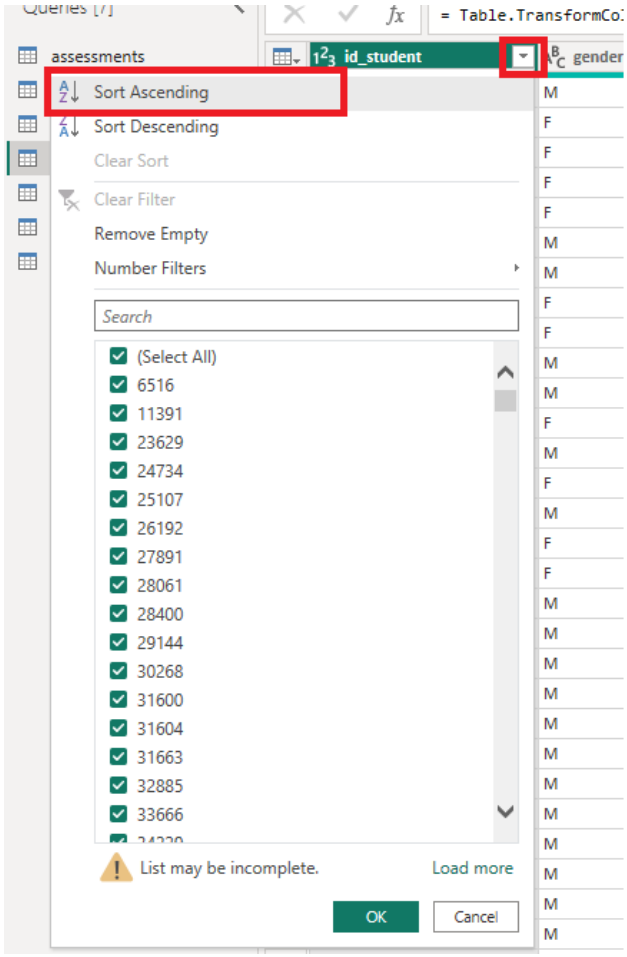


Fig 19 : Sorting Rows

This can be done by clicking on the drop-down next to column name and then selecting the ‘sort ascending’ option. This sorts the rows of the table in the ascending order of student ids.

Alternatively, this can be achieved using the following M language code.

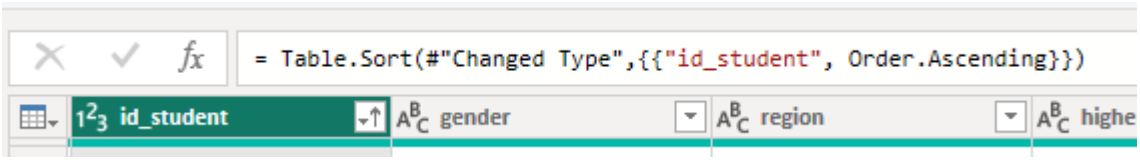


Fig 20 : Sorting Rows using M language

Filtering Rows :

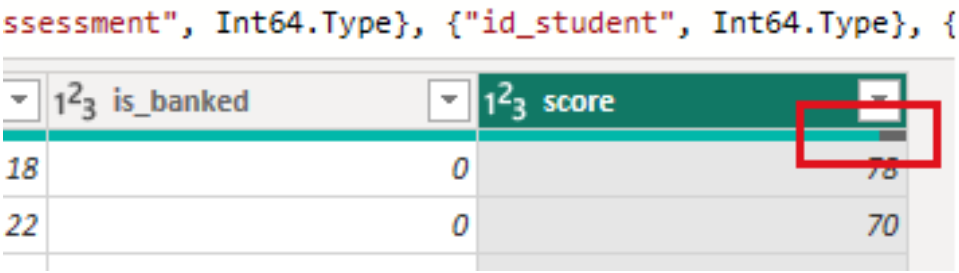


Fig 21 : Invalid column values

The bar not being full indicates that the ‘score’ column has some invalid values in it such as - ‘null’, ‘ ’, ’-’, ‘N/A’, etc.

The drop-down can be clicked to have a look at all the unique values present in that column. It is seen that ‘null’ values are selected, they can be deselected to exclude the rows containing null value in the ‘score’ column as done in Fig 22.

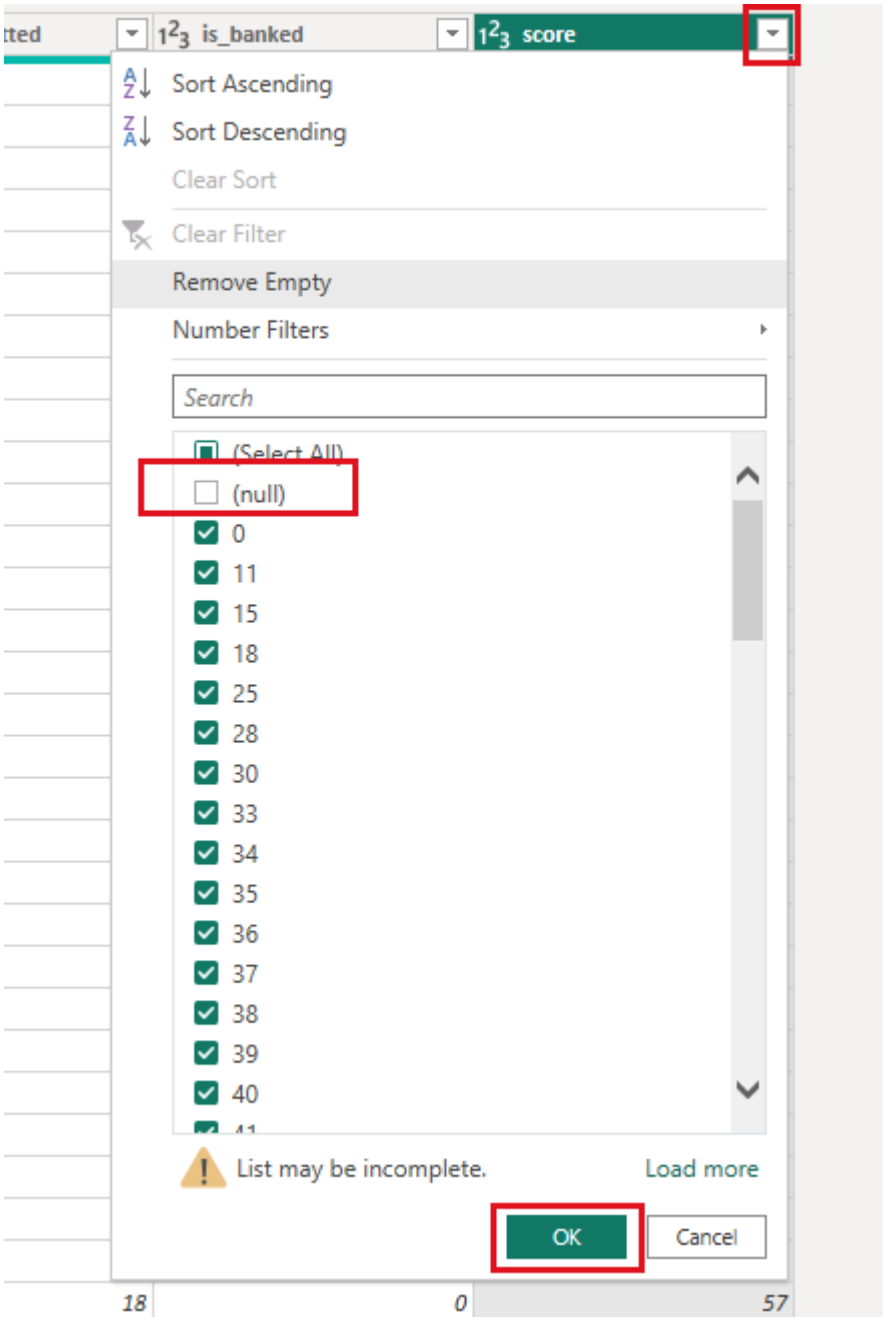


Fig 22 : Filtering Rows

The same result can be achieved using M Language code refer Fig 23

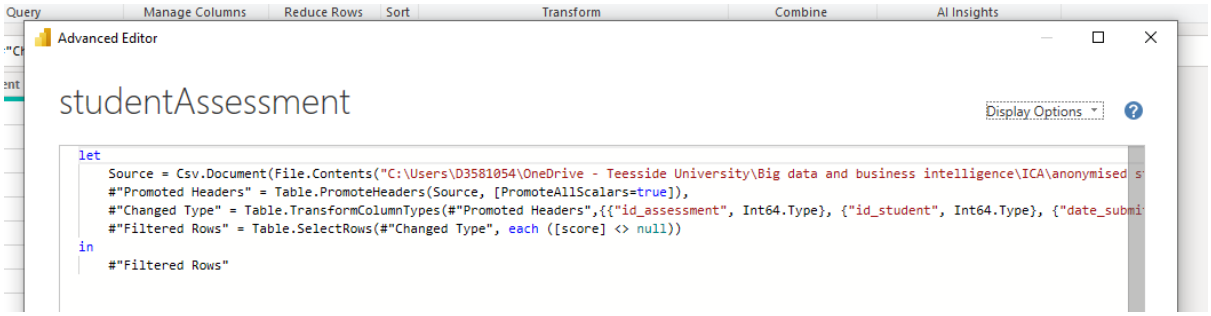


Fig 23 : Filtering Rows using M language

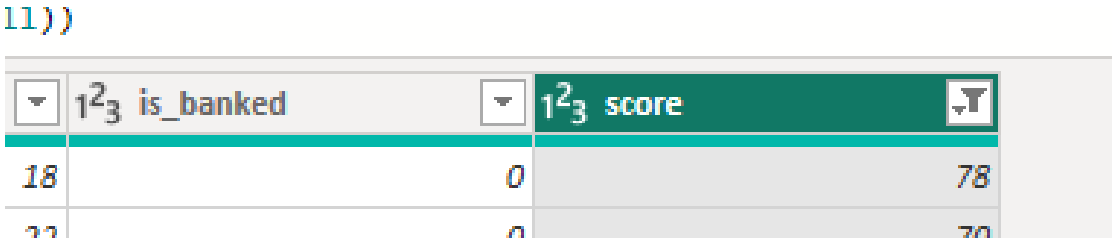


Fig 24 : Filtered Rows

The null values are removed, it is evident by the bar below the column name which is filled now as displayed in Fig24.

3. Data Modelling

Fig.25 shows how the data model is looking after the data cleaning process has been completed.

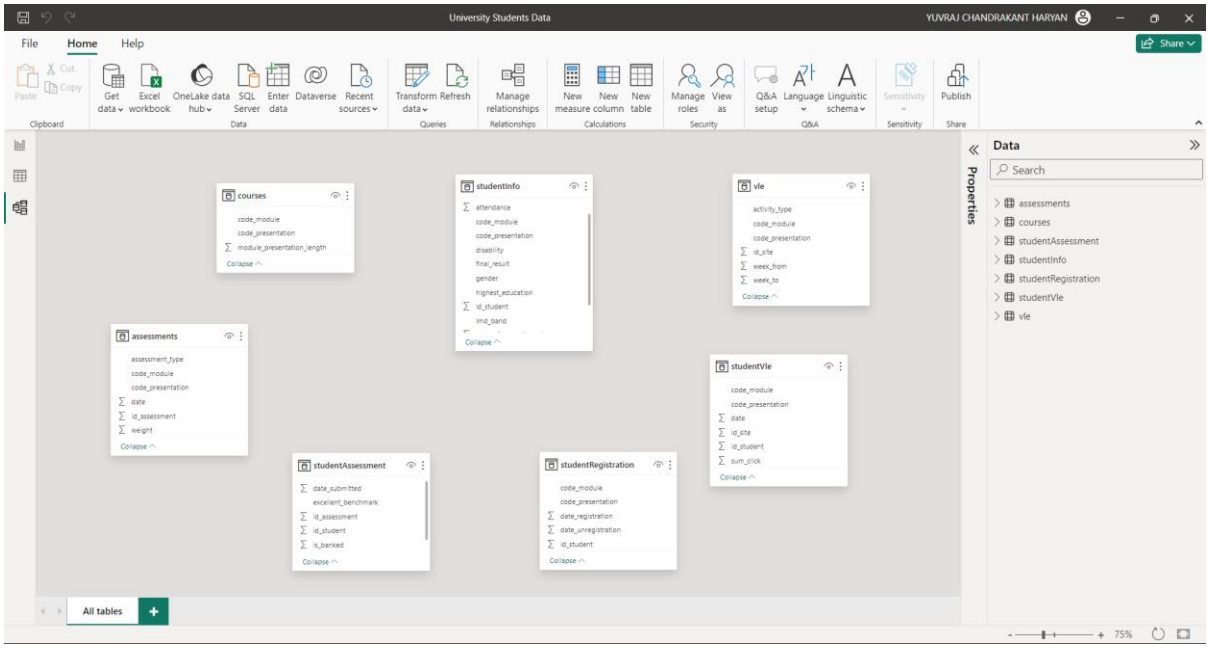


Fig 25 : Data Model after Data cleaning

It can be evidenced that there no active relationships among the tables yet.

A relationship between ‘studentInfo’ and ‘studentAssessment’ table can be created as ‘id_student’ the primary key of studentInfo table is referred as a foreign key in the studentAssessment table.

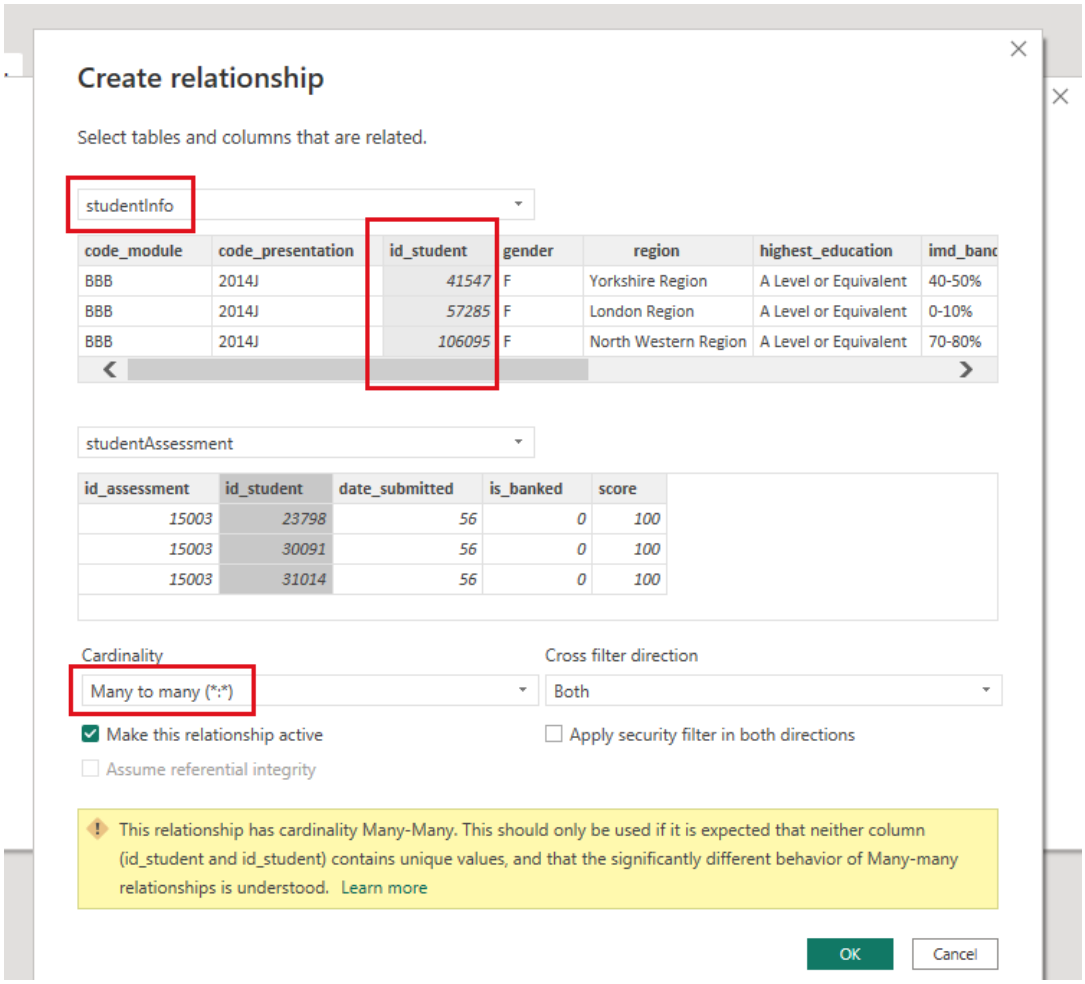


Fig 26 : Many to many cardinality while creating relationships

As the studentInfo table and studentAssessment table both have repetitive values in the id_student column, the cardinality ratio is shown as many to many in Fig 26.

Deleting the duplicate values in the id_student column of the studentInfo table to make all its value unique is displayed in the Fig 27. Now id_student can be considered as the primary key column of the studentInfo table.

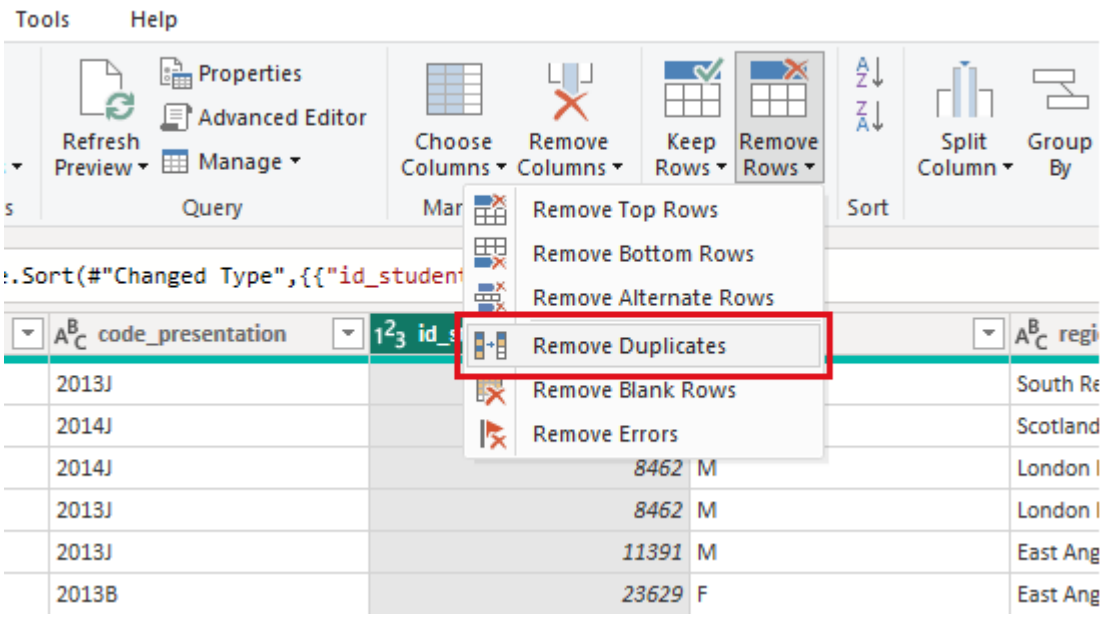


Fig 27 : Removing Duplicate Rows

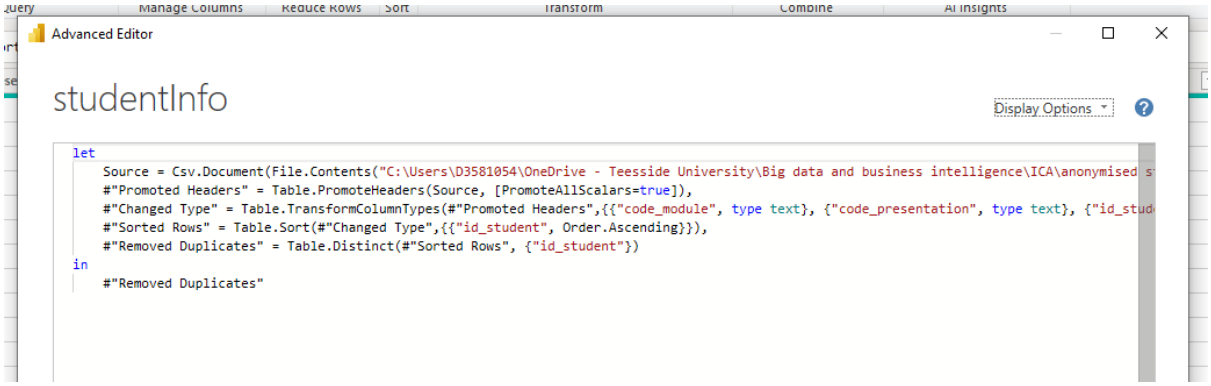


Fig 28 : Removing Duplicate Rows using M language

The tables in our model need to be related to one another in order to carry out the analysis needed for this project. In order to accomplish this, choose “Manage relationships” from the data model section's home tab. When a dialogue box appears, ‘New’ is chosen. Based on the common column “id_student” the first relationship is between studentInfo (fact table) and studentAssessment (dimension table) table.

Now it is seen in Fig 29 that it is allowed to form a one to many relationship between the studentInfo and studentAssessment table.

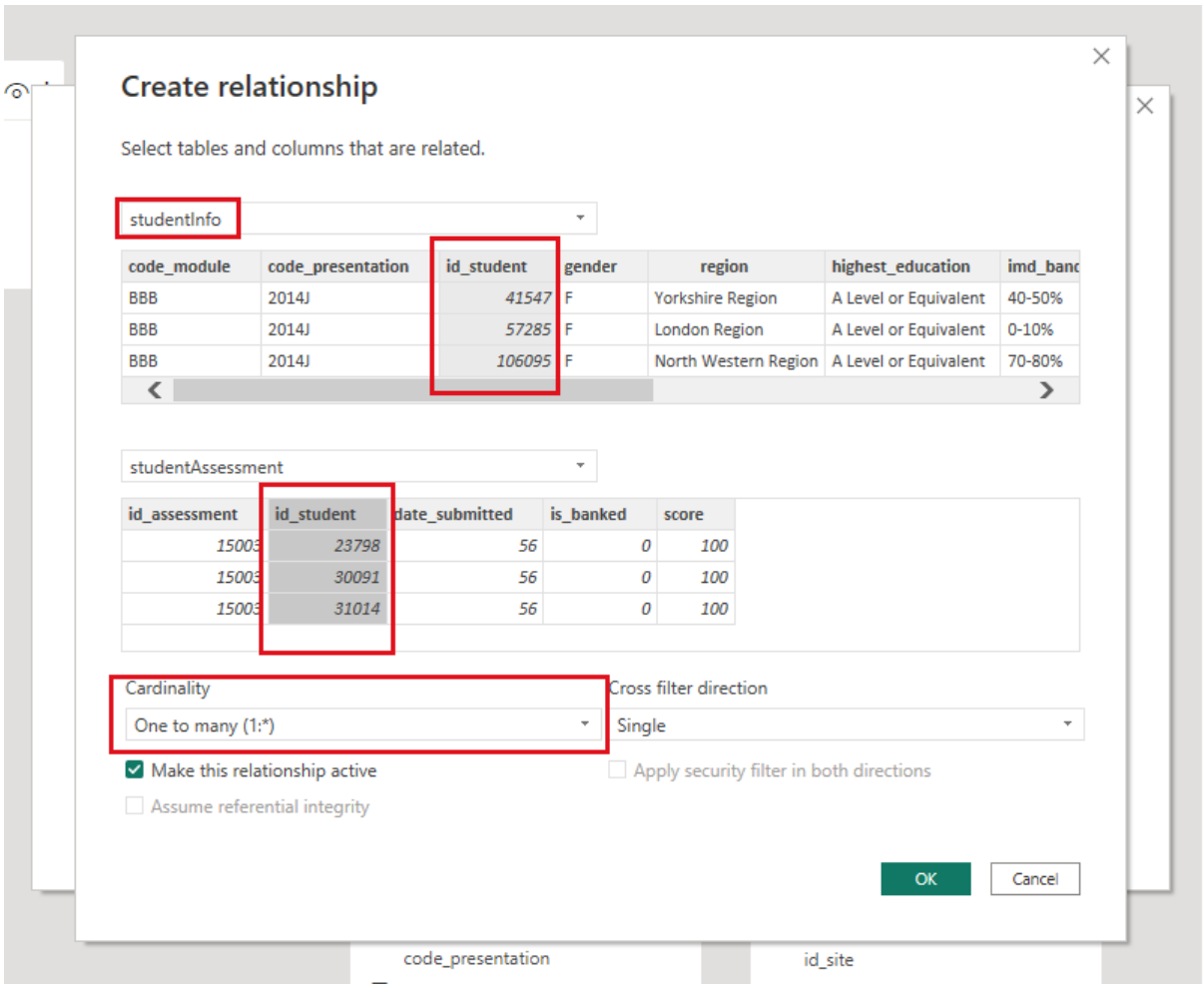


Fig 29 : Creating a one to many relationship

Same steps can be repeated to form one to many relationships between ‘studentInfo’ table and ‘studentRegistration’, ‘studentVle’ tables respectively based on common column ‘id_student’. Also, relationships are formed between ‘assessments’ and ‘studentAssessment’, ‘vle’ and ‘studnetVle’ tables on the basis of columns ‘id_assessment’ and ‘id_site’ respectively.

All of these active relationships are showcased in Fig 30.

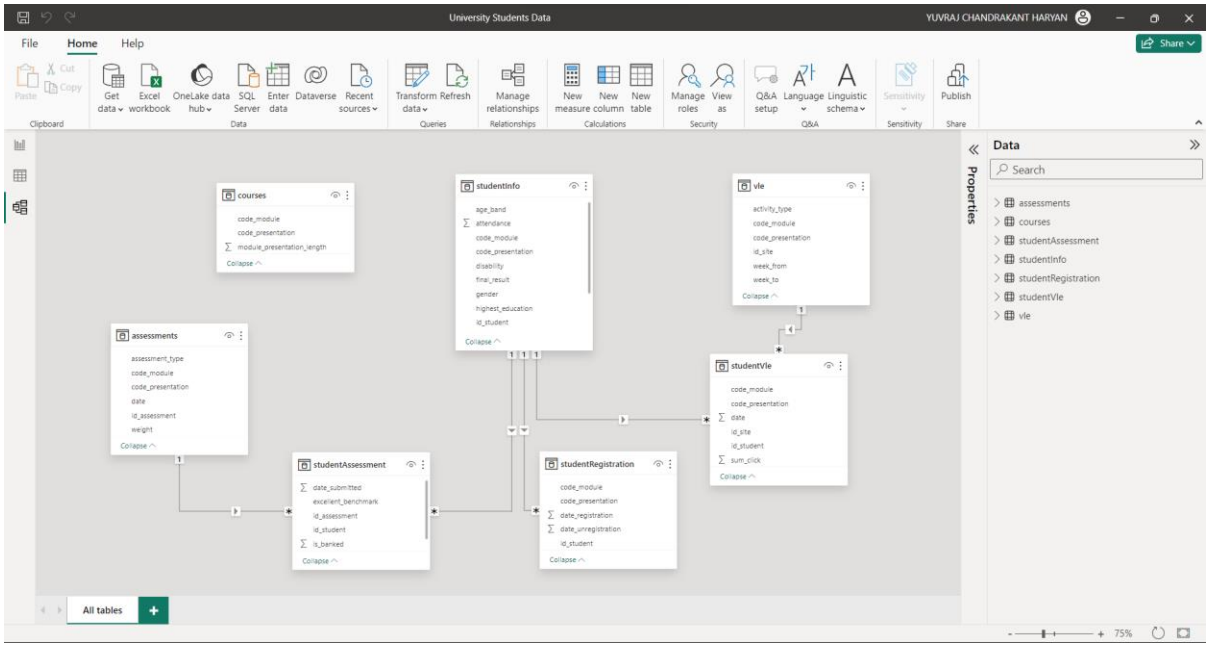


Fig 30 : Data Model with Active Relationships

UNIVERSITY STUDENTS DATA ANALYSIS

SECTION 2: BUSINESS INTELLIGENCE SOLUTION

NAME: YUVRAJ CHANDRAKANT HARYAN

STUDENT ID: D3581054

SUBMISSION DATE : 09/01/2024

1. Executive Summary

1.1 Introduction

The report seeks to answer the following questions:

1. How does the student's previous education level impact their performance? (KPI: Average score based on highest education level)

2. How well do students are making use of the Virtual Learning Environment (VLe) ?

3. Are there any trends in the way students use the VLe and are there any resource types that are popular among students?

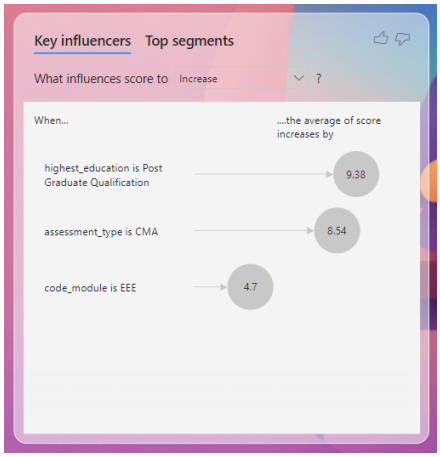
4. Is there any trend in the way of students choosing the courses?

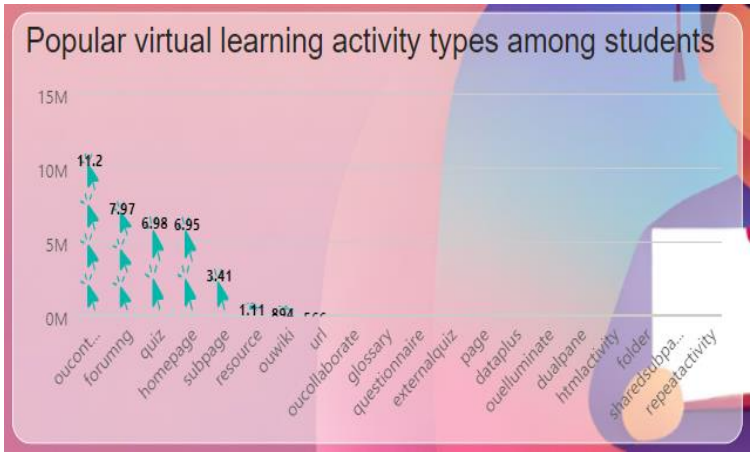
5. From what region are most of the students at the University?

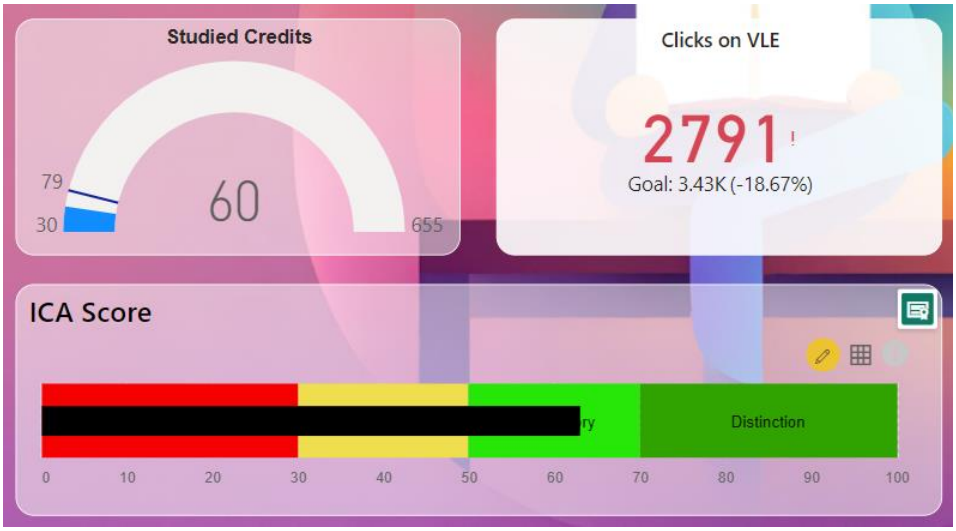
6. Can a portal-like page where all the student data is viewed at a glance can be made using power BI?

1.2 Key Findings

- ✓ The highest education level being ‘Post Graduate’, increases the average of score by 9.38
- ✓ A figure representing every students ‘number of clicks’ on the VLe resources is pinned on the Students Portal page. The number for each student can be compared with a goal of ‘3.43K’ clicks which is the average clicks on the VLe by students.
- ✓ The most popular VLe resource unveiled to be ‘oucontent’ and the students are found highly using the VLe for the ‘FFF’ course.
- ✓ ‘BBB’ and ‘FFF’ were found to be the most common courses opted by student while ‘AAA’ being the least popular. Majority of females are pursuing the BBB course meanwhile most males belong to the FFF course.
- ✓ Majority of the students in the university turned out to be domestic students i.e from the UK itself.
- ✓ A ‘Students Portal’ page has been created on the dashboard displaying detailed information regarding the student. The details displayed include his/her age band, gender, region, academic credits, number of clicks on VLe, attendance, ICA score and final result.







1.3 Recommendations

- The number of clicks on VLe for each student should be visible to them, so that they can judge how active they are on the VLe.
- Unpopular VLe resources should be made interesting taking inspirations from the ‘oucontent’ resource.
- Infrastructure in the BBB and FFF courses should be prepared for large number of students as these are the most popular courses and are chances that remain popular in the future also.
- The AAA course syllabus should be scrutinized and modified to increase its student count.
- Scholarships and bursaries should be examined or introduced to encourage the admission of international students in the university.

2. Introduction

Embarking on a Business Intelligence (BI) analysis of the University Students Dataset signifies a strategic exploration into the realms of academic data. This initiative aims to uncover valuable insights that can inform decision-makers, optimize educational strategies, and enhance the overall student experience. Through a focused examination of assessments, courses, student information, and virtual learning activities, this analysis seeks to distil actionable knowledge from the dataset, contributing to informed decision-making in the dynamic landscape of higher education.

The analysis aims to address the following queries :

1. How does the student's previous education level impact their performance? (KPI: Average score based on highest education level)
2. How well do students are making use of the Virtual Learning Environment (VLe) ?
3. Are there any trends in the way students use the VLe and are there any resource types that are popular among students?
4. Is there any trend in the way of students choosing the courses?
5. From what region are most of the students at the University?
6. Can a portal-like page where all the student data is viewed at a glance can be made using power BI?

2.1 Dataset

The Anonymised University Students Dataset downloaded from the analyse.kmi.open.ac.uk includes the "studentAssessment", "studentInfo", "studentVle", "vle", "assessments" and "courses" tables. Section 1 of this report includes a detailed description of the dataset along with a link to the source.

2.2 Data Model

The final data model after the data modelling process is shown below in the Fig 1. The first section of this report contains a detailed explanation of the data modelling procedure. The dataset includes 6 main tables of which studentInfo is the fact table table and other are dimension tables. As primary key in the studentInfo table is referred as foreign key in most of the other tables.

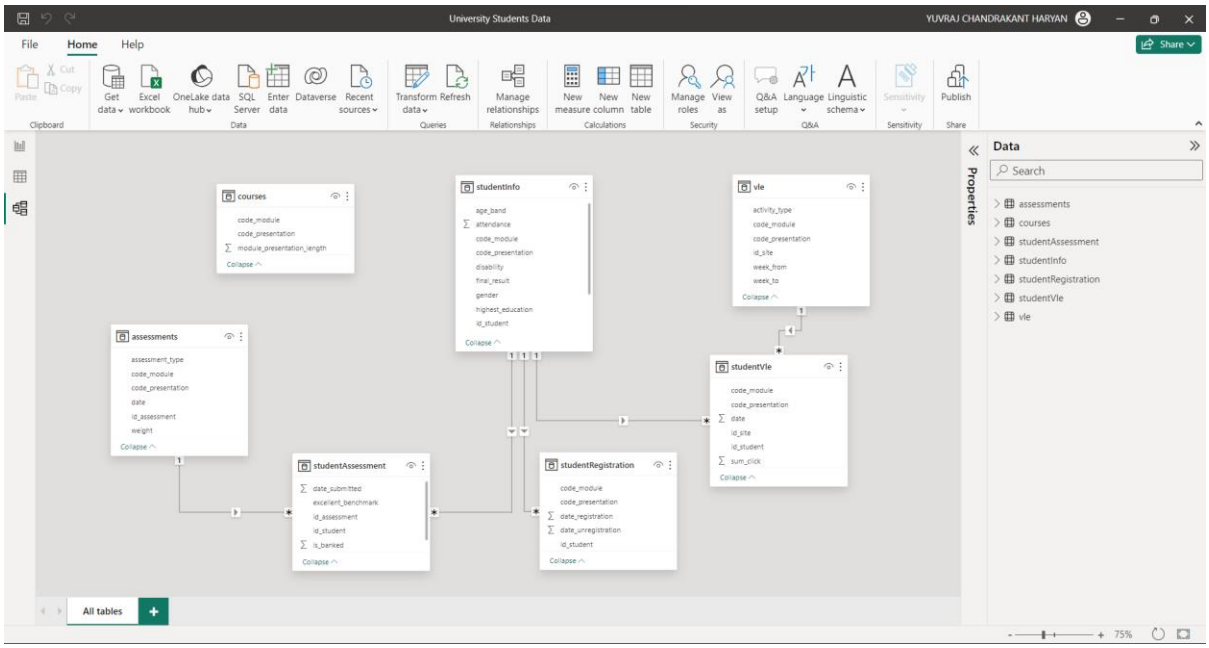


Fig 1 : Data model

3. Findings based on analysis and evaluation

Various analysis were carried out and some visualisations were generated in order to answer the Business Intelligence questions mentioned above.

Question 1 : How does the student's previous education level impact their performance? (KPI: Average score based on highest education level)

The key influencer chart generated in the BI dashboard has coined that the average of score increases by 9.38 if the previous highest qualification is Post Graduate. It can be evidenced below in the Fig 2.

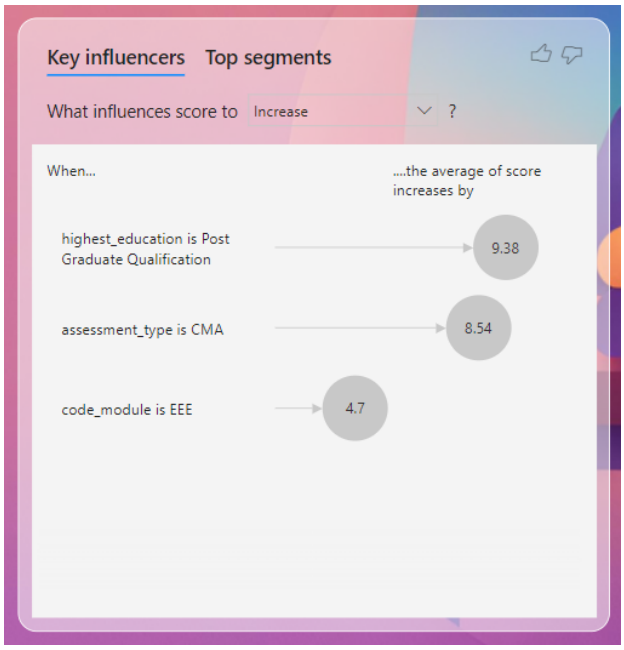


Fig 2 : Key influencers in increasing the average of score

Question 2 : How well do students are making use of the Virtual Learning Environment (VLe) ?

On the Students Portal page, a figure that represents each student's "number of clicks" on the VLe resources is pinned. Each student's number can be compared to the target of "3.43K" clicks, which represents the average number of clicks made by students on the VLe.

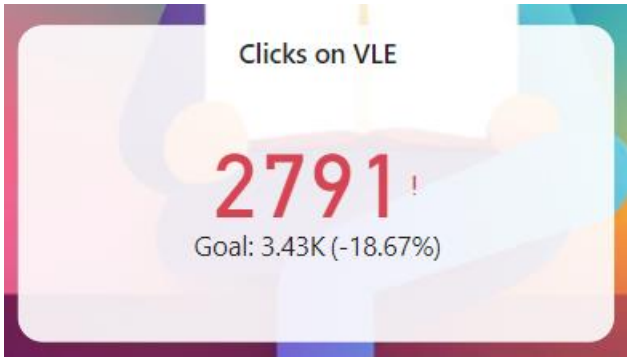


Fig 3 : number of clicks on VLe by a student vs the goal

Fig 3 shows a card visual which is pinned on the student portal. It shows the number of clicks of an individual student and the color represents if its below the target or above. Red means the number of clicks are lacking the goal while green means that the target is achieved.

For calculating the target 2 custom columns were introduced to the dataset. These columns were calculated using the following DAX code :

- **average_clicks_col =**
  ROUND(
  AVERAGE('studentVle'[average_clicks_per_student]),
  2
)

  This column is the Goal/target number of clicks to be achieved. The code calculates the average of ‘average clicks per student column’ i.e. the average clicks by all university students together and rounds them up to 2 decimal places.

- **average_clicks_per_student =**
  ROUND(
  CALCULATE(
  SUM('studentVle'[sum_click]) /
  COUNTROWS(VALUES('studentVle'[id_student])),
  ALLEXCEPT('studentVle', 'studentVle'[id_student])
  ),
  2
)

  This column is the average clicks on VLe by every individual student. The code calculates the total sum of clicks by a single student and divides it by the number of row entries that the student has in the table (i.e. calculates the average of sum clicks per student) and round it up to 2 decimal places.

Question 3 : Are there any trends in the way students use the VLe and are there any resource types that are popular among students?

"oucontent" emerged as the most popular VLe resource. Students are found to be heavily utilising the VLe for the "FFF" course.

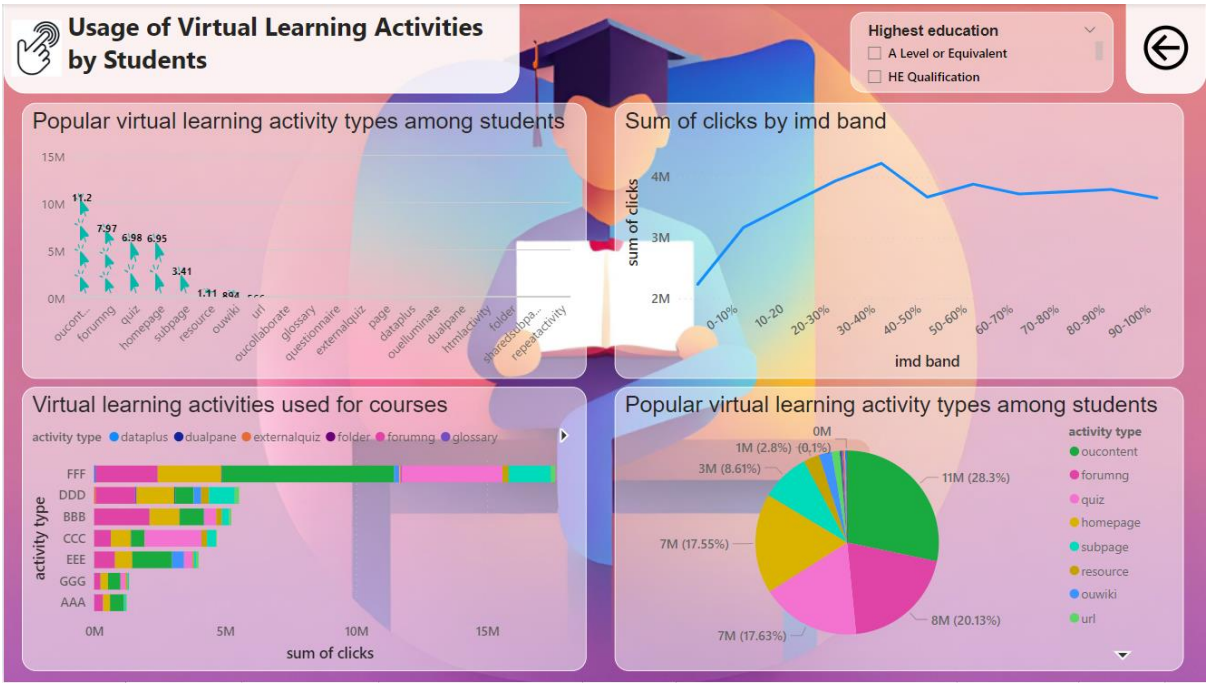


Fig 4 : VLe activities dashboard page

Fig 4 displays the dashboard page which holds the visualisations used to explain the above Business Intelligence question.

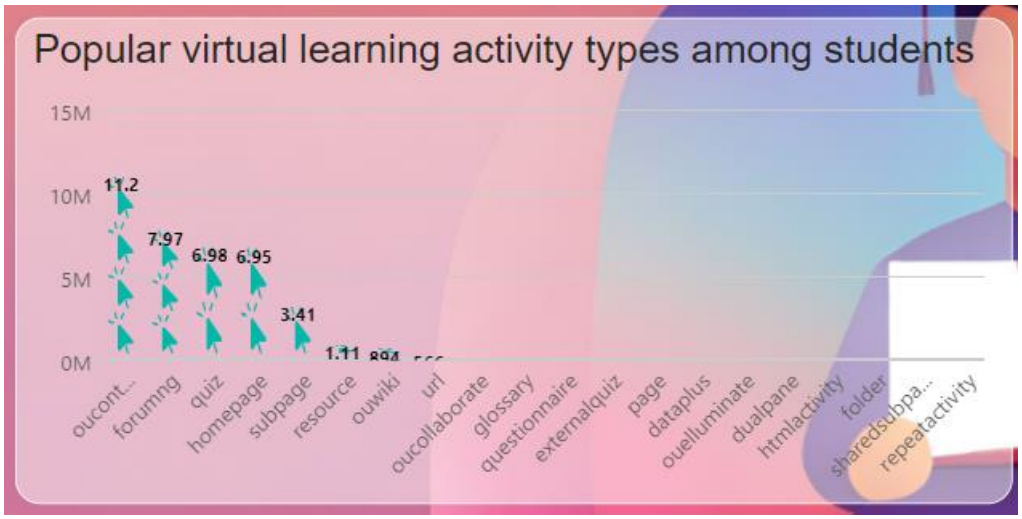


Fig 5 : Infographic column chart representing Virtual Learning activity vs sum of clicks

Fig 5 shows that ‘oucontent’ activity type has clearly outperformed all the other VLe activity types. It leads the race by 11M sum of clicks by the students.

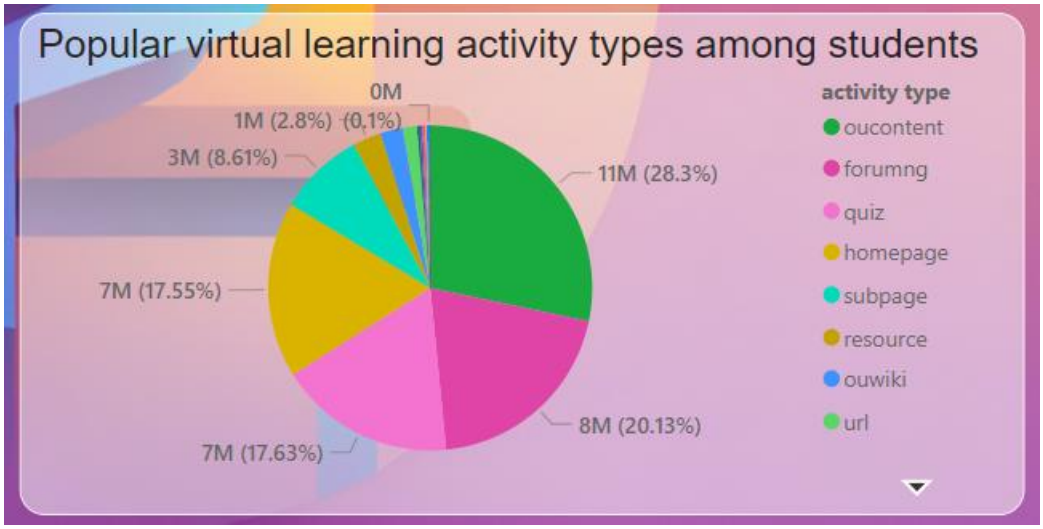


Fig 6 : Pie chart of sum of clicks for the VLe activity types

Even Fig 6 confirms that ‘oucontent’ has secured the largest portion within the sum of clicks on VLe activities by the students. Making it undeniably the most popular VLe resource among the students

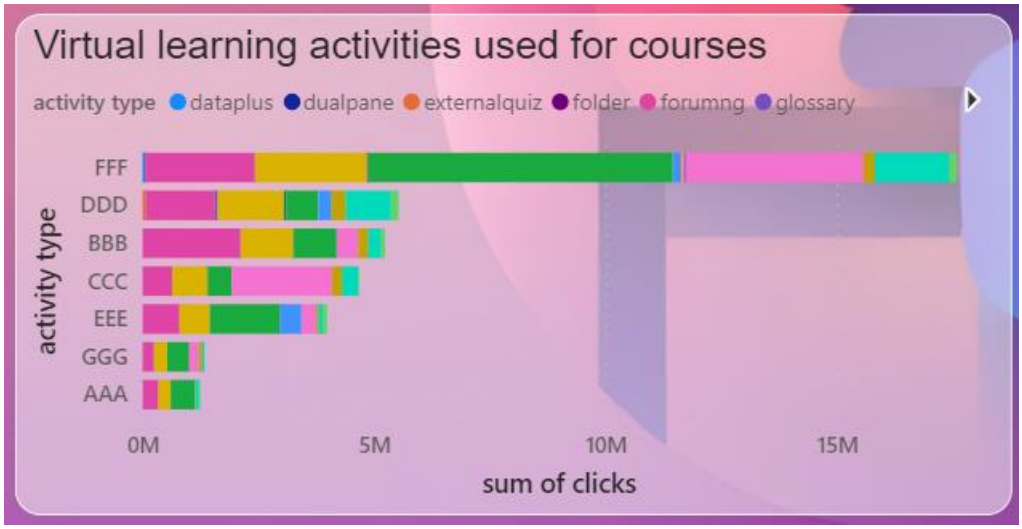


Fig 7 : Bar chart representing the use of different types of VLe activities in various courses

Observing Fig 7 it can be seen that VLe activites are largely used in the FFF course by the students. The sum of clicks on the VLe activities for the FFF course surpasses 15M.

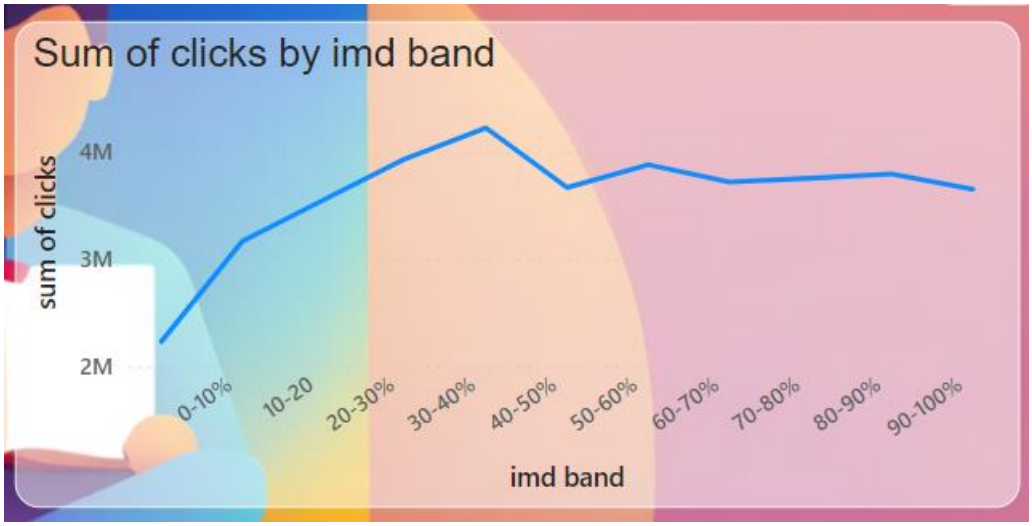


Fig 8 : Line graph representing imd band of students vs there number of clicks on VLe

In Fig 8 it can be seen that VLe is most used by the students which fall in the 30-40 % imd band. On the other hand is least used by students who fall under the 0-10% imd band.

Question 4 : Is there any trend in the way of students choosing the courses?

It was discovered that students most frequently chose "BBB" and "FFF" courses, with "AAA" being the least preferred. While most males are enrolled in the FFF course, the majority of females are pursuing the BBB course.

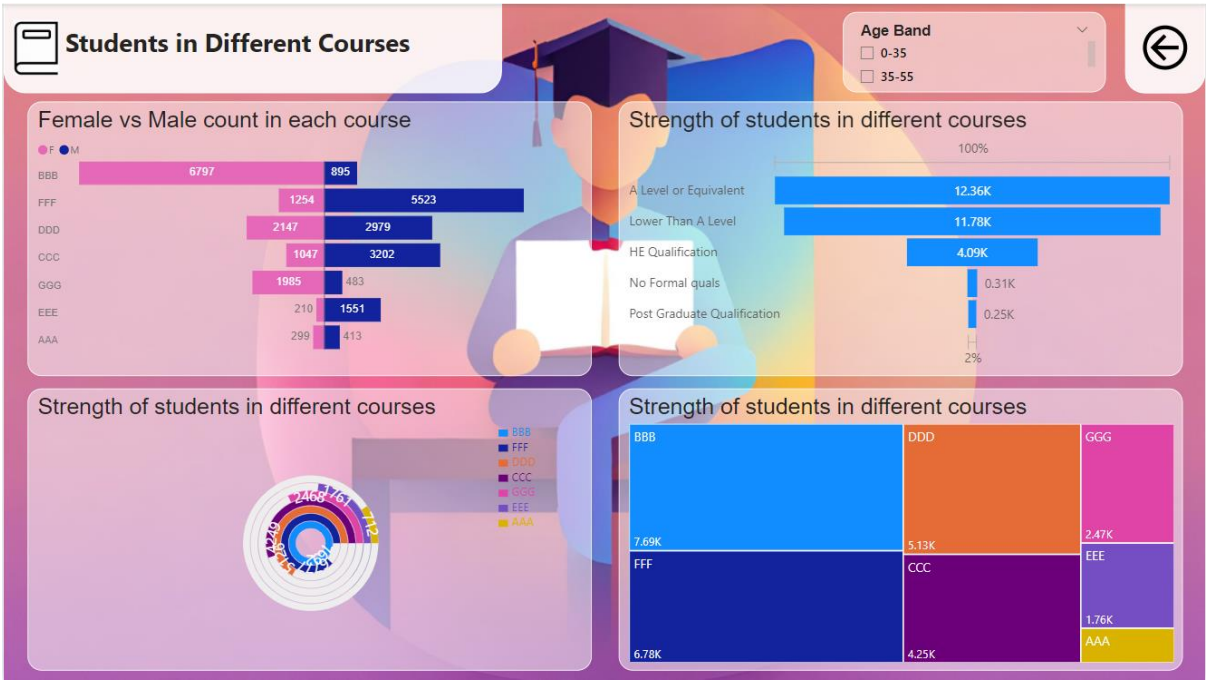


Fig 9 : Courses Dashboard page

Fig 9 displays the dashboard page which holds the visualisations used to explain the above Business Intelligence question.

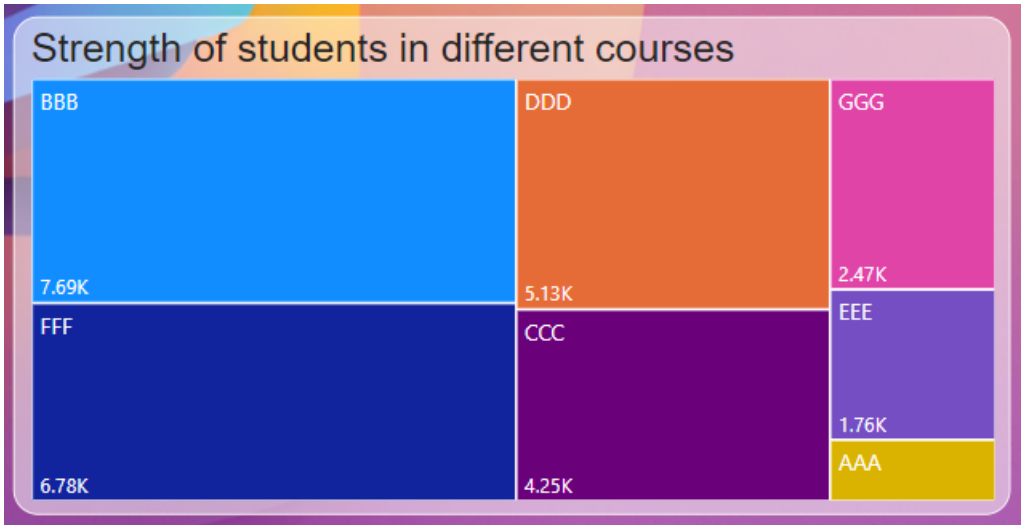


Fig 10 : Treemap of number of students in each course

In Fig 10 it's visible that courses BBB and FFF share the largest portion of student count with 7.69K and 6.78K students respectively. While AAA holds the least student count.

Fig 11 : Radial chart of number of students in each course

Even Fig 11 bargains for the same showing BBB and FFF as the most popular courses among students while AAA stands to be the least popular one.

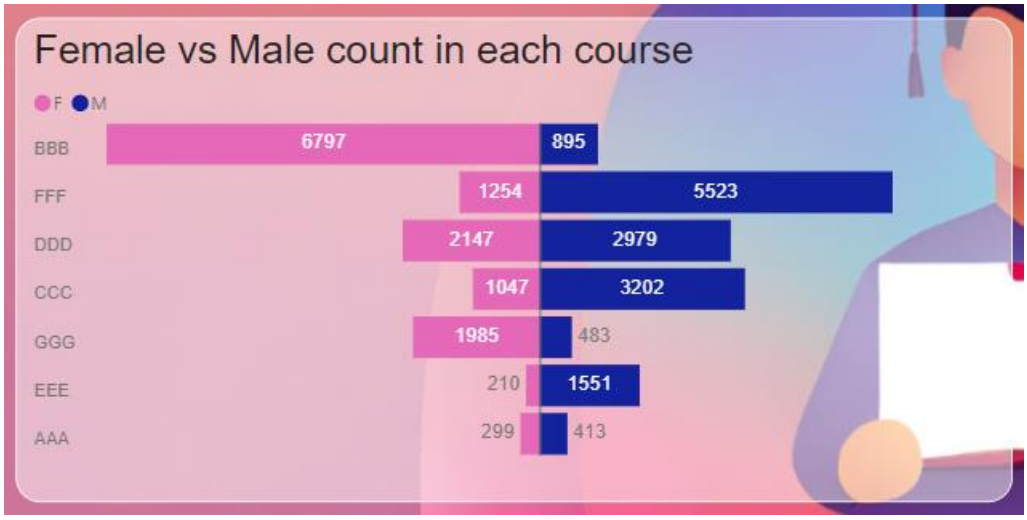


Fig 12 : Tornado chart comparing number of males and females in each course

It can be seen as majority of females have opted for the BBB course with a count of 6.79K. While most males have opted for the FFF course with a count of 5.52K.

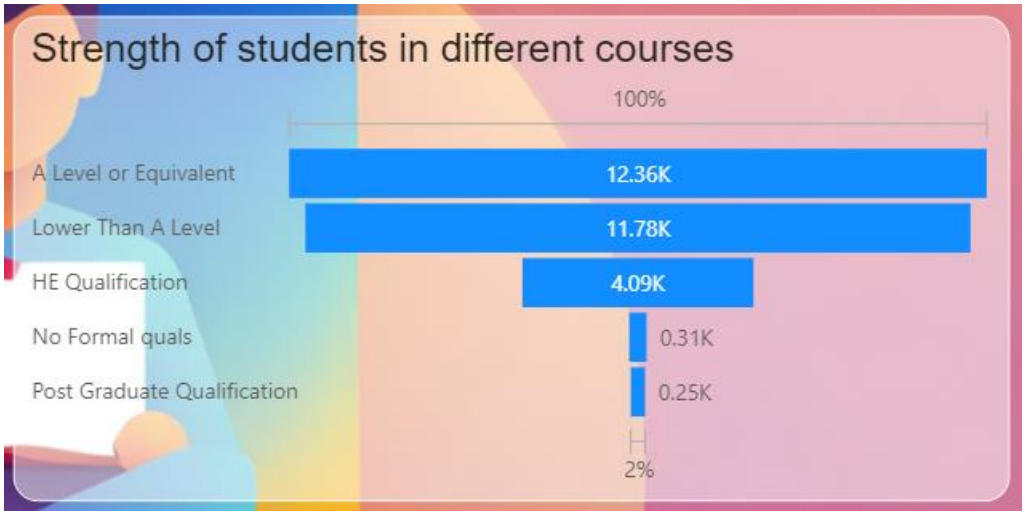


Fig 13 : Funnel chart representing the count of previous highest education of students in different courses

By looking at Fig 13 it is evident that majority of students in the university have their previous highest education as ‘A Level or Equivalent’ or ‘Lower Than A Level’. The university has negligible count of Post Grads enrolled.

Question 5 : From what region are most of the students at the University?

It turned out that domestic students, or those from the UK, made up the majority of the university's student body.

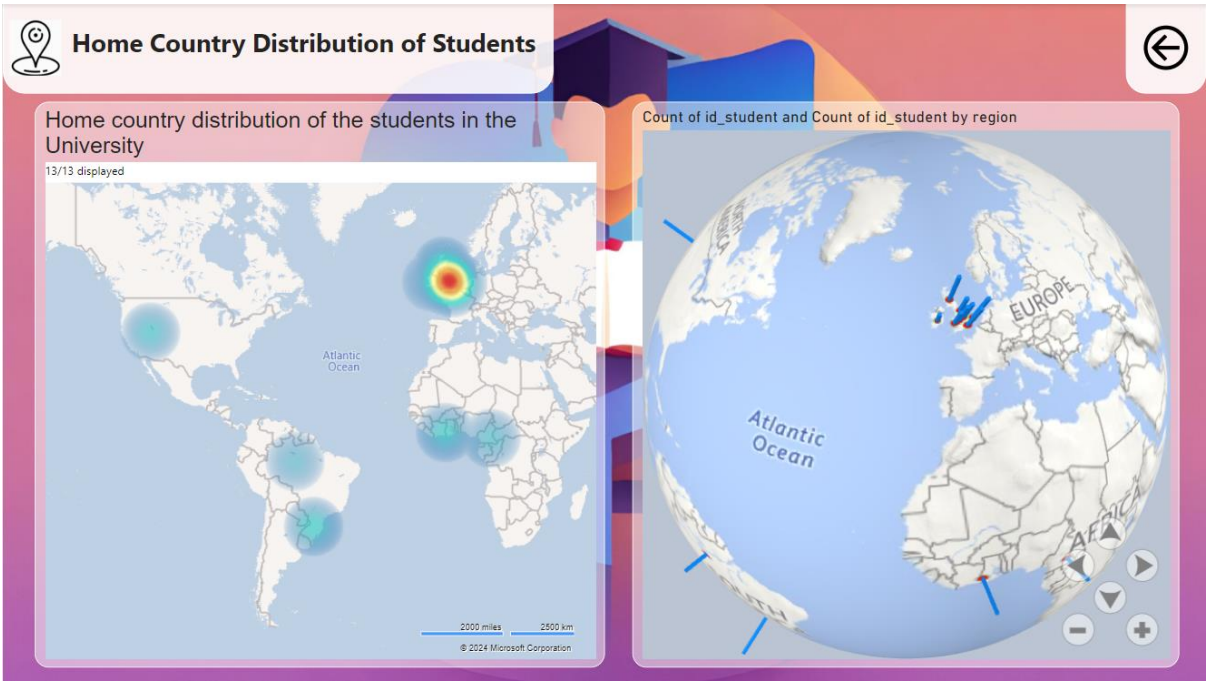


Fig 14 : Home country distribution of students Dashboard page

Fig 14 displays the dashboard page which holds the visualisations used to explain the above Business Intelligence question.

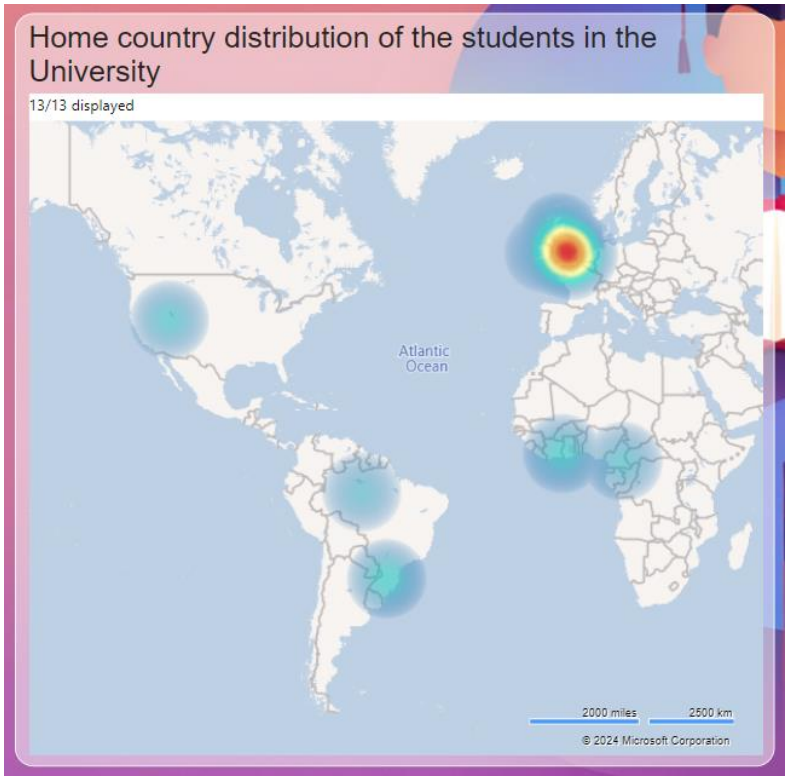


Fig 15 : Heat map of Home country distribution of students

Observing Fig 15 it can be known that most of the students studying in the university belong to the UK, as a hotspot can be seen on the world map on the UK region. This implies that there is a majority of domestic students in the university.

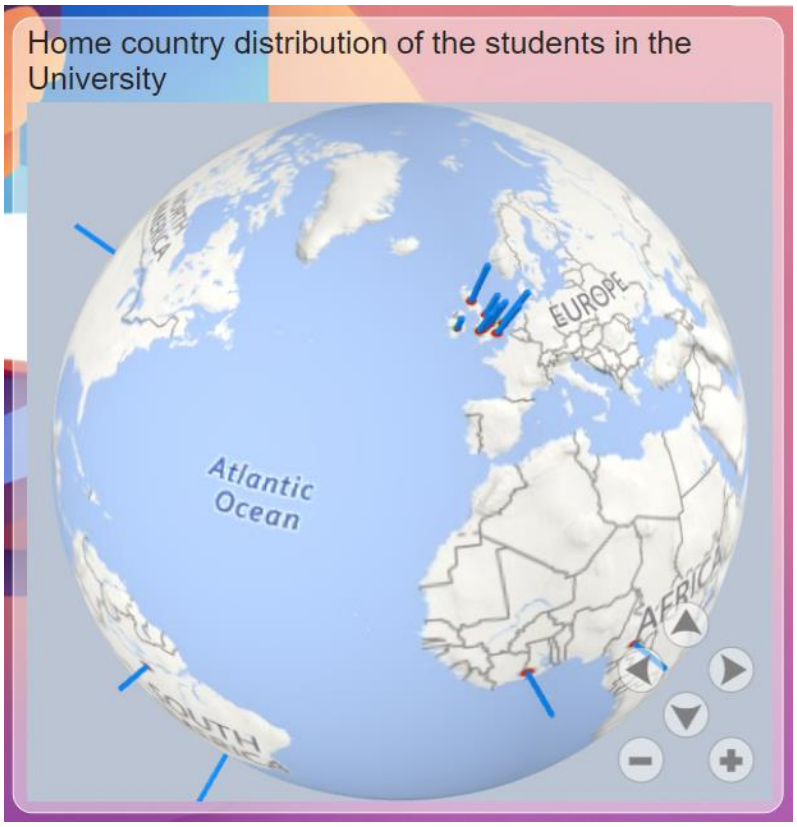


Fig 16: Globe map of Home country distribution of students

Even the Fig 16 suggests the same as most of the candles are concentrated in the Uk region. Here candles represent the count of students from that region.

Question 6 : Can a portal-like page where all the student data is viewed at a glance can be made using power BI?

On the dashboard, a "Students Portal" page has been created that offers comprehensive student information. His or her age band, gender, region, academic credits, number of clicks on VLe, attendance, ICA score and final result are among the details shown.

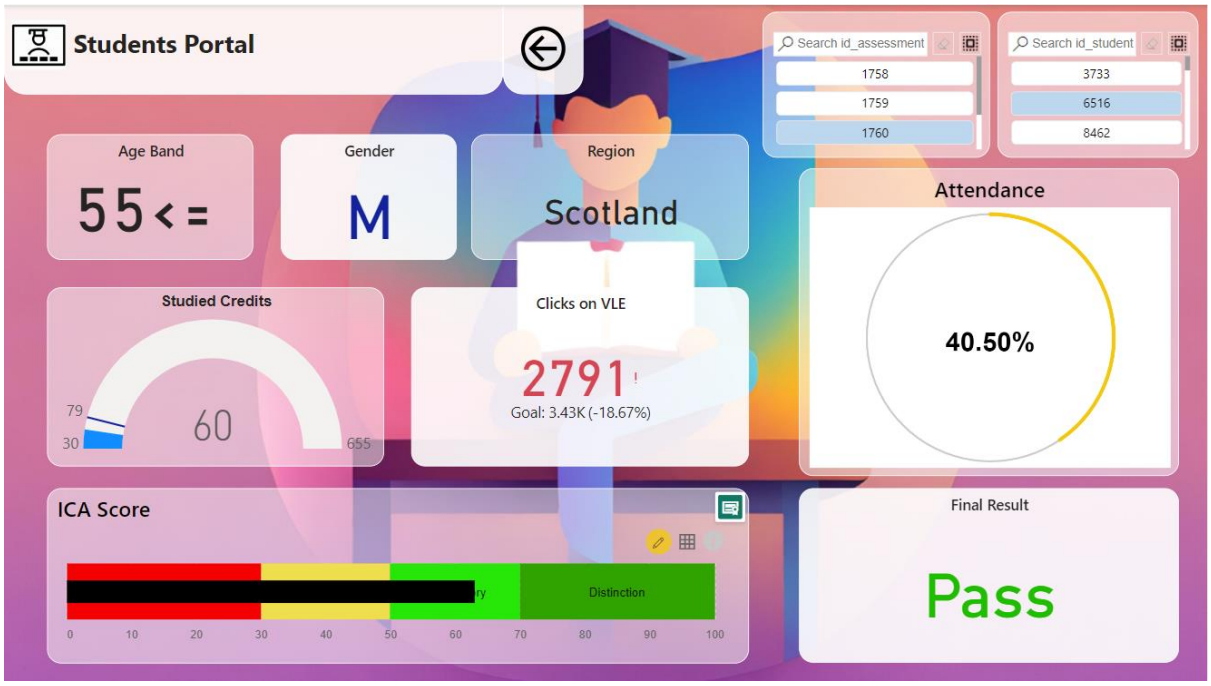


Fig 17: Students Portal Dashboard page

Fig 17 displays the dashboard page which holds the visualisations used to explain the above Business Intelligence question.

Fig 18 : Slicers for selecting student id and assessment id

The data displayed on the student portal is of an individual students and his performance in an individual ICA. For this 2 slicers have been used as shown in Fig 18. These select a particular student by his id and a specific element by its id. After selecting values in these slicers the data of an individual student is displayed on the Student Portal.

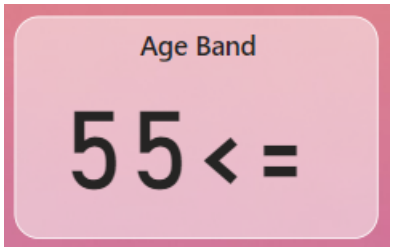


Fig 19 : Card visual displaying age band of the student

Fig 19 shows the card visual that is used to display the age band in which the selected student falls.

Fig 20 : Card visual displaying gender of the student

In Fig 20 the card usual that is used for showing the gender of the selected student is been displayed. Here M stands for Male and F stands for Female.

Fig 21 : Card visual displaying region of the student

The card visual that is used to know from which country does the selected student belong is shown in Fig 21.

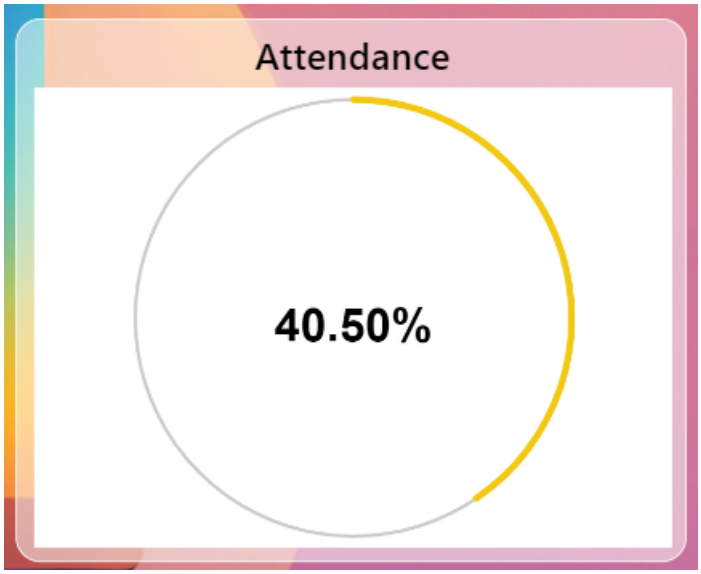


Fig 22 : KPI Donut chart displaying attendance of the student

Fig 22 shows a KPI Donut chart which is used to display the attendance of the selected student. Notice the color of the circle, It has red, yellow and green colors for highlighting low, moderate and high attendance respectively.

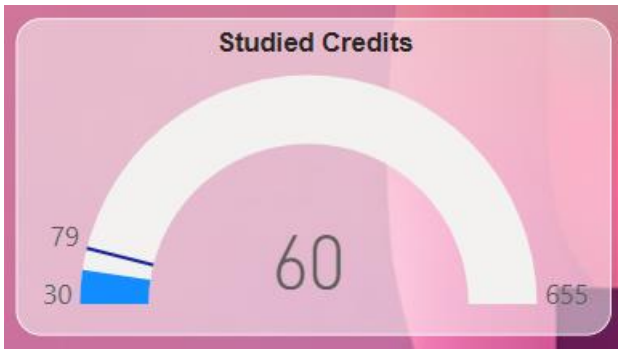


Fig 23 : Gauge displaying studied credits of the student

Fig 23 displays a Gauge showing studied credits of the selected student for an average credit target and a max credit.

For this 2 custom columns were created using the following DAX code :

- **max_studied_credits_col = MAX(studentInfo[studied_credits])**
  This column is the maximum studied credits column. The code calculates the maximum of the studied credits that the students have received.

- **avg_studied_credits_col = ROUND(AVERAGE(studentInfo[studied_credits]), 0)**
  This column is for the average studied credits of the students. Th code calulates the average of studied credits scored by the students and rounds it up to 0 decimal places.

- **min_studied_credits_col = MIN(studentInfo[studied_credits])**
  This column is the minimum studied credits column. The code calculates the minimum of the studied credits that the students have received.

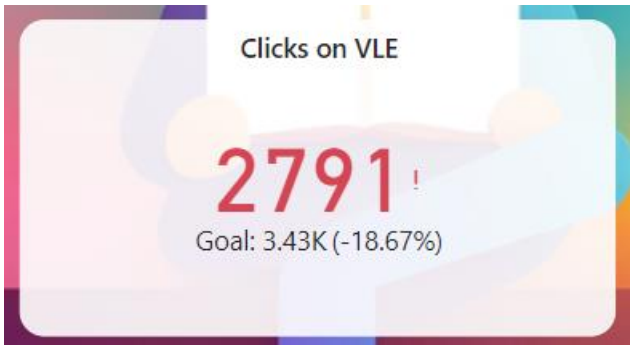


Fig 24 : KPI card for displaying number of clicks by student on VLe vs goal

A card visual that is pinned to the student portal is depicted in Fig. 24. Each student's click count is displayed, with a colour indicating whether they are above or below the goal. Green indicates that the target has been met, while red indicates that there are still too many clicks to reach the goal.

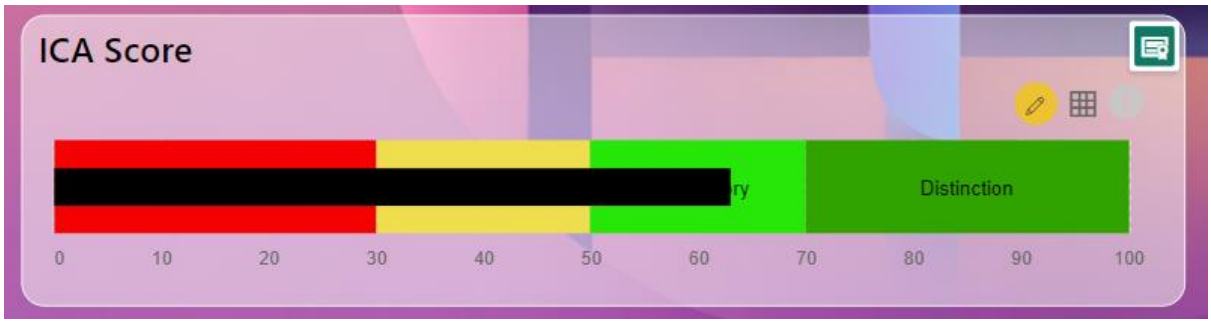


Fig 25 : Bullet chart displaying Score of student in ICA

A Bullet chart displaying the ICA score of the selected student is shown in Fig 25. The bullet chart is divided into cohorts of marks with a label for each cohort i.e. if the score of a student is between 0-30 marks , he/she is considered as fail. While 30-50 marks is considered as poor where improvement is needed. The 50-70 range is considered as satisfactory where performance is up to the mark. Distinction lies in the range of 70-100 which represents an excellent job by the student.

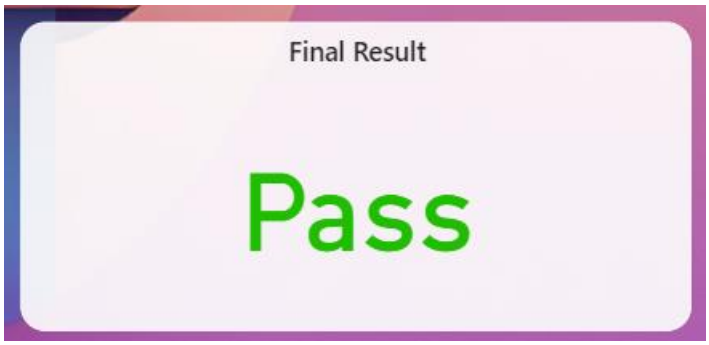


Fig 26 : Card visual displaying the final result of the student

Fig 26 shows a card visual displaying the final result of the selected student. It has 4 values. Pass, which tells that the student has passed. If he/she does and excellent job and clear the course then they might receive a Distinction. Fail, representing failure in clear the course. Withdrawn, if he/she isn’t able to complete the course due to any reasons.

Home Page and Navigation

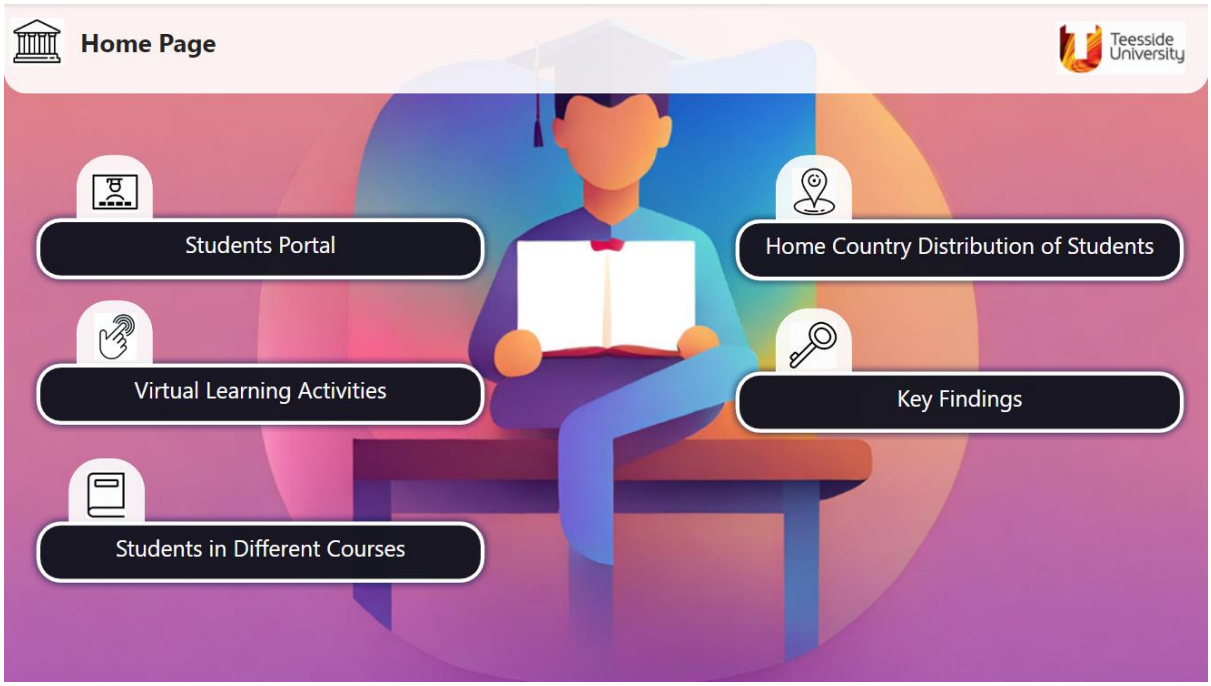


Fig 27 : Home page of the Dashboard

Home page can be referred as the cover page or the starting page of the Dashboard. The purpose of this page is to facilitate the ease of access to other dashboard pages. It technically stores the link to other pages in the form of buttons.

In Fig 27 five buttons can be seen namely ‘Students Portal’, ‘Virtual Learning Activities’, ‘Students in Different Courses’, ‘Home country Distribution of Students’ and ‘key findings’.

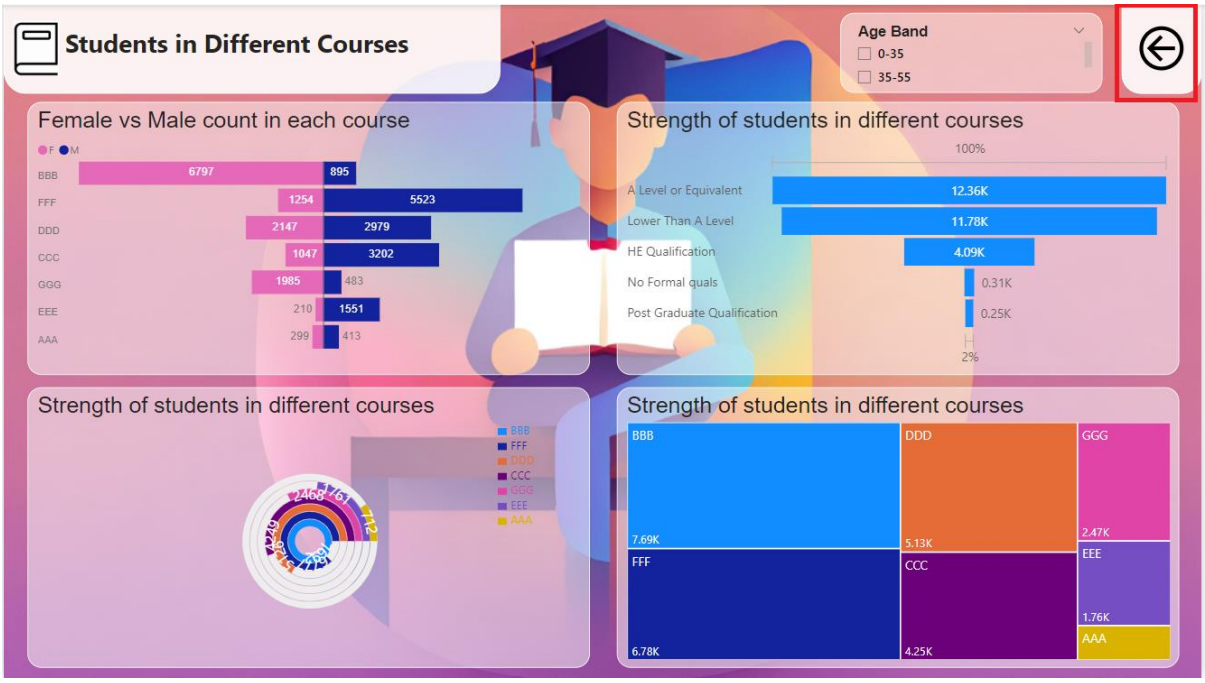


Fig 28 : Back button on dashboard pages

Fig 28 displays the back button provided on each dashboard page which can be clicked on to get back to the home page.

Key findings Dashboard Page

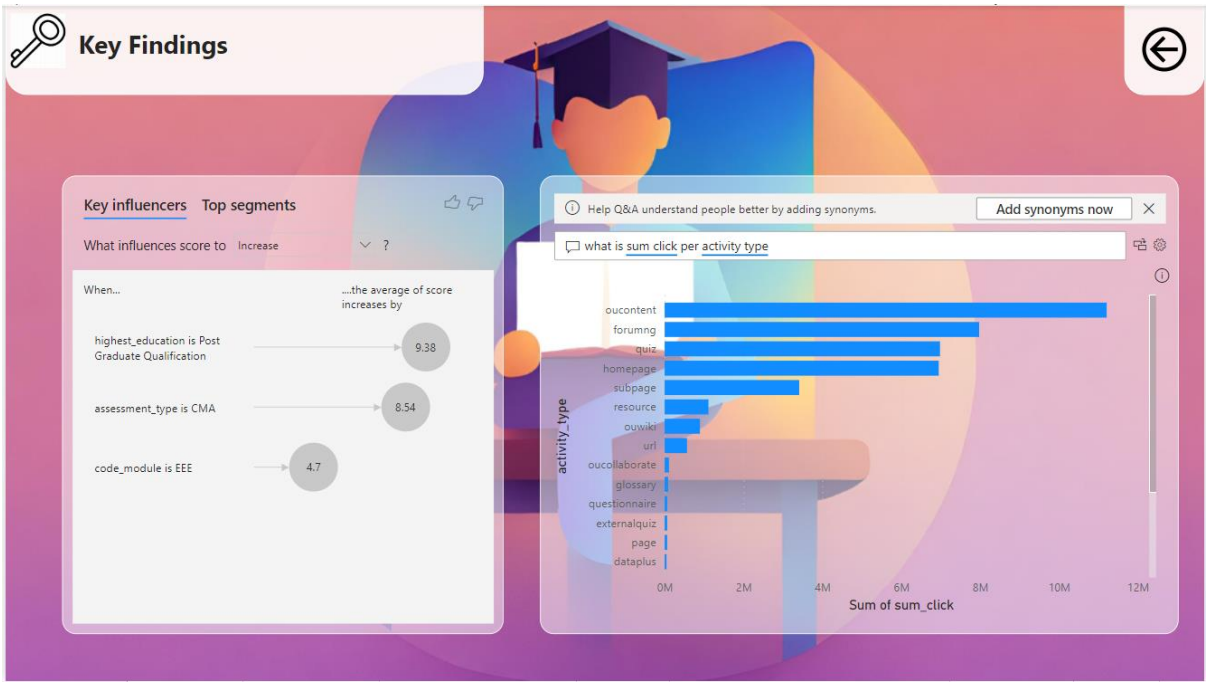


Fig 29 : Key Findings Dashboard Page

The key findings page shown in Fig 29 has a ‘key influencer chart’ and a ‘Q&A chart’. This page is basically to display some of the conclusions of the analysis.

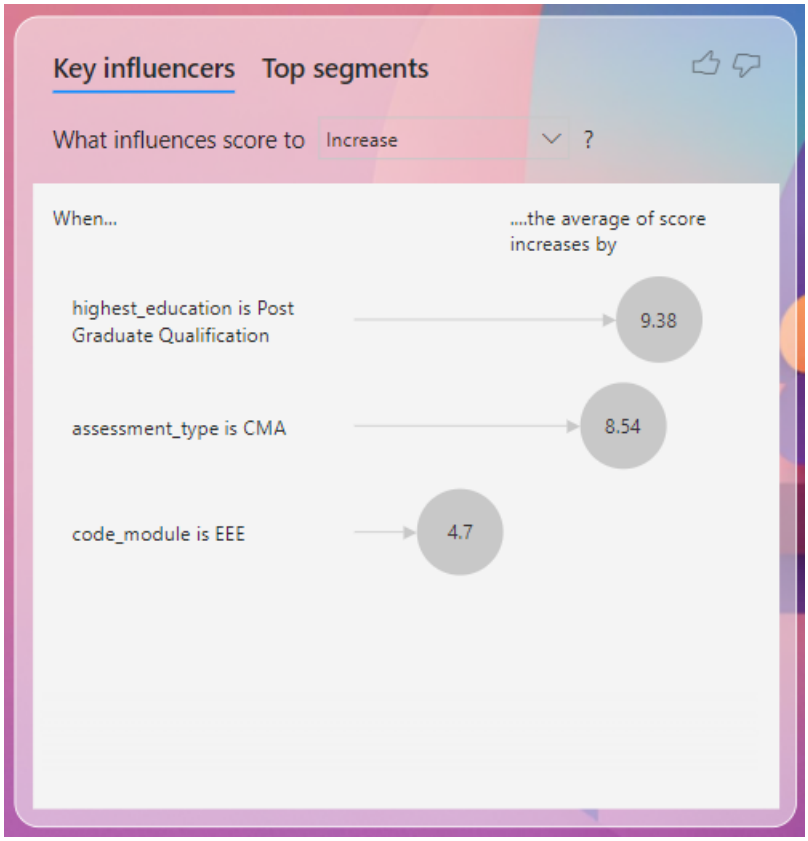


Fig 30 : Key influencers chart on the Key Findings Dashboard page

A Key Influencers chart is a type of visual in Power BI that assists you in determining and comprehending the factors that have the greatest influence on a specific result or metric. It is especially helpful when examining the relationships between the various variables in your dataset.

According to the key influencer chart produced in the Power BI dashboard, if the prior highest qualification was a postgraduate degree, the average score of the students increases by 9.38.

Also, it turned out that the average of score increases by 8.54 I the type of the assessment is CMA.

Furthermore, the students of the EEE course on an average score 4.7 more that the students belonging to the other courses.

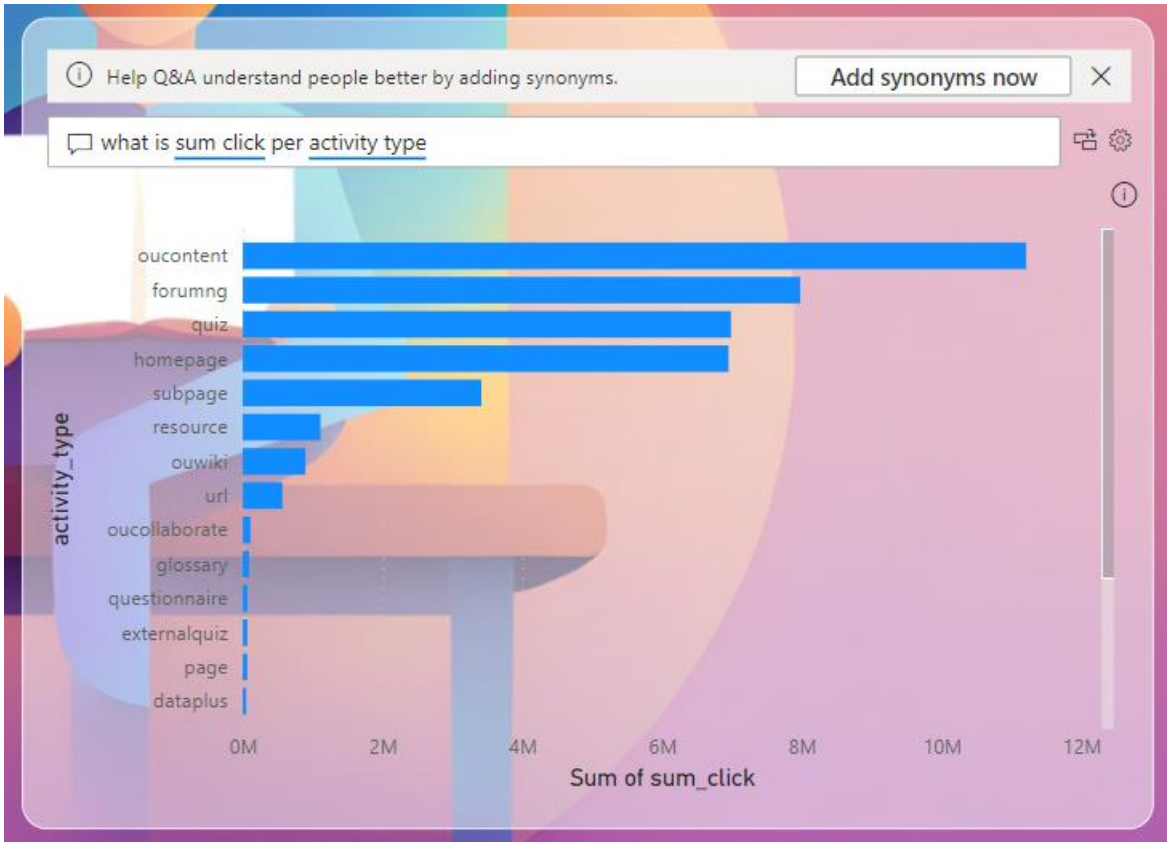


Fig 31 : Q&A chart

The Q&A (Question and Answer) feature in Power BI allows users to ask questions about their data in natural language and receive visualizations as answers. Its purpose is to simplify data exploration and analysis for users who might not know how to create conventional charts or queries.

As seen in Fig 31, there is an input bar at the top of the visual which receives the questions related to data from the user in native language. Then provides the answer to that particular question in the format of a data visualisation so that it is easily and effortlessly comprehended by the end user.

As the above screenshot shows that a question related to the amount of sum clicks for each VLe activity type was asked and the Q&A feature generated a bar chart representing the sum of clicks for each Activity type.

4. Summary

- The highest level of education, "Post Graduate," raises the average score by 9.38 points.
- On the Students Portal page, a figure that represents each student's "number of clicks" on the VLe resources is pinned. Each student's number can be compared to the target of "3.43K" clicks, which represents the average number of clicks made by students on the VLe.
- The most widely used VLE resource has been revealed as "oucontent," and students are found to be heavily utilising the VLe for the "FFF" course.
- It was discovered that students most frequently chose courses are "FFF" and "BBB," with "AAA" being the least preferred. While most males are enrolled in the FFF course, the majority of females are pursuing the BBB course.
- It was discovered that the majority of the university's students were domestic, meaning they were from the United Kingdom.
- On the dashboard, a new page called "Students Portal" has been made that offers comprehensive details about the student. His or her age band, gender, region, academic credits, number of clicks on VLe, attendance, ICA score and final result are among the details shown.

5. Recommendation

- Students should be able to see how many clicks each student has made on the VLe in order to assess their level of activity.
- Use the 'oucontent' resource as inspiration to make unpopular VLe resources engaging.
- Because BBB and FFF are the most popular courses and have the potential to remain popular in the future, their infrastructure should be ready to accommodate a large number of students.
- To accommodate more students, the syllabus for the AAA course needs to be carefully examined and adjusted.
- Bursaries and scholarships ought to be reviewed or implemented in order to promote the university's admission of international students.

6. Conclusion

The University Students Dataset's Business Intelligence analysis and Power BI dashboard creation offered a comprehensive learning opportunity covering data preparation, modelling, language competency, visualisation techniques, and critical thinking. Taking power BI into account specifically, I learned transformation of data in it, use of power queries, then creating fact and dimensions table while forming the data model. Also, learnt to manipulate data using power BI interface as well as by writing M Language queries. Furthermore, when a requirement of custom measures arose, it gave me a chance to try my hands on the DAX code. It helped me improve my skills in Business Intelligence as well as data analytics. These transferable skills highlight the value of data-driven insights in forming well-informed strategies and decisions in the education sector.

7. Reference

analyse.kmi.open.ac.uk. *Open University Learning Analytics dataset*. [online] Available at: <https://analyse.kmi.open.ac.uk/open_dataset#:~:text=This%20page%20introduces%20the%20anonymised,selected%20courses%20(called%20modules).>
[Accessed 8 October 2023].

Self-assessment

Report Section	Description	Grade your work from 0 to 100
Report Structure	The report is well-written, and it contains all the relevant sections	95
Data Pre-processing and Data Modelling	Many pre-processing steps have been applied. The data model is well-structured	95
Dax and M language	Both DAX and M Language have been extensively used in the report	90
Dashboard Design	The dashboard contains a variety of charts, including advanced ones not covered in the module.	100
Average		Add below the average of the four cells above: 95