**APPENDIX B**



**DATA STORYTELLING (CAA1S06)**

Final Report (Part 2)

Submitted by

Chan Jun Qi, Julius (2082753A)

Date of Submission: 29/08/2021

# DATA STORYTELLING (CAA1S06)

**SGUS-AAI Apr Semester**

**Assignment**

**Practical Class** : PCO

**Submitted by** : <2082753A> <Chan Jun Qi, Julius>

**Date:** dd /08/2021

**“By submitting this work, I am / we are declaring that I am / we are the originator(s) of this work and that all other original sources used in this work has been appropriately acknowledged.**

**I / We understand that plagiarism is the act of taking and using the whole or any part of another person’s work and presenting it as my/ our own without proper acknowledgement.**

**I / We also understand that plagiarism is an academic offence and that disciplinary action will be taken for plagiarism.”**

**Name and Signature of student : chan jun qi ,julius**



\*Where PXX is the practical class number

**Revised Data Preparation for Assessments.csv**

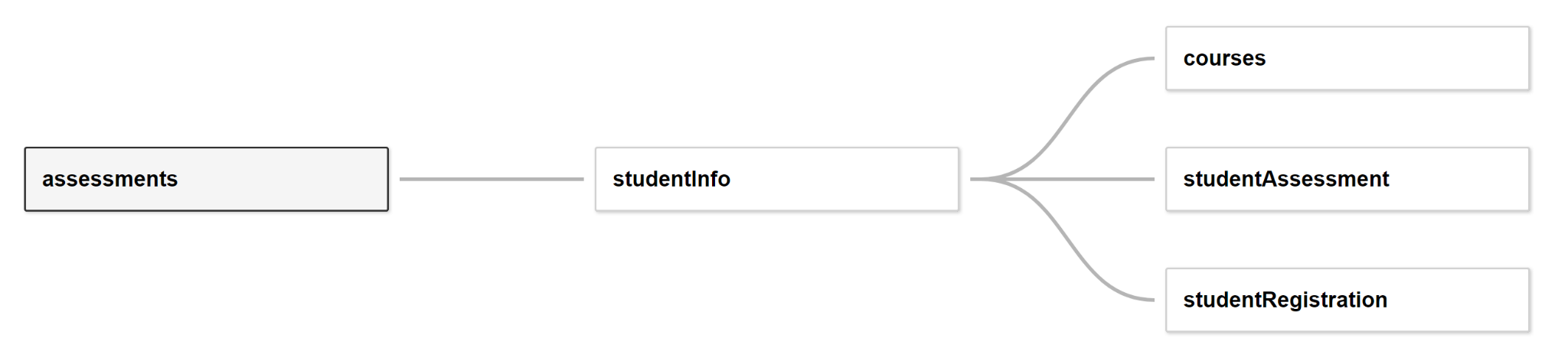
For Assessments.csv, “if the information about the final exam date is missing, it is at the end of the last presentation week”. As date was an important field for the analysis, null values were replaced with the corresponding values of the course duration referenced from module\_presentation\_length field in Courses.csv.

Assessments.csv

|  |  |  |  |
| --- | --- | --- | --- |
| Row # | code\_module | code\_presentation | date |
| 7 | AAA | 2013J |  |
| 13 | AAA | 2014J |  |
| 25 | BBB | 2013B |  |
| 37 | BBB | 2013J |  |
| 49 | BBB | 2014B |  |
| 55 | BBB | 2014J |  |
| 64 | CCC | 2014B |  |
| 65 | CCC | 2014B |  |
| 74 | CCC | 2014J |  |
| 75 | CCC | 2014J |  |
| 110 | DDD | 2014J |  |

|  |  |  |  |
| --- | --- | --- | --- |
| Row # | code\_module | code\_presentation | module\_presentation\_length |
| 2 | AAA | 2013J | 268 |
| 3 | AAA | 2014J | 269 |
| 4 | BBB | 2013J | 268 |
| 5 | BBB | 2014J | 262 |
| 6 | BBB | 2013B | 240 |
| 7 | BBB | 2014B | 234 |
| 8 | CCC | 2014J | 269 |
| 9 | CCC | 2014B | 241 |
| 11 | DDD | 2014J | 262 |

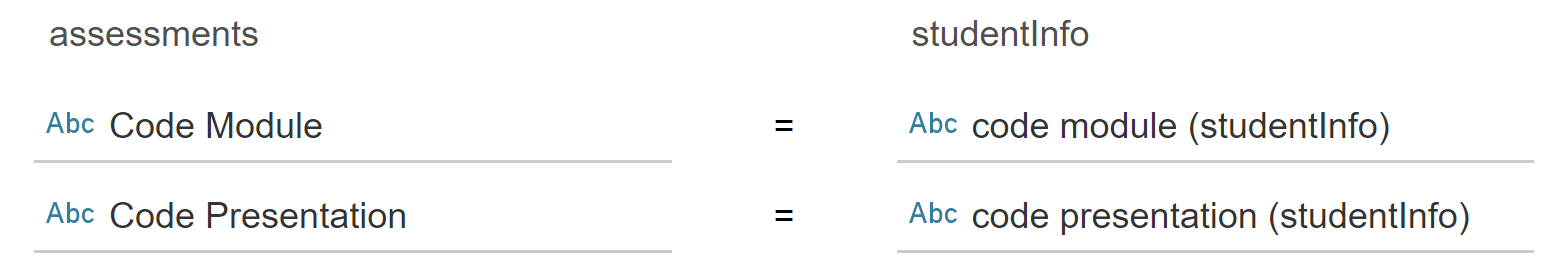
**Data Integration**



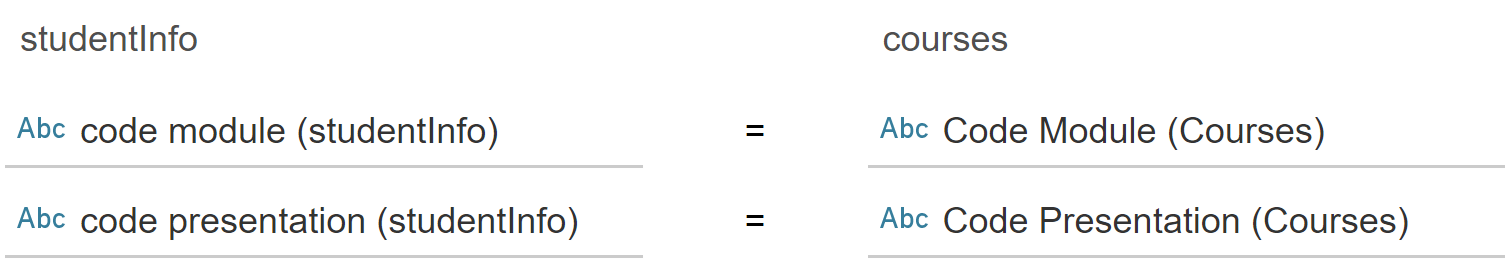
The 5 data sets are combined using Tableau’s data modelling feature. Every data set forms a relationship to studentInfo due to it having primary keys that can reference to all other datasets and is also the base of our analysis.

The relationships are shown below:

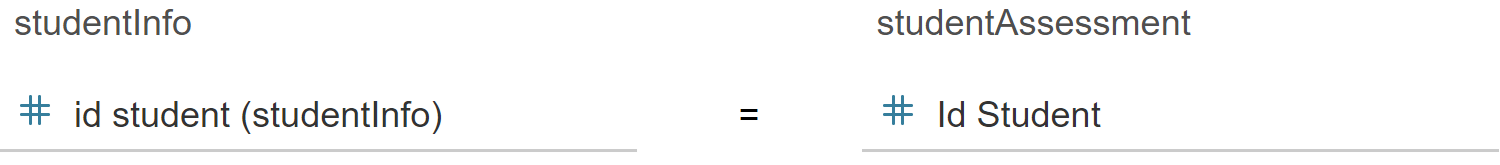
**Assessments.csv and studentInfo.csv**



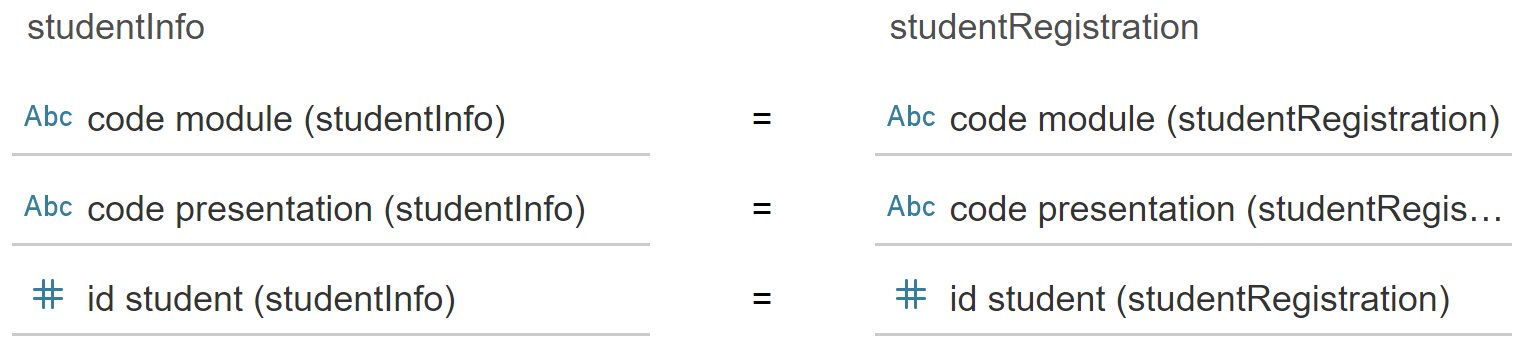
**courses.csv and studentInfo.csv**



**studentAssessment.csv and studentInfo.csv**



**studentRegistration.csv and studentInfo.csv**

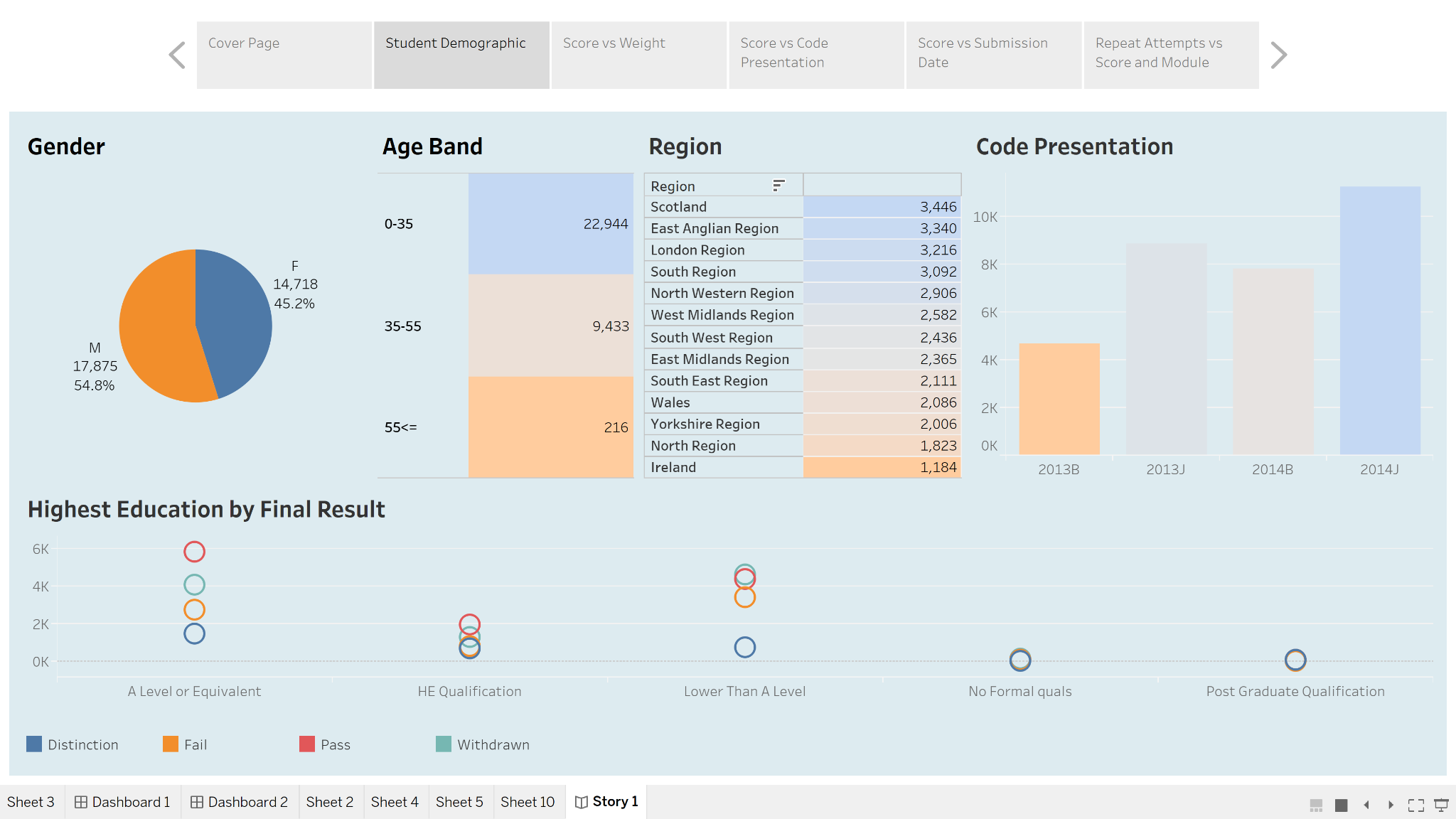


**Storyboard Cover Page**



The first story point serves as a cover page. This is a requirement for the storyboard design as you are presenting to the Open University during the introduction and explanation. The cover page makes it engaging to everyone instead of jumping into charts and figures right away in subsequent story points.

**Dashboard**



This dashboard seeks to analyze the student demographic across all course\_presentation. It covers both reporting of general statistics such as gender, age band and region distribution and cohort size by code presentation. It also seeks to identify relationships through a visualization of highest\_education by final\_result.

The dashboard is also interactive where the user can make selections by clicking any or multiple fields he/she wishes to drill down into, for example males from Scotland with an A level or equivalent.

Graphical user interface, application

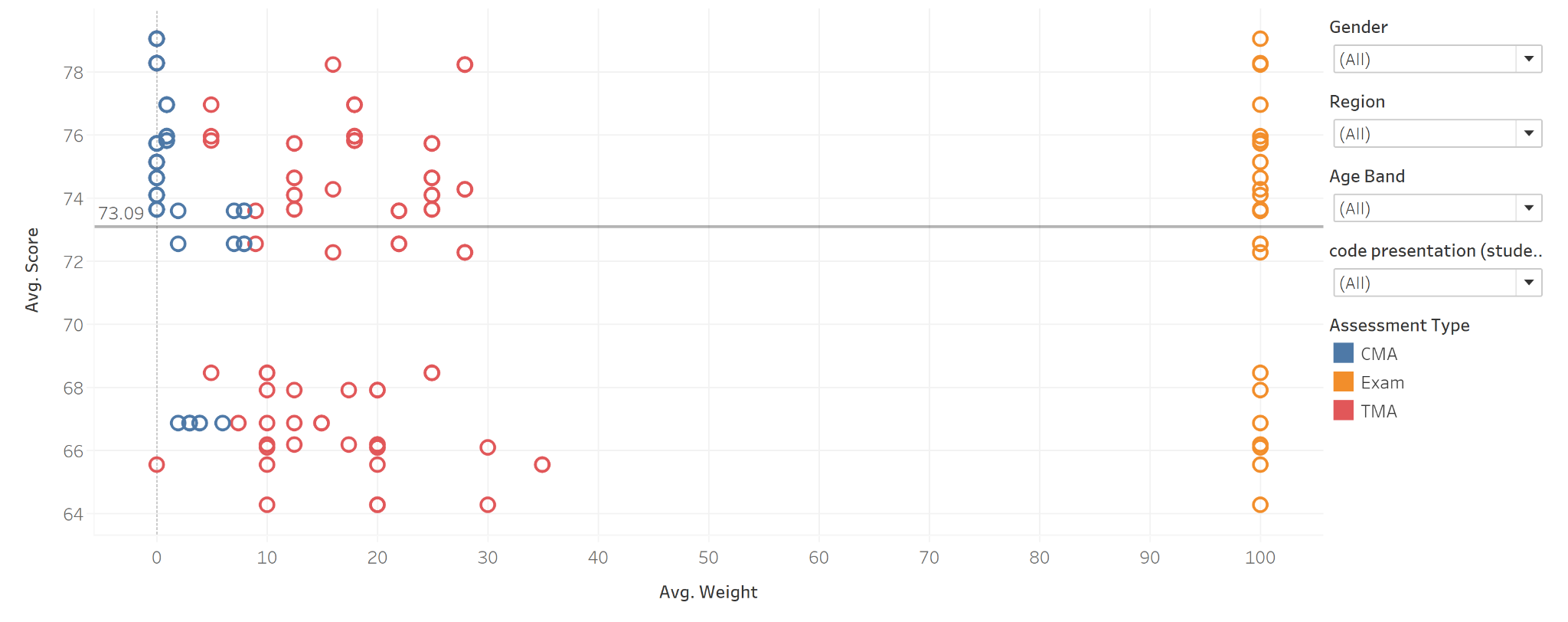
Description automatically generated

According to the dashboard.

* The majority of the pupils are male.
* The majority of the pupils are between the ages of 0 and 35.
* The majority of the students in the region come from Scotland.
* The majority of students in higher education are A level or Equivalent.

This is significant since these students are the ones that give the majority of the data insights.

**Average Score by Assignment Type and Weight**



This scatterplot seeks to answer the question whether the Weight of an assessment affects the average score of the students. It is created by using Average Weight as Columns and Average Score as rows, with Id Assessment as the details.

The colour legend is by assessment type. A reference line showing average score is also created.

It can be seen from the reference line that the average score is 73.09.

Also, it is observed that CMA tend to score higher on average even though they are of lower or have no weightage. TMA scores are more scattered and have higher fluctuations regardless of weightage.

An interesting observation can be spotted for ID 15020 where the average score is 66.94 even though it has no weightage, compared to higher scoring TMA with no weightage. For exams, there is a variation of scores across the courses as well. However, there is a higher concentration of courses achieving an average score of above 70 compared to those with lower.

There is also a filter panel with factors like Gender, Region, Age Band and also the Code Presentation for users to drill down and make a deeper level of analysis. For example, if a comparison was done for genders, the filter for gender will be toggled respectively to give the following results.

**Males Average Score**

Chart, scatter chart

Description automatically generated

Males have a variation of average scores across all assessment types regardless of weightage from 63.90 to 81.28. Majority of the results fall above the 70 mark range and have an average score of 73.20.

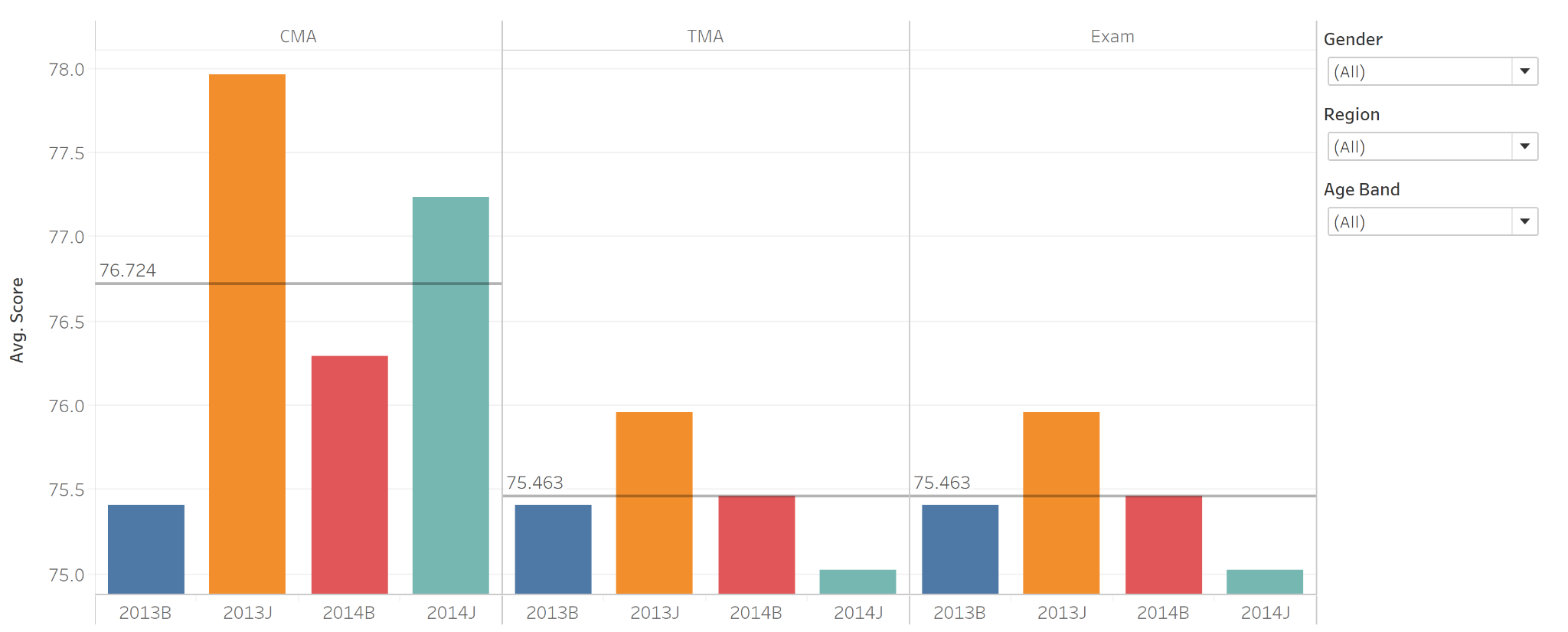
**Females Average Score**

Chart, scatter chart

Description automatically generated

Females have a variation of average scores from 61.89 to 90.50. Majority of scores are above 74, however there is also a large concentration of scores below 70, resulting in an average score of 74.14.

**Average Score by Assignment Type and Code Presentation**



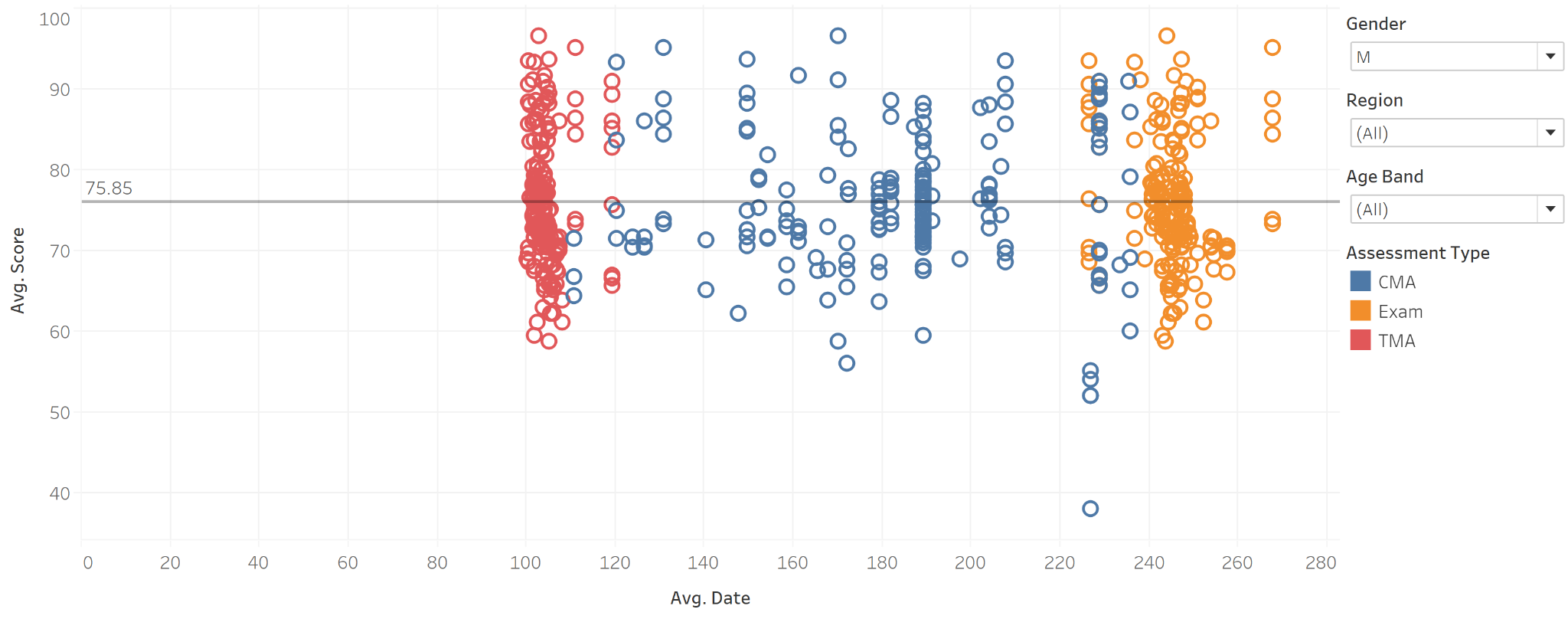
This visualization seeks to answer the question of whether the average score of assignments is affected by the Code Presentation. It is created using bar charts with Assessment Type and Code Presentation as columns and average scores as rows. A reference line showing average score by Assessment Type is also created. The colour legend indicates different Code Presentation.

**Recommendations**

It can be seen that across all assessment types, 2013J which indicates presentations starting in October having the highest average scores. Regarding the assessments, the study can look into factors on Code 2013J such as method of teaching, teachers and other external factors resulting in this disparity in grades.

Another interesting finding is that Code 2014J's average score has plummeted to the lowest level average score during TMA & Exam. Reviewing student needs, such as allowing students to provide feedback via a survey form or course materials, is one step that may be performed.

**Score vs Days from Start Date**



This scatterplot seeks to answer the question of whether the submission date affects the score. It is created using the average score of assessments against the average date of submission by assessment id. A reference line for the average score is also created.

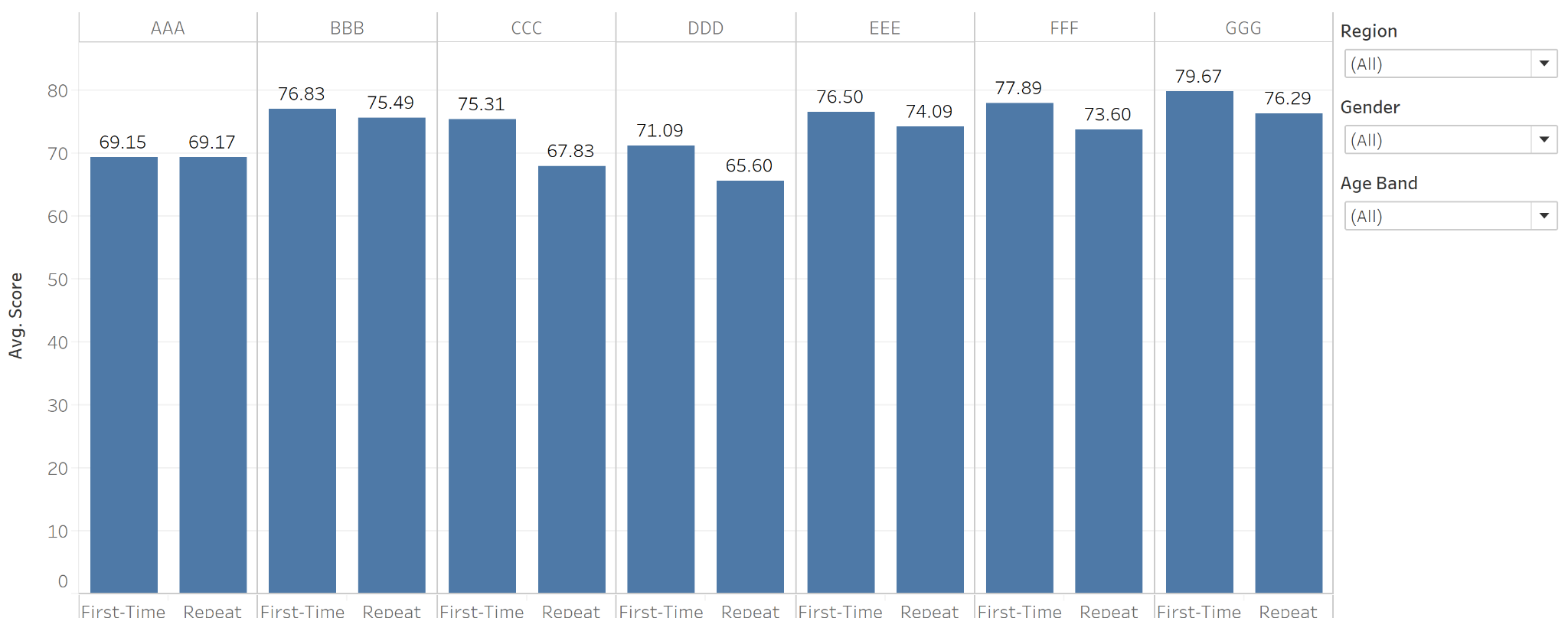
It can be seen that regardless of date, the distribution of marks is fairly consistent. However, it can be seen that CMA near examination dates tend to score much lower compared to other times. This can be already seen on the bar chart of the Average score by Assessment type and Code presentation.

This brings about the question on whether students tend to focus on higher weightage assessment, and whether CMA during this period is necessary as it may lead to additional workload and stress for students.

**Recommendations**

Increased time between CMA and Exams on a later date would assist to balance the workload, maybe resulting in better scores for students.

**Scores comparison between First-Timers and Repeated Attempts**



This bar chart seeks to find out whether students with repeated attempts tend to score better than first timers. It is created by creating a set for students with above 1 attempt and editing the alias as “Repeat” and analyzed by Code Module.

Despite undergoing the module for repeated times, one would expect a student to do better in subsequent attempts.

However, it can be seen that students with repeated attempts do not score better than students with only one attempt.

**Recommendations**

More emphasis such as reviewing their teaching method and providing remedial lessons to these students. Using the filter can allow the user to drill down into specific factors affecting the trends so more focused care can be given.