**Bigquery Activity**

**Goal**

Practice using Bigquery, along with data upload, data warehouse architecture, and querying.

**Description**

You are given three csv files:

student\_exam\_results.csv

* Shows math, reading, and writing scores for each student

student\_exam\_survey.csv

* Shows results of a short survey each student was asked to complete. Includes parental level of education, lunch plan, and test preparation course. The values in these three columns include numerically encoded values that equate to these:

|  |  |  |
| --- | --- | --- |
| **Parental Level of Education** | **Lunch Plan** | **Test Preparation Course** |
| 1 = Some High School | 1 = Free/Reduced | 0 = None |
| 2 = High School | 2 = Standard | 1 = Completed |
| 3 = Some College | - |  |
| 4 = Associate’s Degree |  |  |
| 5 = Bachelor’s Degree |  |  |
| 1. = Master’s Degree |  |  |

student\_list.csv

* Shows student information including gender and year group

Your first course of action is to prepare the data. You should design a data warehouse with this data using a star schema. (It might be easiest for you to first generate a single large data with all the data denormalized). You may need to create your own data to include the decoded values for student\_exam\_survey.csv.

Once you have your data warehouse completed, find this information:

* Average scores across subjects of students grouped by year group.
* Average scores across subjects of students grouped by gender
* How does parental level of education affect student test scores?
* How well did the test preparation course help students?

**Presentation**

You will give a 1-2 minute presentation of your findings on Friday, June 2. Be prepared to show an ERD of your data warehouse, as well as the results of your queries.