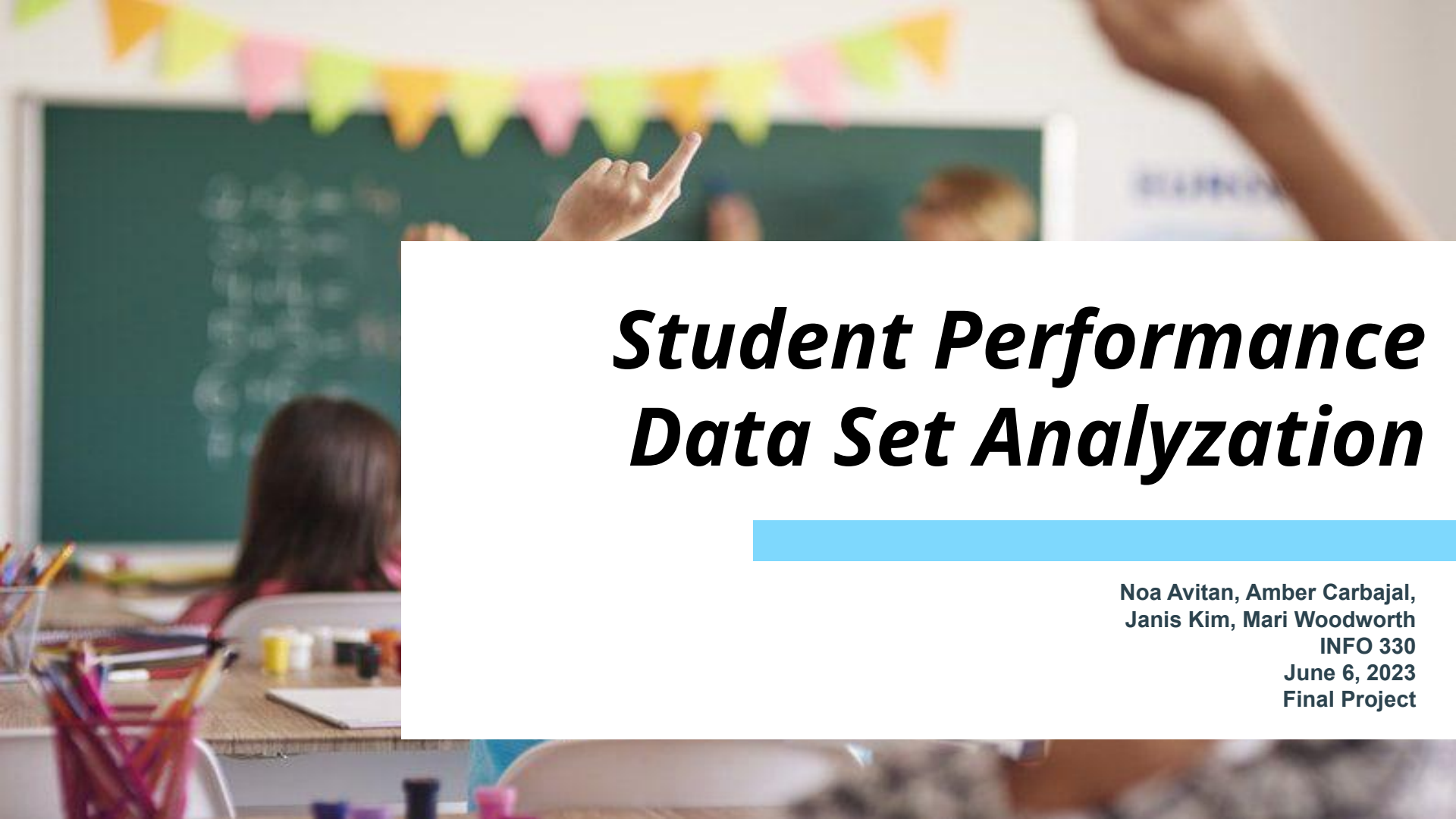


Student Performance Data Set Analyzation

Noa Avitan, Amber Carbajal,
Janis Kim, Mari Woodworth
INFO 330
June 6, 2023
Final Project

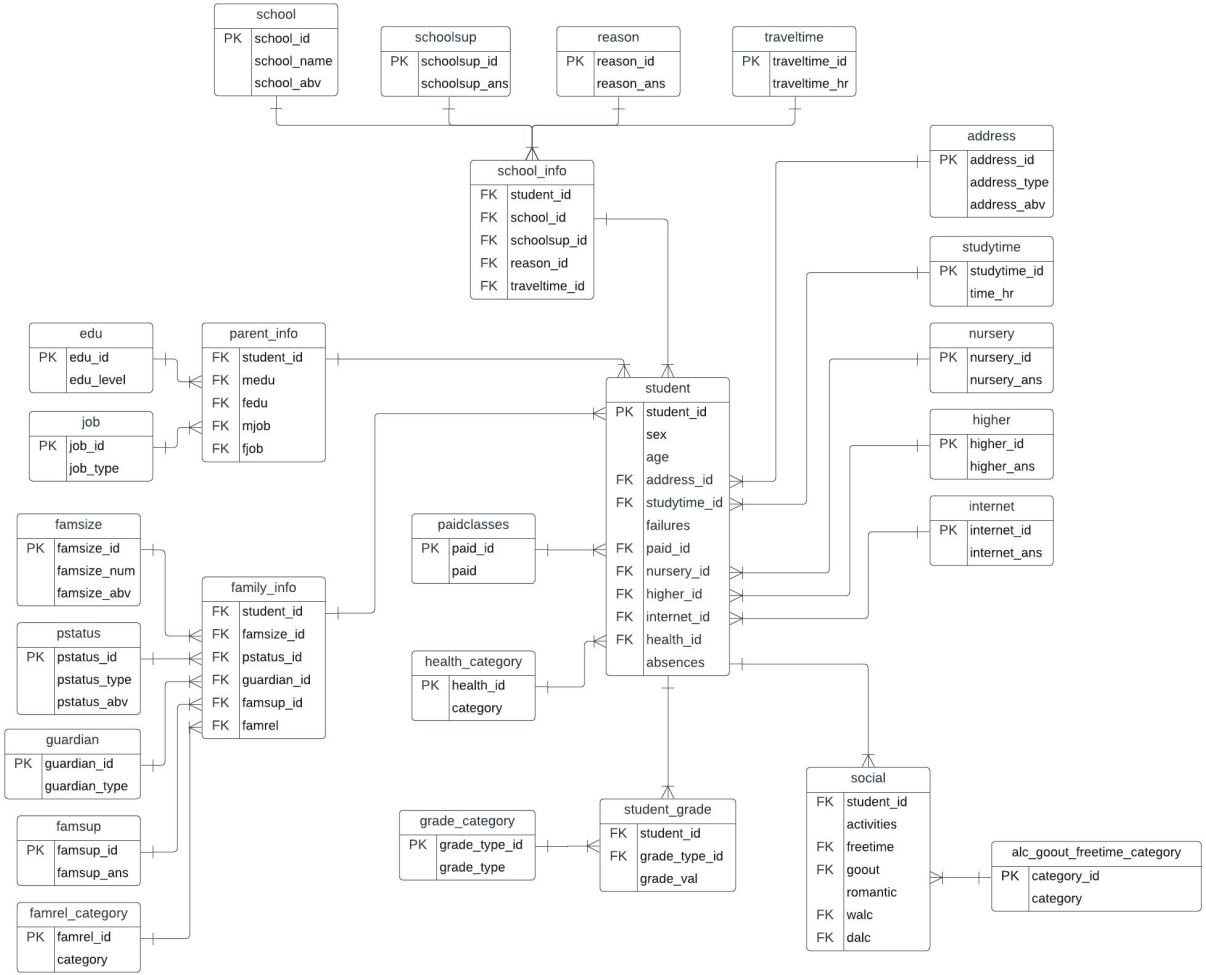
Student Performance Dataset

- This data approach student achievement in secondary education of 2 portuguese schools.
- The data attributes include student grades, demographics, social and school related features!
- The dataset can be found here: https://www.kaggle.com/datasets/larsen0966/student-performance-data-set
- The actual research paper published by Cortez and Silva can be found here: http://www3.dsi.uminho.pt/pcortez/student.pdf

Example

school student's school (binary: 'GP' - Gabriel Pereira or 'MS' - Mousinho da Silveira)	sex student's sex (binary: 'F' - female or 'M' - male)	age student's age (numeric: from 15 to 22)	address student's home address type (binary: 'U' - urban or 'R' - rural)	famsize family size (binary: 'LE3' - less or equal to 3 or 'GT3' - greater than 3)	Pstatus parent's cohabitation status (binary: 'T' - living together or 'A' - apart)	Medu mother's education (numeric: 0 - none, 1 - primary education (4th grade), 2 â€“ 5th to 9th grade, 3 â€“ secondary
GP 65% MS 35%	F 59% M 41%	15 22	U 70% R 30%	GT3 70% LE3 30%	T 88% A 12%	0 3.60 - 4.00 Count: 175
GP	F	18	U	GT3	A	4
GP	F	17	U	GT3	T	1
GP	F	15	U	LE3	T	1
GP	F	15	U	GT3	T	4
GP	F	16	U	GT3	T	3
GP	M	16	U	LE3	T	4
GP	M	16	U	LE3	T	2
GP	F	17	U	GT3	A	4
GP	M	15	U	LE3	A	3
GP	M	15	U	GT3	T	3
GP	F	15	U	GT3	T	4
GP	F	15	U	GT3	T	2

Data Model


school
PK school_id
school_name
school_abv
schoolsup
PK schoolsup_id
schoolsup_ans
reason
PK reason_id
reason_ans
traveltime
PK traveltime_id
traveltime_hr
school_info
FK student_id
FK school_id
FK schoolsup_id
FK reason_id
FK traveltime_id
address
PK address_id
address_type
address_abv
studytime
PK studytime_id
time_hr
nursery
PK nursery_id
nursery_ans
higher
PK higher_id
higher_ans
internet
PK internet_id
internet_ans
edu
PK edu_id
edu_level
job
PK job_id
job_type
parent_info
FK student_id
FK medu
FK fedu
FK mjob
FK fjob
student
PK student_id
sex
age
FK address_id
FK studytime_id
failures
FK paid_id
FK nursery_id
FK higher_id
FK internet_id
FK health_id
absences
famsize
PK famsize_id
famsize_num
famsize_abv
pstatus
PK pstatus_id
pstatus_type
pstatus_abv
guardian
PK guardian_id
guardian_type
famsup
PK famsup_id
famsup_ans
famrel_category
PK famrel_id
category
family_info
FK student_id
FK famsize_id
FK pstatus_id
FK guardian_id
FK famsup_id
FK famrel
paidclasses
PK paid_id
paid
health_category
PK health_id
category
grade_category
PK grade_type_id
grade_type
student_grade
FK student_id
FK grade_type_id
grade_val
social
FK student_id
activities
FK freetime
FK goout
romantic
FK walc
FK dalc
alc_goout_freetime_category
PK category_id
category

Interesting Schema Design

- Some tables such as our alc_goout_freetime_category hold answers to different columns

  Ex: In our social table, our “goout, walc, and dalc” values based on the students response that is stored in the alc_goout_freetime_category table

Table: social

	student_id	activities	freetime	goout	romantic	walc	dalc
	Filter	Filter	Filter	Filter	Filter	Filter	Filter
1	1	no	3	4	no	1	1
2	2	no	3	3	no	1	1
3	3	no	3	2	no	3	2
4	4	yes	2	2	yes	1	1
5	5	no	3	2	no	2	1
6	6	yes	4	2	no	2	1
7	7	no	4	4	no	1	1
8	8	no	1	4	no	1	1
9	9	no	2	2	no	1	1
10	10	yes	5	1	no	1	1

Table: alc_goout_freetime_categor

	category_id	category
	Filter	Filter
1	1	Very low
2	2	Low
3	3	Medium
4	4	High
5	5	Very high

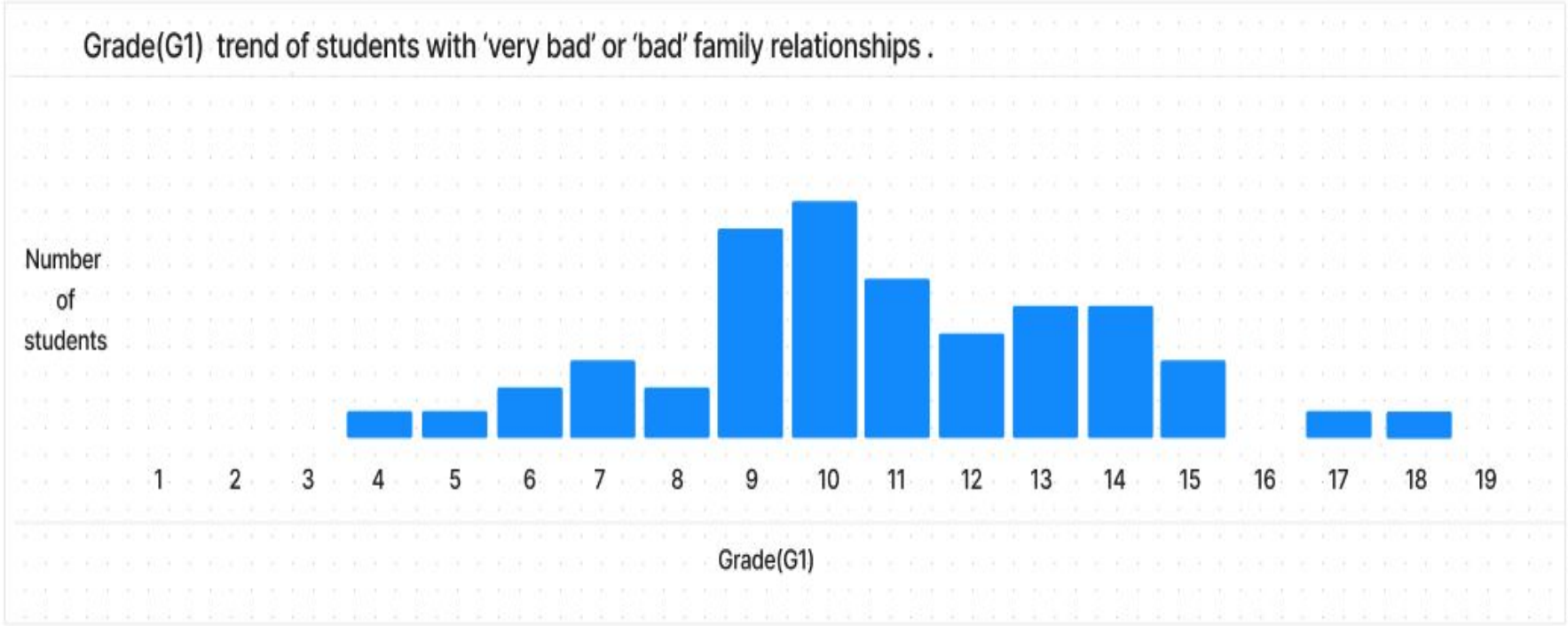
Queries and Visualizations

Query 1

First we want to study the family relationships of the students. Find all students with ‘very bad’ or ‘bad’ family relationships and show their G1 grades. We want to study if the quality of family relationships has a relationship with student performance.

```
SELECT S.student_id, GC.grade_type, SG.grade_val, FI.famrel
FROM student S
JOIN student_grade SG ON S.student_id = SG.student_id
JOIN grade_category GC ON SG.grade_type_id = GC.grade_type_id
JOIN family_info FI ON S.student_id = FI.student_id
JOIN famrel_category FC ON FI.famrel = FC.famrel_id
WHERE (famrel_id=1 OR famrel_id=2) AND GC.grade_type='G1';
```

Visualizations for Query 1



Query 2

Find all students who are living apart from their parents and find if they have failed any classes. Following the previous query, we want to study if living apart from parents affects class failure. Show student information and their grades. (Does having parents physically at home improve students' performance?)

```
SELECT S.student_id, S.failures, GC.grade_type, SG.grade_val, PS.pstatus_abv
FROM student S
JOIN student_grade SG ON  S.student_id = SG.student_id
JOIN grade_category GC ON SG.grade_type_id = GC.grade_type_id
JOIN family_info FI ON S.student_id = FI.student_id
JOIN pstatus PS ON FI.pstatus_id = PS.pstatus_id
WHERE PS.pstatus_type ='Living apart' AND GC.grade_type='G1';
```

Query 3

Find all the female students who have a study time greater than 5 hours and their G3 grade (show their id, sex, studytime, grade value, grade category).

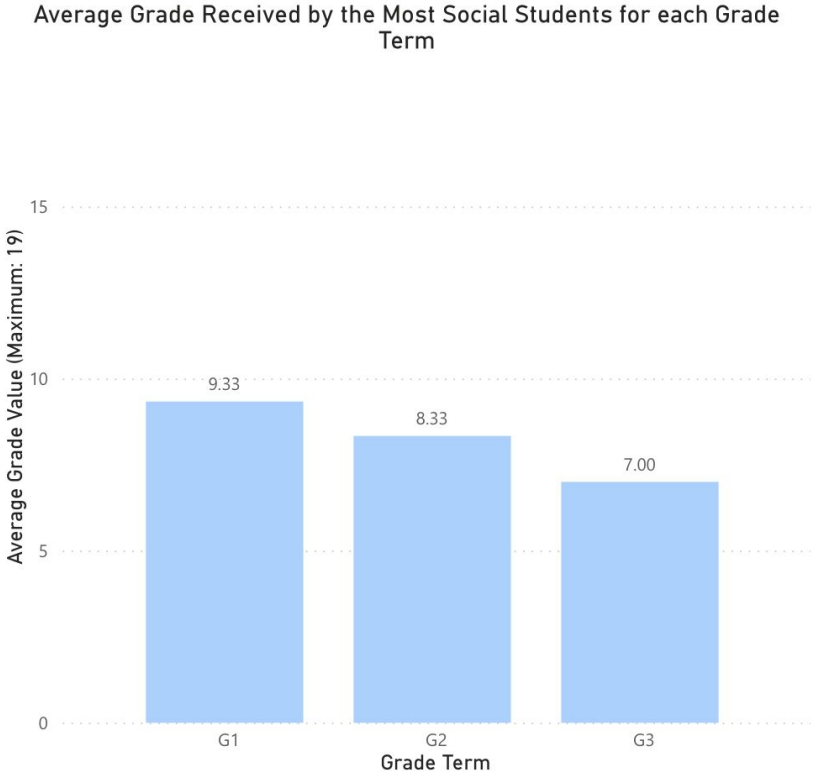
```
CREATE TABLE f_g3_temp AS
SELECT student.student_id, student.sex, studytime.time_hr, grade_category.grade_type, student_grade.grade_val
FROM student
JOIN studytime
ON student.studytime_id = studytime.studytime_id
JOIN student_grade
ON student.student_id = student_grade.student_id
JOIN grade_category
ON student_grade.grade_type_id = grade_category.grade_type_id
WHERE student.sex = 'F'
AND (studytime.time_hr = '5 - 10 hours'
OR studytime.time_hr = '>10 hours')
AND grade_category.grade_type = 'G3';
```

Query 4

Find the students who allocate their time away from academics and determine what their final grade is and show their id, activities, freetime, goout, and all of their grades. (In other words, find students that have a very high going out rate (5), very high free time after school (5), and no extra-curricular activities).

```
CREATE TABLE student_grade_activites_temp AS
SELECT student.student_id, social.activities, social.freetime, social.goout, grade_category.grade_type, student_grade.grade_val
FROM student
JOIN student_grade
ON student.student_id = student_grade.student_id
JOIN grade_category
ON student_grade.grade_type_id = grade_category.grade_type_id
JOIN social
ON student.student_id = social.student_id
WHERE social.activities = 'no'
AND social.goout = '5'
AND social.freetime = '5';
```

Visualizations for Query 4



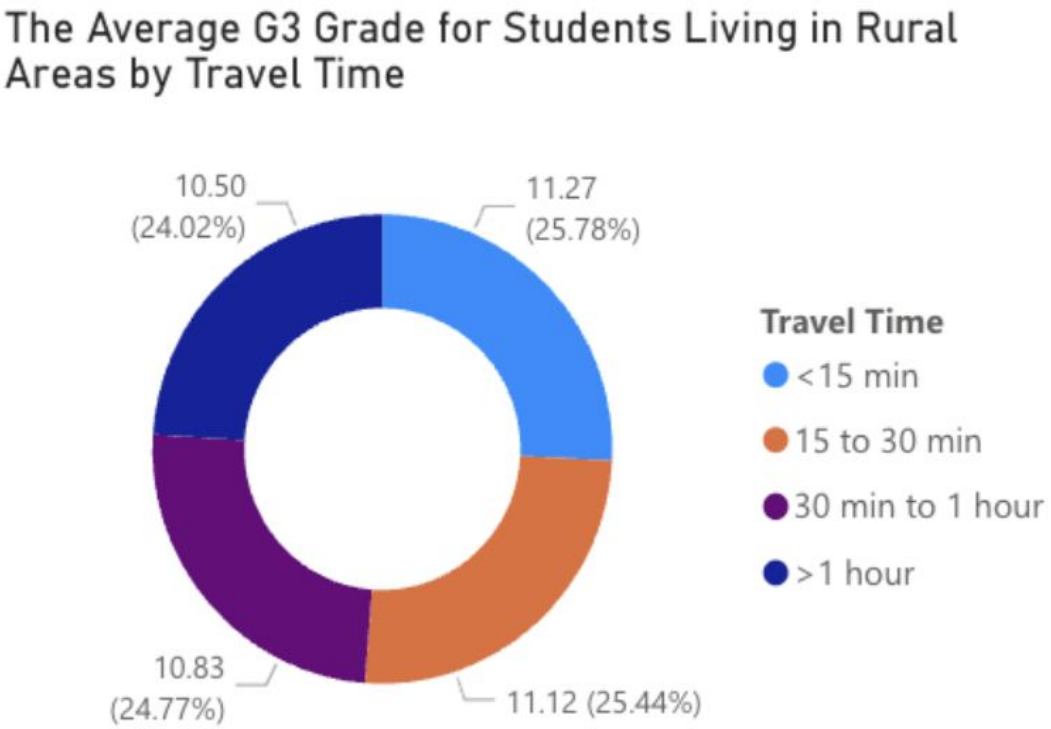
Query 5

Select all students who live in a rural area and find their travel time to school, internet access, mother and father's jobs, and their grades over the semesters and final grade (address, traveltime, internet, mjob, fjob, and G3).

```
CREATE TABLE temp_job AS
SELECT parent_info.*, job.*
FROM student
JOIN parent_info ON parent_info.student_id = student.student_id
JOIN job ON parent_info.mjob = job.job_id

SELECT DISTINCT student.student_id, traveltime.traveltime_hr, internet.internet_ans,
job.job_type AS fjob_type, temp_job.job_type AS mjob_type, student_grade.grade_val AS g3_grade
FROM student
JOIN school_info ON student.student_id = school_info.student_id
JOIN traveltime ON school_info.traveltime_id = traveltime.traveltime_id
JOIN internet ON student.internet_id = internet.internet_id
JOIN parent_info ON student.student_id = parent_info.student_id
JOIN job ON parent_info.fjob = job.job_id
JOIN temp_job ON temp_job.mjob = parent_info.mjob
JOIN student_grade ON student.student_id = student_grade.student_id
WHERE student.address_id ='2'
AND student_grade.grade_type_id = '3';
```

Visualizations for Query 5



Query 6

For students under Portugal's legal drinking age (18) that have a Walc [weekend alcohol consumption] or Dalc [workday alcohol consumption] > 3, determine if they want to pursue a higher education and show their id, higher [if they want to pursue higher education], and G3 [final grade].

```
SELECT student.student_id, higher.higher_ans, student_grade.grade_val AS g3_grade
FROM student
JOIN social ON student.student_id = social.student_id
JOIN higher ON student.higher_id = higher.higher_id
JOIN student_grade on student_grade.student_id = student.student_id
WHERE age < '18' AND walc > '3' AND dalc > '3' AND grade_type_id = '3';
```

Query 7

Select the id, name of school, sex, and age for the top student based on their final grade [final grade] [sex] [age] [name of school]. This query will help us determine what the demographic of the top performing student is.

```
CREATE TABLE top_students AS
SELECT student.student_id, school.school_name, student.sex, student.age, student_grade.grade_val
FROM school
JOIN school_info ON school.school_id = school_info.school_id
JOIN student ON school_info.student_id = student.student_id
JOIN student_grade ON student.student_id = student_grade.student_id
WHERE grade_type_id = '3'
AND grade_val = '19';
```

Query 8

For the previous student, what is their weekly study time and have they had any past class failures? This will help us relate the correlation of study time to academic performance.

```
CREATE TABLE top_traits AS
SELECT student.student_id, student.failures, studytime.time_hr
FROM student
   JOIN studytime ON student.studytime_id = studytime.studytime_id
WHERE student_id = '339'
OR student_id = '637';
```

Visualizations for Query 8