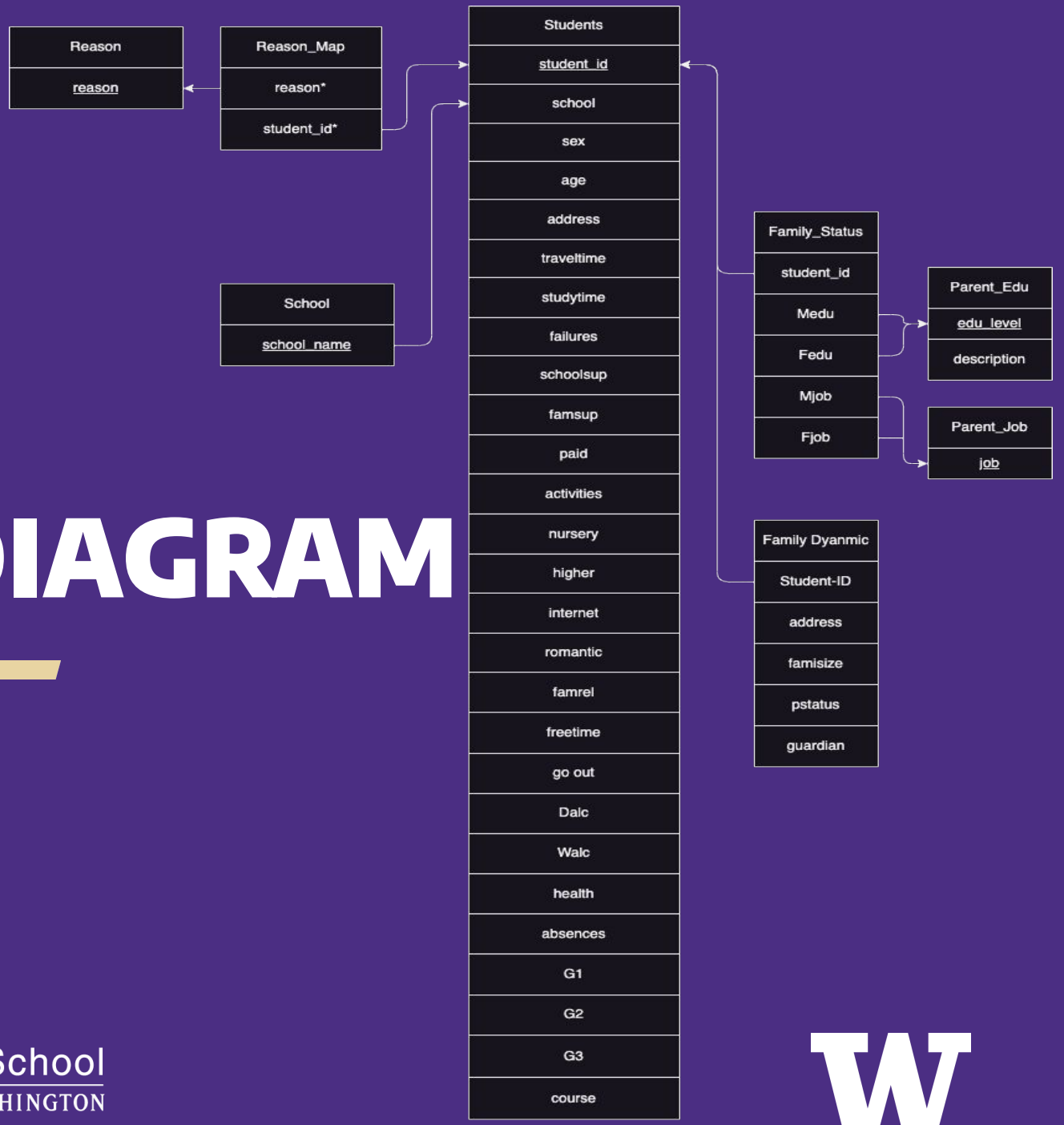
Final Project: MPUDAT40

Information School
UNIVERSITY of WASHINGTON

Dataset: Student Alcohol Consumption

school	sex	age	address	famsize	Pstatus	Medu	Fedu	Mjob	Fjob
GP	F	18	U	GT3	A	4	4	at_home	teacher
GP	F	17	U	GT3	T	1	1	at_home	other
GP	F	15	U	LE3	T	1	1	at_home	other
GP	F	15	U	GT3	T	4	2	health	services
GP	F	16	U	GT3	T	3	3	other	other
GP	M	16	U	LE3	T	4	3	services	other
GP	M	16	U	LE3	T	2	2	other	other
GP	F	17	U	GT3	A	4	4	other	teacher
GP	M	15	U	LE3	A	3	2	services	other
GP	M	15	U	GT3	T	3	4	other	other
GP	F	15	U	GT3	T	4	4	teacher	health
GP	F	15	U	GT3	T	2	1	services	other
GP	M	15	U	LE3	T	4	4	health	services
GP	M	15	U	GT3	T	4	3	teacher	other
GP	M	15	U	GT3	A	2	2	other	other

Small sample of the dataset



ERD DIAGRAM

Information School
UNIVERSITY of WASHINGTON

Queries

Primary reason for a student's alcohol consumption and average age group per category

```
SELECT reason, COUNT(*) AS occurrence
FROM Reason_Map
GROUP BY reason
ORDER BY occurrence DESC;
```

reason	occurrence
course	430
home	258
reputation	248
other	108

Primary reason for a student's alcohol consumption and average age group per category

```sql
SELECT r.reason AS primary_reason, ROUND(AVG(s.age), 1) AS average_age_group
FROM Student s
JOIN Reason_Map rm ON s.student_id = rm.student_id
JOIN Reason r ON rm.reason = r.reason
GROUP BY r.reason;
```

primary_reason	average_age_group
course	16.8
home	16.7
other	16.8
reputation	16.7

Parental status and alcohol consumption

```sql
SELECT fd.Pstatus AS parental_status,
    ROUND((COUNT(*) FILTER (WHERE r.reason = 'home') * 100.0 / COUNT(*)), 2) AS home_percentage,
    ROUND((COUNT(*) FILTER (WHERE r.reason <> 'home') * 100.0 / COUNT(*)), 2) AS other_percentage
FROM Student s
JOIN Reason_Map rm ON s.student_id = rm.student_id
JOIN Reason r ON rm.reason = r.reason
JOIN Family_Dynamic fd ON fd.student_id = s.student_id
WHERE fd.Pstatus = 'A' AND s.Dalc > 0;
```

parental_status	home_percentage	other_percentage
A	28.93	71.07

Correlation between education levels of parents (Medu, Fedu) and student performance (Jesus):

```sql
-- Query 1

SELECT pe1.edu_description AS Medu, pe2.edu_description AS Fedu,
    ROUND(AVG(s.G1), 2) AS G1_average,
    ROUND(AVG(s.G2), 2) AS G2_average,
    ROUND(AVG(s.G3), 2) AS G3_average
FROM Student s
JOIN Family_Status fs ON s.student_id = fs.student_id
JOIN Parent_edu pe1 ON fs.Medu = pe1.edu_level
JOIN Parent_edu pe2 ON fs.Fedu = pe2.edu_level
GROUP BY fs.Medu, fs.Fedu;
```

The impact of personal lifestyle choices on academic performance (Jesus):

```sql
SELECT r.reason, s.romantic,
       ROUND(AVG(s.Dalc), 2) AS average_Dalc,
       ROUND(AVG(s.Walc), 2) AS average_Walc,
       ROUND(AVG(s.goout), 2) AS average_goout,
       ROUND(AVG(s.G1), 2) AS average_G1,
       ROUND(AVG(s.G2), 2) AS average_G2,
       ROUND(AVG(s.G3), 2) AS average_G3
FROM Student s
JOIN Reason_Map rm ON s.student_id = rm.student_id
JOIN Reason r ON rm.reason = r.reason
WHERE s.Dalc > 0 AND s.Walc > 0 -- Filter for students who consume alcohol
GROUP BY r.reason, s.romantic;
```

	reason	romantic	average_Dalc	average_Walc	average_goout	average_G1	average_G2	average_G3
1	course	no	1.41	2.28	3.21	11.06	11.16	11.31
2	course	yes	1.54	2.23	3.18	10.49	10.21	10.34
3	home	no	1.55	2.35	3.16	11.12	11.22	11.37
4	home	yes	1.6	2.24	3.09	11.62	11.56	11.36
5	other	no	1.81	2.63	3.06	10.5	10.84	10.76
6	other	yes	1.91	2.46	3.26	11.07	11.04	10.98
7	reputation	no	1.34	2.16	3.09	12.3	12.5	12.68
8	reputation	yes	1.3	2.19	3.17	11.18	10.96	11.06

How the students home address can impact their academic performance

```sql
/*
  address and student performance(Tyler):
*/

SELECT fd.address AS addy,
  (COUNT(*) FILTER (WHERE s.G3 > 13 AND fd.address = 'U' )) AS urban_high_perf,
  (COUNT(*) FILTER (WHERE s.G3 > 6 AND s.G3 < 14 AND fd.address= 'U')) AS urban_avg_perf,
  (COUNT(*) FILTER (WHERE s.G3 < 7 AND fd.address= 'U')) AS urban_low_perf,
  (COUNT(*) FILTER (WHERE s.G3 > 13 AND fd.address= 'R' )) AS rural_high_perf,
  (COUNT(*) FILTER (WHERE s.G3 > 6 AND s.G3 < 14 AND fd.address = 'R' )) AS rural_avg_perf,
  (COUNT(*) FILTER (WHERE s.G3 < 7 AND fd.address = 'R' )) AS rural_low_perf
FROM Student s
JOIN Family_Dynamic fd ON fd.student_id = s.student_id;
```

urban_high_perf	urban_avg_perf	urban_low_perf	rural_high_perf	rural_avg_perf	rural_low_perf
237	468	54	57	201	27

How the sexual orientation of a student can impact their academic performance

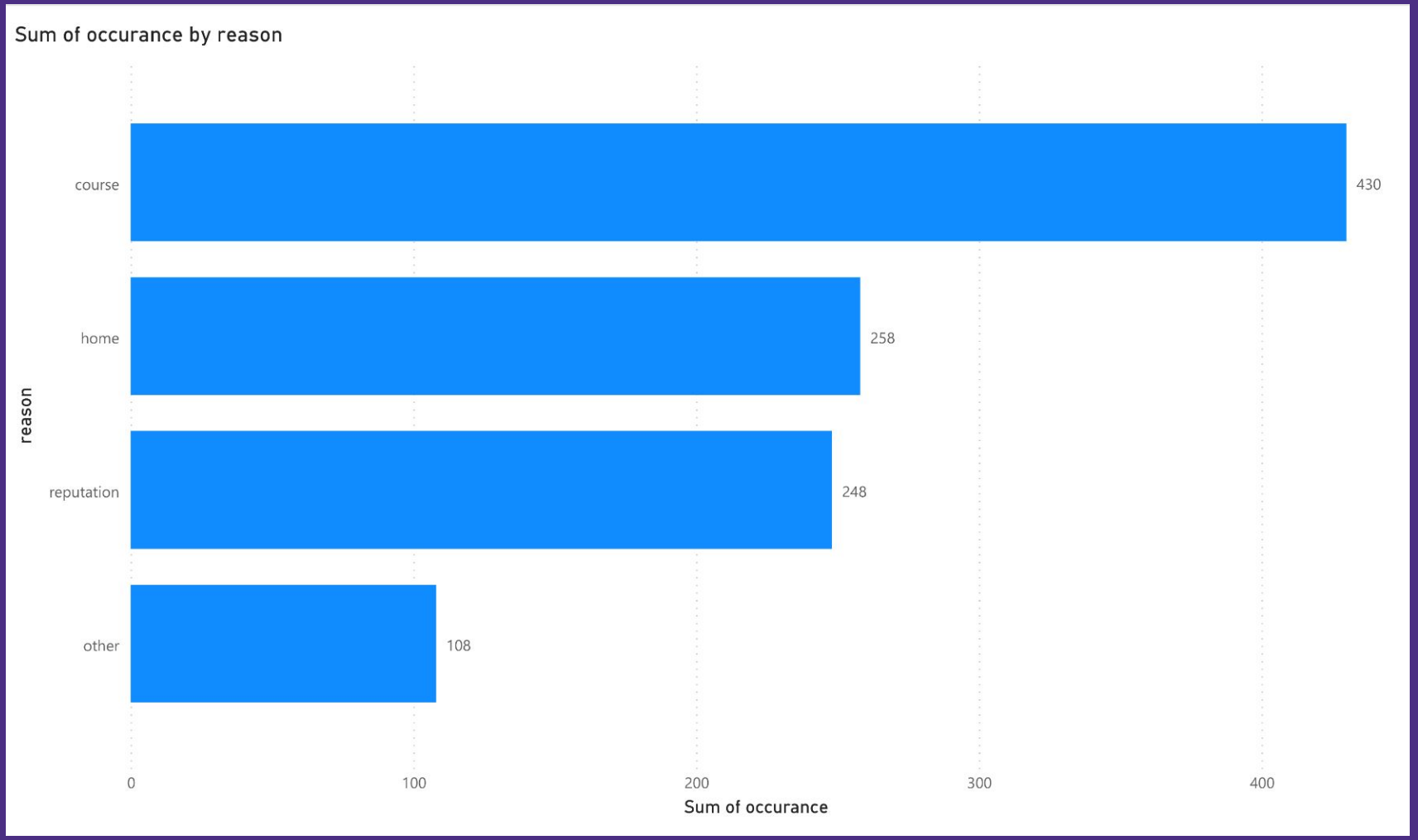
```sql
/*
  sex and student performance(Tyler):
*/

SELECT s.sex AS sex,
  ROUND((COUNT(*) FILTER (WHERE s.G3 > 13 AND s.sex = 'F') * 100.0 / COUNT(*) FILTER(WHERE s.sex = 'F')), 2) AS female_percentage_high_perf,
  ROUND((COUNT(*) FILTER (WHERE s.G3 > 6 AND s.G3 < 14 AND s.sex = 'F') * 100.0 / COUNT(*) FILTER(WHERE s.sex = 'F')), 2) AS female_percentage_avg_perf,
  ROUND((COUNT(*) FILTER (WHERE s.G3 < 7 AND s.sex = 'F') * 100.0 / COUNT(*) FILTER(WHERE s.sex = 'F')), 2) AS female_percentage_low_perf,
  ROUND((COUNT(*) FILTER (WHERE s.G3 > 13 AND s.sex = 'M') * 100.0 / COUNT(*) FILTER(WHERE s.sex = 'M')), 2) AS male_percentage_high_perf,
  ROUND((COUNT(*) FILTER (WHERE s.G3 > 6 AND s.G3 < 14 AND s.sex = 'M') * 100.0 / COUNT(*) FILTER(WHERE s.sex = 'M')), 2) AS male_percentage_avg_perf,
  ROUND((COUNT(*) FILTER (WHERE s.G3 < 7 AND s.sex = 'M') * 100.0 / COUNT(*) FILTER(WHERE s.sex = 'M')), 2) AS male_percentage_low_perf
FROM Student s;
```

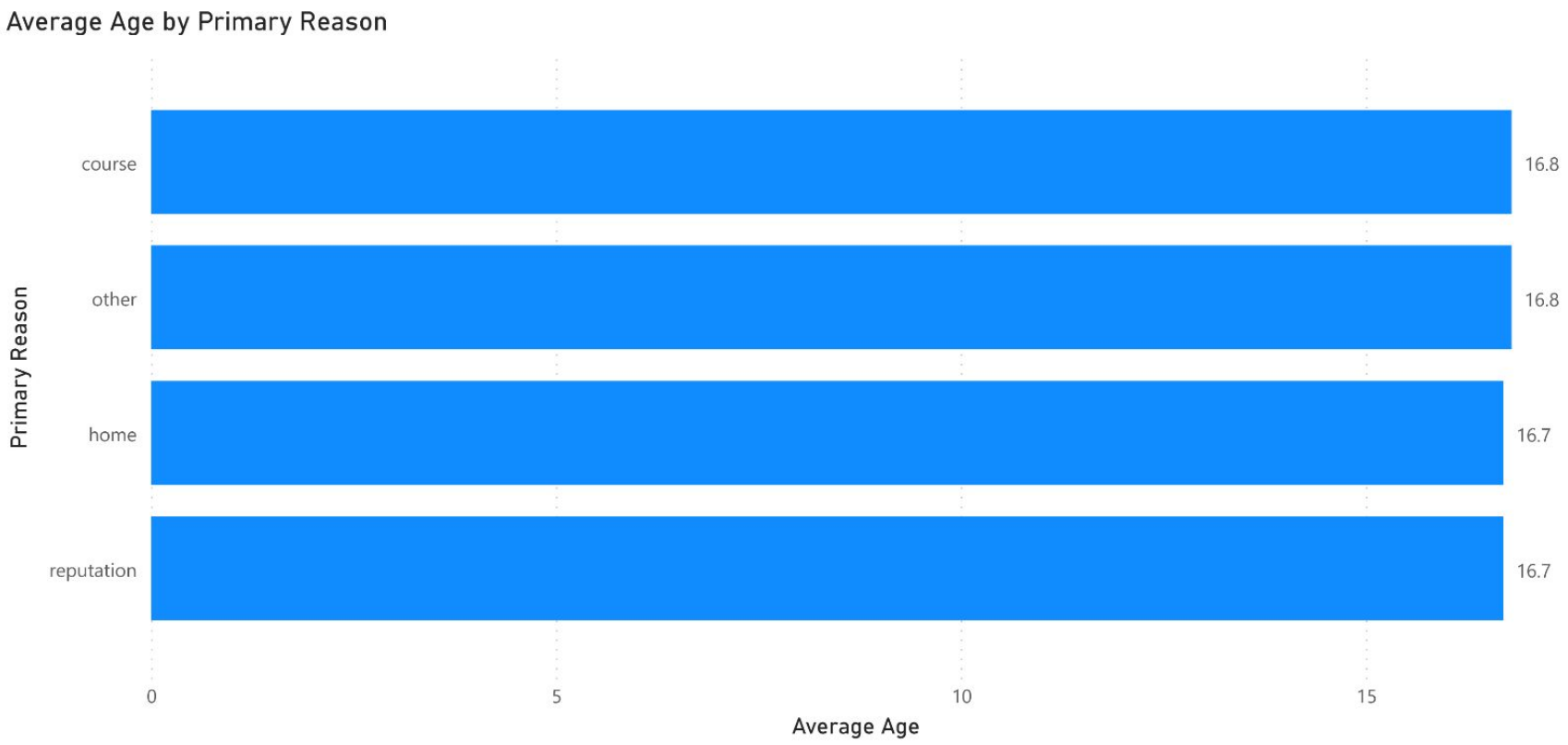
female_percentage_high_perf	female_percentage_avg_perf	female_percentage_low_perf	male_percentage_high_perf	male_percentage_avg_perf	male_percentage_low_perf
29.61	62.44	7.95	26.27	66.23	7.51

Visualizations

Primary reason for a student's alcohol consumption and average age group per category



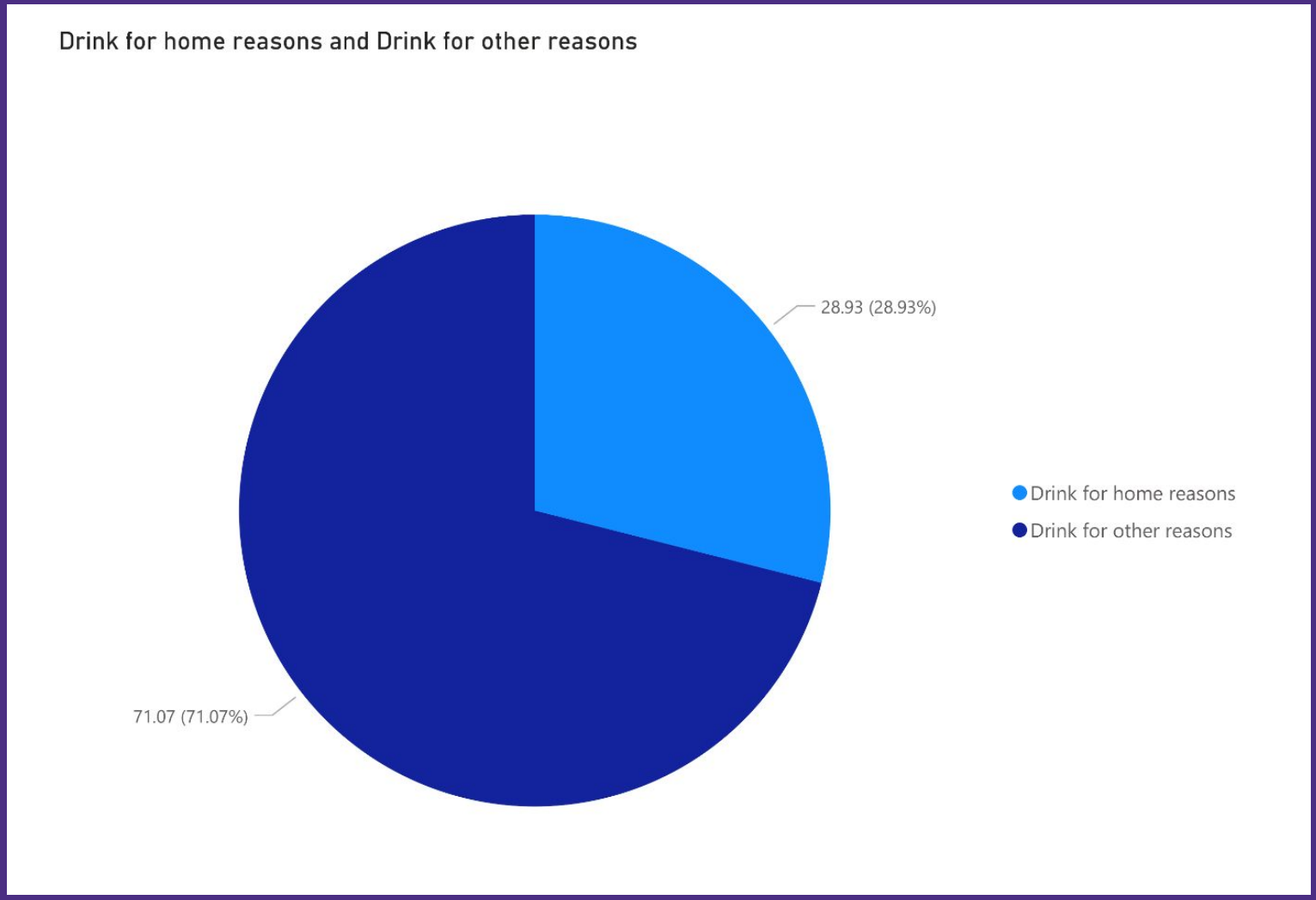
Primary reason for a student's alcohol consumption and average age group per category (cont.)



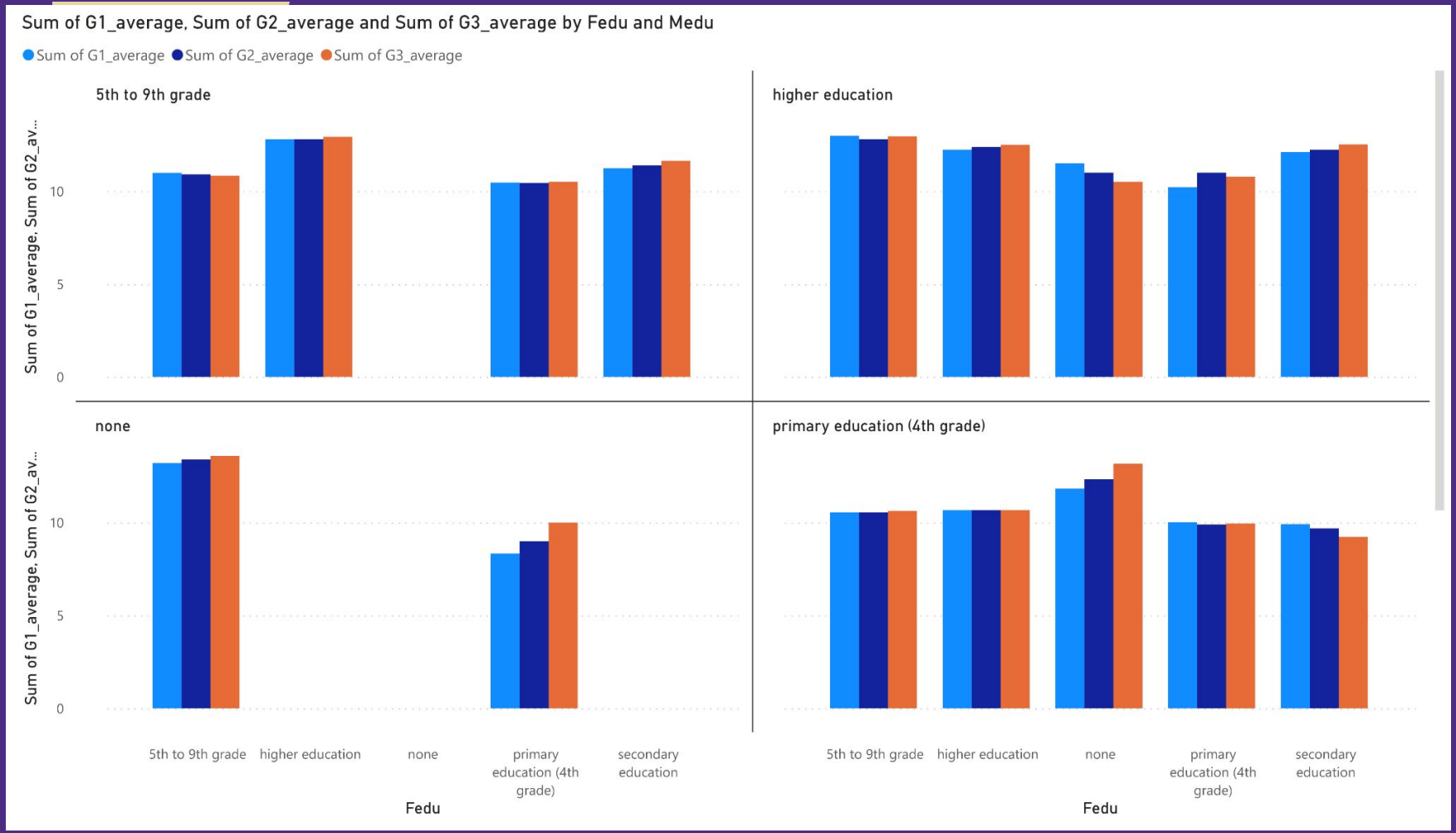
course and other tied for highest Sum of average_age_group at 16.80.

Across all 4 primary_reason, Sum of average_age_group ranged from 16.70 to 16.80.

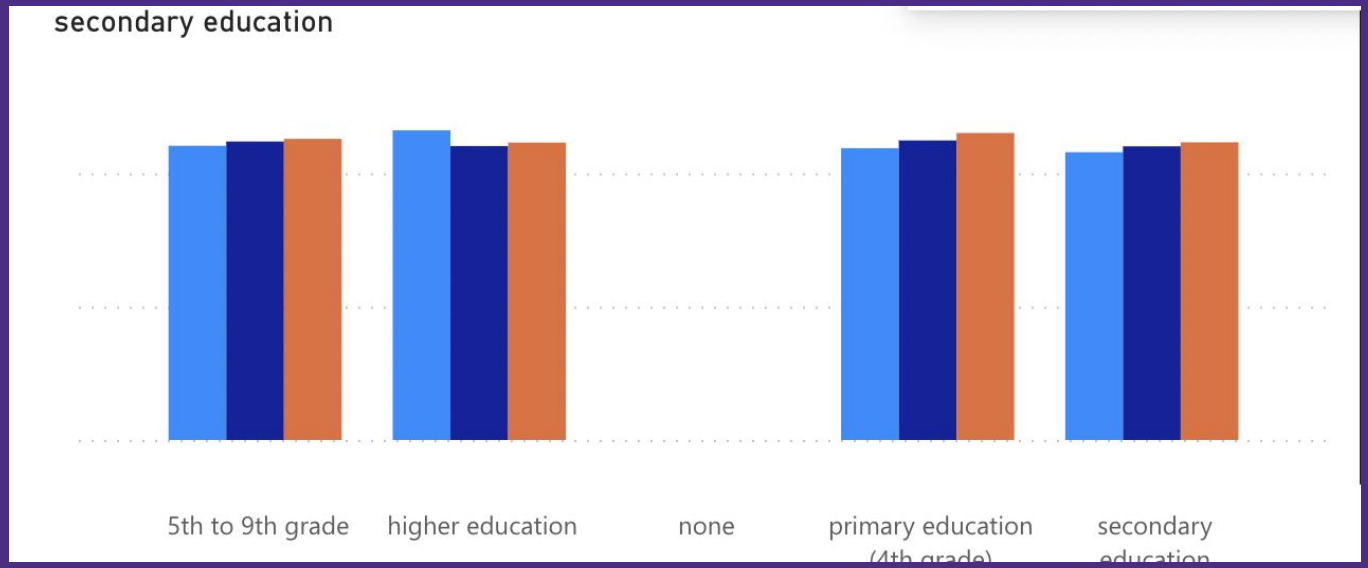
Parental status and alcohol consumption



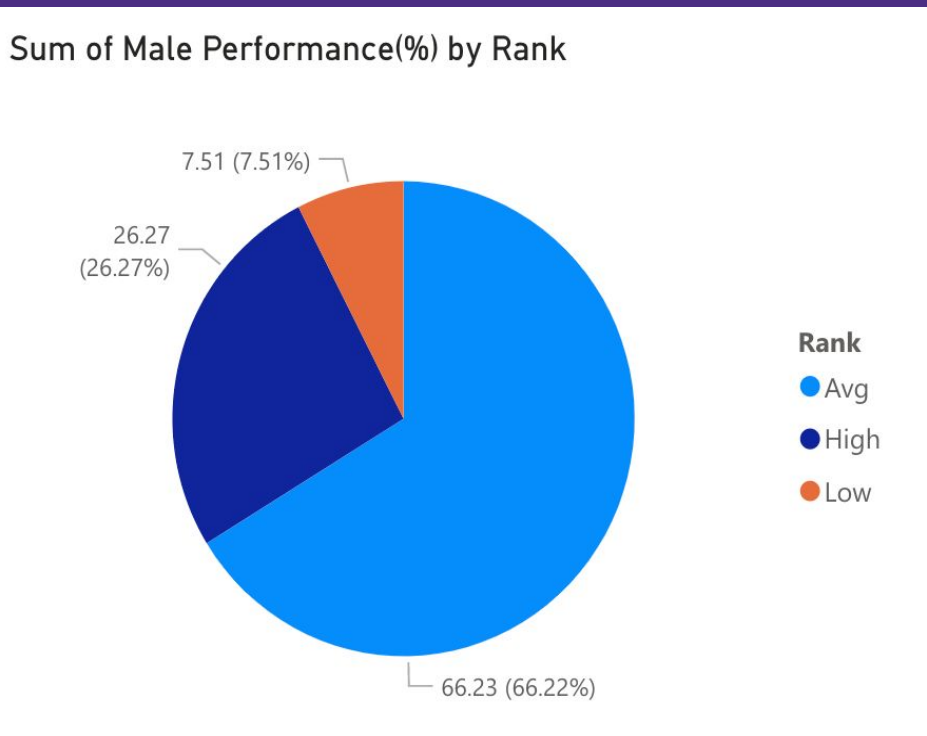
Correlation between education levels of parents (Medu, Fedu) and student performance (Jesus):

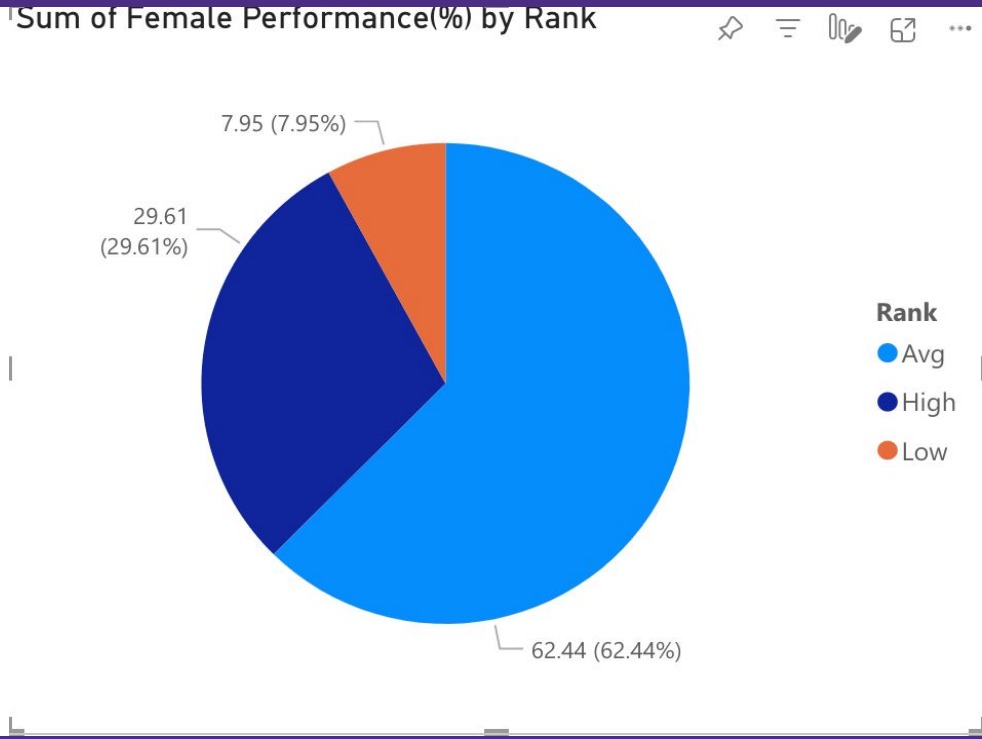


Correlation between education levels of parents (Medu, Fedu) and student performance (Jesus):



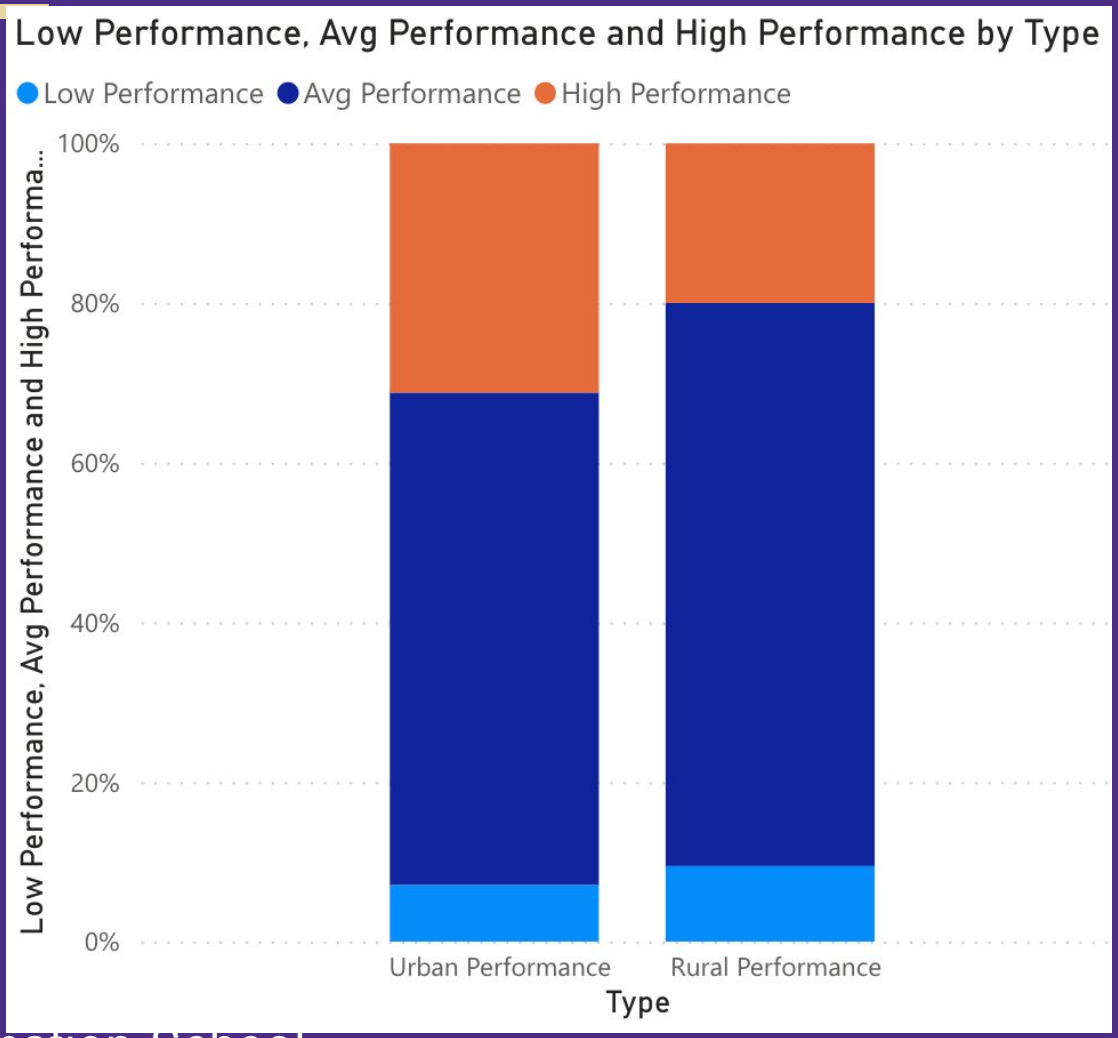
How the sexual orientation of a student can impact their academic performance





Information School
UNIVERSITY of WASHINGTON

How the students home address can impact their academic performance visualized



Information School
UNIVERSITY of WASHINGTON