

What makes a student perform well?

Binal Modi (bjm470)

Apurv Srivastav (as11419)

Project goal

Using data collected from [Center for Machine Learning and Intelligent Systems](#), we aim to explore the dataset with different attributes of high performing students and poor performing students to understand the roles of these associated attributes in a student's performance. To measure performance, we will consider average grades combining maths and Portuguese (native language) subjects' grades drawn from two Portuguese schools.

Data Set

Attributes:

Attribute	Description	Type	Range and Values
sex	Gender of Student 'F' - female or 'M' - male	Categorical	F, M
age	Student's age	Quantitative	15 - 22
address	student's home address type 'U' - urban or 'R' - rural	Categorical	U, R
Pstatus	parent's cohabitation status 'T' - living together or 'A' - apart	Categorical	T, A
Medu	mother's education 0 - without higher education, 1 - with higher education Although the original dataset has a range of values, we will consider them under educated category only if they have higher education for our purpose, so the new categories would be 0(without higher education) or 1(with higher education)	Categorical	0 - 1
Fedu	father's education 0 - without higher education, 1 - with higher education Although the original dataset has a range of values, we will consider them under educated category only if they have higher education for our purpose, so the new categories would be 0(without higher education) or 1(with higher education)	Categorical	0 - 1

Mjob	mother's job 'teacher', 'health' care related, civil 'services' (e.g. administrative or police), 'at_home' or 'other'	Categorical	Multiple
Fjob	father's job 'teacher', 'health' care related, civil 'services' (e.g. administrative or police), 'at_home' or 'other'	Categorical	Multiple
reason	reason to choose this school close to 'home', school 'reputation', 'course' preference or 'other'	Categorical	Multiple
studytime	weekly study time 1 - <2 hours, 2 - 2 to 5 hours, 3 - 5 to 10 hours, or 4 - >10 hours)	Quantitative	2 - 10
failures	number of past class failures n if $1 \leq n < 3$, else 4	Quantitative	1 - 4
schoolsup	extra educational support yes or no	Categorical	Yes No
famsup	family educational support yes or no	Categorical	Yes No
activities	extra-curricular activities yes or no	Categorical	Yes No
romantic	with a romantic relationship yes or no	Categorical	Yes No
famrel	quality of family relationships 1 - very bad to 5 - excellent	Ordinal	1 - 5
freetime	free time after school 1 - very low to 5 - very high	Ordinal	1 - 5
goout	going out with friends 1 - very low to 5 - very high	Ordinal	1 - 5
Dalc	workday alcohol consumption 1 - very low to 5 - very high	Ordinal	1 - 5
Walc	weekend alcohol consumption 1 - very low to 5 - very high	Ordinal	1 - 5
Grade	G3 - final grade numeric: from 0 to 20, output target	Quantitative	Multiple

Analytical Questions and Proxy Tasks

How does family background influence a student's performance?

- Does parents education (having a higher secondary education or above) contribute to an individual's performance?
 - medu (Mother's education)
 - fedu (Father's education)
 - Grade
- If the parents' education does contribute to the performance, does the quality of family relationship also matters to improve their performance?
 - pstatus (parent's cohabitation status)
 - famrel (quality of family relationships)
 - Grade

How lifestyle contributes to their performance in school?

- Does having free time after school improve their performance?
 - Freetime (free time after school)
- Having free time but being in a romantic relationship or outing or alcohol consumption takes a toll on the performance?
 - romantic (with a romantic relationship)
 - goout (going out with friends)
 - dalc (workday alcohol consumption)
 - walc (weekend alcohol consumption)
 - Grade
- Even though frequent alcohol consumption, but if they have sufficient study time is the performance affected?
 - studytime (weekly study time)
 - dalc (workday alcohol consumption)
 - walc (weekend alcohol consumption)
 - Grade
- Studios students(with >10 hours study time) but no extracurricular activities have better performance than students with moderate (4-10 hours) study time and extracurricular activities?
 - activities (extra-curricular activities)
 - studytime (weekly study time)
 - Grade

How school and facilities contribute to students performance?

- Does the reason for choosing the school influence their performance?
 - reason (reason to choose this school)
 - Grade
- Does extra support from school improve their performance?

- schoolsup (extra educational support from school)
 - Grade
- Although extra support from school but no family education support affect their performance?
 - schoolsup (extra educational support from school)
 - famsup (family educational support)
 - Grade

How student's gender and personal experiences influence their performance?

- How performance varies according to the sex and age of a student?
 - sex (gender of student)
 - age (age of student)
 - Grade
- Does the performance vary for students of the same age if they come from a rural or urban area?
 - address (student's home address type)
 - sex (gender of student)
 - age (age of student)
 - Grade
- If previous setbacks and failure affect their performance?
 - failures(number of past class failures)
 - Grade

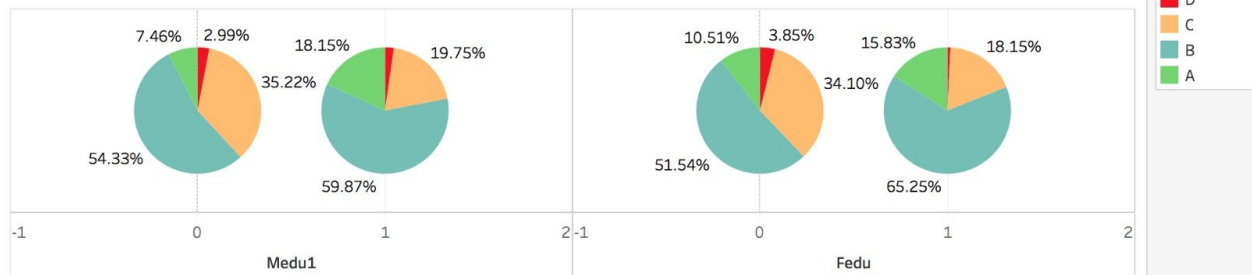
Story Design

Data Analysis

How does family background influence a student's performance?

- Does parents education (having a higher secondary education or above) contribute to an individual's performance?
 - medu (Mother's education)
 - fedu (Father's education)
 - Grade

Parent Education



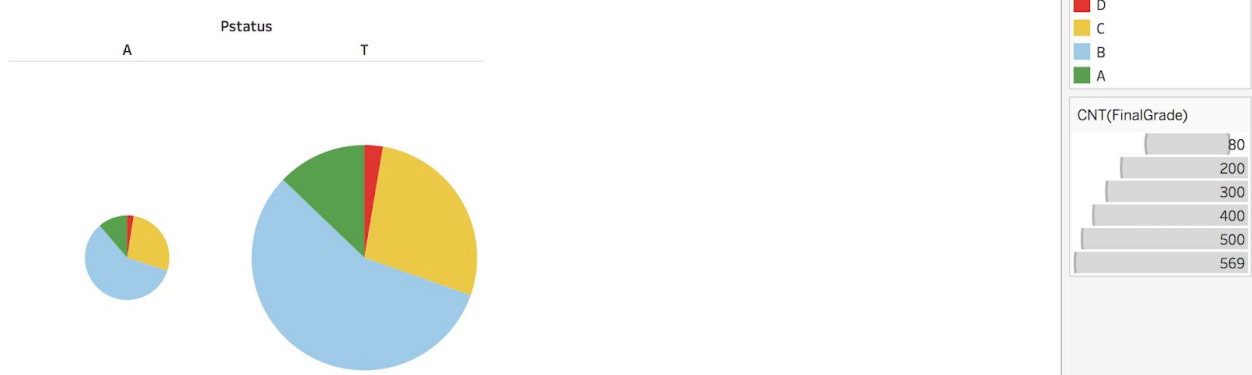
- If the parents' education does contribute to the performance, does the quality of family relationship also matters to improve their performance?
 - pstatus
 - famrel (quality of family relationships)
 - Grade

Family Relation



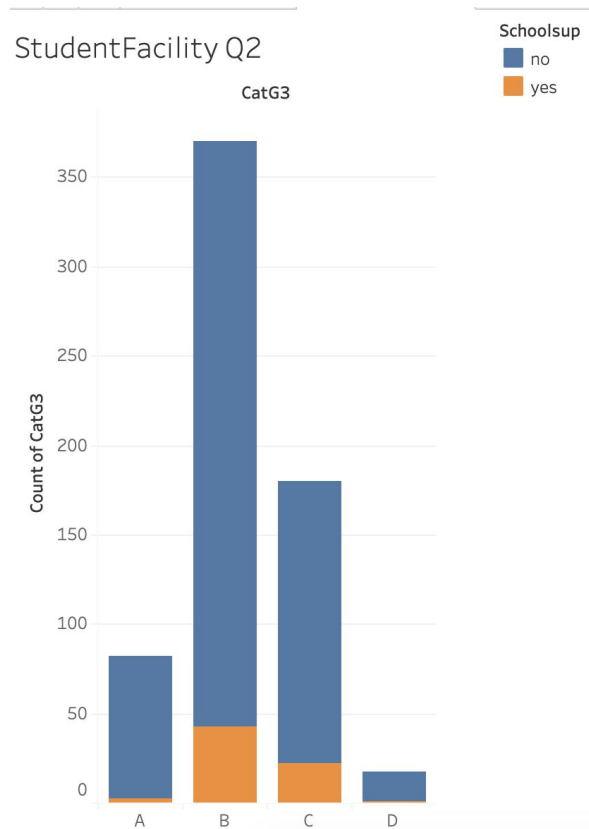
This shows family relationship and their effect on grades

Paresnt Status



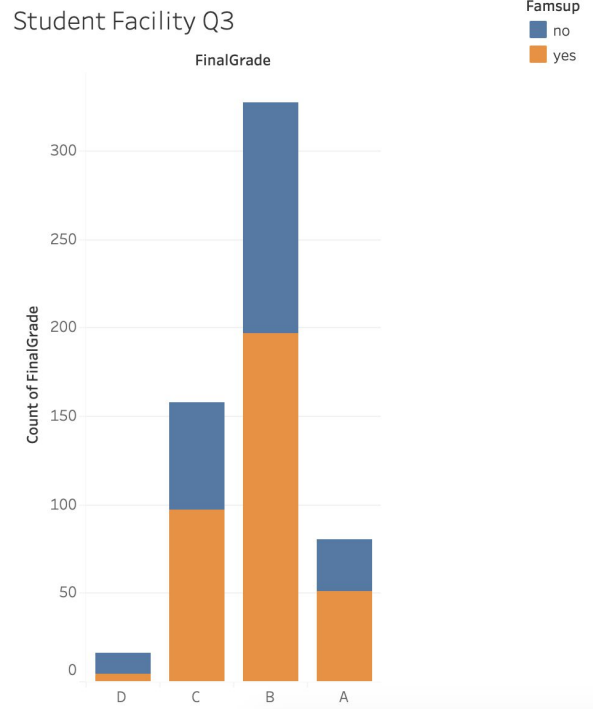
How school and other facilities contribute to students performance?

- Does extra support from school improve their performance?



We can clearly see from the graph above that having school support is not making a significant impact on performance. Students with higher grades have had no school support and also for average scorers the group of students with no school support dominate.

- Although no school support but does having family education support affect their performance?
(Note: We changed the question a bit, because we thought it would be more interesting to see for the major scorers)
 - schoolsup (extra educational support from school)
 - famsup (family educational support)
 - Grade



How student's gender and personal experiences influence their performance?

- How performance varies according to the sex and age of a student?

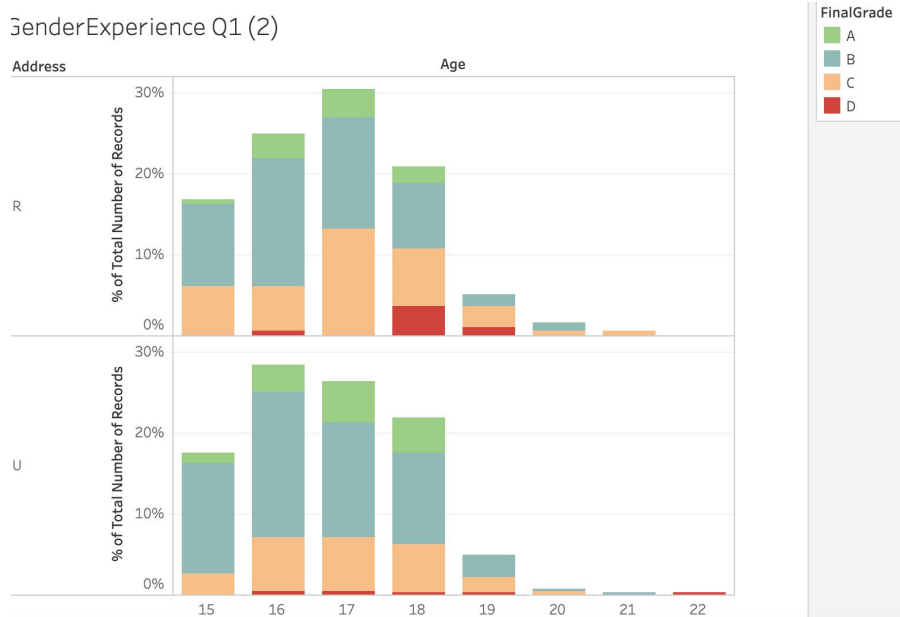


Unlike usual prejudice, here we can see that there is no significant difference in the performance of students when comparing genders. However, as the age

increases we see students who score A reduces.

- Does the performance vary for students of the same age if they come from a rural or urban area?
 - address (student's home address type)
 - age (age of student)
 - Grade

GenderExperience Q1 (2)

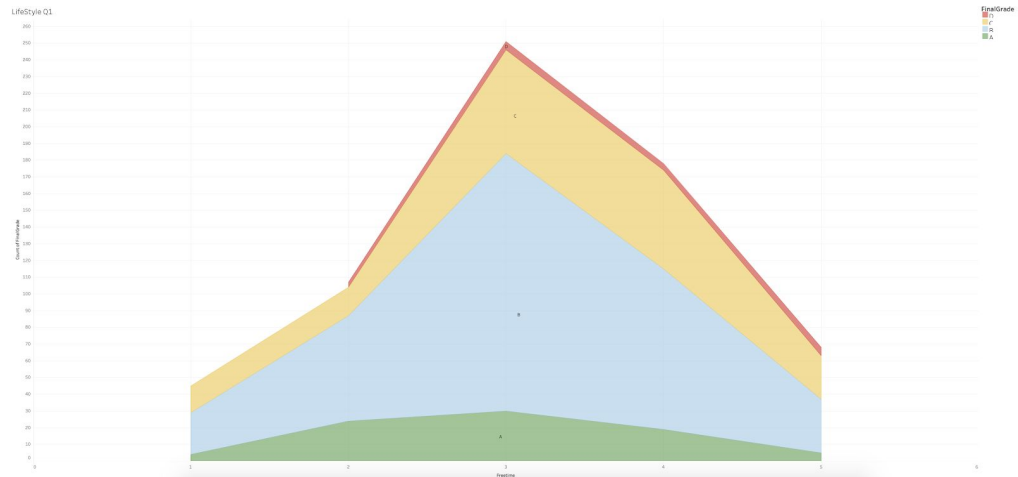


This visualisation tells that there are usually less students scoring D from the Urban areas and also that students from urban areas have usually have A or B grades

- If previous setbacks and failure affect their performance?
 - failures(number of past class failures)
 - Grade

How lifestyle contributes to their performance in school?

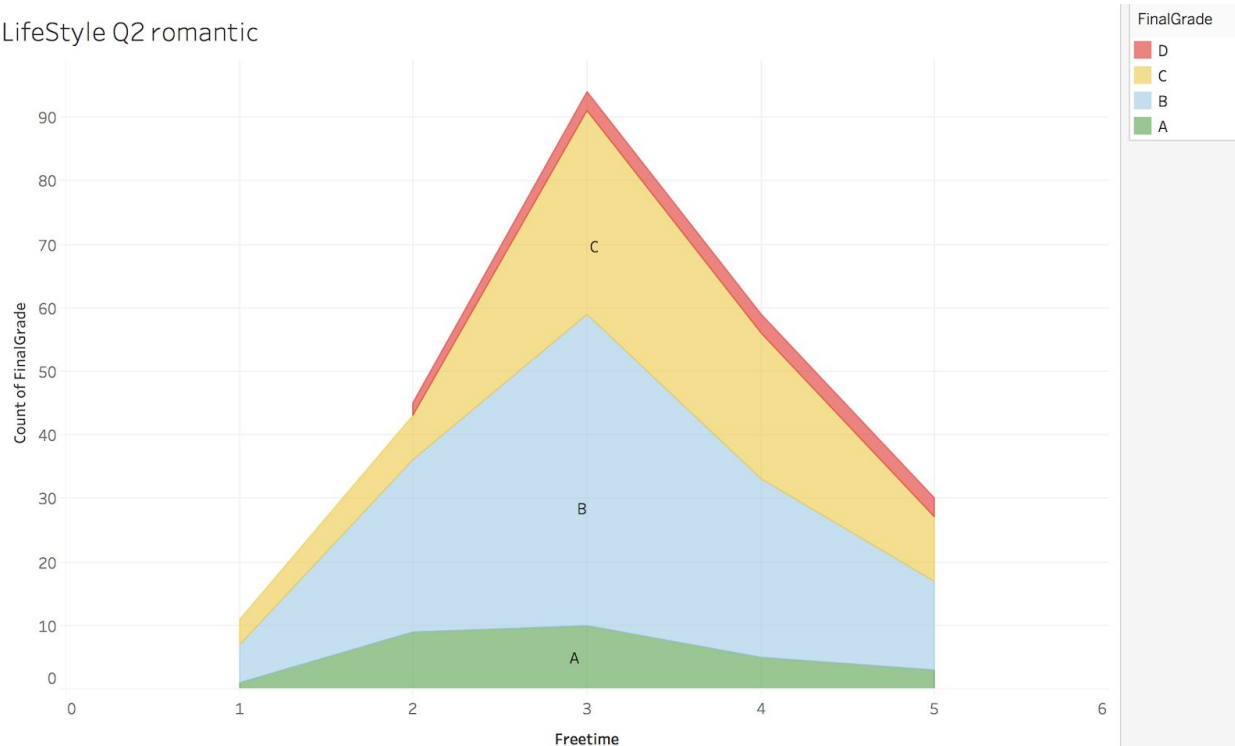
- Does having free time after school improve their performance?
 - Freetime (free time after school)



This shows that, even if they have free time after a certain amount, it doesn't contribute positively to their performance.

- Having free time but being in a romantic relationship or outing or alcohol consumption takes a toll on the performance?
 - romantic (with a romantic relationship)
 - goout (going out with friends)
 - dalc (workday alcohol consumption)
 - walc (weekend alcohol consumption)
 - Grade

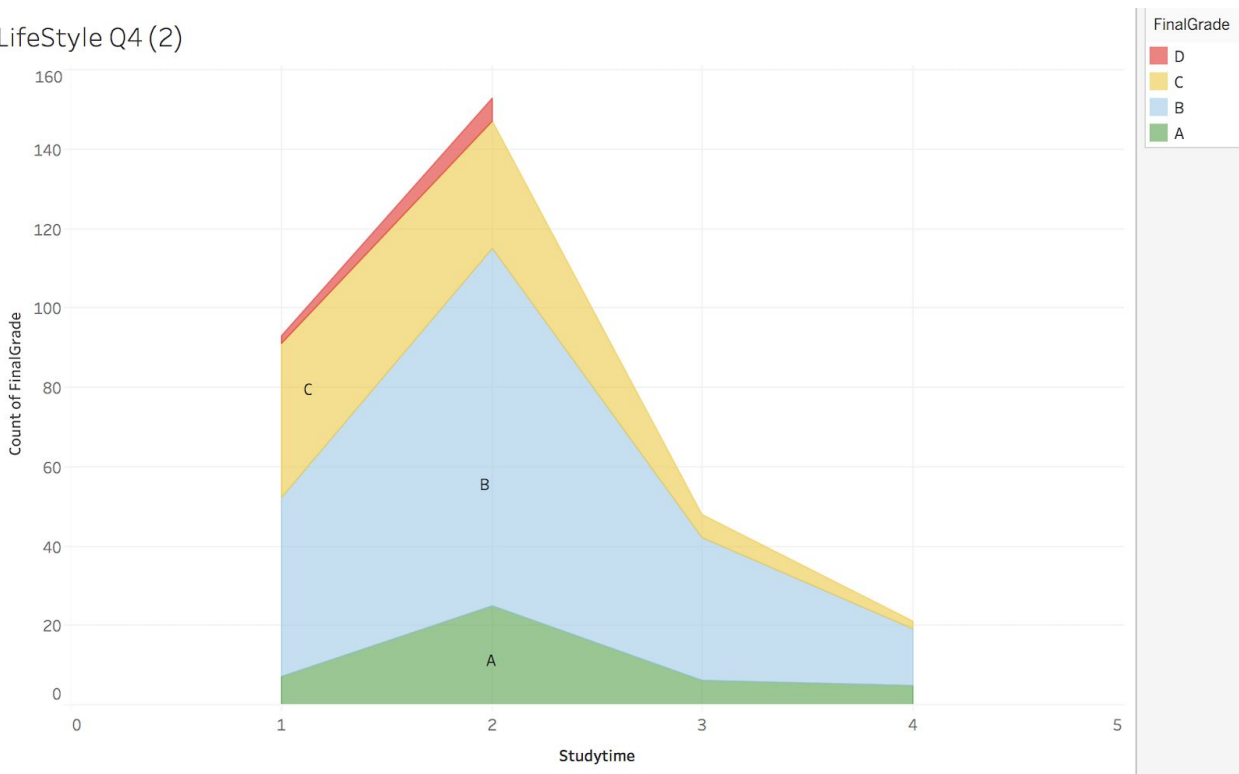
LifeStyle Q2 romantic



This shows that having a romantic relationship does not contribute much when filtered with freetime and shows same kind of trend.

- Even though frequent alcohol consumption, but if they have sufficient study time is the performance affected?
 - studytime (weekly study time)
 - dalc (workday alcohol consumption)
 - walc (weekend alcohol consumption)
 - Grade

LifeStyle Q4 (2)



Alcoholic students with 3-4 study time failed to improve their grade though they managed to avoid D grade.

Storyboard

Note- Here, we have written how our story would go about for the visualisation project and with every bullet point we will move to a new page and discover more about the analysis.

Topic - What makes a student perform well?

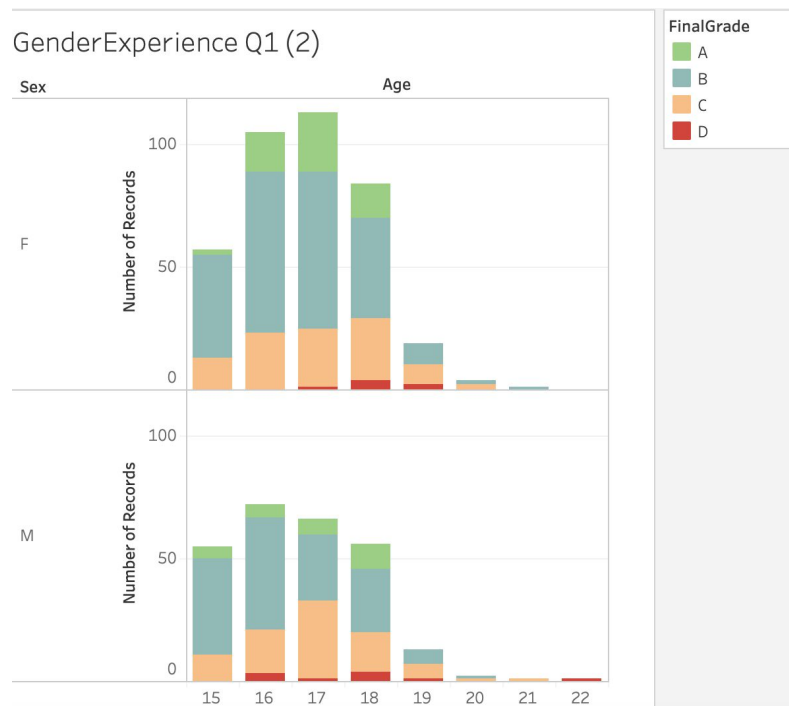
INTRODUCTION

While exploring our dataset of students performance for two portuguese schools, we came across the idea of exploring the data and analysing the factors that contribute to their performance.

ANALYSIS

Of course, there are multiple factors that impacts a students performance. However, we wanted to take into consideration some of the major factors that would be very essential for a students performance such as and students demographics, their family background, school support, family support and lifestyle.

- Peeling the layers one by one, we first want to check if the **demographics** make a difference when it comes to grades.

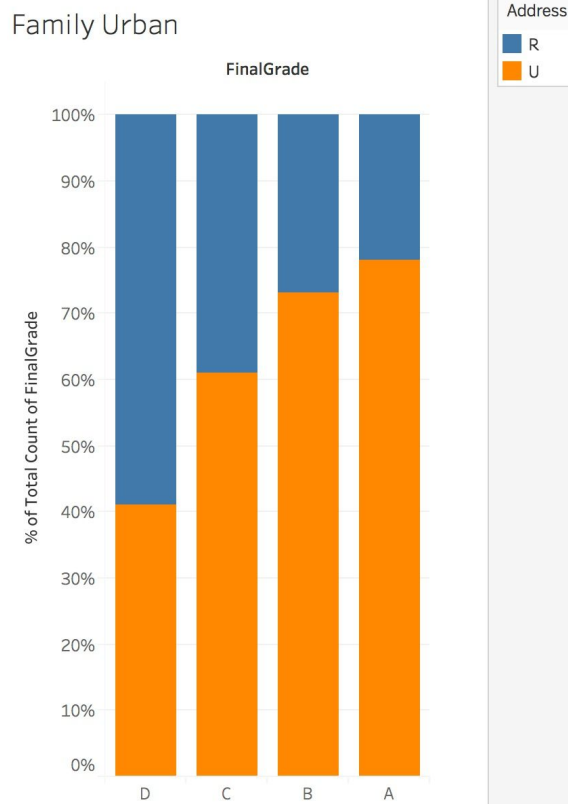


(We could plot a graph showing the grades division for each gender with different bars for different age. We plan to add the category 20,21 and 22 in age since they have very less values and show it as a combined category of 20+)

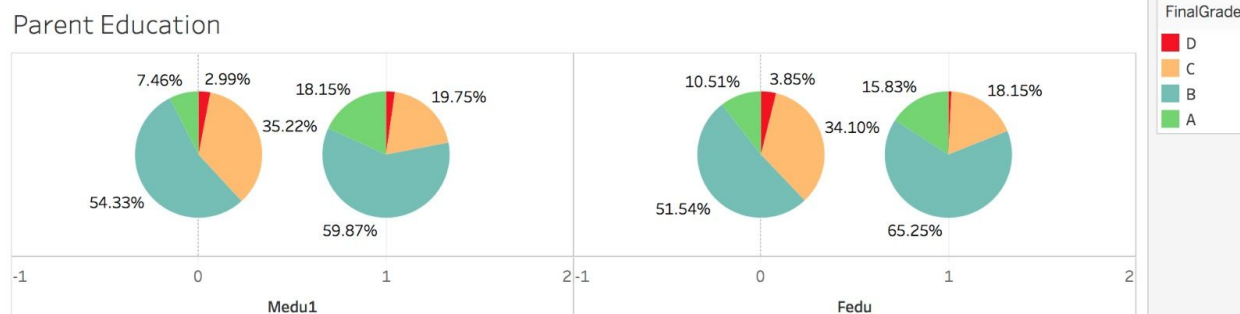
This visualisation does not show much difference of distribution between male and female category but it however helps us rule out other misbeliefs like males perform poor and females are more serious, since we can see a similar kind of distribution for both.

- We then go deeper and check if the **students background** has an impact on his/her performance. The two things that come into our mind when we think about background is a person's hometown and their family background. So we analyse on those further:

- Below, we will show how grades vary when students come from rural or urban areas.

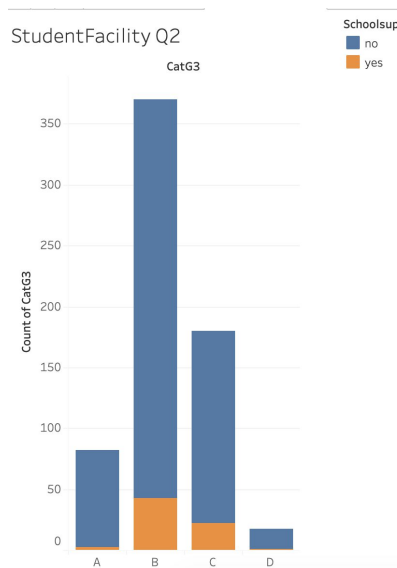


- It will also be interesting to show how fathers and mothers education impacts a students performance.



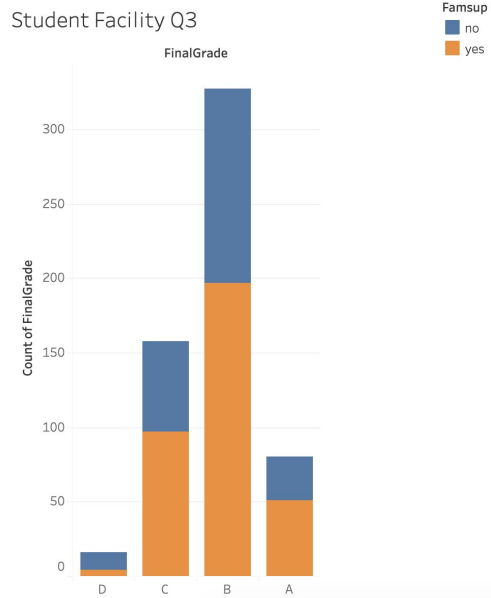
This can be seen from the graph that, if the the mother is educated, we see more students scoring A and B grades while under the other category when mother is not educated,we see more of B and C grades. Same pattern is observed in father's education but it is a little less mildly correlated.

- The previous analysis helped us to explore students performance according to their family background. But we also know that students also need **support** sometimes and that may impact their performance. So, we decided to analyse that aspect as well.
 - First, we checked if having an extra education support from the school helps students to perform better. We plotted the graph below that gives us an insight on the above.

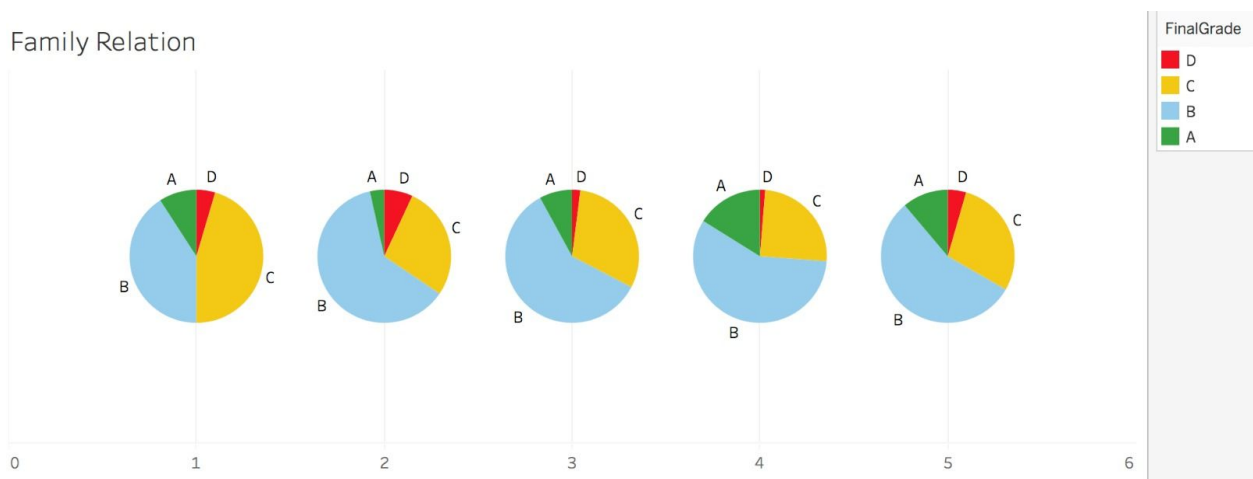


This visualisation tells us that actually it doesn't make a significance difference if there is an extra school support since the count of the students dominate in A,B,C grades scoring same even with no educational support.

- Since the students with no educational support dominated, we thought it would be interesting to if actually those students are performing well as they have family support and maybe that has a relation. So, we filtered the students with no extra school support and checked for their family education support to gain deeper knowledge.

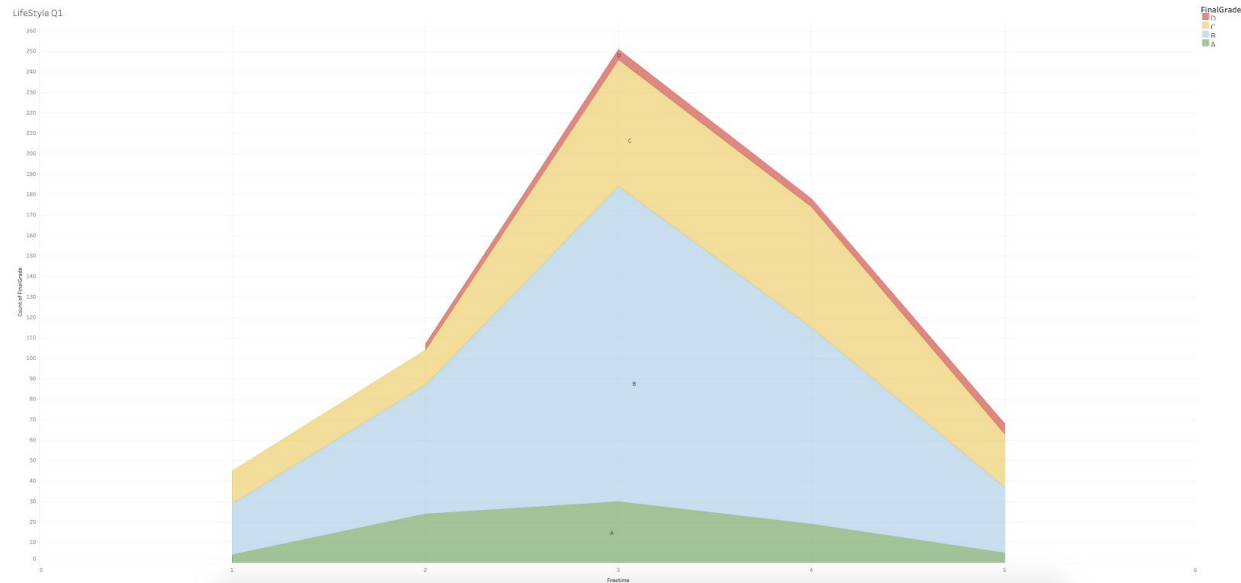


- With the educational support, it is also necessary how students relationships are with their parents and we also took that into consideration and plotted the below graph



We can note with this that as the as the family relationship gets better we see more students with A and B grades.

- Apart from the support that students get, we know other personal factors also impact on their overall performance and thus we analysed some of those
 - First aspect we thought would be interesting is if they have enough free time after school and if that helps their performance?



With the increase in area we can see that the performance increases until its in category 3 but after that even if they more free time it doesn't necessarily have a positive affect on their performance.

Changelog

Include any changes you did on the first proposal you submitted.

We made a couple of changes -

Dataset -

- We felt it would be best to club the G3 grading to A,B,C,D rather than having the whole range to have good quantity of data to analyse according to category. So we picked the range,
 A: 0 - 6
 B: 6 - 11
 C: 11 - 16
 D: 16 - 21

- For father's and mother's education, we saw that we were not able to create equal samples to show if is educated or not educated. So, we changed the the metrics for education as 0 (from not educated until 9th grade) and 1(secondary and higher education)

Proxy tasks

We have also changed a few questions after analysing the analysis questions that we had listed and we have changed those questions in the story board.