Final Project

Meghaan Binkley-Hopper

February 22, 2024

Section 1. Introduction

This project analyzes student performance in a class. Different characteristics and factors are given for each student along with the total number of failures that they have had in the class. This study compares several of the characteristics and factors to the total number of failures. It is the goal of this analysis to provide indicators as to whether or not a student is likely to fail.

Project available via Tableau Public: https://public.tableau.com/views/MBinkley-Hopper_FinalProject/Goal1?:language=en-US&publish=yes&:sid=&:display_count=n&:origin=viz_share_link

Project available via GitHub: https://github.com/Meghaan-Binkley-Hopper/DataVisualization-FinalProject.git

Section 2. Data Description

Data Domain Description: Student Performance Data

Data Source Description: https://www.kaggle.com/datasets/devansodariya/student-performance-data

Number of Records: 395

Detailed Explanation of Attributes:

- sex: text; options of F *(female)* or M *(male)*
- age: number; ranging from 15-22
- famsize: text; family size, GT3 represents greater than 3 and LE3 represents less than or equal to 3
- Medu: number; mother's education, ranked 1-4
- Fedu: number; father's education, ranked 1-4
- Mjob: text; mother's job, options are at_home, health, services, teacher, and other
- Fjob: text; father's job, options are at_home, health, services, teacher, and other
- studytime: number; number of hours 1-4
- failures: numbers; number of failures
- activities: text; yes or no on whether the student participates in activities
- nursery: text; yes or no on whether the student participated in nursery school
- internet: text; yes or no on whether the student has access to internet in their home
- absences: number; number of absences

Assumptions: None

Goals:

- Compare the student's sex to the number of failures
- Compare the student's age to the number of failures
- Compare the amount of study time to number of failures
- Compare the class absences to number of failures
- Compare mother's job to number of failures

- Compare father's job to number of failures

Section 3. Data Cleaning Strategies

This data set was complete and clean upon download. There were some columns that the creator of the dataset did not explain completely, so they were deleted.

Section 4. Clean Dataset

The screenshot below shows an excerpt of the data used in this project. Overall, there were 395 students included in the dataset. each record of the dataset below depicts a student along with how many times that they failed an exam within the class as well as some different features about them. The features about each student that are included are sex, age, family size, mother's education, father's education, mother's job, father's job, amount of study time, involvement in activities, participation in nursery school, internet access in their home, and number of absences.

	A	B	C	D	E	F	G	H	I	J	K	L	M	N
1	sex	age	famsize	Medu	Fedu	Mjob	Fjob	studytime	failures	activities	nursery	internet	absences	
2	F	18	GT3	4	4	at_home	teacher	2	0	no	yes	no	6	
3	F	17	GT3	1	1	at_home	other	2	0	no	no	yes	4	
4	F	15	LE3	1	1	at_home	other	2	3	no	yes	yes	10	
5	F	15	GT3	4	2	health	services	3	0	yes	yes	yes	2	
6	F	16	GT3	3	3	other	other	2	0	no	yes	no	4	
7	M	16	LE3	4	3	services	other	2	0	yes	yes	yes	10	
8	M	16	LE3	2	2	other	other	2	0	no	yes	yes	0	
9	F	17	GT3	4	4	other	teacher	2	0	no	yes	no	6	
10	M	15	LE3	3	2	services	other	2	0	no	yes	yes	0	
11	M	15	GT3	3	4	other	other	2	0	yes	yes	yes	0	
12	F	15	GT3	4	4	teacher	health	2	0	no	yes	yes	0	
13	F	15	GT3	2	1	services	other	3	0	yes	yes	yes	4	
14	M	15	LE3	4	4	health	services	1	0	yes	yes	yes	2	
15	M	15	GT3	4	3	teacher	other	2	0	no	yes	yes	2	
16	M	15	GT3	2	2	other	other	3	0	no	yes	yes	0	
17	F	16	GT3	4	4	health	other	1	0	no	yes	yes	4	
18	F	16	GT3	4	4	services	services	3	0	yes	yes	yes	6	
19	F	16	GT3	3	3	other	other	2	0	yes	yes	no	4	
20	M	17	GT3	3	2	services	services	1	3	yes	yes	yes	16	
21	M	16	LE3	4	3	health	other	1	0	yes	yes	yes	4	
22	M	15	GT3	4	3	teacher	other	2	0	no	yes	yes	0	
23	M	15	GT3	4	4	health	health	1	0	no	yes	yes	0	
24	M	16	LE3	4	2	teacher	other	2	0	yes	yes	yes	2	
25	M	16	LE3	2	2	other	other	2	0	yes	yes	yes	0	
26	F	15	GT3	2	4	services	health	3	0	yes	yes	yes	2	
27	F	16	GT3	2	2	services	services	1	2	no	no	yes	14	
28	M	15	GT3	2	2	other	other	1	0	no	yes	yes	2	
29	M	15	GT3	4	2	health	services	1	0	no	yes	yes	4	
30	M	16	LE3	3	4	services	other	2	0	yes	yes	yes	4	
31	M	16	GT3	4	4	teacher	teacher	2	0	yes	yes	yes	16	
32	M	15	GT3	4	4	health	services	2	0	no	no	yes	0	
33	M	15	GT3	4	4	services	services	2	0	yes	yes	yes	0	
34	M	15	GT3	4	3	teacher	at_home	2	0	yes	yes	yes	0	
35	M	15	LE3	3	3	other	other	2	0	yes	no	yes	0	
36	M	16	GT3	3	2	other	other	1	0	no	no	yes	0	
37	F	15	GT3	2	3	other	other	1	0	yes	yes	no	0	

student_data | student performance

Section 5. Visualization Tools

For the purpose of this project, I utilized Tableau to visualize my data. I used a cleaned Excel file as my data source. From there, I used Tableau to create tables, treemaps, line graphs, and column charts. I also created a dashboard and storyboard. I chose to use Tableau for my visualizations because it provides a professional looking product, and I wanted to get more experience with utilizing it.

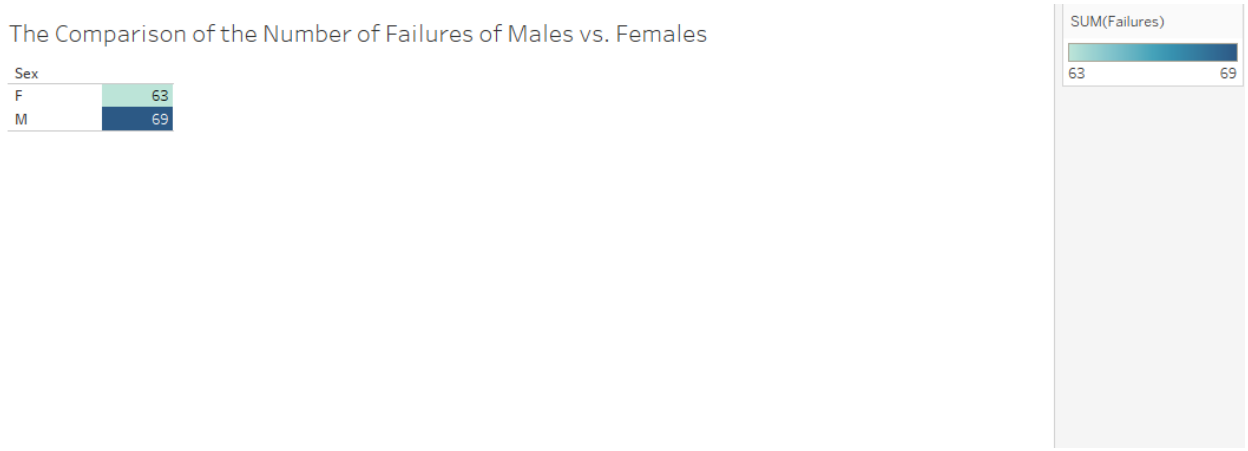
Project available via Tableau Public: https://public.tableau.com/views/MBinkley-Hopper_FinalProject/Goal1?:language=en-US&publish=yes&:sid=&:display_count=n&:origin=viz_share_link

Project available via GitHub: https://github.com/Meghaan-Binkley-Hopper/DataVisualization-FinalProject.git

Section 6. Visualizations and Stories

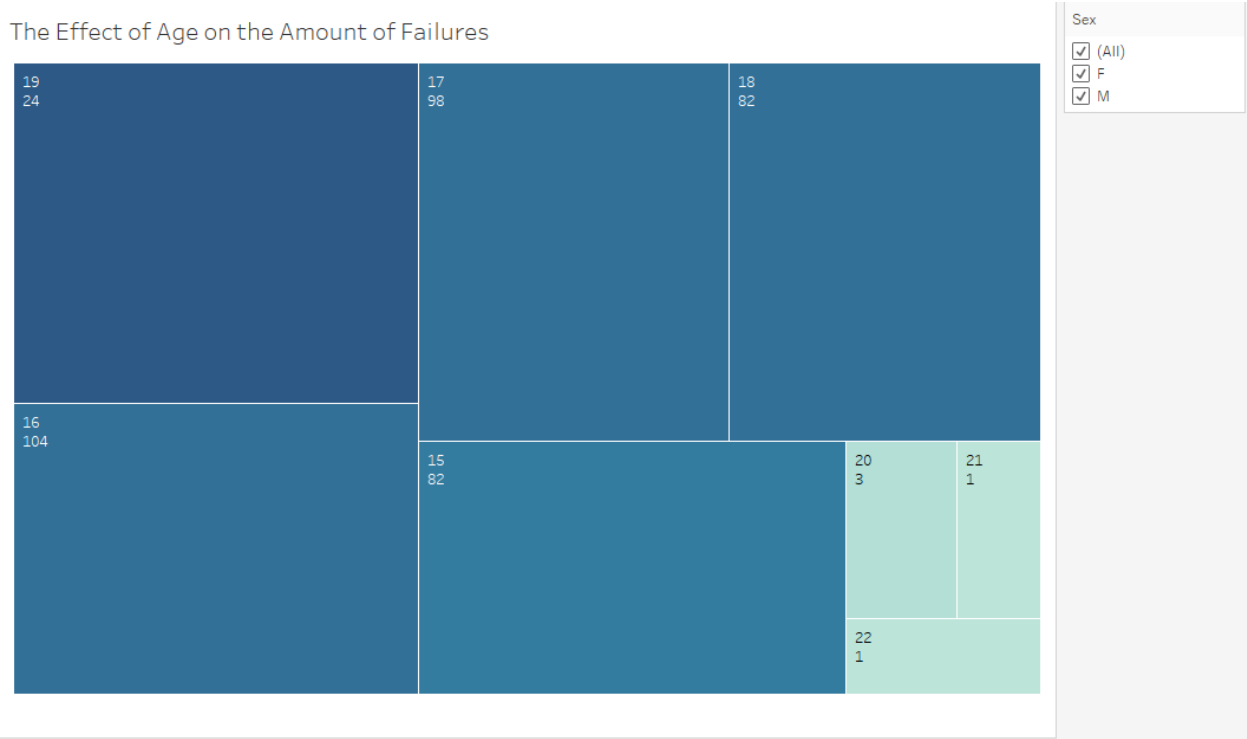
Goal #1: Compare the student's sex to the number of failures

Story: In this dataset, there are 395 students enrolled in a class. Amongst the female students, there were 63 total failures, and, amongst the male students, there were 69 total failures.



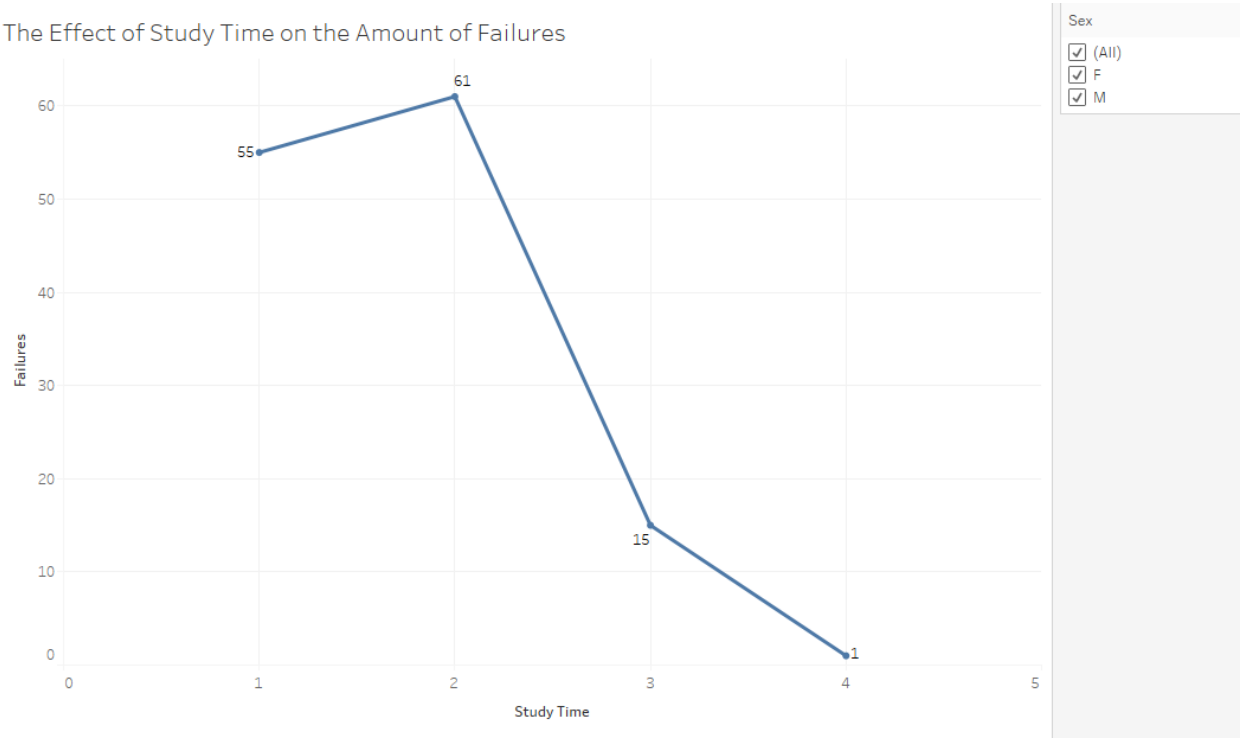
Goal #2: Compare the student's age to the number of failures

Story: This treemap shows that, except for a few outliers, the amount of failures based on age stayed consistent.



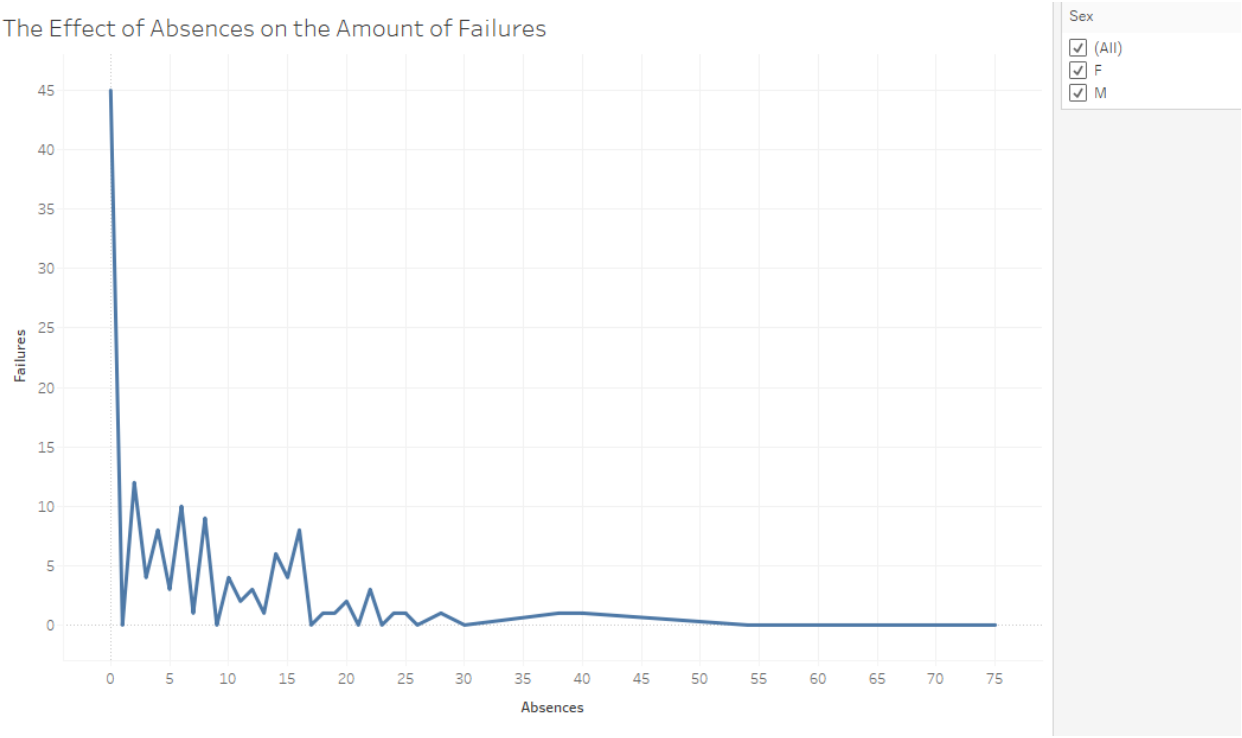
Goal #3: Compare the amount of study time to number of failures

Story: Study time played a big role in how many failures students had. Students who studied 2 or less hours were a lot more likely to fail than students who studied 3-4 hours.



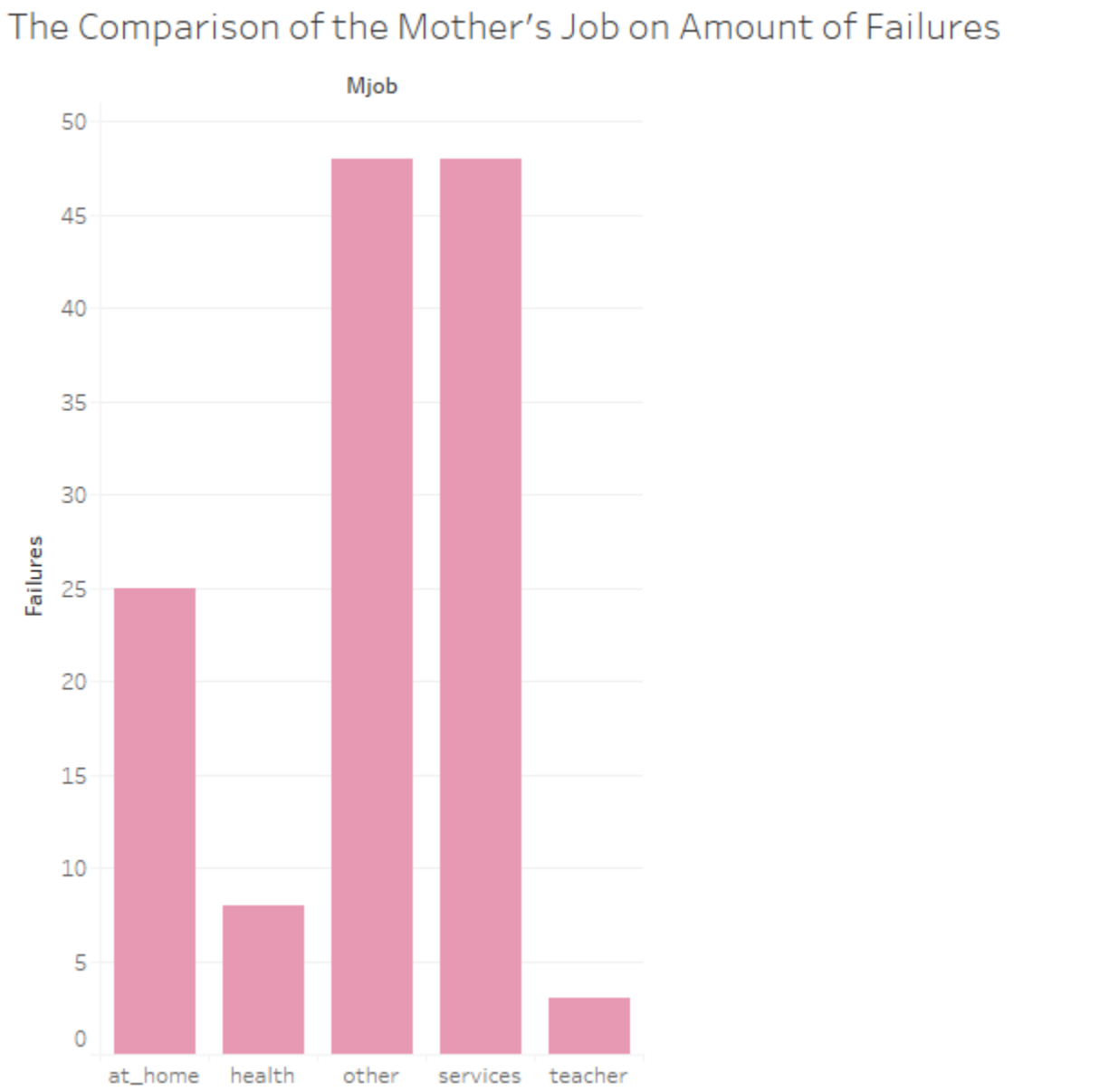
Goal #4: Compare the class absences to number of failures

Story: This chart shows there was no correlation between absences and failures.



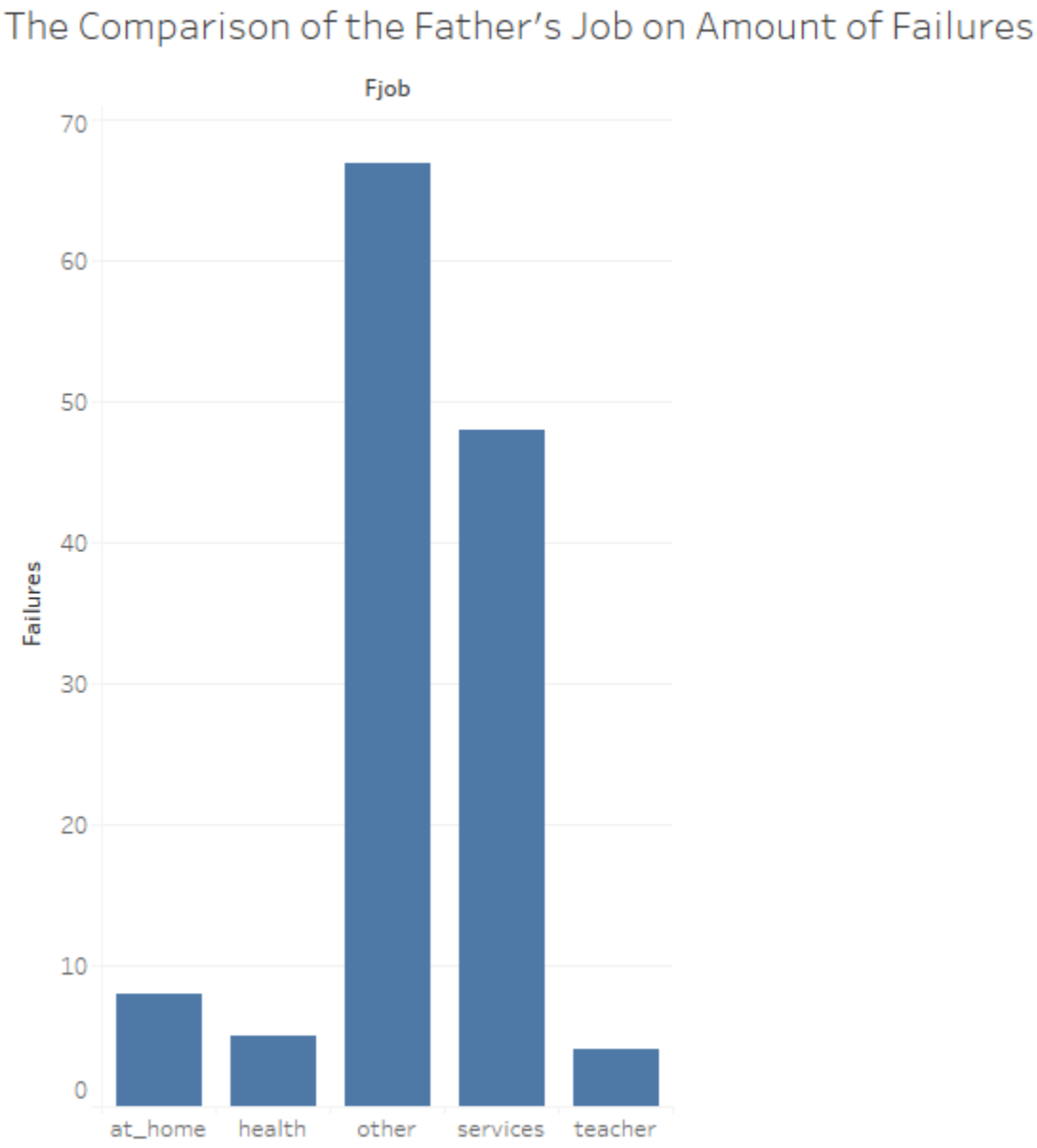
Goal #5: Compare mother's job to number of failures

Story: This chart demonstrates that a mother's job impacted the likelihood that a student failed.



Goal #6: Compare father's job to number of failures

Story: This chart demonstrates that a father's job impacted the likelihood that a student failed.



Section 7. Conclusions

After completing this analysis, it is apparent that the amount of study time and the parents' jobs have the most impact on how many failures a student had in their class while age, sex, and absences had little impact. Overall, there were 132 failures on exams in this class. There were 395 total students, and 83 of those students had at least 1 failure.

After analyzing the data, it is clear that students who studied 3 to 4 hours for their exams were less likely to fail. Out of the 162 failures, 55 of them only studied for 1 hour, and 61 of them studied for only 2 hours. Students who studied 3 hours only had a total of 15 failures, and students who studied for 4 hours had only 1 failure. These results show that the more prepared the students are for the exam, the more likely they are to pass the exam.

The other factor that had strong correlation to the number of failures that students had were their parents' jobs. The analysis shows that students whose parents had jobs in either the services or other category were more likely to receive failing grades. The top 2 job categories for the parents of students with the least amount of failures were teacher and health. This indicates that students who had parents whose jobs required more education were more likely to succeed in their classes.

In conclusion, key indicators for the success of a student are the amount of time that they study and their parents' profession. Absences, sex, and age are not good indicators of a student's likelihood to fail.