Short Story

To give some context, this dataset was taken from Kaggle (<https://www.kaggle.com/datasets/rkiattisak/student-performance-in-mathematics>) and stores student grades in math, writing, and reading categories. This data is a summary of student performances from three different American high schools. In the data set, each student is listed with their gender, ethnicity, parental background in education, what type of meal they ate, and whether they prepared for their tests.

For myself, I’m a student that performs relatively well in their university studies. However, I feel that I could be doing better grades-wise and I’m at times inconsistent with keeping my grades very high. I hope that this dataset can reveal to me what factors may be related to good academic performance.

To process the data, I separated the students in three categories: one was the whole student body, the next was the group of “A Students”, where all of their grades in each section was over 80%, and the last was the “D Students” group, where all students in this group scored lower than 60% in each category. I next examined every potential independent variable in this dataset, where I explored the distributions in each category (gender, ethnicity, parental background, etc.) in the groups for A students, D students, and all students. This method could give an indication to whether each independent variable could impact grades.

Now here are the results:

Gender distribution:

A blue and orange circle with text

Description automatically generated A blue and orange circle with text

Description automatically generated A blue and orange circle with text

Description automatically generated

Ethnicity distribution:

A colorful circle with text on it

Description automatically generated A colorful circle with text

Description automatically generated A colorful circle with text

Description automatically generated

Parental background distribution:

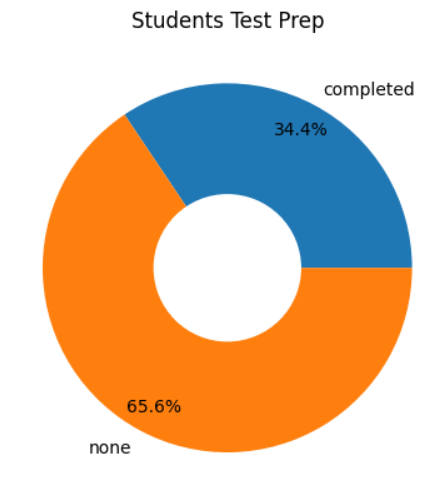
A pie chart with text on it

Description automatically generated A pie chart with text on it

Description automatically generated A pie chart with text on it

Description automatically generated

Test preparation distribution:

 A blue and orange circle with text

Description automatically generated A pie chart with text on it

Description automatically generated

Lunch distribution:

A pie chart with text and numbers

Description automatically generated A pie chart with text on it

Description automatically generated A blue and orange circle with text

Description automatically generated

These distributions have some interesting revelations. For instance, it’s clear that successful students tend to not reduce the quality of their lunch, are more likely to prepare for tests, have parents who are likely to have taken a degree in post-secondary education, and and are more likely to be female and in the D and E ethnicity groups.

While many of these factors (gender, ethnicity, background) are outside of my control, I can learn from this dataset to make sure I do the necessary test prep and also eat a proper meal before a test.

In the end, there are definitely more factors outside of these listed ones in this dataset. It could’ve been interesting to look at performance in other subjects or look into how the number of hours of sleep before a test impact performances. We could also change the approach to this dataset as well. While pie charts was one way to make a conclusion about student performances. We could instead compare average grades between different groups. We could also set a tighter range or better define the grade boundaries instead of just examining a “D” or “A” student.

All in all, this dataset does reveal that while some grades could be outside of your control, it is still important to do the necessary test preparation and eat well beforehand.