Challenge 1

Ouchen Oussama

Part I
Data Exploration & Data Visualization Design

1 Key Visualizations

The dataset contains a variety of information, including personal details, academic performance, and demographic information. To make the most of this data, I will select the following key elements to visualize:

1. Age Distribution: A histogram or bar chart would be an appropriate choice to visualize the age distribution of the students. This can provide insights into the age range and potential trends within the student population.
2. Academic Performance: The dataset includes several academic performance metrics, such as math score, arabic score, and scores in the first language. I will create a line chart to display these performance indicators. This can help identify strengths and weaknesses in different academic areas, as well as patterns in academic performance.
3. Demographic Composition: The data includes information about whether students are interns or externs. A pie chart would be useful to visualize this aspect.
4. Family Assistance Impact: The dataset includes details about the students' family's interference with studies (help or no help). A simple bar chart could be used to visualize these relationships and identify any patterns or correlations.

2 Color Scheme and Insights Extraction

For the color scheme, I will use a combination of blues, greens, and oranges to maintain a professional and easy-to-interpret visual style. The size of the charts and graphs will be adjusted to ensure optimal readability and balance within the overall visualization layout.
By creating these visualizations, I aim to extract the following insights:

1. Understand the age distribution of the student population and identify any potential trends or outliers.
2. Analyze the academic performance of students across different subjects and identify areas of strength and weakness.
3. Explore the demographic composition of the student body and identify any significant imbalances or diversity within the population.
4. Investigate the relationship between family assistance and academic performance, which could inform future interventions or support programs.

These insights can be valuable for the school administration, educators, and policymakers in understanding the student population and making informed decisions to improve educational outcomes.

Part II
Coding and Deployment

3 Data Preparation

To prepare the data for analysis, I first removed **null values** to ensure data integrity. Next, I standardized the text by unifying **case sensitivity** (e.g., converting entries like "oui" and "Oui" or "non" and "Non" to a consistent format). Additionally, I translated Arabic entries into French for uniformity in **language**. I also trimmed column names to remove **unnecessary spaces**, enhancing compatibility with subsequent processing. Finally, I converted the cleaned data from Excel format (.xlsx) to **CSV**, as D3.js requires CSV files for visualization.

4 Interactive Data Visualization with D3.js

In this section, we will outline the process of building the visualizations using D3.js using the JavaScript, HTML, and CSS code necessary to create the interactive visualizations described earlier. Each visualization will use D3.js's powerful data manipulation and rendering capabilities to bring the dataset to life.

JavaScript Code:

The JavaScript code has been divided into several distinct parts (functions), each serving a specific purpose in the analysis and visualization of the dataset :

```
1   d3.csv("dataset.csv").then(function(data) {
2
3   >   const processAgeData = (data) => {…
10      };
11
12  >   const createAgeDistributionChart = (data) => {…
74      };
75
76  >   const createAcademicScoresChart = (data) => {…
136     };
137
138 >   const createPieChart = (data) => {…
202     };
203
204 >   const createFamilyImpactChart = (data) => {…
292     };
293
294     processAgeData(data);
295     createAgeDistributionChart(data);
296     createAcademicScoresChart(data);
297     createPieChart(data);
298     createFamilyImpactChart(data);
299 });
```

The resulting charts are as follows :

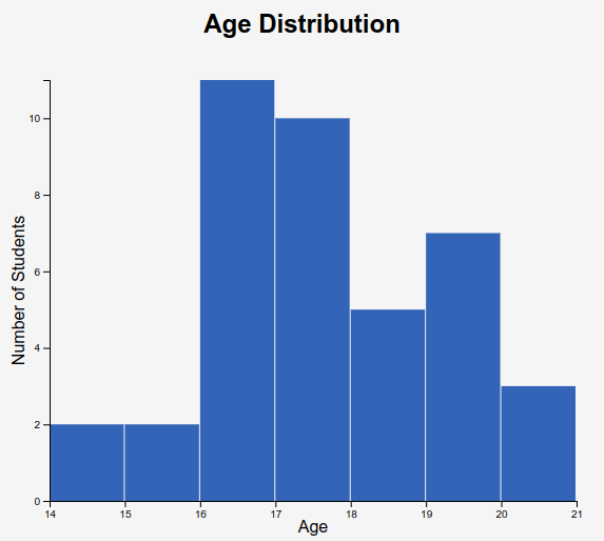

Figure 1: Age Distribution Histogram

The age distribution indicates that most students are between 16 and 18 years old, with a noticeable decline in the number of students above 18. This suggests the dataset is focused on a younger population, likely high school students, which may affect the interpretation of academic performance trends.

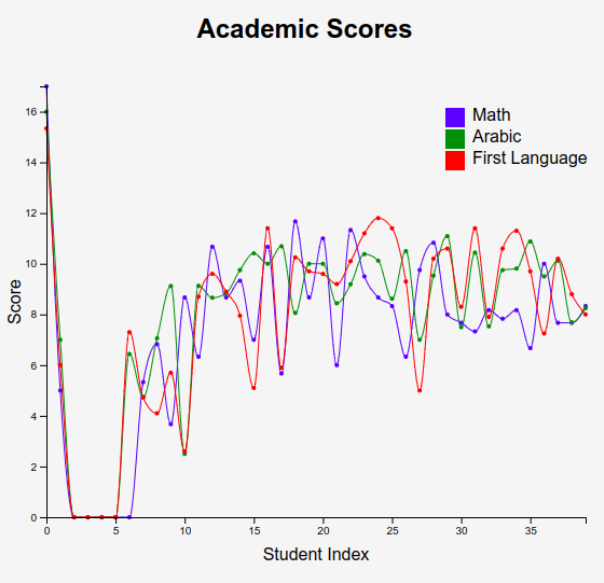

Figure 2: Academic Performance Line Chart

The comparison of scores in Math, Arabic, and the first language shows fluctuations across students. While scores generally align, there are occasional divergences where some subjects outperform others. This could point to varying levels of difficulty or teaching effectiveness in specific subjects.

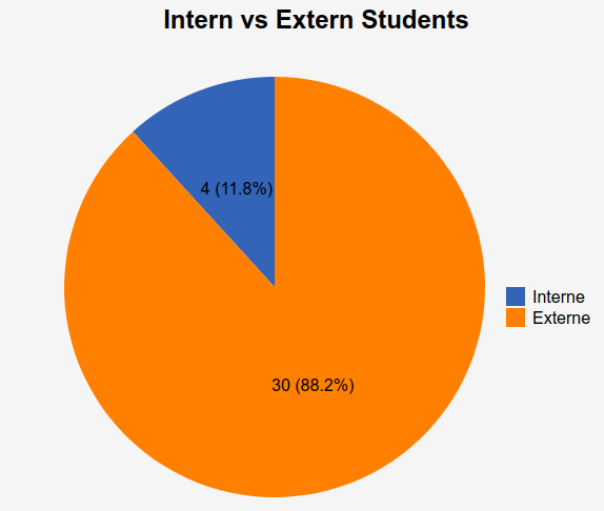

Figure 3: Externe v. Interne Students Pie Chart

Most students are externs (88.2%), while a smaller proportion (11.8%) are interns. This distribution might indicate that external students dominate the sample, and their experiences or performance may disproportionately influence the data trends.

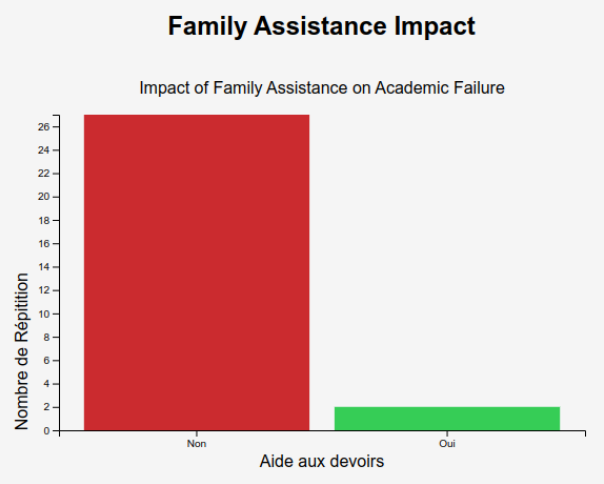

Figure 4: Family Assistance Impact Bar Chart

The chart shows that the majority of students who do not receive help with their homework ("Non") have a higher rate of academic failure compared to those who receive assistance ("Oui"). This suggests that family support can have a positive impact on academic success.

5 Deployment to ObservableHQ

After creating the visualizations, they will be deployed to ObservableHQ, an interactive platform for sharing and exploring data visualizations. ObservableHQ enables seamless integration of JavaScript and D3.js code, allowing the visualizations to remain interactive and easily shareable.

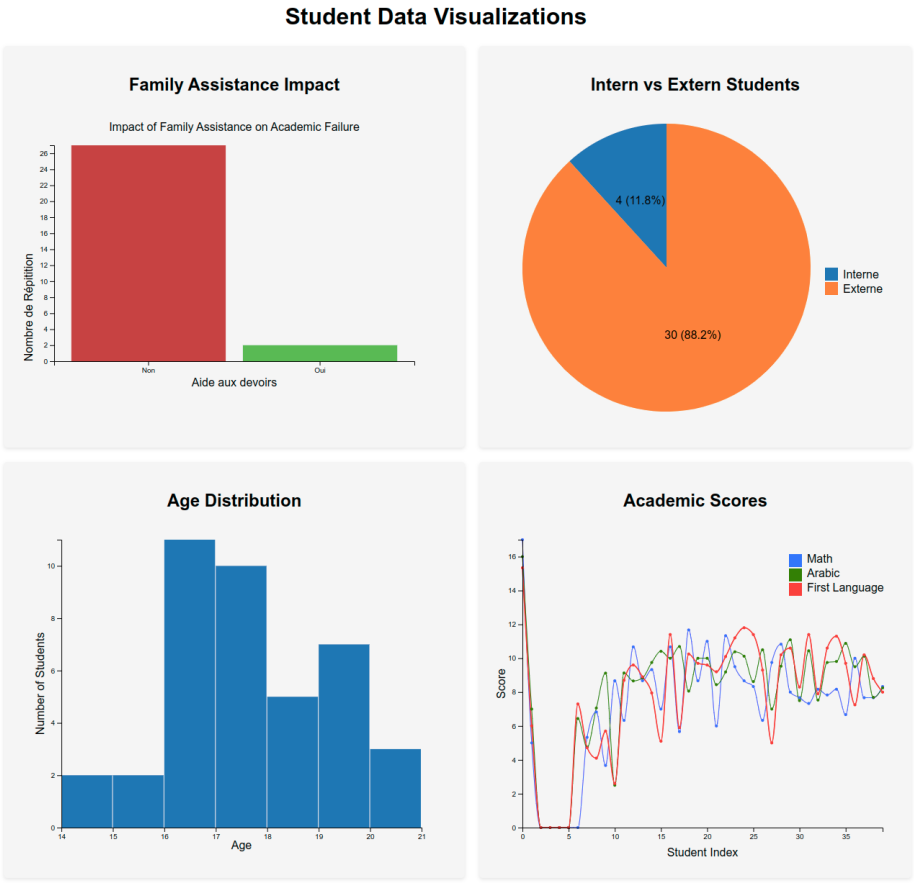

Figure 5: Final Dashboard

List of Figures

Contents