Name:	Khushi Bharat Jain
UID:	2021700030
Experiment:	10

Design Big Data Dashboards using Tableau on Education Sector Dataset

Aim

To design interactive Big Data dashboards in Tableau using datasets from the Education sector, focusing on enrollment rates, academic performance, funding, and demographics, among other factors. This project aims to reveal insights on trends, disparities, and key metrics in education.

Objectives

1. Create basic and advanced charts in Tableau to visualize data relevant to the Education sector.

2. Develop dashboards that convey meaningful insights on various education metrics.

3. Derive observations and trends from each visualization to understand the educational landscape better.

4. Document findings for each visualization and provide a basis for future research or recommendations.

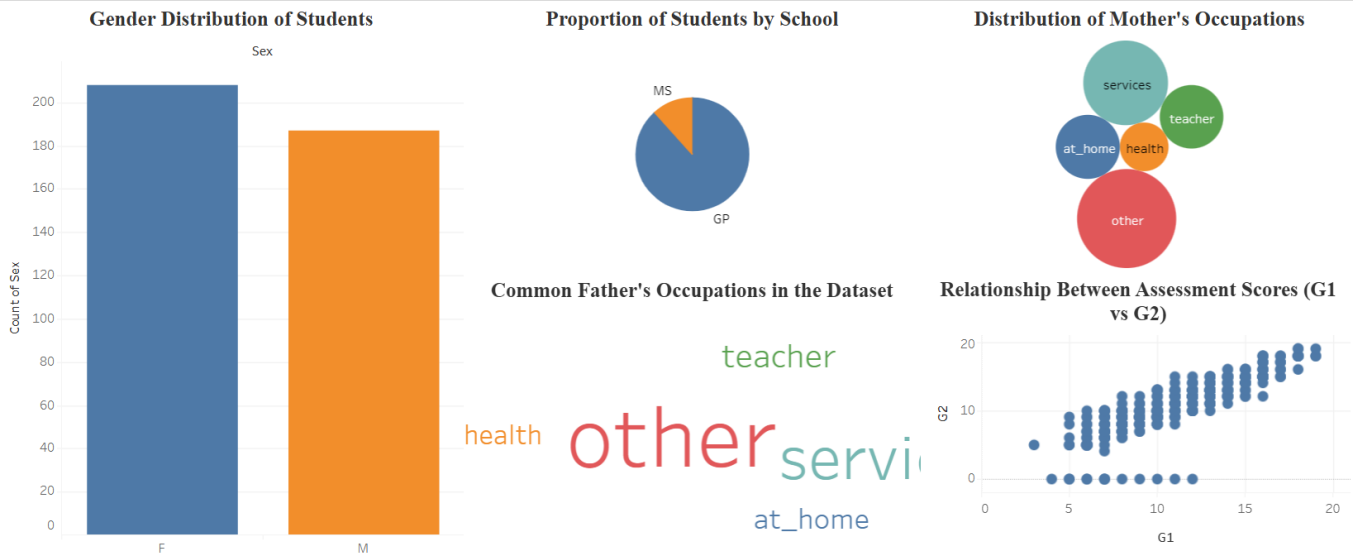
Dataset Link:

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

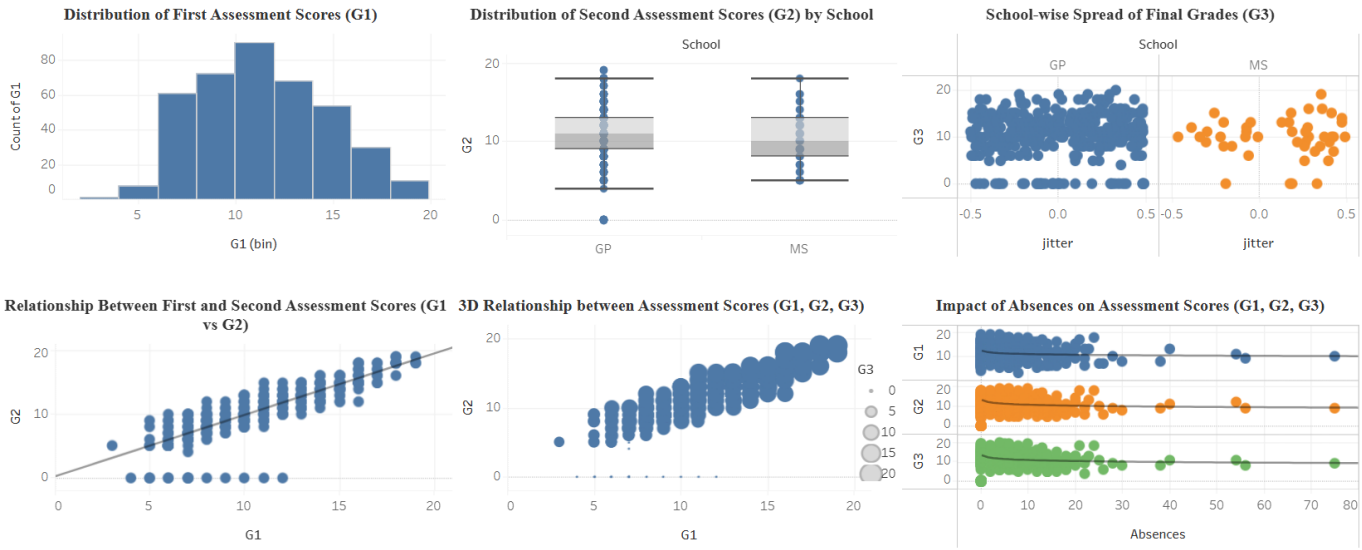
Tableau Public Link:

https://public.tableau.com/app/profile/khushi.jain4331/viz/adv_exp10/Story1

Dashboard 1:

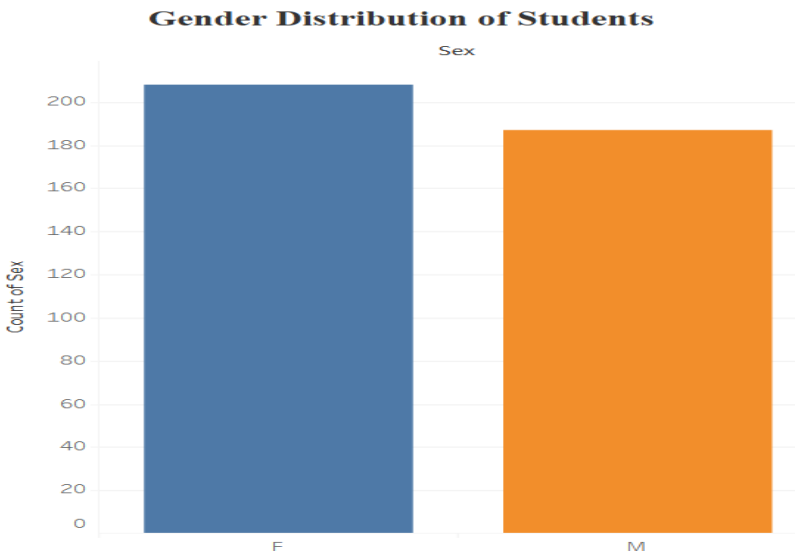


Dashboard 2:



Analysis:

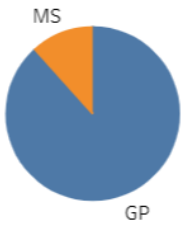
Bar Chart:



This bar chart compares the number of male and female students. It shows that there are slightly more female students than male students in this dataset.

This distribution indicates that the sample or school environment has a balanced or slightly higher representation of female students.

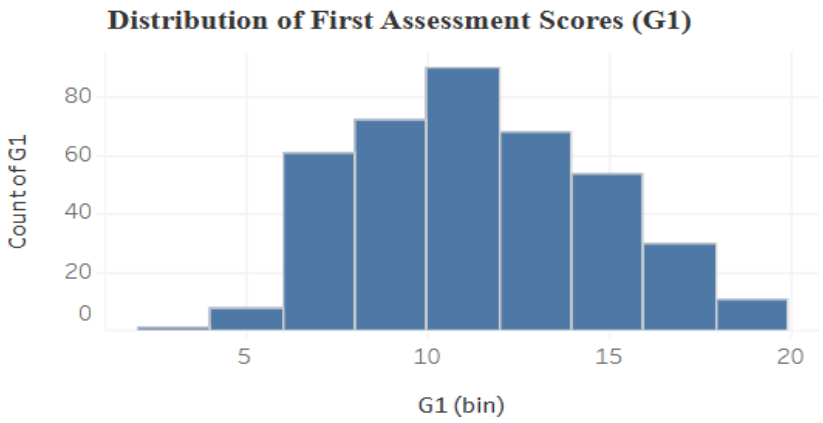
Pie Chart:



The majority of students attend school GP, with a smaller percentage attending school MS.

The significant difference in school representation may impact comparative analysis between the schools. With most data coming from one school, generalizations about the entire student population might lean more towards the characteristics of school GP.

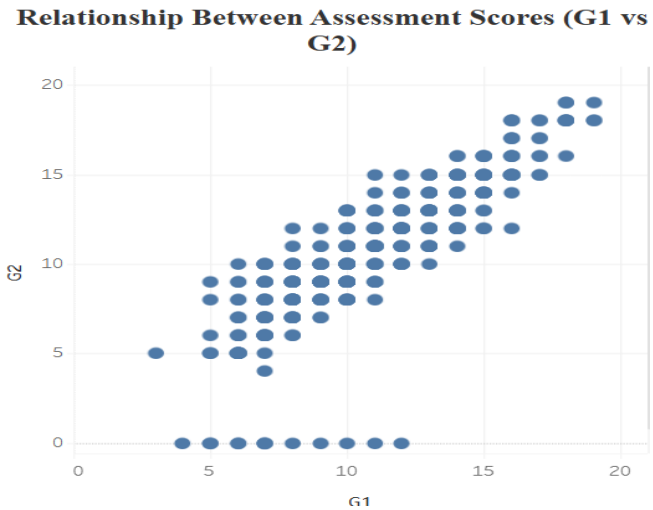
Histogram:



This histogram shows the distribution of scores for the first assessment (G1). Most students scored between 8 and 14, with a peak around 10-12.

The distribution suggests that the majority of students performed moderately well in the first assessment, with fewer students scoring at the extremes (low or high). This may indicate an average level of difficulty for the assessment, with room for improvement among the students.

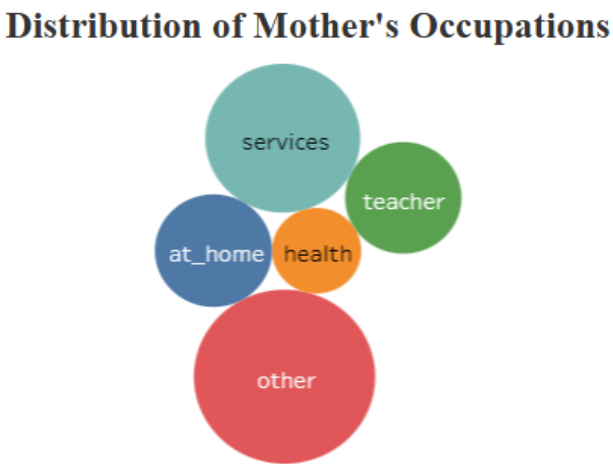
Scatter Plot:



This scatter plot displays the relationship between the first (G1) and second (G2) assessment scores. There is a positive trend, with higher G1 scores generally associated with higher G2 scores.

The positive correlation suggests that students who performed well in the first assessment tended to maintain or improve their performance in the second assessment. This might indicate consistency in students' performance or effectiveness of teaching methods over time.

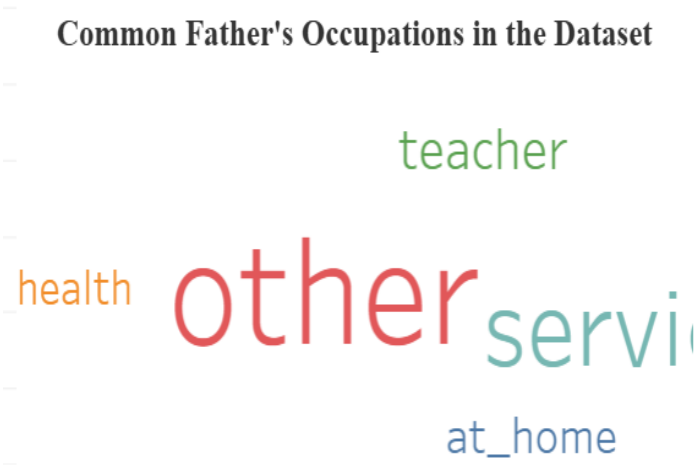
Bubble Plot:



This bubble chart displays the distribution of mothers' occupations, with bubble sizes representing the frequency of each occupation. "Other" is again the largest, followed by "services" and "teacher."

Similar to the father's occupation chart, this visualization highlights that mothers in this dataset are employed in a variety of jobs. This variety could be used to analyze if or how parental occupation impacts student performance.

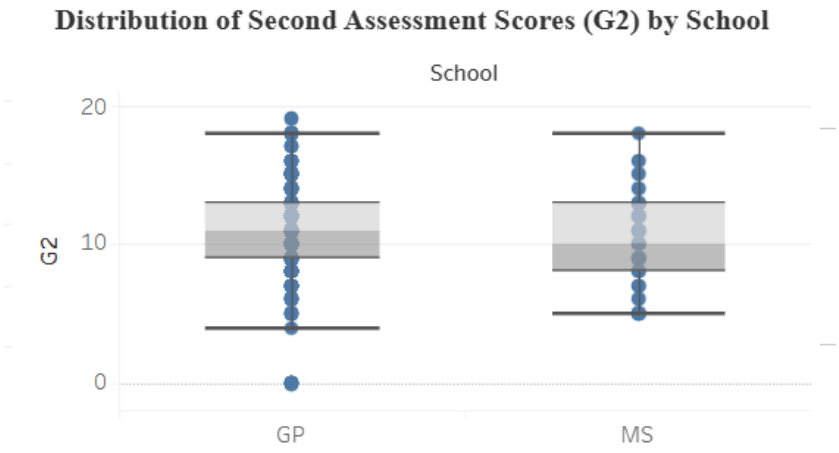
Word Chart:



"Other" is the most common occupation, followed by "service" and "teacher."

The diversity in parental occupations may suggest varying socio-economic backgrounds among the students.

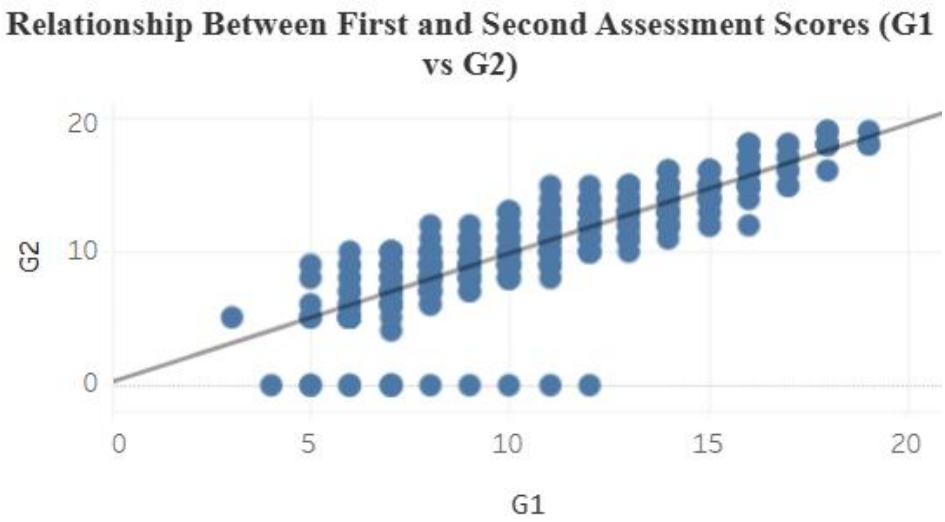
Box and Whisker Plot:



The box plot compares the distribution of second assessment scores (G2) between the two schools, GP and MS. Both schools appear to have a similar distribution.

There is little difference between the schools in terms of G2 scores, which may indicate consistent teaching or assessment standards.

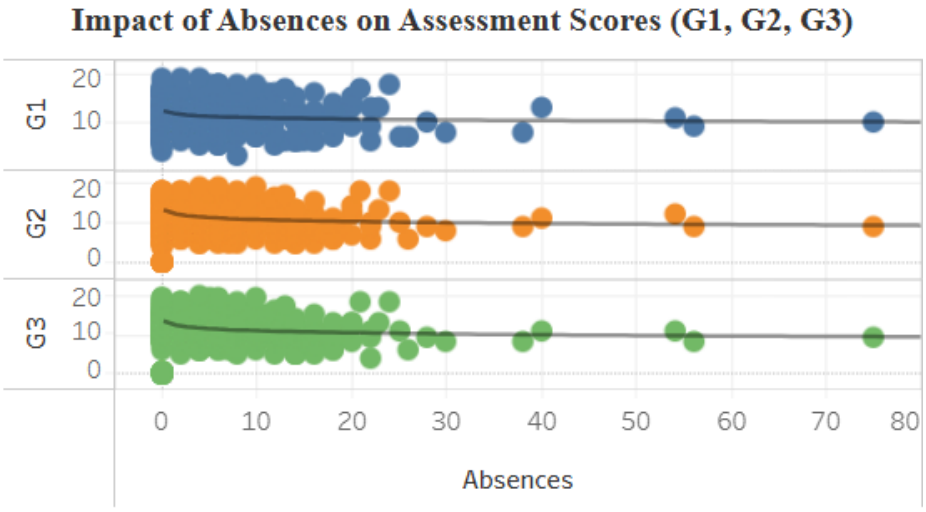
Linear Regression Plot:



There is a positive trend, with higher G1 scores generally associated with higher G2 scores.

The positive correlation suggests that students who performed well in the first assessment tended to maintain or improve their performance in the second assessment. This might indicate consistency in students' performance or effectiveness of teaching methods over time.

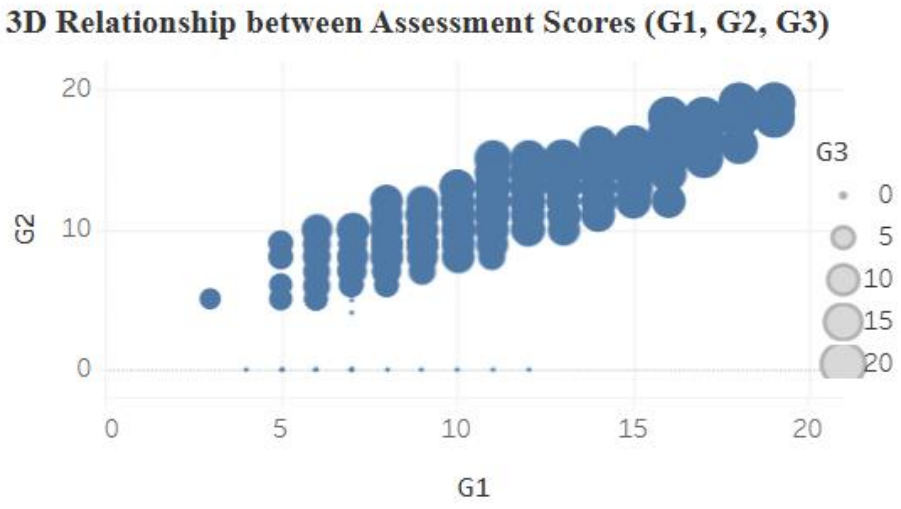
Non Linear Regression Plot:



For all three assessment scores (G1, G2, G3), there appears to be a slightly negative trend as absences increase. This suggests that higher absence rates may correlate with lower assessment scores, though the effect is not very strong for lower levels of absences.

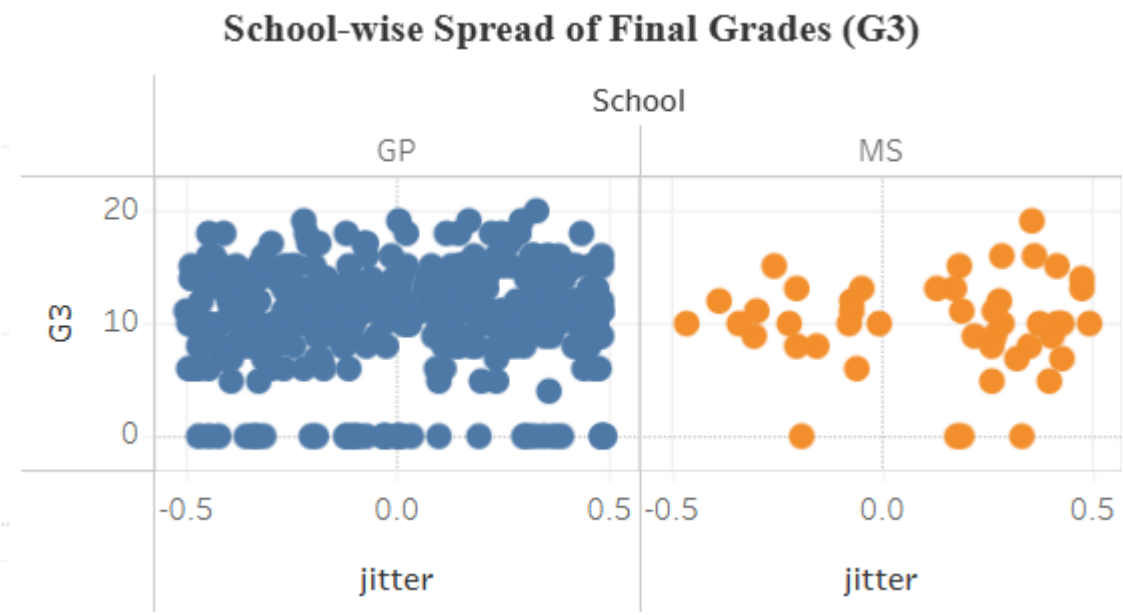
The trend lines indicate a gradual decrease in scores.

3d Chart:



The 3D plot helps in visualizing trends across all three assessments. It appears that students with higher G1 and G2 scores also tend to have higher G3 scores. This cumulative trend suggests that consistent academic performance across assessments correlates with better final grades.

Jitter Plot:



This jitter plot shows the spread of final grades (G3) across the two schools, GP and MS.

Both have spread out distribution.

Conclusion

The lab exercise demonstrates the effectiveness of using Tableau for educational data visualization. The visualizations allow insights into trends, disparities, and areas of improvement, facilitating a comprehensive understanding of the educational landscape.