Student Performance Analysis

I. Introduction

Students are the main asset for various Universities. Universities and students play an important role in producing graduates of high qualities with its academic performance achievement. Academic performance achievement is the level of achievement of the students' educational goal that can be measured and tested through examination, assessments and other form of measurements. However, the academic performance achievement varies as different kind of students may have different level of performance achievement.

The student academic performance is usually stored in student management system, in different formats such as files, document, records, images and other formats. These available students' data could be extracted to produce useful information. However, the increasing amount of students' data becomes hard to be analysed by using traditional statistic techniques and database management tools. Thus, a tool is necessary for universities to extract the useful information. This useful information could be useful to predict the students' performance.

Student performance analysis and prediction using datasets has become an essential component of modern

Education systems. With the increasing availability of data on student demographics, academic history, and other relevant factors, schools and universities are using advanced analytics and machine learning algorithms to gain insights into student performance and predict future outcomes. This approach helps educators identify areas of improvement, personalize learning experiences, and provide targeted support to struggling students. Furthermore, student performance analysis and prediction can also aid in decision-making processes for school administrators and policymakers, helping them allocate resources more effectively. In this article, we will explore the benefits of using datasets for student performance analysis and prediction and discuss some of the methods and tools used in this field.

→ Parental Level Education:-

Students whose parents stay involved in school have better attendance and behaviour, get better grades, demonstrate better social skills and adapt better to school.

Parental involvement is the active, ongoing participation of a parent or primary caregiver in the education of a child.

Parents can demonstrate involvement at home by:

- reading with children;
- helping with homework;
- discussing school events;
- attending school functions, including parent-teacher meetings; and
- volunteering in classrooms.

"Involvement" is the first step towards engagement. It includes participation in school events or activities, with teachers providing learning resources and information about their student's grades. With involvement, teachers hold the primary responsibility to set educational goals.

With engagement, home and school come together as a team. Schools empower parents and caregivers by providing them with ways to actively participate, promoting them as important voices in the school and removing barriers to engagement. Examples include encouraging families to join the family-teacher association or arranging virtual family-teacher meetings for families with transporation issues.

→ Exploratory Data Analysis:-

The data set consists of the marks secured in various subjects by high school students from the United states, which is accessible from kaggle Student performance in Exams. There are 1000 occurences and 8 columns:

- gender
- race / ethnicity
- Parental level of education
- lunch
- test preparation course
- math score
- reading score
- Writing score

We will be checking out the performance of the class in each subject, the effect of Parent level of education on the student performance, and also the relationship between Grender and student performance.

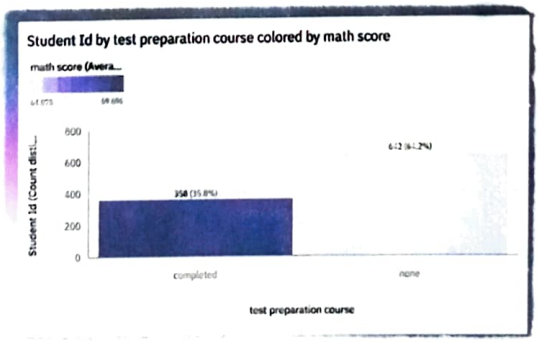
- Grender:- Avoid segregating boys and girls into separate lines, separate sports activities and mix seating up in the classroom.

-: Visualizations :-

→ Stacked Bar chart:-

A stacked bar chart is a type of bar graph that represents the proportional contribution of individual data points in comparison to a total. The height or length of each bar represents how much each group contributes to the total.

A stacked bar chart should be used when there is a need to see the different parts that make up each bar of a bar graph. A stacked bar chart is used to show the total or average of each category. Stacked Bar chart are best used when showing comparisons between categories.

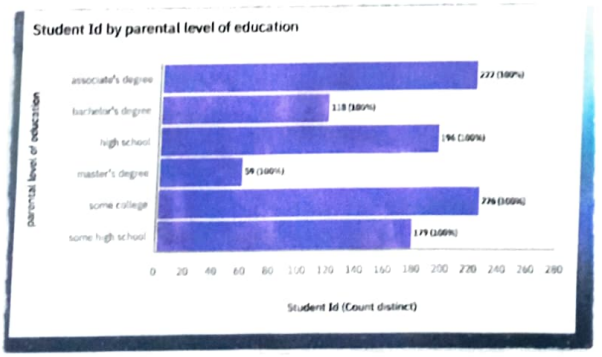


- In this chart we can see that the student ID by race / ethnicity is colored by gender.

→ Column chart:

A column chart is a data visualization where each category is represented by a rectangle, with the height of the rectangle being proportional to the values being plotted. Column charts are also known as vertical bar charts. They allow easy comparisons among a number of items and trends analysis.

Column charts are useful for showing data changes over a period of time or for illustrating comparisons among items. In column charts, categories are typically organized along the horizontal axis and values along the vertical axis.

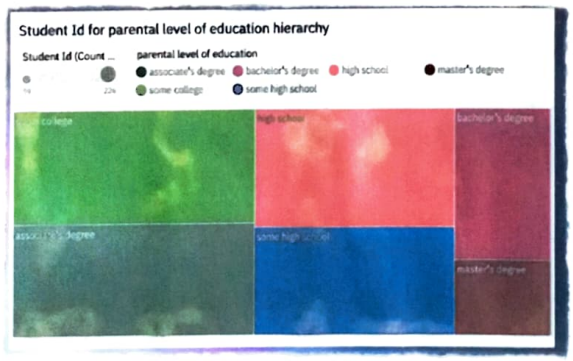


- In this chart we can see that the student In test preparation course is colored by math score.

→ Hierarchy treemap:-

A treemap charts provide a hierarchy (hierarchial) view of your data and makes it easy to spot patterns, such as which items are a store's best sellers. The tree branches are represented by rectangles and each sub-branch is shown as a smaller rectangle.

The treemap functions as a visualization composed of nested rectangles. These rectangles represent certain categories within a selected dimension and are ordered in a "hierarchy" or "tree". Quantities and patterns can be compared and displayed in a limited chart space. Tree maps represent part to whole relationships.

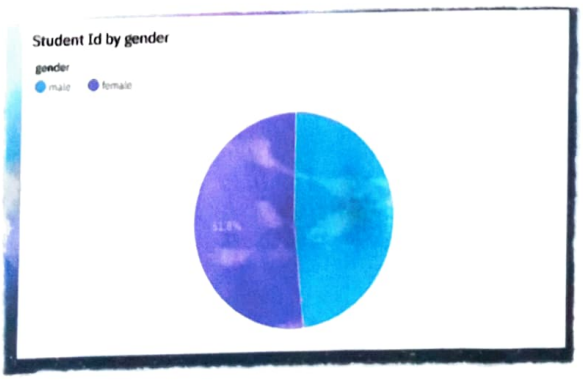


- In this chart we can see that student Id for parental level of education.

→ Pie chart:-

A pie chart, is a circular statistical graphic which is divided into slices to illustrate numerical proportion. A pie chart, sometimes called a circle chart, is a way of summarizing a set of nominal data or displaying the different values of a given variable. (eg: percentage distribution). This type of chart is a circle divided into a series of segments. Each segment represents a particular category.

A pie chart helps organize and show data as a percentage of a whole. True to the name, this kind of visualization uses a circle to represent the whole, and slices of that circle, or "pie", to represent the specific categories that compose the whole.

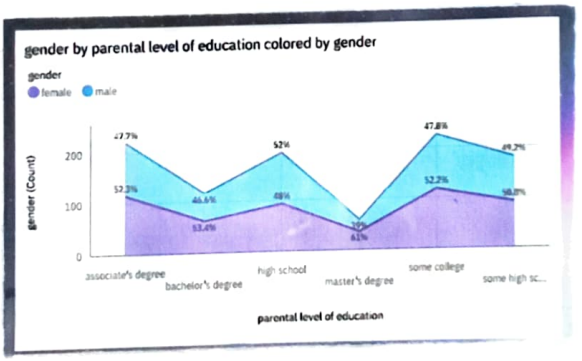


- In this pie chart we can see that student Id by gender.

→ Area chart:-

An area chart combines the line chart and bar chart to show how one or more groups numeric values change over the progression of a second variable, typically that of time. An area chart is distinguished from a line chart by the addition of shading between lines and a baseline, like in a bar chart.

Area charts are used to represent cumulated totals using numbers or percentages (stacked area charts in this case) over time. Use the area chart for showing trends over time among related attributes.

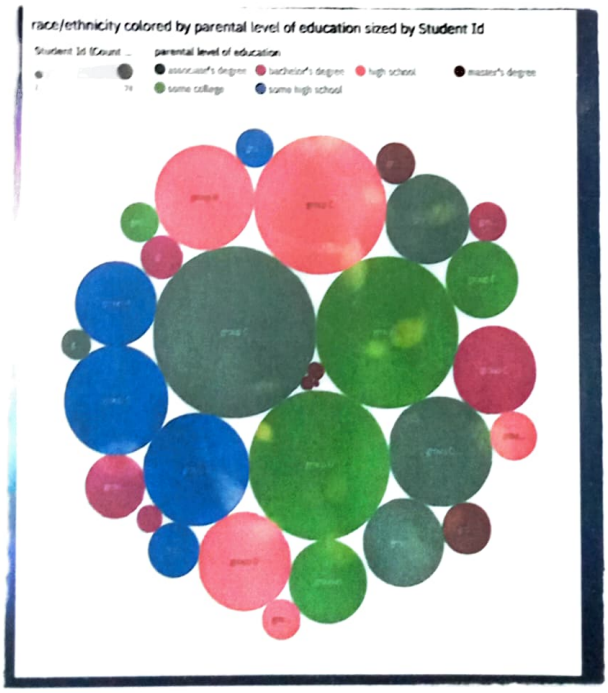


- In this we can see that gender by parental level of education is colored by gender.

* Hierarchy Bubble chart:-

A hierarchy bubble visualization shows a large amount of data in a small space. The size of each bubble shows a quantitative dimension of each data point. It shows many levels within a hierarchy and relationships between groups based on assigned attributes.

A bubble chart is primarily used to depict and show relationships between numeric variables. They are a great tool to establish the relationship between variables and examine relationships between key business indicators, such as cost, value and risk.

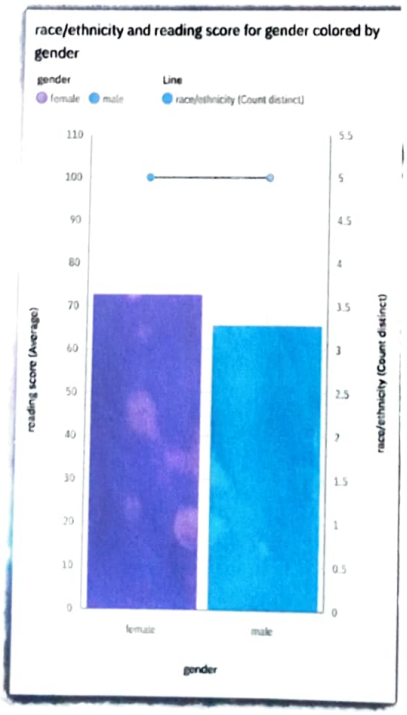


- In this chart we can see that the parental level of education hierarchy is colored by parental level of education and sized by student ID.

→ Line chart:-

A line chart is a type of chart used to show information that changes over time. Line charts are created by plotting a series of several points and connecting them with a straight line. Line charts are used to track changes over short and long periods.

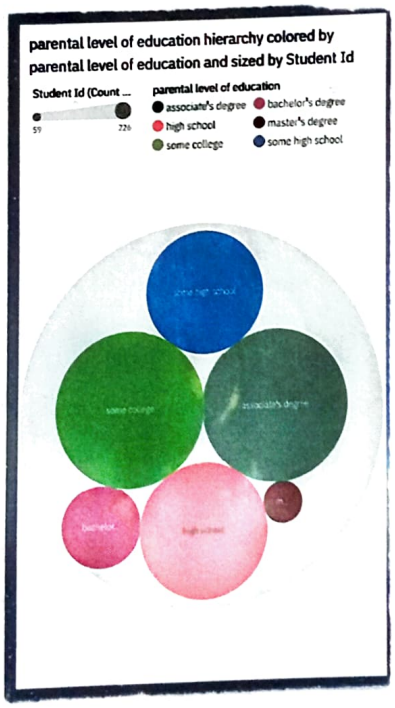
A line chart is used to show the change in information over time. The horizontal axis is usually a time scale. Linear graphs are used to analyze and predict future markets and opportunities.



→ Packed bubble chart:-

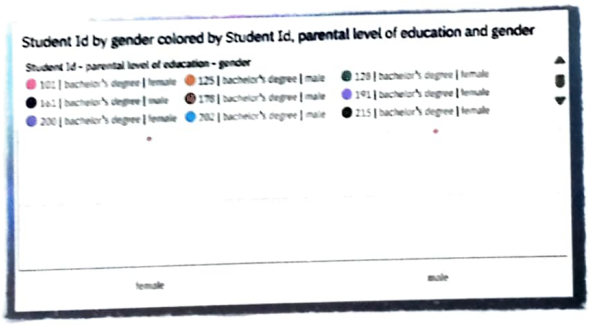
Packed bubble charts are visualizations where the size and optionally the color of the bubbles are used to visualize the data. The positioning of the bubbles is not significant, but is optimized for Compactness.

Packed bubble, also known as circular Treemap or circle packing, is great for visualizing hierarchical data as nested circles of different sizes and colors.

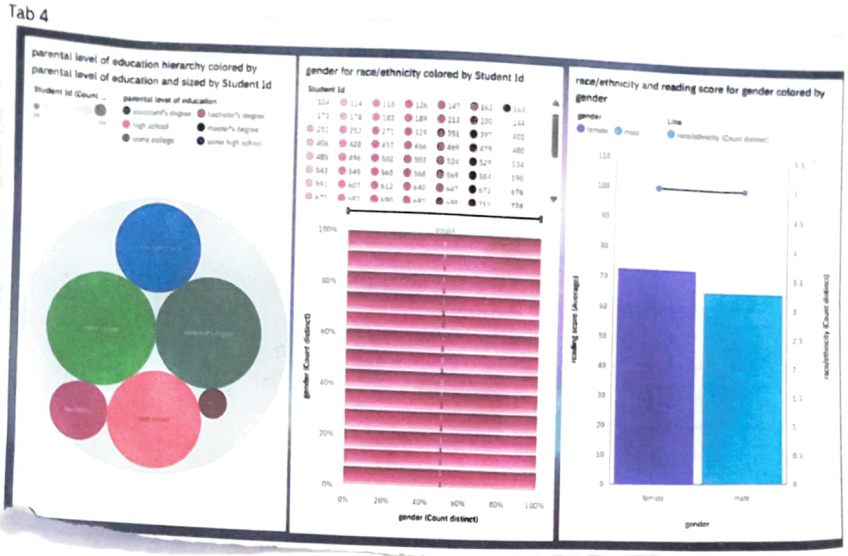


→ Point chart:-

Point and figure is a charting technique used in technical analysis. Point and figure charting does not plot price against time as time-based charts do. Contrary to some other types of charts, like candlesticks, which mark the degree of an asset's movement over set time periods, P&F charts utilize columns consisting of stacked X's or O's, each of which represents a set amount of price movement. The X's illustrate rising prices, while O's represent a falling price.

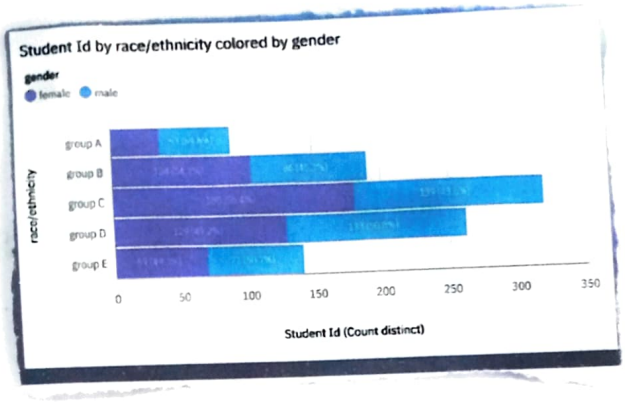


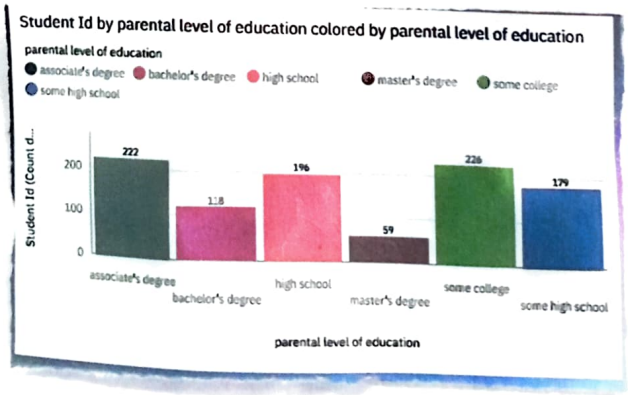
Tab 4


parental level of education hierarchy colored by parental level of education and sized by Student Id
gender for race/ethnicity colored by Student Id
Student Id
gender (Count distinct)
race/ethnicity and reading score for gender colored by gender
gender
Line
reading score (Average)
race/ethnicity (Count distinct)
female
male

Tab 2


Student Id and gender
Student Id
Student Id
gender
101 female
102 female
103 female
104 male
105 male
106 female
107 female
108 male
109 male
110 female
111 male
112 male
113 female
114 male
115 female
116 female
117 male
118 female
119 male
Student Id for parental level of education hierarchy
Student Id by gender colored by Student Id, parental level of education and gender
female
male

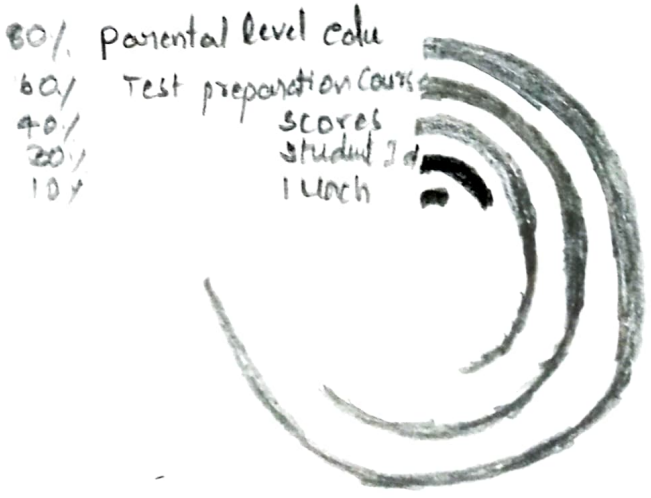

Student Id by race/ethnicity colored by gender
gender
female
male
group A
group B
group C
group D
group E
race/ethnicity
0
50
100
150
200
250
300
350
Student Id (Count distinct)


Student Id by parental level of education colored by parental level of education
parental level of education
associate's degree
bachelor's degree
high school
master's degree
some college
some high school
Student Id (Count d...
200
100
0
222
118
196
59
226
179
associate's degree
bachelor's degree
high school
master's degree
some college
some high school
parental level of education

→ Radial bar chart:-

A Radial / Circular Bar chart simply refers to a typical Bar chart displayed on a polar coordinate system, instead of a cartesian system. It is used to show comparisons among categories by using a circular shape.

A Radial chart (also called a circular bar chart) uses circular shapes to compare different categories. Radial charts are essentially bar charts displayed on a polar coordinate system instead of a cartesian system. Radial charts can be used for aesthetic reasons while simple bar charts are better for comparing values.



→ Marimekko chart:

A marimekko chart is a graphical representation that uses stacked bar graphs of varying widths to visualize categorical data. A marimekko chart is also known as the mosaic plot, or simply, Mekko charts. They are ideal for representing categorical sample data.

A mekko chart let's you use the width of the columns to bring in an additional variable. In this example the width of the columns represents the total amount of revenue for each company.

Conclusion

This analysis has taught us that, factors such as Parental level of education, socioeconomic disadvantage, test preparation courses affect the students performances in the exams. But there are many exceptions as well. There are students with a low parental level of education scoring full marks. Also, some students have not completed the test preparation course getting full marks. These students may have their own strategies for test preparations. Socioeconomic disadvantage also has many exceptions. These students did not allow economic obstacles to affect their efforts. So, many factors are affecting the students' performances. Some have great effects while not some. Also, there are other factors to be considered as well which are not mentioned in the data set.

Factors such as facilities at school, the quality and methods of teaching, peer pressures, hours spent studying, diets, sleeping patterns etc. Such factors also affect the performance of the students.