***PROJECT NAME:***

**Unleashing The Potential Of Our Youth: A Student Performance Analysis**

A country's growth is strongly measured by quality of its education system. Education sector, across the globe has witnessed sea change in its functioning. Today it is recognized as an industry and like any other industry it is facing challenges, the major challenges of higher education being decrease in students' success rate and their leaving a course without completion.

Analysing student work is an essential part of teaching. Teachers assign, collect and examine student work all the time to assess student learning and to revise and improve teaching. Ongoing assessment of student learning allows teachers to engage in continuous quality improvement of their courses. Many factors can influence a student's performance, including the influence of the parents' educational background, test preparation and so on.

The dataset contains the marks secured by 1000 students from a school. This project analyses and correlates student performance with different attributes. The analysis aims to understand the influence of important factors such as parental level of education, the status of test preparation course etc. on the performance of the students in the exams.

# **Introduction:**

Data analysis and data visualization are essential components of data science. Actually, before the machine learning era, all data science was about the interpretation and visualization of data with different tools and making conclusions about the nature of data. Nowadays, these tasks are still present. They just became one of many miscellaneous data science jobs. Very often, the so-called EDA (exploratory data analysis) is a required part of the machine learning pipeline. It allows a better understanding of data, its distribution, purity, features, etc. Also, visualization is recommended to present the results of the machine learning work to different stakeholders. They may not be familiar with sophisticated data science principles, but it is convenient for them to look at graphs and charts. Besides, data analysis and visualization can be done as standalone tasks if there is no need to dig deeper into the data. In any case, a good data scientist should know how to analyze and visualize data.

In this ,we will show how to analyze data and how to build nice and informative graphs. We will use popular Python libraries for the visualization, namely matplotlib and seaborn. Also, we will use Pandas as a tool for manipulating dataframes. The dataset we will work with is the [Student Performance Data Set](https://archive.ics.uci.edu/ml/datasets/Student+Performance).

## What Is Data Analytics?

Data analytics is the science of analyzing raw data to make conclusions about information. Many of the techniques and processes of data analytics have been automated into mechanical processes and [algorithms](https://www.investopedia.com/terms/a/algorithm.asp) that work over raw data for human consumption.

### KEY TAKEAWAYS

* Data analytics is the science of analyzing raw data to make conclusions about that information.
* Data analytics help a business optimize its performance, perform more efficiently, maximize profit, or make more strategically-guided decisions.
* The techniques and processes of data analytics have been automated into mechanical processes and algorithms that work over raw data for human consumption.
* Various approaches to data analytics include looking at what happened (descriptive analytics), why something happened (diagnostic analytics), what is going to happen (predictive analytics), or what should be done next (prescriptive analytics).
* Data analytics relies on a variety of software tools including spreadsheets, data visualization, reporting tools, data mining programs, and open-source languages for the greatest data manipulation.

## Understanding Data Analytics

Data analytics is a broad term that encompasses many diverse types of data analysis. Any type of information can be subjected to data analytics techniques to get insight that can be used to improve things. Data analytics techniques can reveal trends and metrics that would otherwise be lost in the mass of information. This information can then be used to optimize processes to increase the overall efficiency of a business or system.

For example, [manufacturing](https://www.investopedia.com/terms/m/manufacturing.asp) companies often record the runtime, downtime, and work queue for various machines and then analyze the data to better plan workloads so the machines operate closer to peak capacity.

Data analytics can do much more than point out [bottlenecks](https://www.investopedia.com/terms/b/bottleneck.asp) in production. Gaming companies use data analytics to set reward schedules for players that keep the majority of players active in the game. Content companies use many of the same data analytics to keep you clicking, watching, or re-organizing content to get another view or another click.

Data analytics is important because it helps businesses optimize their performances. Implementing it into the business model means companies can help reduce costs by identifying more efficient ways of doing business and by storing large amounts of data.

## Data Analysis Steps

The process involved in data analysis involves several steps:

1. The first step is to determine the data requirements or how the data is grouped. Data may be separated by age, demographic, income, or gender. Data values may be numerical or divided by category.
2. The second step in data analytics is the process of collecting it. This can be done through a variety of sources such as computers, online sources, cameras, environmental sources, or through personnel.
3. The data must be organized after it's collected so it can be analyzed. This may take place on a spreadsheet or other form of software that can take statistical data.
4. The data is then cleaned up before analysis. It's scrubbed and checked to ensure that there's no duplication or error and that it is not incomplete. This step helps correct any errors before it goes on to a data analyst to be analyzed.

## Types of Data Analytics

Data analytics is broken down into four basic types:

1. **Descriptive analytics:**This describes what has happened over a given period of time. Have the number of views gone up? Are sales stronger this month than last?
2. **Diagnostic analytics:**This focuses more on why something happened. It involves more diverse data inputs and a bit of hypothesizing. Did the weather affect beer sales? Did that latest marketing campaign impact sales?
3. **Predictive analytics:**This moves to what is likely going to happen in the near term. What happened to sales the last time we had a hot summer? How many weather models predict a hot summer this year?
4. **Prescriptive analytics:** This suggests a course of action. We should add an evening shift to the brewery and rent an additional tank to increase output if the likelihood of a hot summer is measured as an average of these five weather models and the average is above 58%,

Data analytics underpins many quality control systems in the financial world, including the ever-popular [Six Sigma](https://www.investopedia.com/terms/s/six-sigma.asp) program. It's nearly impossible to optimize something if you aren’t properly measuring it, whether it's your weight or the number of defects per million in a production line.

The [sectors](https://www.investopedia.com/terms/s/sector.asp) that have adopted the use of data analytics include the travel and hospitality industry where turnarounds can be quick. This industry can collect customer data and figure out where problems, if any, lie and how to fix them.

Healthcare combines the use of high volumes of structured and unstructured data and uses data analytics to make quick decisions. Similarly, the retail industry uses copious amounts of data to meet the ever-changing demands of shoppers. The information that retailers collect and analyze can help them identify trends, recommend products, and increase profits.

## Data Analytics Techniques

Data analysts can use several analytical methods and techniques to process data and extract information. Some of the most popular methods include:

* [**Regression analysis**](https://www.investopedia.com/terms/r/regression.asp) entails analyzing the relationship between dependent variables to determine how a change in one may affect the change in another.
* [**Factor analysis**](https://www.investopedia.com/terms/r/random-factor-analysis.asp) entails taking a large data set and shrinking it into a smaller data set. The goal of this maneuver is to attempt to discover hidden trends that would otherwise have been more difficult to see.
* **Cohort analysis** is the process of breaking a data set into groups of similar data, often into a customer demographic. This allows data analysts and other users of data analytics to further dive into the numbers relating to a specific subset of data.
* [**Monte Carlo simulations**](https://www.investopedia.com/terms/m/montecarlosimulation.asp) model the probability of different outcomes happening. They're often used for risk mitigation and loss prevention. These simulations incorporate multiple values and variables and often have greater forecasting capabilities than other data analytics approaches.
* **Time series analysis** tracks data over time and solidifies the relationship between the value of a data point and the occurrence of the data point. This data analysis technique is usually used to spot cyclical trends or to project financial forecasts.

## Data Analytics Tools

Data analytics has rapidly evolved in technological capabilities in addition to a broad range of mathematical and statistical approaches to crunching numbers. Data analysts have a broad range of software tools to help acquire data, store information, process data, and report findings.

Data analytics has always had loose ties to spreadsheets and Microsoft Excel. Data analysts also often interact with raw programming languages to transform and manipulate databases.

Data analysts also have help when reporting or communicating findings. Both Tableau and Power BI are data visualization and analysis tools used to compile information, perform data analytics, and distribute results via dashboards and reports.

Other tools are also emerging to assist data analysts. SAS is an analytics platform that can assist with [data mining](https://www.investopedia.com/terms/d/datamining.asp). Apache Spark is an open-source platform useful for processing large sets of data. Data analysts have a broad range of technological capabilities to further enhance the value they deliver to their company.

## The Role of Data Analytics

Data analytics can enhance operations, efficiency, and performance in numerous industries by shining a spotlight on patterns. Implementing these techniques can give companies and businesses a competitive edge. The process includes four basic steps of analysis.

### Data Mining

As the name suggests, this step involves “mining” or gathering data and information from across a broad spectrum of sources. Various forms of information are then recreated into the same format so they can eventually be analyzed. The process can take a good bit of time, more than any other step.

### Data Management

Data requires a database to contain, manage, and provide access to the information gathered through mining. The next step in data analytics is therefore the creation of such a database to manage the information. SQL was a common tool used for this purpose in the early days of data analytics and it's still widely used in 2023. Created in 1979, this computing language allows relational databases to be queried and the resulting data sets to be more easily analyzed.2

### Statistical Analysis

The third step is statistical analysis. It involves the interpretation of the gathered and stored data into models that will hopefully reveal trends that can be used to interpret future data. This is achieved through [open-source](https://www.investopedia.com/terms/o/open-source.asp) programming languages such as Python. More specific tools for data analytics, like R, can be used for statistical analysis or graphical modeling.

### Data Presentation

The results of the data analytics process are meant to be shared. The final step is formatting the data so it’s accessible to and understandable by others, particularly those individuals within a company who are responsible for growth, analysis, efficiency, and operations. Having access can be beneficial to shareholders as well.

## Importance and Uses of Data Analytics

Data analytics provide a critical component of a business’s probability of success. Gathering, sorting, analyzing, and presenting information can significantly enhance and benefit society, particularly in fields such as healthcare and crime prevention. But the uses of data analytics can be equally beneficial for small enterprises and startups that are looking for an edge over the business next door, albeit on a smaller scale,

## **Why Is Data Analytics Important?**

Implementing data analytics into the business model means companies can help reduce costs by identifying more efficient ways of doing business. A company can also use data analytics to make better business decisions.

## **What Are the 4 Types of Data Analytics?**

Data analytics is broken down into four basic types. Descriptive analytics describes what has happened over a given period. Diagnostic analytics focuses more on why something happened. Predictive analytics moves to what is likely going to happen in the near term. Finally, prescriptive analytics suggests a course of action.

## **Who Is Using Data Analytics?**

Data analytics has been adopted by several sectors where turnarounds can be quick, such as the travel and hospitality industry. Healthcare is another sector that combines the use of high volumes of structured and unstructured data, and data analytics can help in making quick decisions. The retail industry also uses large amounts of data to meet the ever-changing demands of shoppers.