The Interesting Dataset That I Have Found Recently.

The interesting dataset that I have discovered recently is *Student Performance Data Set* which is available on *UCI Machine Learning Repository*. This dataset is prepared by *Paulo Cortez, University of Minho, Portugal.*

LINK TO THE DATASET: https://archive.ics.uci.edu/ml/datasets/student+performance

This dataset consists of the performance of the students of the two schools. The data was conducted to figure out the performance of the students with respect to various attributes — school, age, home to school travel time, activities, gender, food habits, parents' education, and many more.

The dataset consists of thirty-three attributes and 649 rows. The data is collected through school reports and questionnaires.

Following are the conclusion of the study of the dataset:

1. The students, who have done the test preparation course, perform better than those who have not done that.
2. Most of the students got B and C grades followed by A grades in the third rank.
3. Students living in urban areas are more likely to score better than students in rural areas.
4. The students, who live near the school, perform better than those who live far from school.
5. Students who engage in extra curriculum activities perform well than the students who don't.
6. Students who take good/healthy lunches do perform better than those who don't.
7. Parents' education does not affect much of the performance of the student, but if the parents have master's degrees then most likely the student will have overall high performance.

File System and Database System

File systems and database systems may have a single motivation, that is to store data, but they come up with different promising factors that one has to note while working with them.

- If we talk about file systems, it stores data on a disk of the computer. They are less expensive and less complex than DBMS. They are not meant to do several transactions at a time. Only one user can access a file at a time. The file system provides the detail of the data representation and storage of data. There is no efficient query processing in the file system. Constraints are difficult to maintain in the file system. File systems provide less security in comparison to DBMS. It doesn't provide backup and recovery of data if it is lost. There are more chances of data redundancy and data inconsistency in this.

DBMS, which uses software to store and retrieve the data, is more expensive and more complex than the file system, but this complexity comes from the allowance of many transactions at a time. Many users can access the database at the same time. DBMS gives an abstract view of data that hides the details. There exists query processing language that helps to access and retrieve data efficiently. Constraints are easy to maintain in DBMS. It is more secure than the file system. It provides backup and recovery of data even if it is lost. Data consistency is controlled by the use of normalization, and there is no data redundancy in this.

About Unix Command Line

I do know the basics of shell scripting. For shell scripting, I have audited the course from Edx named *Shell Programming: A necessity for all Programmers*.

Course link: https://www.edx.org/course/shell-programming-a-necessity-for-all-programmers

I do know how to handle text, filters, and directories. I have learned about input, output, processes, and substitutions. I do know how to use conditional statements, loops, and functions. I have used shell tools like find, sed, lsof, curl, wget, ssh, and more. I do know how to use text filters and regular expressions.