**Assignment: 1**

**Problem Statement:**

Get Admission dataset from following link.

[Admission dataset](https://drive.google.com/file/d/10BcQF6yQc4aBRGPaatboEc-3ndLgR5fZ/view?usp=classroom_web&authuser=0)

Perform the following operations using Python on suitable data sets:

a) read data from different formats (like csv, xls)

b) indexing and selecting data, sort data,

c) describe attributes of data, checking data types of each column,

d) counting unique values of data, format of each column, converting variable data type

(e.g. from long to short, vice versa),

e) identifying missing values and fill in the missing values

**Objective:**

This assignment aims to demonstrate proficiency in data manipulation and analysis using Python. Specifically, it focuses on performing common data operations such as reading data from different formats, indexing, selecting, sorting, describing attributes, checking data types, counting unique values, converting variable data types, and handling missing values.

**Software and Hardware Requirements:**

* Operating System: Windows, Linux, macOS
* Python Environment: Anaconda, Jupyter Notebook
* Libraries and Packages: Pandas, NumPy

**Theory:**

1. **Pandas Introduction:** Pandas is a popular Python library used for data manipulation and analysis tasks.

**2.Data Structures:** Pandas provides two main data structures - Series and DataFrame.

* Series: It is a one-dimensional array that can hold any data type and is labeled.
* DataFrame: It is a two-dimensional labeled data structure with columns of potentially different types, similar to a spreadsheet or SQL table.

**3.Functionality:** Pandas allows users to perform various operations on data, including loading data from different file formats (such as CSV, Excel, SQL), manipulating data (sorting, filtering, grouping), and performing statistical and analytical tasks.

**4.Ease of Use**: Pandas is known for its ease of use, providing intuitive data structures and functions that simplify complex data operations.

**Methodology:**

**a) Reading Data from Different Formats:**

Utilize libraries like Pandas to read data from CSV, Excel, or other formats into data structures like Data Frames.

**b) Indexing, Selecting, and Sorting Data:**

Use Pandas indexing methods like loc[] for label-based indexing and sorting functions like sort\_values().

**c) Describing Attributes of Data and Checking Data Types:**

Leverage Pandas' describe() function for summary statistics and dtypes attribute for data type inspection.

**d) Counting Unique Values, Checking Format, and Converting Variable Data Type:**

Employ Pandas' functions like nunique() for counting unique values, dtypes attribute for format checking, and astype() for data type conversion.

**e) Identifying Missing Values and Filling Them:**

Use Pandas isnull() and fillna() functions to identify and fill missing values respectively.

**Advantages and Applications:**

These operations facilitate efficient data manipulation, preprocessing, and analysis, crucial for exploratory data analysis (EDA) and modeling tasks.

They enable researchers, data scientists, and analysts to gain insights, prepare data for modeling, and build predictive models for various domains like finance, healthcare, marketing, etc.

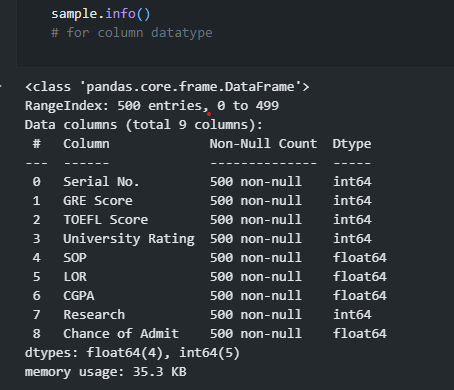
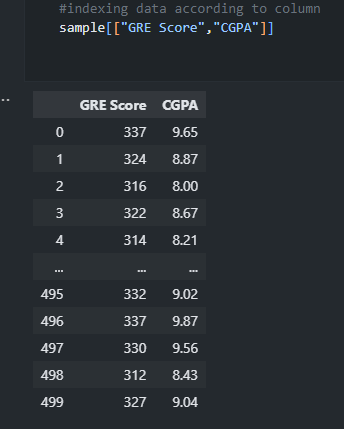
**Limitations with Example:**

One limitation could be the computational overhead and memory usage, especially with large datasets. For instance, sorting or indexing large datasets may require significant computational resources and time.

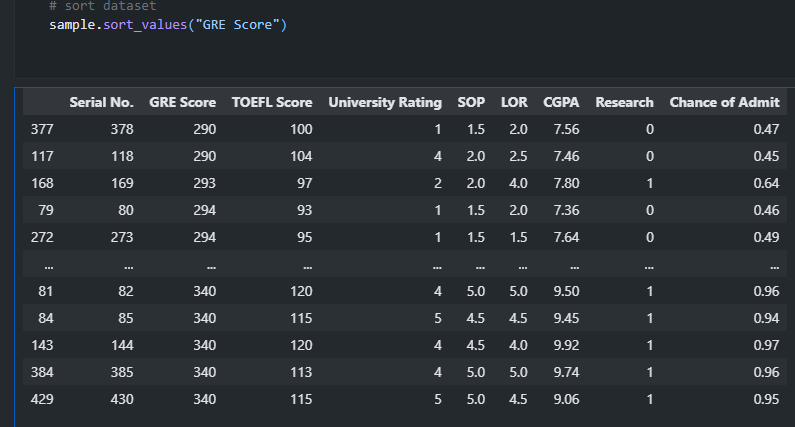
**Working and Algorithm:**

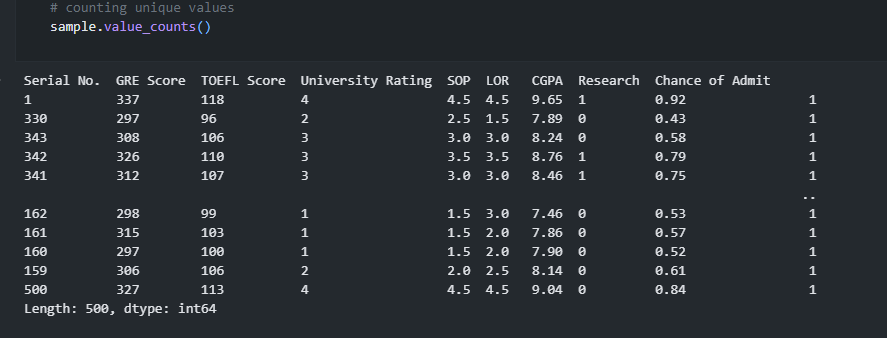
1. Read data from different formats (CSV, Excel) using Pandas.
2. Index, select, and sort data using Pandas methods.
3. Describe attributes of data and check data types using Pandas.
4. Count unique values, check format, and convert variable data types using Pandas.
5. Identify missing values and fill them using Pandas methods.

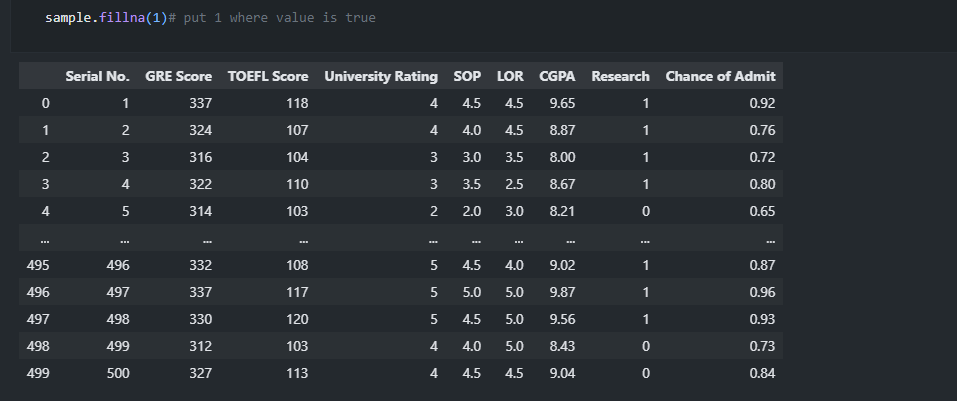
**Diagram:**

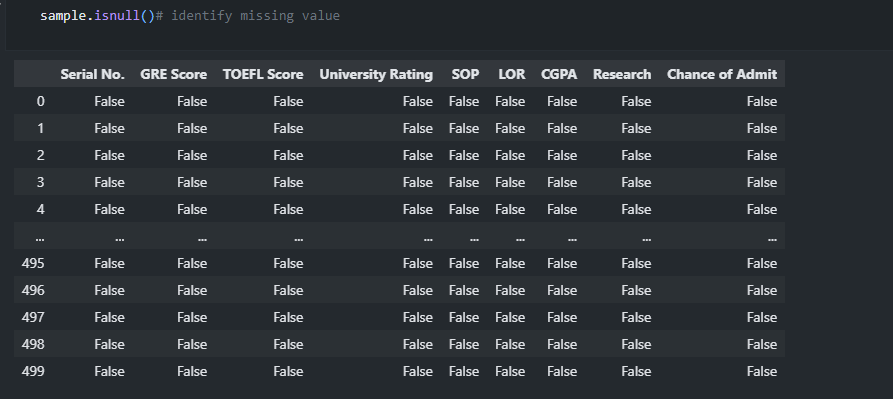
**** ****











**Conclusion:**

Python, with its powerful libraries like Pandas, provides a comprehensive set of tools for data manipulation and analysis. By mastering these operations, analysts can efficiently handle and analyze data, making informed decisions and deriving valuable insights from datasets.