**National University of Computer and Emerging Sciences**



**Lab Manual 07**

Department of Computer Science

**Objective:**

In the lab students will learn:

* Datatype Conversions

**String to Numeric**:

The pd.to\_numeric() function is used to convert string data to numeric data types (either integers or floats). This is necessary when dealing with numerical values stored as strings due to data import issues.



The errors='coerce' argument is useful for handling invalid parsing, converting them to NaN.

**Datetime Conversions**:

The pd.to\_datetime() function converts string representations of dates into pandas datetime objects. This is essential for time-series analysis and date-based operations.



**Float to Integer**:

The astype() function is used to change data types. When converting from floats to integers, it truncates the decimal points.



|  |  |  |
| --- | --- | --- |
| Function | Purpose | Syntax |
| pd.to\_numeric() | Converts string to numeric | pd.to\_numeric(df['string\_col'], errors='coerce') |
| pd.to\_datetime() | Converts string to datetime | pd.to\_datetime(df['string\_date\_col']) |
| astype() | Converts a column to another datatype | df['float\_col'].astype(int) |

**Questions:**

1. Write a function to fill missing values in a DataFrame with the mean of the column, but only for numeric columns.
2. How would you drop rows where any two specific columns both contain missing values? (e.g., drop rows where both Score and Age have missing values)
3. Identify the top 5 outliers in a DataFrame column using the z-score method, and drop these outliers from the DataFrame.
4. Write a code to remove duplicate rows but keep the last occurrence of each duplicate.
5. Convert a column of string dates to datetime, extract the year from each date, and store it in a new column called Year.