Day 05

Topic Covered: Data Frames and Factors in R

Summary:

Today’s session focused on two important data structures in R — data frames and factors. These structures are widely used in data analytics and statistical modeling. A data frame is similar to a spreadsheet, storing data in rows and columns, where each column can have a different data type. This makes data frames ideal for real-world datasets.

Factors are used to store categorical data, such as gender or city names. They store unique levels, helping classify and analyze categories more efficiently. Both data frames and factors are essential for data cleaning, transformation, and visualization, forming the foundation for more advanced analytics tasks.

New Concepts Learned:

- **Data Frames: Two-dimensional tables created using data.frame().**
  Key functions:
  - summary(): Provides data summary
  - length(): Returns number of columns
  - dim(): Shows rows and columns
  - nrow() and ncol(): Count rows and columns
  - cbind() and rbind(): Add columns or rows
    Accessing data using indexing, column names, or the $ operator.
- **Factors: Used to represent categorical data with unique levels.**
  Key functions:
  - factor(): Create a factor
  - levels(): Displays categories
  - length(): Number of elements
    Factors support adding and modifying levels.

Activity:

- Created and manipulated data frames using data.frame(), cbind(), and rbind()
- Practiced accessing columns and rows using indexing and the $ operator
- Used summary(), nrow(), and ncol() to analyze structure
- Created factors for categorical data and worked with levels() and length()

Challenges Faced:

Understanding the difference between numeric values and factor levels was slightly confusing. Managing data frames with mixed data types also required careful handling to avoid inconsistencies.

Key Takeaway:

Data frames and factors are essential for organizing and categorizing data in R. They form the core of data preprocessing and are crucial for preparing datasets for meaningful analysis and visualization.