DAY 3

* **Data Transformation**
* It is the process of converting data from one format, structure, or value into another to make it suitable for analysis, storage, or further processing.
* It involves converting, cleansing, and structuring data into a usable format which is used to analyzed to support decision-making processes.
* **Various methods of data transformation**

1. **Normalization**

* It scales numbers so they are between 0 and 1.
* Makes different types of data easier to compare, even if they have different ranges.
* Ex. Changing income from 10,000 – 1,00,000 to a scale of 0 to 1.

1. **Standardization**

* Helps some machine learning models work better by making all data behave the same way.
* Ex. Adjusting test scores so they all follow a consistent scale.

1. **Encoding**

* Converts words or categories into numbers.
* Computers can understand numbers, not words, so this makes data usable for analysis.
* Ex. Changing "Yes" to 1 and "No" to 0.

1. **Discretization**

* Turns continuous data (like age or salary) into groups or categories.
* Makes complex data easier to analyze by grouping it into simple categories.
* Ex. Grouping ages into "Under 18," "18-35," "36-60," and "Above 60."

1. **Attribute Generation**

* Creates new information from existing data.
* Gives more details for analysis, which can help find patterns.
* Ex. Creating an "age" value from a person’s birth year.

1. **Revising**

* Fixes mistakes or inconsistencies in the data.
* Ensures the data is accurate and reliable.
* Ex. Removing duplicate entries or fixing wrong dates.

1. **Manipulation**

* Changes or calculates new values based on existing data.
* Helps to make data more useful or easier to work with.
* Ex. Combining first and last names into a full name.

1. **Separating**

* Breaks down a piece of information into smaller parts.
* Makes it easier to filter or analyze specific details.
* Ex.Splitting an address into street, city, and zip code.

1. **Combining/Integrating**

* Brings together data from different sources into one place.
* Gives a complete view of the data by merging information.
* Ex.Merging customer data from different databases into a single record.
* **Advantages of Data Transformation**
* Better Data Quality – Fixes errors and makes data more accurate.
* Easy to Use – Converts data into a format that different systems can understand.
* Faster Analysis – Organizes data so it can be processed quickly.
* Saves Space & Time – Removes unnecessary data and improves efficiency.
* More Secure – Helps protect sensitive information.
* **Limitations of Data Transformation**
* Takes Time & Effort – Can be slow and complex, especially for big data.
* Risk of Mistakes – Poor transformation can lose or change important data.
* Expensive – Requires special tools and skilled people.
* Hard to Scale – Some methods may not work well with very large data.
* Can Introduce Errors – If not done correctly, it may create biased or misleading results.
* **Why do businesses need data transformation?**

Organizations generate a huge amount of data daily. However, it is of no value unless it can be used to gather insights and drive business growth. Organizations utilize data transformation to convert data into formats that can then be used for several processes. There are a few reasons why organizations should transform their data.

* Transformation makes disparate sets of data compatible with each other, which makes it easier to aggregate data for a thorough analysis
* Migration of data is easier since the source format can be transformed into the target format
* Data transformation helps in consolidating data, structured and unstructured
* The process of transformation also allows for enrichment which enhances the quality of data

The ultimate goal is consistent, accessible data that provides organizations with accurate analytic insights and predictions.