**Vivekanand Education Society’s Institute of Technology Department of AI&DS**

****

**Subject: Social Media Analytics**

**Class: D16ad**

| **Roll No: 30** | **Name:** [**SUHANEE KANDALKAR**](mailto:2021.suhanee.kandalkar@ves.ac.in) |
| --- | --- |
| **Practical No: 3** | **Title: Data cleaning and pre processing** |
| **DOP:** | **DOS:** |
| **Grades:** | **LOs Mapped:** |
| **Signature:** |  |

**Title:** Social Media Data Cleaning

**Aim:** To perform Data Cleaning and Preprocessing operation on Social media data for business perspective ( Using Python, MongoDB, R, etc).

**Theory:**

#### What is Data Cleaning and Preprocessing?

Data cleaning and preprocessing involve several operations aimed at improving the quality of raw data before it is used for analysis or model-building. This process involves:

* **Removing or correcting inaccurate, incomplete, or irrelevant data.**
* **Handling missing values.**
* **Removing duplicates.**
* **Normalizing and scaling data.**
* **Converting categorical data into numerical format.**
* **Dealing with outliers.**

These steps are crucial because high-quality, clean data ensures that analysis or models built on it are accurate, efficient, and useful.

#### Need for Data Cleaning and Preprocessing

Data cleaning and preprocessing are critical for the following reasons:

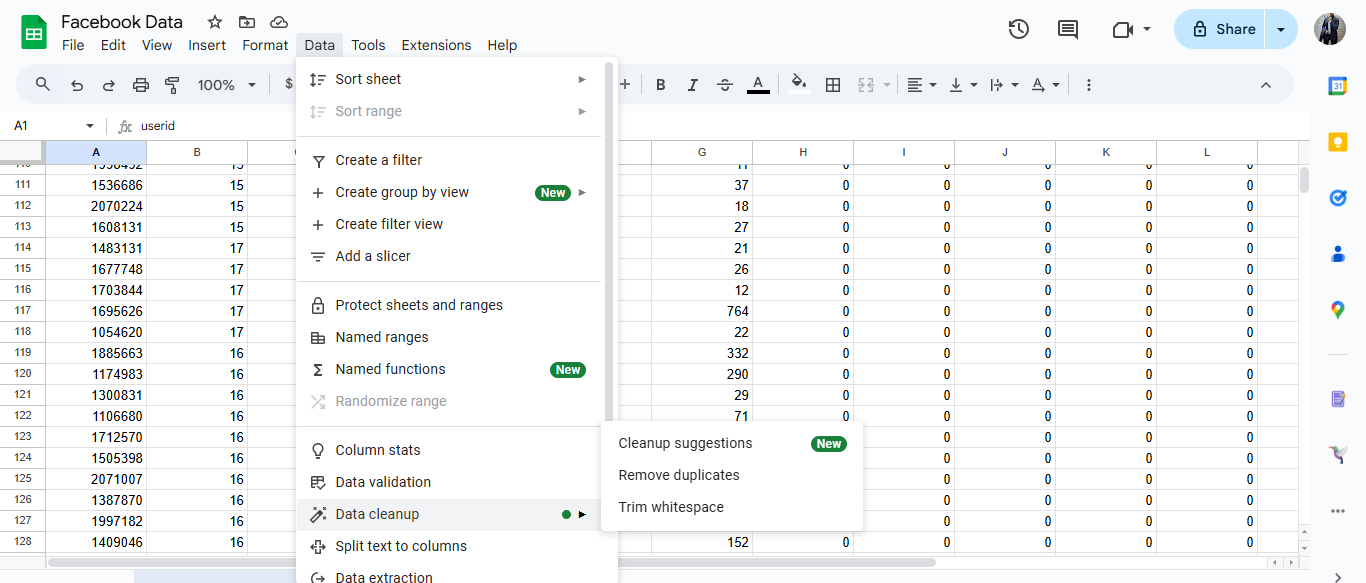
1. **Accuracy:** Raw data may contain errors or inconsistencies. Cleaning removes these issues to ensure that the analysis is based on reliable information.
2. **Improvement of Data Quality:** Unclean data can lead to misleading conclusions. Preprocessing enhances the quality and structure of data.
3. **Model Efficiency:** Machine learning models perform better when data is clean and well-structured.
4. **Better Insights for Business Decisions:** Clean data provides more accurate insights, which can guide business strategies, marketing campaigns, or product developments.
5. **Handling Missing or Corrupt Data:** Social media data, especially from platforms like Twitter or Facebook, can be incomplete or corrupted, requiring preprocessing steps like imputing missing values or removing irrelevant information.

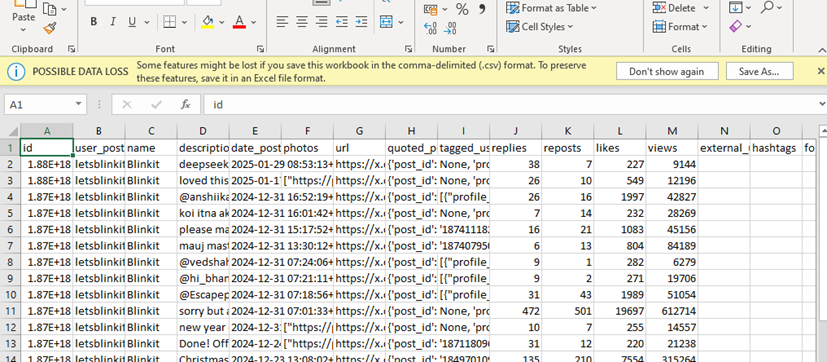
**OUTPUT:**

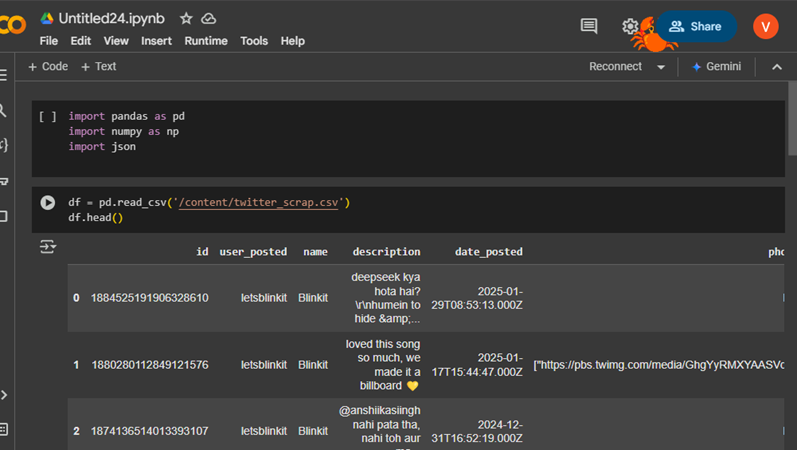
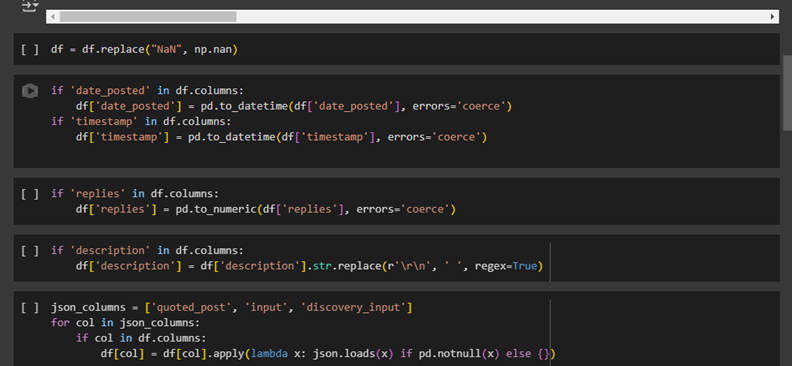
Screenshots

Data Cleaning and prep

**Google Sheets**



**Microsoft Excel**

**Using Python **





**Conclusion:**

* **Data Cleaning is crucial**: Raw social media data is often noisy and inconsistent. By applying various cleaning and preprocessing techniques, we enhance the quality of the data for subsequent analysis or modeling.
* **Business Use Case**: Clean data allows businesses to gain accurate insights from social media platforms, whether it's for sentiment analysis, brand monitoring, or customer feedback.
* **Efficiency in Analysis**: Using tools like Python, MongoDB, and data visualization libraries helps streamline the data cleaning process, ensuring the data is in an ideal state for analysis.