



Task- 1(Easy)

Task: Data cleaning

Description:

Perform data cleaning steps on a dataset given to you. It involves analyzing and cleaning the data.

Requirements:

- Understanding of Python and libraries like Pandas and NumPy.
- Jupyter Notebook or any Python environment.

Steps to Follow:

- **Dataset:** Download the dataset from given link on your pc.
- Load the dataset on jupyter notebook or any python environment.
- **Data cleaning:** Clean the data by handling missing values, removing duplicates.
- Fill the missing values using mean, median and mode of that column.
- Remove duplicate columns.

What You'll Learn:

- Techniques for handling missing data.
- Identifying and removing duplicates.

- Standardizing data formats.
- Dealing with outliers and data transformation.

Additional Suggestions(Optional):

- Experiment with different imputation strategies for missing values.
- Visualize data distributions before and after cleaning for comparison.
- Utilize Pandas functions efficiently to simplify the cleaning process.

Data set link:

<https://drive.google.com/file/d/1TaUffBPO7GqLjzSvl-LjvWujZwdhmuqM/view?usp=sharing>

Conclusion:

This project will help you gain valuable skills in data preprocessing, a critical step in any data science or analysis project. Cleaning messy data is an essential skill for ensuring the accuracy and reliability of your analyses.