**Idea: Cleaning Data**

**Description:**

[Data cleaning](https://www.springboard.com/blog/data-analytics/data-cleaning/) is the process of fixing or removing incorrect, corrupted, duplicate, or incomplete  
data within a dataset. Messy data leads to unreliable outcomes. Cleaning data is an essential  
part of data analysis, and demonstrating your data cleaning skills is key to landing a job. Here  
are some projects to test out your data cleaning skills:

**Dataset 1** [**Link**](https://www.kaggle.com/datasets/dgomonov/new-york-city-airbnb-open-data)

**Dataset 2** [**Link﻿**](https://www.kaggle.com/datasets/datasnaek/youtube-new)

**Key Concepts and Challenges:**

1. Data Integrity: Ensuring the accuracy, consistency, and reliability of data throughout thecleaning process.
2. Missing Data Handling: Dealing with missing values by either imputing them or makinginformed decisions on how to handle gaps in the dataset.
3. Duplicate Removal: Identifying and eliminating duplicate records to maintain datauniqueness.
4. Standardization: Consistent formatting and units across the dataset for accurate analysis.
5. Outlier Detection: Identifying and addressing outliers that may skew analysis or modelperformance.