Data_Cleaning_Process.docx

Title: Data Cleaning Process for Hip Hop Streaming Analysis

1. Introduction

- Data cleaning is crucial for ensuring the accuracy and reliability of the analysis. This document outlines the steps taken to clean the datasets used in this project.

2. Initial Data Inspection

- Loaded the `Hip_Hop_Streaming.csv` file into Excel to inspect the data structure.
- Identified columns: Playlist Name, Album Playlist, Views, Estimated Revenue, etc.

3. Cleaning Steps

Removing Duplicates: Checked for and removed duplicate entries.
excel
Copy code

```
Data > Remove Duplicates
```

-
- **Handling Missing Values:**

Replaced missing revenue values with 0 using the formula:
excel
Copy code

```
=IF(ISBLANK(B2), 0, B2)
```

   -
- **Separating Revenue for Different Artists:**
   - Used filters to isolate revenue entries for specific artists.
- **Standardizing Formats:** Ensured consistent formatting for all entries, particularly for numerical values.
- **Extracting Categories:**

Used formulas to categorize data:
excel
Copy code

```
=IF(A2="Kendrick Lamar", "Kendrick Lamar", IF(A2="J. Cole", "J. Cole", "Other"))
```

-

4. Final Review

- Reviewed cleaned data for accuracy and completeness.
- Saved the cleaned dataset as `Processed_Hip_Hop_Streaming.csv` for analysis.

5. Tools Used

- Microsoft Excel for initial cleaning.
- Python and Pandas for further processing and analysis.