🧹 AI Data Cleaner using Gemini

🔑 Enter your Gemini API Key   ⦾

•••••••••••••••••••••••••••••••••••••   👁

✅ API Key configured successfully!

📁 Upload your dataset (CSV or Excel)

☁ **Drag and drop file here**
Limit 200MB per file • CSV, XLSX   Browse files

📄 **my_file (1).csv** 2.1KB   ✕

🔍 Original Dataset Preview

	Rank	Peak	All Time Peak	Actual gross	Adjusted gross (in 2022 dollars)	Artist	Tour title
0	1	1	2	$780,000,000	$780,000,000	Taylor Swift	The Eras Tou
1	2	1	7[2]	$579,800,000	$579,800,000	Beyoncé	Renaissance
2	3	1[4]	2[5]	$411,000,000	$560,622,615	Madonna	Sticky & Swe
3	4	2[7]	10[7]	$397,300,000	$454,751,555	Pink	Beautiful Tra
4	5	2[4]	None	$345,675,146	$402,844,849	Taylor Swift	Reputation S
5	6	2[4]	10[9]	$305,158,363	$388,978,496	Madonna	The MDNA T
6	7	2[10]	None	$280,000,000	$381,932,682	Celine Dion	Taking Chan
7	7	None	None	$257,600,000	$257,600,000	Pink	Summer Car
8	9	None	None	$256,084,556	$312,258,401	Beyoncé	The Formati
9	10	None	None	$250,400,000	$309,141,878	Taylor Swift	The 1989 Wo

Original Shape: 20 rows × 11 columns

📝 Cleaning Summary Report

Here's a summary report of the data cleaning performed:

Data Cleaning Summary Report

Dataset Overview: The dataset initially contained 20 rows and 11 columns. After the cleaning process, the dataset maintains the same structure, with 20 rows and 11 columns. No rows or columns were added or

removed.

Cleaning Focus: The primary objective of this cleaning pass was to address missing values in two key columns: 'Peak' and 'All Time Peak'.

Detailed Cleaning Steps & Justification:

1. **Column: 'Peak'**

   - **Problem Identified:** This column had **11 missing values**.
   - **Action Taken:** All 11 missing values in the 'Peak' column were filled using the **mode** of the column, which was **'1'**.
   - **Why this approach?** We chose to impute with the mode because 'Peak' often represents a specific rank, level, or a discrete category. Using the most frequent existing value ('1') ensures that the imputed data remains consistent with the most common observation in the dataset, avoiding the introduction of artificial values that could skew the distribution if numerical averages (mean/median) were used. This approach preserves the typical representation of 'Peak' within the dataset.

2. **Column: 'All Time Peak'**

   - **Problem Identified:** This column had a more significant number of missing values, totaling **14**.
   - **Action Taken:** All 14 missing values in the 'All Time Peak' column were filled using the **mode** of the column, which was **'10[7]'**.
   - **Why this approach?** Similar to 'Peak', the values observed in 'All Time Peak' (such as '10[7]') suggest a categorical or specific identifier format rather than purely continuous numerical data. In such cases, the mode is the most appropriate imputation strategy. It guarantees that the filled-in values are actual valid entries found within the column and represent the most common 'All Time Peak' status, thereby maintaining the integrity and most frequent characteristic of this specific data point.

Conclusion: The cleaning process successfully addressed all identified missing values in the 'Peak' and 'All Time Peak' columns through targeted mode imputation. Your dataset is now fully populated in these critical fields, ensuring it is complete and ready for further analysis or model training without the issues previously caused by incomplete data.

Cleaned Shape: 20 rows × 11 columns

📊 Cleaned Dataset Preview

	Rank	Peak	All Time Peak	Actual gross	Adjusted gross (in 2022 dollars)	Artist	Tour title
0	1	1	2	$780,000,000	$780,000,000	Taylor Swift	The Eras Tou
1	2	1	7[2]	$579,800,000	$579,800,000	Beyoncé	Renaissance
2	3	1[4]	2[5]	$411,000,000	$560,622,615	Madonna	Sticky & Swe
3	4	2[7]	10[7]	$397,300,000	$454,751,555	Pink	Beautiful Tra
4	5	2[4]	10[7]	$345,675,146	$402,844,849	Taylor Swift	Reputation S
5	6	2[4]	10[9]	$305,158,363	$388,978,496	Madonna	The MDNA T
6	7	2[10]	10[7]	$280,000,000	$381,932,682	Celine Dion	Taking Chan
7	7	1	10[7]	$257,600,000	$257,600,000	Pink	Summer Car
8	9	1	10[7]	$256,084,556	$312,258,401	Beyoncé	The Formati
9	10	1	10[7]	$250,400,000	$309,141,878	Taylor Swift	The 1989 Wo

⬇️ Download Cleaned Dataset

⬇️ Download Cleaning Report

💡 **Tips:**

- The tool automatically removes duplicates and fills missing values
- Numeric columns use mean for missing values
- Text columns use mode (most common value) for missing values
- Get your free Gemini API key from Google AI Studio