1. Remove useless columns
2. Column name correction
3. Description of the dataset
   1. Dim
   2. range / unique
   3. Certain pattern for str col(to be done later)
4. Check duplicate values (Rows and Column): Checking object type(json strings)
5. Check missing values
   1. Check rows first and then column
   2. Leave blank or fill with average or previous row
6. Check bad values
   1. Out ranged
   2. Wrong type

Rating:

The proportion of error type in each column

Data science question:

1.How does the sound feature of music affect the categorization and its relationship between trending?

Potential Analysis:

1. Description of the dataset
2. Explain variables
3. Directions
   1. What are the similarities of sound features for sleeping type music and dining music?
   2. What is the most important feature in party music?
   3. What kind of music playlist affect people’s emotion most?
   4. What is the difference among R&B, Blues, and soul music?
   5. What is the best type of music for gaming?
4. The rating procedure

Citation:

Broeck et al. Data Cleaning: Detecting, Diagnosing, and Editing Data Abnormalities. PLOS One – Medicine, 2005.<http://journals.plos.org/plosmedicine/article?id=10.1371/journal.pmed.0020267>

[http://cs229.stanford.edu/proj2016/report/RidleyDumovic-ClassificationOfArtistGenrethroughSupervisedLearning-report.pdf](http://journals.plos.org/plosmedicine/article?id=10.1371/journal.pmed.0020267)

没用的：

Disc\_number, explicit, external\_ids,external\_urls, href, linked\_from,preview\_url, uri, analysis\_url, track\_href, type.1, uri.1, track\_id