* Spotify Top 100 USA Podcasts with EP
  + Documented Changes from Original:
    - Basic formatting like bolding column names
    - Of the original dataset I choose to use the data from the 2 CSV files named spotify\_usa\_top100\_podcasts\_episodes and spotify\_usa\_top100\_podcasts as the data from the CSV file titled spotify\_usa\_podcasts\_ranks was repetitive and unnecessary
    - Deleted the first row from both files as it was just a repetitive numbering of rows
    - spotify\_usa\_top100\_podcasts\_episodes
      * Removed columns title ‘id’, ‘is\_externally\_hosted’, ‘is\_playable’, ‘language’, ‘languages’, and ‘restriction\_reason’ as the data in these rows is irrelevant to our analysis
    - spotify\_usa\_top100\_podcast
      * Removed columns titled ‘copyrights’, ‘languages’, ‘is\_externally\_hosted’, and ‘available\_markets’ as the data in these rows is irrelevant to our analysis
* Spotify User Behavior Dataset
  + Documented Changes from Original:
    - Basic formatting like bolding column names
    - Removed columns C-N and H because they contained data about music or other data irrelevant to the analysis of podcasts. Remaining rows include: Age, Gender, pod\_lis\_frequency, fav\_pod\_genre, preffered\_pod\_format, pod\_host\_preference, preffered\_pod\_duration
    - Removed 130 rows where pod\_lis\_frequency was ‘Never’ as people who never listen to podcasts are not helpful for determining what will make a podcast more popular or higher quality
    - Removed 4 rows where Age was either “6-12” or “60+” as the number of entries in these groups is so small