Billboard Project Process Map:

1. Develop hypothesis
   1. Question: “why do we like what we like?”
   2. Assumptions:
      1. Given the available resources for the project, the Billboard Hot 100 is the best measurement of “what we like.”
      2. Billboard’s methodology must change over time, if for no other reason than emerging technology.
         1. Focus was on change in December 1998:
            1. Additional radio formats, including R&B, mainstream rock, AAA, and country music were added to data collection.
            2. Weighting of sales in the ranking calculation declined from 40% to 25%.
            3. Songs unavailable for sale as singles would now be included.

Record labels would provide albums to radio stations without releasing

* + 1. If the basis of measurement for what we like changes over time, then before we can consider the impact of the qualities of specific entries (e.g. artist’s gender, song genre, duration, seasonality, etc.) we need to consider the impact of the system itself.
  1. Hypothesis: Changes in methodology have a significant impact on key measurements in ranking on the Billboard Hot 100.
  2. Risks:
     1. While the change in methodology should impact the resulting report, individual entries could be outliers, skewing the data.

1. Acquire data:
   1. Internet search uncovered the below links, which included a database of the Billboard Hot 100 from 1940-2015
      1. <http://www.modestinsights.com/analyzing-the-billboard-hot-100/>
         1. <http://www.modestinsights.com/wp-content/uploads/2015/03/all_billboard_data.txt>
2. Import and Clean data
   1. Convert pipe-delimited txt file to csv
      1. Found Stack Overflow code to convert
         1. Originally had done conversion in Excel
   2. Reformat “chart date” column to include only the year
   3. Convert Python list of lists to Pandas dataFrame
   4. Convert appropriate columns to numeric dtype.
3. Data analysis
   1. Create pivot table showing the mean, min, and max for entry position and overall weeks on chart.
      1. Original hypothesis focused on the change in methodology in 1998 but data revealed several other occasions where trends shifted. Revisited original analysis to include other significant shifts:
         1. 1991: Compilation moves from submitted reports by radio stations and retailers to observed airplay via electronic monitoring in 85 markets and actual sales, both collected by Nielsen.