# **Introduction**

* Why did you choose this topic and questions we're interested in studying
* List team members and a description of what everyone did

# **Description of Data**

* Describe how data was collected/accessed
  + Who is included in this study
  + When was this query processed
  + Definition of any features
  + How is data collected?
  + Estimated accuracy rate?
* Merge the data / Match
  + Match song name from Billboard - Spotify: lowercase everything -- Derek (over spring break)
* Rpub: everyone should check out the analyses that other people have done
* Everyone should link their columbia emails to github
* Other noteworthy features

# **Analysis of Data Quality**

* Graphs, textual description
  + Univariate graphs/distributions -- Manksh + Louis
  + Outliers
  + Clusters

# **EDA**

* Findings
* Data Cleaning Process
* Approaches that worked / didn't
* Data challenges

# **Conclusion**

* Limitations,
* Next Steps