**Project Requirements**

 Use Pandas to clean and format your data set(s)

Group

 Create a Jupyter Notebook describing the **data exploration and cleanup** process

Amber- where to find?

 Create a Jupyter Notebook illustrating the **final data analysis (Conversational)**

Amber

 Create at least 6-8 visualizations (ideally 2 per "question" asked)

Group

 Save images of your visualizations (png or svg) to include in a markdown document (?) with findings of your project (and to distribute to the class and instructional team), and for inclusion in your presentation = Read Me document and Presentation

Amber

 Use at least one API

Steve

 Create a write-up summarizing your major findings. This should include a heading for each "question" you asked of your data, and under each heading, a short description of what you found and any relevant plots. This should be continually updated as part of your repository's readme.md file.

Amber

 Use at least one statistical test to accept or reject a hypothesis described in your project objective

Steve/Christian

**Presentation Requirements**[**¶**](http://localhost:8891/lab#Presentation-Requirements)

The presentation requirements for Project 1 are as follows.

Your presentation must:

*  Be at least 8-10 min. long
*  Describe the core message or hypothesis for your project.
*  Describe the questions you and your group found interesting, and what motivated you to answer them
*  Summarize where and how you found the data you used to answer these questions
*  Describe the data exploration and cleanup process (accompanied by your Jupyter Notebook)
*  Describe the analysis process (accompanied by your Jupyter Notebook)
*  Summarize your conclusions. This should include a numerical summary (i.e., what data did your analysis yield), as well as visualizations of that summary (plots of the final analysis data)
*  Discuss the implications of your findings. This is where you get to have an open-ended discussion about what your findings "mean".
*  Tell a good story! Storytelling through data analysis is no different than in literature. Find your narrative and use your analysis and visualization skills to highlight conflict and resolution in your data.

**Technical Requirements**

* The technical requirements for Project 1 are as follows.
*  Use Pandas to clean and format your data set(s)
*  Create a Jupyter Notebook describing the **data exploration and cleanup** process
*  Create a Jupyter Notebook illustrating the **final data analysis**
*  Use Matplotlib (or other Python plotting and graph library) to create at least 6-8 visualizations of your data (ideally, at least 2 per "question" you ask of your data)
*  Save images of your visualizations (png or svg) to include in a markdown document that communicates the findings of your project (and to distribute to the class and instructional team), and for inclusion in your presentation
*  Use at least one API, (assuming you can find an API with data pertinent to your primary research questions
*  Create a write-up summarizing your major findings. This should include a heading for each "question" you asked of your data, and under each heading, a short description of what you found and any relevant plots. This should be continually updated as part of your repository's readme.md file.
*  Use at least one statistical test to accept or reject a hypothsis described in your project objective

**IDEAS for Tasks**

Replay ability stats per artist- Amber (?)----Ask about this – where to find?

* ***Question 1- Which countries’ users engage the most with the app?***
  + Top 20 countries by listener average – Steve- Done
  + Type of chart: bar
  + Type of chart: Stacked bar
* ***Question 2- Which artist are most popular based on their country play count?***
  + Top 20 artists by count of countries- Steve- Done
  + Type of chart: Bar
* ***Question 3- Which 5 artists are played most often globally? \*\*\*\*Why does this matter?\*\*\****
  + 5 top artists (+ other) globally by sum of play counts- Amber
  + Type of chart: Pie
  + Look at Steve’s dataframe
  + Top 5 artists by playcount (or listener?)
* ***Question 4- Who are the top artists and listener count in Hungary?***
  + Top artist and the listeners by (1) top country chart (Hungary): Scatter- Ishamael (?)
* ***Question 5- Who are the most popular artists by listener per country?***
  + Top 5 artist per country (geomap with click action)- Christian
  + Type of visualization: Geo map
* ***Question 6- What are the most popular genres by listener per country?*** 
  + Top 5 genres per country – Steve- dataframe/Christian - geomap
  + Type of visualization: geomap with click action- group effort
* ***Question 7- What is the GDP and play count of each country?***
  + GDP and playcount- Christian
  + R2 – 0.01, so zero correlation
  + We fail to reject the null hypothesis, there seems to be no correlation between GDP and play counts/listeners (?)

**Hypothesis**

By the data set, lastfm will not be used (Listener or Play Count?) less in lower GDP countries.

Countries with lower GDP will have fewer listeners == If GDP is lower last.fm will have a higher listenership. (cheaper option)

(Creates null and alternative Hypothesis)

Type of analysis: Statistical

**Limitations-**

* No time frame or trending ability
* Only Last.fm data
* *Duplicate data*
* *API documentation is not up to date*
* Listeners versus population (could be a GDP correlation)—Access issue
* No demographic information

**Note:**

Top 50 artists based on listeners- country based

Top artists- play count and listeners and genre

Switch to GDP

Note for Amber: Look up Last Fm listener count and Spotify Listener count

Could create an md document? (Steve) –

I should have the repo on my github page—

Who speaks to what? – Are there

Google maps geo plot-

Notes from Jaime-

* What is high, medium, low in terms of listenership?
* Who used which CSVs?
* How did each person contribute?
* Did we see a relationship among genre and artist?
* Lok at color having meaning for our visualization…. So that people can make an easy inference
* Pinpoint where it started?
* Color code things based on locality or genre?
* Everyone create a github repo for a portion of our data
* What are some barriers to this app?
* Ask for questions- have one in the pocket.