# DS5110 Final Project

## Iteration #2

Peiying Li    Marc-Andy Mexil

# 1. Project Kickoff

- **Do you have a dataset ready to use for the current project?**
- **What are the specific goals of this project?**

We plan to use the [Top Spotify Podcast Episodes](#) dataset from Kaggle. This dataset contains  29 variables, including information such as date, rank, region, description, and creator information about podcast episodes. The dataset is updated daily and already contains 180k rows (about 370 MB in size).  We chose this dataset since it is barely used (currently only has one associated notebook) and contains a decent amount of diverse information.

Our main goal is to extract key insights from this dataset regarding the top trending podcasts. For example: what are people listening to and learning in the U.S. and around the world (what are the top trending podcast topics in the U.S. and around the world)? What is a good length for a successful episode? What are the catchwords for podcast titles? We may also use the following ideas from Kaggle as inspiration:
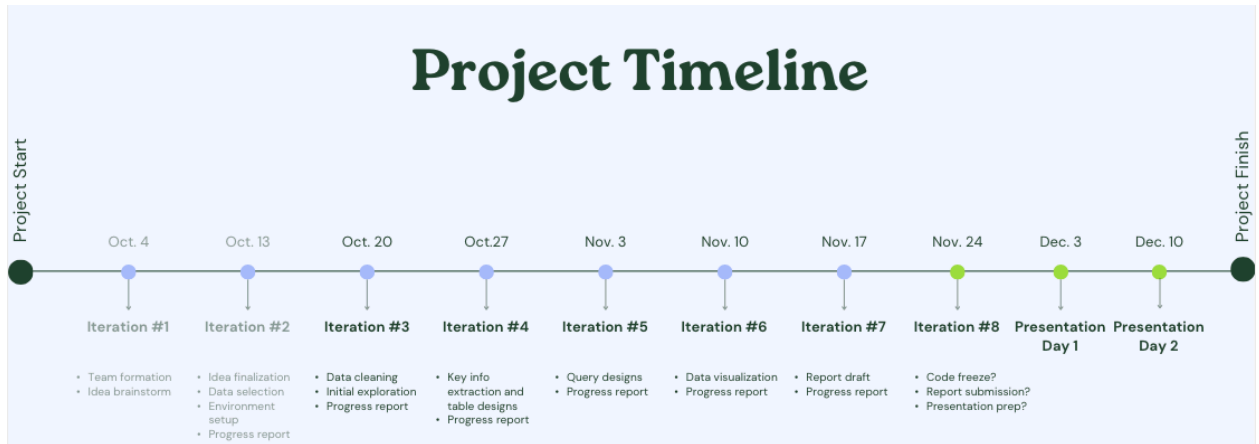
- Trending Podcasts: Discover the most popular podcast episodes in different regions.
- Cultural Preferences: Compare podcast popularity to understand cultural influences on podcast consumption.
- Trend Analysis: Track how podcast rankings change over time to identify emerging trends.
- Content Analysis: Dive into episode metadata for sentiment analysis, topic modeling, or genre classification.

To achieve our goals, we plan to conduct exploratory data analysis, data cleaning, data table designs (either in SQL or Python depending on the requirement), data querying, as well as creating insightful visualizations and/or dashboards to report our findings.

- **How do we define the project scope clearly to avoid scope creep?**

  We first discussed our plan within our group and documented our goals, tasks and concerns. Then seek guidance and validation from the teaching team to make sure our project is within the scope.

- **What deliverables must be completed at different phases?**
- **What are the major milestones, and what deadlines should we set?**



Project Timeline

- **Do the team's capabilities align with these goals? Are there any gaps that need to be addressed early on?**

  Based on each member's skills (documented below), we think that the team's capabilities align with our goals, and no gaps need to be addressed early on.

## 2. Team Discussions

- **What are the core skills each team member brings to the table?**
- **What tools do we have experience with, and what do we need to learn?**

  Peiying: Version control (in GitHub), data processing and visualizations in Python & R, relational DB design and data querying in SQL, project planning, technical report & presentation

  Marc-Andy: SQL, Tableau, presentation experience, need to learn Python

- **How will each person's expertise contribute to specific tasks?**

  **Coding:**

  Marc-andy: SQL queries for retrieving and answering questions about our data

Peiying: data loading, cleaning/processing, data visualization in Python

**Presentation:** We will work as a team to create an engaging presentation, with relevant visuals and a succinct summary of our findings.

**Report**: We will work as a team to create a report that accurately represents our findings in a way that can be understood and easily digested. It will contain an overview of our project and what we set out to do, the answers to the questions we asked, an analysis of those answers, the conclusion we reached due to our analysis, what actions should be taken from that conclusion, and any sources that we referenced for the project.

- **What skills are missing that may cause delays or challenges?**

  New skills needed to be gained from the course, and other guidance from the Prof./TA on business insights and project scope control.

- **What programming languages and platforms should we select based on our project needs and team experience?**

  We will use SQL and Python for relevant data processing and querying. We may use other Python packages, R or Tableau for data visualizations.

## 3. Skills & Tools Assessment

- **Are there external resources or team members with expertise in the areas where we lack skills?**

  Where Marc-Andy lacks in Python and R Peiying makes up for it with her expertise in said languages. He could also reference other documentation sheets for the languages.

- **Which tools, frameworks, and libraries are most suitable for the project's scope?**

  GitHub, Python (Pandas, Numpy, Seaborn, Matplotlib), R (Shiny), SQL, Tableau

- **How can we ensure that each team member is comfortable with the tools selected?**

  We have discussed each member's skills and will make sure to communicate and seek help when learning new skills on our own and practice with each other.

- **Have specific tasks been assigned based on individual strengths, and are team members clear on their roles?**

We have decided on a general task breakdown based on each member's strength, and will assign specific tasks as iterations go on, especially after our project idea has been reviewed by the teaching team.

## 4. Initial Setup

- **What development environment setup is necessary for this project?**

  Github, SQL/SQLite, Python

- **Have we successfully configured version control (such as Git)? Does everyone have access to the repository?**

  Yes. Please find our [Git repo here](#).

- **Have we installed and configured all required software, libraries, and tools?**

  In progress

- **What testing can we run to ensure that the development environment is functioning correctly?**

  We can use the command line to verify that our environment setups are the same (same version of the software, etc.).

- **What troubleshooting steps should we take if the setup does not work as expected?**

  We will help each other, Google, and or seek help from the Prof. /TA/ other classmates.

## 5. Progress Review

- **What has been achieved so far? Have we completed the initial setup and repository configuration?**
    1. Project idea selection & agreement
        a. Potential dataset selected
    2. Tasks, milestones, and timelines clear
    3. Environment setup done
    4. Report template done
- **Have there been any issues or blockers, and how can we address them quickly?**

  We faced challenges in deciding on the project topic and finding a suitable dataset. We chatted with the Prof. and each looked for potential datasets before finally

deciding on one.  When facing challenges, we should always seek help early and often so that we can address them quickly.

- **Is each team member contributing as expected, and does everyone understand their role?**
- **Are we on track with the timeline and milestones, or do we need to adjust them?**

    Yes and Yes.

- **How does the progress align with the project's overall objectives?**

    For this iteration, our main goal is to decide on a project topic and choose a dataset to work with, which we have completed and documented. The Prof. will guide us with the topic selection otherwise we are all set.

## 6. Plan Revision

- **Based on progress so far, do we need to adjust the project timeline or milestones?**
- **Are any tasks delayed or requiring reassignment due to workload or skill gaps?**

    No.

- **How can we ensure that all members are clear on the revised plan and their next steps?**

    We will meet and communicate regularly - we plan to meet briefly every other day to update each other on our progress.

- **What communication strategies can we implement to avoid future delays or misunderstandings?**

    Regular progress check for each other via teams/text, offer to help, communicate challenges, and seek help in time so that the other person can help, escalate to Prof. TA if further adjustments are needed.

- **How will we track progress going forward and maintain alignment with the revised plan?**

    Use our milestone and timeline as a guide and follow it.

## 7. Submission for This Iteration

- **What specific tasks need to be documented for this iteration's submission?**

    Idea selection and dataset, noting what skills are at our disposal.

- **Have we detailed the challenges faced, the solutions implemented, and any adjustments to the plan?**

  Yes.

- **If your data is available online, please provide a link to access it.**

  [Top Spotify Podcast Episodes](#)

- **Is the PDF using the Overleaf template, and does it reflect the team's actual progress?**

  Yes, it does. There is an outline of what the report will consist of currently.

  [Project Template](#)

- **Has everyone tracked their progress using the Excel file, and are we submitting it along with the PDF?**

  [Progress Tracker](#)

- **Does the submission meet all the project requirements, and is it ready for review by stakeholders?**

  Yes.