Project Iteration 2

Name: Pavan Kumar

Professor: Nafa, Fatema

1. Project Kickoff

Specific Goals: The goal of this project is to build a Spotify data pipeline that analyzes the "Top 100 Hits" playlist and generates a Power BI dashboard for data visualization. We aim to handle large-scale data processing, clean and transform the data, and produce actionable insights through visual reports.

Project Scope: The scope includes setting up cloud integration, ETL processes, data warehousing, and visualizing the data in Power BI. Scope creep will be avoided by sticking strictly to playlist data and limiting reporting to key insights such as trends, song popularity, and streaming behavior. Non-essential features, such as deep artist analysis or historical comparison, are excluded.

Deliverables:

Phase 1: Cloud setup and Spotify API integration.

Phase 2: Data extraction and transformation pipeline.

Phase 3: Power BI Dashboard creation.

Milestones and Deadlines:

Phase 1: Complete by October 25.

Phase 2: Complete by November 10.

Phase 3: Complete by November 20.

Team Capabilities: Our team's expertise aligns well with the project goals. I will handle cloud integration, ETL, and Power BI visualization, while Ankur will focus on QA, API integration, and data processing. However, we lack some advanced Power BI skills for more sophisticated visualizations, which could cause minor delays.

Dataset: The dataset is readily available through Spotify's API. We will be using their publicly available "Top 100 Hits" playlist data.

2. Team Discussions

Core Skills:

Pavan: Expertise in cloud integration, ETL, and data visualization using Power BI.

Ankur: Skilled in API integration, QA, and data cleaning.

Task Assignment:

I will handle the data pipeline, cloud integration, and visualizations.

Ankur will focus on ensuring data quality, API data extraction, and report generation.

Missing Skills: We need to improve our Power BI and data transformation skills, which could cause challenges in generating advanced reports.

Tools: We have experience with cloud platforms (Azure/AWS), Python for ETL, and basic Power BI. However, we need to learn more advanced Power BI features for better insights.

Languages and Platforms: We will use Python for ETL and API integration, Power BI for dashboarding, and cloud services for storage and processing.

3. Skills & Tools Assessment

External Resources: We will consult online Power BI resources and tutorials to cover gaps in visualization skills.

Tools/Frameworks: Python (ETL), Power BI (visualization), Azure (cloud storage and processing), and Spotify API (data source).

Comfort Level: Each member is comfortable with Python and API integrations. However, advanced Power BI techniques will require additional learning.

Task Assignments: Based on individual strengths, tasks are assigned as follows:

I will lead ETL and visualization.

Ankur will focus on data extraction and QA.

4. Initial Setup

Development Environment: We have set up an AWS cloud environment, configured version control (GitHub), and initialized a Python-based ETL process.

Version Control: GitHub is configured, and both team members have access to the repository.

Software and Libraries: Python, Spotify API, Power BI, and AWS cloud services have been installed and configured. The necessary Python libraries (requests, pandas, etc.) are installed.

Testing: We've run basic tests on Spotify API connectivity and confirmed that data can be extracted. Power BI is functioning properly with test data.

Troubleshooting: In case of setup issues, we will refer to the error logs, re-check API configurations, and use GitHub Issues for collaborative troubleshooting.

5. Progress Review

Achievements:

Completed cloud setup, version control, and initial Spotify API integration.

Data extraction from Spotify is functional, and the team is on track with the initial setup phase.

Blockers:

The advanced Power BI features are causing minor delays, but we are addressing this by focusing on learning.

Team Contribution: Each member is contributing as expected, and both are aligned with their roles and the project goals.

Timeline Review: We are on track with the current milestones, but some delays may occur due to Power BI challenges.

Alignment: Progress aligns well with project objectives, although slight adjustments to deadlines may be necessary.

6. Plan Revision

Timeline Adjustment: Given the current Power BI learning curve, we may push the final dashboard deadline by a few days.

Task Reassignment: If necessary, Pavan will focus more on Power BI while Ankur assists with additional data processing.

Communication Strategy: We will use daily check-ins to monitor progress and avoid misunderstandings. We also plan to communicate in-person or via WhatsApp.

Tracking: We will continue tracking progress through GitHub Issues and task boards.

7. Submission for This Iteration

Documented Tasks: This iteration includes the project setup, version control configuration, Spotify API integration, and initial Power BI configuration. Challenges related to Power BI learning have been documented, along with solutions and task adjustments.

Data Access: Spotify API is the data source; there is no public dataset link since the data is fetched dynamically.