

Determine Business Objectives

- Background
 1. Spotify, as a leading music and podcast streaming platform, faces several challenges in its ecosystem. While it boasts an extensive catalog of music from established artists, it also hosts a multitude of smaller, independent artists who struggle to gain visibility and reach wider audiences. These emerging artists often face the following issues:
 - Lack of Exposure: Smaller artists find it challenging to compete for attention in a sea of content. Their songs may get lost amidst popular tracks, limiting their chances of being discovered by listeners.
 - Data-Driven Decision Making: In today's digital age, data plays a crucial role in decision-making. However, many independent artists lack the resources or expertise to analyze their streaming data effectively. They miss out on valuable insights that could help them tailor their strategies and grow their audience.
 - Monetization and Sustainability: For independent artists, streaming revenue is a significant source of income. Maximizing their monthly streams directly impacts their financial sustainability. Yet, without data-driven guidance, they may struggle to optimize their music for better performance.
- Business Activity & Category
 - Category
 - Improve Strategy: Allows the Artist's Team at Spotify to optimize artists' presence and improve popularity.
 - Activity
 - D03: Improve identification of opportunities
- OKR (Objectives and Key Results)
 - Objective
 - Empower independent artist at Spotify to increase their popularity on the platform
 - Key Result
 - At least 50% of artists who work with the Artist Team at Spotify will achieve a 20% increase, on average, in popularity by the end of Q4
 - OMTM
 - Popularity
- OKR Initiatives
 - Develop and deploy a descriptive analytics dashboard to recommend music attributes based on current popular trends.
 - Implement a diagnostic analytics dashboard to identify areas for improvement for artists.
 - Build and automate a data pipeline to extract, transform, and load data from Spotify's API.
 - Develop a predictive model to surface growing artists, genres, and music trends.

Assess Situation

- Inventory of Resources
 - Personnel
 - Christy - Data Analyst
 - Jada - Data Engineer
 - Sebastian - Data Scientist
 - Leia - Legal Manager
 - Brenda - Data Governance
 - Christian - Artist Team Consultant (sales)
 - Lily - Machine Learning Engineer
 - Keith - Privacy and Audit Manager
 - Data
 - Billboard 100 API
 - Spotify Web API ([Web API | Spotify for Developers](#)) through Spotipy library ([Welcome to Spotipy! — spotipy 2.0 documentation](#))
 - Billboard charts data through billboard-charts ([GitHub - guoguo12/billboard-charts: Python API for downloading Billboard charts](#))
 - Computing Resources
 - AWS RDS
 - Software
 - MySQL
 - Tableau Desktop
 - Python (Jupyter Notebook)
- Requirements, Assumptions, and Constraints
 - Access to Spotify API is reliable
 - Access to Billboard 100 API is reliable
 - Data quality from the API is consistent
 - No budget for additional software or resources
- Risks and Contingencies
 - API changes affecting data access
 - Limit to the number of requests to Spotify's API
 - Legal concerns around training data that will be used to train models and support competitor artists
 - Skill gaps within the team
 - Storage restraints with AWS RDS
- Terminology
 - ETL: Extract, Transform, Load
 - API: Application Programming Interface
 - Billboard Hot 100: a chart that ranks the most popular songs in the United States, regardless of performer
 - Track Features

- Acousticness: This measures the likelihood that a track is acoustic. A score of 1.0 indicates high confidence that the track is acoustic, while 0.0 indicates low confidence.
- Danceability: This feature describes how suitable a track is for dancing based on factors like tempo, rhythm stability, beat strength, and overall regularity. It ranges from 0.0 (least danceable) to 1.0 (most danceable).
- Energy: Energy measures the intensity and activity level of a track. Higher values represent tracks that are more energetic, while lower values represent calmer and quieter tracks. The range is from 0.0 to 1.0.
- Instrumentalness: This feature predicts whether a track contains no vocals. A score close to 1.0 suggests that the track is instrumental, while a score close to 0.0 indicates that the track likely contains vocals.
- Liveness: Liveness detects the presence of an audience in the recording. A value above 0.8 suggests that the track is likely to be a live recording, while values below 0.8 indicate a studio recording.
- Loudness: Loudness measures the overall loudness of a track in decibels (dB). The values typically range from -60 dB to 0 dB.
- Speechiness: This feature detects the presence of spoken words in a track. Values above 0.66 suggest that the track is probably entirely spoken words, while values below 0.33 suggest that it's likely music and instrumental content.
- Tempo: Tempo measures the overall estimated tempo of a track in beats per minute (BPM). The range can vary widely depending on the music genre but typically falls between 50 BPM and 220 BPM.
- Valence: Valence measures the overall positivity or happiness of a track. Higher values represent tracks with a more positive mood, while lower values represent tracks with a more negative mood. The range is from 0.0 to 1.0.
- Costs and Benefits
 - Costs based on a typical hourly rate for each role identified under personnel and the expected hours required to complete the project
 - Costs
 - Personnel costs
 - Christy - Data Analyst: \$77,195 yr
 - Jada - Data Engineer: \$120,000 yr
 - Sebastian - Data Scientist: \$142,000 yr
 - Leia - Legal Manager: \$85,352 yr
 - Brenda - Data Governance: \$155,831yr
 - Christian - Artist Team Consultant (sale): \$80,000 yr
 - Lily - Machine Learning Engineer: \$142,000yr
 - Data Access to Spotify's API
 - Upfront costs: hiring the data team to conduct the MDM project, software costs, training costs
 - Unexpected costs: if the project takes longer than expected -> higher time

- Benefits
 - Artists can generate \$57,600 in revenue annually
 - 1M streams a month with a 20% increase can improve monthly revenue by \$4,8000 therefore increasing annual revenue to \$57,600
 - Spotify can indirectly generate \$576,00 annually (est. is from just one artist)
 - Assuming that Spotify can make 10x artist revenue through ad space and retention
 - Provide opportunities for Artists to increase their popularity
 - Improved efficiency in the Artist Team's recommendations
 - Improved user experience through discovery
 - Benefits based on the Data Periodic Table's Business Activity Categories. Try to quantify as best as you can, i.e. for labor hours saved. If you cannot quantify the benefit, still identify the benefit.

Determine Data Mining Goals

- Data Mining Goals aka Enabling Activities aka Data Activities - Identify **ALL** relevant data activities from the Data Periodic Table
 - F11: Increased understanding and insight of customers
 - F22: More accessible data
 - F17: Improved use of data science models
- Data Mining Success Criteria
 - Successfully develop and deploy a new artist data-driven recommendation system that increases new artists' popularity on Spotify.

Produce Project Plan

- Initial Assessment of Tools (Technology) and Techniques (Analytics)
- Data Toolkit Items
 - Descriptive analytics dashboard
 - Diagnostic analytics dashboard
 - Spotify API
 - Billboard API
 - Recommendation model
- Project Plan
 - Sprint 1
 - Data Acquisition
 - Set up API access and define data extraction procedures
 - ETL
 - Extract, transform, and load the data from python, redshift, and then to our MySQL database
 - Data Preparation
 - Clean the data

- Develop the descriptive analytics dashboard
- Sprint 2
 - ETL
 - Automate the pipeline
 - Diagnostic Analytics
 - Develop the dashboard
 - Predictive model
 - Build a recommendation system that will support the growth of small artist
 - A/B Test
 - A/B Test the UX/UI of the recommender
 - Create a Data Manifesto
- Work will be documented in this [Trello board](#).