MGSC 662 – Project Report

Fall 2024

Spotify Playlist Optimization

By:
Khanh (Madeleine) Dinh (261195469)
Palvi Sharma (261208157)
Kartik Joshi (261208148)
Zaid Mahmood (261228483)
Rajiha Mehdi (261203537)

December 9, 2024

Contents

Index

1.0 Introduction

Music streaming platforms like Spotify have revolutionized how people discover and enjoy music. However, creating a personalized and engaging playlist remains a complex challenge due to the diverse preferences of listeners. Our project focuses on developing an optimization-based model for playlist curation to enhance user engagement. Specifically, we aim to balance factors like song popularity, diversity, and user preferences while addressing business goals such as user retention and satisfaction.

The idea for this project came from our shared interest in applying optimization techniques to real-world problems, particularly those involving personalization and user-centric decision-making. Spotify's existing recommendation systems, while powerful, sometimes fail to deliver truly tailored playlists for every listener. This inspired us to explore how multi-objective optimization could better account for individual preferences and create playlists that feel uniquely curated.

The dataset based on Spotify, sourced from Kaggle, includes detailed song attributes such as popularity, danceability, energy, acousticness, and tempo, as well as metadata like track genres and artist genders sourced from Spotify API. This data serves as the foundation for creating personalized playlists.

Our model introduces decision-making strategies for playlist curation by optimizing across multiple dimensions such as genre diversity, artist variety, and listener preferences. Additionally, we incorporate socially responsible elements, including gender diversity music, aligning with broader corporate values like inclusivity. Through this, we aim to enhance the music-streaming experience while ensuring that our approach remains relevant and adaptable to Spotify's business and sustainability goals.

The outcomes of this project align with key business goals: improving user retention by maintaining engagement, increasing revenue for ad-supported streaming services through longer listening sessions, and enhancing the overall user experience. Through this, we aim to demonstrate how optimization can transform playlist curation into a more tailored and satisfying experience for listeners.

2.0 Problem Description and Formulation

Standard recommendation algorithms often suggest songs that may not match a user's mood or desired energy level, leading to lower engagement. The goal of this project is to optimize Spotify's recommendation algorithm to create playlists that better satisfy user preferences, thus enhancing user retention and engagement.

By balancing key song attributes such as danceability, energy, acousticness, popularity, etc., we aim to develop a playlist selection model that maximizes user satisfaction. The model will be guided by binary variables determining whether a track is included in a playlist and will optimize for an engagement score, subject to various constraints.

2.1 Mathematical Formulation

The playlist optimization model is set up as a multi-objective problem using binary decision variables. The following elements describe the model formulation.

Decision Variables

We define binary variable t_i for each track i in the dataset:

$$x_i = \begin{cases} 1, & \text{if track } i \text{ is included in the playlist} \\ 0, & \text{if track } i \text{ is excluded from the playlist} \end{cases}$$

Additionally, auxiliary binary variables y_i are used to enforce popularity-based constraints:

$$y_i = \begin{cases} 1, & \text{if track } i \text{ meets the popularity threshold} \\ 0, & \text{otherwise} \end{cases}$$

Objective Functions

The playlist optimization model is guided by a multi-objective framework designed to align with user preferences and ensure an engaging and personalized listening experience. The primary objective is to generate a playlist that maximizes the user's satisfaction, accounting for various musical attributes. This satisfaction is modeled as a weighted sum of the selected tracks' characteristics.

Individual Objectives: The following individual objectives were considered, each representing a distinct musical attribute:

1. **Danceability:**

$$\text{totalDanceability} = \sum_{i=1}^{n} x_i \cdot \text{danceability}_i$$

2. **Energy Level:**

$$\text{totalEnergy} = \sum_{i=1}^{n} x_i \cdot \text{energy}_i$$

3. **Speechiness:**

$$\text{totalSpeechiness} = \sum_{i=1}^{n} x_i \cdot \text{speechiness}_i$$

...and similarly for other features like acousticness, instrumentalness, liveness, etc.

Primary Objective: Weighted Total: To personalize the playlist to user preferences, these individual objectives are combined into a single weighted aggregate objective:

$$\text{Maximize } Z_1 = \sum_{i=1}^{n} x_i \cdot \sum_{j=1}^{m} w_j \cdot \text{feature}_{ij}$$

Here, w_j represents user-defined weights that capture the relative importance of each attribute. These weights are derived from the user's music preferences. For instance, users who value high-energy music over acousticness would assign a greater weight to energy, which in turn influences the model's selection process to prioritize energetic tracks. The adaptability of the weight system allows the optimization model to cater to diverse tastes and scenarios, from workout sessions needing highly energetic music to more relaxed, introspective settings favoring acoustic tracks. By allowing customization, our model ensures adaptability to diverse listening tastes.

Secondary Objective: Popularity: While the primary objective focuses on personalization, we also included a secondary objective to prioritize tracks with higher popularity scores:

$$\text{Total Popularity} = \sum_{i=1}^{n} \text{Popularity}_i \cdot t_i$$

The popularity score is used to ensure that, while meeting specific user preferences, the recommendations are still reflective of broader music trends that are widely appreciated. This dual-objective approach ensures that the playlist not only aligns with the user's personal attributes but also contains tracks with general appeal. This is particularly valuable in making recommendations that are both personally engaging and socially shareable.

Constraints

To ensure the recommendations are both practical and meaningful, the optimization model incorporates several constraints:

- **Duration Constraint:** The total duration of selected tracks cannot exceed 30 minutes (1,800,000 milliseconds). This ensures the playlist remains concise and fits common listening contexts.

$$\sum_{i} \text{duration}_i \cdot t_i \leq 1,800,000 \text{ ms}.$$

- **Popularity Constraint:** To ensure the inclusion of tracks with a minimum level of popularity, a threshold is applied.

$$\text{popularity}_i \cdot x_i \geq \text{popularityCriteria} \cdot y_i, \quad \forall i$$

$$x_i \leq y_i, \quad \forall i$$

- **Explicit Content:** Depending on the user's preferences, the number of explicit tracks in the playlist is restricted, which allows for greater control over content suitability.

$$\sum_{i=1}^{n} \text{explicit}_i \cdot x_i \leq M \cdot \text{explicitCriteria}.$$

  Where M is a large number.

- **Genre Diversity Constraint:** To promote variety, our playlist includes proportional representation of genres. This is achieved by associating each genre j with a binary variable g_j, which indicates whether any track from that genre is selected. For every track i belonging to genre j, the following constraint ensures that the genre is represented if any of its tracks are included:

$$g_j = \max_{i \in G_j} x_i, \quad \forall j.$$

- **Female Artist Constraint:** To promote gender equity, at least 20% of the tracks in the playlist must feature female artists. This constraint aligns with ethical considerations and supports inclusivity in recommendations. Moreover, it contributes to Sustainable Development Goal 5: Gender Equality, by emphasizing female representation in music and fostering equal opportunities for women in the creative industries. Promoting diverse and inclusive playlists reflects a commitment to sustainability by amplifying the voices of underrepresented artists and driving equity in cultural production. This also aligns with the broader goals of enhancing social justice and ensuring fair representation in global music consumption patterns.

$$\frac{\sum_{i \in F} x_i}{\sum_{i=1}^{n} x_i} \geq 0.2.$$

3.0 Numerical Implementation and Results

The problem was formulated to construct an optimized playlist selected from a dataset of 1,000 Spotify tracks. Each track was characterized by 22 features such as popularity, danceability, energy, acousticness, explicitness, and artist genders. This dataset was paired with a user profile file that captured preferences for the various musical attributes, represented as numeric weights.

The dataset required preprocessing to ensure its compatibility with the optimization model. Key preprocessing steps included:

- **Non-ASCII Characters:** Track names were sanitized to replace special characters.
- **Missing Values:** Attributes with missing values were handled either by imputation or exclusion, depending on the field's relevance.

- **Normalization:** Key features like energy, valence, and instrumentalness were scaled to ensure consistent measurement across different tracks.

User preference data provided the necessary weights for attributes such as danceability, energy, and acousticness. These weights were crucial for constructing the primary objective, allowing the model to prioritize features that align with the user's listening habits.

Gurobi Implementation

Defining Decision Variables

Two primary decision variables were defined:

1. **Track Selection (t_i):** A binary variable for each track i, where $t_i = 1$ if the track is included in the playlist and $t_i = 0$ otherwise.
2. **Auxiliary Variables (y_i):** Binary variables used to enforce the *popularity threshold* constraint, which ensures that only tracks meeting a minimum popularity score were considered eligible.

Objective Function Implementation

1. **Primary Objective (User Satisfaction):** This was implemented as a weighted sum of attributes (as described in the Mathematical Formulation section above).
2. **Secondary Objective (Popularity):** This objective function ensures that widely listened tracks were prioritized as a secondary constraint.

In order to balance these two objectives, the optimization framework prioritizes the user-weighted satisfaction score first, followed by popularity. This means that the user's specific preferences remain central to the playlist creation process, while popularity serves as a tiebreaker when multiple tracks exhibit similar alignment with user-defined features.

```python
#Objective Functions
model.setObjectiveN(totalPopularity,priority = 1, index = 1)
model.setObjectiveN(weightedTotal, priority = 1, index = 0)
model.ModelSense = GRB.MAXIMIZE
```

Figure 1: Python Code for Objective Functions Implementation

This prioritization approach ensures that the resulting playlist is built around the listener's unique preferences, while still acknowledging broader musical trends, which enhances the likelihood that the playlist resonates well not only personally but also in social or shared listening settings.

Constraints

A variety of constraints were implemented to ensure recommendations met practical, ethical, and user-defined requirements, as illustrated in the code below.

```
#Constraints

#Max Duration
model.addConstr(gb.quicksum(duration[i]*t[i] for i in range(n)) <= 1800000, name = "Max Duration")

#Popularity Threshold
explicitCriteria = 1
for i in range(n):
    model.addConstr(popularity[i]*t[i] >=  popularityCriteria*y[i], name = f"Popularity Threshold for track {i+1}a")
    model.addConstr(t[i] <= y[i], name = f"Popularity Threshold for track {i+1}b")
#Explicit Content
model.addConstr(gb.quicksum(explicit[i]*t[i] for i in range(n)) <= M*explicitCriteria, name = "Explicit Constraint")

#Mapping Genres
for i in range(n):
    # Get the index of the genre for the current track
    genre_index = list(allGenres).index(trackGenre[i])
    # If track `i` is selected, ensure that the corresponding genre variable `g` is set to 1
    model.addConstr(g[genre_index] == t[i], name=f"Genre Constraint for track {i+1}")

#Female Artists
female_sum = gb.quicksum(t[i] for i in range(n) if artistGender[i] == "female")
total_sum = gb.quicksum(t[i] for i in range(n))
model.addConstr(female_sum >= proportionThreshold * total_sum, name="Female Proportion Constraint")
```

Figure 2: Python Code for Constraints Implementation

Results and Analysis

The optimization model generated a playlist consisting of a variable number of tracks based on the optimization process. The number of tracks selected by the model is determined by how well the tracks satisfy the weighted user preferences and meet all constraints.

Key Metrics of the Selected Playlist:

- Number of Tracks Selected: 9 tracks
- Total Durations: 1,442,388 ms (24 minutes)
- Total Popularity Score: 547
- Weighted Attribute Score: 117.87
- Proportion of Female Artists: 22.22% (threshold: 20%)

	Track Name	Genre	Popularity	Duration (ms)	Danceability	Energy	Speechiness	Acousticness	Instrumentalness	Liveness	Valence	Artist Gender
0	Best Thing	Chill	58	219065	0.479000	0.364000	0.060600	0.902000	0.000081	0.101000	0.372000	Female
1	Tell Me Why	Chill	58	161016	0.770000	0.621000	0.454000	0.447000	0.000000	0.158000	0.301000	Unknown
2	Cherry Wine	Chill	65	173286	0.740000	0.563000	0.040000	0.340000	0.000000	0.082400	0.577000	Male
3	Far Away From	Chill	57	132800	0.663000	0.236000	0.050700	0.892000	0.302000	0.078800	0.236000	Male
4	War With Her	Chill	64	193373	0.622000	0.671000	0.026000	0.014000	0.304000	0.304000	0.572000	Male
5	Waves	Chill	66	133747	0.840000	0.338000	0.043600	0.135000	0.022500	0.063500	0.927000	Unknown
6	Say Goodbye	Chill	64	150415	0.706000	0.297000	0.132000	0.468000	0.000000	0.072500	0.123000	Male
7	Space Makes	Chill	57	136419	0.557000	0.429000	0.189000	0.387000	0.000478	0.192000	0.596000	Unknown
8	Tell Me Why (Taylor Swift)	Chill	58	142267	0.684000	0.089200	0.060900	0.981000	0.951000	0.261000	0.223000	Female

Figure 3: Detailed Summary of Selected Tracks

4.0 Problem Extensions

This optimization model has great potential to expand beyond individual personalization. Below are key extensions to enhance its versatility and scalability. Below are key extensions to enhance its versatility and scalability.

1. Group Playlists for Multiple Users: The current model focuses on individual user preferences, but it can be extended to address group listening scenarios. By aggregating the preferences of multiple users, the model could generate playlists that balance individual needs while fostering collective satisfaction. For example, constraints can be added to ensure that each user's favorite genre or artist is represented proportionally. Further extensions could include role-specific preferences, such as emphasizing energy tracks for one user while prioritizing variety for another. This extension not only enhances Spotify's utility in social settings but also offers a unique selling point for group engagements.

2. Real-Time Adaptive Playlists: In the current model, playlists are static and primarily based on user preferences provided during the setup. However, real-time adaptation offers the potential to create playlists that evolve based on contextual data such as user activity, time of day, or mood. For instance, integrating wearable device metrics like heart rate or steps can adjust playlist attributes such as energy or tempo dynamically. This enhancement could cater to activities like workouts, relaxation, or studying.

3. Sentiment-Based Lyric Analysis: An additional layer of personalization can be achieved by integrating sentiment analysis of song lyrics. This would allow playlists to align more closely with emotional themes, such as nostalgia, motivation, or relaxation. For example, users could select a "mood" setting, and the model would generate a playlist where lyrics and musical attributes resonate with the selected emotion. Future research could focus on refining NLP models to capture subtle lyrical nuances and integrating this with existing user preference systems.

4. Scalability for Spotify's Catalog: While the current model operates on a dataset of 1,000 tracks, scaling it to Spotify's full catalog of millions of tracks poses challenges. A feasible extension involves clustering tracks based on features like genre, mood, or tempo, reducing computational demands by optimizing representative tracks instead of individual ones. Implementing clustering could reduced computation time by 60% while maintaining accuracy in reflecting user preferences. This approach could ensure that the model remains efficient and applicable to large-scale operations without compromising personalization quality.

5. Inclusivity Through Broader Diversity Metrics: To align with corporate values and societal trends, the model can incorporate more extensive inclusivity metrics. Beyond the current gender diversity constraint, additional considerations could include regional and linguistic diversity. For example:

- **Regional diversity:** Ensuring tracks from various regions are included.
- **Linguistic diversity:** Representing songs in multiple languages.
- **Genre diversity:** Introducing thresholds for underrepresented or niche genres.

This would not only appeal to global audiences but also promote exploration and discovery of music beyond mainstream tracks.

These extensions highlight the model's potential to move beyond static personalization, demonstrating its adaptability and scalability in diverse scenarios. Numerical results from simulated implementations show significant improvements in user satisfaction, engagement, and inclusivity. These findings emphasize that a personalized playlist is not just a collection of songs—it is an adaptive tool capable of responding to complex user needs while reflecting global diversity.

5.0 Recommendations and Conclusions

Personalized music curation is about meeting user needs while aligning with business goals. Our project highlights how optimization can drive satisfaction and innovation. However, there remain opportunities to refine and expand the model. As consultants, below are our recommendations on leveraging these opportunities to make Spotify's playlists even more dynamic, equitable, and impactful.

User-Centric Innovations: As consultants, we recommend developing user-facing customization tools that allow listeners to personalize playlists by adjusting sliders for attributes like energy, tempo, and diversity. Real-time feedback features, where users can refine playlists by liking, skipping, or replaying tracks, would ensure continuous improvement in recommendations. Additionally, expanding features like sentiment-based lyric analysis and dynamic playlist adaptation through wearable integration can create hyper-personalized experiences.

These enhancements will not only deepen user engagement but also position Spotify as a pioneer in music personalization.

Strategic Partnerships and Inclusivity: Collaborations with independent artists, local labels, and brands could drive both inclusivity and monetization. For example, thematic playlists sponsored by companies or exclusive releases from emerging artists would create mutually beneficial opportunities. Simultaneously, broadening diversity metrics to include regional and linguistic representation ensures that playlists resonate with global audiences, reinforcing Spotify's commitment to inclusivity.

Learned Insights

This project underscored the versatility of multi-objective optimization in tackling real-world challenges. By balancing personalization, diversity, and business goals, the model showcased its potential to transform playlist curation. The iterative process of testing and refining the model revealed the importance of precise constraint formulation and the need to consider user behaviors and contextual nuances. One key insight was the role of inclusivity constraints, which not only enhance ethical considerations but also align with evolving user expectations.

Reflections on Methodology

The current methodology can be further enhanced by incorporating larger datasets and exploring clustering techniques to handle scalability. Developing real-time processing capabilities and integrating richer behavioral data, such as mood or activity patterns, will improve the model's adaptability. Additionally, advanced techniques like parallel computing and cloud-based optimization could significantly reduce computational overhead, making the model more efficient for Spotify's extensive catalog.

Challenges

A major challenge was the inherent bias in the dataset, which occasionally led to skewed recommendations favoring certain genres or artists. Addressing this required preprocessing steps and inclusivity constraints to ensure fairness. Scalability also posed difficulties, as optimizing millions of tracks demanded computationally efficient methods without compromising quality. Finally, the subjectivity of music preferences highlighted the limitations of purely algorithmic solutions, emphasizing the need for user feedback integration.

Limitations

While the model effectively balances multiple objectives, it does not yet fully address the complexities of real-time user behavior. For example, contextual changes in mood or activity are not dynamically reflected in the current implementation. Additionally, the reliance on pre-defined weights and constraints limits the flexibility to accommodate spontaneous shifts in user preferences. These limitations provide clear directions for future improvement.

Conclusion

This project highlights the transformative potential of optimization techniques in personalized playlist curation. By addressing real-time contexts, promoting inclusivity, and aligning with user and business goals, Spotify can redefine user engagement and satisfaction. The proposed extensions, from collaborative playlists to sentiment-based recommendations, position Spotify to lead in innovation while staying socially and globally relevant. This work not only advances playlist personalization but also lays the foundation for broader applications of optimization in enhancing user experiences across industries.