Increasing Spotify's User Engagement Through Weather Based Recommendation

Fall 2022 MGMT 59000 - 011 LEC

Team

Thannir Malai
Annamalai Kumar

Nikhitha
Siddi

Yi Chen
Chu

Agenda

- Business Problem
- Process Flow & Analysis
- Recommendations
- Limitations & Further Analysis
- Appendix

Business Problem | Spotify currently doesn't have a weather based recommendation.

Current Recommendation System

- Content-based filtering- Ex- Genre, style, danceability

- Collaborative Filtering- Ex- User's past interest & listening habits

Spotify Users Pain Point

Spotify Community reviews highlight lack of recommendations based on local weather

Competitors In This Space

- YouTube Music

- Weatherify: Recommends music and songs based on the weather.

Question:

- Users are looking for weather-based recommendations. But will the addition of weather-based section in recommendation, increase the Spotify's user engagement significantly?

Process Flow

Data Collection

Kaggle:
- Spotify Europe dataset

Web Scraping:
- Parent-genre & sub-genre map
- Spotify user's pain points

Data Cleaning & Pre-Processing

- Cleaning reviews to identify pain points

- Categorizing weather condition

- Map parent-genre to sub-genre

- **Metric:** number of streams

Data model & Hypothesis Testing

Text Analysis:
To identify top user pain points

Two tailed t-test:
To prove "average number of streams for different parent genres is significantly different for diverse weather"

Lift Analysis

Analyzing the lift or dip in average of number of streams for each genre vs weather

Data Collection & Pre-Processing | European Spotify dataset was processed to analyze impact of weather on genre of songs played

Kaggle Dataset :
- Level of data: Date - Region – Song track – Number of streams
- Region level climate information for Europe (rain, temperature, cloud, humidity etc.), sub-genre of each song etc.

Web Scraping:
- Scraped Spotify parent genre & the sub genres under each to map it out
- Scraped Spotify Community ongoing issues

Weather Condition Flags: Based on research
- Humid Day: Humidity >55%
- Cloudy Day: >62.5% opaque clouds (5/8)
- Sunny Day: <37.5% opaque clouds (3/8)
- Rainy Day: 1/0 flag (in dataset)
- Snowy Day: 1/0 flag (in dataset)

Data Collection & Pre-Processing | Top 6 parent genre contributed to 95% of total streams.
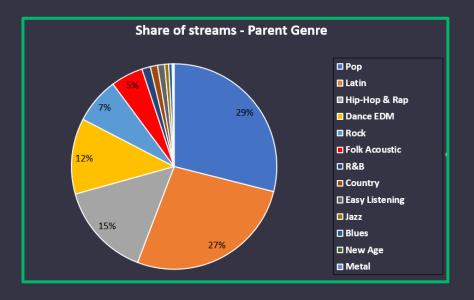
Parent-Genre & Sub-Genre Mapping:
- 537 sub-genre available were mapped to 15 parent-genre (Pop, Rock, Latin etc.)

- Of the 15 parent-genre, top 6 contributed to 95% of the streams and were chosen for further analysis (Pop, Latin, Hip-Hop& Rap, Dance EDM, Rock, Folk Acoustic)

Share of streams - Parent Genre

- Pop — 29%
- Latin — 27%
- Hip-Hop & Rap — 15%
- Dance EDM — 12%
- Rock — 7%
- Folk Acoustic — 5%
- R&B
- Country
- Easy Listening
- Jazz
- Blues
- New Age
- Metal

Text Analysis | Pain point analysis using user reviews results in "Recommendations" (13% weightage).

<u>Word cloud based on the frequency of occurrence of words:</u>

- Most frequent words include- Song, spotify, artist, genre, listen, recommendation, playlist, podcast

- Suggests that recommendations is a significant pain point in the ongoing issues

<u>Further analysis of reviews concerning recommendations:</u>

- Most frequent words include- playlist, new tracks, weather, listen, album

- Suggests that recommendations concerning weather may have potential to improve user experience

Text Analysis | Weather-based albums, playlists, songs are frequently requested by users
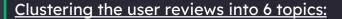
Clustering the user reviews into 6 topics:

- Clustering using LDA to classify ongoing issues data to various topics.

- Topic 5 also carries a high weightage as it ranks 3 of the 6 clusters formed

	Word 0	Word 1	Word 2	Word 3	Word 4
Topic 0	spotify	song	music	time	like
Topic 1	new	playlist	release	indie	new music
Topic 2	playlist	release	new	indie	daily
Topic 3	spotify	song	app	playlist	account
Topic 4	pop	answer	maybe	accept	appreciate
Topic 5	playlist	album	song	http	weather

	Topic Num	Num Documents
2	5	44
4	4	13
1	3	113
3	2	14
5	1	12
0	0	115

Hypothesis Testing | Test if weather significantly impacts the genre of songs played by users.

Hypothesis to be tested:
Weather significantly impacts the genre of songs played by users.

Sample Hypothesis: On a snowy day, if average streaming of Pop songs is significantly different than average streaming of Non-Pop songs.

- Null Hypothesis: Difference in number of streams between Pop and Non-Pop songs on a snowy day is the same as on a non-snowy day

- Alternate Hypothesis: Difference in number of stream between Pop and Non-Pop songs on a snowy day is significantly different than on a non-snowy day

Metrics	Snowy Day	Not Snowy Day
Average Difference in number of streams of Pop and Non-Pop	-3,85,087	-3,53,773
Observations	3,729	3,548
Hypothesized Mean Difference	0	
df	7,275	
t Stat	2.832287536	
P(T<=t) two-tail	0.004634328	
t Critical two-tail	1.960290123	

Result:
- We can conclude with 95% confidence that, the difference in number of streams between Pop and Non-Pop songs on a snowy day is significantly different than on a non-snowy day.

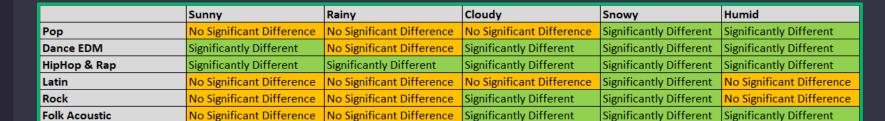- This establishes the significance of Pop songs on a snowy day

Hypothesis Testing | Daily weather conditions significantly impact the genre of songs played by the users

To establish significance, two tailed t-test was done for each genre-weather combination.

	Sunny	Rainy	Cloudy	Snowy	Humid
Pop	No Significant Difference	No Significant Difference	No Significant Difference	Significantly Different	Significantly Different
Dance EDM	Significantly Different	No Significant Difference	Significantly Different	Significantly Different	Significantly Different
HipHop & Rap	Significantly Different	Significantly Different	Significantly Different	Significantly Different	Significantly Different
Latin	No Significant Difference	No Significant Difference	No Significant Difference	Significantly Different	No Significant Difference
Rock	No Significant Difference	No Significant Difference	Significantly Different	Significantly Different	No Significant Difference
Folk Acoustic	No Significant Difference	No Significant Difference	Significantly Different	Significantly Different	Significantly Different

Sunny Weather

- Dance EDM
- Hip-Hop & Rap

Rainy Weather

- Hip-Hop & Rap

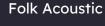## Cloudy Weather

- Dance EDM
- Hip-Hop & Rap
- Rock
- Folk Acoustic

Snowy Weather

All the genres have a significant difference in the number of streams

Humid Weather
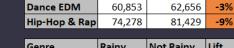
- Pop
- Dance EDM
- Hip-Hop & Rap
- Folk Acoustic

Lift Analysis | An average of 14% absolute lift in number of streams was observed for different weather conditions in Europe

After proving significance, the lift/dip in number of streams for each genre was studied for difference in weather conditions.

Genre	Sunny	Not Sunny	Lift
Dance EDM	60,853	62,656	-3%
Hip-Hop & Rap	74,278	81,429	-9%

Genre	Rainy	Not Rainy	Lift
Hip-Hop & Rap	84,912	71,210	16%
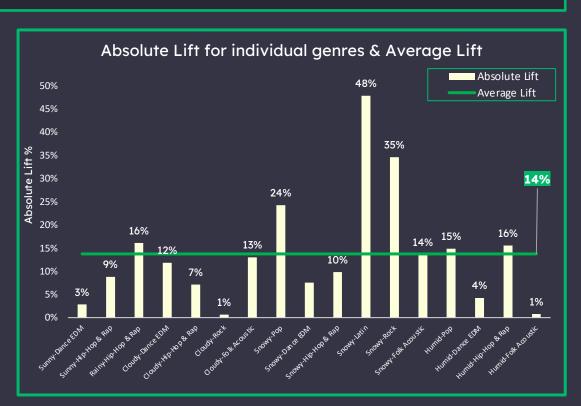
Genre	Cloudy	Not Cloudy	Lift
Dance EDM	68,338	60,260	12%
Hip-Hop & Rap	72,342	77,521	-7%
Rock	37,129	37,366	-1%
Folk Acoustic	29,448	25,612	13%

Genre	Snowy	Not Snowy	Lift
Pop	1,51,222	1,14,573	24%
Dance EDM	61,225	65,845	-8%
Hip-Hop & Rap	76,338	83,830	-10%
Latin	1,42,173	74,077	48%
Rock	38,042	24,870	35%
Folk Acoustic	26,382	22,789	14%

Genre	Humid	Not Humid	Lift
Pop	1,38,675	1,59,297	-15%
Dance EDM	62,850	60,165	4%
Hip-Hop & Rap	83,386	70,421	16%
Folk Acoustic	26,295	26,086	1%

Absolute Lift for individual genres & Average Lift

Impact & Recommendations | In Europe Weather based recommendation will increase user engagement by 1.67Bn minutes

Metric	Europe
Annual Users (Mn)	121
Average Daily Usage (mins)	99
Daily engagement (Mn mins)	11,979
Updated User engagement (Mn mins)	13,656
User engagement increase (Mn mins)	1,677

- Current daily user engagement in Europe is ~12Bn minutes

- The weather based recommendation potentially can increase user engagement by ~1.67Bn minutes

Central Recommendation

Add a section recommending local weather based playlist *(Customer Engagement)*

Peripheral Recommendation

Changing the color of user interface based on weather for power users* *(Customer Retention)*

Increased customer engagement and retention leads to increased probability of :
- Current premium user base to extend the subscription
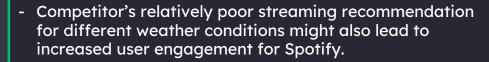- Non premium user base to subscribe for premium services

*Power users are users who fall in the top 10 percentile of user engagement distribution in Spotify

Limitations | Lack of statistical power to identify any causality and incapability to test the significance with A/B testing.

Causality

- Competitor's relatively poor streaming recommendation for different weather conditions might also lead to increased user engagement for Spotify.

- Unavailability of unique list of songs in each genre

Lack of Statistical Power

- Lack of hourly streaming data.
- Hourly streaming data will help better validate the significance of weather on genre of songs streamed.
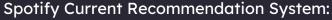
Perform A/B Testing

- Performing an A/B testing for test group (receiving weather based recommendation) vs control group will help assert that weather impacts the genre of songs streamed.

Appendix | References

Spotify Current Recommendation System:
- https://tinuiti.com/blog/performance-display/spotify-algorithm/#:~:text=Powered%20by%20AI%2C%20Spotify's%20algorithm,features%2C%20and%20past%20listening%20habits.
- https://www.music-tomorrow.com/blog/how-spotify-recommendation-system-works-a-complete-guide-2022
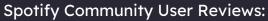
Classification of Weather:
- https://www.weather.gov/arx/why_dewpoint_vs_humidity#:~:text=less%20than%20or%20equal%20to,in%20the%20air%2C%20becoming%20oppressive
- https://www.mentalfloss.com/article/56820/whats-difference-between-mostly-sunny-and-partly-cloudy
- https://weather.com/science/weather-explainers/news/common-weather-terms-used-incorrectly

Spotify Community User Reviews:
- https://community.spotify.com/t5/forums/searchpage/tab/message?q=weather&noSynonym=false&page=4&collapse_discussion=true

Spotify User Statistics (Annual Users, Average Daily Usage):
- https://www.businessofapps.com/data/spotify-statistics/