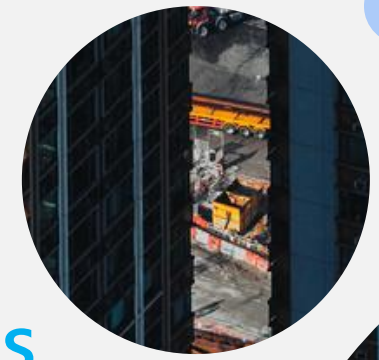
KKCOMPANY

Competition Report

Main speaker: An Syu Li

目錄

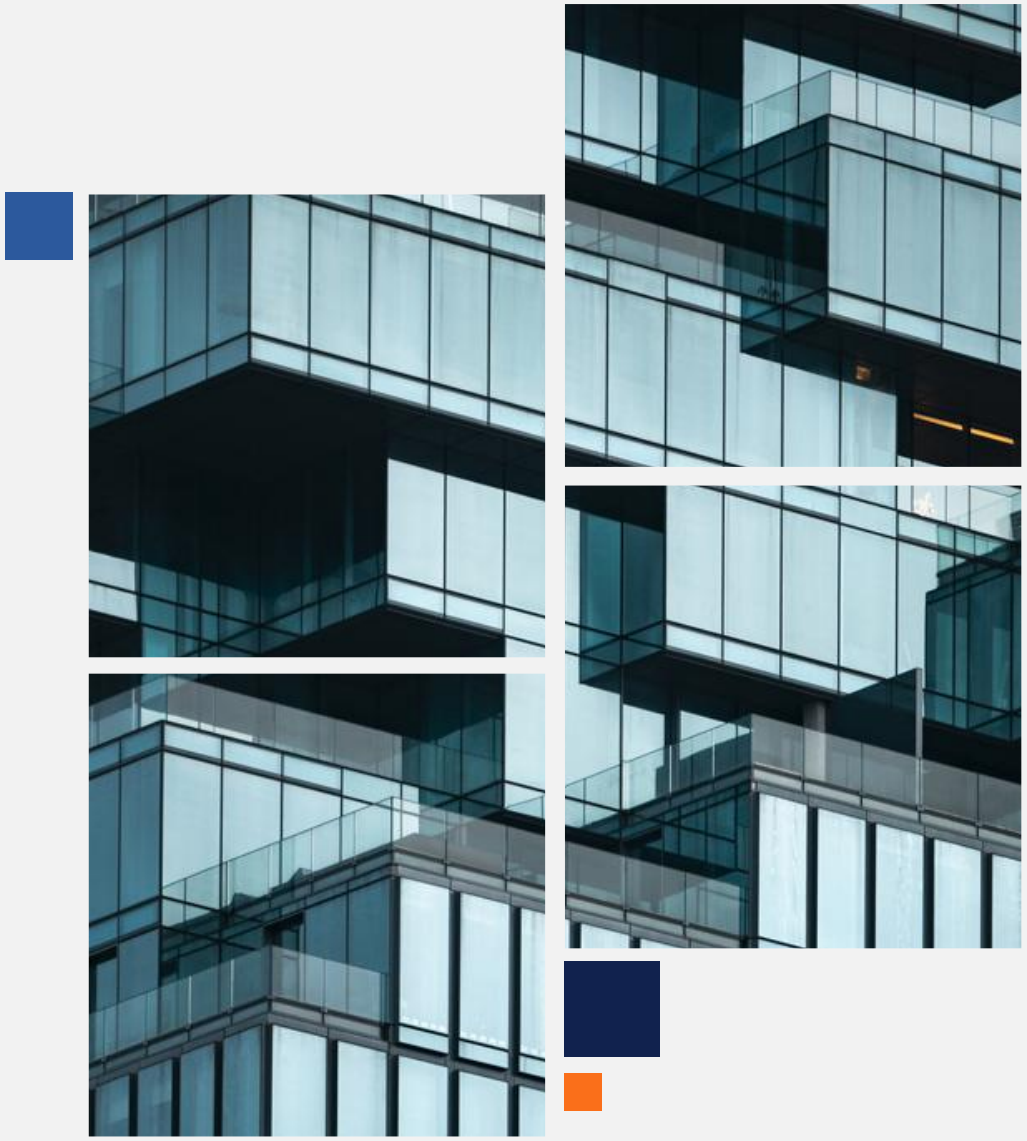


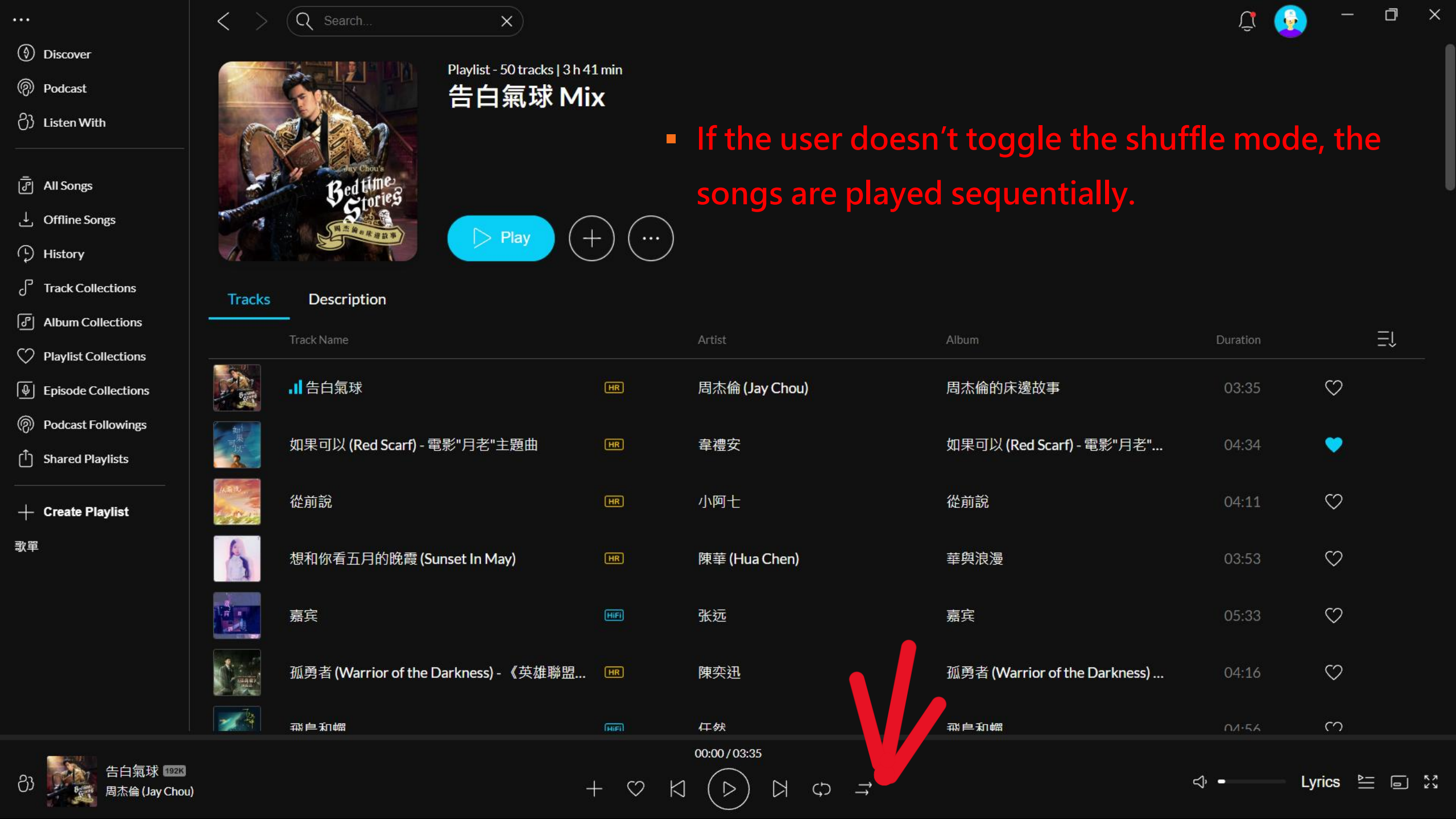
- User and Data Observations
- Feature Extraction
- Modeling
- Performance Analysis

User and Data Observations

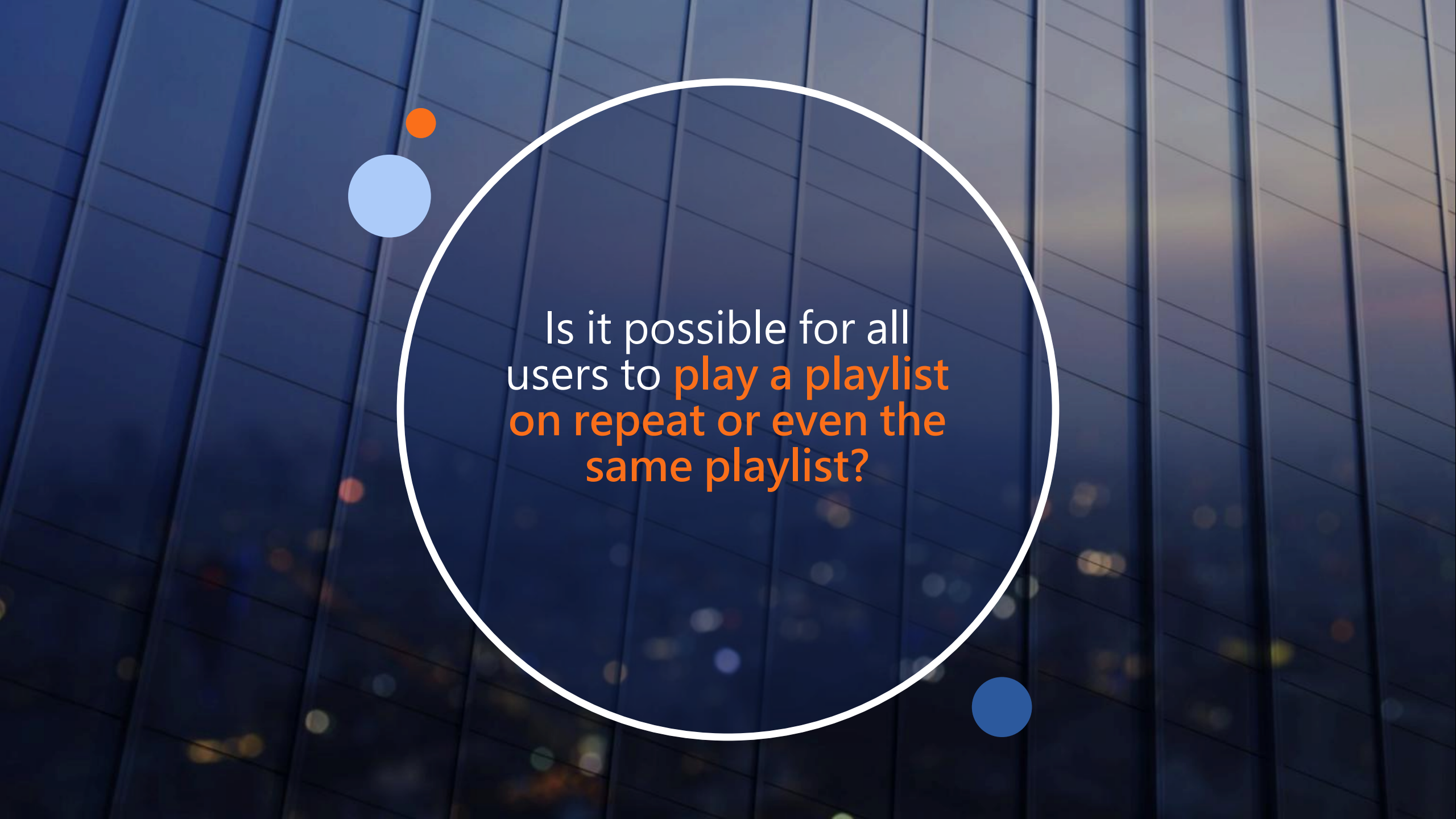
Ask about experience

- Me： How do you use a Music App
- Friend: Search for a song I like and play it
- Me: And, after the song ends…?
- Friend: Let the next song be played.





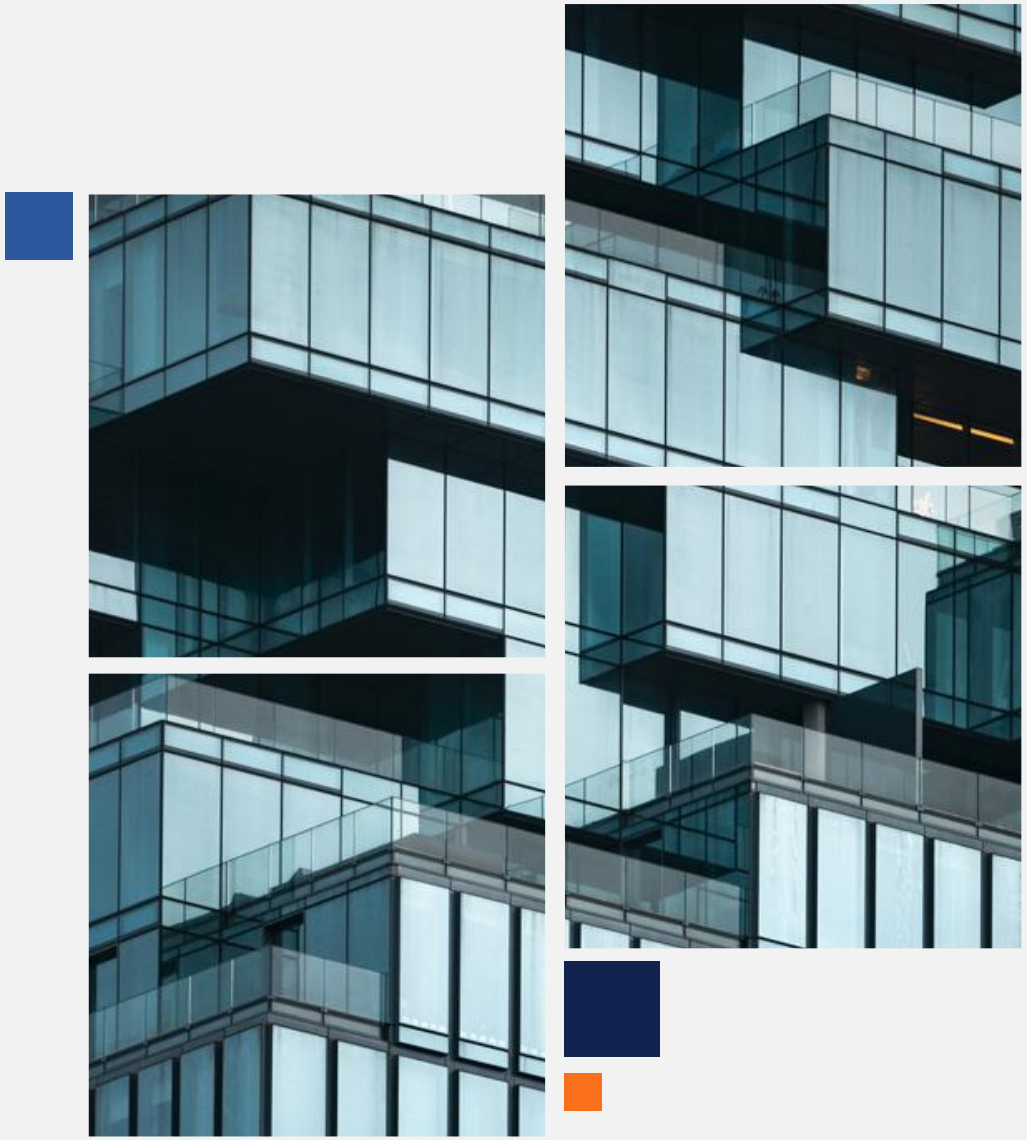
- If the user doesn't toggle the shuffle mode, the songs are played sequentially.


Is it possible for all users to play a playlist on repeat or even the same playlist?

Data Observation and Implementation -Part 1

Define the problem：The universal playlist?

- Sol： Find sessions where the **same songs are played**
- Implementation： If the first 20 songs in the testing set match those in the training set, **the last 5 songs from the training set will be used as predictions.**



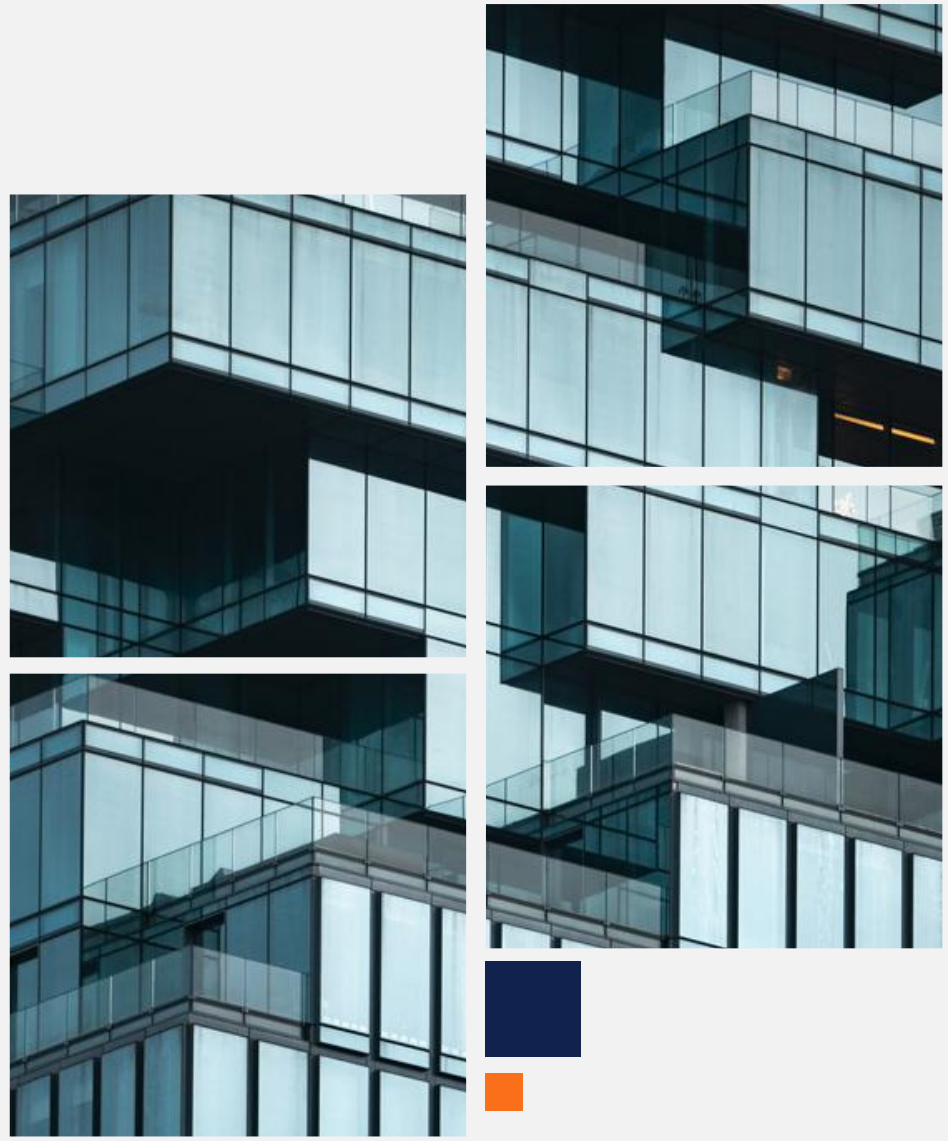
Score for this method

Submission and Description	Private Score	Public Score	Selected
same_df_withNoneFilled.csv Complete (after deadline) · Autoencoder0621 · 37...	0.15775	0.16242	☐

Data Observation and Implementation -Part 1

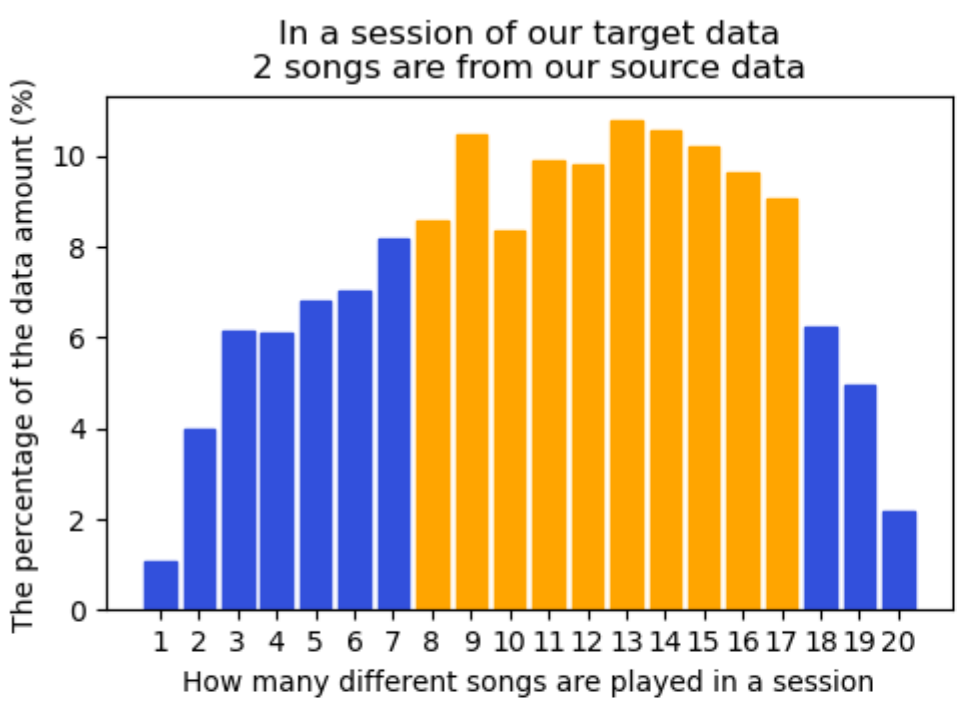
Define the problem：Who has many replays ?

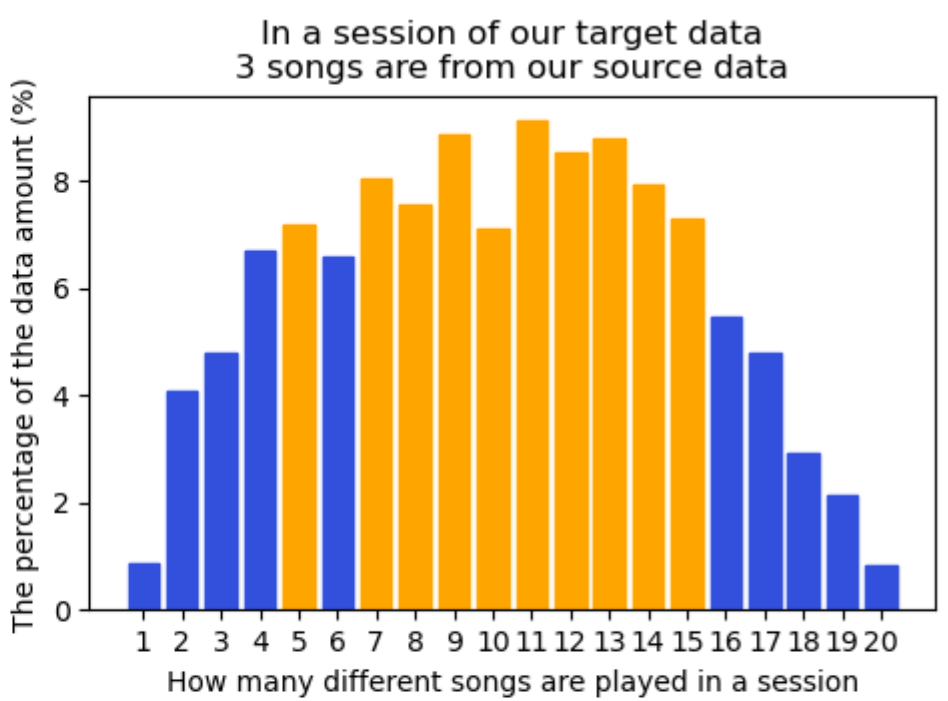
- Sol：Identify whether **taking 20 songs as predictions** is worth it
- Implementation：Clustering according to **the number of unique songs**

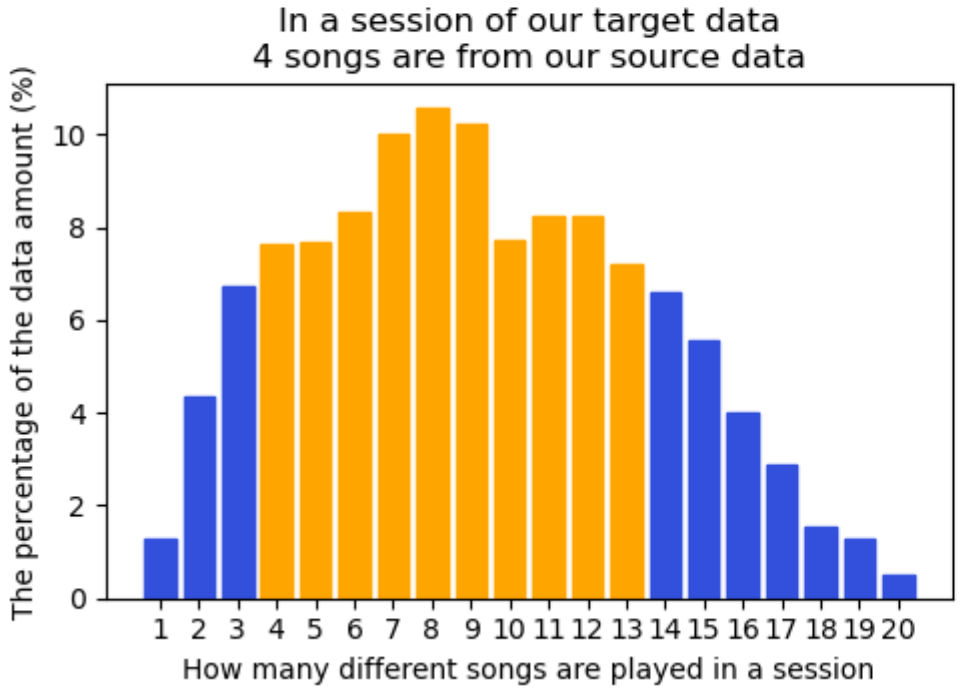


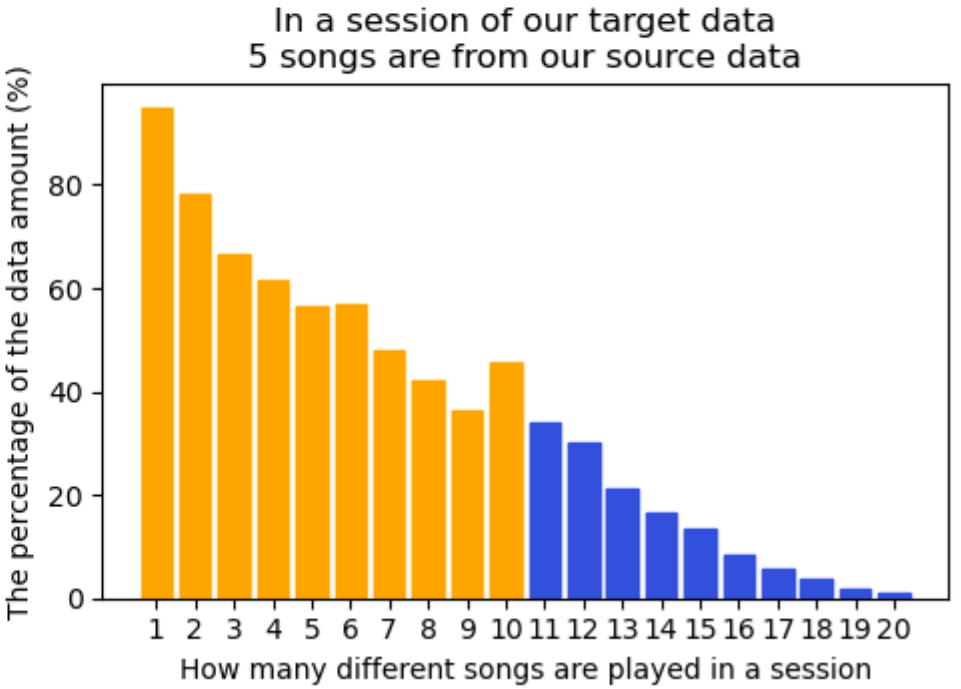
Steps

- Cluster according to the number of unique songs
- Calculate how many come from the first 20.
- Visualization


In a session of our target data
2 songs are from our source data
The percentage of the data amount (%)
0
2
4
6
8
10
1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20
How many different songs are played in a session

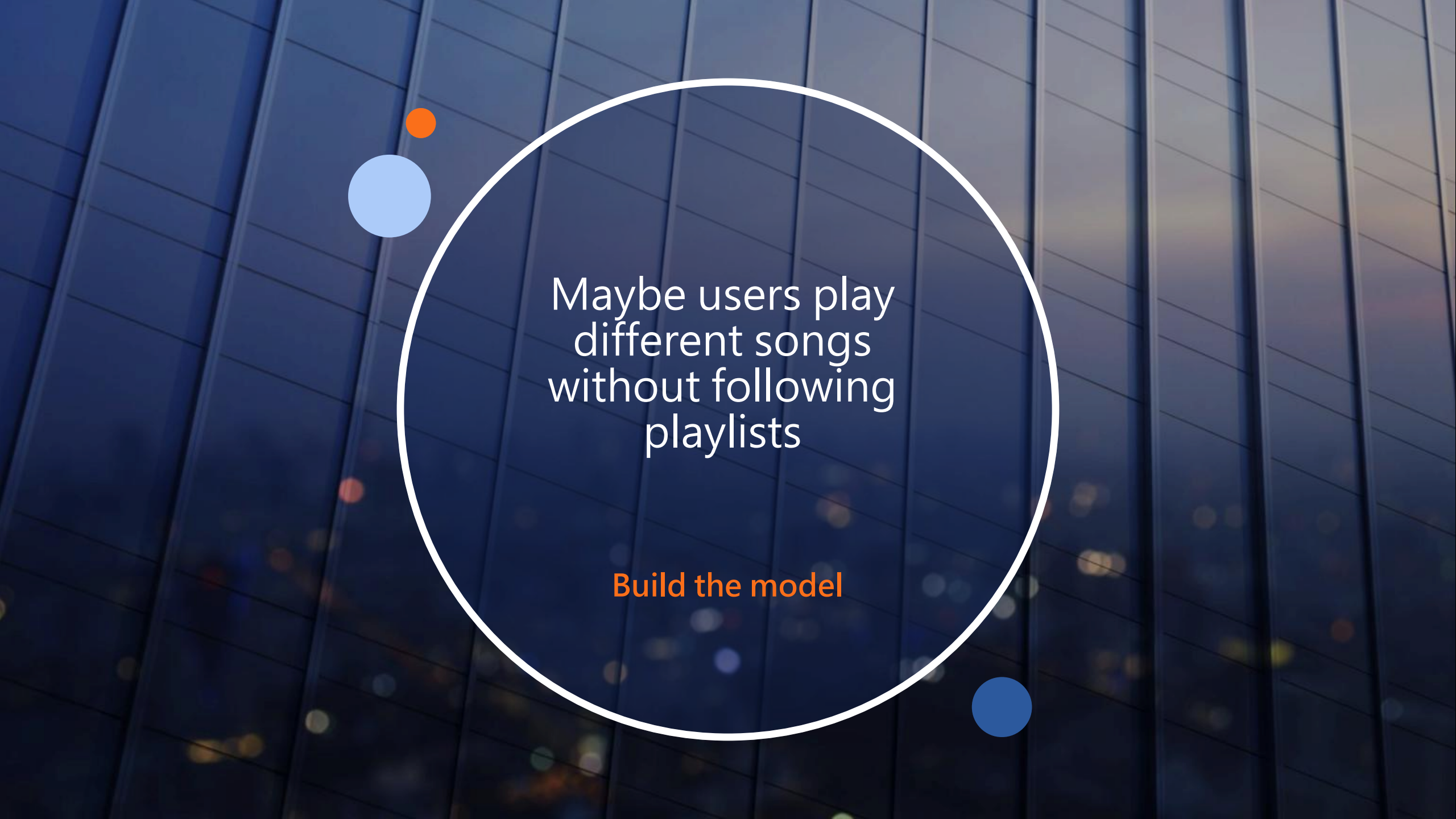

In a session of our target data
3 songs are from our source data
The percentage of the data amount (%)
0
2
4
6
8
1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20
How many different songs are played in a session


In a session of our target data
4 songs are from our source data
The percentage of the data amount (%)
0
2
4
6
8
10
1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20
How many different songs are played in a session


In a session of our target data
5 songs are from our source data
The percentage of the data amount (%)
0
20
40
60
80
1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20
How many different songs are played in a session

Result revealed for this method

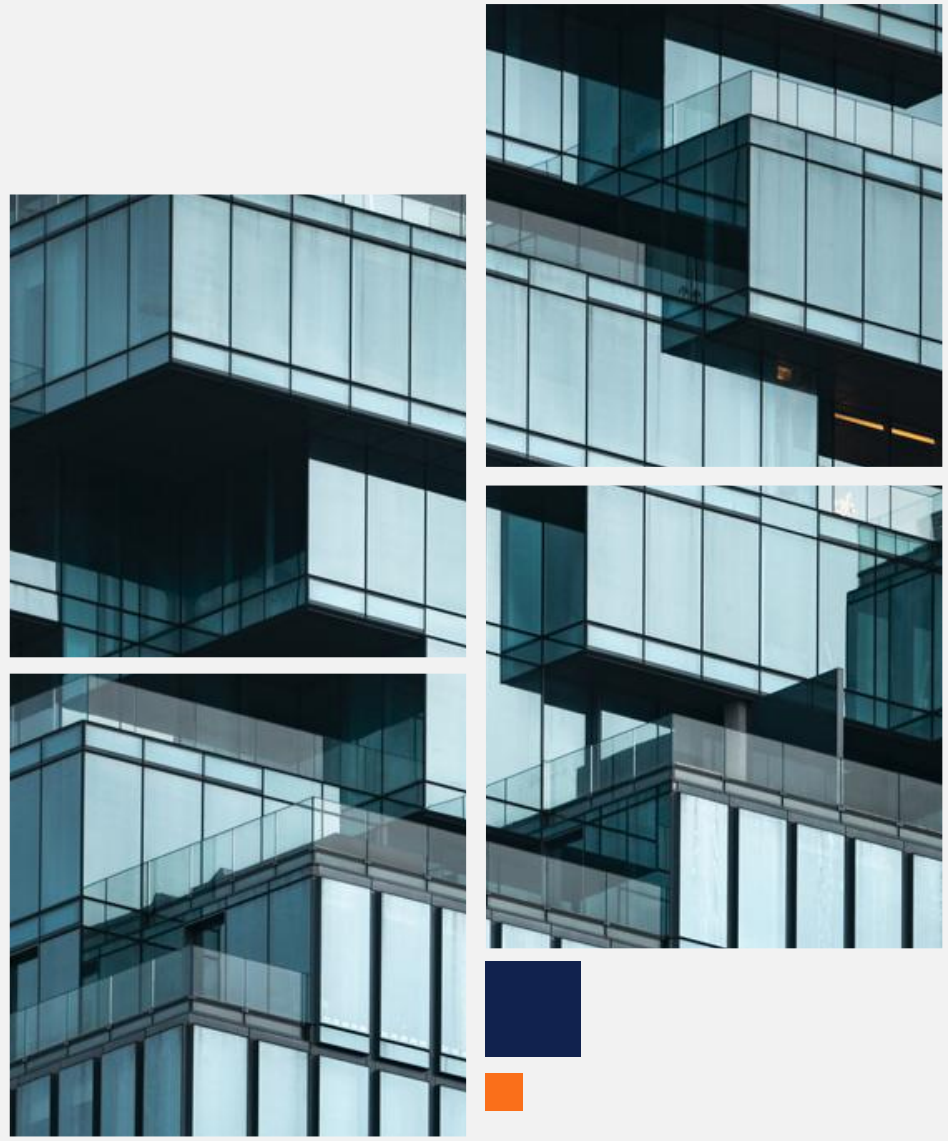
Submission and Description	Private Score	Public Score	Selected
onlyFrequentlyPlayed.c... Complete (after deadline) · Autoe...	0.14145	0.14403	☐


Maybe users play different songs without following playlists
Build the model

Feature Extraction

Define the problem：How do transform input ?

- Sol： TF-IDF
- Implementation：Transform the **first 20 songs' IDs** and their corresponding metadata

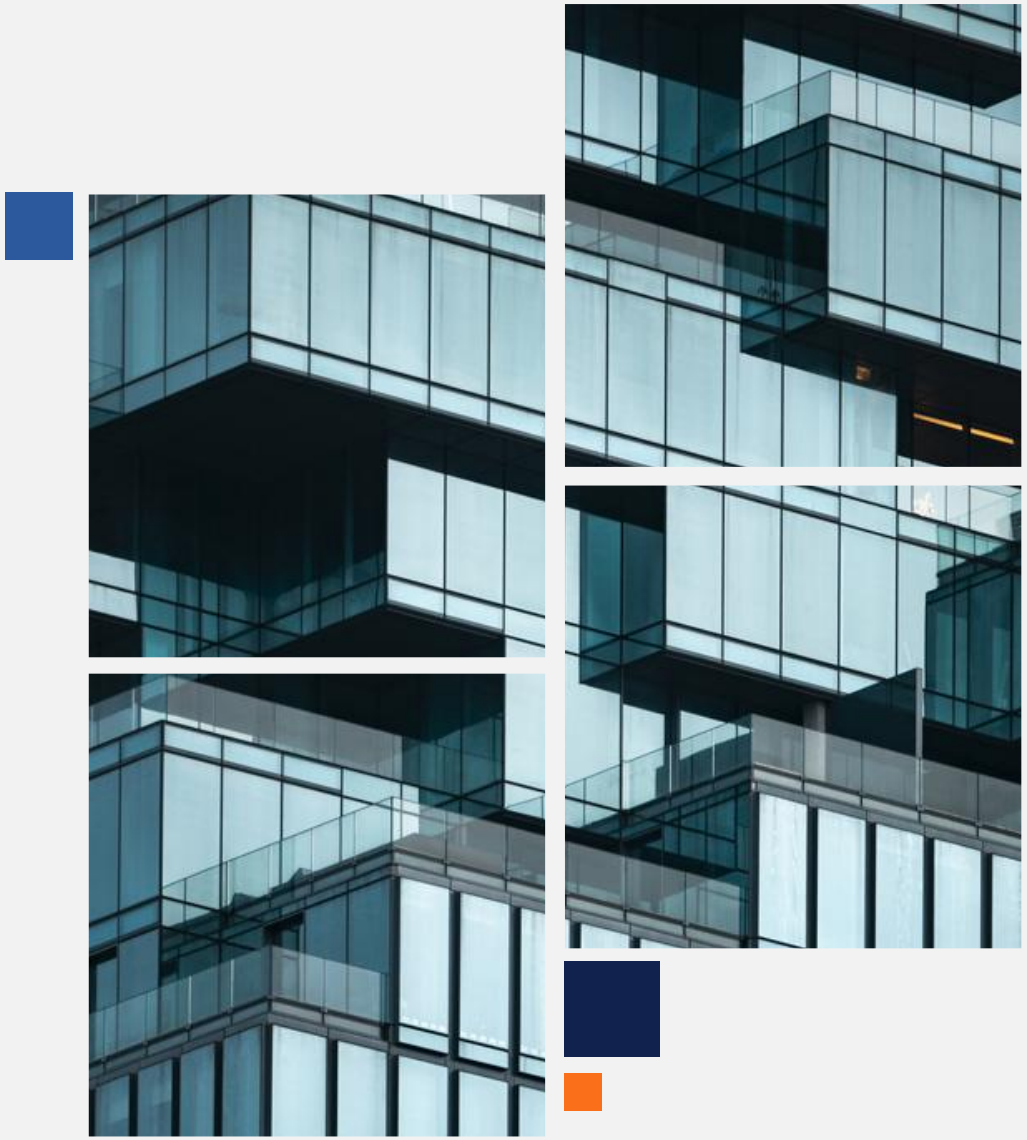


Modeling

Someone has to do the dirty job

Define the problem：High dimension and modeling ?

- Sol：sparse metric(for storage)
- Implementation： using KNN and cosine similarity to calculate distance without using the similarity matrix.



Performance Analysis

Method 1(predicting by matching)

Submission and Description	Private Score	Public Score	Selected
same_df_withNoneFilled.csv Complete (after deadline) · Autoencoder0621 · 37...	0.15775	0.16242	☐

Method 2(predicting by frequency)

Submission and Description	Private Score	Public Score	Selected
onlyFrequentlyPlayed.c... Complete (after deadline) · Autoe...	0.14145	0.14403	☐

Method 3(modeling)

Submission and Description	Private Score	Public Score	Selected
modelPredictingResult.csv Complete (after deadline) · Autoencoder0621 · ...	**0.06008**	**0.05931**	☐

The Amount of Data Used for different methods