Hit Song Project - Findings

By Harriet Peng

EDA - Summary

Hit songs on average have higher:

- Danceability
- Energy
- Loudness
- Musical Positiveness (Valence)
- Longer duration (duration_ms)
- Beats per minute (tempo)

And lower:

- Speechiness level
- Acousticness level
- Instrumentalness level
- The chorus would start early (chorus_hit)
- Number of sections (sections)

Model Performance

Great Model Performance:

- Preliminary XGB:
  - Average ROC-AUC: 0.850
  - Average Accuracy: 0.767
- Optimized XGB:
  - Average ROC-AUC: 0.857
  - Average Accuracy: 0.769

Feature Importance:

- In terms of Cover: **Sections** is the most important, followed by **Acousticness**, **Mode** and **Danceability**.
- In terms of Gain: **Instrumentalness** is the most important, followed by **Danceability**, **Acousticness** and **Speechiness**.
- In terms of Weight: **Duration_ms** is the most important, , followed by **Acousticness**, **Valence** and **Speechiness**.

Assessment of the data:

- Descriptive analysis shows a general picture of how hit songs look like;
- This Optimized XGB model can definitely help with hit song identification.