**HO CHI MINH UNIVERSITY OF SCIENCE**

**FACULTY OF INFORMATION TECHNOLOGY**

**BÁO CÁO ĐỒ ÁN CUỐI KỲ**

**Subject: Data Visualization Class:** 22KHDL  
**Group:** 5

**Students:**

Dương Nhật Huy – 22127148  
Nguyễn Đình Kiên - 22127216  
Trương Thuận Kiệt – 22127224 Phạm Minh Mẫn - 22127257  
Hồ Đăng Phúc - 22127492

**TP. Hồ Chí Minh, tháng 4, năm 2025**

Contents

[**1. Thông tin nhóm và phân công công việc 4**](#_Toc197781788)

[**2. Giới thiệu tập dữ liệu 5**](#_Toc197781789)

[**2.1. Thông tin cơ bản 5**](#_Toc197781790)

[**2.1.1. Vietnamese\_only\_songs\_cleaned 5**](#_Toc197781791)

[**2.1.2. position\_streams\_by\_time\_daily\_total\_peak\_cleaned and position\_streams\_by\_time\_weekly\_total\_peak\_cleaned 5**](#_Toc197781792)

[**2.1.3. Vietnamese\_artists 6**](#_Toc197781793)

[**2.2. Mô tả chi tiết 6**](#_Toc197781794)

[**2.2.1. Vietnamese\_only\_songs\_cleaned 6**](#_Toc197781795)

[**2.2.2. position\_streams\_by\_time\_daily\_total\_peak\_cleaned and position\_streams\_by\_time\_weekly\_total\_peak\_cleaned 7**](#_Toc197781796)

[**2.2.3. Vietnamese\_artists 7**](#_Toc197781797)

[**3. Khám phá và phân tích thống kê cho dữ liệu 8**](#_Toc197781798)

[**3.1. Vietnamese\_only\_songs\_cleaned 8**](#_Toc197781799)

[**3.1.1. Dữ liệu thiếu 8**](#_Toc197781800)

[**3.1.2. Phân tích thống kê 9**](#_Toc197781801)

[**3.2. position\_streams\_by\_time\_daily\_total\_peak\_cleaned and position\_streams\_by\_time\_weekly\_total\_peak\_cleaned 10**](#_Toc197781802)

[**3.2.1. Dữ liệu thiếu 10**](#_Toc197781803)

[**3.2.2. Phân tích thống kê 10**](#_Toc197781804)

[**3.3. Vietnamese\_artists 13**](#_Toc197781805)

[**3.3.1. Dữ liệu thiếu 13**](#_Toc197781806)

[**3.3.2. Phân tích thống kê 13**](#_Toc197781807)

[**4. Mối tương quan giữa các biến 14**](#_Toc197781808)

[**4.1. Vietnamese\_only\_songs\_cleaned 14**](#_Toc197781809)

[**4.2. position\_streams\_by\_time\_daily\_total\_peak\_cleaned and position\_streams\_by\_time\_weekly\_total\_peak\_cleaned 16**](#_Toc197781810)

[**4.3. Vietnamese\_artitsts 17**](#_Toc197781811)

[**5. Tương quan stat model 17**](#_Toc197781812)

[**5.1. Giới thiệu 17**](#_Toc197781813)

[**5.2. Phương pháp 18**](#_Toc197781814)

[**5.3. Stream\_counts vs các biến còn lại 18**](#_Toc197781815)

[**5.4. Popularity vs các biến còn lại 20**](#_Toc197781816)

[**6. Trực quan hóa dữ liệu 21**](#_Toc197781817)

[**6.1. Tích hợp AI trong việc thiết kế dashboard 21**](#_Toc197781818)

[**6.2. Dashboard tổng quan 25**](#_Toc197781819)

[**6.2.1. Tổng thể dashboard 25**](#_Toc197781820)

[**6.2.2. Các Chỉ số Chính (KPIs): 25**](#_Toc197781821)

[**6.2.3. Xu hướng theo Thời gian: 25**](#_Toc197781822)

[**6.2.4. Phân bố: 26**](#_Toc197781823)

[**6.2.5. Bộ lọc: 26**](#_Toc197781824)

[**6.2.6. Kết luận: 26**](#_Toc197781825)

[**6.3. Songs dashboard 26**](#_Toc197781826)

[**6.3.1. Tổng thể dashboard 26**](#_Toc197781827)

[**6.3.2. Các nhận xét biểu đồ tổng quát: 27**](#_Toc197781828)

[**6.3.3. Các chi tiết về một số bài hát nổi bật (lấy các bài Dù cho tận thế, Chuyện đôi ta). 27**](#_Toc197781829)

[**6.3.4. Bộ lọc: 28**](#_Toc197781830)

[**6.3.5. Kết luận: 28**](#_Toc197781831)

[**6.4. Artists dashboard 29**](#_Toc197781832)

[**6.4.1. Tổng thể dashboard 29**](#_Toc197781833)

[**6.4.2. So sánh tổng quát: 29**](#_Toc197781834)

[**6.4.3. Các phân tích chi tiết (xét các ca sĩ nổi bật như Sơn Tùng, HIEUTHUHAI): 30**](#_Toc197781835)

[**6.4.4. Bộ lọc: 31**](#_Toc197781836)

[**6.4.5. Kết luận: 31**](#_Toc197781837)

[**6.5. Dashboard thể loại 32**](#_Toc197781838)

[**6.5.1. Tổng thể dashboard 32**](#_Toc197781839)

[**6.5.2. Các Chỉ số Chính (KPIs): 32**](#_Toc197781840)

[**6.5.3. Phân tích chi tiết theo Thể loại: 32**](#_Toc197781841)

[**6.5.4. Bộ lọc: 33**](#_Toc197781842)

[**6.5.5. Kết luận: 33**](#_Toc197781843)

[**7. Tích hợp AI trong việc phân tích 33**](#_Toc197781844)

[**7.1. Hỏi đáp 33**](#_Toc197781845)

[**7.2. Chatbot 33**](#_Toc197781846)

[**8. Tích hợp AI trong việc đánh giá dashboard 34**](#_Toc197781847)

[**8.1. Dashboard Tổng quan 34**](#_Toc197781848)

[**8.2. Dashboard Bài hát 35**](#_Toc197781849)

[**8.3. Dashboard Thể loại 35**](#_Toc197781850)

[**8.4. Dashboard Nghệ sĩ 35**](#_Toc197781851)

[**8.5. Đánh giá tổng quát 36**](#_Toc197781852)

1. **Thông tin nhóm và phân công công việc**

|  |  |  |  |
| --- | --- | --- | --- |
| MSSV | Họ và Tên | Công việc | Mức độ hoàn thành |
| 22127148 | Dương Nhật Huy | * Trực quan hóa Summary Dashboard * Thiết kế, trang trí các dashboard * Tích hợp AI | 100% |
| 22127216 | Nguyễn Đình Kiên | * Trực quan hóa Genres Dashboard * Viết báo cáo * Làm powerpoint | 100% |
| 22127224 | Trương Thuận Kiệt | * Thu thập dữ liệu * Tiền xử lý dữ liệu * Trực quan hóa Songs Dashboard * Viết báo cáo | 100% |
| 22127257 | Phạm Minh Mẫn | * Thu thập dữ liệu * Trực quan hóa Artists Dashboard * Tích hợp AI | 100% |
| 22127492 | Hồ Đăng Phúc | * Trực quan hóa Genres Dashboard * Viết báo cáo * Làm powerpoint | 100% |

1. **Giới thiệu tập dữ liệu**
   1. **Thông tin cơ bản**
      1. **Vietnamese\_only\_songs\_cleaned**

* Nguồn gốc: Thu thập từ [Spotify API](https://developer.spotify.com/documentation/web-api) dành cho developers
* Chủ đề: Thông tin chi tiết về các bài hát tiếng việt và các bài hát ở thị trường Việt Nam

A screenshot of a web api

Description automatically generated

* + 1. **position\_streams\_by\_time\_daily\_total\_peak\_cleaned and position\_streams\_by\_time\_weekly\_total\_peak\_cleaned**
* Nguồn gốc: Thu thâp từ [Kworb](https://kworb.net/), một trang web về xếp hạng cũng như thông tin về 1 bài hát hoặc 1 nghệ sĩ.
* Chủ đề: Xếp hạng và số lượng phát cao nhất trong ngày và tuần.

A screenshot of a computer

Description automatically generated

* + 1. **Vietnamese\_artists**
* Nguồn gốc: Thu thập từ Spotify API dành cho developers
* Chủ đề: Thông tin về các nghệ sĩ Việt Nam như tên, độ phổ biến, …
  1. **Mô tả chi tiết**
     1. **Vietnamese\_only\_songs\_cleaned**
* Mỗi dòng diễn tả thông tin về các bài hát bao gồm: tên, thể loại và thuộc album nào.
* Ý nghĩa của các cột

|  |  |
| --- | --- |
| **Cột** | **Ý nghĩa** |
| Id | Mã Spotify của mỗi bài hát. |
| Album | Tên album mà bài hát xuất hiện. |
| Album\_type | Loại album. Giá trị cho phép: "**album**", "**single**", "**compilation**". |
| Album\_popularity | Độ phổ biến của album. Giá trị nằm trong khoảng từ 0 đến 100, với 100 là phổ biến nhất. |
| Artists | Tên các nghệ sĩ biểu diễn bài hát |
| Disc\_number | Số đĩa. |
| Duration\_ms | Độ dài của bài hát. |
| Explicit | Bài hát có chứa lời bài hát tục tĩu hay không. |
| Name | Tên bài hát. |
| Popularity | Độ phổ biến của bài hát. Giá trị nằm trong khoảng từ 0 đến 100, với 100 là phổ biến nhất.  Độ phổ biến của bài hát là một giá trị từ 0 đến 100, với 100 là phổ biến nhất. Độ phổ biến được tính bằng thuật toán và chủ yếu dựa trên tổng số lần phát bài hát và thời gian gần đây của các lần phát đó.  Nói chung, các bài hát đang được phát nhiều hiện tại sẽ có độ phổ biến cao hơn so với các bài hát được phát nhiều trong quá khứ. Các bài hát trùng lặp (ví dụ: cùng một bài hát từ một đĩa đơn và một album) được đánh giá độc lập. Độ phổ biến của nghệ sĩ và album được suy ra từ độ phổ biến của bài hát thông qua tính toán.  **Lưu ý: giá trị độ phổ biến có thể chậm hơn so với độ phổ biến thực tế vài ngày: giá trị này không được cập nhật theo thời gian thực.** |
| Track\_number | Số thứ tự của bài hát trong album. Nếu album có nhiều đĩa, số thứ tự bài hát là số trên đĩa được chỉ định. |
| Release\_date | Ngày album được phát hành lần đầu tiên. |
| Genres | Thể loại của bài hát. |
| Stream\_counts | Tổng số lần bài hát được phát trên Spotify. |

### **position\_streams\_by\_time\_daily\_total\_peak\_cleaned and position\_streams\_by\_time\_weekly\_total\_peak\_cleaned**

* Mỗi dòng diễn tả thông tin về sự thể hiện của các bài hát trên thị trường Việt Nam.
* Ý nghĩa của các cột:

|  |  |
| --- | --- |
| **Cột** | **Ý nghĩa** |
| Id | Mã Spotify của từng bài hát. |
| TotalStream | Tổng số lượt phát theo tuần/ngày. |
| PeakStream | Lượt phát cao nhất theo tuần/ngày. |
| PeakPosition | Vị trí cao nhất theo tuần/ngày. |

### **Vietnamese\_artists**

* Mỗi dòng diễn tả thông tin về nghệ sĩ Việt Nam ở thị trường Việt Nam.
* Ý nghĩa của các cột:

|  |  |
| --- | --- |
| **Cột** | **Ý nghĩa** |
| Id | Mã Spotify của từng nghệ sĩ. |
| Name | Tên của nghệ sĩ. |
| Followers | Số lượt theo dõi của nghệ sĩ. |
| Genres | Thể loại âm nhạc của nghệ sĩ. |

1. **Khám phá và phân tích thống kê cho dữ liệu**
   1. **Vietnamese\_only\_songs\_cleaned**

A screenshot of a computer

Description automatically generated

* Bộ dữ liệu chứa tổng cộng 4846 dòng và 14 cột.
* Trong đó có cột stream\_counts có mất mát.
* Chuyển cột release\_date về datetime.

### **Dữ liệu thiếu**

|  |  |  |
| --- | --- | --- |
| **Cột** | **Số lượng thiếu** | **Tỷ lệ (%)** |
| Stream\_counts | 29 | 0.598 |

* + Ở đây, ta có thể thấy được các giá trị thiếu là vô cùng ít, đồng thời ở đây ta cũng biết được các giá trị thiếu là do khi thu thập một số bài hát có **số lượng nghe < 1000** nên sẽ không có.
  + Và để giữ sự đa dạng của các thuộc tính nên nhóm sẽ giữ lại các dòng này.

### **Phân tích thống kê**

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
|  | **album\_popularity** | **disc\_number** | **duration\_ms** | **popularity** | **track\_number** | **stream\_counts** |
| count | 4846.000000 | 4846.000000 | 4.846000e+03 | 4846.000000 | 4846.000000 | 4.817000e+03 |
| mean | 25.919934 | 1.003714 | 2.381399e+05 | 27.985349 | 3.120718 | 4.121705e+06 |
| std | 15.573992 | 0.060839 | 6.362212e+04 | 14.618266 | 3.771611 | 1.772154e+07 |
| min | 0.000000 | 1.000000 | 1.952800e+04 | 0.000000 | 1.000000 | 1.027000e+03 |
| 25% | 14.000000 | 1.000000 | 1.990928e+05 | 18.000000 | 1.000000 | 1.545430e+05 |
| 50% | 25.000000 | 1.000000 | 2.328505e+05 | 27.000000 | 1.000000 | 6.609920e+05 |
| 75% | 37.000000 | 1.000000 | 2.725415e+05 | 38.000000 | 4.000000 | 2.770964e+06 |
| max | 83.000000 | 2.000000 | 1.813776e+06 | 80.000000 | 41.000000 | 8.532404e+08 |

* + **album\_popularity:**
    - Trung bình: 25.92, với độ lệch chuẩn là 15.57.
    - Giá trị dao động từ 0 đến 83.
    - Phần lớn các album có độ phổ biến thấp (25% dưới 14), nhưng có một số album rất phổ biến (max = 83).
  + **disc\_number:**
    - Hầu hết các bài hát nằm trên đĩa đầu tiên (median = 1, max = 2).
    - Điều này cho thấy dữ liệu chủ yếu liên quan đến các album đơn giản, không có nhiều đĩa.
  + **duration\_ms:**
    - Thời lượng trung bình của các bài hát là khoảng 238 giây (2.38e+05 ms).
    - Thời lượng dao động từ 19.5 giây (min = 1.95e+04 ms) đến khoảng 30 phút (max = 1.81e+06 ms).
    - Phần lớn các bài hát có thời lượng trong khoảng 199 giây đến 273 giây (25% đến 75%).
  + **popularity:**
    - Trung bình: 27.98, với độ lệch chuẩn là 14.62.
    - Giá trị dao động từ 0 đến 80.
    - Phần lớn các bài hát có độ phổ biến thấp (25% dưới 18), nhưng có một số bài hát rất phổ biến (max = 80).
  + **track\_number:**
    - Trung bình: 3.12, với độ lệch chuẩn là 3.77.
    - Số thứ tự bài hát dao động từ 1 đến 41.
    - Phần lớn các bài hát nằm ở vị trí đầu tiên trong album (median = 1).
  + **stream\_counts:**
    - Trung bình: 4.12 triệu lượt stream, nhưng độ lệch chuẩn rất lớn (17.72 triệu).
    - Giá trị dao động từ 1,027 đến hơn 853 triệu lượt stream.
    - Điều này cho thấy có một số bài hát rất phổ biến (outliers), trong khi phần lớn các bài hát có lượt stream thấp hơn (75% dưới 2.77 triệu).

## position\_streams\_by\_time\_daily\_total\_peak\_cleaned and position\_streams\_by\_time\_weekly\_total\_peak\_cleaned

|  |  |
| --- | --- |
| **Weekly** | **Daily** |
| A computer screen with white text  Description automatically generated | A screen shot of a computer  Description automatically generated |

* Đối với bộ dữ liệu hằng tuần có **772 dòng** và hằng ngày là **726 dòng**, cả hai đều có **4 cột.**
* Ở các cột như vn\_position và vn\_stream\_counts thiếu dữ liệu.

### **Dữ liệu thiếu**

|  |  |  |
| --- | --- | --- |
| **Cột** | **Số lượng thiếu** | **Tỷ lệ (%)** |
| Vn\_position (Weekly) | 1 | 0.129 |
| Vn\_stream\_count (Weekly) | 1 | 0.129 |
| Vn\_position (Daily) | 1 | 0.138 |
| Vn\_stream\_count (Daily) | 1 | 0.138 |

* + Các dòng thiếu dữ liệu là vô cùng ít vì thế nên nhóm vẫn sẽ giữ lại các dòng này.

### **Phân tích thống kê**

* + **Weekly**

|  |  |  |  |
| --- | --- | --- | --- |
|  | **TotalStreamCounts** | **PeakStreamCounts** | **PeakPosition** |
| count | 7.710000e+02 | 7.710000e+02 | 771.000000 |
| mean | 5.851733e+06 | 2.680038e+05 | 66.584955 |
| std | 9.287612e+06 | 3.178791e+05 | 58.811201 |
| min | 4.698000e+03 | 4.698000e+03 | 1.000000 |
| 25% | 2.867365e+05 | 9.259600e+04 | 12.000000 |
| 50% | 1.499057e+06 | 1.797870e+05 | 51.000000 |
| 75% | 6.552288e+06 | 3.087205e+05 | 111.000000 |
| max | 5.269765e+07 | 3.445288e+06 | 200.000000 |

* + **TotalStreamCounts:**
    - Trung bình: 5.85 triệu lượt stream mỗi tuần.
    - Độ lệch chuẩn: 9.29 triệu, cho thấy sự biến động lớn giữa các tuần.
    - Min/Max: Dao động từ 4,698 đến 52.69 triệu lượt stream.
    - Phân phối:
    - 25% số tuần có lượt stream dưới 286,736.
    - 50% số tuần (median) có lượt stream khoảng 1.49 triệu.
    - 75% số tuần có lượt stream dưới 6.55 triệu.
  + **PeakStreamCounts:**
    - Trung bình: 268,003 lượt stream cao nhất trong tuần.
    - Độ lệch chuẩn: 317,879, cho thấy sự biến động lớn giữa các tuần.
    - Min/Max: Dao động từ 4,698 đến 3.44 triệu lượt stream cao nhất trong tuần.
    - Phân phối:
    - 25% số tuần có lượt stream cao nhất dưới 92,596.
    - 50% số tuần có lượt stream cao nhất khoảng 179,787.
    - 75% số tuần có lượt stream cao nhất dưới 308,720.
  + **PeakPosition:**
    - Trung bình: Vị trí cao nhất trung bình là 66.58.
    - Độ lệch chuẩn: 58.81, cho thấy sự phân tán lớn.
    - Min/Max: Dao động từ vị trí 1 đến 200.
    - Phân phối:
    - 25% số tuần có vị trí cao nhất dưới 12.
    - 50% số tuần có vị trí cao nhất khoảng 51.
    - 75% số tuần có vị trí cao nhất dưới 111.
  + **Daily**

|  |  |  |  |
| --- | --- | --- | --- |
|  | **TotalStreamCounts** | **PeakStreamCounts** | **PeakPosition** |
| count | 7.250000e+02 | 725.000000 | 725.000000 |
| mean | 5.583042e+06 | 43500.244138 | 50.513103 |
| std | 9.160756e+06 | 47062.807284 | 45.455326 |
| min | 1.704000e+03 | 1364.000000 | 1.000000 |
| 25% | 3.157060e+05 | 16963.000000 | 9.000000 |
| 50% | 1.464363e+06 | 31830.000000 | 39.000000 |
| 75% | 6.013064e+06 | 50816.000000 | 83.000000 |
| max | 5.294350e+07 | 455604.000000 | 189.000000 |

* + **TotalStreamCounts:**
    - Trung bình: 5.58 triệu lượt stream mỗi ngày.
    - Độ lệch chuẩn: 9.16 triệu, cho thấy sự biến động lớn giữa các ngày.
    - Min/Max: Dao động từ 1,704 đến 52.94 triệu lượt stream.
    - Phân phối:
    - 25% số ngày có lượt stream dưới 315,706.
    - 50% số ngày (median) có lượt stream khoảng 1.46 triệu.
    - 75% số ngày có lượt stream dưới 6.01 triệu.
  + **PeakStreamCounts:**
    - Trung bình: 43,500 lượt stream cao nhất trong ngày.
    - Độ lệch chuẩn: 47,062, cho thấy sự biến động lớn giữa các ngày.
    - Min/Max: Dao động từ 1,364 đến 455,604 lượt stream cao nhất trong ngày.
    - Phân phối:
    - 25% số ngày có lượt stream cao nhất dưới 16,963.
    - 50% số ngày có lượt stream cao nhất khoảng 31,830.
    - 75% số ngày có lượt stream cao nhất dưới 50,816.
  + **PeakPosition:**
    - Trung bình: Vị trí cao nhất trung bình là 50.51.
    - Độ lệch chuẩn: 45.45, cho thấy sự phân tán lớn.
    - Min/Max: Dao động từ vị trí 1 đến 189.
    - Phân phối:
    - 25% số ngày có vị trí cao nhất dưới 9.
    - 50% số ngày có vị trí cao nhất khoảng 39.
    - 75% số ngày có vị trí cao nhất dưới 83.

## Vietnamese\_artists

A screen shot of a computer

Description automatically generated

* Tất cả các cột đều ở đúng kiểu dữ liệu và không có bất kỳ dữ liệu thiếu nào.

### **Dữ liệu thiếu**

* + Không có bất kỳ dữ liệu bị thiếu nào

### **Phân tích thống kê**

|  |  |  |
| --- | --- | --- |
|  | **popularity** | **followers** |
| count | 1115.000000 | 1.115000e+03 |
| mean | 30.951570 | 8.933268e+04 |
| std | 14.979501 | 3.273770e+05 |
| min | 0.000000 | 0.000000e+00 |
| 25% | 20.000000 | 6.185000e+02 |
| 50% | 32.000000 | 4.320000e+03 |
| 75% | 42.000000 | 4.008950e+04 |
| max | 68.000000 | 6.549380e+06 |

* + **popularity**
    - Trung bình: 30.95, với độ lệch chuẩn là 14.98.
    - Min/Max: Dao động từ 0 đến 68.
    - Phân phối:
    - 25% số nghệ sĩ có độ phổ biến dưới 20.
    - 50% số nghệ sĩ (median) có độ phổ biến khoảng 32.
    - 75% số nghệ sĩ có độ phổ biến dưới 42.
    - Nhận xét:
    - Độ phổ biến của các nghệ sĩ khá đa dạng, nhưng phần lớn tập trung ở mức trung bình (20-42).
    - Một số nghệ sĩ có độ phổ biến rất thấp (min = 0), có thể là các nghệ sĩ ít được biết đến hoặc mới xuất hiện.
  + **followers**
    - Trung bình: 89,332 người theo dõi, với độ lệch chuẩn là 327,377.
    - Min/Max: Dao động từ 0 đến 6.54 triệu người theo dõi.
    - Phân phối:
    - 25% số nghệ sĩ có dưới 618 người theo dõi.
    - 50% số nghệ sĩ (median) có khoảng 4,320 người theo dõi.
    - 75% số nghệ sĩ có dưới 40,089 người theo dõi.
    - Nhận xét:
    - Số lượng người theo dõi có sự phân tán rất lớn, với một số nghệ sĩ có hàng triệu người theo dõi (max = 6.54 triệu), trong khi nhiều nghệ sĩ có rất ít hoặc không có người theo dõi (min = 0).
    - Điều này cho thấy sự chênh lệch lớn giữa các nghệ sĩ nổi tiếng và ít nổi tiếng.

1. **Mối tương quan giữa các biến**
   1. **Vietnamese\_only\_songs\_cleaned**

A colorful squares with white text

Description automatically generated

* **Album\_popularity**
  + **Tương quan mạnh với popularity (0.81):**
    - Điều này cho thấy album phổ biến thường chứa các bài hát phổ biến. Đây là mối quan hệ hợp lý vì độ phổ biến của album thường phụ thuộc vào độ phổ biến của các bài hát trong album.
  + **Tương quan trung bình với track\_number (0.28):**
    - Các bài hát ở vị trí đầu tiên trong album có thể đóng góp nhiều hơn vào độ phổ biến của album.
  + **Tương quan trung bình với stream\_counts (0.25):**
    - Album phổ biến thường có tổng lượt stream cao hơn.
* **Disc\_number**
  + Không có mối tương quan đáng kể với các biến khác:
    - Hệ số tương quan gần bằng 0 với tất cả các biến, cho thấy số đĩa trong album không ảnh hưởng nhiều đến các yếu tố như độ phổ biến hay lượt stream.
* **Duration\_ms**
  + Tương quan âm yếu với album\_popularity (-0.10) và popularity (-0.06):
    - Thời lượng bài hát không ảnh hưởng nhiều đến độ phổ biến của album hoặc bài hát. Tuy nhiên, các bài hát quá dài có thể ít phổ biến hơn.
  + Không có mối tương quan đáng kể với các biến khác:
    - Thời lượng bài hát không phải là yếu tố chính ảnh hưởng đến các biến khác như lượt stream hay vị trí bài hát.
* **Popularity**
  + Tương quan mạnh với album\_popularity (0.81):
    - Như đã phân tích ở trên, bài hát phổ biến thường nằm trong album phổ biến.
  + Tương quan trung bình với stream\_counts (0.33):
    - Các bài hát phổ biến thường có lượt stream cao hơn, nhưng không phải lúc nào cũng vậy. Có thể có những bài hát ít phổ biến nhưng vẫn có lượt stream cao do các yếu tố khác như quảng bá hoặc viral.
* **Track\_number**
  + Tương quan trung bình với album\_popularity (0.28):
    - Các bài hát ở vị trí đầu tiên trong album có thể đóng góp nhiều hơn vào độ phổ biến của album.
  + Không có mối tương quan đáng kể với các biến khác:
    - Vị trí bài hát trong album không ảnh hưởng nhiều đến độ phổ biến hay lượt stream.
* **Stream\_counts**
  + Tương quan trung bình với popularity (0.33):
    - Các bài hát có lượt stream cao thường phổ biến hơn, nhưng không phải lúc nào cũng vậy.
  + Tương quan trung bình với album\_popularity (0.25):
    - Album có tổng lượt stream cao thường phổ biến hơn.
  1. **position\_streams\_by\_time\_daily\_total\_peak\_cleaned and position\_streams\_by\_time\_weekly\_total\_peak\_cleaned**

|  |  |
| --- | --- |
| **Weekly** | **Daily** |
| A diagram of a heatmap  Description automatically generated | A diagram of a heatmap  Description automatically generated |

* **Mối quan hệ giữa TotalStreamCounts và PeakStreamCounts:**
  + Weekly có hệ số tương quan cao hơn (0.55 so với 0.50), cho thấy trong dữ liệu Weekly, tổng lượt stream và lượt stream cao nhất có mối liên hệ chặt chẽ hơn.
* **Mối quan hệ giữa TotalStreamCounts và PeakPosition:**
  + Weekly có hệ số tương quan âm mạnh hơn (-0.51 so với -0.45), cho thấy trong dữ liệu Weekly, tổng lượt stream ảnh hưởng nhiều hơn đến vị trí trên bảng xếp hạng.
* **Mối quan hệ giữa PeakStreamCounts và PeakPosition:**
  + Weekly và Daily có hệ số tương quan tương tự (-0.47 và -0.46), cho thấy mối quan hệ giữa lượt stream cao nhất và vị trí trên bảng xếp hạng không thay đổi nhiều giữa hai tập dữ liệu.
  1. **Vietnamese\_artitsts**

A red and blue squares

Description automatically generated

* **Mối tương quan giữa popularity và followers**
  + Hệ số tương quan: 0.41.
  + Đây là mối tương quan dương trung bình, cho thấy rằng các nghệ sĩ có nhiều người theo dõi hơn thường có độ phổ biến cao hơn.
  + Tuy nhiên, hệ số tương quan không quá mạnh, điều này có thể do:
    - Một số nghệ sĩ có nhiều người theo dõi nhưng không phổ biến (ít hoạt động hoặc không có bài hát mới).
    - Một số nghệ sĩ có độ phổ biến cao nhưng chưa thu hút được nhiều người theo dõi (nghệ sĩ mới nổi hoặc viral).
* **Mối tương quan nội tại**
  + popularity với chính nó: Hệ số tương quan là 1.00, vì đây là mối tương quan hoàn hảo (một biến với chính nó).
  + followers với chính nó: Hệ số tương quan là 1.00, tương tự như trên.

1. **Tương quan stat model**
   1. **Giới thiệu**

* Trong bài toán này, nhóm sẽ phân tích mối tương quan giữa số lượng phát (stream\_counts) và các biến độc lập còn lại và sự phổ biến (popularity) và các biến độc lập còn lại.
  1. **Phương pháp**
* Ở đây nhóm sẽ sử dụng mô hình OLS (Ordinary Least Squares) và mô hình sẽ có dạng như sau:

Biến Mục Tiêu = *β*0 +*β*1 *×*[Biến Độc Lập]+*ε* (1)

* Trong đó*,*(*β*0) là hệ số chặn, (*β*1) là hệ số hồi quy cho [biến độc lập], và (*ε*) là sai số ngẫu nhiên.

## Stream\_counts vs các biến còn lại

A screenshot of a computer

Description automatically generated

* **Tổng quan về mô hình**
  + **R-squared (R²): 0.113**
    - Chỉ số này cho biết mô hình giải thích được 11.3% sự biến thiên của biến phụ thuộc (stream\_counts). Đây là một giá trị thấp, cho thấy mô hình chưa phù hợp để dự đoán tốt stream\_counts.
    - Có thể còn nhiều yếu tố quan trọng khác chưa được đưa vào mô hình.
  + **Adj. R-squared (R² điều chỉnh): 0.112**
    - Giá trị này gần bằng R², cho thấy số lượng biến độc lập trong mô hình không gây ra quá nhiều nhiễu.
  + **F-statistic: 123.0 (p-value: 6.49e-123)**
    - Giá trị này rất nhỏ, cho thấy mô hình tổng thể có ý nghĩa thống kê. Ít nhất một biến độc lập có ảnh hưởng đáng kể đến biến phụ thuộc.
* **Ý nghĩa của các biến:**
  + **album\_popularity:**
    - Hệ số: -6.514e+04.
    - Mối quan hệ âm: Khi độ phổ biến của album tăng thêm 1 đơn vị, số lượt stream giảm trung bình 65,140 lượt, giữ các yếu tố khác không đổi.
    - P-value: 0.036 (< 0.05), có ý nghĩa thống kê.
  + **disc\_number:**
    - Hệ số: -3.519e+05.
    - Không có ý nghĩa thống kê (p-value: 0.929), cho thấy số đĩa không ảnh hưởng đáng kể đến số lượt stream.
  + **duration\_ms:**
    - Hệ số: -12.6594.
    - Mối quan hệ âm: Khi thời lượng bài hát tăng thêm 1 ms, số lượt stream giảm trung bình 12.66 lượt.
    - P-value: 0.002 (< 0.05), có ý nghĩa thống kê.
  + **popularity:**
    - Hệ số: 4.596e+05.
    - Mối quan hệ dương: Khi độ phổ biến của bài hát tăng thêm 1 đơn vị, số lượt stream tăng trung bình 459,600 lượt.
    - P-value: 0.000 (< 0.05), có ý nghĩa thống kê mạnh.
  + **track\_number:**
    - Hệ số: 1080.9226.
    - Không có ý nghĩa thống kê (p-value: 0.988), cho thấy vị trí bài hát trong album không ảnh hưởng đáng kể đến số lượt stream.

## Popularity vs các biến còn lại

A screenshot of a computer screen

Description automatically generated

* **Tổng quan về mô hình**
  + **R-squared (R²): 0.737**
    - Mô hình giải thích được 73.7% sự biến thiên của biến phụ thuộc (popularity). Đây là một giá trị khá cao, cho thấy mô hình phù hợp để dự đoán popularity.
  + **Adj. R-squared (R² điều chỉnh): 0.737**
    - Giá trị này gần bằng R², cho thấy số lượng biến độc lập trong mô hình không gây ra quá nhiều nhiễu.
    - F-statistic: 2699 (p-value: 0.000)
    - Giá trị này rất nhỏ, cho thấy mô hình tổng thể có ý nghĩa thống kê. Ít nhất một biến độc lập có ảnh hưởng đáng kể đến biến phụ thuộc.
* **Ý nghĩa của các biến:**
  + **const**:
    - Hệ số: 17.2967.
    - Đây là giá trị trung bình của popularity khi tất cả các biến độc lập khác bằng 0.
  + **album\_popularity:**
    - Hệ số: 0.8051.
    - Mối quan hệ dương: Khi độ phổ biến của album tăng thêm 1 đơn vị, popularity tăng trung bình 0.8051 đơn vị.
    - P-value: 0.000 (< 0.05), có ý nghĩa thống kê mạnh.
  + **disc\_number:**
    - Hệ số: -9.6738.
    - Mối quan hệ âm: Khi số đĩa tăng thêm 1, popularity giảm trung bình 9.6738 đơn vị.
    - P-value: 0.000 (< 0.05), có ý nghĩa thống kê mạnh.
  + **duration\_ms:**
    - Hệ số: 1.038e-05.
    - Mối quan hệ dương: Khi thời lượng bài hát tăng thêm 1 ms, popularity tăng trung bình 0.00001038 đơn vị.
    - P-value: 0.000 (< 0.05), có ý nghĩa thống kê mạnh.
  + **track\_number:**
    - Hệ số: -1.0464.
    - Mối quan hệ âm: Khi số thứ tự bài hát trong album tăng thêm 1, popularity giảm trung bình 1.0464 đơn vị.
    - P-value: 0.000 (< 0.05), có ý nghĩa thống kê mạnh.
  + **stream\_counts:**
    - Hệ số: 9.132e-08.
    - Mối quan hệ dương: Khi số lượt stream tăng thêm 1 đơn vị, popularity tăng trung bình 0.00000009132 đơn vị.
    - P-value: 0.000 (< 0.05), có ý nghĩa thống kê mạnh.

1. **Trực quan hóa dữ liệu**

## Tích hợp AI trong việc thiết kế dashboard

* Ở đây nhóm đã nhờ sự giúp đỡ của [Grok](https://grok.com/) một chatbot AI để có thể gợi ý dashboard, trong đó:

A screenshot of a chat

Description automatically generated

A screenshot of a computer

Description automatically generated

A screenshot of a computer

Description automatically generated

A screenshot of a computer

Description automatically generated

A screenshot of a web page

Description automatically generated

A screenshot of a computer

Description automatically generated

* Kế đến sẽ là phần chi tiết của các dashboard

## Dashboard tổng quan

### **Tổng thể dashboard**

**A screenshot of a computer

Description automatically generated**

**Giao diện dashboard tổng quan**

* + Dashboard này cung cấp một cái nhìn tổng thể về thị trường âm nhạc Việt Nam bao gồm quy mô, các đặc điểm trung bình, xu hướng theo thời gian và phân bố của các yếu tố chính như thể loại và độ phổ biến.

### **Các Chỉ số Chính (KPIs):**

* **Quy mô:** Tập dữ liệu bao gồm 1409 nghệ sĩ, 4735 bài hát, và 195 thể loại khác nhau. Tổng lượt phát (streams/plays) đạt 4.1 triệu.
* **Đặc điểm trung bình:**
* Thời lượng trung bình của một bài hát là 4.0 phút.
* Độ phổ biến trung bình là 28.

### **Xu hướng theo Thời gian:**

* **Số lượng bài hát:** Biểu đồ đường "Số lượng bài hát qua từng năm" cho thấy một sự gia tăng đột biến và cực kỳ mạnh mẽ về số lượng bài hát được ghi nhận trong những năm gần đây (đặc biệt là khoảng sau 2015 và đỉnh điểm vào đầu những năm 2020). Giai đoạn trước năm 2000 có số lượng bài hát rất ít. Điều này có thể phản ánh sự bùng nổ của thị trường âm nhạc hoặc sự thay đổi trong phương pháp thu thập dữ liệu.
* **Thời lượng bài hát:** Biểu đồ "Thời lượng bài hát qua từng năm" có một xu hướng thú vị: thời lượng trung bình đạt đỉnh cao (khoảng 8 phút) vào đầu những năm 1980, sau đó giảm mạnh và ổn định quanh mức 4 phút từ cuối những năm 1990 cho đến nay, với một xu hướng giảm nhẹ trong những năm gần nhất. Điều này có thể phản ánh sự thay đổi trong thị hiếu âm nhạc, từ các định dạng dài hơn sang các bài hát ngắn gọn, phù hợp với radio hoặc streaming hơn.

### **Phân bố:**

* **Thể loại nhạc:** v-pop chiếm tỷ trọng lớn nhất (29.74%), theo sau là vinahouse (19.68%), vietnamese lo-fi (19.64%), vietnamese hip hop (15.8%), và vietnam indie (15.14%). Tổng cộng 5 thể loại này chiếm gần như toàn bộ thị phần trong biểu đồ này.
* **Độ phổ biến:** Phần lớn các bài hát tập trung ở mức độ phổ biến trung bình (cao nhất ở khoảng 30-40). Có ít bài hát ở mức rất thấp (0-10) hoặc rất cao (50+). Phân bố này khá chuẩn, tập trung quanh giá trị trung bình là 28.
  + 1. **Bộ lọc:**
* Có thể lọc dữ liệu theo Năm, Khoảng phổ biến, Thể loại và Thời lượng.
  + 1. **Kết luận:**
* Các thể loại như v-pop, vinahouse, lo-fi, hip hop và indie chiếm ưu thế áp đảo. Thời lượng bài hát trung bình đã ổn định ở mức khoảng 4 phút trong nhiều năm. Dữ liệu về độ phổ biến cho thấy hầu hết các bài hát có mức độ phổ biến trung bình.

## Songs dashboard

* + 1. **Tổng thể dashboard**

A screenshot of a graph

AI-generated content may be incorrect.

* + Dashboard cung cấp thông tin về từng bài hát với các phần chính: tên bài hát, nghệ sĩ trình diễn, độ phổ biến, thể loại bài hát, vị trí cao nhất bài hát từng lên, lượt phát cao nhất mà bài hát đạt được.
  + Cùng với các thông tin tổng quan về các bài hát trên thị trường như: biểu đồ xếp hạng bài hát theo độ phổ biến, biểu đồ xếp hạng lượt stream của bài hát, và biểu đồ cho thấy mối quan hệ giữa độ phổ biến với độc phổ biến album.
    1. **Các nhận xét biểu đồ tổng quát:**
  + Những bài hát có độ phổ biến cao nhất là: raining for hours (80), Kyoto (76), La Diabla (73), Mất kết nối (70),… Những bài hát có lượt stream cao nhất là: La Diabla (853M), raining for hours (645M), Andree Right Hand xin chào (199M),…
* Có thể thấy các bài hát có độ phổ biến cao sẽ thường có lượt stream cao và nằm trong top đầu, điều này là thực tế, vì độ phổ biến cao sẽ có nhiều người nghe dẫn tới lượt stream cao.
* **Về biểu đồ mối tương quan giữa độ phổ biến bài hát và độ phổ biến album:** Biểu đồ thể hiện mối tương quan dương tuyến tính khá rõ ràng giữa độ phổ biến của bài hát và độ phổ biến của album chứa bài hát đó. Điều này có nghĩa là khi album phổ biến, khả năng cao các bài hát trong album đó cũng phổ biến, và ngược lại.
* Dường như không có nhiều bài hát phổ biến xuất hiện trong những album ít phổ biến. Điều này phản ánh tầm quan trọng của thương hiệu và độ lan tỏa của toàn bộ album đối với thành công của từng bài hát.

### **Các chi tiết về một số bài hát nổi bật (lấy các bài Dù cho tận thế, Chuyện đôi ta).**

A screenshot of a computer

AI-generated content may be incorrect.

* Bài "Dù Cho Tận Thế" của ERIK đạt độ phổ biến cao với vị trí cao nhất là số 1 và lượt phát cao nhất 347K, độ phổ biến bài hát là 67 và album là 57, cho thấy sức hút mạnh mẽ của ca khúc; ERIK – một trong những nghệ sĩ nổi bật của V-pop – đã thể hiện rõ khả năng tạo hit và duy trì độ phổ biến ổn định qua nhiều bài hát.

A screenshot of a computer

AI-generated content may be incorrect.

* Tiếp theo là bài hát có lượt stream nằm trong top là ”Chuyện đôi ta” của Da LAB – nhóm nhạc đình đám. Với độ phổ biến bài hát và album lần lượt là 61, 51. Bài hát đã leo tới vị trí cao nhất 3 với số lượt phát cao nhất là 73K.

### **Bộ lọc:**

* Cho phép ta lọc theo năm, theo tên bài hát, cho phép tìm kiếm

### **Kết luận:**

* Dashboard cho thấy rằng một số bài hát nổi bật như La Diabla, Raining for hours, và Kyoto đạt được độ phổ biến cao và lượt stream ấn tượng (trên hàng trăm triệu), chứng minh sức hút mạnh mẽ và tính lan truyền cao của các bản hit này. Và việc các bài hát có độ phổ biến cao sẽ thường có lượt stream nằm trong top là thực tế
* Biểu đồ tương quan giữa độ phổ biến bài hát và độ phổ biến album cho thấy mối liên hệ tuyến tính dương, tức album phổ biến thường chứa các bài hát phổ biến. Một điểm nữa đó là gần như không có nhiều bài hát phổ biến cao xuất hiện trong các album có độ phổ biến thấp, điều này cho thấy tầm quan trọng của thương hiệu

## Artists dashboard

### **Tổng thể dashboard**

A screenshot of a computer

AI-generated content may be incorrect.

**Giao diện dashboard nghệ sĩ**

* Dashboard tập trung vào phân tích chi tiết từng nghệ sĩ cụ thể. Bao gồm các chỉ số như độ nổi tiếng, lượt theo dõi, tổng lượt phát, và thể loại nhạc liên quan. Ngoài ra, dashboard còn đưa ra so sánh vị trí của nghệ sĩ với các nghệ sĩ khác trên thị trường âm nhạc.
  + 1. **So sánh tổng quát:**
  + Tree map bên dưới so sánh các nghệ sĩ có nhiều lượt theo dõi nhất như Sơn Tùng M-TP (6.5M), Đen Vâu (4.3M), Vũ (2.4M), HIEUTHUHAI (2.2M),…
  + Mặt dù có nhiều lượt theo dõi nhất nhưng về độ nổi tiếng Sơn Tùng MTP (65) chỉ xếp ở vị trí thứ 6 sau HIEUTHUHAI (68), Anh Trai Say Hi (67), Dangrangto (67), tlinh (66).
  + Thậm chí những cái tên có nhiều lượt theo dõi khác như Đen Vâu, Vũ, Mỹ Tâm cũng có thứ hạng rất thấp trong độ nổi tiếng. Điều này cho thấy rằng các nghệ sĩ gạo cội, có hoạt động nghệ thuật lâu năm (những người đã có nhiều người theo dõi) sẽ không nổi tiếng bằng các hiện tượng mới nổi đương thời. Điều này là do khán giả trên nền tảng spotify chủ yếu là giới trẻ, sự thay đổi trong xu hướng âm nhạc. Một nghệ sĩ có thể từng rất nổi (nhiều người follow), nhưng nếu không có bài hit mới hoặc ít hoạt động gần đây, thì độ nổi tiếng hiện tại thấp.

### **Các phân tích chi tiết (xét các ca sĩ nổi bật như Sơn Tùng, HIEUTHUHAI):**

A screenshot of a computer

AI-generated content may be incorrect.

* + Sơn Tùng là ca sĩ có nhiều lượt theo dõi nhất với 6.5M lượt theo dõi, 865M lượt phát, 65 điểm độ nổi tiếng. Với các thể loại âm nhạc đa dạng như v-pop, vinahouse, hiphop. Cùng các bài hát quen thuộc trong giới trẻ như: Chúng ta của hiện tại, Muộn rồi mà sao còn,… đã góp phần làm nên tên tuổi của MTP.

A screenshot of a computer

AI-generated content may be incorrect.

* HIEUTHUHAI – cái tên cực kì quen thuộc trong cộng đồng người yêu vpop, vietnamese hiphop. Không bất ngờ khi nhà sản xuất âm nhạc này có độ nổi tiếng cao nhất, vượt mặt các cái tên đình đám như tlinh, Dangrangto,… Với những bài hit như Không thể say, Exit Sign, Ngủ một mình,… mang về cho anh 68 điểm nổi tiếng, 2.2M lượt theo dõi và 840M lượt phát.

### **Bộ lọc:**

* Cho phép lọc theo tên nghệ sĩ, cho phép tìm kiếm theo tên.

### **Kết luận:**

**Insight: Tại sao một nghệ sĩ có lượt theo dõi cao nhưng lại không đứng top trong bảng xếp hạng độ nổi tiếng?**

* Lượt theo dõi phản ánh sự trung thành: **Lượt theo dõi (followers)** là **chỉ số tích lũy theo thời gian**, phản ánh sự trung thành của fan, mức độ được yêu thích bền vững trong quá khứ
* Độ nổi tiếng phản ánh xu hướng: **Độ nổi tiếng (popularity)** là một **chỉ số động**, có thể tính theo số lượt phát gần đây, sự viral trên mạng xã hội, vị trí trên bảng xếp hạng gần nhất,...
* Một nghệ sĩ có thể từng rất nổi (nhiều người follow), nhưng nếu không có bài hit mới hoặc ít hoạt động gần đây, thì độ nổi tiếng hiện tại thấp.
* Nghệ sĩ “huyền thoại” vs “ngôi sao hiện tại”: Một số nghệ sĩ như Mỹ Tâm, Đen Vâu, Sơn Tùng có fanbase lâu dài, ổn định, nên lượng theo dõi cao. Trong khi đó, các nghệ sĩ như HIEUTHUHAI, tlinh, hay các hiện tượng có thể viral nhất thời, khiến độ nổi tiếng tăng vọt – nhưng lượng follower tích lũy chưa kịp theo kịp.

## Dashboard thể loại

### **Tổng thể dashboard**

**A close-up of a graph

Description automatically generated**

**Giao diện dashboard thể loại**

* Dashboard này đi sâu vào phân tích các thể loại nhạc, xem xét số lượng bài hát, tổng lượt phát và số lượng nghệ sĩ liên quan đến từng thể loại.

### **Các Chỉ số Chính (KPIs):**

* Bộ dữ liệu có tổng cộng 194 thể loại và 4701 bài hát.
  + 1. **Phân tích chi tiết theo Thể loại:**
* **Tổng lượt phát:** Mặc dù có nhiều thể loại, lượt phát tập trung chủ yếu vào 5 thể loại hàng đầu đã thấy ở trang Tổng quan. v-pop dẫn đầu với khoảng 19.8 tỷ lượt phát (19839M), theo sau là vietnam indie (~15.3 tỷ), vietnamese hip hop (~14.9 tỷ), vietnamese lo-fi (~12.3 tỷ), và vinahouse (~10 tỷ). Các thể loại khác như vietnamese bolero, bolero, lo-fi ... có tổng lượt phát thấp hơn đáng kể.
* **Số lượng bài hát:** Số lượng của 5 thể loại chính quá khác biệt so với các thể loại còn lại: v-pop có nhiều bài hát nhất (4683), tiếp theo là vietnamese hip hop (3096), vietnam indie (3092), vietnamese lo-fi (2479), và vinahouse (2389). Các thể loại khác có số lượng bài hát ít hơn nhiều.
* **Số lượng nghệ sĩ:** V-pop liên kết với nhiều nghệ sĩ nhất (1378), theo sau là vietnamese hip hop (1048), vietnam indie (1012), vietnamese lo-fi (881), và vinahouse (850). Điều này cho thấy sự tập trung không chỉ về bài hát mà còn nghệ sĩ trong các thể loại này.
  + 1. **Bộ lọc:**
* Cho phép lọc theo Năm và chọn cụ thể các Thể loại.
  + 1. **Kết luận:**
* Thể hiện rõ sự thống trị của 5 thể loại nhạc Việt Nam hiện đại (v-pop, vietnamese hip hop, vietnam indie, vietnamese lo-fi, vinahouse) trên cả ba phương diện: tổng lượt phát, số lượng bài hát và số lượng nghệ sĩ. v-pop nổi bật là thể loại dẫn đầu về tổng lượt phát và cũng có nhiều bài hát/nghệ sĩ nhất. Sự chênh lệch giữa 5 thể loại hàng đầu và phần còn lại là rất lớn, cho thấy một thị trường tập trung cao độ vào các xu hướng âm nhạc hiện đại.

1. **Tích hợp AI trong việc phân tích**

A screenshot of a chat

Description automatically generated

* 1. **Hỏi đáp**

A screenshot of a phone

Description automatically generated

* + Ở đây, nhóm sẽ sử dụng phần Q&A có sẵn trong PowerBI.
  + Và đồng thời, chúng ta cũng có thể hỏi các câu hỏi để Q&A trả lời câu hỏi và ở trên là ví dụ.
  1. **Chatbot**

Screens screenshot of a chat

Description automatically generated

* + Nhóm sử dụng một dạng visual mới là **AI Lens:** Tích hợp OpenAI cho phép đặt câu hỏi bằng ngôn ngữ tự nhiên và sẽ đưa ra insights về dữ liệu. Tuy nhiên về mặt khuyết điểm: Đây là text generation AI, vì thế sẽ bị hạn chế về mặt trực quan.

# **Tích hợp AI trong việc đánh giá dashboard**

* + Ở đây nhóm em sẽ sử dụng Chat GPT để có thể có được đánh giá tổng quan đồng thời để có được góc nhỉn cải thiện của GPT.

## Dashboard Tổng quan

* + Do ở phần bên trái, dashboard đã có filter thể loại vì thế nhóm sẽ không chỉnh sửa thêm phần tooltip.
  + Và đường trung bình thời lượng, nhóm sẽ không để vào do sẽ gây khó nhìn trong lúc trực quan hóa.

A screenshot of a computer

Description automatically generated

## Dashboard Bài hát

* + Ở các tên bài hát các ký tự đặc biệt $ là các tên của nghệ sĩ nên nhóm em cũng sẽ không xử lý tên có ký tự đặc biệt.

A screenshot of a computer

Description automatically generated

## Dashboard Thể loại

* + Nhóm đã điều chỉnh lại sắp xếp các thể loại giảm dần theo tổng lượng phát.

A screenshot of a computer

Description automatically generated

## Dashboard Nghệ sĩ

* + Như đã đề cập phía trước tên nghệ sĩ không có ký tự nào bị lỗi vì thế nhóm em vẫn giữ nguyên tên.
  + Bảng các bài hát do nhóm em muốn nhấn mạnh vào lượt phát và tên các bài hát của nghệ sĩ nên nhóm sẽ không thêm năm phát hành và thể loại để tránh coginitve overload.

A screenshot of a computer

Description automatically generated

## Đánh giá tổng quát

A screenshot of a review

Description automatically generated