**README file**

**Spotify 2024 Streaming data**

**1. Data Pre-processing**

The initial step involved cleaning and organizing the data to ensure accuracy and consistency. This included:

* Checking for and correcting all spelling errors across datasets.
* Creating normalized tables for Albums, Artists, and Countries to enhance data structure.
* Removing duplicate entries to maintain data integrity.
* Establishing primary keys in each supporting table.
* Adding corresponding foreign keys in the main fact table to enable relationships.
* Converting all relevant datasets into proper data table formats.
* Performing a thorough review to verify data consistency.

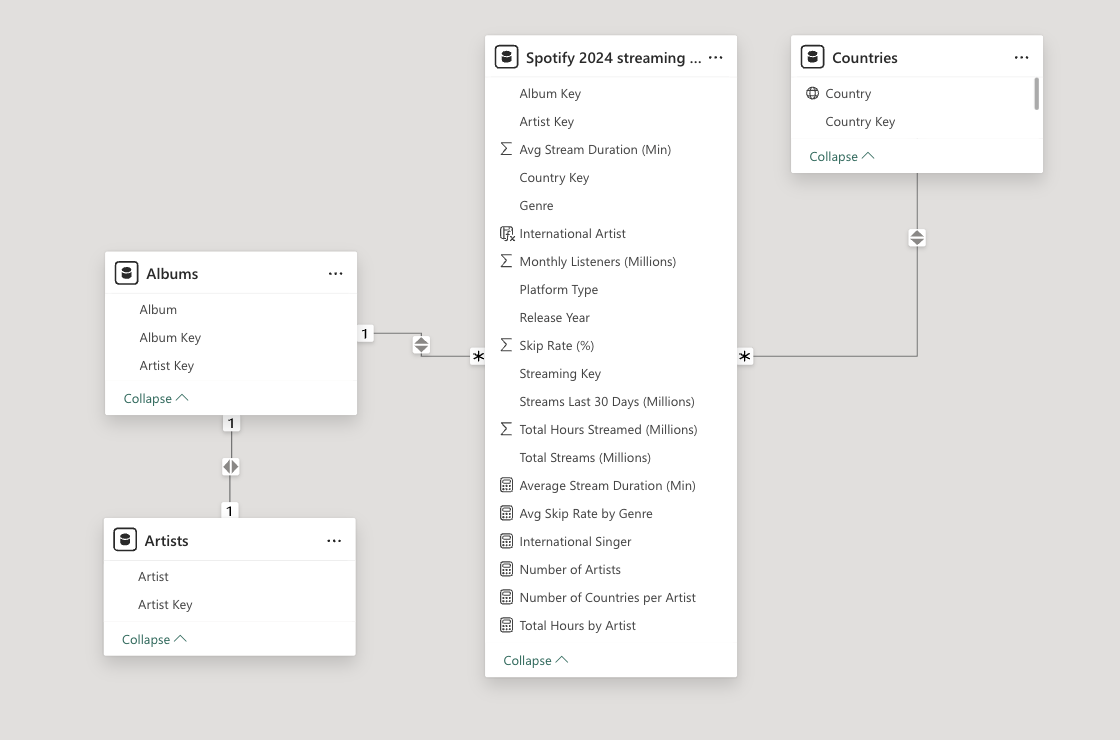
**2. Data Modelling**

Once pre-processing was complete, a structured data model was created.

* Relationships were defined between the fact table and dimension tables using primary and foreign keys.
* A series of DAX measures and calculated columns were developed to support the report's analytical needs:
  + A measure was created to calculate the number of countries each artist is streamed in.
  + A calculated column labelled "International Singer" was added to indicate whether an artist was streamed in more than two countries.
  + An average skip rate measure was implemented to analyse genre-based skip behaviour.
  + A measure was introduced to summarize total streaming hours per artist.
  + A measure to compute average stream duration (in minutes) was also created, handling blank values appropriately.
  + A measure to count the number of unique artists was added using distinct artist keys.

| **Type** | **Name** | **Formula** |
| --- | --- | --- |
| **Measure** | Number of Countries per Artist | CALCULATE(DISTINCTCOUNT('Spotify 2024 streaming data'[Country Key]), ALLEXCEPT('Spotify 2024 streaming data', 'Spotify 2024 streaming data'[Artist Key])) |
| **Calculated Column** | International Singer | IF([ArtistCountryCount] > 2, "Yes", "No") |
| **Measure** | Avg Skip Rate by Genre | CALCULATE(AVERAGE('Spotify 2024 streaming data'[Skip Rate (%)]), ALLEXCEPT('Spotify 2024 streaming data', 'Spotify 2024 streaming data'[Genre])) |
| **Measure** | Total Hours by Artist | CALCULATE(SUM('Spotify 2024 streaming data'[Total Hours Streamed (Millions)]), ALLEXCEPT('Spotify 2024 streaming data', 'Spotify 2024 streaming data'[Artist Key])) |
| **Measure** | Average Stream Duration (Min) | VAR Result = AVERAGE('Spotify 2024 streaming data'[Avg Stream Duration (Min)]) RETURN IF(ISBLANK(Result), 0, Result) |
| **Measure** | Number of Artists | DISTINCTCOUNT(Artists[Artist Key]) |

Final data model:

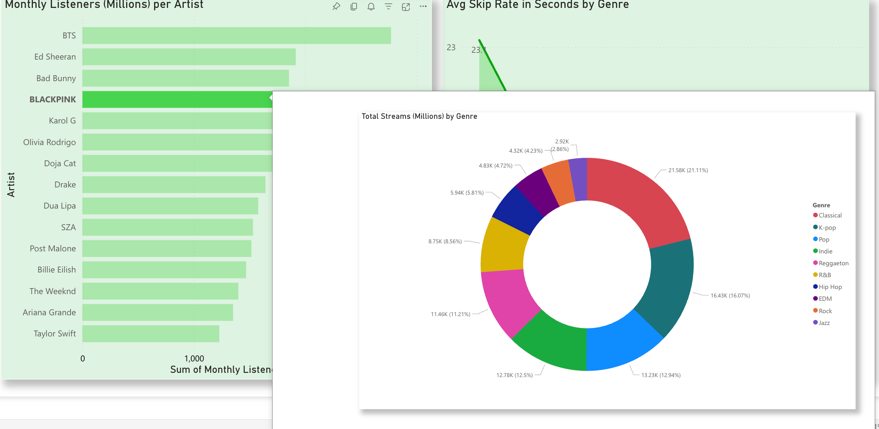


**3. Report Creation**

The Power BI report was developed over multiple pages to present insights clearly and interactively.

* **Page 1** features two KPI tiles showing the number of artists and the average stream duration. It also includes:
  + A bar chart displaying monthly listeners per artist.
  + A shaded line chart visualizing the average skip rate (in seconds) by genre.
  + A slicer allowing filtering by country.
* **Page 2** displays:
  + A map chart with bubble sizes representing total yearly streams per country.
  + Two slicers for filtering by genre and artist.
* **Pages 3 and 4** serve as tooltip pages:
  + Page 3 includes a table (Country, Total Streams) used as a tooltip for the map chart.
  + Page 4 contains a donut chart showing the number of streams per artist, used as a tooltip for the bar chart on Page 1.
* A drill-down table was created to show artist, country, 30-day stream totals, and annual stream totals. This table is linked to the monthly listeners bar chart via drill-through functionality.
* Report interactivity was carefully managed:
  + KPI tiles were protected from being influenced by slicer selections.
  + The Selection Pane was used to organize visuals, and tab order was refined for a better user experience.
* A bookmark was created specifically for the "Argentina" country view, enabling quick access.

Example of a tooltip:

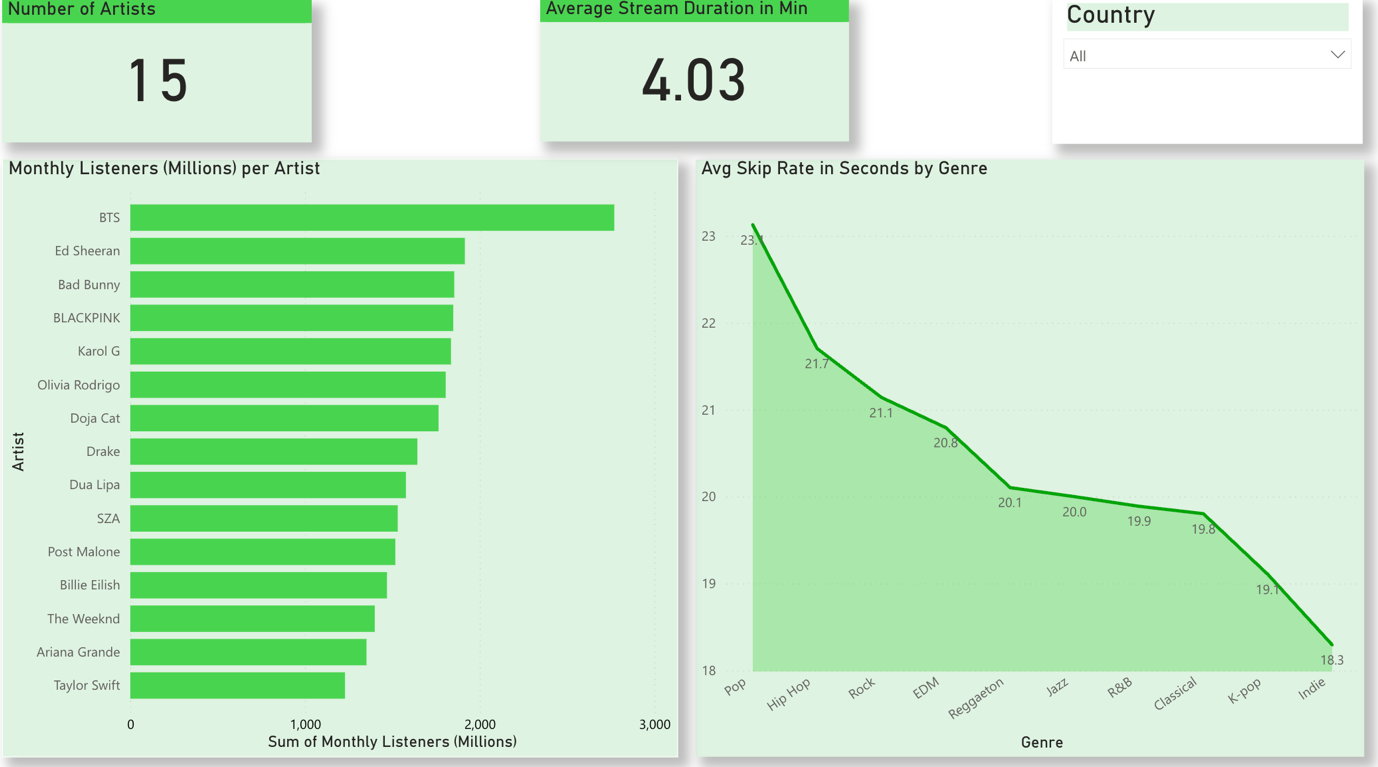


**4. Visualization & Formatting**

Visual consistency was maintained throughout the report by:

* Applying a uniform colour theme and formatting style.
* Adding background colours and shading to key figures and charts to enhance readability and emphasis.

Example of Page 1 – Overview. The whole file will be attached.



**5. Data insights**

Based on the 2024 Spotify streaming dataset, the following insights have been identified. These represent key trends, but are not exhaustive.

1. Artist Trends

* BTS leads the analysed artists in both total stream hours and monthly listeners, indicating strong global fan engagement.

1. Genre Popularity by Region

* Rock records its highest number of streams in Turkey, suggesting a regionally strong listener base.
* K-Pop is most streamed in Argentina, pointing to a concentrated market for Korean music in South America.

1. Listener Behaviour

* Pop has the lowest skip rate, with listeners skipping after an average of 23.1 seconds, indicating higher early retention.
* Indie is skipped fastest, at an average of 18.3 seconds, suggesting lower initial engagement.