BUS MATRIX ARCHITECTURE

SHRADDHA SAHA

DATAPLAY

Contents

- Objective
- Data collection
- Key concept
- Work flow
  - Business Goal
  - Business Processes
  - Relevance of the business processes
  - Dimensions
  - Attributes of the dimensions
  - Bus matrix
- Inference

Objective

- The objective of this project is to create a Business (Bus) Matrix Architecture from the provided dataset. A Bus Matrix Architecture is a detailed matrix that shows the relationship between business processes and the dimensions of the data.
- We will construct this tabular form in MS Excel .

Data collection

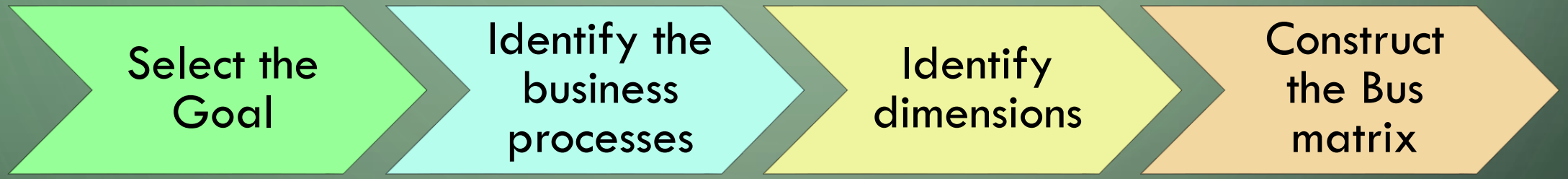
- The dataset is provided by DataPlay on “Most Streamed Spotify Songs 2024”
- The dataset is an Excel worksheet in CSV format.
- One can download it from Kaggle too.

Key concept

BUS MATRIX

- It is a consolidated tabular view that depicts the relationships between dimension and fact tables.
- A bus matrix enables one to identify easily which facts share the same dimension(s). By using this information, we can assess which tables can be combined into one report.
- It is particularly useful in the design phase of a dimensional data warehouse, helping to ensure that all necessary data is captured and properly organized.

Work flow


Select the Goal
Identify the business processes
Identify dimensions
Construct the Bus matrix

Business Goal

- I am taking the **BUSINESS GOAL** **to analyze the track performance on Spotify only.**
- From this analysis, will try to understand which kind of music is enjoyed the most and build more tracks of that kind
- This will be helpful in improving the track performance in the future.

Business Processes

Business Processes	Spotify Popularity
	Spotify Streams
	Spotify Playlist Count
	Spotify Playlist Reach
	Deezer Playlist Reach
	Amazon Playlist Count
	Shazam Counts
	Pandora Track Stations

- To analyze the track performance on Spotify
- To analyze the relative track performance using these different platforms

Relevance of the business processes

**As our Goal is to analyze the track performance on Spotify only; we will surely consider all of the below mentioned Spotify metrics for 'business processes':

- Spotify Popularity : The Spotify popularity score is typically a value from 0 to 100, with 100 being the most popular.
- Spotify Streams : A stream is counted when a track is played for at least 30 seconds. If a listener plays the same track multiple times, each instance is counted as a separate stream.
- Spotify Playlist Count : This metric counts every distinct playlist in which the track appears. It reflects how widely the track is curated by users and playlist editors.
- Spotify Playlist Reach : Playlist reach is calculated by summing up the follower counts of all the playlists that include the track. It estimates how many users could potentially see or hear the track if they listen to the playlists in which it appears.

Relevance of the business processes (cont.)

**We need the following business processes to analyze the relative track performance, we have also indicated why we took these metrics only from a particular App:

- Deezer Playlist Reach : Deezer is another music streaming app. Now Reach is more focused on the potential audience size. The total number of unique listeners who have access to a song through the playlists it's featured on.. So, *not taken playlist Count*.
- Amazon Playlist Count : Amazon Music is also a music streaming service. It delivers a library of more than 100 million songs; only Playlist count is available in our data for this app.
- Shazam Counts: Shazam uses song recognition technology to detect and recognize a song that is playing remotely ; taken Counts into relevance as it's the only metric available for this dataset.
- Pandora Track Stations: Pandora is a subscription-based music streaming service; one user can personalize their track station, so we took this (not Pandora streams).

Relevance of the business processes (cont.)

Now there might be a question about **why we did not take other Apps into account:

- We are analyzing a music streaming app, so we haven't considered social media apps like YouTube and TikTok.
- Soundcloud empowers independent artists with the tools, services and resources they need to help them build and grow their careers; and our goal is track performance based, so left this also.
- SiriusXM is not taken as we used Pandora for analysis which is owned by the same broadcasting corporation Sirius XM; and Pandora holds more weightage when it comes to total monthly active users (according to 4th quarter of 2023).
- Apple Music plays well in Apple devices only, so it is not helpful for this analysis of overall track performance.
- AirPlay is owned by Apple, so same problem as Apple Music.
- TIDAL popularity could not be used as there is a huge missing value problem in the dataset.

Dimensions

Identified dimensions	Track Information
	Album Info
	Artist data
	Time
	Law

- Determined the dimensions that are relevant to the business processes identified

Attributes of the dimensions

** We have documented the desirable attributes for each of the dimensions we have selected:

- Track Information : [Track name, genre, length, purpose, etc.]
- Album Info : [album name, producer involved, studios, total album length, marketing partners involved, etc.]
- Artist data : [lead vocalist name, award history, support teams, lyricists, instrument artists, etc.]
- Time : [Release date, month, year, time, song duration, occasional effect, all-time popularity, etc.]
- Law : [ISRC code, date of issue, issuing authority, copyright cases if any, etc.]

Bus matrix

	Track Information	Album Info	Artist data	Time	Law
Spotify Popularity	X		X	X	
Spotify Streams	X	X	X	X	X
Spotify Playlist Count	X	X	X		X
Spotify Playlist Reach	X		X	X	
Deezer Playlist Reach	X			X	
Amazon Playlist Count	X			X	
Shazam Counts	X				
Pandora Track Stations	X				

* The Bus Matrix is designed for the chosen Business Goal; it will surely change if the Goal, Business Processes and Dimensions differ.
** It may change on the data provided and on others' perspective also.

Business processes are listed along the vertical axis (rows)

Dimensions are listed along the horizontal axis (columns)

X is noted to show a relationship between the business processes and the dimensions. Blank cells refer to no relationship.

Inference

❖ We know that the **Bus Matrix is an essential tool for planning and designing data warehouse architecture.**

Keeping this in mind, we have drawn the following insights:

- Since most of the business processes are related to **Track Information**, this dimension is central to analyzing our business goal and should be well-defined and consistent across different platforms.
- There is a need to enrich **Album Info** dimension with more data.
- The **Artist Data** dimension is linked to multiple grains, indicating that understanding artist performance across platforms is crucial.
- **Time** dimension also highlights the importance of it's attributes for track performance.

Thank you !