

ASSIGNMENT 4 - ADDRESSING COMPLEXITY

Declaration:

I have read and I understand the plagiarism provisions in the General Regulations of the University Calendar for the current year, found at <http://www.tcd.ie/calendar>. I have also completed the Online Tutorial on avoiding plagiarism 'Ready Steady Write', located at <http://tcd- ie.libguides.com/plagiarism/ready-steady-write>.

Dataset Details:

- **The Tools/Technologies employed:** Google Collaboratory, implemented entirely using Python (Plotly, matplotlib, pandas, and numpy libraries to support the Data Visualisation task).
- **The Dataset:**
 - **General Description:** "Spotify Set" dataset has a shape of (32834, 23). The motivation for visualising this dataset elicited from playlist tracks' segmentation and analysis, coupled with artist's popularity.
 - **Data Types, Attribute Types:**
 - *Categorical Attribute:* track_artist, playlist_name, playlist_genre, playlist_subgenre
 - *Qualitative Attribute:* track_album_release_date
 - *Quantitative Discrete Attributes:* track_popularity, danceability, energy, key, loudness, mode, speechiness, acousticness, instrumentalness, liveness, valence, tempo, and duration_ms.
- **TECHNICAL CONTRIBUTIONS OF THE PAPER:**
 - Data Import, Exploration, and Descriptive Statistics.
 - Visualisation with Plotly and matplotlib:
 - Plotly Express has been redeemed for the creation of '**parallel coordinates**' and '**sunburst**' plots for manifesting correlation, associations, and alliances between the multitudinous variables (categorical, qualitative, and quantitative attributes) and their impressions on the '**track popularity**' (holds quantitative discrete data type). This is furthered by visualisation of playlist_genre popularity hierarchy.
 - Predominantly, 3 plots are constructed for the exhibition and demonstration of 'playlist artist popularity', 'track segmentation to identify 'track popularity'', and 'playlist genre popularity'.
 - Insights through Visualisations:
 - The insights are extracted from the dataset, which could have future purposes, for instance, *Playlist Curation, Marketing and Content Strategy, Artist Collaboration and Licensing, User Engagement, and Data-Driven Insights.*
 - Data interpretation, incorporating a plethora of visual encoding channels & idioms, is strengthened by rendering a visually appealing & illuminative parade of plots contributing to the recognition of patterns, trends, movements, leitmotifs, & intercorrelation (relationships) in the data.

- **FRAME 1: SUNBURST PLOT - WHEEL REPRESENTING ARTIST'S RECORDS AND POPULARITY**
Visual Encoding Channels along with the Justification:: A diverse assortment of visual encoding channels
- **Position:**
 - The position of the different music records(playlist_name) on the outer wheel is dictated by their parent artist's name(playlist_artist), positioned by their popularity(track_popularity). For instance, upon clicking on the artist "Billie Eilish", the structure blossoms and we have music records: "lovely(with Khalid)", "bad guy", and "everything i wanted".

- Now, the sequential order of these music records are dictated by the measure of track_popularity in the increasing degree, found in the metadata upon hovering on the radial slice.
- **Size:**
 - The size of the 'track_artist's radial pie segment(Qualitative data(Categorical)) & the music records' radial slice is indicated by their respective track_popularity value under the same segment. For the playlist artist's name, it's the songs' count which fall under them and the track_popularity. And subsequently, they are located on the wheel in relation to this parameter.
 - For instance, under the artist DaBaby, music records: Suge, VIBEZ, and BOP are placed adjacent to each other in the increasing chain, their 'track_popularity' being: 87, 88, and 93.
- **Color:**
 - The colors of the songs are indicated by their nature of track_popularity, in cahoots with the color palette on the right. Track_popularity drives the color assignment task with smooth operation.
 - For eg., all songs under Sech are colored according to the color associated with the artist: Sech.

Idioms along with the Justification::

- *Sunburst*: I exercised the Sunburst plot for presenting a nested structure lying amongst all the music artists & their tracks. This acknowledges the hierarchical representation & compact visualisation.
- Color coding & Interactivity helped me with the presentation & storytelling that is evinced by the dataset.
- *Critical Analysis*: *Although the Sunburst plot acted like a pillar for expressing diverse information under a single visualisation, label overlapping & limited scalability were the weaknesses I worked with, as it became cluttered with an increased no. of data points. Hence, I picked the top 50 rows for this visualisation.*

FRAME 2: PARALLEL COORDINATES PLOT FOR ASSESSING TRACK POPULARITY

Visual Encoding Channels along with the Justification:: A diverse assortment of visual encoding channels:

- **Position:**
 - For each of the dataset's quantitative attributes, axes are plotted with vertical arrangement with their respective spectrum of values from the dataset. This neatly covers all-round aspects aimed by this plot.
 - Quantitative Discrete Attributes are positioned on the X-axis.
- **Color:**
 - The popular tracks are encoded in a lighter shade, while less popular tracks are encoded in a darker shade, reflected by the adjacent color strip. In the resultant grid of lines, one string of chain of lines is a collection of inputs from each of the attributes. A movement in either of the attribute values along the longitudinal frame, which holds quantitative continuous interval attributes, results in the shift in the track popularity.

Idioms along with the Justification::

- *Parallel coordinates plot*:
 - Exercising a parallel coordinates plot allowed me to deep-dive into the attributes, relationships, & their correlation. For eg, track popularity lying in the proximity of 100 holds water if danceability, loudness, & acousticness lie the the upper ranges of their resp. parallel axes' spectrum.
 - Data Clustering comes to life, allowing me to discern and perceive the potential clusters, what important information they hold, and how that influences the track popularity. For instance, with loudness (-10 to 0), speechiness (0 to 0.2), instrumentality (0), and liveness (0 to 0.2), leads to a fair conclusion that the track popularity could fall under 80-100.
 - Concept of dynamic view and interactive display has been blended with this plot, which is crucial to the dataset storytelling with its insights & judgments.

- *Critical Analysis:* While parallel coordinates plot serves as a useful idiom, Overplotting & Complexity & Cognitive Load become an issue as there's an increase in data points, leading to complex visualisation & would fail to convey crucial info., which can overwhelm the viewers.
- *Line graph:*
 - Here, the ordinal data is encoded & plotted for each of the adjacent song's attributes in a parallel fashion.
 - This helped me show how the diverse set of attributes gave me the authority to manifest the change in the track popularity over the course of all the songs.
- *Color gradient:*
 - Track popularity is tracked by virtue of the concept of color gradient. Here, I've embodied a Continuous Color Scale: Sunsetdark.
 - Color strip on right(adjacent to the plot) signifies spectrum of colors embedded to denote track popularity.
 - The lighter-tone colors at the top of the gradient represent more popular songs, while the darker-toned colors at the bottom of the gradient represent less popular songs.

● **FRAME 3: SUNBURST PLOT - WHEEL FOR PLAYLIST GENRE POPULARITY**

Visual Encoding Channels along with the Justification:: A diverse assortment of visual encoding channels.

- *Position:*
 - The position of the different playlist_subgenres on the outer wheel is dictated by their parent genre(Categorical attribute), positioned their count.For instance, under the playlist_genre Latin, we have playlist_subgenres, Reggaeton, Latin pop, Tropical, and Latin hip hop.
 - Now, the sequential order of these playlist_subgenres are dictated by their count in the increasing degree, found in the metadata upon hovering on the slice.
- *Size:*
 - Size of the subgenres is indicated by their respective 'count' value, so that the difference is palpable in all .
 - For instance, under the playlist_genre Latin, playlist_subgenres: Reggaeton, Latin pop, Tropical, & Latin hip hop are placed adjacent in the chain arrangement, 'count': 949, 1262, 1288, & 1656.
- *Color:*
 - Color of genres is indicated by overall aggregate track_popularity of music songs lying under resp. genre.
 - For eg, all rap subgenres are colored according to their track_popularity color palette. Rap subgenres: Trap, Hip Hop, Gangster Rap, and South. Each subgenre is assigned a color from the color palette on the right according to their track_popularity. Taking mean of the subgenre's aggregate popularity values gives us the track_popularity of the genre, thereby assigning the designated color to it from the color spectrum.

Idioms along with the Justification::

- *Sunburst:* I exercised the Sunburst for presenting a nested structure that lies amongst all the genres and subgenres. This acknowledges the hierarchical representation and compact visualisation.
- Color coding & Interactivity helped me with the presentation & storytelling evinced by the dataset.
- Sunburst plot acted like a pillar for expressing diverse information under a single visualisation.
- *Critical Analysis:* Although it's perfect to use the Sunburst plot in my case, it might be unsuitable for every data structure. Limited interactivity & deep hierarchical complexities come into play further on.

Visualisation's Strengths: The visualisation cogently captures all the relationships between the elements for genre_popularity, track_popularity, and playlist_artists' work insights.

Visualisation's Weaknesses: Parallel Coordinates is a tad bit cluttered & confusing. Large data rows can get the best of a Sunburst plot, so I visualised the data by taking a segment of the data points (50). Deep hierarchical complexities & limited interactivity dictate the terms. This plot can also be unbefitting to other data structures.

[The Output Video Link](#) | [Github](#) | [Project Files](#)

