**Project Requirements:**

**Data validation and Preprocessing:** Clean andpreprocess the data. This step includes handling any missing values, outliers and ensuring that the data is ready for analysis.

**Chart 1. Distribution of Released Year**

* Use ML algorithms to analyse the distribution of released years for tracks on Spotify.
* Visualize the distribution using a histogram to show the frequency of tracks released in different years.

**Chart 2. Relationship between Released Year and Streams**

* Perform predictive analysis to understand the relationship between the released year and the number of streams using ML regression techniques.
* Develop a predictive model to estimate the number of streams based on the released year.
* Create a scatter plot visualization to depict the relationship between released year and streams.

**Chart 3. Distribution of Streams by Playlist Presence**

* Aggregate streams by the presence of tracks in Spotify playlists (in\_spotify\_playlists) using ML techniques.
* Visualize the distribution using a stacked bar chart, where each bar represents playlist presence and the stacked segments represent the proportion of streams from each category.

**Chart 4. Streams Trend Over Time**

* Analyse the trend of streams over time using ML time series analysis techniques.
* Visualize the trend using a line chart, where the x-axis represents time (e.g., months or years) and the y-axis represents the number of streams.

**Chart 5. Correlation Matrix Heatmap**

* Generate a correlation matrix using ML techniques to explore the relationships between different variables (released year, streams, BPM, danceability, valence, energy, etc.).
* Visualize the correlation matrix using a heatmap, where each cell represents the correlation coefficient between two variables.

**Chart 6. Distribution of Streams by Artist Count**

* Aggregate streams by the count of artists involved in each track (artist\_count) using ML techniques.
* Visualize the distribution using a bar chart, where each bar represents the number of artists and the height represents the number of streams.

**Chart 7. Distribution of Streams by Key**

* Aggregate streams by musical key (key) using ML techniques.
* Visualize the distribution using a pie chart, where each slice represents a key and the size represents the proportion of streams.

**Chart 8. Danceability vs. Valence Scatter Plot**

* Analyse the relationship between danceability and valence using ML techniques.
* Create a scatter plot visualization to depict the relationship between danceability and valence, where each point represents a track.

**Note:** *Please use (appropriate) ML models wherever necessary even if not specified – apply you critical thinking and analytical skills*

**TECHNICAL EXPECTATIONS:**

* Use appropriate tools and libraries when necessary
* Data Analysis – Pandas, NumPy, etc.
* Visualization – Matplotlib, Seaborn, Plotly
* Interactive Dashboard – Dash/Streamlit

**CAPSTONE PRESENTATION:**

* Include the preprocessing methods employed in your project
* Provide the insights and interpretations of the visualizations
* Explain in detail the analysis and the visualization techniques used and why?
* Provide explanation for the choice of the ML algorithm