

1.4 Assignment: Capstone Project

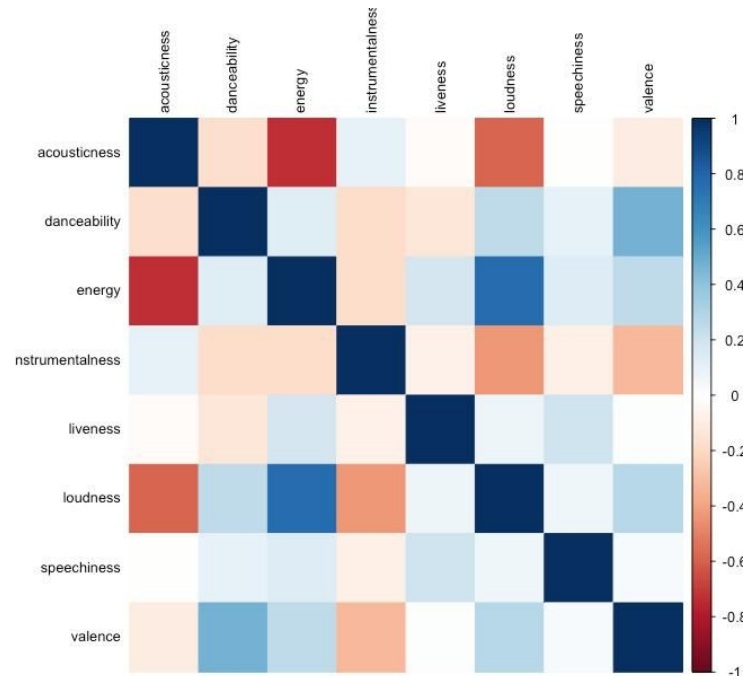
By: Varsha Iyengar, Jeshal Mohapatra,
Krish Tiwari, Hugo Vides

Heatmap Analysis

Tools used: Tidyverse, ggplot2, corrplot

Relationships:

- Acousticness vs Energy
- Danceability vs Acousticness
- Loudness vs Acousticness
- Energy vs Loudness
- Instrumentalism vs Danceability
- Instrumentalism vs energy
- Valence vs Danceability
- Liveness vs Speechiness



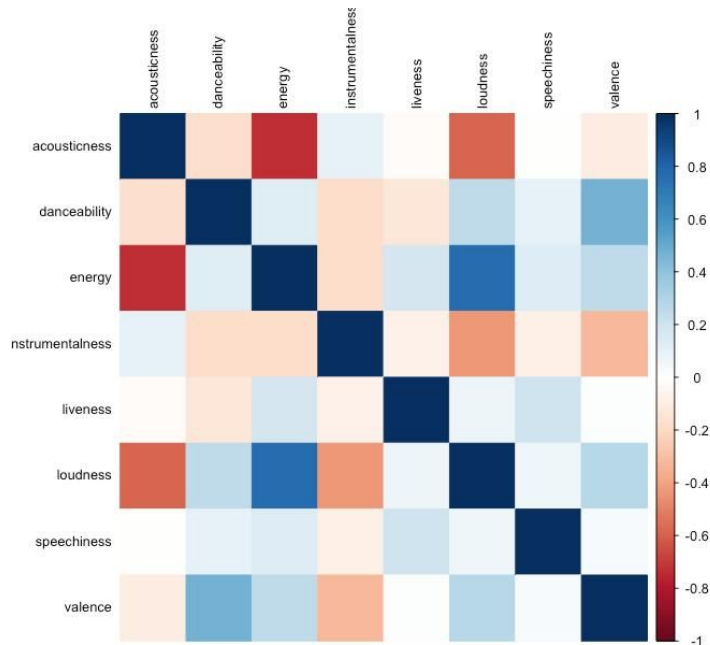
Heatmap Analysis (Continued)

Steps:

- Imported and explored the dataset.
- Visualized the distributions.
- Calculated and visualized correlations.
- Created a correlation plot to show relationships and correlation strengths.

Range:

- 1 is a strong positive correlation or a more blue color.
- -1 is a strong negative correlation or a more red color.
- 0 is no correlation or a more whiter color.



Model Enhancement


Data Cleaning:

- Action: Removed outliers
- Benefit: Data is cleaned leading to more accurate clustering and recommendations.

Advanced Clustering Algorithms:

- Algorithm used: DBSCAN, Agglomerative Clustering
- Benefit: Better at finding clusters in the data with varying densities and shapes.

Feature Scaling:


- Action: Standardized data in StandardScaler.
 - Benefit: Ensures that all features contribute equally to the clustering process.
- 

Finding from Data Exploration

Distribution of Features:

- Histogram: Plotted distributions for features.
- Histogram helps understand the spread and commonality of feature values in the dataset.

Correlation Analysis:


- Correlation Matrix: Visualized the correlation between features using a color-coded matrix.
 - Correlation Matrix identifies which features are strongly related or independent.
- 

Comparison of Different Clustering Algorithms

DBSCAN:

- Parameters: Epsilon, Minimum Samples
- Evaluation:
 - Silhouette Scores: Measures how similar an object is to its own cluster compared to other clusters.
 - Silhouette scores vary with different epsilon values showing DBSCAN's sensitivity to parameter settings.

Agglomerative Clustering:

- Parameters: Number of Clusters
 - Evaluation:
 - Silhouette Scores: Provides fixed cluster count
 - Consistent silhouette scores result for a fixed number of clusters.
- 

Incorporating User Feedback in the Future

Current Feedback Collection:

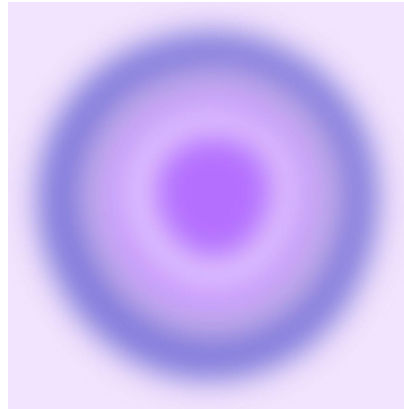
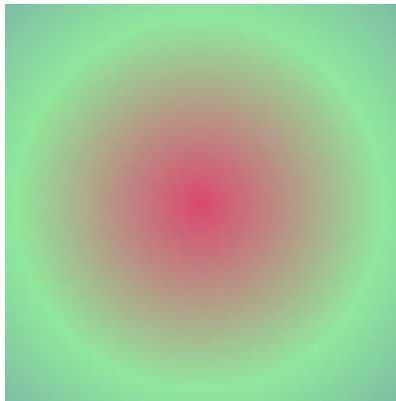
- Method: Users rate recommended songs on a scale of 1 to 5
- Process: Collects ratings and feedback for continuous improvements.



Innovation

We will implement a system where you enter your emotion you are feeling into the code and based on that it will change the recommended songs to match the vibe of your emotions.

The moodboard:



The background is a solid pink color. In the top right corner, there is a decorative pattern of overlapping geometric shapes, including triangles and squares, in various shades of pink and magenta.

Thank you for listening!