Aybars Salim Güney 27958

Spotify Stream Analysis Code Report

Executive Summary

This code report provides a detailed overview of the analysis conducted on my Spotify stream history from 2019. The project aims to explore music preferences through seasonal trends and genre analysis. The analysis encompasses data cleaning, exploratory data analysis, visualizations, hypothesis testing, and reflection on findings.

Motivation

The motivation behind this project is to gain a deeper understanding of my music preferences and how they evolve over different seasons. By analyzing the Spotify stream history, I aim to uncover patterns, discover the most common genres, and explore how my music choices vary throughout the year.

Data Source

The data for this project originates from my personal Spotify stream history of 2019. The dataset includes information such as timestamps, track details, and popularity scores. The data was collected directly from the Spotify platform.

Data Analysis

Techniques Used

The analysis employs a combination of data cleaning, exploratory data analysis (EDA), and hypothesis testing techniques. The primary stages of the analysis include:

Data Cleaning: Preprocessing the raw data to handle missing values, empty genres, and popularity scores of 0.

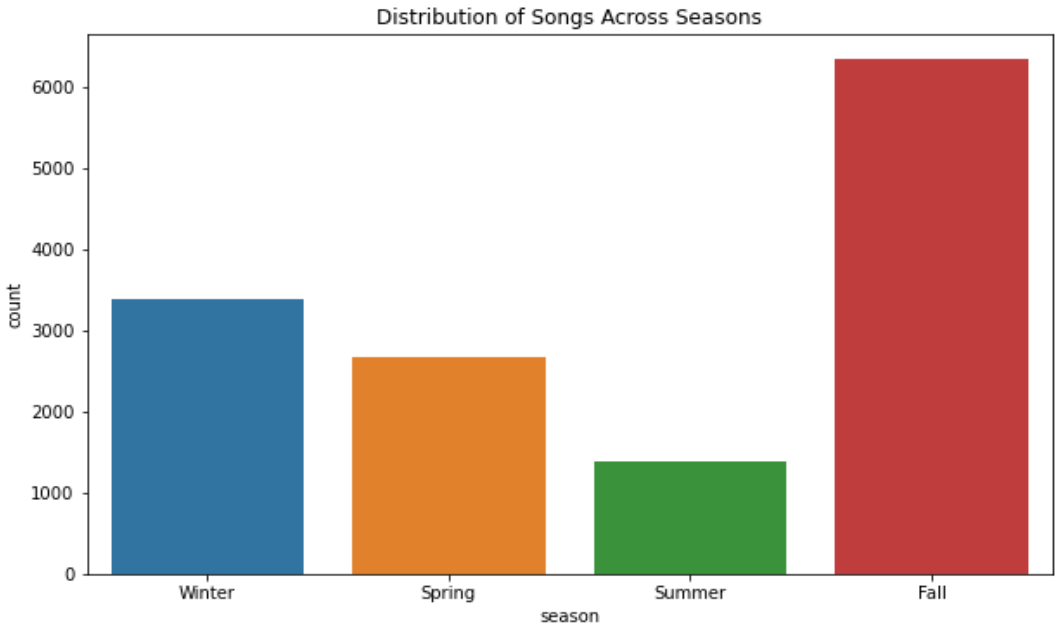
```
# Handling Missing Values
df = df.dropna()

# Handling Empty Genres and Popularity of 0
df = df[(df['track_genres'].notna()) & (df['track_genres'] != '')]
df = df[df['track_popularity'] != 0]
```

Seasonal Analysis: Categorizing the timestamps into seasons to investigate the distribution of songs and the most common genres in each season.

```
# Extract month and season
df['month'] = df['timestamp'].dt.month
seasons = {1: 'Winter', 2: 'Winter', 3: 'Spring', 4: 'Spring', 5: 'Spring', 6: 'Summer', 7: 'Summer', 8: 'Summer', 9: 'f
df['season'] = df['month'].map(seasons)
```

Visualizations: Generating visualizations such as count plots, boxplots, and other charts to represent trends in music preferences.

```
# Explore the Data and Visualize
plt.figure(figsize=(10, 6))
sns.countplot(x='season', data=df)
plt.title('Distribution of Songs Across Seasons')
plt.show()

plt.figure(figsize=(12, 8))
sns.boxplot(x='season', y='track_popularity', data=df_top_genres, hue='track_genres')
plt.title('Distribution of Track Popularity Across Top 3 Genres and Seasons')
plt.show()

plt.figure(figsize=(10, 6))
sns.countplot(x='season', hue='track_genres', data=df_top_genres)
plt.title('Top 3 Genres Occurrence in Each Season')
plt.show()
```

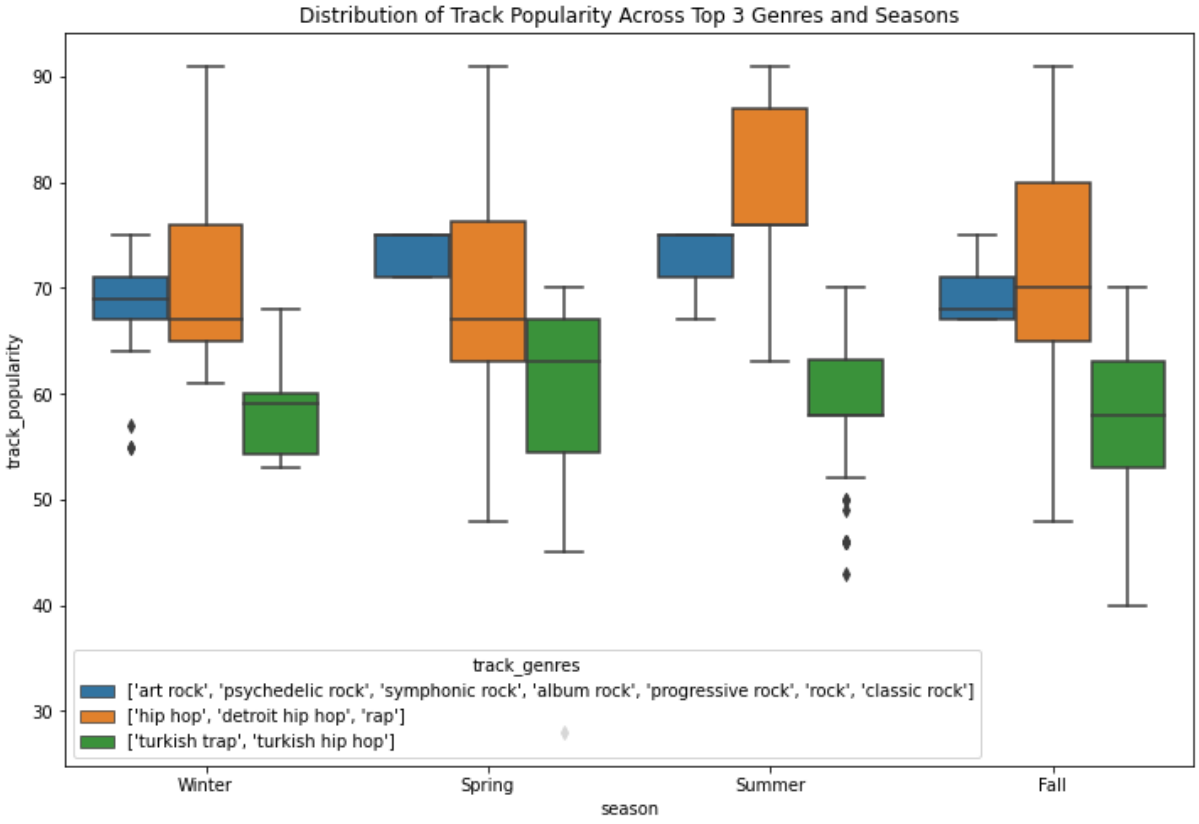


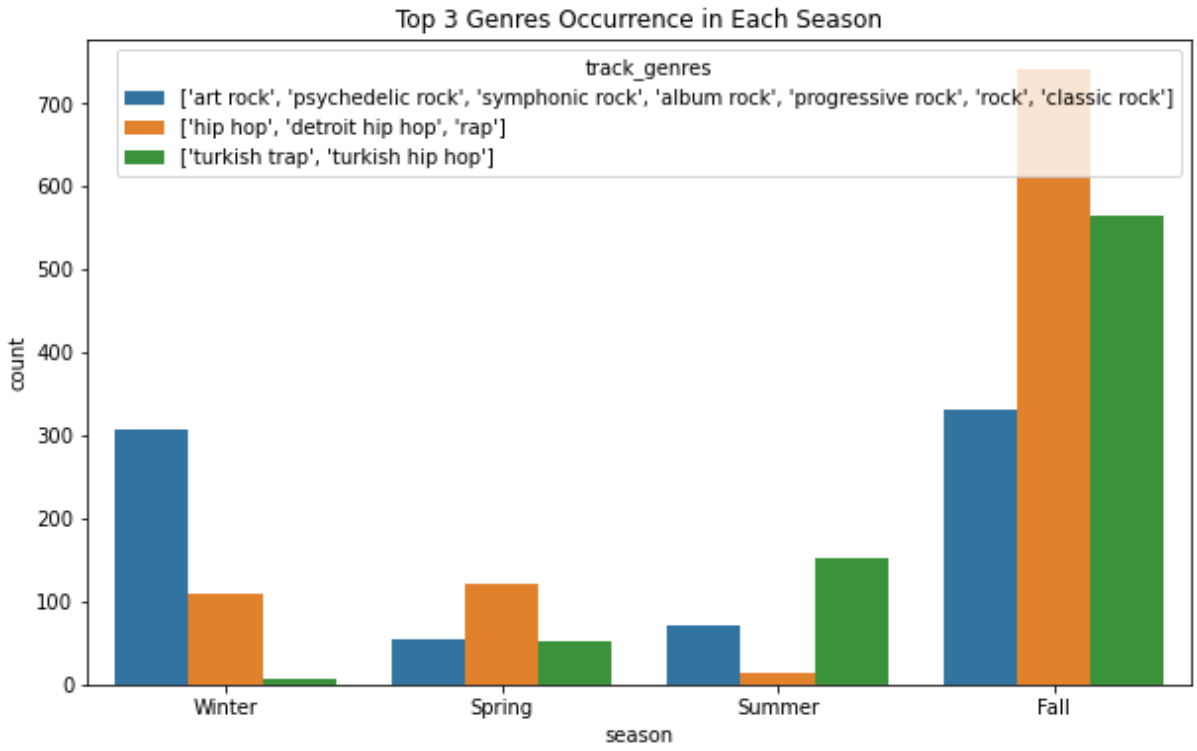
Hypothesis Testing: Employing a chi-square test for independence to examine if the most common genres change seasonally.

```
# Hypothesis:
# Null Hypothesis (H0): There is no significant evidence that the most common genres remain constant across seasons.
# Alternative Hypothesis (H1): There is evidence that the most common genres change over time seasonally.
contingency_table = pd.crosstab(df_top_genres['season'], df_top_genres['track_genres'])

# Perform Chi-Square Test for Independence
chi2_stat, p_value, _, _ = chi2_contingency(contingency_table)

# Print Results
print("Chi-Square Test Statistic:", chi2_stat)
print("P-value:", p_value)

# Decide whether to accept or reject the null hypothesis
alpha = 0.05  # significance level
if p_value < alpha:
    print("Reject the null hypothesis. There is evidence that the most common genres change over time seasonally.")
else:
    print("Fail to reject the null hypothesis. There is no significant evidence that the most common genres change over
```

Findings

The analysis reveals insights into my music preferences, highlighting the most common genres and their variations throughout the seasons. This includes understanding which genres are more prominent during specific times of the year.





Limitations and Future Work

Limitations

- The analysis is based on historical data from 2019 and may not reflect current music preferences.

- The dataset is limited to the available Spotify stream history and does not include external factors influencing music choices.

Future Work

- Explore real-time data analysis to understand current music preferences.

- Incorporate additional external data sources, such as weather data or mood indicators, to enhance the analysis.

- Extend the analysis to cover more recent years for a comprehensive view of evolving music tastes.

Conclusion

The Spotify stream analysis project provides valuable insights into music preferences and their seasonal variations. The combination of data cleaning, exploratory data analysis, visualizations, hypothesis testing, and reflection on findings contributes to a comprehensive understanding of the dataset.