

Exploratory Data Analysis Report: Netflix Content Global Trends

Student Roll No: SU92-MSAIW-S25-013

Subject: Advanced Data Analytics

Project Link: [Kaggle Notebook](#)

Dataset: [Netflix Movies and TV Shows](#)

1. Executive Summary

This report analyzes the composition of Netflix's library, totaling 8,797 titles. The objective was to identify dominant content types, primary geographic contributors, and significant categorical relationships using frequency distribution and Chi-Square statistical testing.

2. Univariate Analysis: Categorical Distributions

2.1 Content Type: Movies vs. TV Shows

Netflix remains a movie-centric platform.

- **Movies:** 6,131 titles (**69.69%**)
- **TV Shows:** 2,666 titles (**30.31%**)

2.2 Geographic Dominance

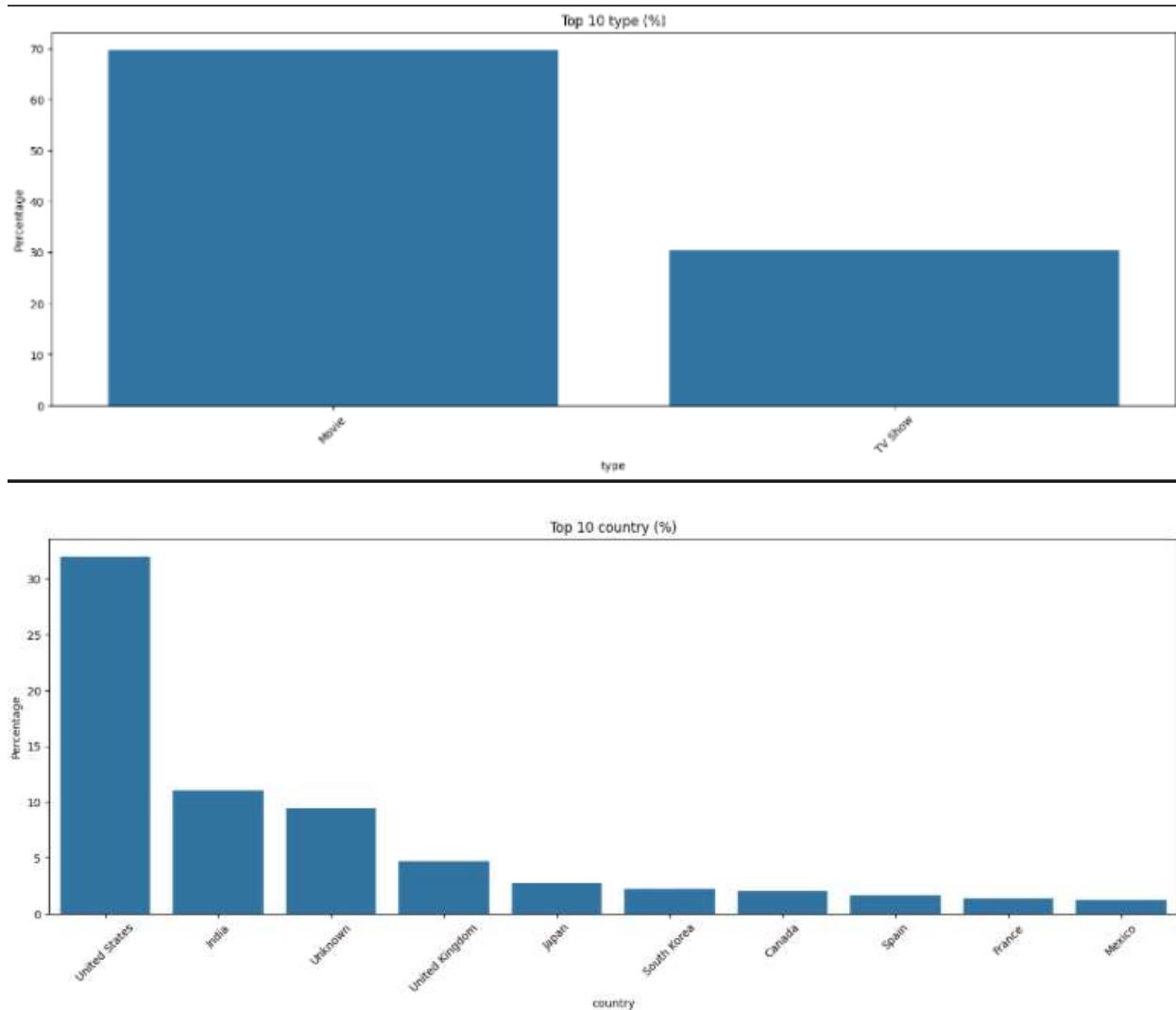
The United States is the primary content provider, followed by India. A notable portion of the data (**9.44%**) contains "Unknown" origins, which suggests a need for further data cleaning in future iterations.

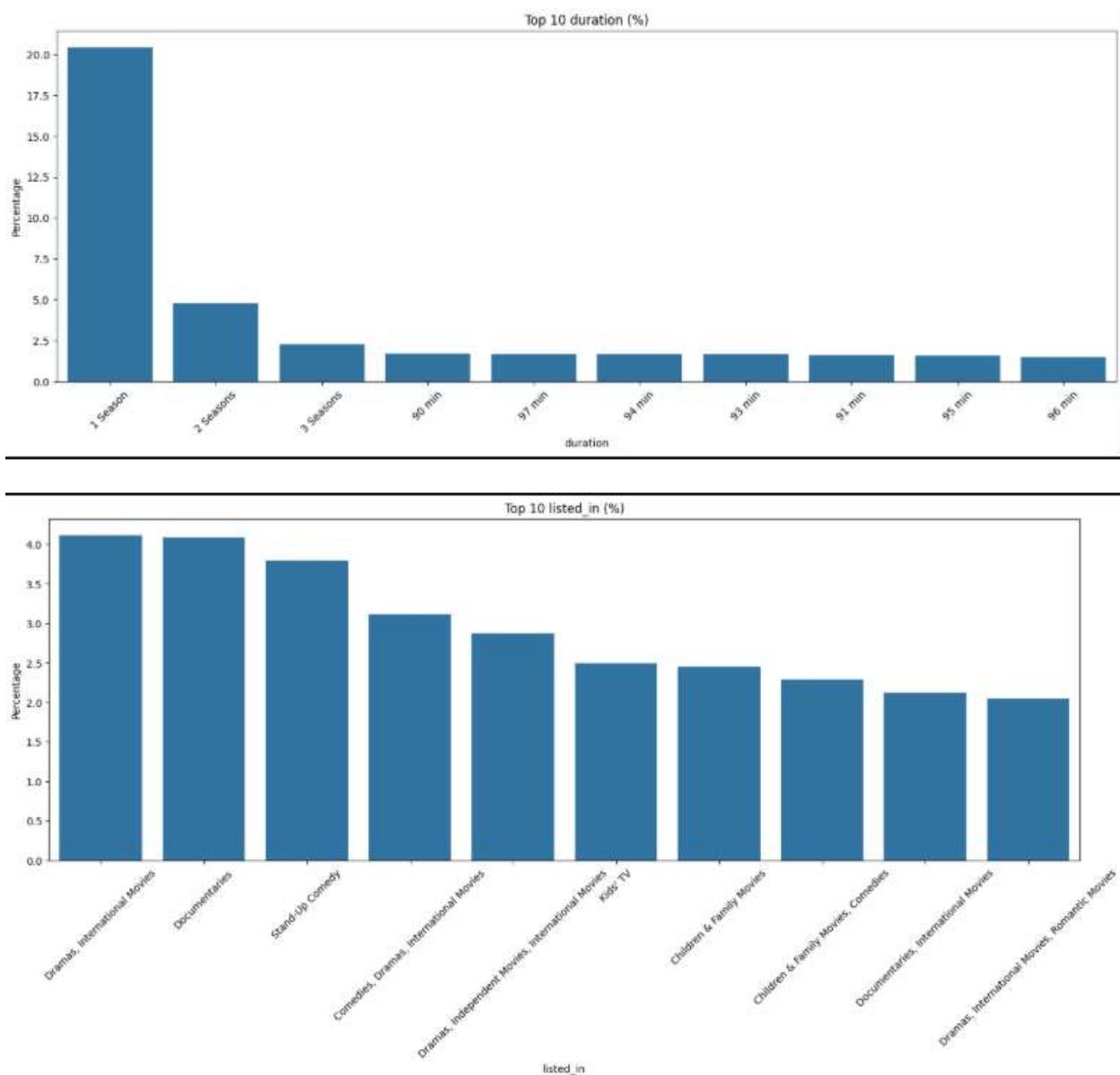
- **Top 3 Countries:** USA (31.97%), India (11.05%), Unknown (9.44%).

2.3 Duration and Genre (Listed In)

- **Duration:** The most frequent format is the "**1 Season**" TV show (20.42%), indicating a high volume of limited series or new acquisitions that haven't yet reached a second season.

- **Genres: Dramas and International Movies** are the most common listings (4.12%), highlighting Netflix's focus on non-English and dramatic storytelling to appeal to a global audience.





3. Bivariate Analysis: Statistical Relationships

To understand if these categories are independent or related, we performed **Chi-Square Tests of Independence**.

3.1 Significant Correlations ($p < 0.05$)

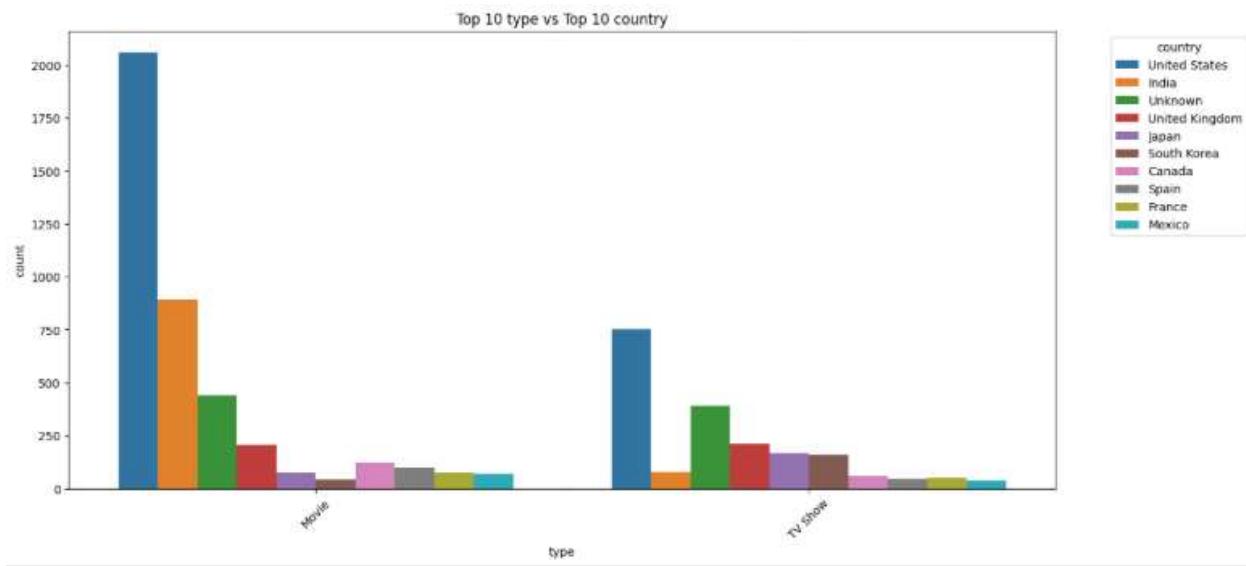
- **Type vs. Country, Duration, and Listed_in ($p=0.0000$):** There is a highly significant relationship between the content type (Movie/TV Show) and its country of origin and

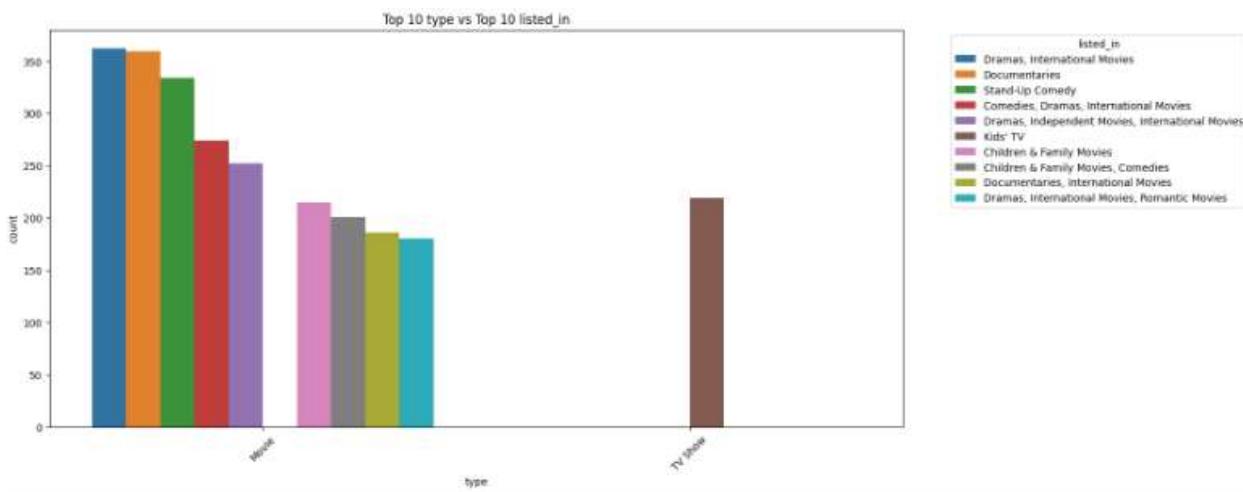
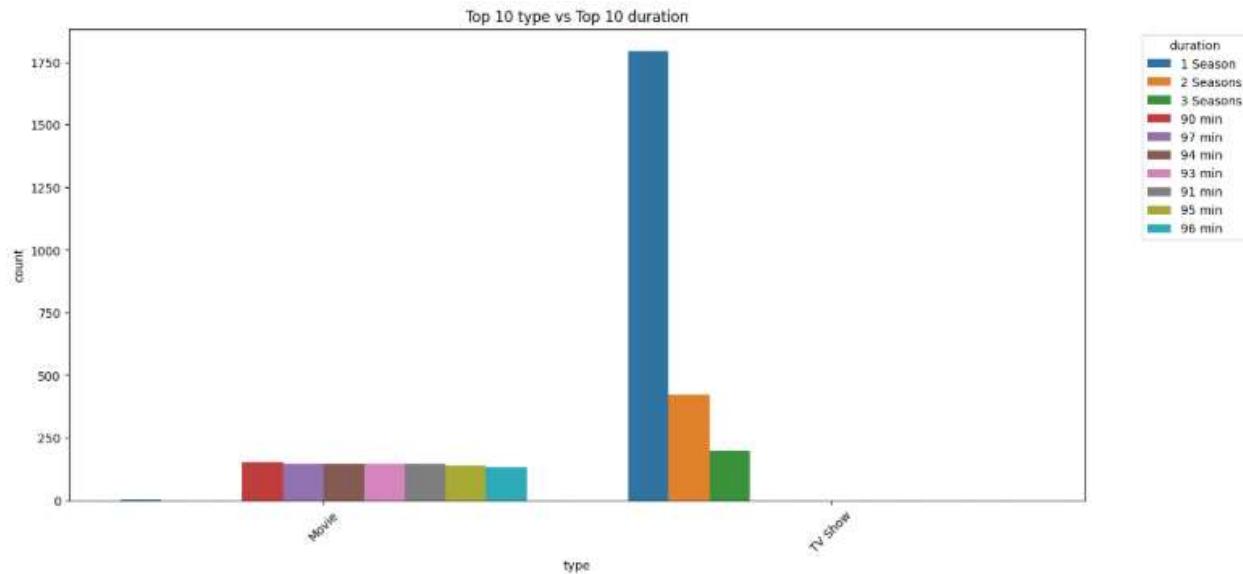
genre. For example, India produces a vastly higher ratio of Movies compared to TV Shows compared to Japan.

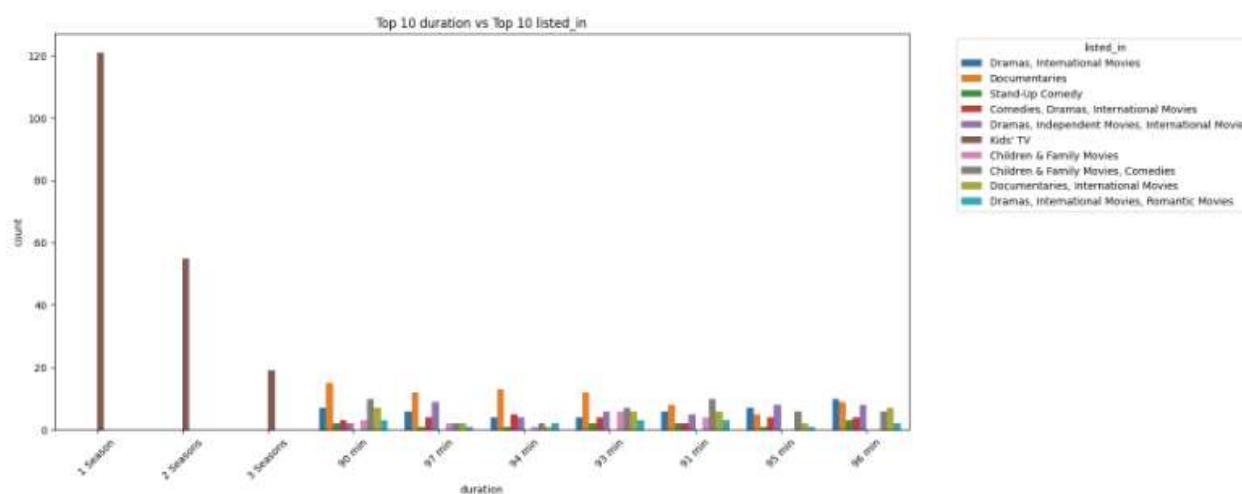
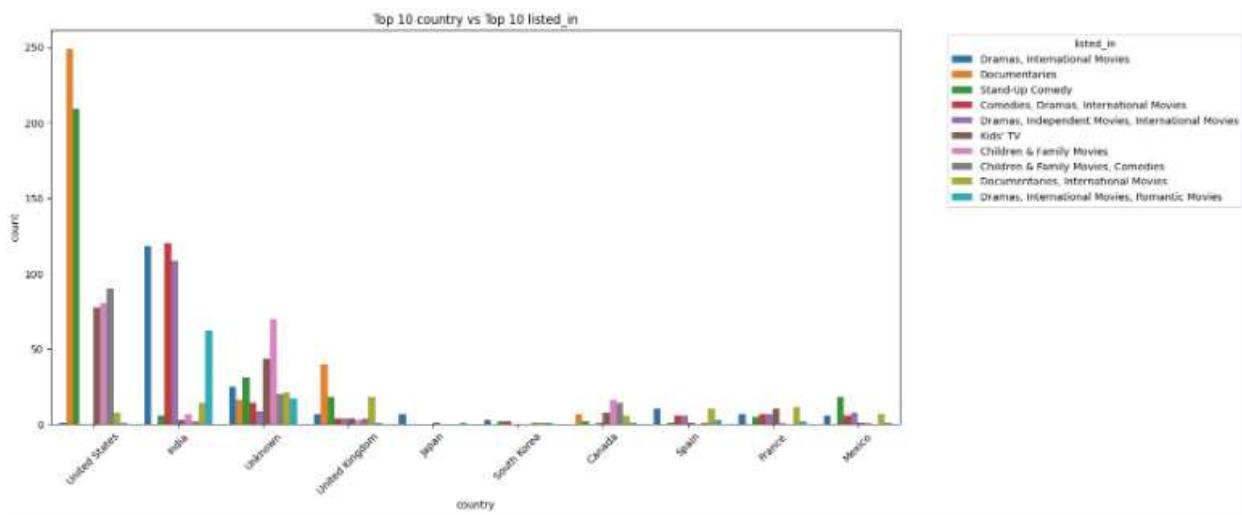
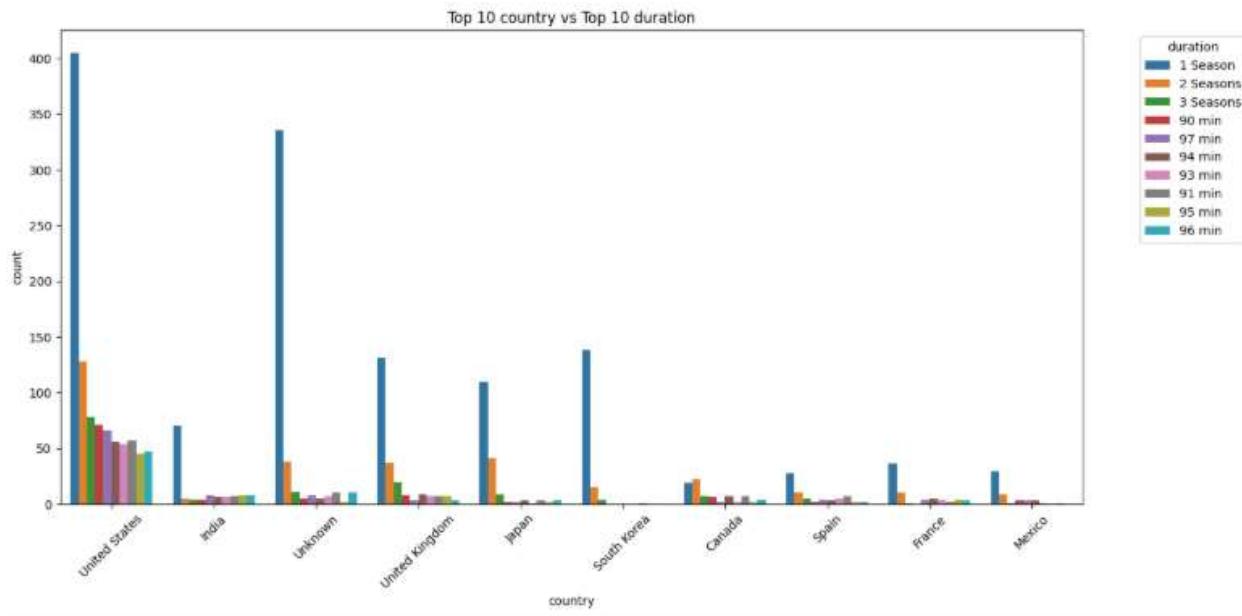
- **Country vs. Duration (\$p=0.0000\$):** The length of content is heavily dependent on the country. Western markets show higher variance in movie runtimes, while international markets lean heavily into multi-season TV shows.

3.2 Non-Significant Relationships

- **Country vs. Listed_in (\$p=1.0000\$):** Interestingly, the country of origin does not statistically dictate the genre. This suggests that Netflix maintains a diversified genre portfolio (Comedies, Dramas, Documentaries) across all major regions.
- **Duration vs. Listed_in (\$p=0.0838\$):** This relationship is not significant at the 5% level, suggesting that a genre's popularity isn't necessarily tied to how many seasons or minutes it lasts.







4. Key Insights & Conclusion

1. **Movie Dominance:** Despite the rise of "binge-watching" TV culture, movies still make up nearly 70% of the Netflix library.
2. **Global Expansion:** International content (Dramas/International Movies) is the largest genre category, proving that Netflix's growth strategy relies heavily on non-US markets.
3. **Statistical Dependency:** The Chi-Square results prove that content strategy is not random; the type of content (Movie/TV) is strictly tied to the country of origin and the intended duration.

Chi-Square Heatmap

