

# THE CURIOSITY CUP 2022

## A Global SAS® Student Competition

### Team Data Campers

#### Influence of Netflix Shows on European Tourism

### Introduction

Over the years content consumption has migrated to various online sources. Streaming platforms became the main source of entertainment for billions of people. Netflix alone grew its membership from 2.5 million in 2005 to 167.1 million in 2019 (*Netflix Historical financials and membership growth*, 2022), making it one of the most popular streaming services around the globe.

Meantime, international tourism has become incredibly accessible, compared to the previous century. In 1990, The United Nations World Tourism Organization (UNWTO) estimated that there were 439.4 million international tourist arrivals, whereas in 2018, this number grew to 1.4 billion worldwide (Roser, 2017).

And as content is created and filmed in various countries, fans are attracted to visit these places and live the experience of their favorite shows and movies.

### Aim

This report aims to find the correlation between Netflix shows and tourism in Europe. The purpose of our project is to examine whether tourism prospers after a successful release of a movie or show on Netflix in the country where it was filmed. By doing so, we can establish the significance a Netflix show/movie has on foreign traveler's country of destination.

### Data

To be able to study the correlation between Netflix shows and tourism in Europe, we worked with a main dataset that gave us the needed information on the different TV shows and their details such as: title, cast, director, country, rating, duration, date that it was added on Netflix, etc. We then merged with this dataset another one containing information on tourism in Europe from 1990 to 2021, specifying the different countries in Europe and the number of arrivals at tourism accommodation establishments during each month. Further on in this report we will explain the steps taken to clean both data sets and merge them. The data set for the Netflix shows was found on Kaggle (source: <https://www.kaggle.com/shivamb/netflix-shows>) , and the dataset for tourism was found on the Eurostat website ([https://appsso.eurostat.ec.europa.eu/nui/show.do?dataset=tour\\_occ\\_arm&lang=en](https://appsso.eurostat.ec.europa.eu/nui/show.do?dataset=tour_occ_arm&lang=en) ).

Variable name	Description	Non missing values	Additional information
show_id	Unique ID for every Movie / Tv Show	100%	Key - unique identifier for the shows
type	Identifier - A Movie or TV Show	100%	2 categories (Tv Shows - Movies)
title	Title of the Movie / Tv Show	100%	
director	Director of the Movie	70%	
cast	Actors involved in the movie / show	91%	List of actors
country	Country where the movie / show was produced	91%	List of countries
date_added	Date it was added on Netflix	100%	
release_year	Actual Release year of the move / show	100%	
rating	TV Rating of the movie / show	100%	17 categories
duration	Total Duration - in minutes or number of seasons	100%	
listed_in	Genre	100%	List of genres
description	The summary description	100%	
<b>Total observations</b>		8807	

Variable name	Description	Non missing values	Additional information
TIME	Month and year	100%	From January 1990 to December 2021
GEO	Country	100%	38 countries
C_RESID	Reporting country	100%	
UNIT	Number	100%	
NACE_R2	Types of accomodation	100%	5 Types of accomodation
Value	Tourism value	100%	

## Data Cleaning

We had to clean two datasets: the Netflix dataset and the tourism dataset.

For the Netflix dataset, we started by removing the missing values. Then for some of the movies, we had more than one country. We split the countries into separate columns and transposed them and encoded for each movie whether it was shot in each of the countries (1) or not (0). Moreover, we extracted the month and the year that each show was released in. Also, the shows had the length either in duration if it were a movie or number of seasons if it were a TV show, so we also separated into two different columns and the average was calculated for each country. For the cast we calculated the count of actors in each show and displayed the total number, then also calculated the average per country. The variables were grouped by country and by month.

For the tourism dataset we used Enterprise Guide® software to remove all the observations for the local travel, as we aim to discover international travel only. Then, as we do not require the division by the accommodation type, we removed the variable specifying the type of accommodation establishment. Dataset includes percentage changes compared to the same period last year and two years ago, which was excluded as well. As the raw dataset had country code, and not the full country name, we had to extract full country names to an excel file and merge it with the Eurostat data. Dataset also contains missing values and special symbols, which we removed to clean the dataset. We also removed total for local and international travel, as well as aggregated numbers for the whole Europe. Finally, we only kept data for 2019, as this is the year for which we would like to make an analysis on travel data. The variables for this dataset were also grouped by country, and then we were able to merge the two datasets using an inner join on the "Country" variable.

## Analysis

To conduct this project, we used the product SAS Viya for learners by uploading the cleaned dataset. We included some additional variables in the section Prepare Data and then in the Section Explore and visualize we created an eight-tab dashboard to visualize interesting relations between the variables of the Netflix shows and the tourism information. We were able to construct different chart types as a Word Cloud, correlation matrix, bar charts, timeline plots and histograms.

### **WORD CLOUD AND TV SHOWS VS MOVIES BY COUNTRY**

As an initial step in the analysis, it is possible to observe in figure 1 of the appendix (word Cloud) the relevance of specific countries in the tourism industry. Some of them are France, Spain, Italy, and the UK, followed by Turkey, Austria, and Portugal. Comparing this information with the number of TV shows movies (figure 2) (tv shows vs movies by country), we can observe that the country with highest number of shows is UK, followed by France and Spain. This fact shows a relation between tourism and Netflix shows added produced in those 4 countries in Europe.

### **CORRELATION MATRIX**

To go further with the data analysis, the team decided to generate a correlation matrix that would allow us to have an overview of the main correlation between the genres of Netflix movies and total tourism across Europe. By analyzing this matrix, we were able determine the strongest genres correlated to tourism in the region. In that order, the results shown in figure 3 demonstrate that the main variables related to tourism are the following: children-family movies, comedies, international movies and dramas. Clearly, the highest correlated genre is comedies, which holds a correlation of 0.733 versus the variable of tourism, followed by international movies which accounts for 0.697 of correlation. In further figures this correlation is more detailed, zooming into the values for each country of the region.

### **COMPARISON OF MOVIES AND TV SHOWS VS TOURISM**

In figure 3 we can see that the countries with the highest numbers of arrivals have on average higher number of TV shows and movies produced. We can assume that the countries with large exposure on a streaming platform have higher attractiveness for international tourists. And as we mentioned earlier, countries on average have larger number of movies than TV shows.

### **TOURISM VS COMEDIES AND TOURISM VS INTERNATIONAL MOVIES**

The graph on the left shows us tourism vs comedy shows in 2019, and the graph on the right shows tourism vs international movies in 2019. The team decided to study these variables together to compare the impact of comedies and international movies on tourism in Europe. The results show us that there is a correlation between comedies and total tourism as well as between movies and total tourism, as was described in the correlation matrix. However, by analyzing both graphs we can see that the correlation is higher between comedies and total tourism.

## **LENGTH VS TOURISM GRAPH**

In the length vs tourism graph, both movies and tv shows are analyzed in terms of tourism in Europe. The left graph, corresponding to tv shows, exhibits a trend between the number of seasons that tv shows have over a span of one year and the number of tourists that visited the region. As per the graph in the right, tourism follows its seasonality regardless of the average length of the movies. It is valuable to mention that the data studied for each month is the cumulative content on Netflix before the start of each month, whereas the data for tourism is selected for the total period of each month. This structure was chosen in order to have a timespan to display a possible effect on tourism after certain content in the platform.

## **TV SHOWS VS TOURISM BY COUNTRIES**

France, Spain, Italy and UK have the largest number of visitors in the EU. We can observe that for the first half of the year 2019, number of TV shows correlate with the number of tourists in each of these countries. However, as tourism declines after the seasonal boom, the number of TV shows continue to rise, suggesting that release dates of the TV shows do not correspond to the seasonality of tourism.

## **MOVIES VS TOURISM BY COUNTRIES**

The four graphs display movies vs tourism in the following countries: France, Spain, Italy, and U.K. As mentioned before, these countries were chosen since they had the largest number of visitors in Europe during 2019. Up until mid-summer, around the month of July, we can see that tourism increased as the total number of movies released was also increasing. After the summer peak, we can see a decline in tourism but a steady increase in movies released, also suggesting that the tourism peak could be related to other factors.

## **Conclusion**

To conclude, we have not found sufficient evidence suggesting that Netflix shows influence the tourism in European countries, although we do not discard the fact that they do affect tourism to some extent. To further study the effect of Netflix shows on tourism and understand the relationship better, a few suggestions could be comparing tourism over different time spans and choosing several specific shows and analyzing tourism in their countries a few months later.

## References

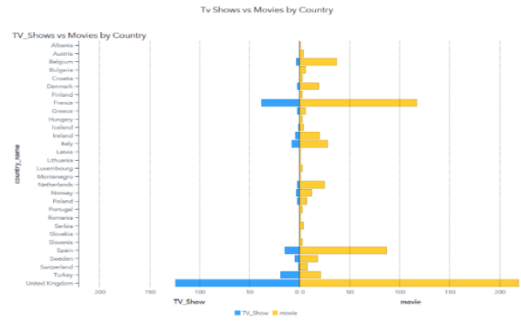
Roser, M. (2017, April 24). *Arrivals by world region*. Our World in Data. Retrieved January 28, 2022, from <https://ourworldindata.org/tourism>

Wikimedia Foundation. (2022, January 27). *Netflix Historical financials and membership growth*. Wikipedia. Retrieved January 28, 2022, from [https://en.wikipedia.org/wiki/Netflix#Historical\\_financials\\_and\\_membership\\_growth](https://en.wikipedia.org/wiki/Netflix#Historical_financials_and_membership_growth)

# Appendix



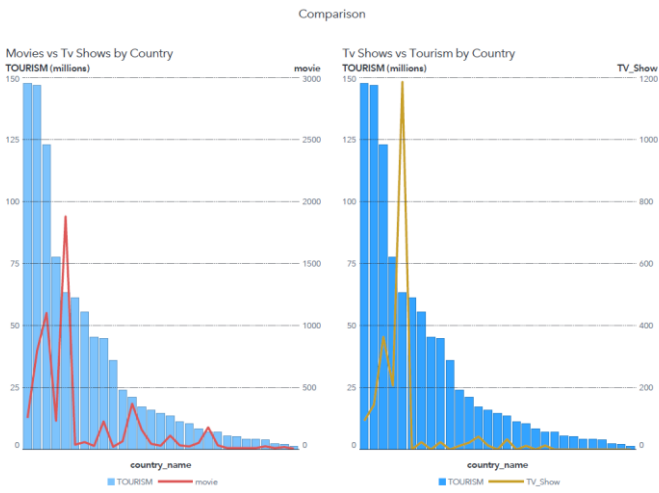
**Figure 1. Word Cloud**



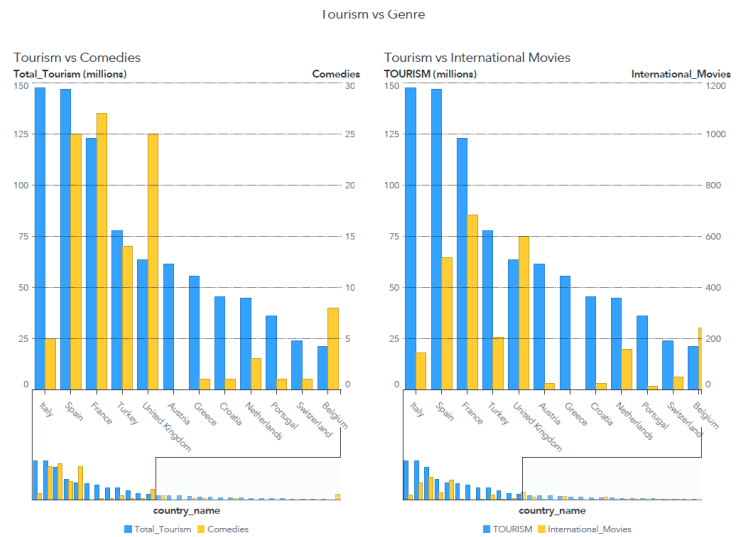
**Figure 2. Tv shows vs movies by country**



**Figure 3. Correlation matrix**



**Figure 4. Movies and tv shows vs tourism**



**Figure 5. Tourism vs top movies genres**

Length vs Tourism

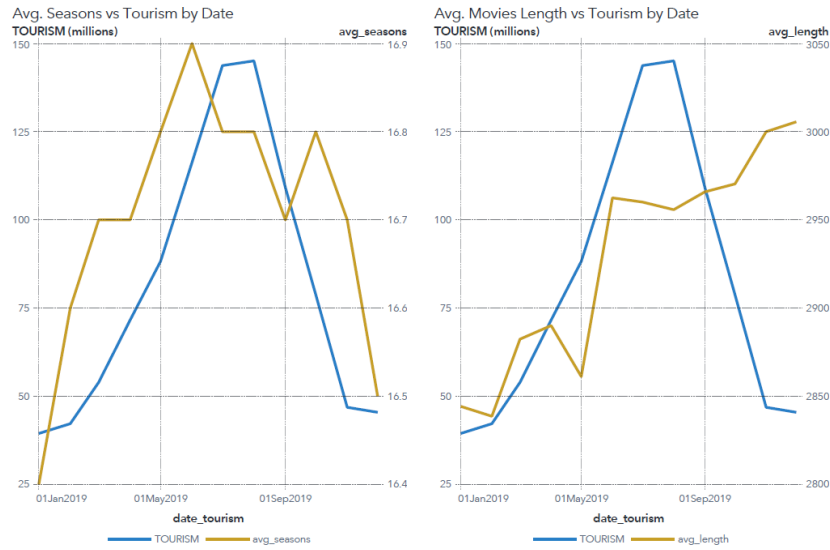


Figure 6. Movies and tv shows length vs tourism

Tv Shows vs Tourism

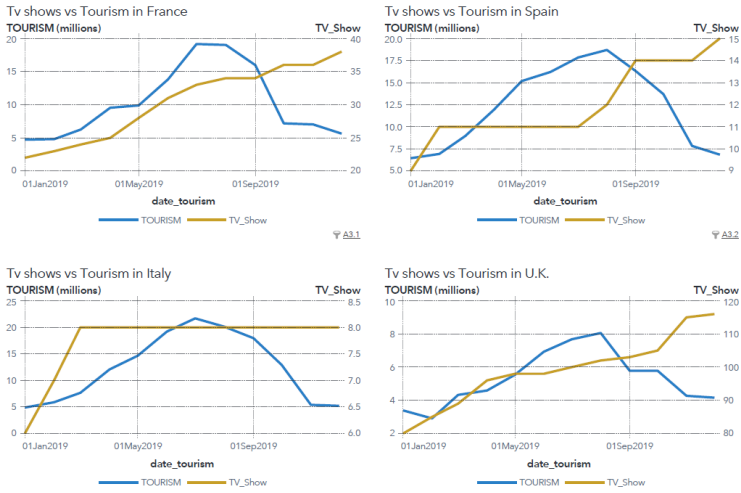


Figure 7. Tv shows vs tourism in top countries

Movies vs Tourism

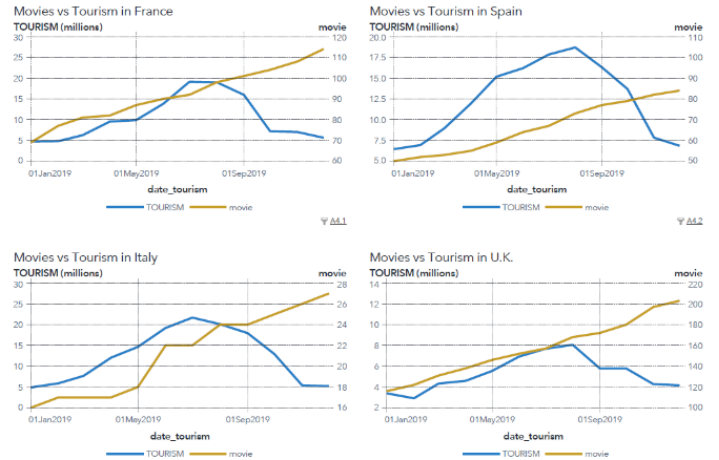


Figure 8. Movies vs tourism in top countries