

# HAPPY COUNTRIES AND NETFLIX



**Group members: Estefany Amado, Christina Laigo, Dean Tolmachoff, Ed Torrente**

QUESTION:

IS THERE A CORRELATION BETWEEN NETFLIX USAGE AND A COUNTRY'S HAPPINESS INDEX?



# DATASETS USED

- Kaggle

- Netflix with country availability:  
<https://www.kaggle.com/syedmubarak/netflix-dataset-latest-2021>
- Happiness Index by country:  
<https://www.kaggle.com/sonam10/world-happiness-based-on-un-data>
- Avg Internet Price by country:  
<https://www.kaggle.com/cityapiio/world-cities-average-internet-prices-2020>

# DATA CLEANUP, EXPLORATION, & LIMITATIONS

## Data Cleanup:

- Netflix data: Deleted columns
- Original dataset saved as 3 workbooks: Genre, Languages, Country
- Converted each workbook to csv
- It was followed up by clean ups in Jupyter notebook:
  - Merges
  - Groupbys

```
In [ ]: import warnings
warnings.filterwarnings("ignore")
import pandas as pd
import scipy.stats as st
import numpy as np
import matplotlib.pyplot as plt

In [ ]: netflix_language = pd.read_csv("Resources/netflix_language.csv")
netflix_language

In [ ]: languagecount = netflix_language.groupby(by='Title')['Languages'].count()

In [ ]: language = languagecount.to_frame()
language.rename(columns={"Languages": "Number of Languages"}, inplace=True)
language

In [ ]: netflix_country = pd.read_csv("Resources/netflix_country.csv")

In [ ]: newnetflix = netflix_country[['Title', 'Country Availability', 'Hidden Gem Score', 'IMDb Score', 'Rotten Tomatoes Score', 'Metacritic Score']]
newnetflix

In [ ]: fullset = pd.merge(newnetflix, language, how='inner', on='Title')
fullset

In [ ]: avg_language_per_title = fullset.groupby('Country Availability')['Number of Languages'].mean()
avg_language_per_title

In [ ]: h_df = pd.read_csv('Resources/happiness_index_2021.csv')
h_df.head()
h_df = h_df[['Country name', 'Ladder score']]
h_df.rename(columns={'Country name': 'Country Availability'}, inplace=True)
```

|   | A                          | B            | C   | D                | E               | F                | G  |     |
|---|----------------------------|--------------|---|------------------|-----------------|------------------|--|-----|
| 1 | Title                      | Genre        | Tags  | Languages        | Series or Movie | Hidden Gem Score | Country Availability   | Row |
|   |                            |              | Social Issue Dramas, Dramas, Movies Based on Books, Period Pieces, Classic Movies, Swedish Movies | Scanian, Swedish |                 |                  | Lithuania, Poland, France, Italy, Spain, Greece, Czech Republic, Belgium, Portugal, Hungary, Slovakia, Netherlands, Germany, Romania, Switzerland, United Kingdom, Iceland | 1-2 |
| 2 | The Simple Minded Murderer | Drama        |   |                  | Movie           | 7.8              |  |     |
|   |                            |              |   |                  |                 |                  | Lithuania, Poland, France, Italy, Spain, Greece, Czech Republic, Belgium, Portugal, Hungary, Slovakia, Netherlands, Germany, Romania, Switzerland, United Kingdom, Iceland | < 3 |
| 3 | To Kill a Child            | Short, Drama | Dramas, Swedish Movies  | Spanish          | Movie           | 8.8              |  |     |
|   |                            |              | Dark  |                  |                 |                  |  |     |

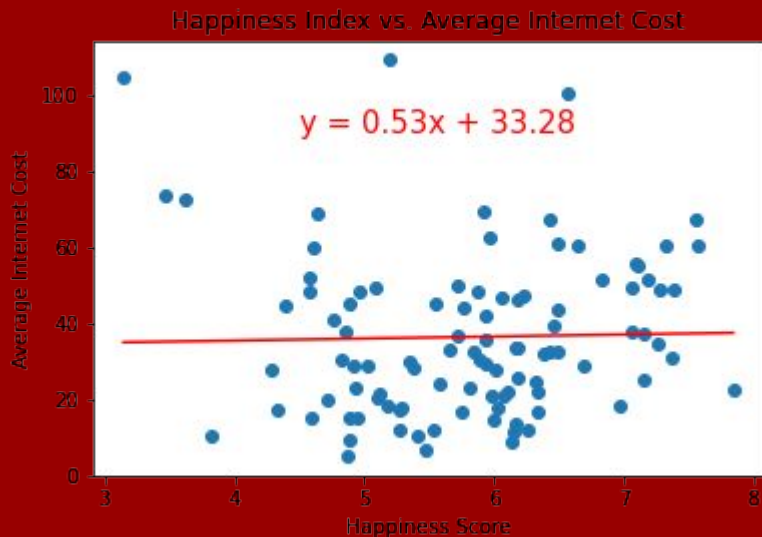
## Exploration & limitations of data:

- Netflix usage unavailable
- Nulls
- Titles last updated in March 2021
- Unique no. of countries incomplete
- Languages incomplete

## RESEARCH QUESTIONS:

- ❑ Does the average show rating correlate with happiness index?
- ❑ Does happiness index correlate with average internet price?
- ❑ Does Netflix offerings indicate diversity in a country? If so, how does diversity correlate with happiness index?
- ❑ Does the number of languages offered for a title impact the hidden gem score?
- ❑ Is there a relationship between languages for a title and country availability?
- ❑ Is there a correlation between countries who have less Netflix offerings and freedom to make life choices rating on the happiness index?
- ❑ Which genres have better critic reviews?

# HAPPINESS INDEX VS. COST OF INTERNET (BY COUNTRY)

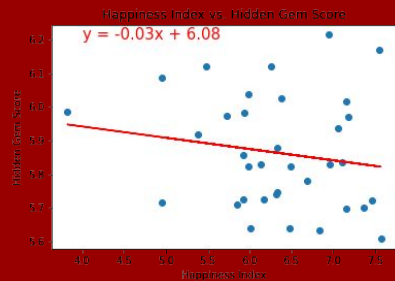


The r-value between the happiness index and average cost of internet is  $\sim .0247$ .

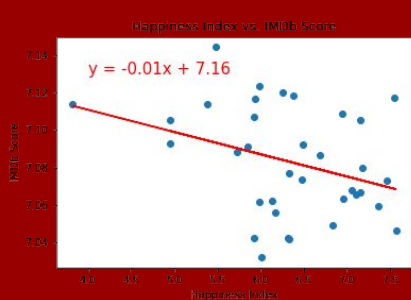
This suggests that there is no correlation between the two datasets.



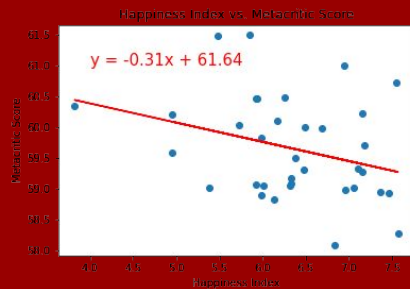
# CRITIC RATINGS VS. HAPPINESS INDEX



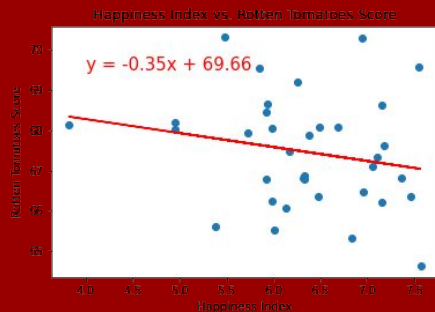
R-value =  $\sim -.1638$



R-value =  $\sim -.3380$



R-value =  $\sim -.3098$

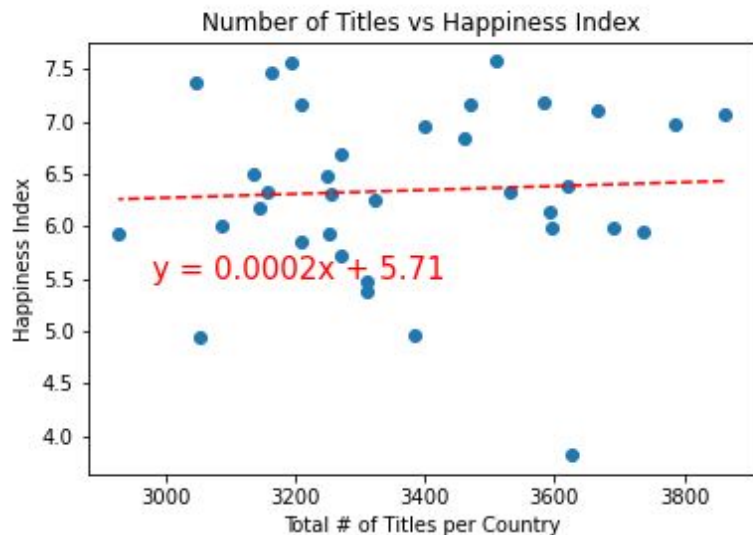


R-value =  $\sim -.2059$

All 4 critic score metrics have a slight negative correlation with the happiness index.

However, given the number of outliers for each score, there is no strong relation between happiness and critic scores.

# NUMBER OF TITLES PER COUNTRY & HAPPINESS INDEX

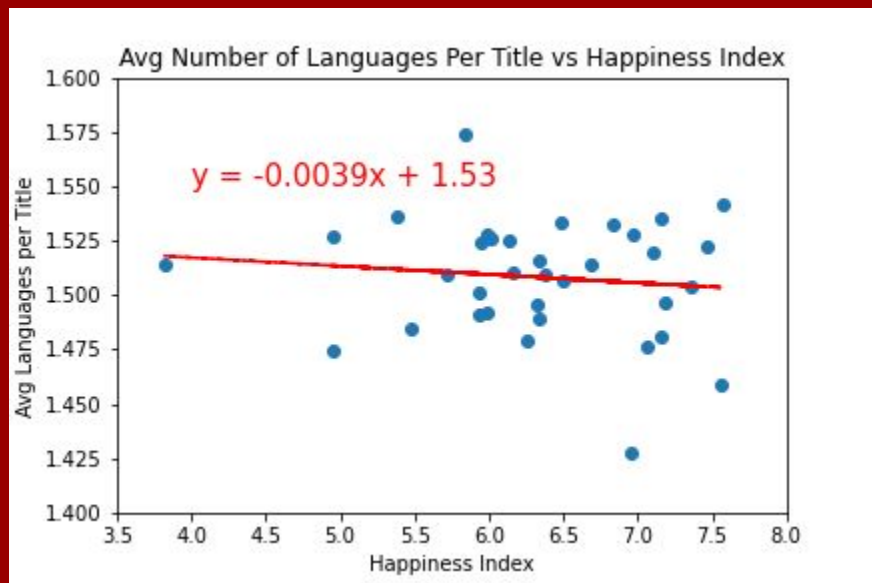


The r-value is 0.053, showing a negligible positive correlation between Happiness Index and Number of Titles available per country.

No correlation between Happiness Index and Number of Titles available per country



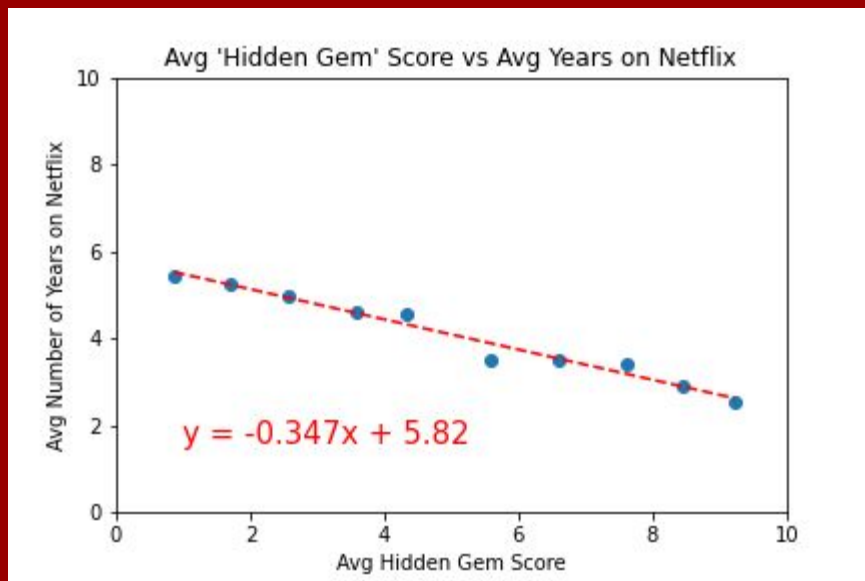
# AVG NUMBER OF LANGUAGES (AS AN INDICATION OF DIVERSITY IN A COUNTRY) & HAPPINESS INDEX



The r-value is -0.12, showing a negligible negative correlation between Happiness Index and Average Number of Languages available per Title

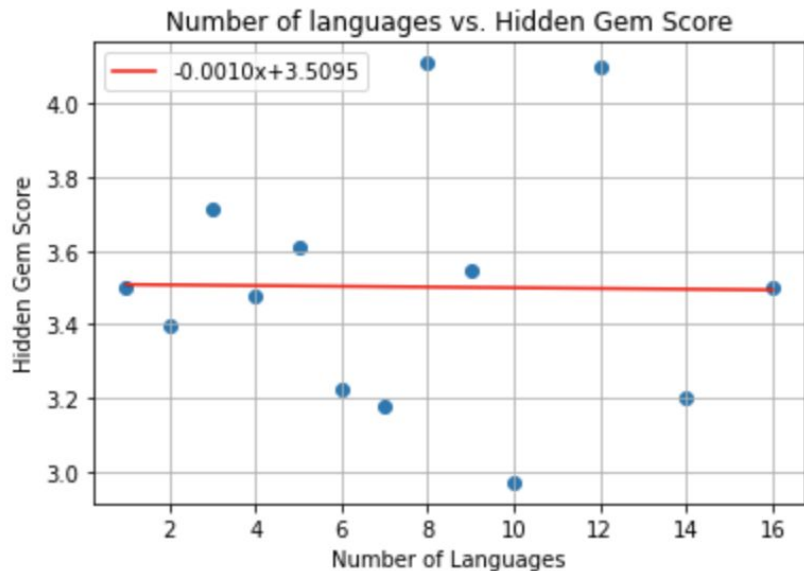
No correlation between Language offerings and Happiness Index

# 'HIDDEN GEM' SCORE & YEARS ON NETFLIX



The r-value is  $-0.98$ , showing a strong negative correlation between 'Hidden Gem' Score and years a title is available on Netflix.

# AVG. NUMBER OF LANGUAGES & 'HIDDEN GEM SCORE'



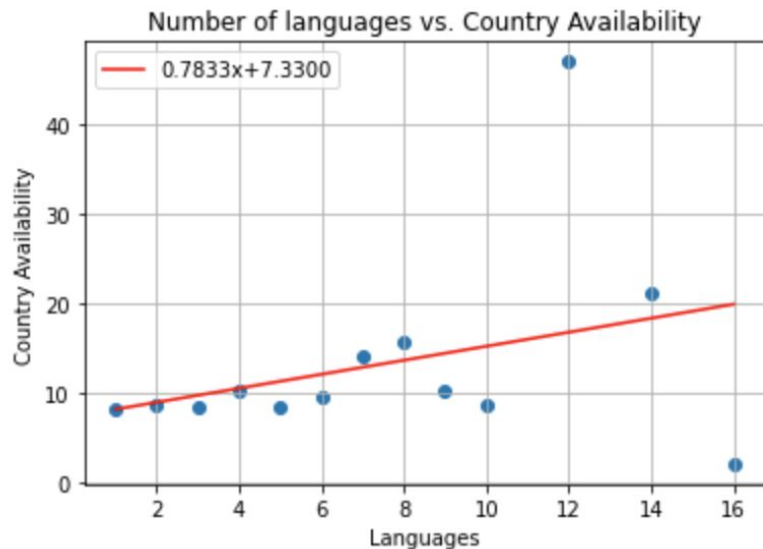
## Correlation and Regression

```
1
2
3 # use the pearson formula and the average calculated above to determin p-value.
4 # if r-value is close to 1, there is a high degree of correlation
5 # if r-value is close to 0, there is no correlation
6 # if r-value is close to -1 suggests there is negative correlation
7
8 correlation_coeff = round(st.pearsonr(group2['Number of Languages'], group2['Hidden Gem Score'])[0], 4)
9 correlation_coeff
```

-0.0134

The r-value is -0.0134, indicating a negative correlation between languages and Hidden Gem Score.

# AVG. NUMBER OF LANGUAGES & 'COUNTRY AVAILABILITY'



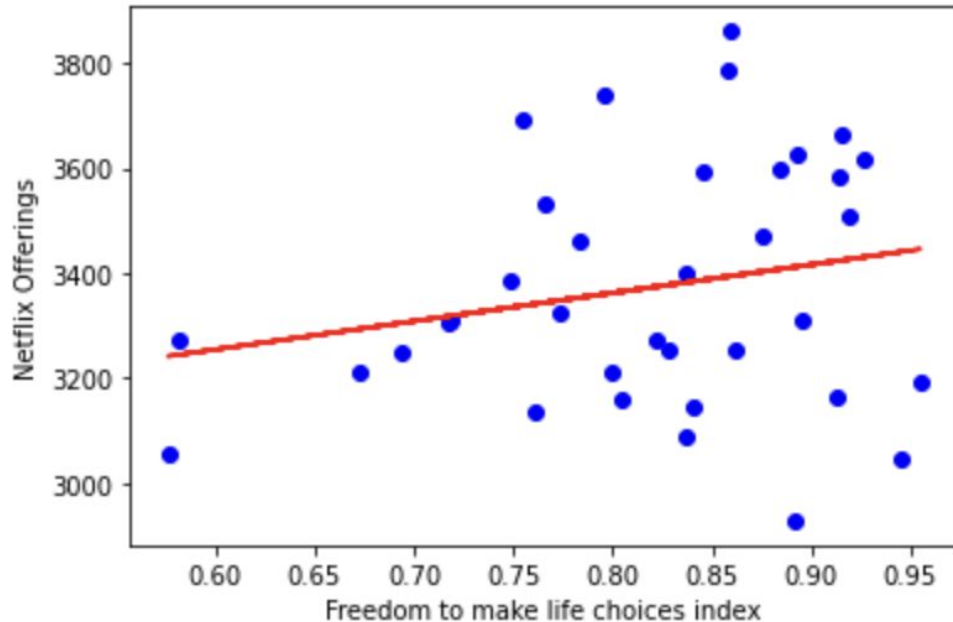
## Correlation and Regression

```
1
2 # use the pearson formula and the average calculated above to determin r-value.
3 # if r-value is close to 1, there is a high degree of correlation
4 # if r-value is close to 0, there is no correlation
5 # if r-value is close to -1 suggests there is negative correlation
6
7 correlation_coeff = round(st.pearsonr(newset_2['Languages'], newset_2['Country Availability'])[0], 4)
8 correlation_coeff
```

0.326

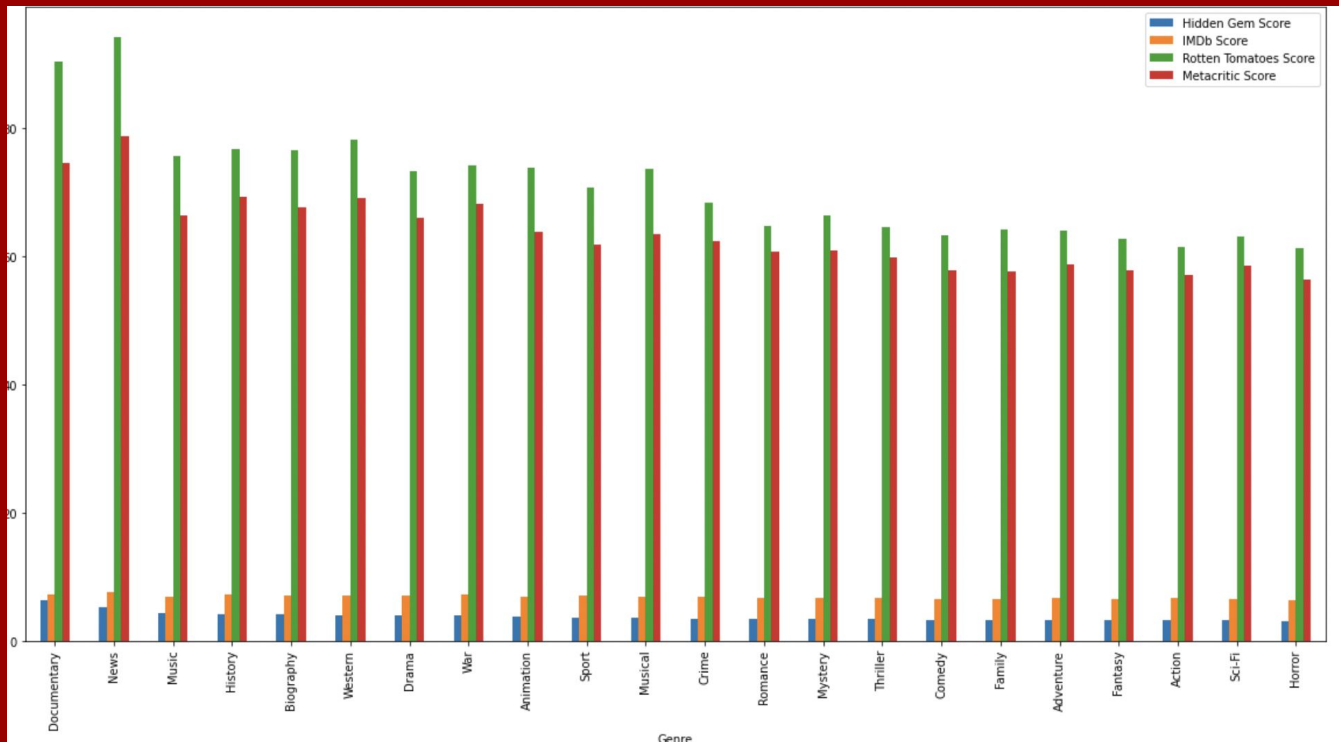
The r-value is 0.326, suggesting little to no correlation between languages and Country availability.

# NETFLIX AVAILABILITY AND HAPPINESS INDEX (FREEDOM TO MAKE LIFE CHOICES)



- The r-value is 0.22, showing a slight positive correlation between the Freedom to make life choices factor of the Happiness Index and Netflix Offerings

# WHICH GENRES HAVE BETTER CRITIC REVIEWS?



| Total Titles |      |
|--------------|------|
| Genre        |      |
| Drama        | 3428 |
| Comedy       | 2319 |
| Action       | 1829 |
| Thriller     | 1812 |
| Crime        | 1229 |
| Romance      | 1204 |
| Adventure    | 1146 |
| Fantasy      | 968  |
| Animation    | 831  |
| Sci-Fi       | 790  |
| Mystery      | 776  |
| Family       | 647  |
| Horror       | 613  |
| Biography    | 467  |
| Documentary  | 391  |
| History      | 320  |
| War          | 232  |
| Music        | 212  |
| Sport        | 204  |
| Musical      | 127  |
| Western      | 75   |

# FINDINGS AND IMPLICATIONS

- ❑ Netflix and its offerings have little to no correlation with a country's happiness index
- ❑ A correlation was not found between genre and happiness index, but comparison can be drawn between genre and critic reviews.
- ❑ The longer a title stays on available on Netflix the lower it's rated



# POSTMORTEM

- ❑ Explore datasets before settling on one
- ❑ Aggregated datasets are not as flexible to use
- ❑ Frame questions with actionable decision in mind
- ❑ If given more time a possible research topic would be popularity of titles by country not country availability.

QUESTIONS?