

# NETFLIX

**Brandilyn Hall , Kwadwo Asante , Lyndah Mupfunya , Yvonne Martinez**

▶ PLAY

# Inspiration

Netflix is a pastime for many people in the world and even more so during COVID.

We were inspired by Netflix due to the impact Netflix has around the world and personally in our lives as well.

We also gathered inspiration from other prior analysis done on Netflix as well:

- [Kaggle: Autoviz on Netflix Dataset](#)
- [Medium: Data Analysis of Netflix Movies & IMDB rating using Python](#)
- [Datacamp: Netflix Movie Data](#)



# Agenda

Hypotheses

Data

Analysis

Limitations &  
Future  
Analysis

Questions



Hypotheses

Data

Analysis

Limitations  
& Future  
Analysis

Questions

# Hypotheses

We propose the following hypotheses to be accepted or rejected by our data analysis.

1

Null Hypothesis: Genres in each country do not differ in duration.

Alternative Hypothesis: Genres in each country do differ in duration.

2

Null Hypothesis: Countries do not differ in movie lengths.

Alternative Hypothesis: Countries differ in movie lengths.

## ⌚ TOTAL WATCHTIME

~6,124 hrs

## ⌚ TOP GENRES

Drama, Comedies, Documentaries,  
Action & Adventure, Children & Family,  
Stand-Up Comedy, Horror, Thriller



Hypotheses

## Data

Data

Analysis

Limitations  
& Future  
Analysis

Questions

kaggle

NETFLIX

Movies and TV Shows

<https://www.kaggle.com/shivamb/netflix-shows>



Hypotheses

Data

Analysis

Limitations  
& Future  
Analysis

Questions

# Data Cleaning

## Movie & TV Show Data

- Removed TV shows and selected applicable columns
- Renamed columns for easier readability

## Null Values

- Dropped from all rows and columns

## Genre

- Kept the top 7 genres

## Countries

- Cleaned up rows that had multiple countries listed
- Kept the top 3 countries (USA, UK, India)

## Duration

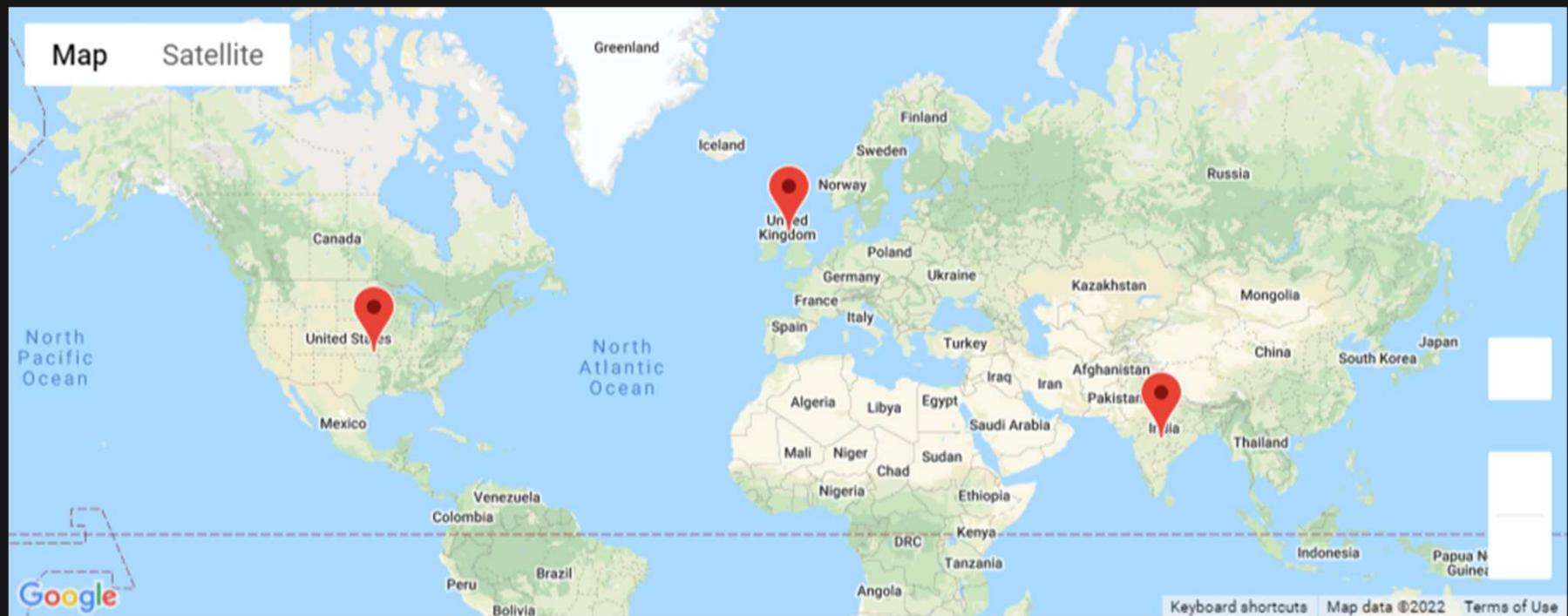
- Defined "movie" as 41+ mins
- Dropped "min" after #

## Ratings

- Consolidated ratings into 6 different ratings



# Top Countries (by number of releases)



Hypotheses

Data

Analysis

Limitations  
& Future  
Analysis

Questions

# Analysis

We analyzed our data to answer the following questions:

**Which countries  
have contributed  
the most movies  
in recent years?**

**Which country  
has the longest  
movie duration?**



Hypotheses

Data

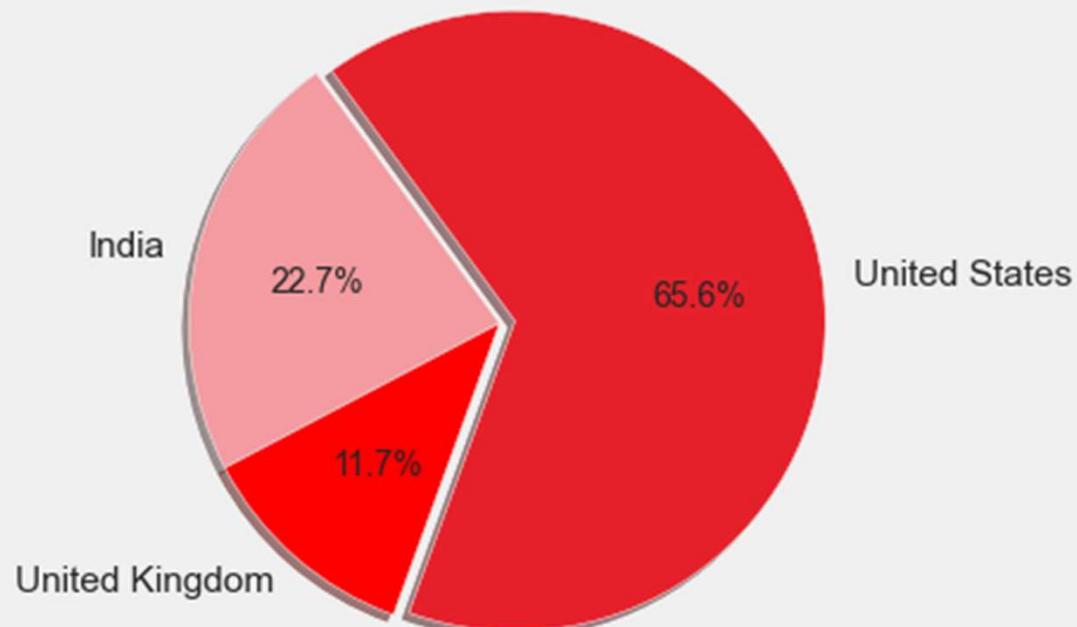
Analysis

Limitations  
& Future  
Analysis

Questions

## Number of Movies by Country

Percentage of Movies by Country



Hypotheses

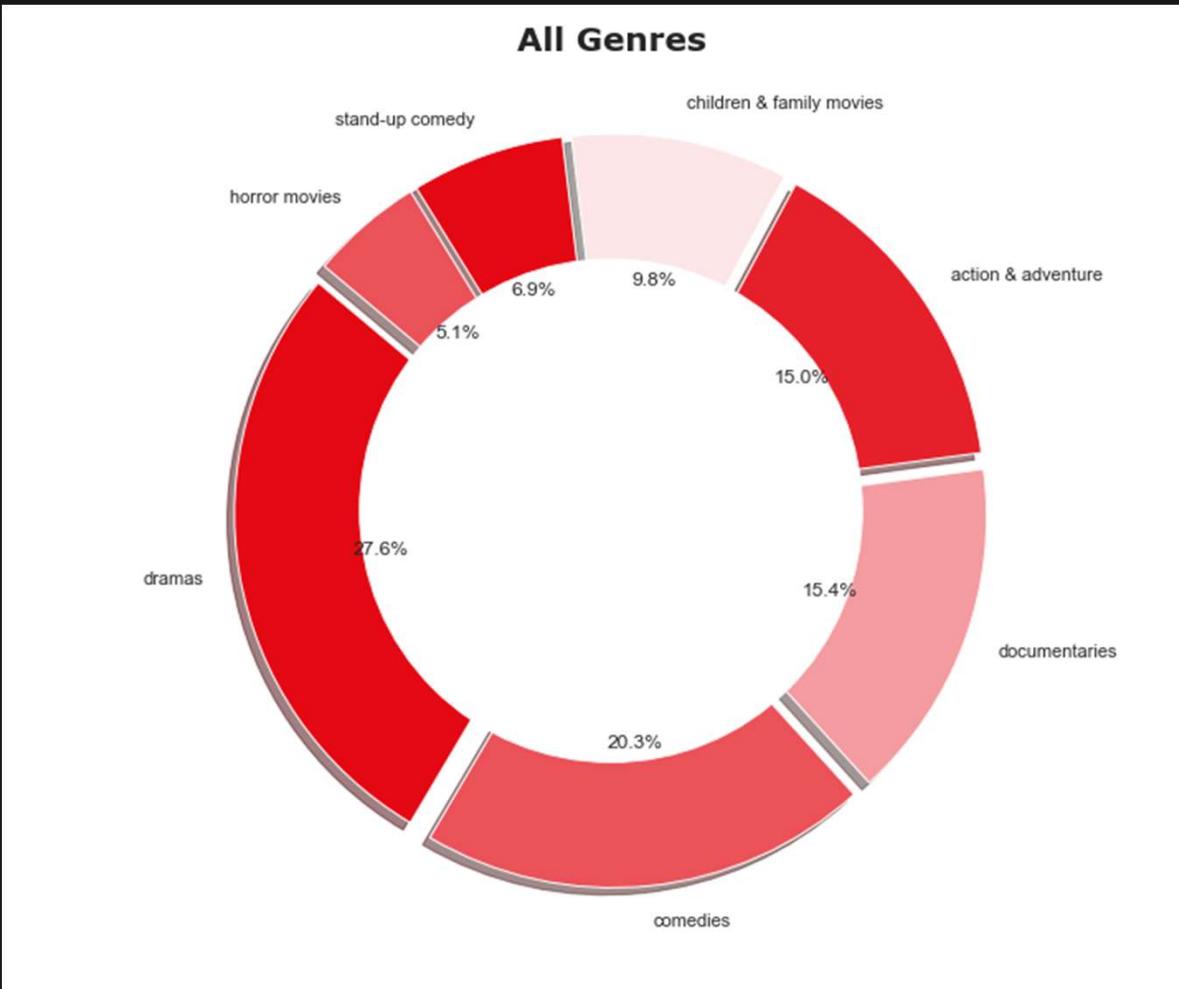
Data

Analysis

Limitations  
& Future  
Analysis

Questions

# Percentage of Movies by Genre



Hypotheses

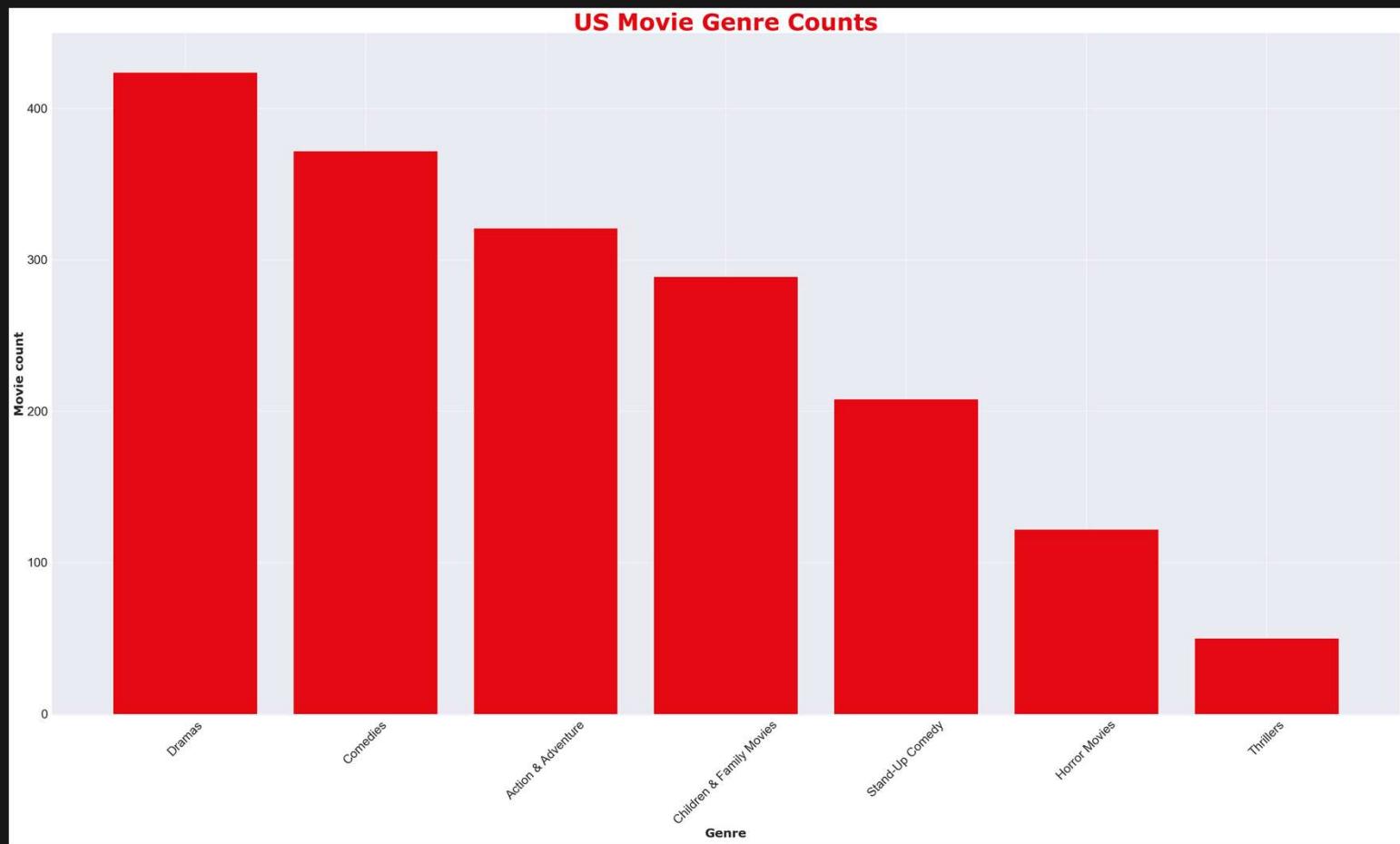
Data

Analysis

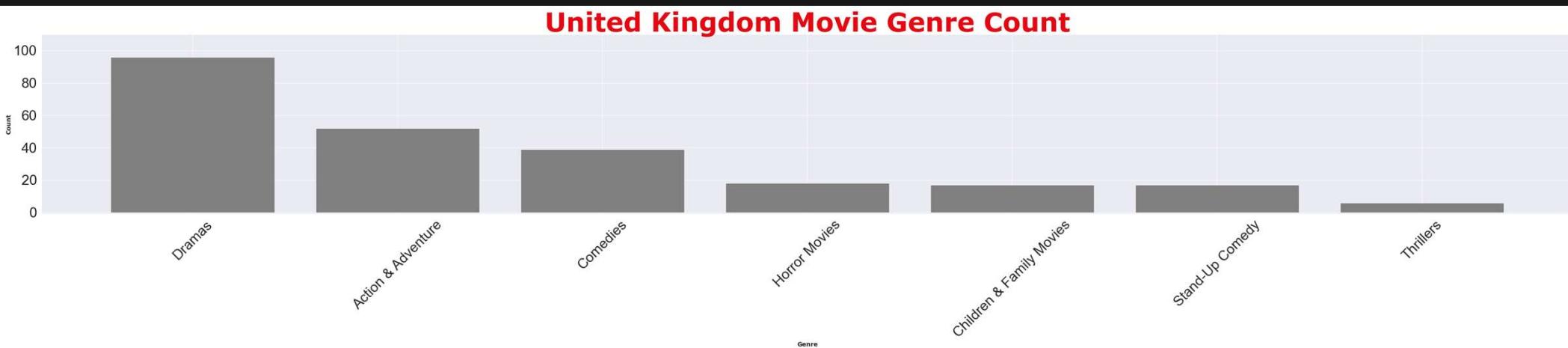
Limitations  
& Future  
Analysis

Questions

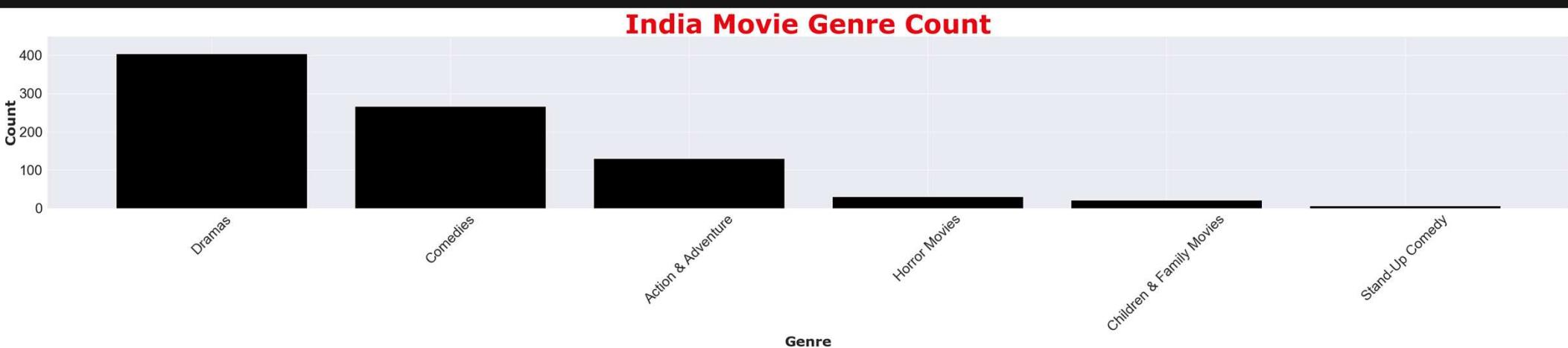
# US Movies by Genre



# Analysis - United Kingdom Movies by Genre



# Analysis - India Movies by Genre



Hypotheses

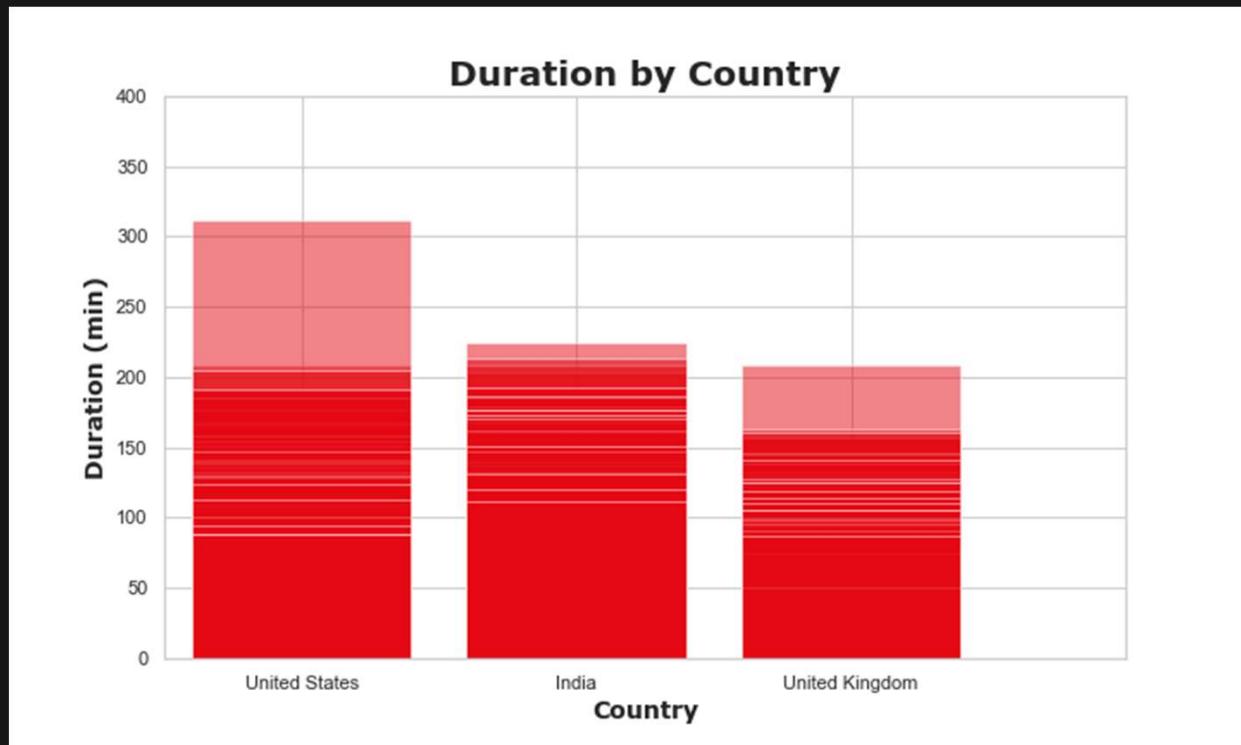
Data

Analysis

Limitations  
& Future  
Analysis

Questions

# Duration by Country



country	mean	max	min	median	std	var
India	126.493151	224.0	41.0	127.0	25.205599	635.322239
United Kingdom	96.611765	208.0	41.0	97.0	23.877190	570.120215
United States	94.167448	312.0	41.0	94.0	21.472312	461.060169



Hypotheses

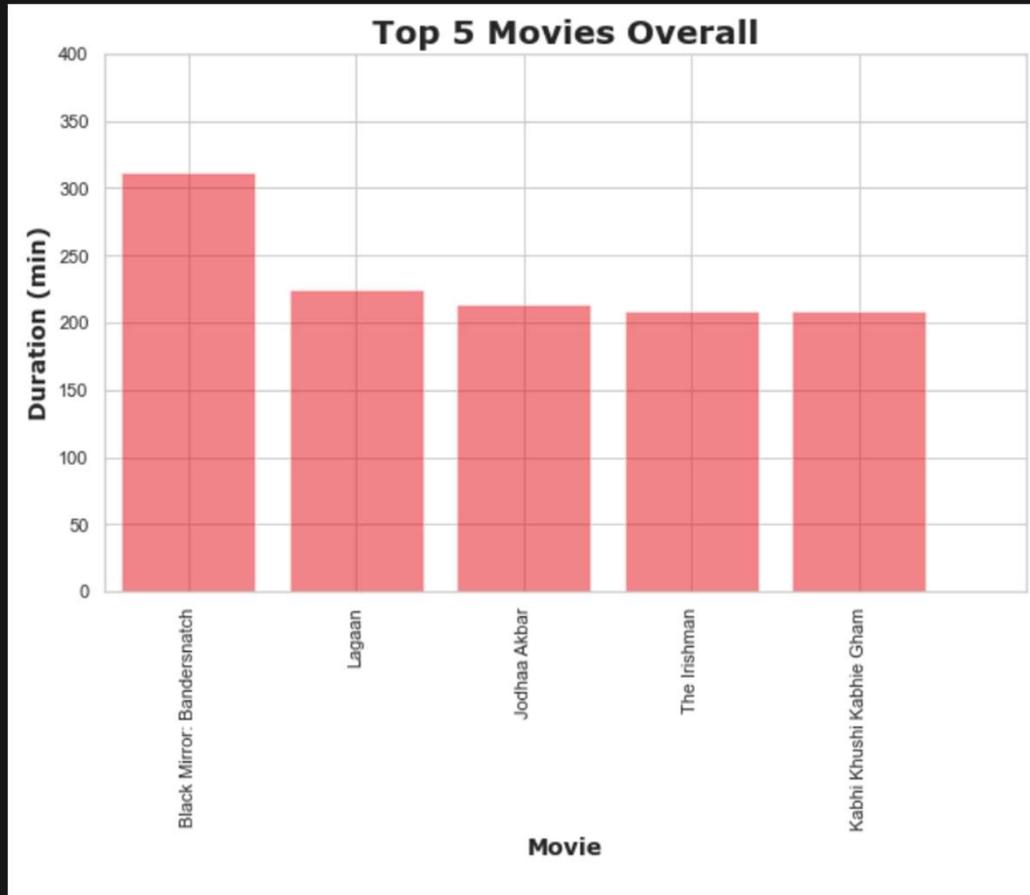
Data

Analysis

Limitations  
& Future  
Analysis

Questions

# Top Movies by Duration



## Top 5 US:

1. Black Mirror : Bandersnatch
2. The Irishman
3. The Gospel of Luke
4. Wyatt Earp
5. The Unbreakable Kimmy Schmidt – Kimmy vs. the Reverend

## Top 5 UK:

1. No Direction Home: Bob Dylan
2. American Honey
3. The Gospel of John
4. Les Miserables
5. Dhan Dhana Dhan Goal

## Top 5 India:

1. Lagaan
2. Jodhaa Akbar
3. Kabhi Kushi Kabhie Gham
4. What's Your Raashee?
5. Kabhi Alvida Naa Kehna



Hypotheses

Data

Analysis

Limitations  
& Future  
Analysis

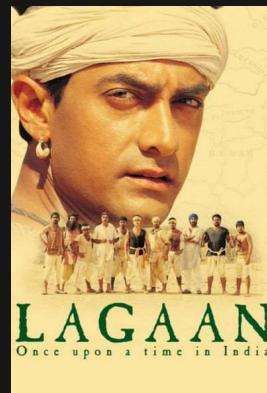
Questions

# Top 5 In the World (By duration)

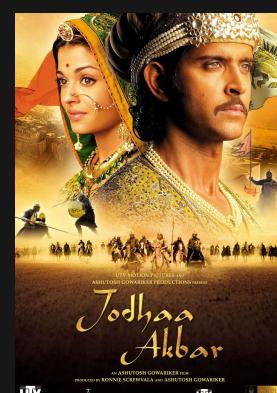
#1 – Black  
Mirror  
Bandersnatch



#2 – Lagaan



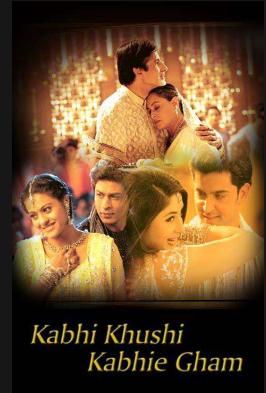
#3 – Jodhaa  
Akbar



#4- The  
Irishman



#5 – Kabhi  
Khushi  
Kabhie Gham



Hypotheses

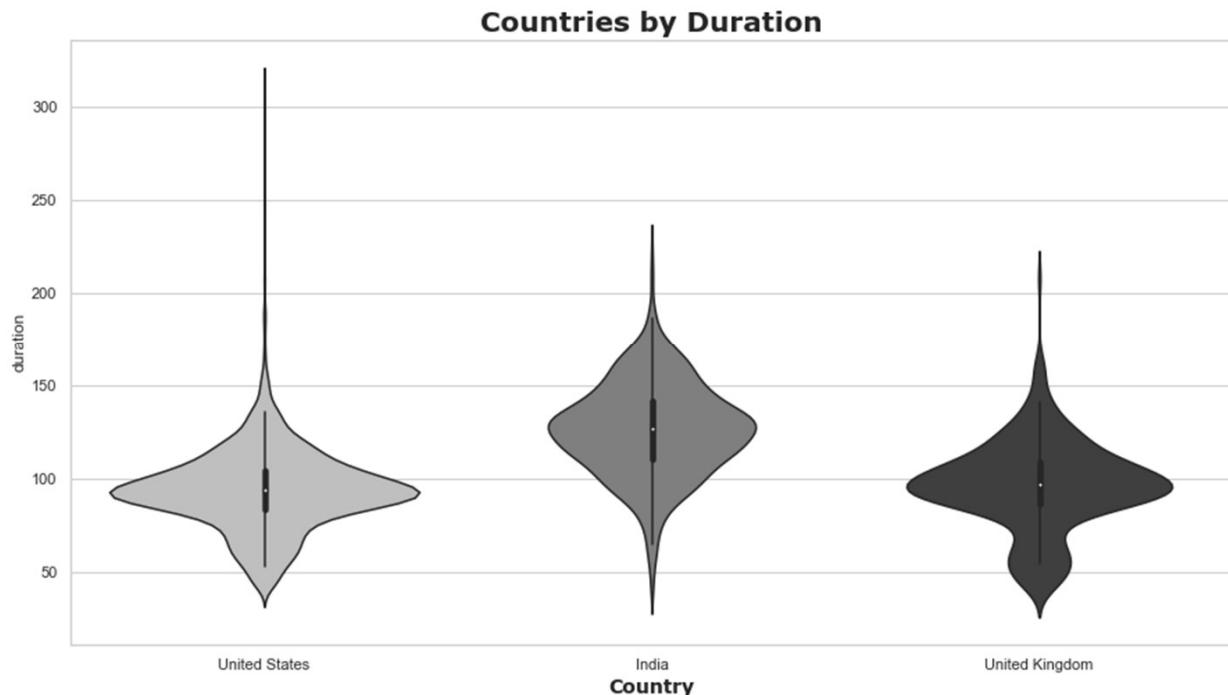
Data

Analysis

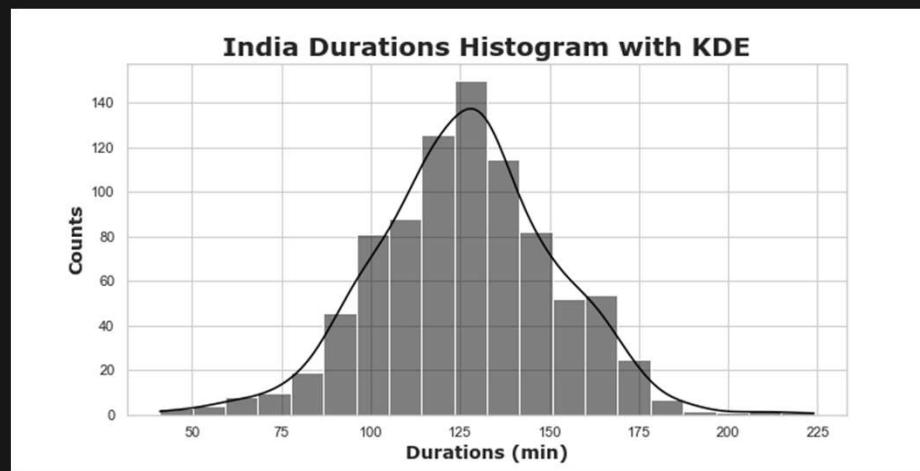
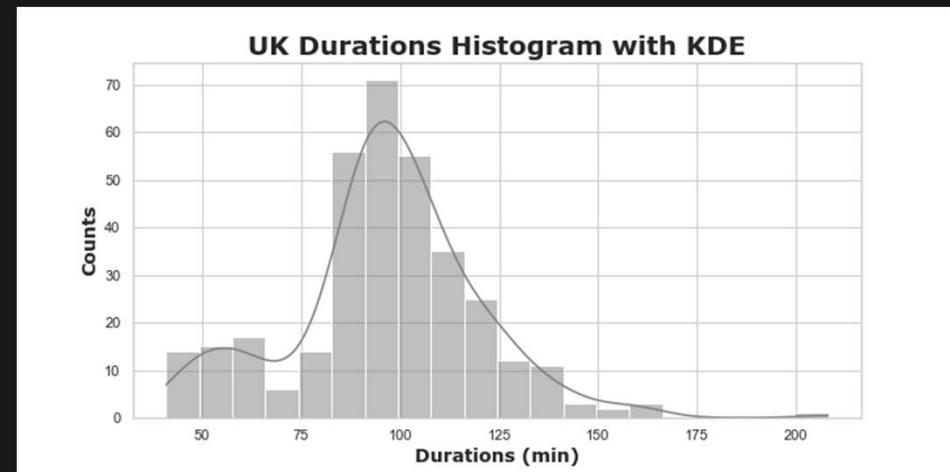
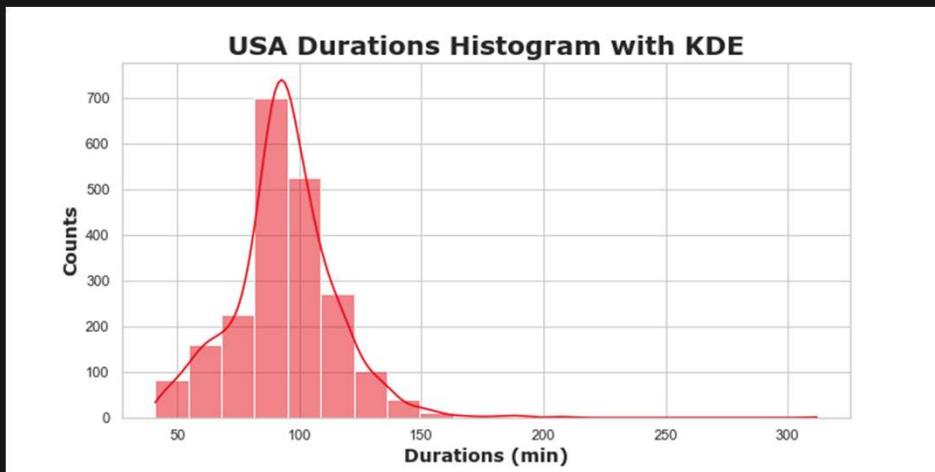
Limitations  
& Future  
Analysis

Questions

# Duration by Country



# Analysis - Number of Movies by Duration



Hypotheses

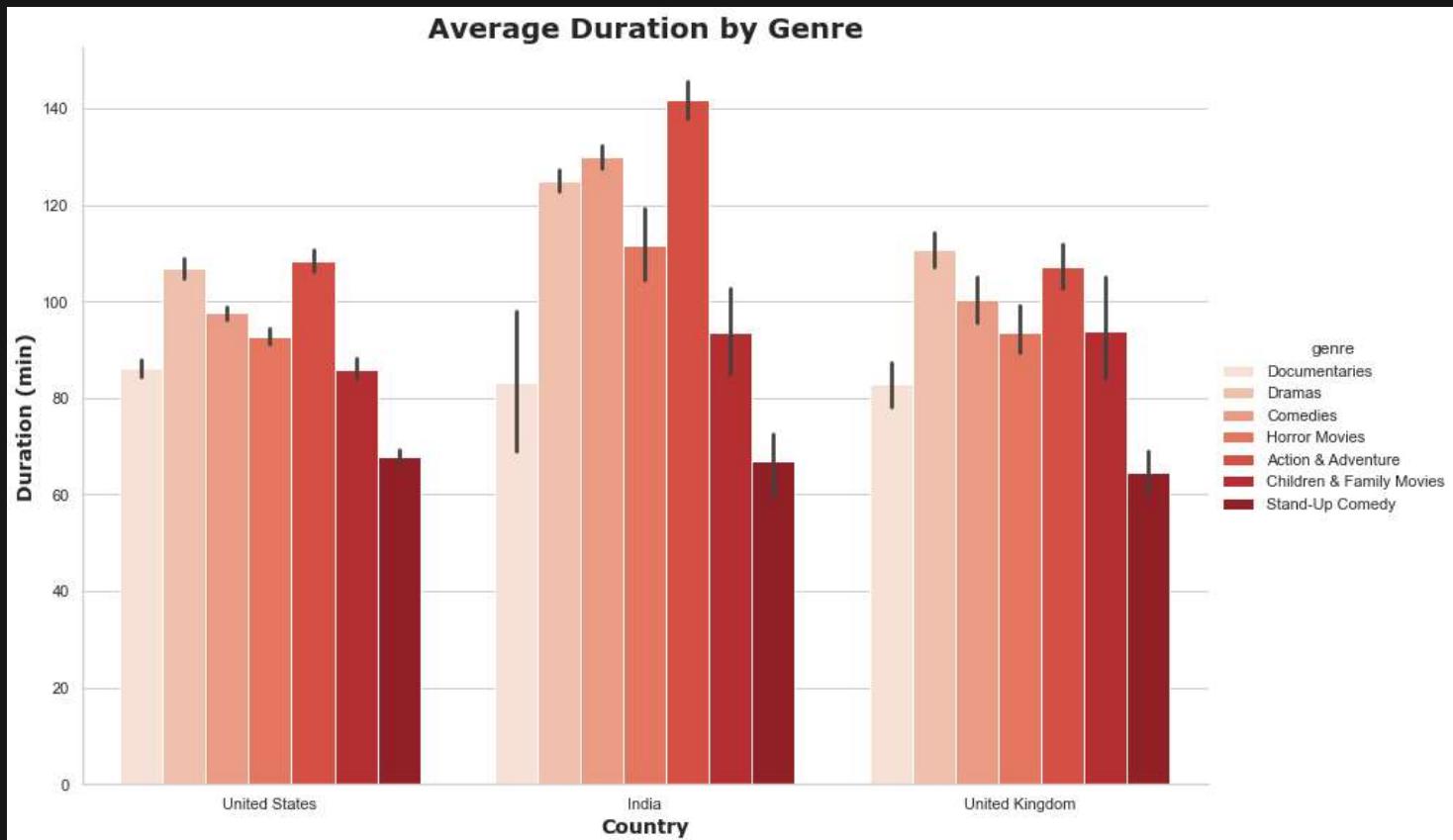
Data

Analysis

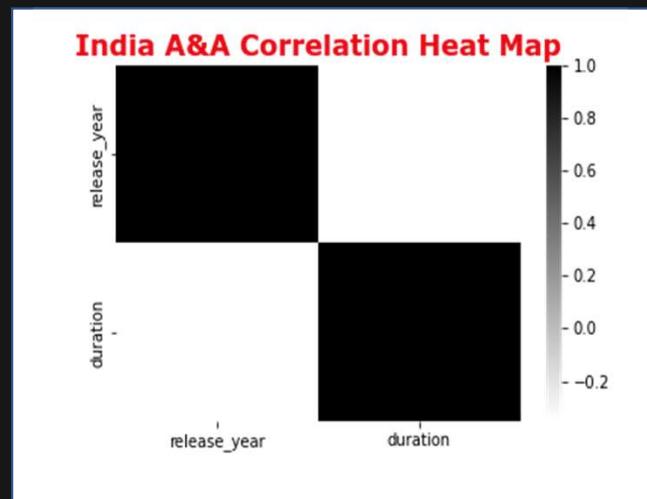
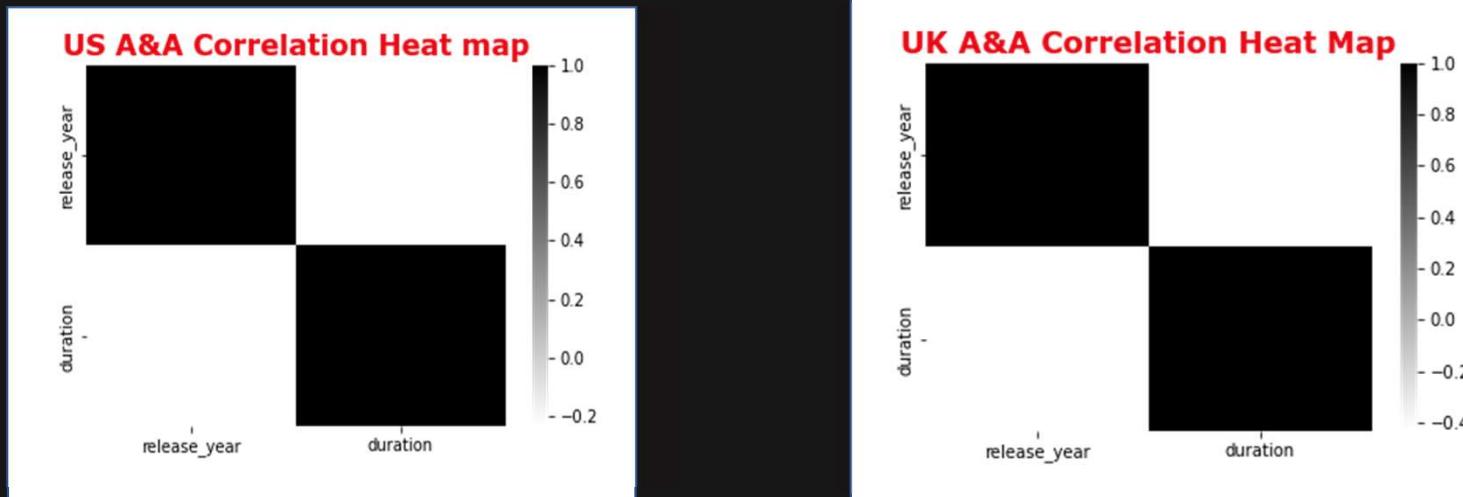
Limitations & Future Analysis

Questions

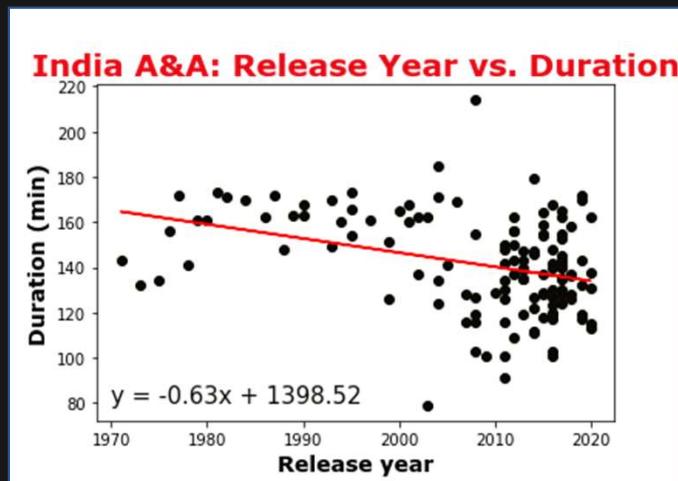
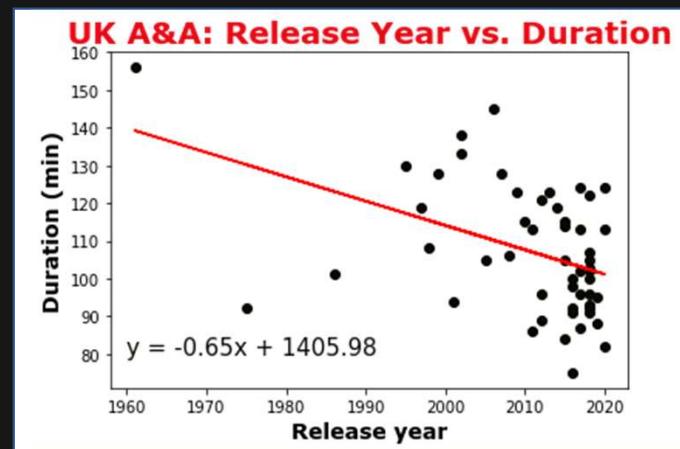
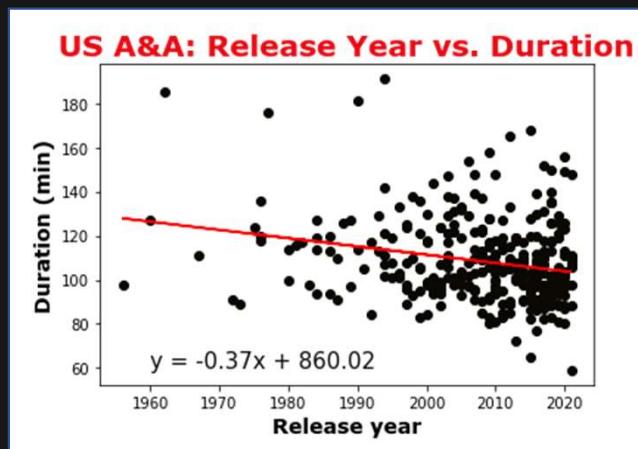
# Duration by Genre



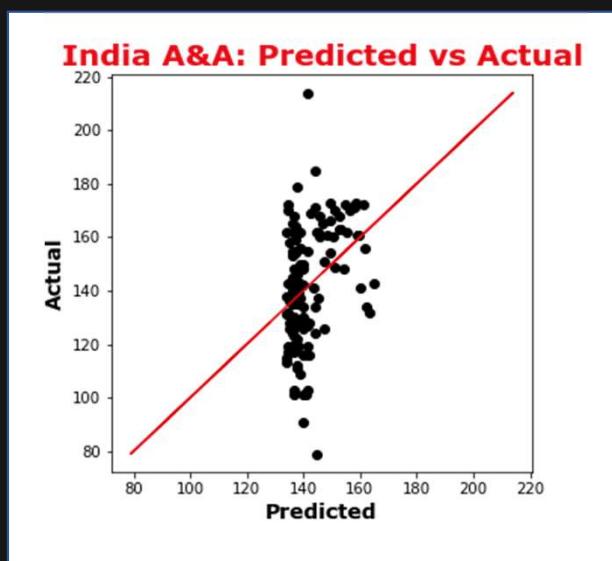
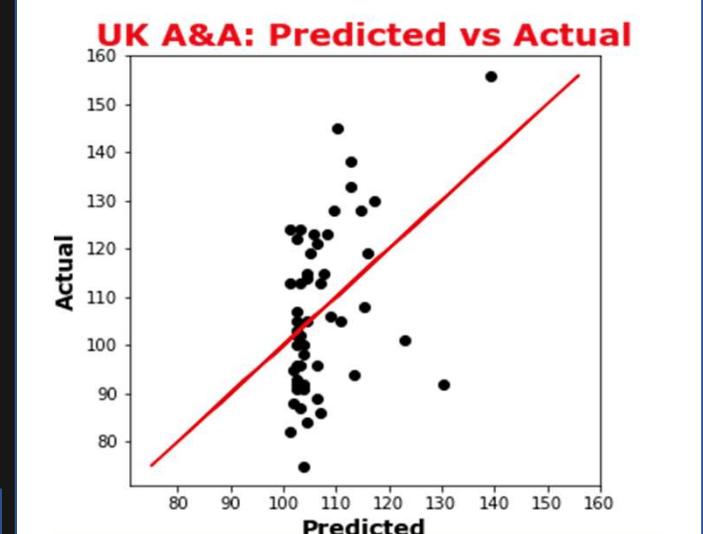
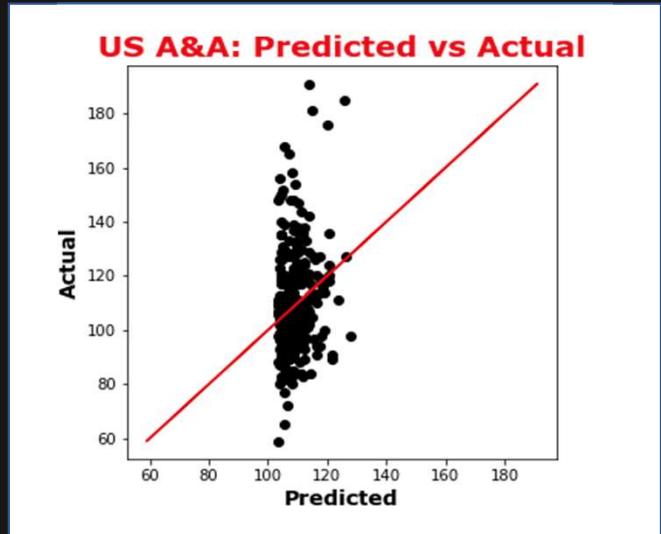
# Analysis - Action & Adventure



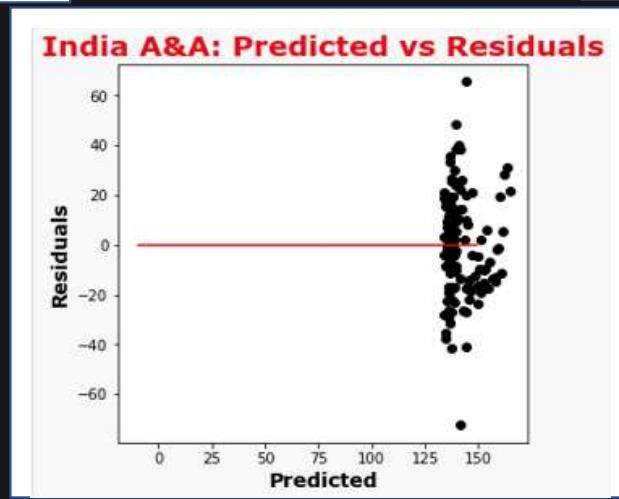
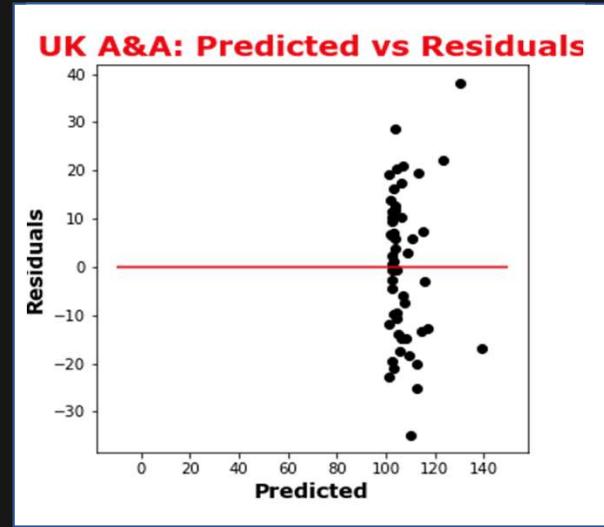
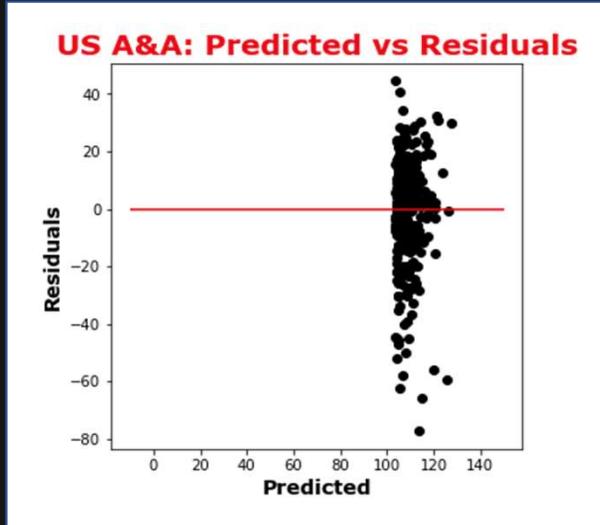
# Analysis - Action & Adventure



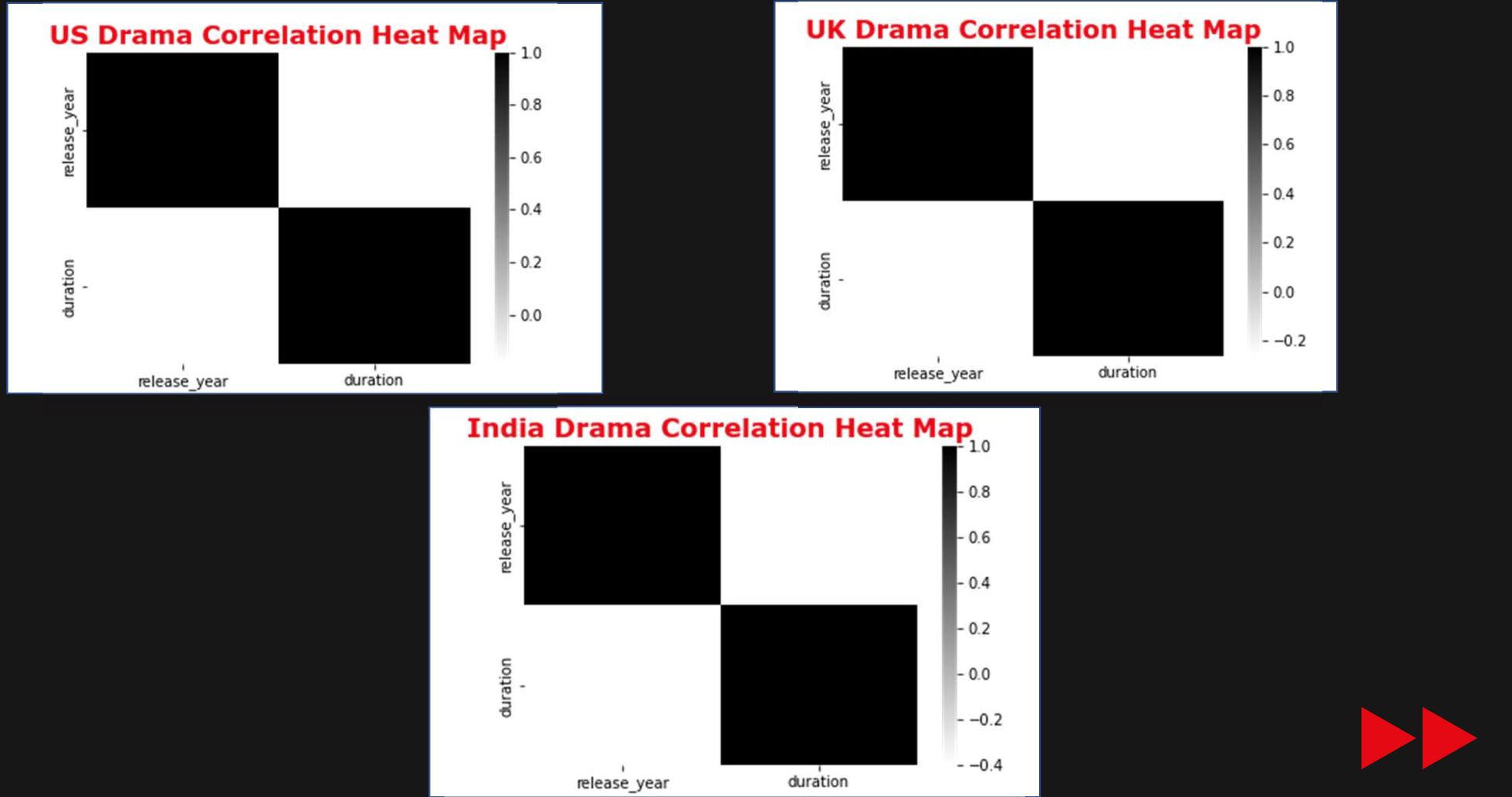
# Analysis - Action & Adventure



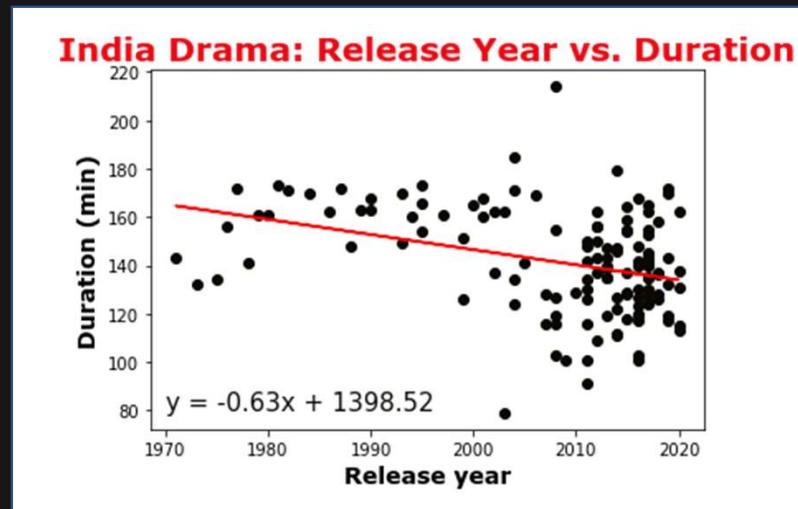
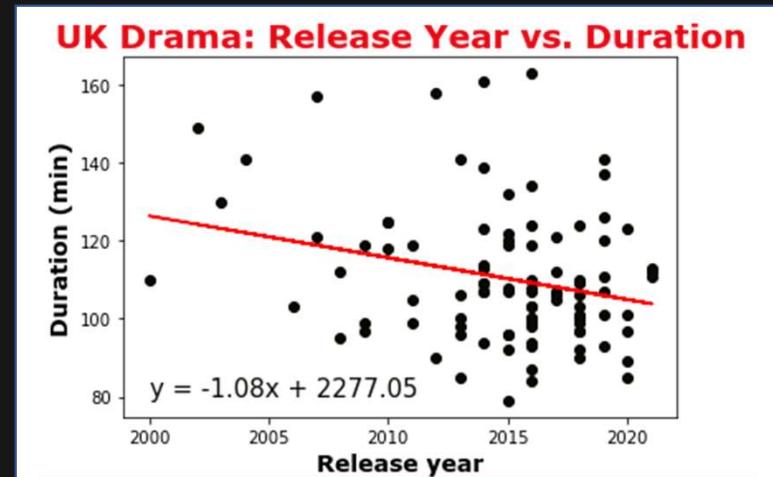
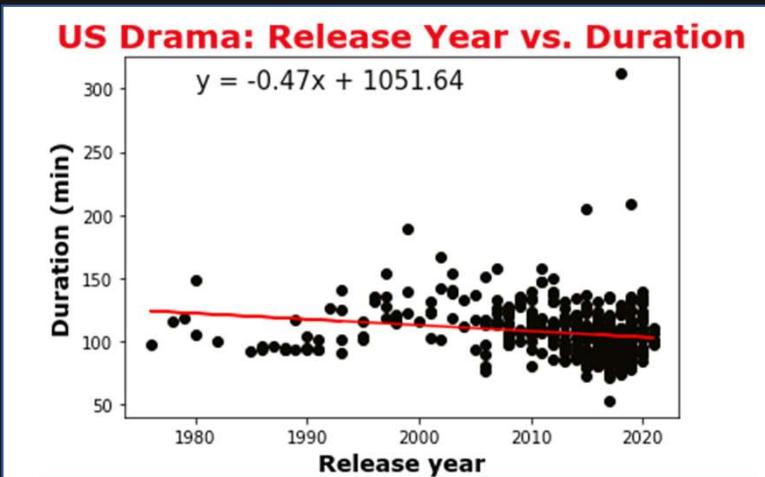
# Analysis - Action & Adventure



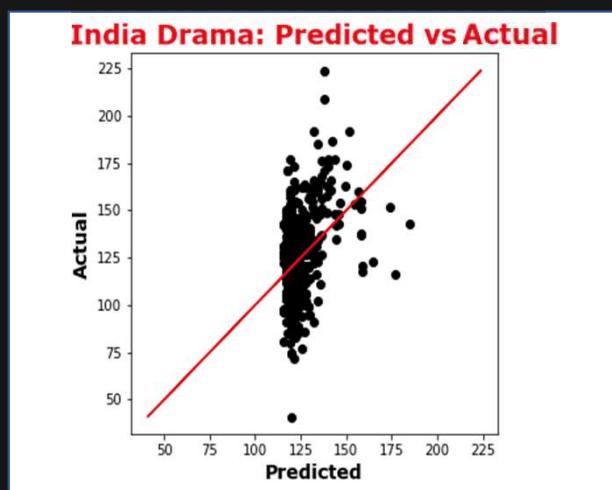
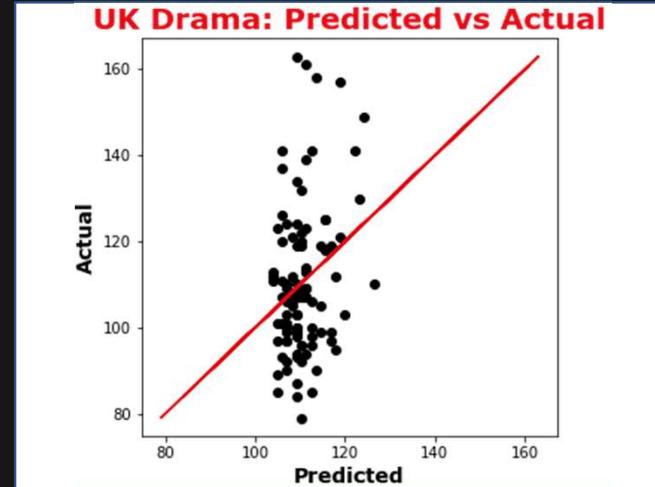
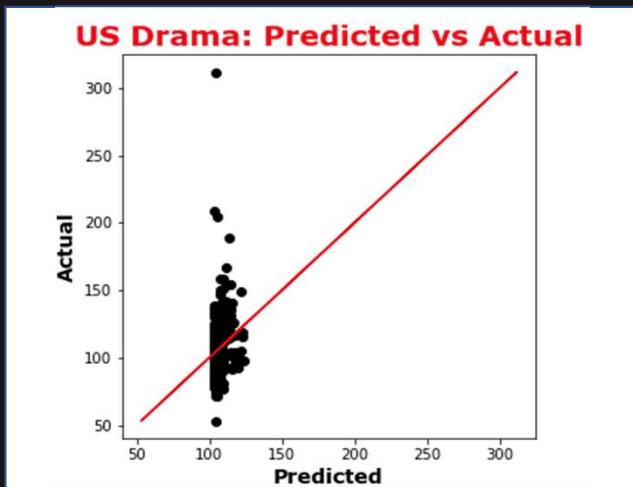
# Analysis - Drama



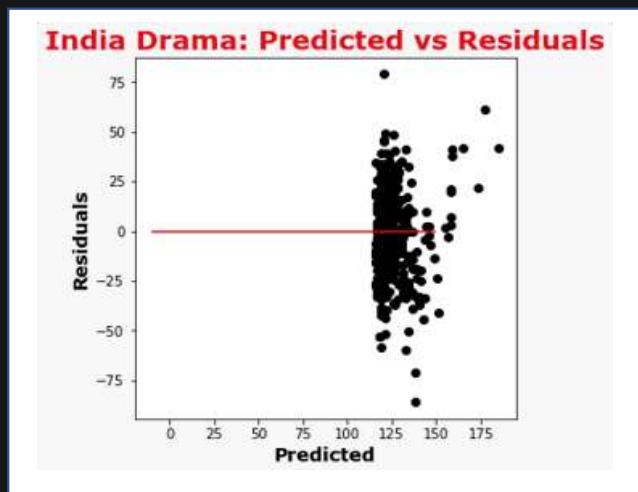
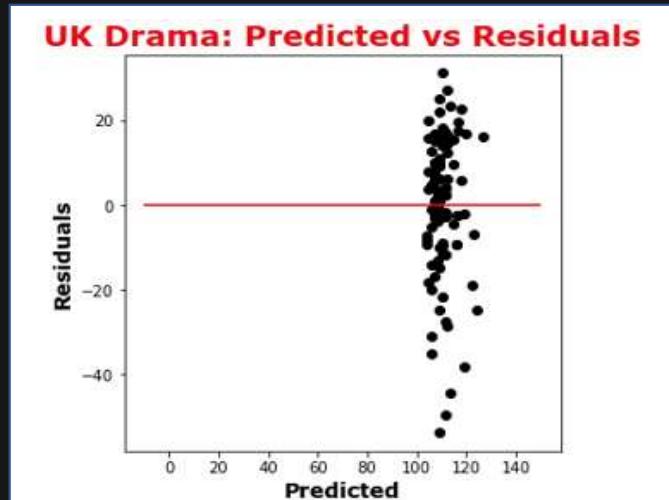
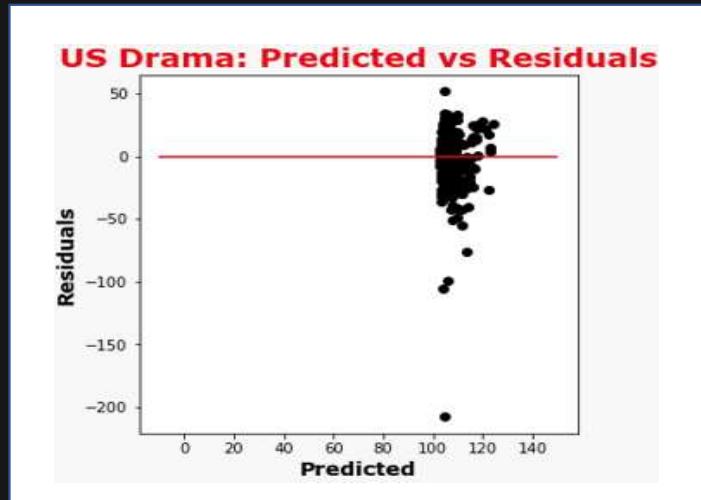
# Analysis – Drama



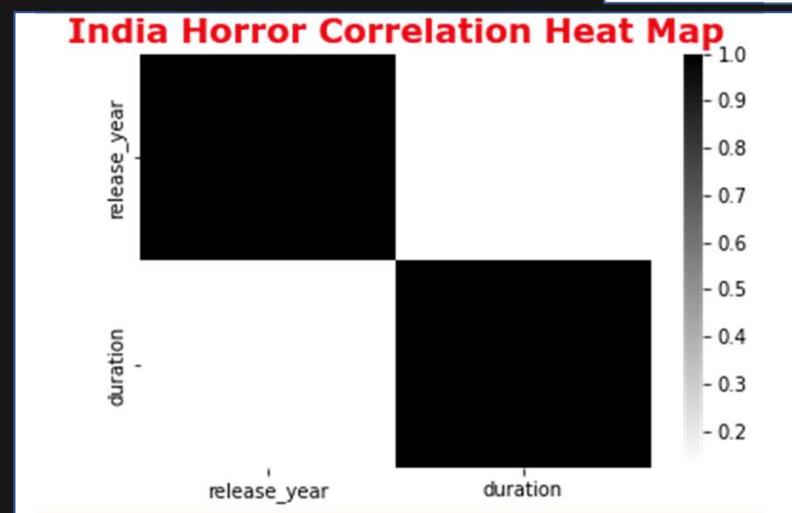
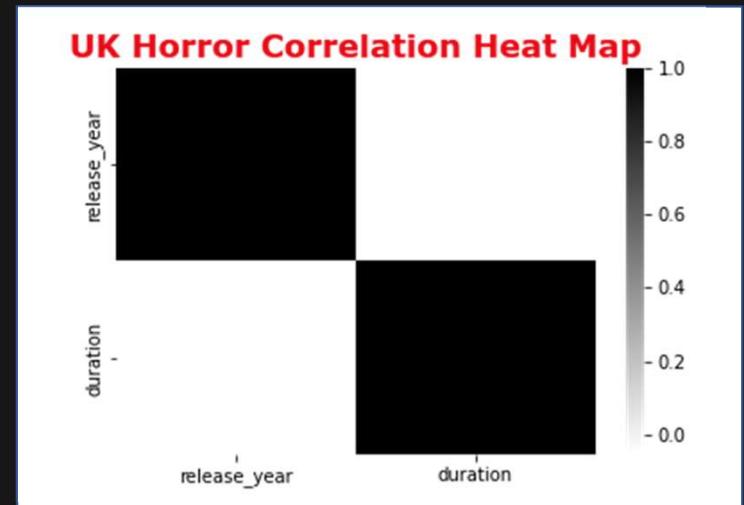
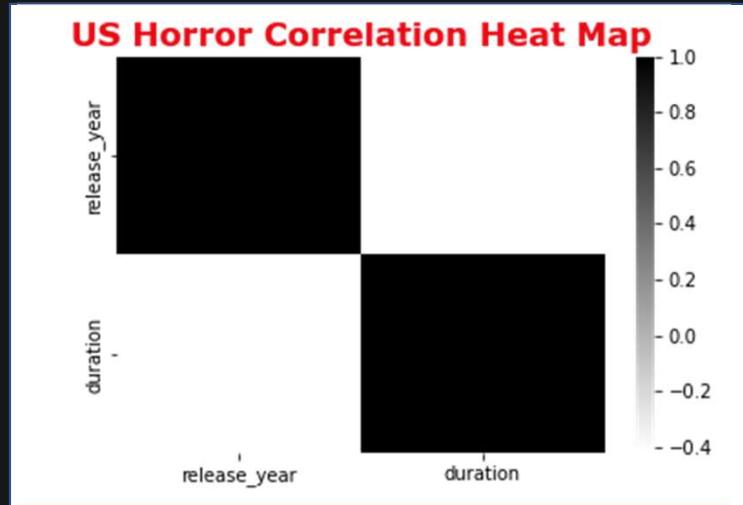
# Analysis - Drama



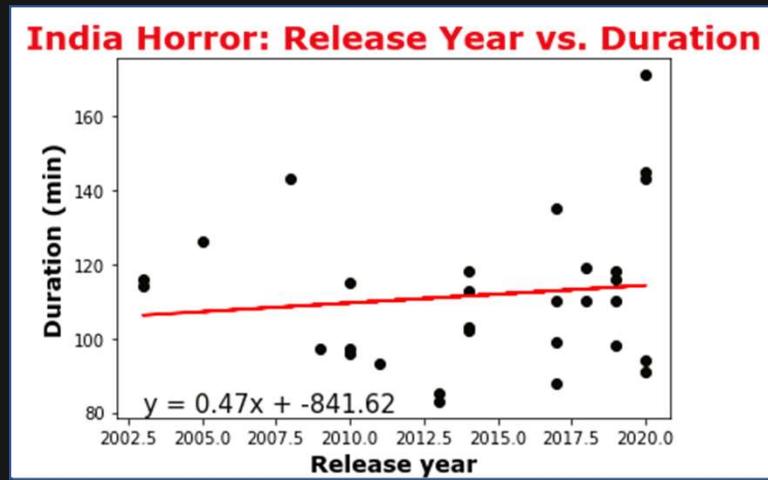
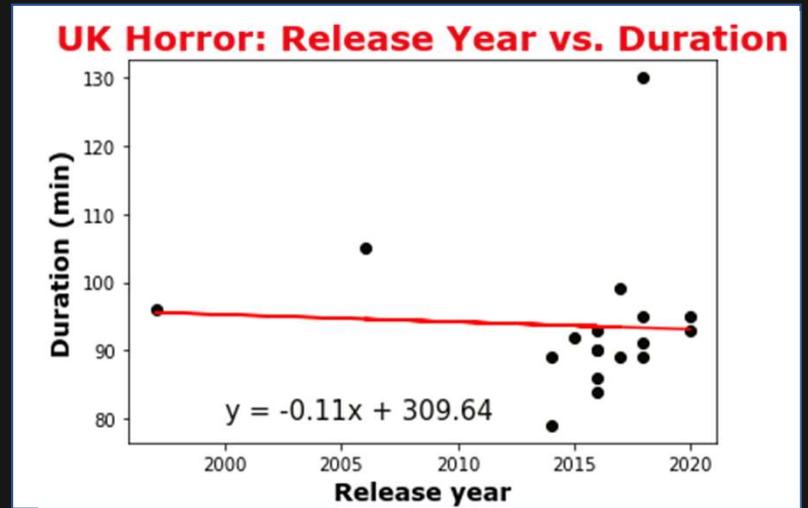
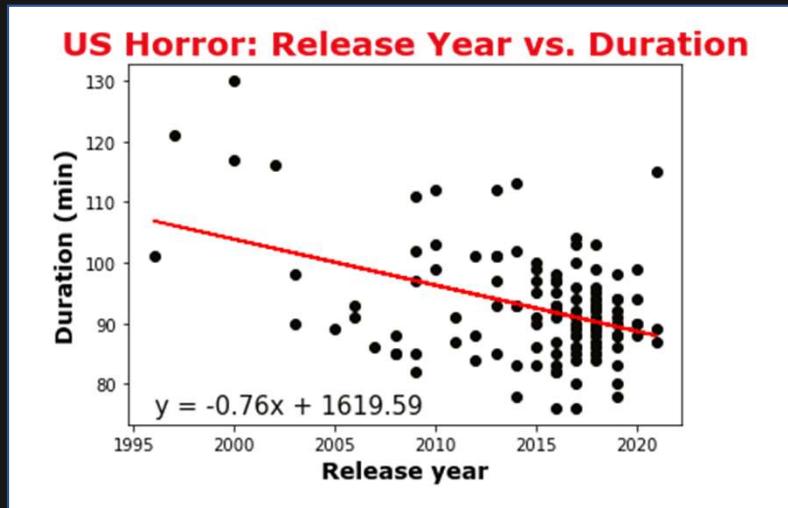
# Analysis - Drama



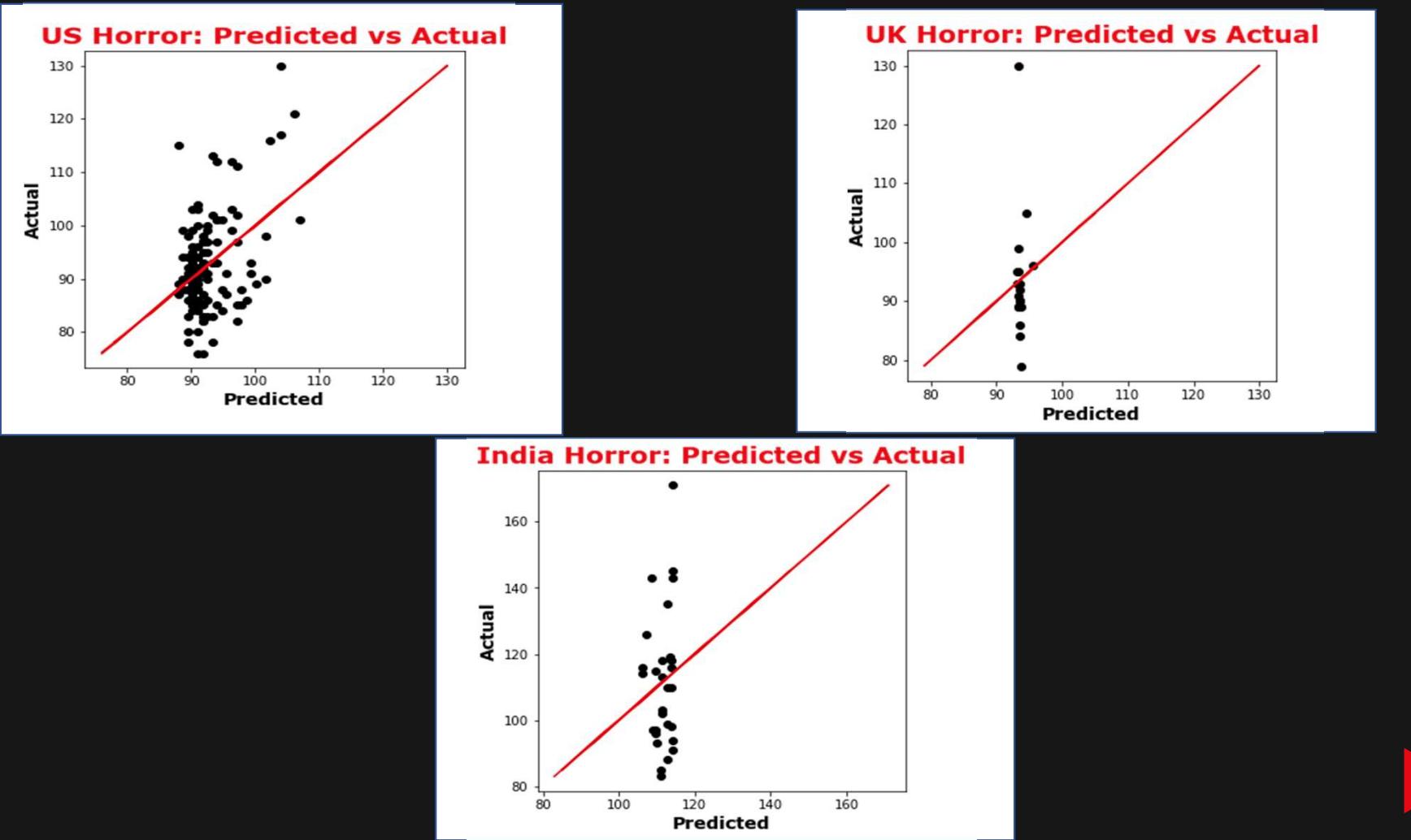
# Analysis - Horror



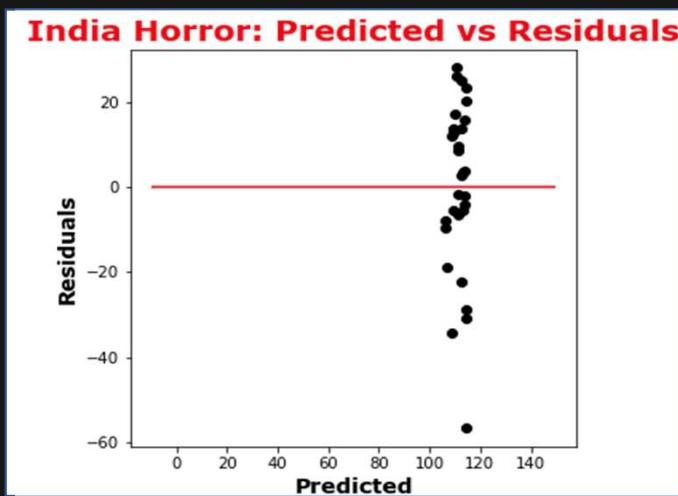
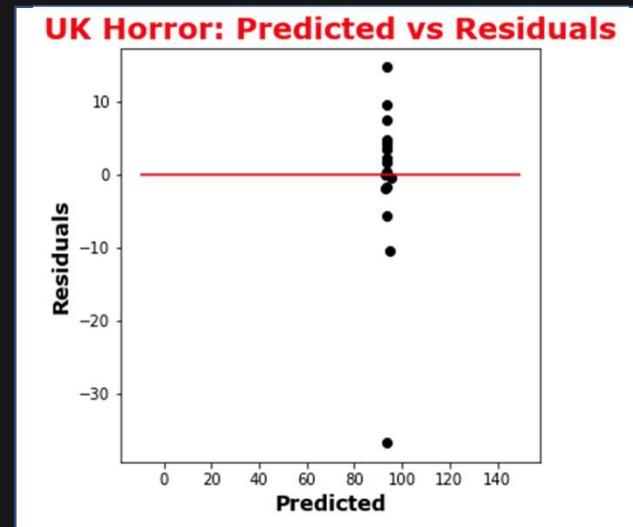
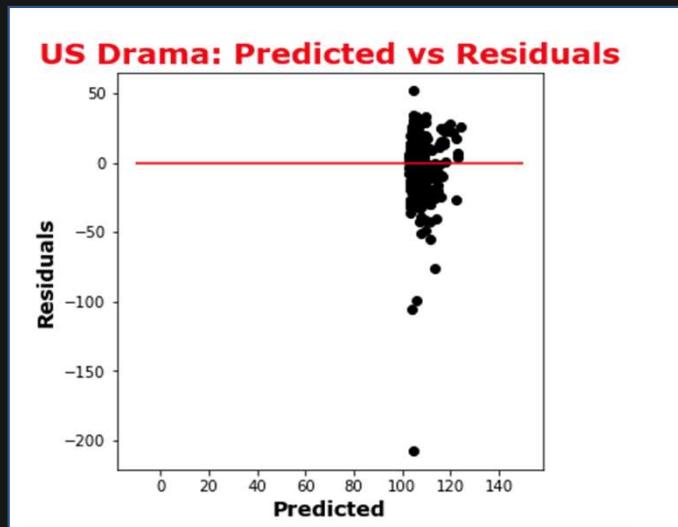
# Analysis - Horror



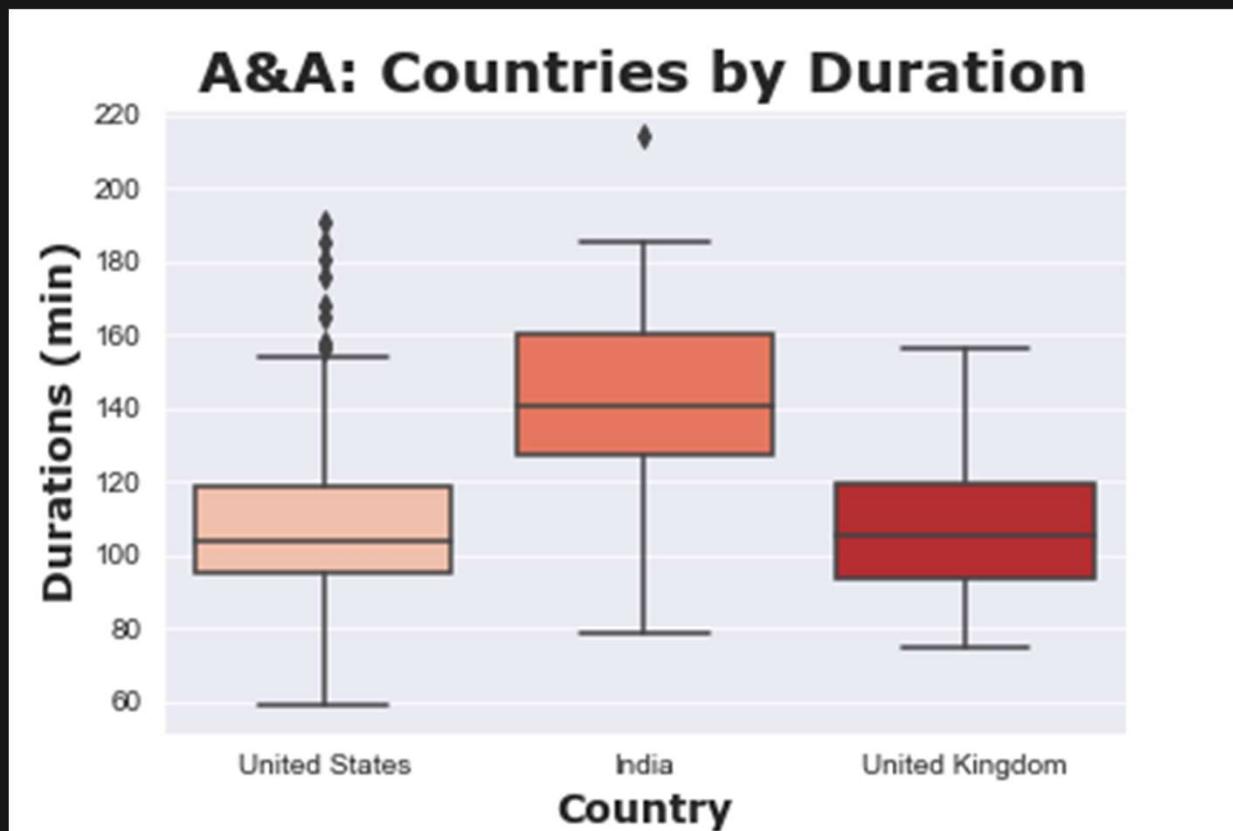
# Analysis - Horror



# Analysis - Horror



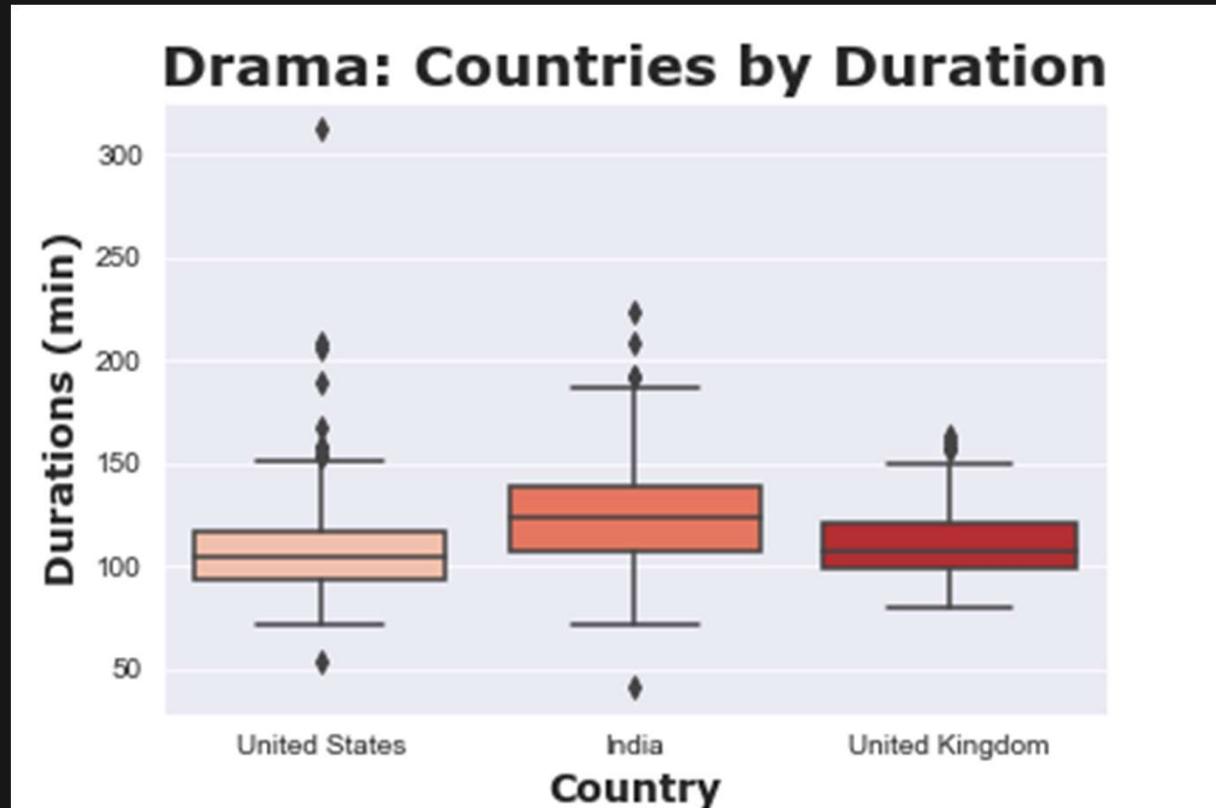
## Anova – Action & Adventure



```
Ttest_indResult(statistic=0.40707701583594236, pvalue=0.6841861826376556)
```



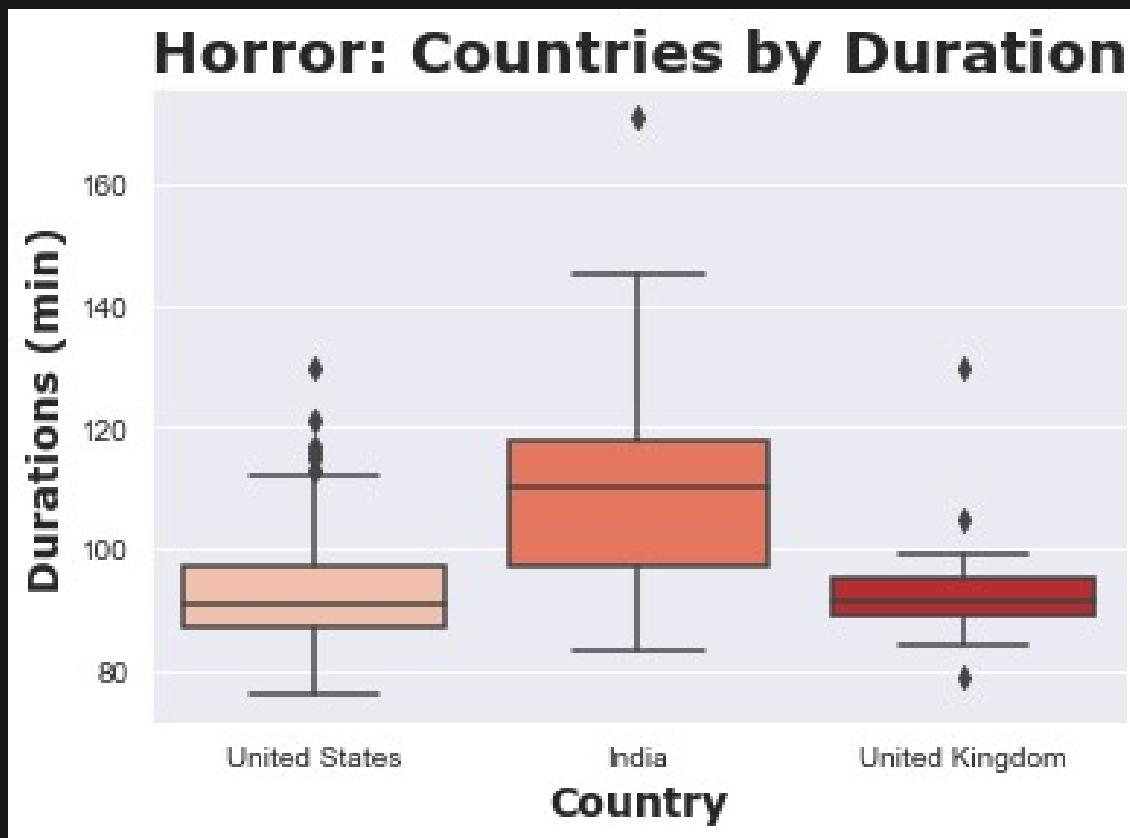
## Anova - Drama



```
Ttest_indResult(statistic=-1.6113572098903137, pvalue=0.10771100114860035)
```



## Anova - Horror



```
Ttest_indResult(statistic=-0.3923435373234593, pvalue=0.6954097017553829)
```



Hypotheses

Data

Analysis

Limitations  
& Future  
Analysis

Questions

## Limitations

Incomplete  
Data

Not all  
Countries  
Participate

Time  
Constraints

## Future Work

Deeper  
Genre &  
Country  
Analysis

Include  
IMDB  
Ratings in  
Analysis

Continue  
Trend  
Analysis on  
Duration



Hypotheses

Data

Analysis

Limitations  
& Future  
Analysis

Questions

# Questions?



# THANKS

► Next Episode