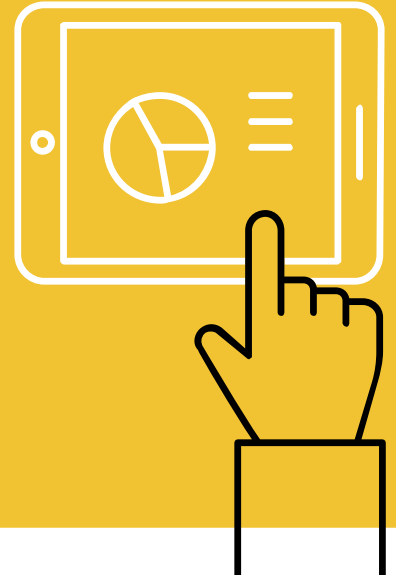
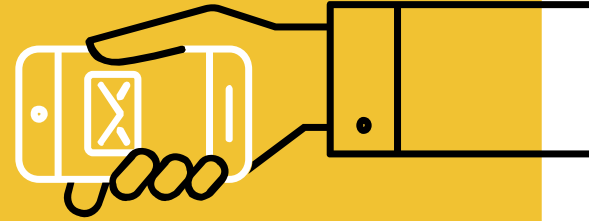
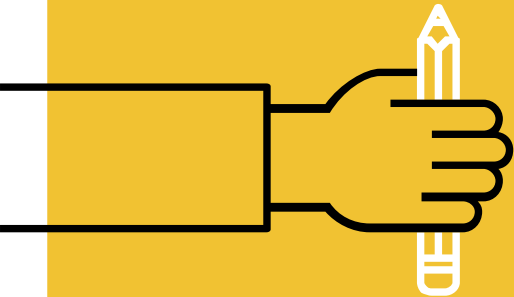
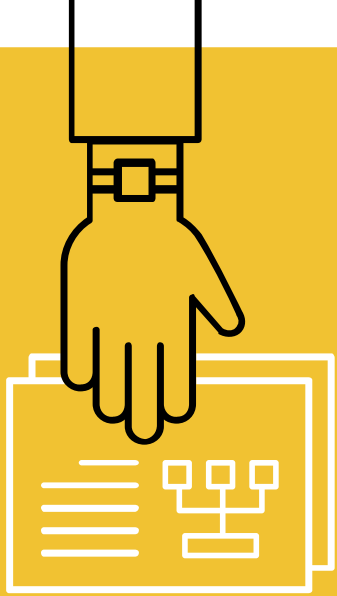
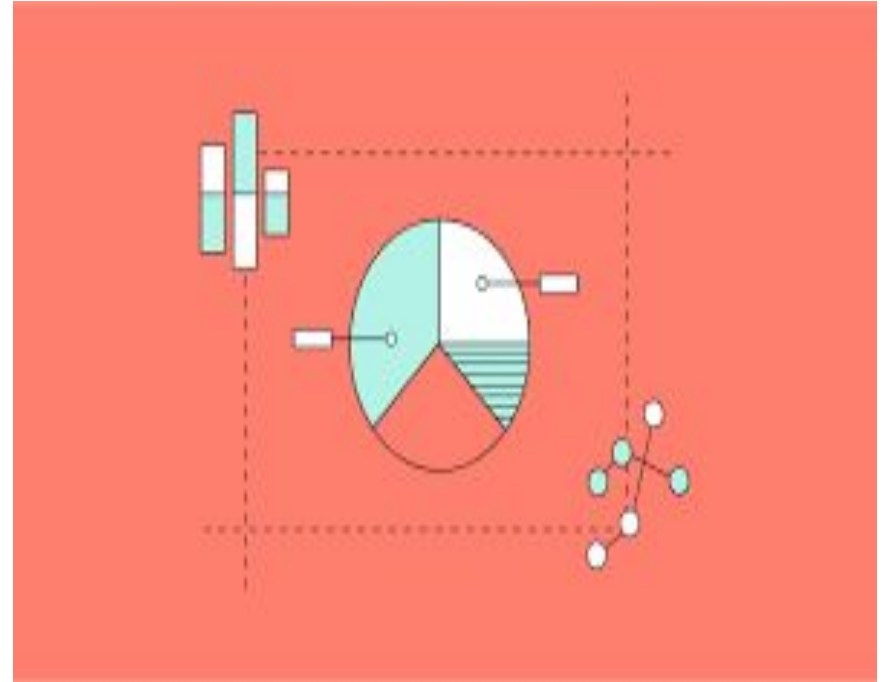


NETFLIX DATA ANALYSIS



Introduction

- Data visualization is a powerful tool that can help us understand complex data patterns and trends more easily.
- Using the visualization tools we are identifying patterns in the content that users tend to enjoy and display it.



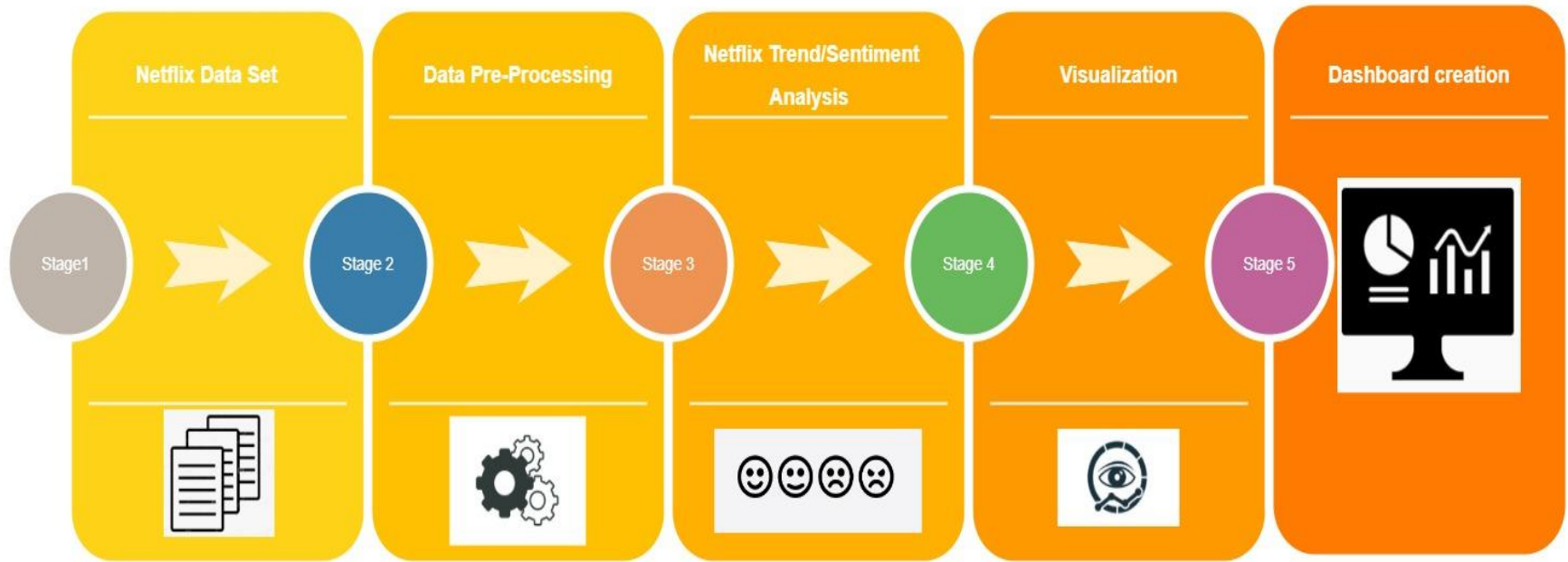
Motivation

- It is to unlock valuable insights for businesses and consumers such as it is possible to understand which countries produce the most content on Netflix, which genres are most popular, and how the content has evolved over time and to understand how major companies like Netflix deal with such huge amounts of data and generate recommendations accordingly.

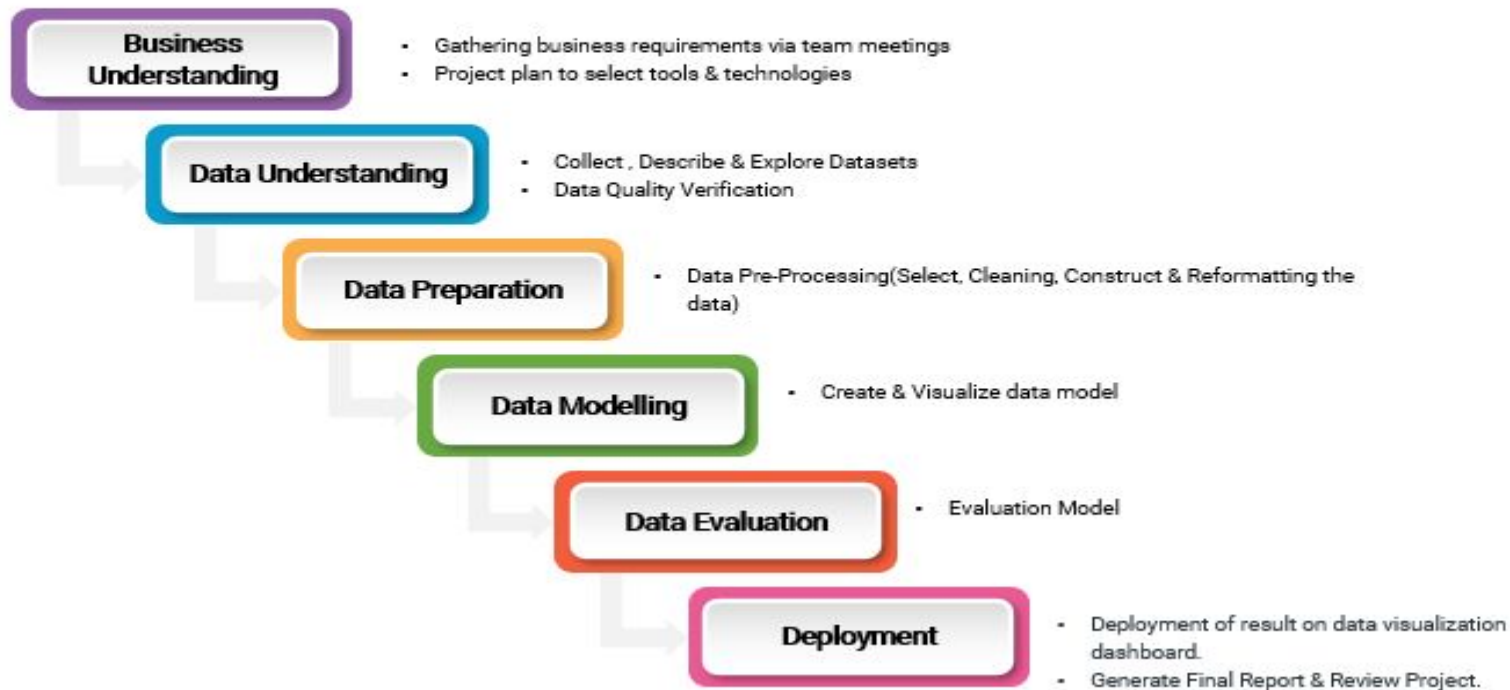
OBJECTIVE

- This dataset contains an exhaustive list of every Netflix TV show and movie that is currently available for streaming. The primary objective of this dataset is to provide information on the various TV episodes and movies that are now available on Netflix as well as the top shows and films by region. Which helps customers sort their list and enjoy
- Business people can get more insights and strengthen their business strategies based on different aspects which we are going to visualize.

Data Processing Diagram



Hybrid CRISP DM & Waterfall Model



Description of the dataset and how it was obtained

- This is a online dataset which was obtained from **kaggle** website.
- The Netflix streaming service's movie and television programming are listed in this dataset. There are specific columns like the title, director, cast, country of origin, year of release, rating, runtime, and genre included (e.g. action, drama, comedy). There are 7787 entries in all of the data.

```
In [4]: # To find unique content on each row
data.nunique()
```

```
Out[4]: show_id      7787
        type         2
        title      7787
        director   4049
        cast       6831
        country    681
        date_added 1565
        release_year 73
        rating     14
        duration   216
        listed_in  492
        description 7769
        dtype: int64
```

```
In [7]: # Counting null values in each column
data.isnull().sum()
```

```
Out[7]: show_id      0
        type         0
        title         0
        director   2389
        cast        718
        country     507
        date_added  10
        release_year 0
        rating       7
        duration     0
        listed_in    0
        description 0
        dtype: int64
```

READING THE DATASET

DATA CLEANING

```
In [9]: data.director.fillna("No Director", inplace=True)
data.cast.fillna("No Cast", inplace=True)
data.country.fillna("Country Unavailable", inplace=True)
data.dropna(subset=["date_added", "rating"], inplace=True)
# dropped the rows with NaN values in "date_added", "rating" columns because they are very few.
```

```
In [10]: data.isnull().sum()
```

```
Out[10]: show_id      0
         type         0
         title        0
         director     0
         cast         0
         country      0
         date_added   0
         release_year  0
         rating       0
         duration     0
         listed_in    0
         description  0
         dtype: int64
```

We have performed data cleaning checking the null columns and inserting data in it.

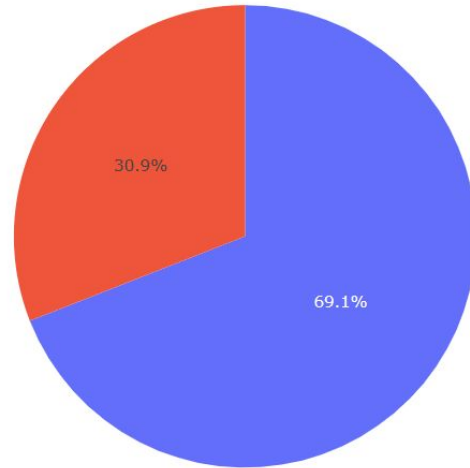
EXPLORATORY DATA ANALYSIS

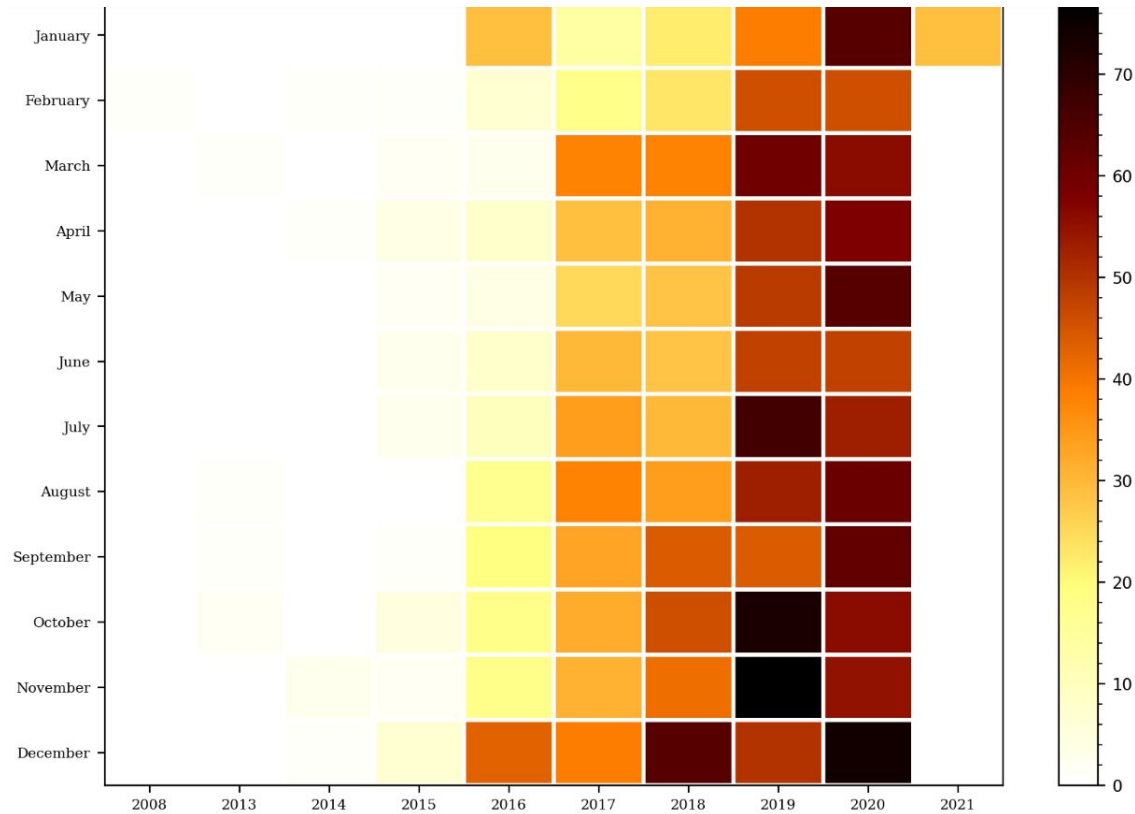
```
In [5]: Visualization = px.pie(values=data['type'].value_counts(),  
                               names=data['type'].value_counts().index,title='Types of Content on Netflix')  
Visualization.show()
```

Types of Content

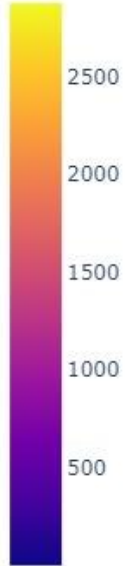
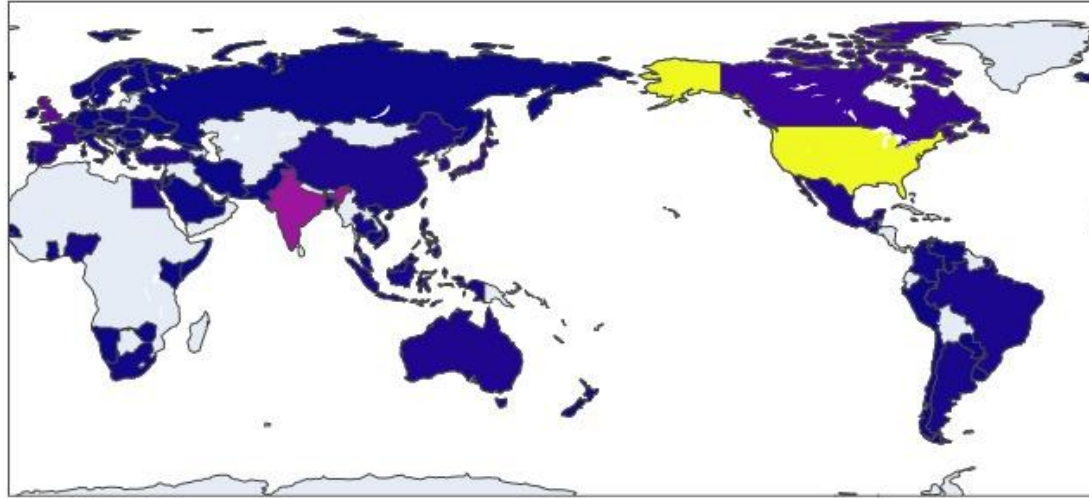
- The PIE CHART compares the content types on NETFLIX
- Overall, the number of movies is significantly larger than the TV show count.

Types of Content on Netflix





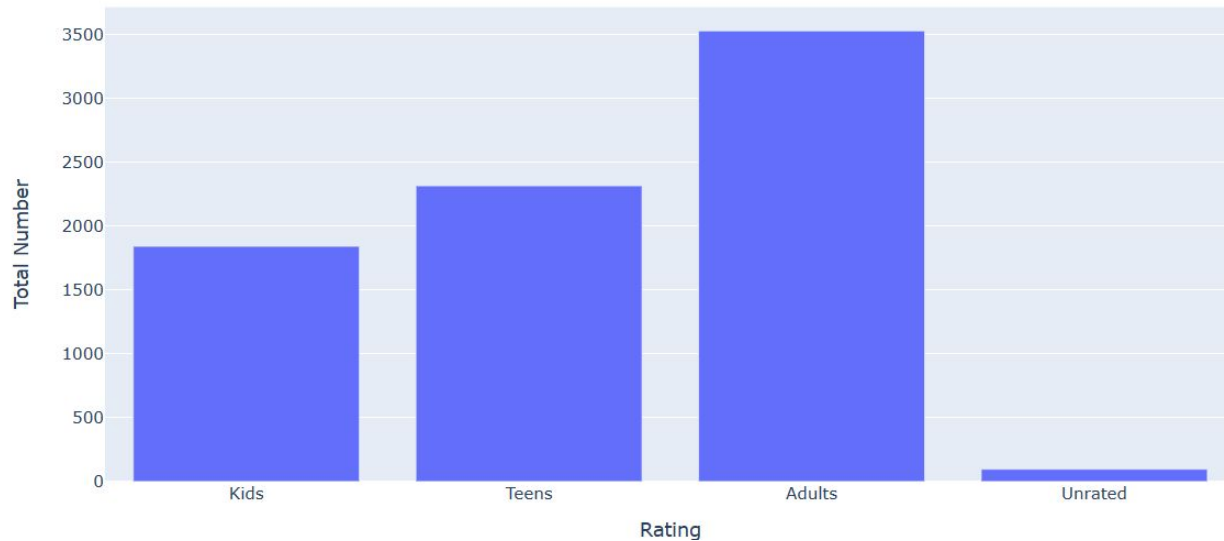
BEST MONTHS TO RELEASE MOVIES



Number of Netflix content viewers per country

```
Adults    3531
Teens     2317
Kids      1843
Unrated    96
Name: rating_group, dtype: int64
```

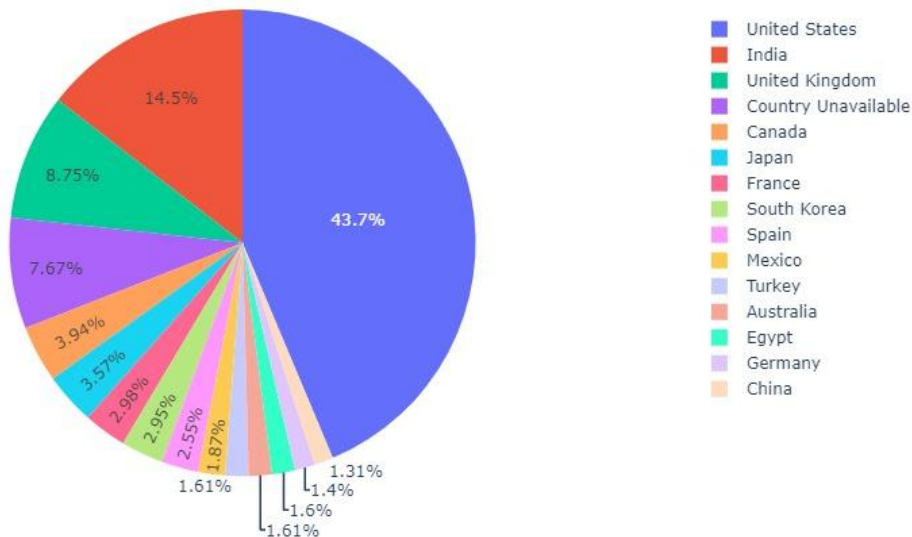
TV-Shows and Movies Rating in Netflix



Distribution of Movie ratings (based on age).

```
In [23]: Visualization = px.pie(values=top15countries,
                                names=top15countries.index,title='Top 15 Countries producing the content to Netflix')
Visualization.show()
```

Top 15 Countries producing the content to Netflix



Count of movies Produced by country

Project Status

WHAT ALL DID WE DO

- Choosing the Dataset
- Data Cleaning
- Data preprocessing.
- Exploratory Data Analysis on the dataset.
- Visualizing the analysed data.

WHAT ALL ARE WE PLANNING ON DOING

- Visualize the data to get more insights.
- Implementing Machine learning techniques to generate movie/tv show recommendations.

Thank you

GROUP-4

- PAVAN
- IRFAN
- SAI
- NAVEEN
- SARAN