**NETFLIX DATA ANALYSIS**

**Mini Project 2021 Report**

**SUBMITTED TO**

**MR. Anurag Mishra SIR**

****

**ABES ENGINEERING COLLEGE (032), GHAZIABAD, UP**

**BY**

**SHIVAM TYAGI**

**1900320100151**

**TABLE OF CONTENTS**

Page No.

DECLARATION ii

CERTIFICATE iii

ACKNOWLEDGEMENTS iv

ABSTRACT v

LIST OF FIGURES viii

**CHAPTER-1INTRODUCTION 10**

1.1 Problem Definition 10

1.2 Motivation 10

1.3 Objective of the Project 11

1.4 Scope of the Project 11

1.5 Need of Work 11

**CHAPTER 2 RELATED WORK 12**

**CHAPTER 3 PROPOSED METHODOLOGY 13**

3.1 Dataset Description 13

3.2 Methods

3.3 Hardware / Software Requirements

3.3 Univariate Analysis

3.4 Bivariate Analysis

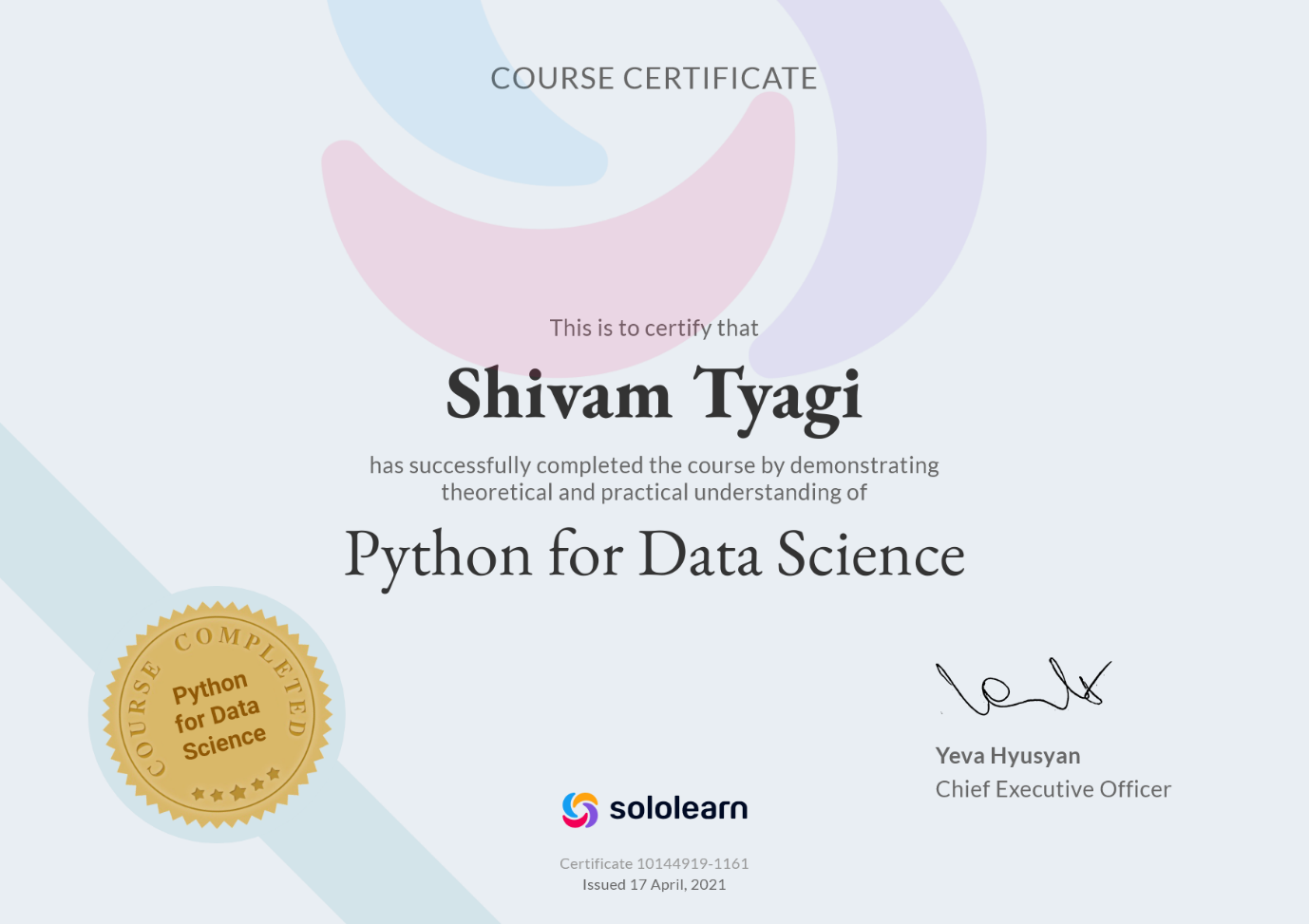
**REFERENCES**

DECLARATION

**I have created a project on data analytics, I am using the ticks and techniques of data analytics which I had learn in my training of data analytics by MR. Gopal Sir in our college. The name of my project is Netflix Data Analyses using Kaggle dataset the name of dataset is Netflix\_title.csv.**

CERTIFICATE

THIS is the course certification which I had completed after getting the tips and techniques from this course.



ACKNOWLEDGEMENTS

I would like to express my deep sense of gratitude and convey thanks to everyone who

helped me and supported during the completion of this project.

This Project is based on Data Science specially on Data Analytics by using Kaggle Dataset. In this project I had learn how to use Kaggle real dataset in data analytics. The Course of Data Analytics is very helpful for me which was organised by ABES ENGINEERING COLLEGE through MR. Gopal Sir.

ABSTRACT

Now days we can watch any movies and web series online on our mobile or TV through OTT platforms like Prime Video, Netflix etc by taking there subscription according to our need and money. In this project I am taking the Netflix to know what type of content they are providing to watch and which country is lunching it and many more.

LIST OF FIGURES

**Missing Values L1**

**Handling missing data L2**

**Analysis**

**Which countries have the highest number of movies and tv shows L3**

**Solve for countries with the highest numbers of movies alone L4**

**Content added over the years L5**

**Calculating the change in content addition during**

**Covid-19 pandemic L6**

**Movie Content? L7**

**TV Content L8**

Chapter 1

Introduction

Problem Definition:

How can subscribers of Netflix can Get the best quality of content which is also beneficial for Netflix to stand out in the field of OTT platforms?

Motivation:

1.I wants to know that **which countries have the highest number of movies and tv shows.**

**2 Solve for countries with the highest numbers of movies alone.**

Objective of the Project:

By this project we can know that what type of content is lunched and which country is lunching it in which year.

Need of Work:

IT can provide an insight that how can we find the best move or web series for us to watch according to our need.

**CHAPTER: 2 RELATED WORK**

**1.Data Information**

**BY df.info() we get where df is the data set name as netflix\_titels.csv.**

<class 'pandas.core.frame.DataFrame'>

RangeIndex: 8807 entries, 0 to 8806

Data columns (total 12 columns):

# Column Non-Null Count Dtype

--- ------ -------------- -----

0 show\_id 8807 non-null object

1 type 8807 non-null object

2 title 8807 non-null object

3 director 6173 non-null object

4 cast 7982 non-null object

5 country 7976 non-null object

6 date\_added 8797 non-null object

7 release\_year 8807 non-null int64

8 rating 8803 non-null object

9 duration 8804 non-null object

10 listed\_in 8807 non-null object

11 description 8807 non-null object

dtypes: int64(1), object(11)

memory usage: 825.8+ KB

2.DATA DESCRIBE

BY using df.describe( )

release\_year

count 8807.000000

mean 2014.180198

std 8.819312

min 1925.000000

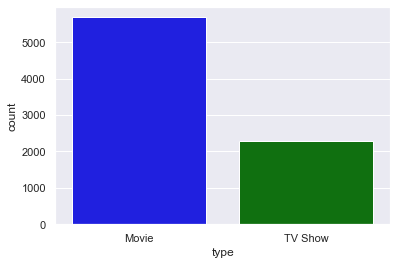
25% 2013.000000

50% 2017.000000

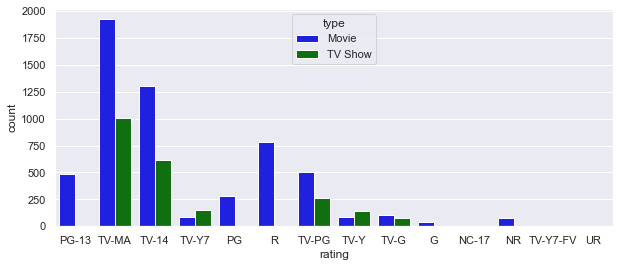
75% 2019.000000

max 2021.000000

3. No of movie and TV show.



4. Relation between movies and shows with their type.



5. RATING AGE COUNT.

0 Teen

1 Adult

4 Adult

7 Adult

8 Adolesent

9 Teen

12 Adult

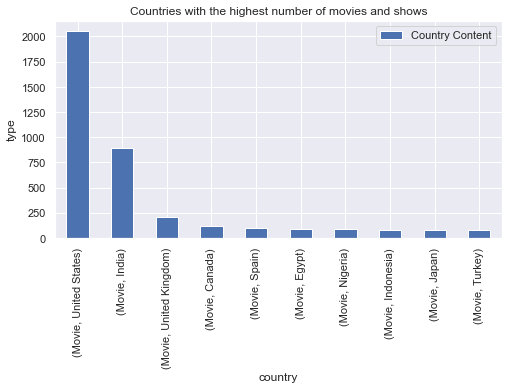
15 Adult

17 Adult

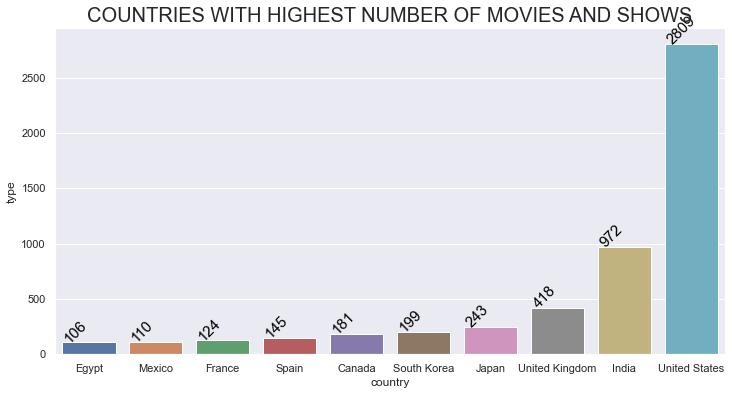
21 Adolesent

Name: rating\_ages, dtype: object

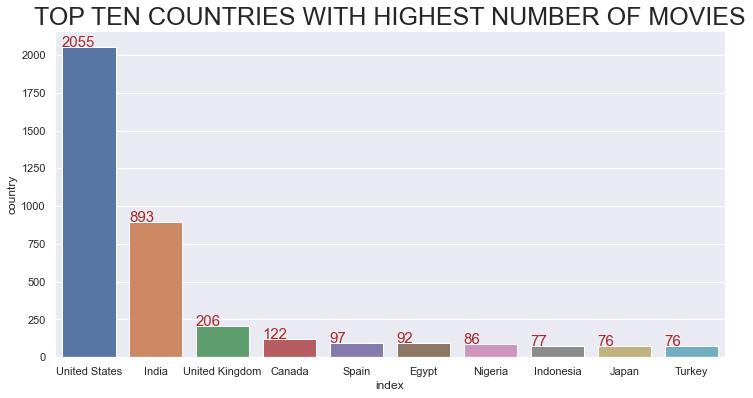
6 Country with the highest no of movies and TV shows.



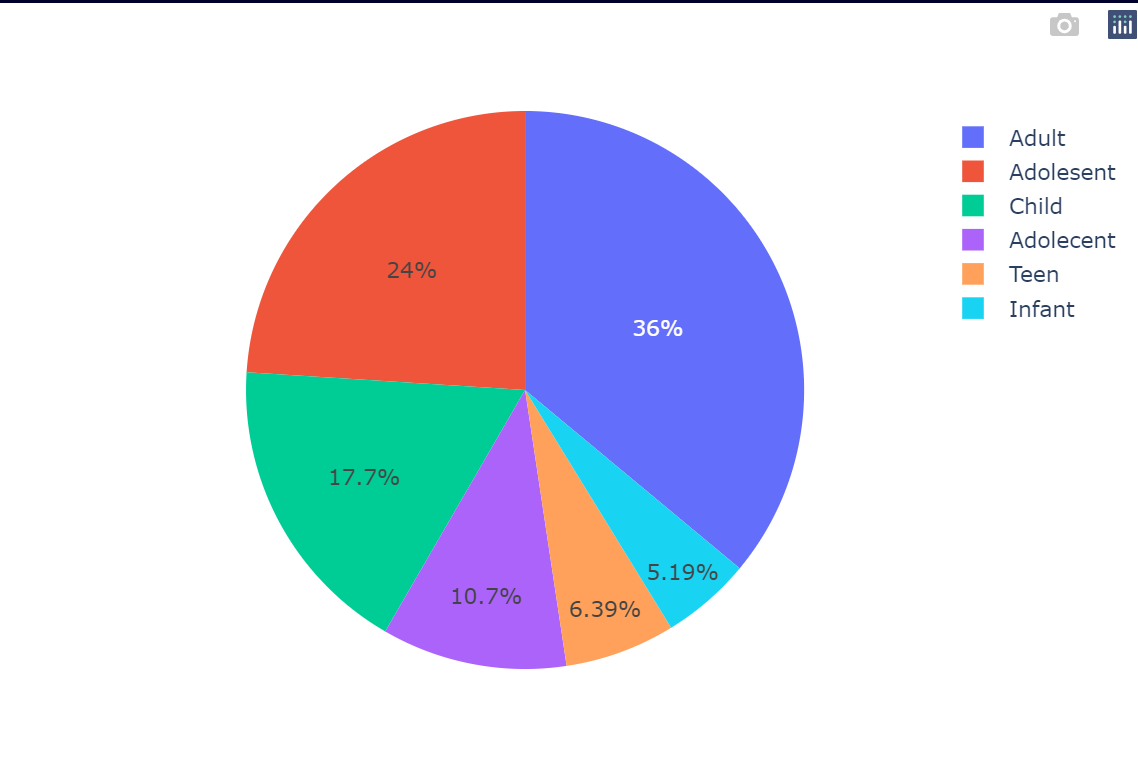
7.



8.



9. Type of content.

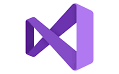
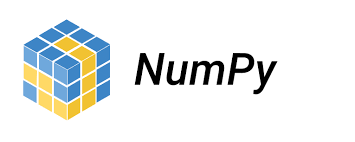
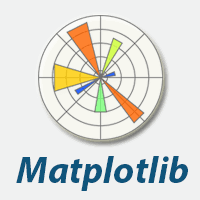
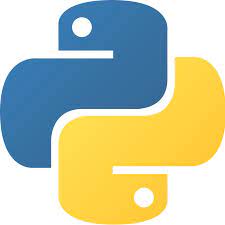




CHAPTER 3

Hardware / Software Requirements

I am using JUPYTER NOTBOOK and Visual Studio with Python language and its packages.

References

I have learned data analytics from this course which very useful for me, by which I have competed different course on Solo Learn app and on Programming Hub also. And Kaggle is very useful for me to getting real data in csv format.