

The data

netflix_data.csv

	Column		Description	
	show_id		The ID of the show	
	type		Type of show	
	title		Title of the show	
	director		Director of the show	
	cast		Cast of the show	
	country		Country of origin	
	date_added		Date added to Netflix	
	release_year		Year of Netflix release	
	duration		Duration of the show in minutes	
	description		Description of the show	
	genre		Show genre	

Import libraries

```
In [1]: # Importing the pandas library, which is used for data manipulation and analysis
import pandas as pd

# Importing the matplotlib library (specifically pyplot), which is used for creating visualizations
import matplotlib.pyplot as plt

In [2]: # Reading a CSV file named "netflix_data.csv" into a pandas DataFrame.
# A DataFrame is a 2-dimensional, table-like data structure (rows and columns); Similar to excel spreadsheet.
df = pd.read_csv("netflix_data.csv")
```

Basic Data Analysis Functions and Methods

1. Data Inspection

```
# head(): View the first few rows of the dataset.
# tail(): View the last few rows of the dataset.
# info(): Get a summary of the dataset, including column data types and non-null counts.

In [3]: # Displaying the first 5 rows of the DataFrame to get a quick overview of the data.
# This helps you understand the structure and content of the dataset.
df.head()
```

	show_id	type	title	director	cast	country	date_added	release_year	duration	description	genre
0	s1	TV Show	3%	NaN	João Miguel, Bianca Comparato, Michel Gomes, R...	Brazil	August 14, 2020	2020	4	In a future where the elite inhabit an island ...	International TV
1	s2	Movie	7:19	Jorge Michel Grau	Demían Bichir, Héctor Bonilla, Oscar Serrano, ...	Mexico	December 23, 2016	2016	93	After a devastating earthquake hits Mexico City...	Dramas
2	s3	Movie	23:59	Gilbert Chan	Tedd Chan, Stella Chung, Henley Hii, Lawrence ...	Singapore	December 20, 2018	2011	78	When an army recruit is found dead, his fellow...	Horror Movies
3	s4	Movie	9	Shane Acker	Elijah Wood, John C. Reilly, Jennifer Connelly...	United States	November 16, 2017	2009	80	In a postapocalyptic world, rag-doll robots hi...	Action
4	s5	Movie	21	Robert Luketic	Jim Sturgess, Kevin Spacey, Kate Bosworth, Aar...	United States	January 1, 2020	2008	123	A brilliant group of students become card-coun...	Dramas

```
In [4]: # Displaying the last 5 rows of the DataFrame to get an overview of the dataset's ending portion.
# This is helpful to check the structure and values at the end of the dataset.
df.tail()
```

	show_id	type	title	director	cast	country	date_added	release_year	duration	description	genre
7782	s7783	Movie	Zozo	Josef Fares	Imad Creidi, Antoinette Turk, Elias Gergji, Car...	Sweden	October 19, 2020	2005	99	When Lebanon's Civil War deprives Zozo of his ...	Dramas
7783	s7784	Movie	Zubaan	Mozez Singh	Vicky Kaushal, Sarah-Jane Dias, Raaghav Chanan...	India	March 2, 2019	2015	111	A scrappy but poor boy worms his way into a ty...	Dramas
7784	s7785	Movie	Zulu Man in Japan	NaN	Nasty C	NaN	September 25, 2020	2019	44	In this documentary, South African rapper Nast...	Documentaries
7785	s7786	TV Show	Zumbo's Just Desserts	NaN	Adriano Zumbo, Rachel Khoo	Australia	October 31, 2020	2019	1	Dessert wizard Adriano Zumbo looks for the nex...	International TV
7786	s7787	Movie	ZZ TOP: THAT LITTLE OL' BAND FROM TEXAS	Sam Dunn	NaN	United Kingdom	March 1, 2020	2019	90	This documentary delves into the mystique behi...	Documentaries

```
In [5]: # Displaying a summary of the DataFrame, including:
# - The number of rows and columns.
# - Column names and their data types (e.g., int, float, object).
# - The number of non-null (non-missing) values in each column.
df.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 7787 entries, 0 to 7786
Data columns (total 11 columns):
 #   Column      Non-Null Count  Dtype
---  --
 0   show_id     7787 non-null   object
 1   type        7787 non-null   object
 2   title       7787 non-null   object
 3   director    5398 non-null   object
 4   cast        7069 non-null   object
 5   country     7280 non-null   object
 6   date_added  7777 non-null   object
 7   release_year 7787 non-null   int64
 8   duration    7787 non-null   int64
 9   description 7787 non-null   object
10   genre       7787 non-null   object
dtypes: int64(2), object(9)
memory usage: 669.3+ KB
```

2. Summary Statistics

```
# describe(): Generates descriptive statistics for numerical columns, such as mean, standard deviation, minimum, and maximum values.
# value_counts(): Counts occurrences of unique values in a column (e.g., genres or ratings).
```

```
In [6]: # Providing a statistical summary of numerical columns in the DataFrame.
# This includes metrics like:
# - Count: Number of non-missing values in each column.
# - Mean: The average value.
# - Std: The standard deviation, showing data spread.
# - Min: The minimum value.
# - 25%, 50%, 75%: Percentiles (25th, median, and 75th).
# - Max: The maximum value.
df.describe()
```

	release_year	duration
count	7787.000000	7787.000000
mean	2013.932580	69.122769
std	8.757395	50.950743
min	1925.000000	1.000000
25%	2013.000000	2.000000
50%	2017.000000	88.000000
75%	2018.000000	106.000000
max	2021.000000	312.000000

```
In [7]: # The line below calculates the frequency of each unique value in the 'release_year' column.
# It returns a summary showing how many movies or shows were released in each year.
# The output is sorted by default in descending order (most frequent years appear first).
df['release_year'].value_counts()
```

```
release_year
2018    1121
2017    1012
2019     996
2016     882
2020     868
      ...
1925         1
1964         1
1966         1
1947         1
1959         1
Name: count, Length: 73, dtype: int64
```

```
In [8]: # The line below calculates the frequency of each unique value in the 'release_year' column.
# It returns a summary showing how many movies or shows were released in each year.
# By default, the result is sorted in descending order (most frequent years appear first).
# If you want to sort the result in ascending order (least frequent years first), use the following:
df['release_year'].value_counts(ascending=True)
```

```
release_year
1947         1
1966         1
1959         1
1964         1
1925         1
      ...
2020     868
2016     882
2019     996
2017    1012
2018    1121
Name: count, Length: 73, dtype: int64
```

3. Sorting and Filtering

```
# sort_values(): Sort the dataset by a specific column in ascending or descending order.
# Filtering rows: Select rows that fulfil a certain criteria
```

```
In [9]: # The line below sorts the DataFrame by the 'release_year' column in descending order.
# This means that the rows with the most recent release years will appear first in the DataFrame.
# If you wanted to sort the values in ascending order (oldest years first), use ascending=True.
df.sort_values('release_year', ascending=False)
```

	show_id	type	title	director	cast	country	date_added	release_year	duration	description	genre
2327	s2328	TV Show	Gabby's Dollhouse	NaN	Laila Lockhart Kraner, Tucker Chandler, Juliet...	NaN	January 5, 2021	2021	1	Cute crafts and colorful magic! J...	Kids
4468	s4469	TV Show	Night Stalker: The Hunt for a Serial Killer	Tiller Russell	NaN	NaN	January 13, 2021	2021	1	Beneath the sunlit glamour of 1985 LA lurks a ...	Crime TV
2753	s2754	TV Show	History of Swear Words	NaN	Nicolas Cage	United States	January 5, 2021	2021	1	Nicolas Cage hosts this proudly profane, funny...	Docuseries
5103	s5104	Movie	Ratones Paranoicos: The Band that Rocked Argent...	Alejandro Ruaz, Ramiro Martinez	Juan Sebastián Gutiérrez, Pablo Cano, Pablo Me...	NaN	January 6, 2021	2021	76	The irrepressible Ratones Paranoicos, Argentin...	Documentaries
1355	s1356	Movie	Chris Rock Total Blackout: The Tamboirne Exten...	Chris Rock	Chris Rock	NaN	January 12, 2021	2021	98	In this extended cut of his 2018 special, Chri...	Stand-Up
...
7616	s7617	Movie	Why We Fight: The Battle of Russia	Frank Capra, Anatole Litvak	NaN	United States	March 31, 2017	1943	82	This installment of Frank Capra's acclaimed do...	Documentaries
7342	s7343	Movie	Undercover: How to Operate Behind Enemy Lines	John Ford	NaN	United States	March 31, 2017	1943	61	This World War II-era training film dramatizes...	Classic Movies
6117	s6118	Movie	The Battle of Midway	John Ford	Henry Fonda, Jane Danwell	United States	March 31, 2017	1942	18	Director John Ford captures combat footage of ...	Classic Movies
4960	s4961	Movie	Prelude to War	Frank Capra	NaN	United States	March 31, 2017	1942	52	Frank Capra's documentary chronicles the rise ...	Classic Movies
4867	s4868	TV Show	Pioneers: First Women Filmmakers*	NaN	NaN	NaN	December 30, 2018	1925	1	This collection restores films from women who ...	TV Shows

7787 rows × 11 columns

```
In [10]: # The line below filters the rows in the DataFrame where the 'genre' column is equal to 'Comedies'.
# It creates a new DataFrame that only contains the rows where the genre is 'Comedies'.
# You can change 'Comedies' to any other genre you want to filter by.
df[df['genre'] == 'Comedies']
```

	show_id	type	title	director	cast	country	date_added	release_year	duration	description	genre
18	s19	Movie	15-Aug	Swapnoneel Jayakar	Rahul Pethhe, Mrunmayee Deshpande, Adinath Koth...	India	March 29, 2019	2019	124	On India's Independence Day, a zany mishap in ...	Comedies
33	s34	Movie	#realityhigh	Fernando Lebrija	Nesta Cooper, Kate Walsh, John Michael Higgins...	United States	September 8, 2017	2017	99	When nerdy high schooler Dani finally attracts...	Comedies
34	s35	Movie	#RoxY	Michael Kennedy	Jake Short, Sarah Fisher, Booboo Stewart, Dan...	Canada	April 10, 2019	2018	105	A teenage hacker with a huge nose helps a cool...	Comedies
36	s37	Movie	#Selfie	Cristina Jacob	Flavia Hojda, Crina Semicic, Olimpia Melinte...	Romania	June 1, 2019	2014	125	Two days before their final exams, three teen ...	Comedies
37	s38	Movie	#Selfie 69	Cristina Jacob	Maia Morgenstern, Olimpia Melinte, Crina Semic...	Romania	June 1, 2019	2016	119	After a painful breakup, a trio of party-lovin...	Comedies
...
7760	s7761	Movie	Zaki Chan	Wael Ihsan	Ahmed Helmy, Yasmin Abdulaziz, Hassan Hosny, H...	Egypt	May 19, 2020	2005	109	An unqualified young man has his work cut out ...	Comedies
7762	s7763	Movie	Zed Plus	Chandra Prakash Dwivedi	Adil Hussain, Mona Singh, K.K. Raina, Sanjay M...	India	December 31, 2019	2014	131	A philandering small-town mechanic's political...	Comedies
7764	s7765	Movie	Zero	Aanand Rai	Shah Rukh Khan, Anushka Sharma, Katrina Kaif, ...	India	May 21, 2019	2019	159	Through his relationships with two wildly diff...	Comedies
7769	s7770	Movie	Zindagi Na Milegi Dobara	Zoya Akhtar	Hrithik Roshan, Abhay Deol, Farhan Akhtar, Kat...	India	December 15, 2019	2011	154	Three friends on an adventurous road trip/back...	Comedies
7778	s7779	Movie	Zombieland	Ruben Fleischer	Jesse Eisenberg, Woody from Toy Story, Emma Stone, ...	United States	November 1, 2019	2009	88	Looking to survive in a world taken over by zo...	Comedies

1074 rows × 11 columns

```
In [ ]: 
```

Goal of the Analysis

The goal of this analysis is to explore whether movies are shorter or longer than they were 30 years ago.

```
In [11]: df.head()
```

	show_id	type	title	director	cast	country	date_added	release_year	duration	description	genre
0	s1	TV Show	3%	NaN	João Miguel, Bianca Comparato, Michel Gomes, R...	Brazil	August 14, 2020	2020	4	In a future where the elite inhabit an island ...	International TV
1	s2	Movie	7:19	Jorge Michel Grau	Demían Bichir, Héctor Bonilla, Oscar Serrano, ...	Mexico	December 23, 2016	2016	93	After a devastating earthquake hits Mexico City...	Dramas
2	s3	Movie	23:59	Gilbert Chan	Tedd Chan, Stella Chung, Henley Hii, Lawrence ...	Singapore	December 20, 2018	2011	78	When an army recruit is found dead, his fellow...	Horror Movies
3	s4	Movie	9	Shane Acker	Elijah Wood, John C. Reilly, Jennifer Connelly...	United States	November 16, 2017	2009	80	In a postapocalyptic world, rag-doll robots hi...	Action
4	s5	Movie	21	Robert Luketic	Jim Sturgess, Kevin Spacey, Kate Bosworth, Aar...	United States	January 1, 2020	2008	123	A brilliant group of students become card-coun...	Dramas

```
In [12]: # The line below filters the rows in the DataFrame where the 'type' column is equal to 'Movie'.
# It creates a new DataFrame containing only the rows where the type is 'Movie'.
# This means it filters out any rows where the 'type' is not 'Movie', leaving us with a subset of the data that
df[df['type'] == 'Movie']
```

	show_id	type	title	director	cast	country	date_added	release_year	duration	description	genre
1	s2	Movie	7:19	Jorge Michel Grau	Demían Bichir, Antoinette Turk, Oscar Serrano, ...	Mexico	December 23, 2016	2016	93	After a devastating earthquake hits Mexico City...	Dramas
2	s3	Movie	23:59	Gilbert Chan	Tedd Chan, Stella Chung, Henley Hii, Lawrence ...	Singapore	December 20, 2018	2011	78	When an army recruit is found dead, his fellow...	Horror Movies
3	s4	Movie	9	Shane Acker	Elijah Wood, John C. Reilly, Jennifer Connelly...	United States	November 16, 2017	2009	80	In a postapocalyptic world, rag-doll robots hi...	Action
4	s5	Movie	21	Robert Luketic	Jim Sturgess, Kevin Spacey, Kate Bosworth, Aar...	United States	January 1, 2020	2008	123	A brilliant group of students become card-coun...	Dramas
6	s7	Movie	122	Yasir Al Yasiri	Amina Khalid, Ahmed Dawood, Tarek Lotfy, Ahmed...	Egypt	June 1, 2020	2019	95	After an awful accident, a couple admitted to a ...	Horror Movies
...
7781	s7782	Movie	Zoom	Peter Hewitt	Tim Allen, Courtney Cox, Chevy Chase, Kate Ma...	United States	January 11, 2020	2006	88	Dragged from civilian life, a former superhero...	Children
7782	s7783	Movie	Zozo	Josef Fares	Imad Creidi, Antoinette Turk, Elias Gergji, Car...	Sweden	October 19, 2020	2005	99	When Lebanon's Civil War deprives Zozo of his ...	Dramas
7783	s7784	Movie	Zubaan	Mozez Singh	Vicky Kaushal, Sarah-Jane Dias, Raaghav Chanan...	India	March 2, 2019	2015	111	A scrappy but poor boy worms his way into a ty...	Dramas
7784	s7785	Movie	Zulu Man in Japan	NaN	Nasty C	NaN	September 25, 2020	2019	44	In this documentary, South African rapper Nast...	Documentaries
7786	s7787	Movie	ZZ TOP: THAT LITTLE OL' BAND FROM TEXAS	Sam Dunn	NaN	United Kingdom	March 1, 2020	2019	90	This documentary delves into the mystique behi...	Documentaries

5377 rows × 11 columns

```
In [13]: # The line below creates a new DataFrame called 'netflix_subset' that contains only the rows where the 'type' c
# This means it filters out any rows where the 'type' is not 'Movie', leaving us with a subset of the data that
df_subset = df[df['type'] == 'Movie']
```

```
In [14]: # The line below creates a new DataFrame called 'netflix_movies' that only includes the selected columns:
# 'title', 'country', 'genre', 'release_year', and 'duration' from the 'netflix_subset' DataFrame.
# This is useful when we are only interested in specific columns and want to reduce the size of the data for fu
netflix_movies = netflix_subset[['title', 'country', 'genre', 'release_year', 'duration']]
```

```
In [15]: # View the first five rows of our data
netflix_movies.head()
```

	title	country	genre	release_year	duration
1	7:19	Mexico	Dramas	2016	93
2	23:59	Singapore	Horror Movies	2011	78
3	9	United States	Action	2009	80
4	21	United States	Dramas	2008	123
6	122	Egypt	Horror Movies	2019	95

```
In [16]: ### Do not worry about this ###
# Define an empty list
colors = []

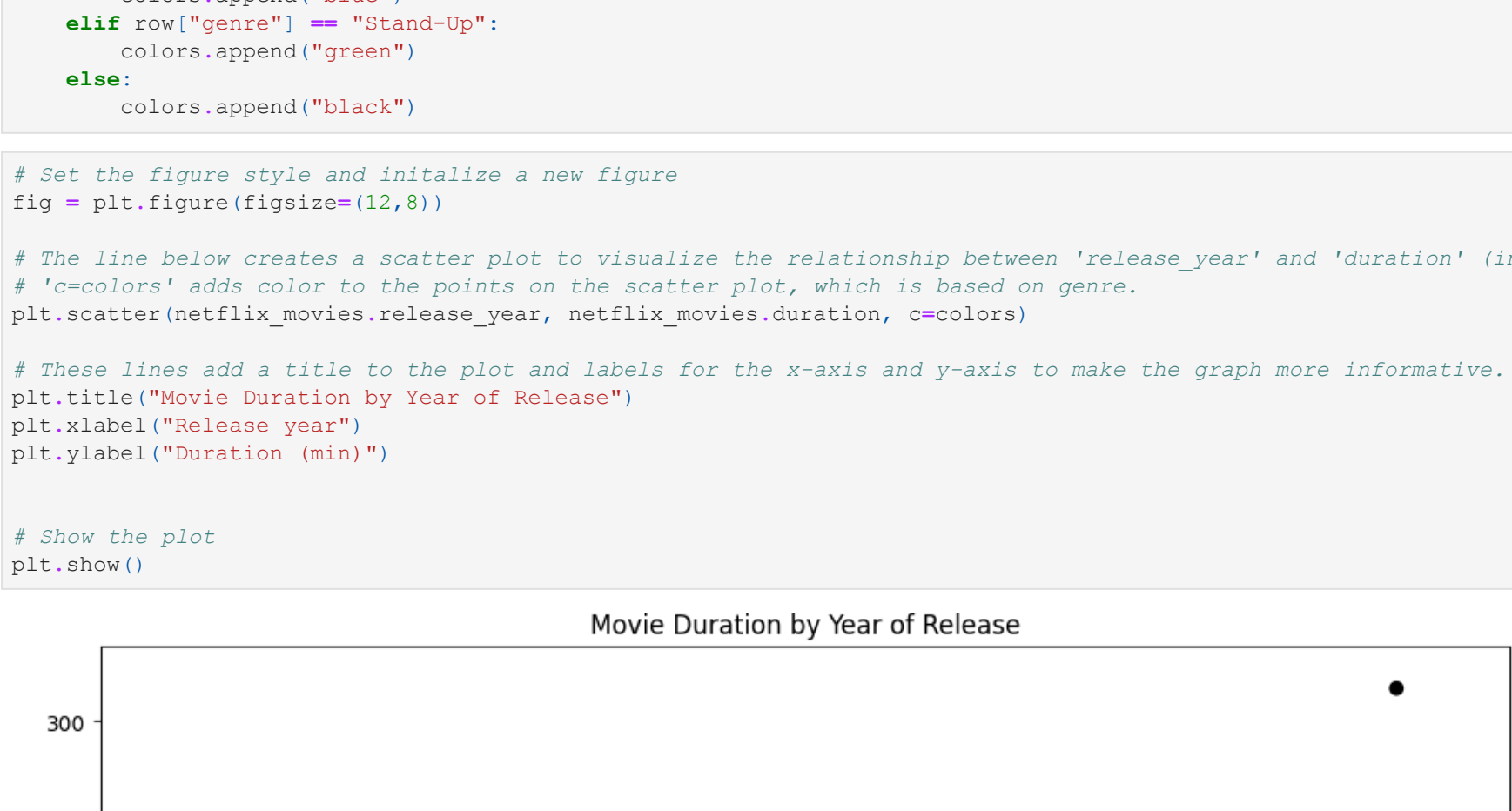
# Iterate over rows of netflix_movies
for label, row in netflix_movies.iterrows():
    if row['genre'] == "Children":
        colors.append("red")
    elif row['genre'] == "Documentaries":
        colors.append("blue")
    elif row['genre'] == "Stand-Up":
        colors.append("green")
    else:
        colors.append("black")
```

```
In [17]: # Set the figure style and initialize a new figure
fig = plt.figure(figsize=(12,8))

# The line below creates a scatter plot to visualize the relationship between 'release_year' and 'duration' (in
# 'c=colors' adds color to the points on the scatter plot, which is based on genre.
plt.scatter(netflix_movies.release_year, netflix_movies.duration, c=colors)

# These lines add a title to the plot and labels for the x-axis and y-axis to make the graph more informative.
plt.title("Movie Duration by Year of Release")
plt.xlabel("Release year")
plt.ylabel("Duration (min)")

# Show the plot
plt.show()
```



```
In [ ]: 
```

