## Solution

# Importing pandas and matplotlib

import pandas as pd

import matplotlib.pyplot as plt

# Read in the Netflix CSV as a DataFrame

netflix\_df = pd.read\_csv("netflix\_data.csv")

# Subset the DataFrame for type "Movie"

netflix\_subset = netflix\_df[netflix\_df["type"] == "Movie"]

# Select only the columns of interest

netflix\_movies = netflix\_subset[["title", "country", "genre", "release\_year", "duration"]]

# Filter for durations shorter than 60 minutes

short\_movies = netflix\_movies[netflix\_movies.duration < 60]

# 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")

# Inspect the first 10 values in your list

colors[:10]

# Set the figure style and initalize a new figure

fig = plt.figure(figsize=(12,8))

# Create a scatter plot of duration versus release\_year

plt.scatter(netflix\_movies.release\_year, netflix\_movies.duration, c=colors)

# Create a title and axis labels

plt.title("Movie Duration by Year of Release")

plt.xlabel("Release year")

plt.ylabel("Duration (min)")

# Show the plot

plt.show()

# Are we certain that movies are getting shorter?

answer = "no"