Netflix Subscriptions Forecasting

Netflix has become one of the most popular streaming platforms worldwide, and its subscription growth is a key metric for understanding the platform's success. By leveraging data-driven techniques like time series forecasting, we can estimate the number of new subscribers Netflix might gain over a specific period. This insight can help Netflix make better business decisions, plan marketing strategies, and allocate resources effectively to sustain its growth.

This project involves collecting historical data on Netflix's subscription growth, cleaning and preprocessing it to remove inconsistencies, and analyzing patterns like seasonality or trends that might influence subscriber counts. Using this analysis, we can select the most appropriate time series forecasting models, such as ARIMA for statistical analysis or LSTM for more advanced deep learning-based predictions.

Netflix can estimate the expected number of new subscribers in a given time period and better understand the growth potential of their business. Below is the process we can follow to forecast subscription counts for Netflix:

- Gather historical Netflix subscription growth data
- Process and clean the data
- Explore and analyze time series patterns
- Choose a time series forecasting model (e.g., ARIMA, LSTM)
- Train the model using the training data
- Forecast future Netflix subscription counts

So the process for forecasting subscriptions for Netflix starts with collecting a dataset based on the historical growth of Netflix Subscribers. I found an ideal dataset for this task.
The dataset link is mentioned here: https://statso.io/forecasting-subscriptions-case-study/, you can download the dataset from here.

Netflix Subscriptions Forecasting using Python

We will start by importing the necessary Python libraries.

```python
# Importing Necessay Python libraries
import pandas as pd
import numpy as np
import matplotlib.pyplot as plt
import plotly.graph_objs as go
import plotly.express as px
import plotly.io as pio
pio.templates.default = "plotly_white"
from statsmodels.tsa.arima.model import ARIMA
from statsmodels.graphics.tsaplots import plot_acf, plot_pacf

# reading the data
data = pd.read_csv(r'C:\Users\sakhi\Downloads\Netflix-Subscriptions.csv')
print(data.head())
```

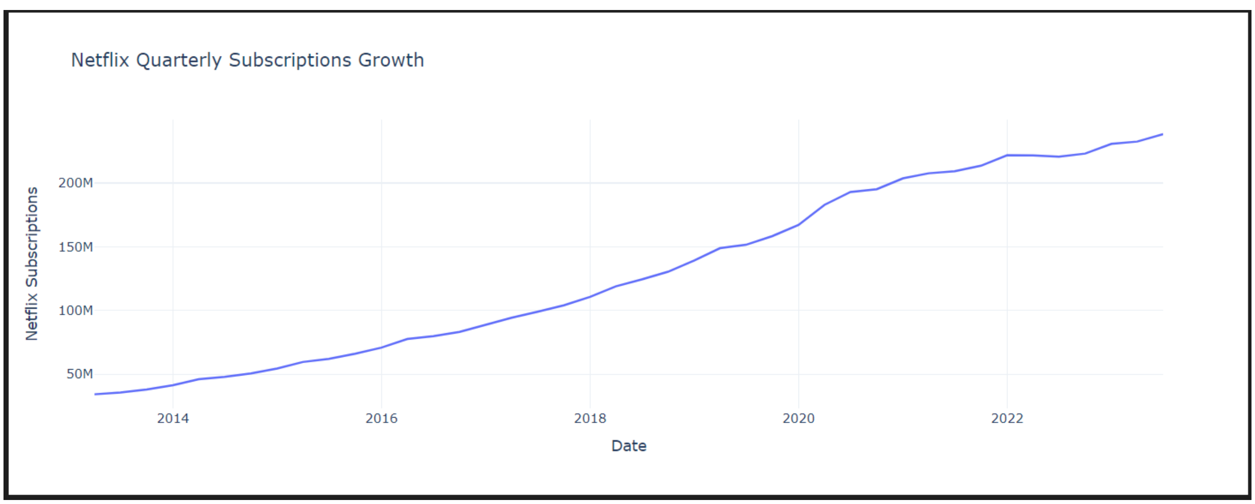
```
  Time Period  Subscribers
0  01/04/2013     34240000
1  01/07/2013     35640000
2  01/10/2013     38010000
3  01/01/2014     41430000
4  01/04/2014     46130000
```

The dataset contains subscription counts of Netflix at the start of each quarter from 2013 to 2023. Before moving forward, let's convert the period column into a DateTime format:

```python
data['Time Period'] = pd.to_datetime(data['Time Period'],
                                     format='%d/%m/%Y')
print(data.head())
```

```
  Time Period  Subscribers
0  2013-04-01     34240000
1  2013-07-01     35640000
2  2013-10-01     38010000
3  2014-01-01     41430000
4  2014-04-01     46130000
```

Now let's have a look at the quarterly subscription growth of Netflix:



The horizontal axis represents time, starting from 2014 and extending up to 2022. It is divided into yearly intervals, indicating the timeline over which Netflix subscriptions have been tracked.

The vertical axis represents the number of Netflix subscriptions in millions (M). Subscription numbers start from just under 50 million in 2014 and gradually increase, eventually surpassing 200 million by 2022.

Trend:

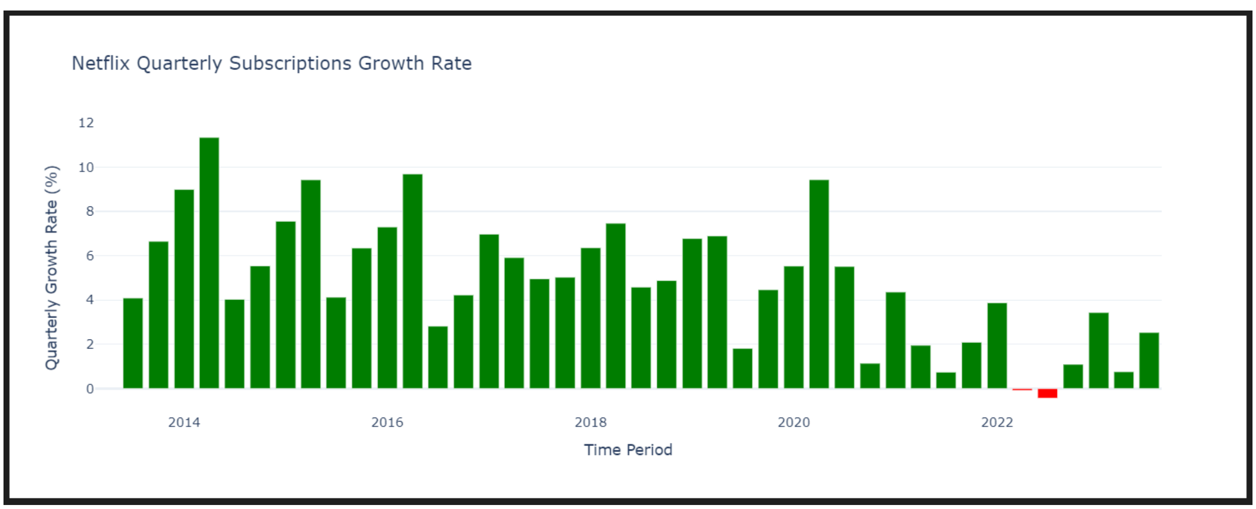
From 2014-2018 the graph shows a steady, consistent increase in subscription numbers. This suggests strong, continuous growth during this period, likely due to Netflix's global expansion and increased investment in original content.
From 2018-2020 the graph shows that the growth rate becomes slightly steeper, indicating an accelerated increase in subscriber counts. This period likely reflects the impact of more widespread adoption, increased content variety, and Netflix's presence in new markets.
From 2020-2022 the graph shows that growth continues, but at a slower rate compared to earlier years. This may signify market saturation in key regions or increased competition from other streaming platforms. Despite this, Netflix maintains a positive trajectory in subscription numbers.

Observations:

While there might be minor fluctuations in some quarters, the overall trend is upward, showing sustained growth. Netflix crossed significant milestones, such as 100M and 200M subscribers, within the observed timeframe.

In the above graph, we can see that the growth of Netflix subscribers is not seasonal. So we can use a forecasting technique like ARIMA in this dataset.

Now let’s have a look at the quarterly growth rate of subscribers at Netflix:



In the above bar graph, the horizontal axis represents the timeline, starting from 2014 and extending to 2022. Each bar corresponds to a quarter within this timeline, grouped by year.
The vertical axis shows the percentage growth rate of Netflix subscriptions for each quarter. Values range from 0% to 12%, indicating the speed of subscriber base expansion (or contraction if negative).

Observations:

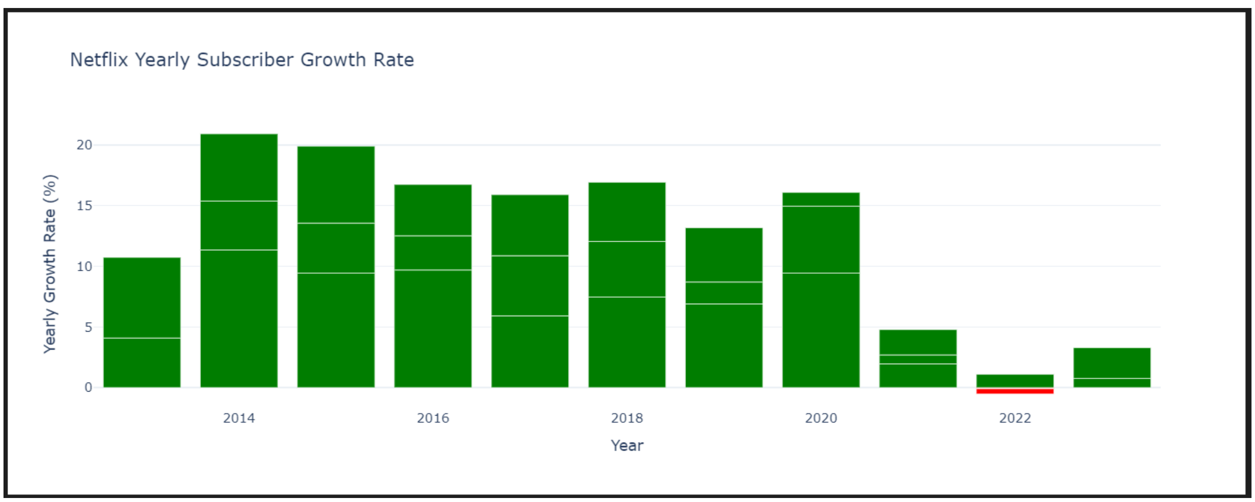
Early Growth (2014-2016): Netflix experienced some of its highest growth rates during this period, with a peak growth rate exceeding 10% in a single quarter. This reflects the initial adoption phase of Netflix's streaming service and its early global expansion.

Steady Growth (2016-2020): Growth rates stabilized to a more consistent range of around 5-7% per quarter. This period likely reflects Netflix's transition to a mature service, with steady subscriber additions driven by its growing content library and market penetration.

Pandemic Spike (2020): There were notable spikes in growth during early 2020, likely due to the COVID-19 pandemic, which increased demand for home entertainment. Some quarters show growth rates returning to the 8-10% range.

Decline and Negative Growth (2022): In 2022, there are instances of negative growth, represented by the red bars, indicating a loss in the subscriber base during those quarters. This likely reflects market saturation, competition from other streaming platforms, or changing consumer preferences.

Now let's have a look at the yearly growth rate:



This graph represents the yearly subscriber growth rate (%) of Netflix from 2014 to 2022. It shows how the rate of subscriber growth has changed over the years, illustrating Netflix's success in attracting new users.

Observations:

2014 to 2016: Strong Growth
Netflix experienced a high growth rate in these years, with yearly growth percentages close to or above 15%. This indicates that Netflix was expanding rapidly, likely due to global market expansion and the rising popularity of streaming services.

2017 to 2019: Steady Growth
The growth rate during this period remained consistently high, staying in the range of approximately 10-15%. This shows Netflix had established itself in many markets and was still gaining subscribers steadily.

2020: Slight Decline but Still Strong
Growth appears slightly lower than the previous years but still significant, around 10%. The pandemic likely contributed to sustained interest in streaming services.

2021 to 2022: Significant Drop
Growth rates dropped sharply, particularly in 2022. The red bar in 2022 highlights a possible decline in growth rate, indicating Netflix may have faced challenges in acquiring new subscribers or possibly losing subscribers in certain markets. Factors like market saturation, competition, or price increases may have contributed to this decline.

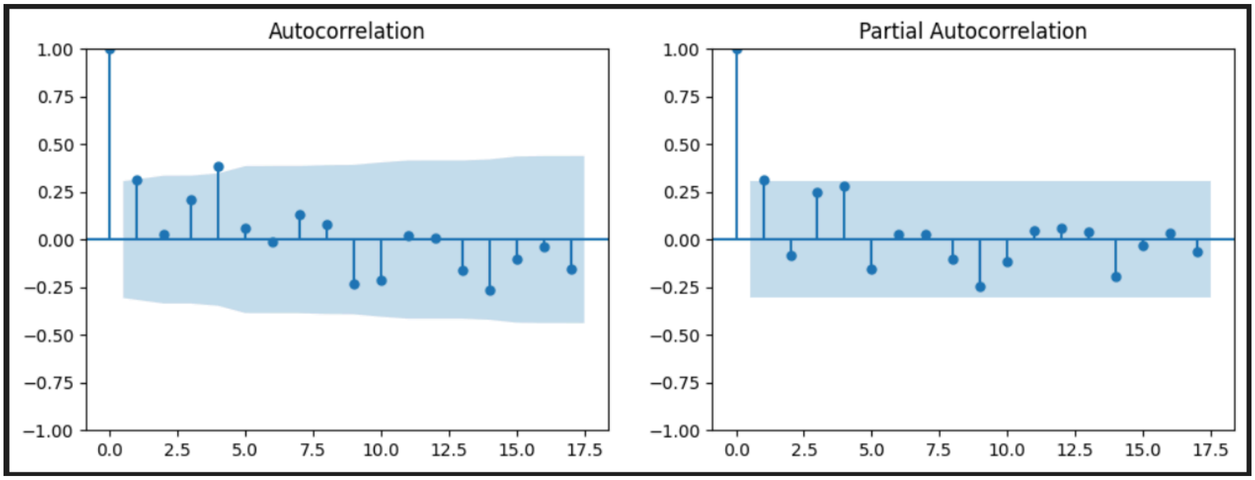
Using ARIMA for Forecasting Netflix Quarterly Subscriptions

Now let's get started with Time Series Forecasting using ARIMA to forecast the number of subscriptions to Netflix using Python. I will start by converting the data into a time series format.

```
time_series = data.set_index('Time Period')['Subscribers']
```

Here we are converting the original DataFrame into a time series format, where the Time Period column becomes the index, and the Subscribers column becomes the data.

Now let's find the value of p and q by plotting the ACF and PACF of differenced time series:



The image shows two plots side by side, representing the Autocorrelation Function (ACF) and the Partial Autocorrelation Function (PACF) of a time series. These are tools used to analyze the patterns and dependencies within time series data.

Autocorrelation Function (ACF) Plot (Left)

The x-axis represents the lag (time difference between observations), ranging from 0 to about 17. The y-axis represents the autocorrelation coefficient, ranging from -1 to 1.

The bar at lag 0 is always 1 (self-correlation). Significant bars (outside the shaded area) indicate a strong correlation at those lags. There are strong positive correlations at lag 1 and a few small positive correlations in early lags (e.g., 2, 3). After lag 5 or so, correlations weaken and mostly fall within the shaded area, indicating no significant autocorrelation at those lags. Shaded Area represents the confidence interval. Bars within this area are not statistically significant.

Partial Autocorrelation Function (PACF) Plot (Right)

This graph measures the correlation between a time series and its lagged values, controlling for the influence of other lags. The x-axis represents the lag, and the y-axis represents the partial autocorrelation coefficient. The bar at lag 0 is again 1. Significant partial autocorrelations are visible for the first few lags (e.g., lags 1, 2, and possibly 3). Beyond lag 3, most bars fall within the shaded confidence interval, indicating weak or no partial autocorrelation at higher lags. This pattern suggests that the first few lags capture most of the dependence structure in the data.

Based on the plots, we find that p=1 and q=1. The ACF plot cuts off at lag 1, indicating q=1, and the PACF plot also cuts off at lag 1, indicating p=1. As there is a linear trend in the subscription growth rate, we can set the value of d as 1 to remove the linear trend, making the time series stationary.

Now here’s how to use the ARIMA model on our data:

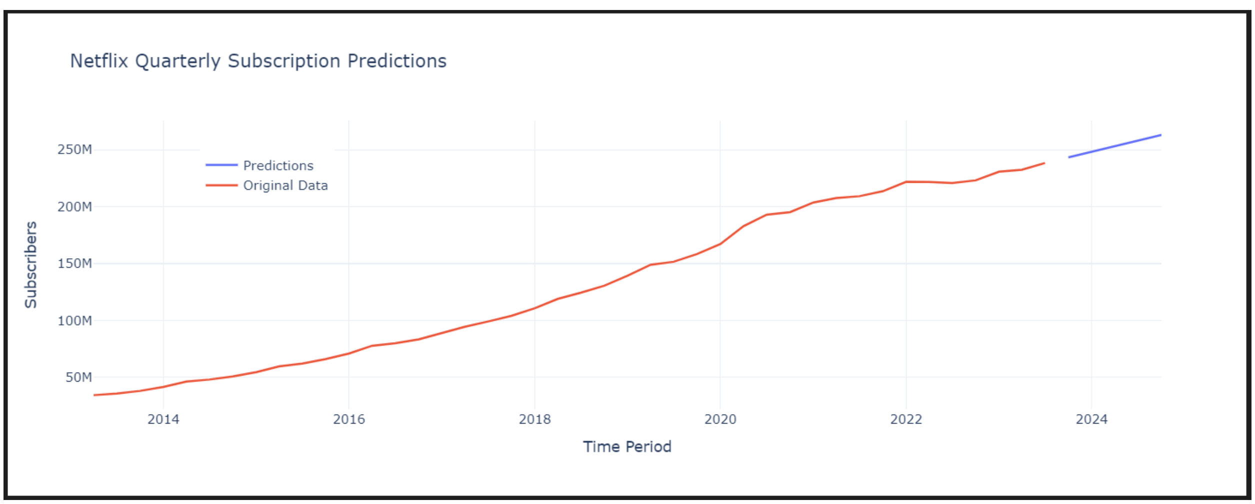
```
                               SARIMAX Results
==============================================================================
Dep. Variable:            Subscribers   No. Observations:                   42
Model:                 ARIMA(1, 1, 1)   Log Likelihood                -672.993
Date:                Wed, 22 Jan 2025   AIC                           1351.986
Time:                        18:47:47   BIC                           1357.127
Sample:                    04-01-2013   HQIC                          1353.858
                         - 07-01-2023
Covariance Type:                  opg
==============================================================================
                 coef    std err          z      P>|z|      [0.025      0.975]
------------------------------------------------------------------------------
ar.L1          0.9997      0.012     80.769      0.000       0.975       1.024
ma.L1         -0.9908      0.221     -4.476      0.000      -1.425      -0.557
sigma2      1.187e+13   1.57e-14   7.57e+26      0.000    1.19e+13    1.19e+13
===================================================================================
Ljung-Box (L1) (Q):                   3.96   Jarque-Bera (JB):                 4.62
Prob(Q):                              0.05   Prob(JB):                         0.10
Heteroskedasticity (H):               7.27   Skew:                             0.54
Prob(H) (two-sided):                  0.00   Kurtosis:                         4.23
===================================================================================
```

Now here’s how to make predictions using the trained model to forecast the number of subscribers for the next five quarters:

```
future_steps = 5
predictions = results.predict(len(time_series), len(time_series) + future_steps - 1)
predictions = predictions.astype(int)
print(predictions)
```

```
2023-10-01    243321465
2024-01-01    248251663
2024-04-01    253180592
2024-07-01    258108254
2024-10-01    263034649
Freq: QS-OCT, Name: predicted_mean, dtype: int64
```

Now let’s visualize the results of Netflix Subscriptions Forecasting for the next five quarters:



This graph highlights Netflix's subscription growth trajectory over a decade (2014 to 2024), emphasizing both historical data and future predictions.

The **red line** represents the actual subscriber numbers, showcasing significant growth from less than 50 million subscribers in 2014 to over 220 million by 2023. The steep incline in earlier years reflects Netflix's aggressive global expansion and increasing popularity of streaming services. However, the line flattens slightly in later years, particularly after 2020, signaling a slowdown likely due to market saturation, increased competition, or other challenges.

The **blue line**, which begins in 2023, shows predicted subscriber growth into 2024. The slope of the blue line suggests that while Netflix's subscriber base is expected to continue growing, the rate of increase is more gradual compared to earlier years. This could imply that future growth depends on strategic efforts such as entering new markets, investing in content, or adopting new business models.

Overall, the graph captures both Netflix's historical success and the tempered expectations for its subscriber growth moving forward.

Conclusion:

Using techniques like time series forecasting, Netflix can estimate the expected number of new subscribers in a given time period and better understand the growth potential of their business. It enhances operational efficiency, financial planning, and content strategy, ultimately contributing to their success and growth in the highly competitive streaming industry. I hope you liked this article on Netflix Subscriptions Forecasting using Python. Feel free to ask valuable questions in the comments section below.

Netflix's subscription growth showcases its journey from rapid expansion in its early years to becoming a mature global streaming leader. While strong early growth (2014-2016) and stable momentum (2016-2020) were bolstered by global expansion and a rich content library, the COVID-19 pandemic in 2020 provided a temporary surge. However, challenges like market saturation, rising competition, and occasional subscriber losses in 2022 highlight the need for strategic innovation. To sustain growth, Netflix must explore new markets, adopt flexible pricing models, invest in diverse and localized content, and enhance customer retention efforts to solidify its position in an increasingly competitive industry.