



School Of Sciences

Supply Chain Management (SCM)

Postgraduate Dissertation

Applying data analysis techniques in the over – the - top streaming industry: the case of the NETFLIX OTT platform.

Eleni Gkalimana

Supervisor: Nikolaos Thomaidis

Patras, Greece, January 2024

Theses / Dissertations remain the intellectual property of students (“authors/creators”), but in the context of open access policy they grant to the HOU a non-exclusive license to use the right of reproduction, customisation, public lending, presentation to an audience and digital dissemination thereof internationally, in electronic form and by any means for teaching and research purposes, for no fee and throughout the duration of intellectual property rights. Free access to the full text for studying and reading does not in any way mean that the author/creator shall allocate his/her intellectual property rights, nor shall he/she allow the reproduction, republication, copy, storage, sale, commercial use, transmission, distribution, publication, execution, downloading, uploading, translating, modifying in any way, of any part or summary of the dissertation, without the explicit prior written consent of the author/creator. Creators retain all their moral and property rights.



# Applying data analysis techniques in the over – the - top streaming industry: the case of the NETFLIX OTT platform.

Eleni Gkalimana

Supervising Committee

Supervisor:

Nikolaos Thomaidis

Associate Professor

Department of Financial and Management

Engineering

University of the Aegean

Patras, Greece, January 2024

*To my beloved husband, Nikos, who has been very supportive the entire time.*

## **Abstract**

NETFLIX is a popular over-the-top streaming platform which has evolved through the years from a DVD rental company to its today's form, as a streaming platform that provides subscription plans to its customers. The dissertation's objective is to distinguish possible patterns in the quarterly published revenues of NETFLIX Streaming Service and to apply different forecasting techniques in order to find the forecasting technique that best fits the data. Also, we will examine how COVID-19 pandemic, the end of the partnership with Disney and the appearance of Disney + platform affect the revenues of NETFLIX.

## **Keywords**

Over the top streaming platform, forecasting techniques, predict with accuracy

# Εφαρμογή τεχνικών ανάλυσης δεδομένων στις διαδικτυακές τηλεοπτικές υπηρεσίες: Η περίπτωση της διαδικτυακής πλατφόρμας Netflix

Ελένη Γκαλιμάννα

## Περίληψη

Το Netflix είναι μία δημοφιλής πλατφόρμα παροχής συνδρομητικών διαδικτυακών τηλεοπτικών υπηρεσιών που έχει εξελιχθεί από μία εταιρία ενοικίασης ταινιών στη ημερινή της μορφή. Ο σκοπός της διπλωματικής είναι μέσω τριμηνιαίων εσόδων που έχουν δημοσιευτεί από το Netflix την περίοδο 2012-2023, να διακριθούν πιθανά μοτίβα εσόδων, να εφαρμοστούν διαφορετικά μοντέλα πρόβλεψης και να βρεθεί εκείνο που ταιριάζει καλύτερα στα δεδομένα και που θα χρησιμοποιηθεί για ακριβή πρόβλεψη. Επίσης, θα εξετάσουμε η πανδημία Covid-19, το τέλος της συνεργασίας με τη Disney και η εμφάνιση της πλατφόρμας Disney+ επηρεάζουν τα έσοδα του Netflix.

## Λέξεις – Κλειδιά

Πλατφόρμα διαδικτυακών τηλεοπτικών υπηρεσιών, μέθοδοι πρόβλεψης, πρόβλεψη με ακρίβεια

## Table of Contents

Abstract .....	v
Περίληψη .....	vi
Table of Contents .....	vii
List of Figures .....	viii
List of Tables .....	ix
List of Abbreviations & Acronyms .....	x
1. Description of the company .....	1
1.1 Introduction .....	1-2
1.1.1 Subscription Plans .....	2-3
1.1.2 Research Questions .....	3
1.2 Literature Review .....	3-4
1.2.1 Research Data .....	4-5
2. Methodology .....	5
2.1 Exponential Weighted Moving Average (EWMA) .....	5-6
2.2 Regression Analysis .....	6
2.3 Performance Metrics .....	6
2.3.1 Mean Absolute Error .....	7
2.3.2 Mean Absolute Percentage Error .....	7
3 Empirical Study .....	7
3.1 Data Processing .....	7-8
3.2 Preliminary Analysis .....	8
3.2.1 Netflix Services .....	8-14
3.2.2 Disney+ .....	14-17
3.3 Variable Significance .....	17-19
3.4 Disney+ effect on Netflix .....	20-21
3.5 Covid-19 Pandemic .....	21-22
3.6 Revenues Forecasting .....	22
3.6.1 Regression Analysis .....	22-26
3.6.2 EWMA .....	26-28
3.6.3 SMA .....	28-30
3.6.4 Model Comparison .....	31
4 Evaluation, Conclusions and Limitations .....	32
4.1 Evaluation and Conclusions .....	32-34
4.2 Limitations .....	35
References .....	36

## List of Figures

Figure 1: Most popular Streaming platforms.....	2
Figure 2: Subscribers per Country .....	3
Figure 3: Accuracy and precision .....	6
Figure 4: Revenues of DVD Rental for the periods 2002-2023.....	9
Figure 5: Logarithm revenue differences DVD Rental.....	10
Figure 6: Box plots per quarter.....	11
Figure 7: Revenues Streaming service for the periods 2010-2022.....	12
Figure 8: Streaming logarithm differences.....	13
Figure 9: Streaming revenue box plot.....	13
Figure 10: Revenue streaming box plot per quarter.....	14
Figure 11: Disney+ revenues.....	15
Figure 12: Disney+ revenue logarithm differences.....	15
Figure 13: Disney + revenue box plot.....	16
Figure 14: Netflix and Disney+ revenues.....	16
Figure 15: Actual and estimated revenues with Regression.....	24
Figure 16: MAPE in relation to $\lambda$ .....	26
Figure 17: Actual and forecasted revenues using EWMA.....	28
Figure 18: Actual and forecasted revenues using SMA.....	30
Figure 19: Model comparison.....	34



## List of Tables

Table 1 Netflix Subscription programs.....	2
Table 2: DVD Rental Revenues.....	8-9
Table 3: Streaming Revenues.....	11-12
Table 4: Disney+ revenues per quarter.....	14
Table 5: Regression analysis result for Streaming service revenues.....	19
Table 6: Regression analysis for Disney+ variable.....	21
Table 7: Regression analysis for Covid-19.....	23
Table 8: Regression analysis for Netflix streaming revenues.....	24
Table 9: Actual and estimated revenues with Regression.....	24
Table 10: Calculated error with regression.....	25
Table 11: Calculated error with EWMA.....	27
Table 12: Error calculation using SMA.....	29
Table 13: Model comparison.....	31

## **List of Abbreviations & Acronyms**

OTT	Over The Top
EWMA	Exponential Weighted Moving Average
SMA	Simple Moving Average
MAE	Mean Absolute Error
MAPE	Mean Absolute Percentage Error

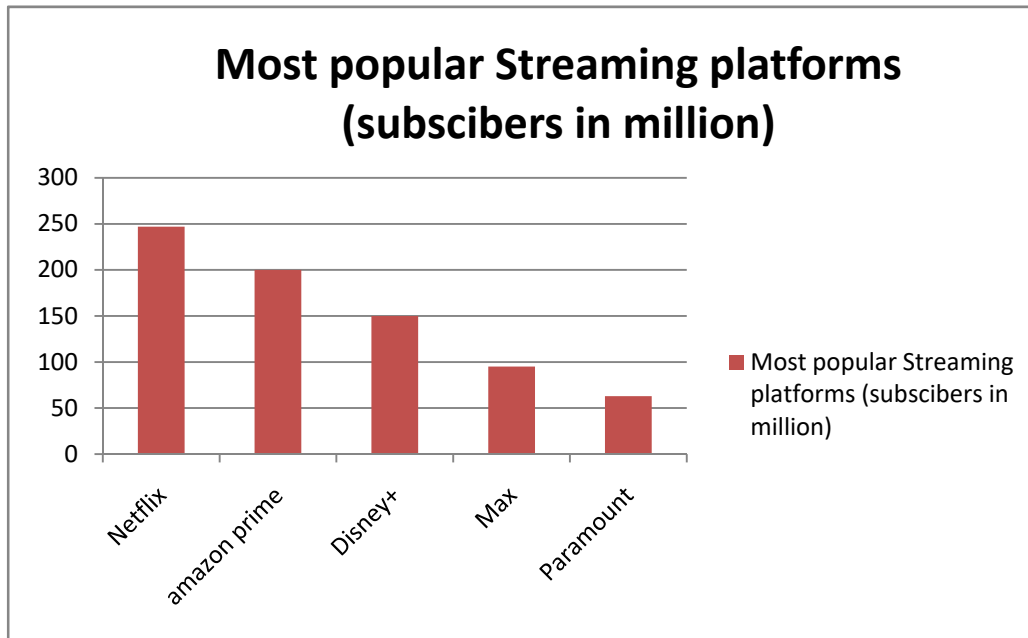
# 1 Description of the company

Netflix is a successful American over the top streaming platform, which offers a wide collection of series, movies, documentaries, shows, etc. to its subscribers. When Netflix first entered the market in 1997, it delivered its DVD content to its American subscribers via mail. Since 2007, Netflix started to convert to its today form, from a rental service to a streaming platform and started to expand worldwide. In 2013 Netflix started to create its own content (movies, series) which was available only to its platform. Even though this move had a lot of risk, Netflix gained more popularity and differentiate itself from other similar platforms. Another innovation that Netflix initiated is to provide personal recommendation to its users, based on their style and their country. This created a friendly environment for their subscribers which raise the satisfaction level and also succeeds not only to maintain their subscribers but also to appeal new ones. Netflix today is the 24<sup>th</sup> most visited website in the world. It is also available in all countries except for Russia and China.

## 1.1 Introduction

The development of the technology has changed the way people entertain themselves. Streaming platforms have replaced the DVD rental stores, since anyone from his home can have access to a wide variety of movies and series for a relatively small fare. The most popular streaming platforms based on the number of subscribers are: Netflix , prime video, Disney +, etc. [2] Netflix seem to win and maintain the competition among other platforms since it offers a wide variety of movies, series and documentaries, original content of movies and series which are also available in many languages and an easy at use environment at which subscribers can watch instantly or download and watch later the show of their choice. Netflix offers 3 different subscription plans and the price range based on ads, library and the devices that can be used simultaneously. Disney plus offers a monthly or a yearly subscription plan with full accessibility to all its content in both programs. At this, point it should be mentioned that Disney was an old partner of Netflix. In 2012, Netflix and Disney signed a valuable partnership in which Disney gave Netflix exclusive rights and access to its content. This partnership ended in 2016, since Disney intended to launch its own streaming platform, which today is the well known Disney+. In 2021, Netflix launched a new service to its platform which is included so far to the subscription fare. This service allows subscribers to play online games through Netflix platform. This is an important move for Netflix and if succeeds, increases by far the competition. For the time being this service is only available in tablets and phones, but it is planned to be also available in TV and computers which will increase by far the competition.

In the figure below, we see the 5 most popular streaming platforms based on number of subscribers. We notice that Netflix is on top with almost 250 million subscribers followed by Amazon Prime (200 million), Disney+ (150 million), Max (95 million) and Paramount (63million)



**Figure 1: Most popular Streaming platforms** (Πηγή: <https://www.digitaltrends.com/home-theater/most-popular-streaming-services-by-subscribers/>)

### 1.1.1 Subscription plans

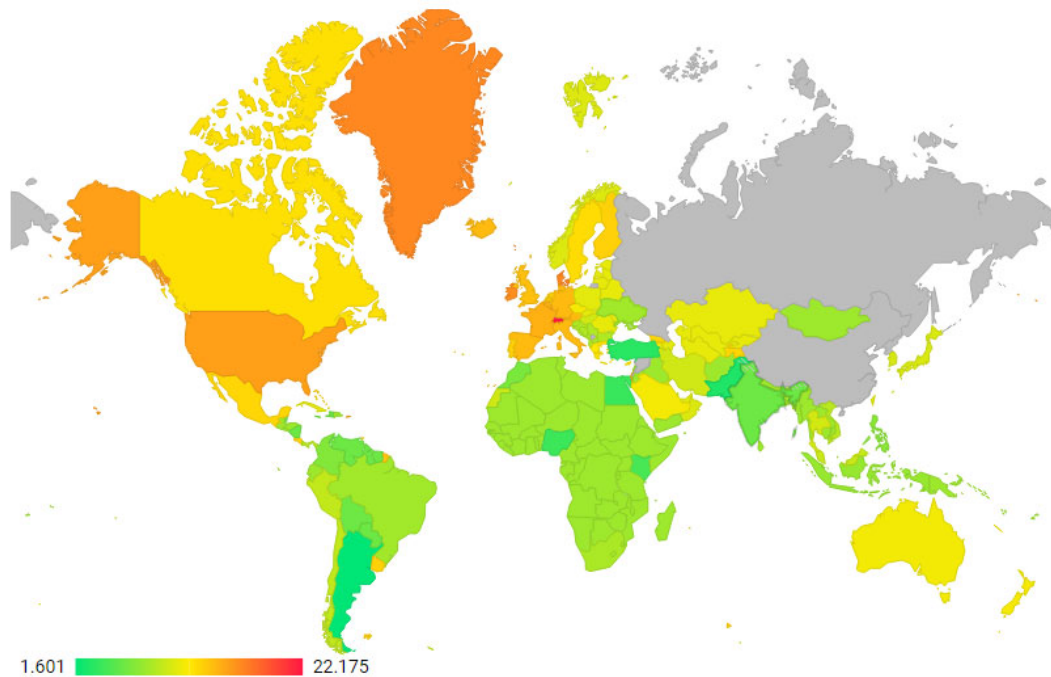
Netflix subscription plans differ per country and so is the price. In the next figure, we observe the 4 different available subscription plans that Netflix offer to its subscribers. So there is the Basic plan, Standard, Premium and mobile. The differences between them are showed in [Table 1]. Most of Netflix subscribers choose the Standard plan since there is balance between price and quality of service.

	Basic	Standard	Premium	Mobile
Simultaneous Tracking Tags	1	2	4	1
Number of devices where content can be downloaded	1	2	4	1
Movies, TV shows and mobile games without restrictions	✓	✓	✓	✓
Watching on phone or tablet	✓	✓	✓	✓
Watching on laptop and TV	✓	✓	✓	✗
HDDefinition	✗	✓	✓	✗
Ultra HD resolution	✗	✗	✓	✗

**Table 1: Netflix subscription programs** (Πηγή: <https://vpnwiki.com/en/netflix-subscription-prices>)

As we mentioned above, price can range per country. The countries which English is spoken have a higher fare in comparison with other countries.

Country	Basic with ads	Basic	Standard	Premium
Australia	\$4.58	\$7.19	\$11.12	\$15.05
Canada	\$4.41	\$7.36	\$12.14	\$15.46
Ireland	–	\$9.65	\$16.08	\$22.52
New Zealand	–	\$7.87	\$11.20	\$15.13
United Kingdom	\$6.20	\$8.68	\$13.65	\$19.86
USA	\$6.99	\$9.99	\$15.49	\$19.99
Global average	–	\$6.43	\$10.09	\$13.03



**Figure 2: Subscribers per Country** (Πηγή: <https://vpnwiki.com/en/netflix-subscription-prices> , accessed on 01/2024)

In the bottom panel of Figure 2, the countries with lower fare are colored with green and with red those with the highest fare. The fare refers to the Standard subscription plan which as we see ranges between 1,601 and 22,175 dollars.

### 1.1.2 Research Questions

In this study, the following questions are going to be addressed and analyzed:

- What are the main statistical features of revenues?
- Is there any trend observed in the data?
- Has the appearance of Disney + and COVID-19 had any significant effect on the revenues of NETFLIX?
- Which of the employed forecasting methods fits best to the data and at with what accuracy rate can revenues be predicted?

By analyzing those questions, we will find out possible trend patterns the behavior of revenues from the beginning of Netflix till today and if the company is becoming more profitable year by year. After analyzing the data, we will examine if Netflix revenues have affected by one of its biggest competitor (Disney+) and the appearance of Covid19 pandemic. Finally, we will implement different forecasting techniques in order to find out which one suits better the data.

## 1.2 Literature Review

The purpose of this research is to examine the statistical features of the most successful streaming platform worldwide, Netflix. By using its past revenues, from the beginning of the company till now, we will track possible patterns, seasonality and also we will have a clear view about how Netflix beats competition and remains to the top all these years. We will also examine if several factors affect Netflix popularity. The first one will be one of its biggest competitors, Disney+ who used to be partners with Netflix and the second one will be Covid-19 pandemic.

Furthermore, we will apply two different forecasting techniques: Regression analysis and Exponential Weighted Moving Average. For the Regression analysis, we will first check which variables are statistically important in order to end up to the final regression model. Finally, we will find out if these models have a good fit to the data with the use of accuracy measures.

### 1.2.1 Research Data

In our research, we will use the revenues of Netflix for the period 2002-2023. Those data are published quarterly to Netflix official site. Unfortunately, there are not separate revenue data for its subscription plan, however all subscription plans refer to the same service and their difference is mostly the simultaneously watch at different devices. We will divide the research in two parts. From 2002-2023 is the period where Netflix provide

the DVD rental as its basic service. The second part is from 2010 till today where Netflix introduced its popular streaming service.

For Disney+ analysis, we will use revenue data which are published to the Walt Disney Company. Disney+ entered the market in 2020 and offers two subscription plans, one monthly and one yearly. Both subscription plans offer the same services.

## 2 Methodology

At first, by using the data mentioned before, we will calculate basic statistical measures (mean and standard deviation) in order to examine how dispersed are the data and how the mean of the revenues change every year. Moreover, we will represent the data in graphs to help us understand visually how the data behave, if there are any trends like upward if the data are rising, seasonality ( if there is a specific time of the year which our data rise or fall) or if the data doesn't have a specific pattern. Furthermore, we will apply several forecasting techniques in order to find the one that bests fits the data and measure the level of accuracy for each one of them. Finally, qualitative techniques will be also used in order to better interpret the data, the results of our calculations and if there is a reasonable explanation between them.

### 2.1 Exponential Weighted Moving Average

Exponential Weighted Moving Average (EWMA) is a quantitative forecasting technique used in time series data, which is using most recent data in order to forecast future ones. The key feature in this technique is that you can choose how much weight is going to be given to the most recent data depending on its importance.

The model that is used in EWMA is the following:

$$\hat{Y}_{t+1} = \lambda * \hat{Y}_t + (1 - \lambda) * Y_t$$

Where,

$\hat{Y}_t$ , is the forecasted value in time t

$Y_t$ , is the revenue at time t

$\lambda$ , weight parameter

So, as we notice from the mathematical model, the actual revenues and the forecasted from the previous period are used for the calculation of the forecasting revenues. The parameter  $\lambda$  can take values between 0 and 1. The closest the  $\lambda$  is to 1, less weight is given

to most recent revenues. In our case, we will pick the value of the  $\lambda$  which minimize the error.

## 2.2 Regression Analysis

Regression analysis is measure which is used to examine if there is a relationship between a dependent variable and the independent variables. So, after this relationship is confirmed, we can calculate a formula which helps us predict the outcome of future variables. In our research, there is one independent variable, so we will use the formula for simple linear regression. The model for simple linear regression is:

$$\hat{Y} = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \dots + \beta_n X_n + et, \text{ where}$$

$\beta_0$  : Y intercept

$\beta_{1..n}$ : Coefficient of  $X_{1..n}$

$X_{1..n}$ : Independent variable

$\hat{Y}$ : Dependent variable

$et$ : error term

With Regression analysis, we can add as many independent variables we think are important to our model and the result of the analysis shows us which of them are more important than others. So, it helps us understand better how our model works and at the same time to end up in a model which best fits the data. In our study we will use multiple regression analysis since there is more than one independent variables. The difference between simple regression and multiple regression is the number of the independent variable. Also, regression analysis gives us the possibility to use a dummy variable. Dummy variable can take only two values 0 and 1 which indicates if the variable is present or not in one category.

## 2.3 Performance Metrics

Performance Metrics are used in forecasting techniques like Regression, EWMA, SMA in order to check if the model is working well for our data. In other words, Performance Metrics calculate the level of accuracy of our model. In our study, we will use two metrics: MAE and MAPE.

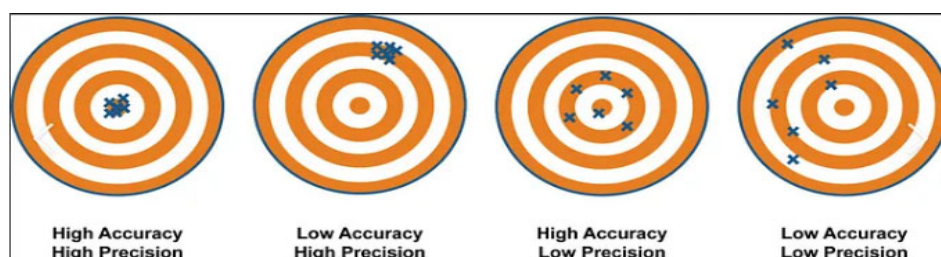


Image showing four target boards and the difference between accuracy and precision



**Figure 3: Accuracy and precision** (Πηγή: <https://88guru.com/library/chemistry/accuracy-and-precision-difference>)

### 2.3.1 Mean Absolute Error (MAE)

In order to calculate MAE, we subtract the forecasted value from the actual and we calculate the average of those differences:

$$MAE = \frac{1}{n} \sum_{i=1}^n (Y_i - \hat{Y}_i)$$

A low value for MAE, indicates a better model performance.

### 2.3.2 Mean Absolute Percentage Error (MAPE)

MAPE is the percentage difference between the actual value and the predicted one. The formula to calculate it is the following:

$$MAPE = \frac{1}{n} \sum_{i=1}^n \frac{Y_i - \hat{Y}_i}{Y_i}$$

The lower the MAPE, the better the performance for our model. MAPE is easier to be used since it shows percentages, while MAE shows the mean error and sometimes can be deceptive since you have to compare it with the range of the actual values.

## 3 Empirical study

In our study, we will mostly use quantitative techniques to represent our data and interpret them like descriptive statistics, graphs and forecasting models. However, it is important to use also qualitative techniques to better understand and interpret them. For example, a sudden raise of revenues could be explained by one phenomenon. Since the study is talking about Streaming Platforms, it is reasonable to believe that a factor that force people to stay home like Covid 19 pandemic would cause an increase in revenues, which would be proved with quantitative techniques. Moreover, a new service or a lower price which could beat the competition in comparison with other companies is also a reason to see a raise in revenues. For example, the appearance of Disney +, Amazon Prime could influence Netflix revenues or the creation of a new service like Netflix originals gives Netflix competitive advantage.

### 3.1 Data processing

For our study, we used the revenues published online by Netflix for the periods 2002-2023. The main problems we faced with those data are the following: the revenues are available quarterly which means that the observations are fewer and there are not separate data for each subscription plan. However, the period of the data availability is long enough to conduct a reliable study. Moreover, there are separate data for the two Netflix services: DVD rental and Streaming. More specifically, there are 36 observations for DVD rental, since from 2020 till 2023 there is only one yearly data available because the service was going to be shut down at late 2023 and for Streaming there are 51 observations available. For Disney+ there are only 15 observations available since the platform joined the market at 2020. For Regression analysis, we converted the data in logarithms in order to normalize them and eliminate the errors.

### 3.2 Preliminary Analysis

At this point of our study, we will use our data to create plots in order to be able to make qualitative assessment. Also, we will examine and interpret the statistical features of our sample by using descriptive statistics and at the same time we will observe if there is any trend, seasonality or any outliers.

#### 3.2.1 Netflix Services

	q1	q2	q3	q4	mean	standard deviation
2002	30,527	36,360	40,731	45,188	38,202	6,258291167
2003	55,669	63,187	72,202	81,185	68,061	11,05599079
2004	100,370	120,321	141,644	143,893	126,557	17,69746543
2005	154,140	164,497	174,317	195,040	171,999	17,43066923
2006	224,126	239,351	255,950	277,233	249,165	22,78241256
2007	305,320	303,693	293,792	303,355	301,540	4,534576551
2008	326,183	337,614	341,269	359,595	341,165	13,86537222
2009	394,098	408,509	423,120	444,542	417,567	21,53545726
2010	493,655	519,819	553,219	592,305	539,750	36,96429281
2011	706,274	769,714	799,152	875,575	787,679	70,25394839
2012	319,701	291,485	271,318	254,368	284,218	28,10269903
2013	243,293	232,381	221,865	213,258	227,699	11,26578888
2014	204,354	194,721	186,597	179,489	191,290	10,70401337
2015	173,200	164,018	157,524	150,995	161,434	9,475875206

2016	144,747	138,732	132,375	126,413	135,567	6,860652388
2017	120,394	114,737	110,214	105,152	112,624	6,492984746
2018	98,751	92,904	88,777	85,157	91,397	5,835362449
2019	33,958	30,347	27,793	31,109	30,802	2,197480987
2020				20,000	20,000	
2021				12,100	12,100	
2022				13,000	13,000	
2023		6,600			6,600	

Table 2: DVD Rental Revenues

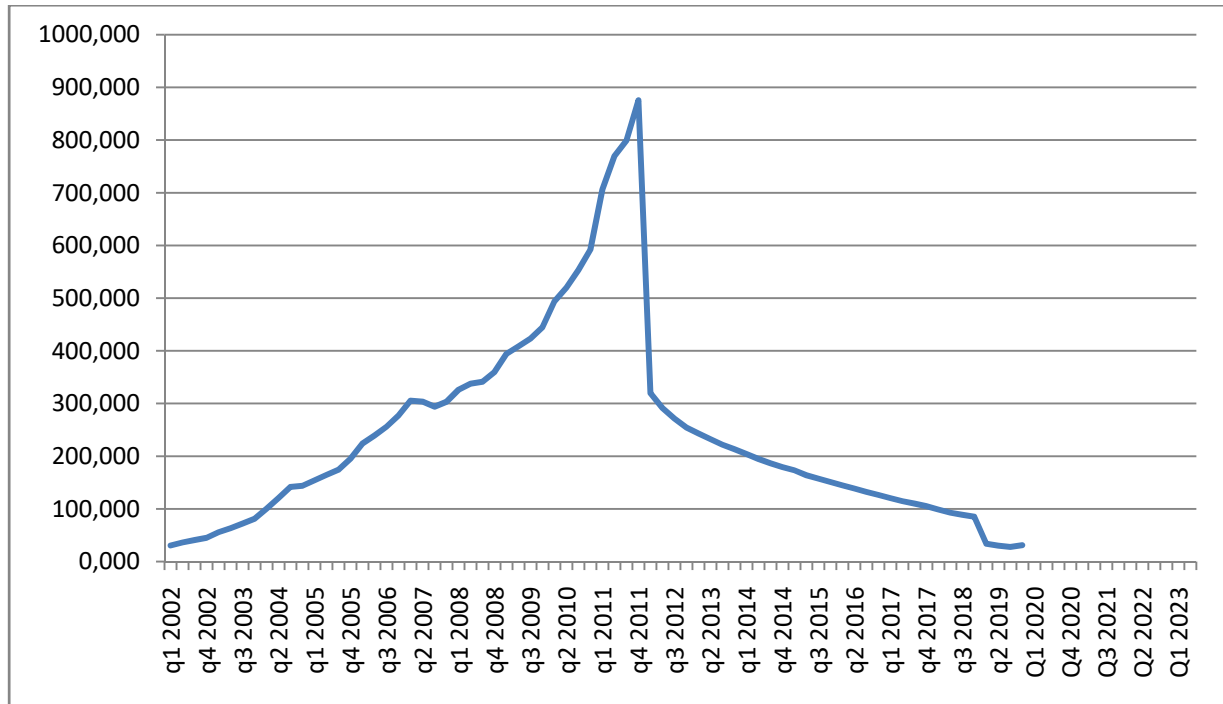


Figure 4: Revenues of DVD Rental for the periods 2002-2023

As we notice from the data and the plot, the revenues started to fall radically since late 2010 where Netflix streaming service was initiated. Also, since the technology is developing rapidly year by year, it is natural that costumers preferred the streaming service instead of the dvd-rental. So, there seems to exist an upward trend till 2011, from 2011-2012 there is an abrupt fall to revenues and from 2012 till 2023 there is a downward trend. The time series could not be characterized as stationary since the mean is changing all the time. Also, there is no sign of seasonality or cyclical pattern.

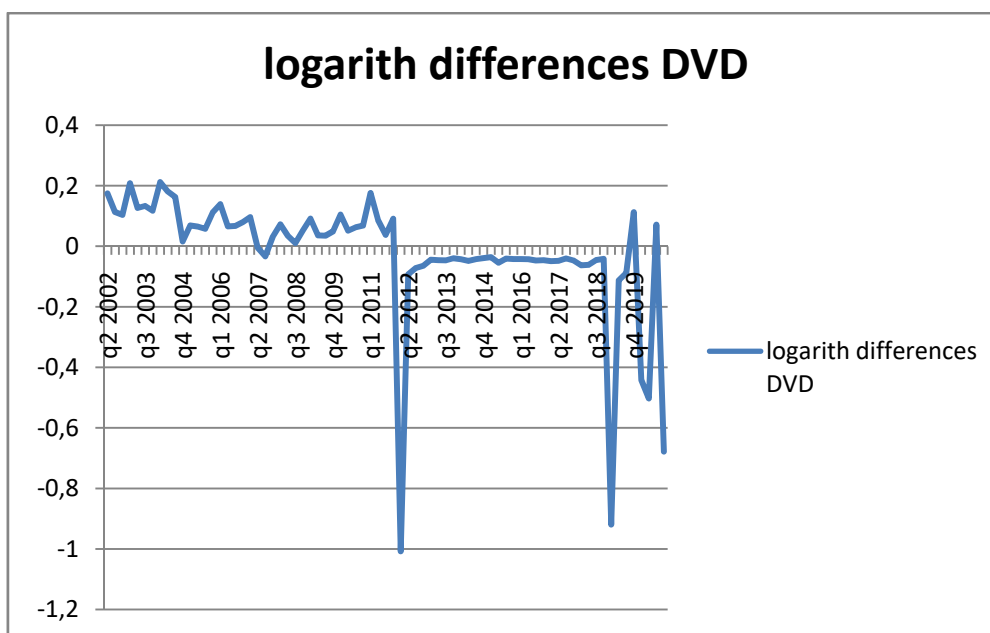
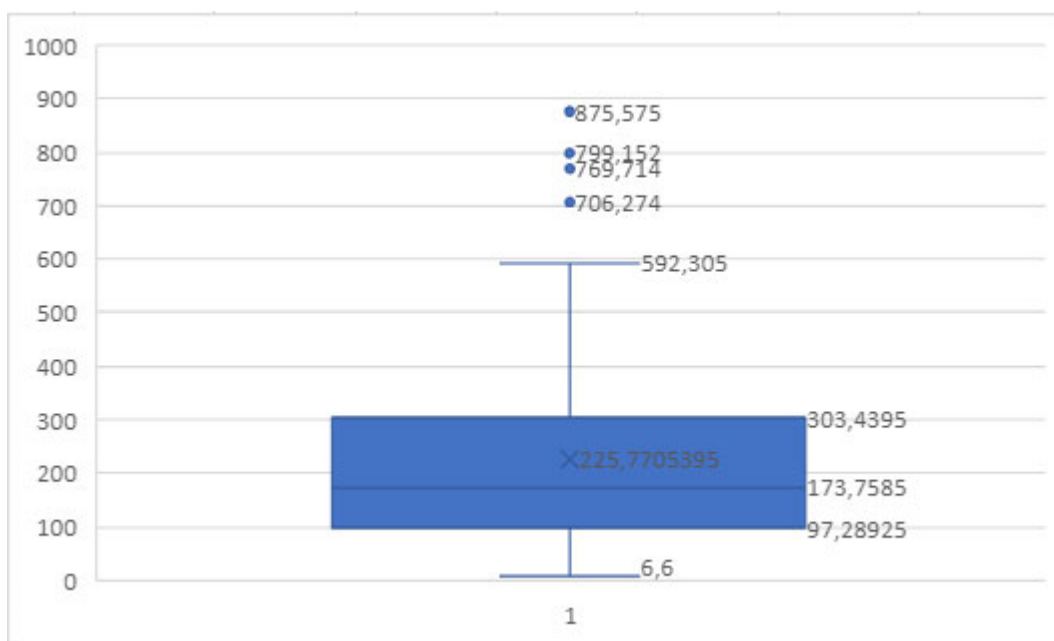


Figure 5: Logarithm revenue differences DVD Rental

As we notice from the graph, the revenues at the beginning were rising either with a bigger rhythm or with a smaller one. However those percentages range between 0 and 0,22. We notice that in year 2012 there was a sudden fall in dvd–rental revenues and there is a possible reasonable explanation for that. Netflix introduced the streaming service which was included free of charge the first years in dvd rental subscription. After 2012, we notice a big fall in DVD rental revenues which seem to stabilize the following years. It seems that DVD rental service is heading to its end.

### Boxplot



As we notice from the Box plot, there outliers in our data which are: 706,274, 769,714, 799,152 and 875,575. Also the data seems to have asymmetry since the line of the median is not in the middle and also there seem to be more data out of the box and in the limits to the upper fence.

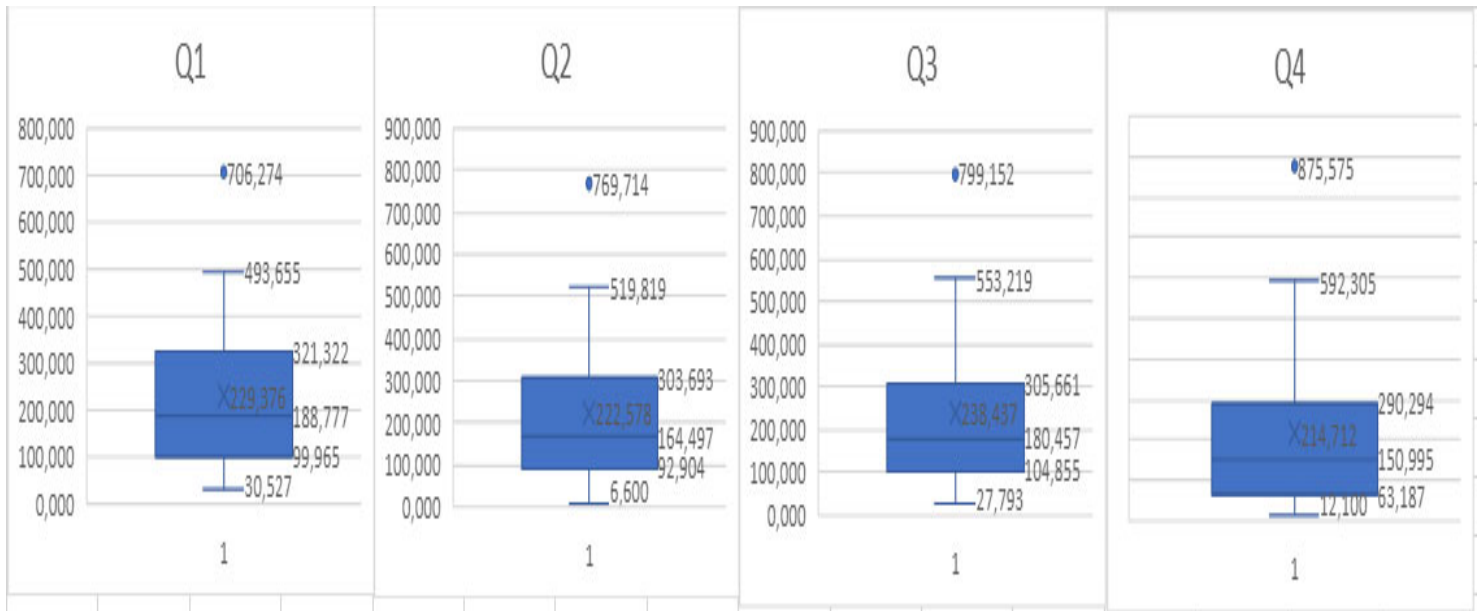


Figure 6: Boxplots per quarter

As we notice from the four Box Plots, there is not a sign of seasonality. In the fourth quarter seem that the revenues are lower than the other three. However the differences are not that remarkable to make an assumption that there is a seasonal pattern.

### Streaming Service

Since 2007, Netflix offer its customers a new service which corresponds with the today development of technology. The subscribers could pick and watch instantly the movie or series of their choice. At first, in order to make this service more approachable to public, it was included for free as an extra feature to the DVD subscription. So, from late 2010 this new service cut off from the DVD rental subscription and the customers could pick between them. DVD rental service and streaming were available only in US at first. The streaming service started gradually to be expanded worldwide.

	Q1	Q2	Q3	Q4	mean	standard deviation
2010				3,617	3,617	
2011	12,279	18,896	22,687	505,322	139,796	243,7219521

2012	550,09	597,678	633,771	690,871	618,1025	59,39589447
2013	780,668	836,991	884,134	961,972	865,9413	76,73010109
2014	1065,735	1145,686	1222,835	1305,239	1184,874	102,7271003
2015	1399,929	1480,676	1580,831	1672,338	1533,444	118,5295857
2016	1812,989	1966,472	2157,813	2351,128	2072,101	233,451728
2017	2516,241	2670,727	2874,645	3180,603	2810,554	287,0630886
2018	3602,105	3814,366	3910,597	4101,684	3857,188	207,7914627
2019	4440,304	4846,916	5173,028	5398,982	4964,808	416,6795733
2020	5767,690	6148,290	6435,640	6644,440	6249,015	379,9162638
2021	7163,280	7341,780	7483,470	7709,320	7424,463	230,7078777
2022	7867,770	7970,140	7925,590	7852,050	7903,888	54,32001619
2023	8161,499	8187,299			8174,399	18,24335495

Table 3: Streaming Revenues

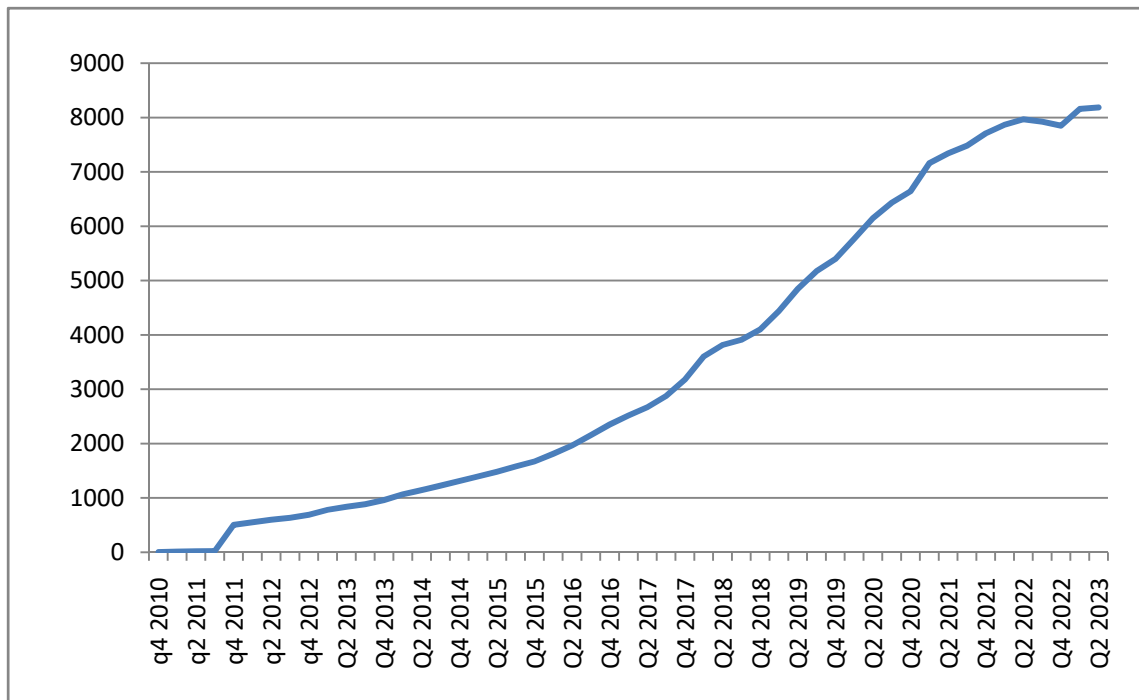


Figure 7: Revenues streaming service for the periods 2010-2022

As we notice from the graph, the revenues are rising so we can assume that there is an upward trend. Also, since the mean and variance are rising too, the time series can't be characterized as stationary. Also, for the same reason the time series hasn't the tendency to go back to the mean since both the mean and variance are rising.

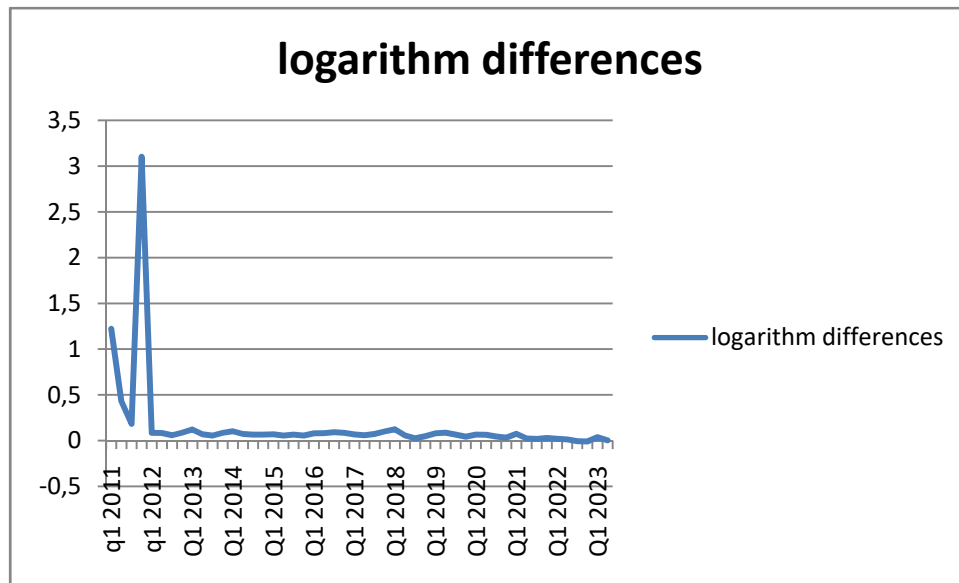


Figure 8: Streaming logarithm differences

As we notice from the graph, there was a huge acceptance at the beginning of the streaming service and as the time passes the raise seems to stabilize.

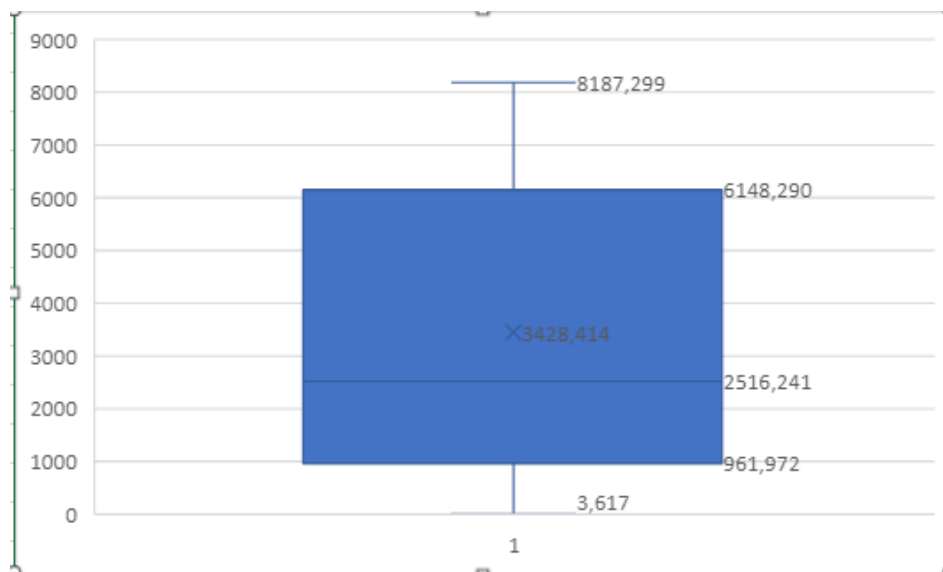


Figure 9: Streaming revenue box plot

It is also confirmed from the Box Plot that there aren't any outliers traced. Also the data seems to have asymmetry since the line of the median is not in the middle and also there seem to be more data out of the box and in the limits to the upper fence.

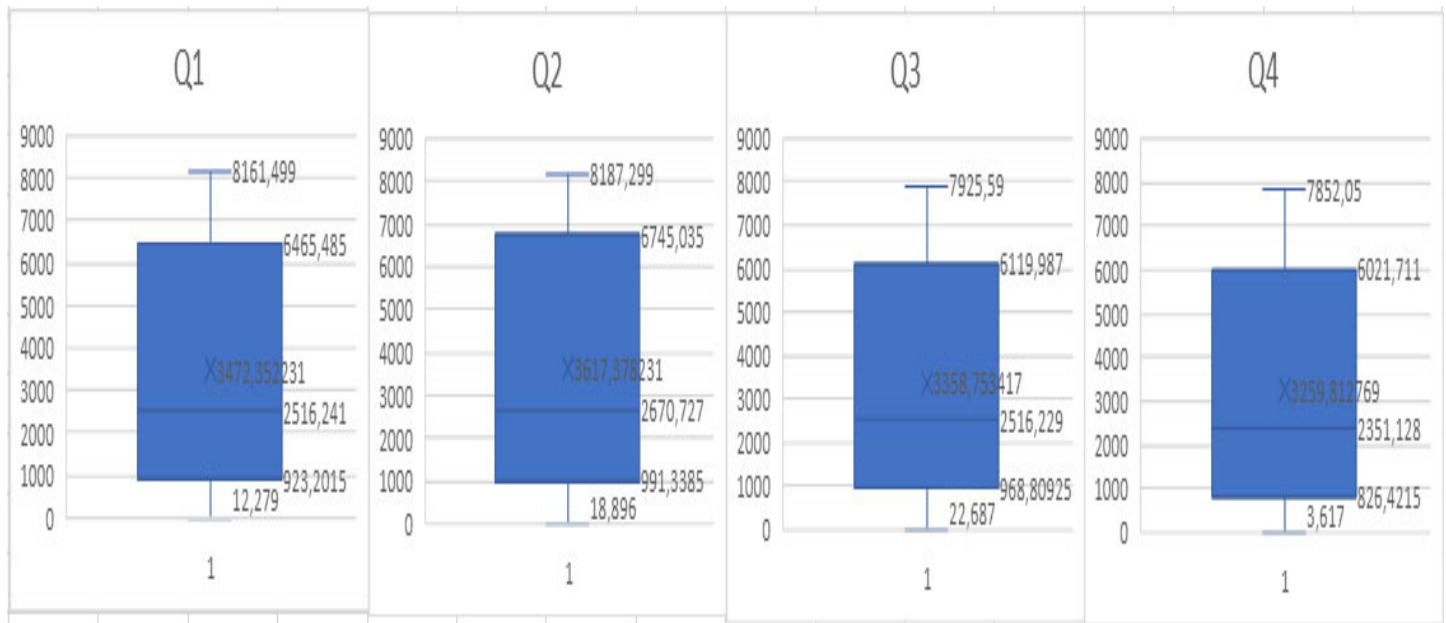


Figure 10: Revenue streaming box plot per quarter

The four Box Plots contain data from Streaming service revenues per quarter. The conclusion from the diagrams regarding any seasonal patterns is that there is no sign of seasonality as the values are in the same levels. The lower values that appear in Box plot were the revenues when Streaming service first entered the market. Q1, Q2, Q3 fluctuate in the same levels with Q2 to have bigger revenues than the other three quarters. However the differences are not that high to make an assumption.

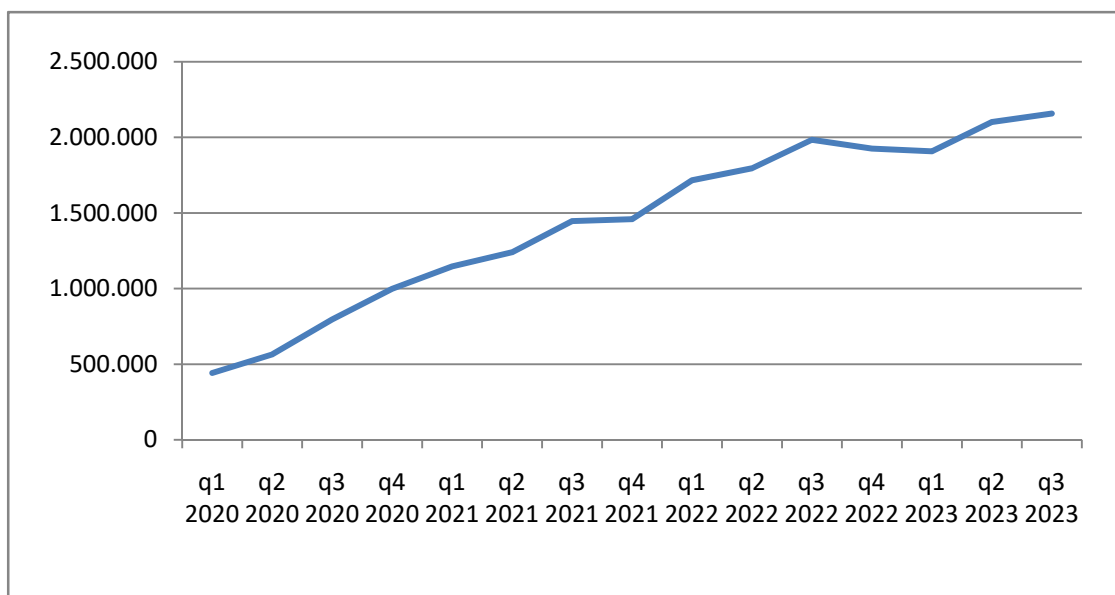
### 3.2.2 Disney+

At this part, we will analyze Disney + data (revenues), one of Netflix biggest competitors.

	1st quarter	2nd quarter	3rd quarter	4th quarter	mean	standard deviation
2020	442.000	565000	796.000	999.000	700.500	247254,9292
2021	1.147.000	1.240.000	1.447.000	1.459.000	1.323.250	154635,8626
2022	1.717.000	1.796.000	1.984.000	1.926.000	1.855.750	121387,465
2023	1.908.000	2.102.000	2.158.000		2.056.000	131194,5121

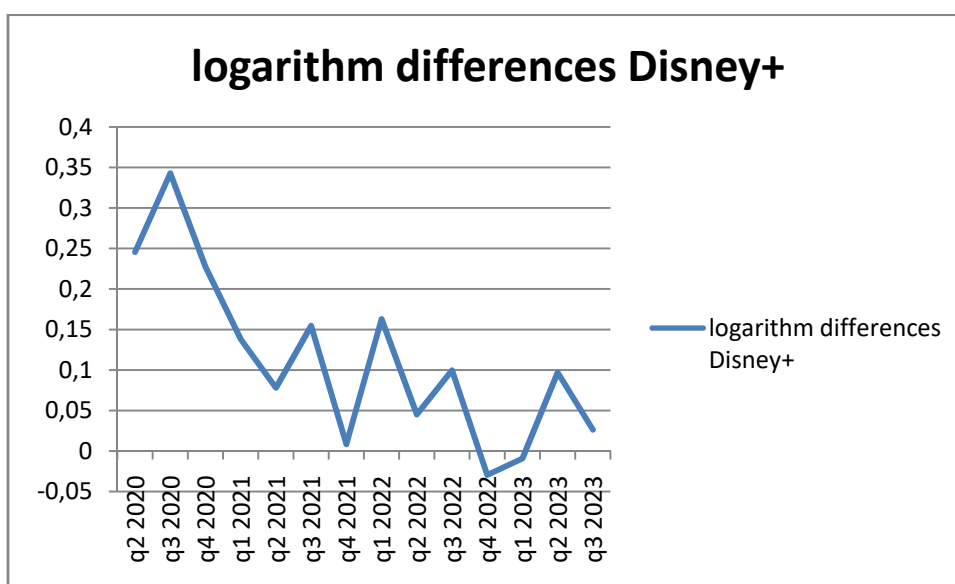
Table 4: Disney+ revenues per quarter





**Figure 11: Disney+ revenues**

The mean and variance are rising, so we can assume that the time series can't be characterized as stationary. Also, for the same reasons, the time series doesn't tend to reverse to the mean. Moreover, from the graph, we can assume that there is an upward trend.



**Figure 12: Disney+ revenue logarithm differences**

As it seems from the graph and the table, Disney + at the beginning has an impressive raise in its revenues. However, it seems that the following years, this raise falls, but still its profits are increasing with a lower rhythm.

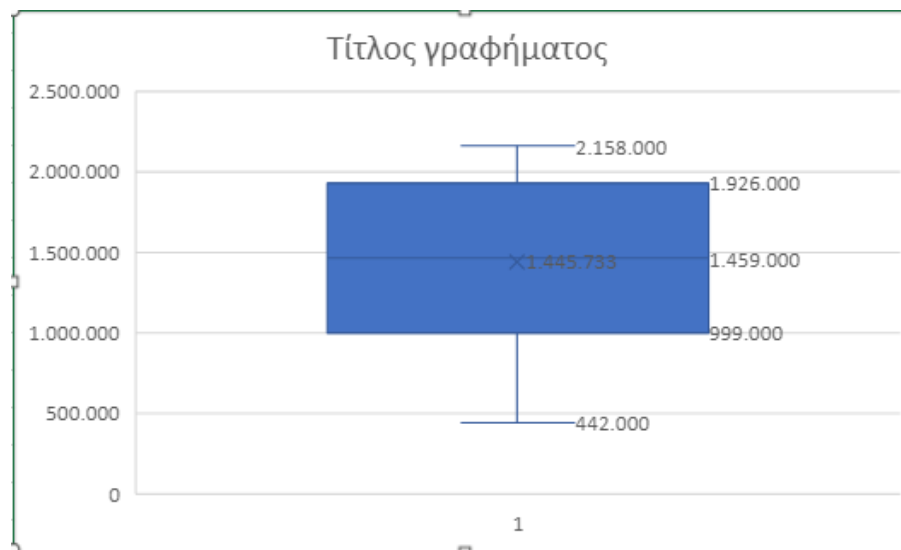


Figure 13: Disney + revenue box plot

From Box Plot, it seems that there aren't any outliers traced and as a result it won't affect the regression model. Also the data seems to have a small asymmetry since the line of the median is not exactly in the middle and also there seem to be more data out of the box and in the limits to the lower fence.

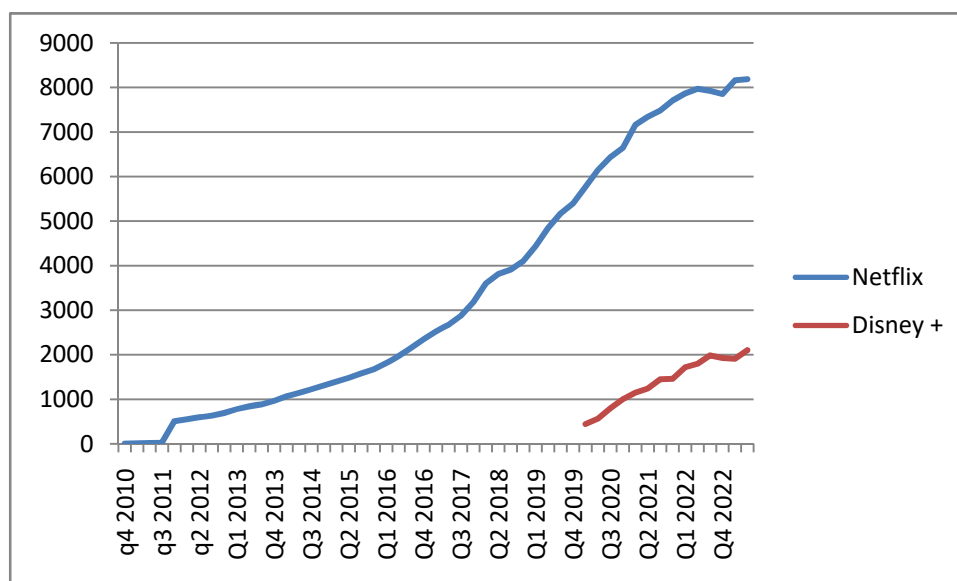


Figure 14: Netflix and Disney+ revenues

As we notice from the graph and the table, Disney+ as enters the market, had a huge raise of its revenues. The following years however, this percentage is diminished rapidly. On the other side, Netflix percentage of raise diminishes year by year. It is obvious, that Netflix

revenues are by far bigger than Disney +. However, Netflix has been in the market since 1997 and gradually transitioned to its current form since 2007, so we can't say if this situation is permanent for the next 15 years.

### 3.3 Variable Significance

We will run hypothesis tests by using regression in order to check which variables that we think are important to our model, are actually important. To begin with, the past observations is a variable worth taking into consideration since by observing how past observations fluctuate; it might be an indicator for future ones. Moreover, another variable that is worth examined is the Quarter. With the quarters, we can check if the revenues rise or fall at a specific quarter. For example, Q1 and Q4 are quarters which include winter months at which people usually prefer staying home and enjoy a movie. Finally, we will also use a time variable in order to see if revenues rise or fall as the time passes.

Hypothesis testing is a statistical method to support a certain hypothesis. In our case, we will check if our variables are statistically important by forming a hypothesis based on probabilities. We will use two hypotheses: null and alternative and we will try to reject null hypothesis. This probability is the P value which measures how probable or not our statement is true or false. In our study, we will use a level of significance ( $\alpha$ )  $\alpha=0,05$ , which means that we are 95% confident that our variable is significant.

We will convert the revenue data into logarithms in order to normalize them.

The regression model including the variables explained before will take the following form:

$$\ln Y_t = \beta_0 + \zeta_1 \ln Y_{t-1} + K_1 Q_{t1} + K_2 Q_{t2} + K_3 Q_{t3} + \gamma t + et$$

where:

$\beta_0$ : Coefficient of  $\ln Y_t$

$\zeta_1$ : Coefficient of  $\ln Y_{t-1,t-2,t-3}$

$\ln Y_{t-1}$ : the logarithm of past observations

$K_{1,2,3}$ : Coefficient of  $Q_{t1,t2,t3}$

$Q_{t1,t2,t3}$ : Seasonal dummy for 1<sup>st</sup>, 2<sup>nd</sup> and 3<sup>rd</sup> quarter (takes the value 1 or 0, which indicates the quarter)

$\gamma$ : slope of the linear trend

t: time trend variable

*et*: error term

We will run hypothesis tests for the variables:  $\ln Y_{t-1}$ ,  $Q_{t1}$ ,  $Q_{t2}$ ,  $Q_{t3}$ ,  $t$

Null Hypothesis:  $H_0: \zeta_1 = 0$  ( the variable  $\ln Y_{t-1}$  is insignificant)

Alternative Hypothesis:  $H_1: \zeta_1 \neq 0$  (the variable  $\ln Y_{t-1}$  is significant)

Null Hypothesis:  $H_0: K_i$  ( the variable  $Q_{ti}$  is insignificant)

Alternative Hypothesis:  $H_1: K_i \neq 0$  (the variable  $Q_{ti}$  is significant)

Null Hypothesis:  $H_0: \gamma = 0$  ( the variable  $t$  is insignificant)

Alternative Hypothesis:  $H_1: \gamma \neq 0$  (the variable  $t$  is significant)

At  $\alpha=0,05\%$  level of significance we get:

SUMMARY OUTPUT								
<i>Regression Statistics</i>								
Multiple R	0,975156							
R Square	0,950929							
Adjusted R Square	0,945353							
Standard Error	0,34804							
Observations	50							
ANOVA								
	<i>Df</i>	<i>SS</i>	<i>MS</i>	<i>F</i>	<i>Significance F</i>			
Regression	5	103,2852	20,65703	170,5333505	1,24E-27			
Residual	44	5,329805	0,121132					
Total	49	108,615						
	<i>Coefficients</i>	<i>Standard Error</i>	<i>t Stat</i>	<i>P-value</i>	<i>Lower 95%</i>	<i>Upper 95%</i>	<i>Lower 95,0%</i>	<i>Upper 95,0%</i>
Intercept	1,964631	0,295551	6,647343	0,000000038	1,368987	2,560276	1,368987	2,560276
$\ln Y_{t-1}$	0,72384	0,053589	13,50714	<b>0,000000000</b>	0,615838	0,831842	0,615838	0,831842
$Q_{t1}$	-0,18152	0,139456	-1,30161	0,199823191	-0,46257	0,099538	-0,46257	0,099538
$Q_{t2}$	-0,23268	0,139334	-1,66995	0,102028936	-0,51349	0,048128	-0,51349	0,048128
$Q_{t3}$	-0,24964	0,142142	-1,75629	0,085997416	-0,53611	0,036826	-0,53611	0,036826
T	0,016007	0,006322	2,532059	<b>0,014984425</b>	0,003266	0,028747	0,003266	0,028747

**Table 5: Regression analysis result for Streaming service revenues**

The variables which are statistical important based on the results are  $\ln Y_{t-1}$  and  $t$

Since  $p(\gamma) = 3 \cdot 10^{-6} < 0,05$ , we reject null hypothesis

$p(\zeta_1) = 3 \cdot 10^{-16} < 0,05$ , we reject null hypothesis

The regression model will take the following form:

$$\ln Y_t = 1,964 + 0,723 \ln Y_{t-1} + 0,016t \text{ (M1)}$$

The results from the regression analysis indicate that the time variable and past observations affect most the data.

### 3.4 Disney+ effect on Netflix

In this point, we will check if Disney+ revenues affect Netflix revenues. The revenues data have been analyzed in the previous paragraphs.

In our previous regression model (M1) we will add an independent variable  $\ln D_t$

$$\ln Y_t = 1,964 + 0,723 \ln Y_{t-1} + 0,016t + s \ln D_t$$

Where  $s$  is the slope of the  $\ln D_t$  variable and  $\ln D_t$  is the variable for Disney+ logarithm revenues.

Let's assume that:

$H_0: s=0$  (the variable is insignificant)

$H_1: s \neq 0$  (the variable is significant)

At  $\alpha=0,05\%$  level of significance we get:

SUMMARY OUTPUT								
<i>Regression Analysis</i>								
Multiple R	0,991487							
R Square	0,983046							
Adjusted R Square	0,97796							
Standard Error	0,01671							
Observations	14							
ANOVA								
	<i>Df</i>	<i>SS</i>	<i>MS</i>	<i>F</i>	<i>significance F</i>			

Regression	3	0,161896	0,053965	193,2772753	3,76E-09			
Residual	10	0,002792	0,000279					
Total	13	0,164688						
	<i>Coefficients</i>	<i>Standard Error</i>	<i>t Stat</i>	<i>P-value</i>	<i>Lower 95%</i>	<i>Upper 95%</i>	<i>Lower 95,0%</i>	<i>Upper 95,0%</i>
Intercept	3,214834	2,107155	1,525675	0,158075202	-1,4802	7,909867	-1,4802	7,909867
T	0,586874	0,293974	1,996348	0,073835882	-0,06814	1,241889	0,06814	1,241889
$\ln Y_{t-1}$	0,000374	0,003361	0,11132	0,913565354	-0,00711	0,007862	0,00711	0,007862
$\ln D_t$	0,066374	0,07447	0,89128	<b>0,39371857</b>	-0,09956	0,232304	0,09956	0,232304

**Table 6: Regression analysis for Disney+ variable**

Since  $P(s)=0,39>0,05$  we can't reject null hypothesis (the variable is insignificant). So the appearance of Disney+ did not had a significant effect on Netflix revenues. It should be mentioned that the revenues data for Disney+ are for the periods 2020 – 2023 since the platform is new to the market. For that reason, we should interpret the result of the previous analysis that Disney+ did not had an effect on Netflix revenues but we are not sure if this situation continues in the future. I also picked this specific streaming platform since as I mentioned before, Disney was a former associate to Netflix and gave full access to its content.

### 3.5 Covid-19 Pandemic

Covid-19 is a pandemic which affected financially many companies in different sectors. Regarding the streaming industry, Covid-19 has mostly affected positively many streaming platforms since people had to stay at home and this kind of entertainment helped them spend time at home. So, we will examine if Covid-19 had an effect in Netflix revenues.

Let's assume that Covid 19 outbreak was between late 2019 and late 2022.

We will check if Covid-19 had an effect on Netflix revenues.

We will apply a Dummy variable for Covid-19, which will take the number 1 for the periods that Covid 19 was in outbreak and the number 0 for the rest of the periods.

$$Y_t = 1,964 + 0,723\ln Y_{t-1} + 0,016t + s_2 Ct$$

Where  $Ct$  is the dummy variable for Covid19

And  $s_2$  is the slope.

Let's assume that

$H_0: s_2 = 0$  (the variable is insignificant)

H1:  $s_2 \neq 0$  (the variable is significant)

At 0,05 % significant level we get:

SUMMARY OUTPUT								
Regression Analysis								
Multiple R	0,972911							
R Square	0,946556							
Adjusted R Square	0,94307							
Standard Error	0,355235							
Observations	50							
ANOVA								
	<i>df</i>	<i>SS</i>	<i>MS</i>	<i>F</i>	<i>significance F</i>			
Regression	3	102,8102	34,27005	271,5711046	2,96E-29			
Residual	46	5,804823	0,126192					
Total	49	108,615						
	<i>Coefficients</i>	<i>Standard Error</i>	<i>t Stat</i>	<i>P-value</i>	<i>Lower 95%</i>	<i>Upper 95%</i>	<i>Lower 95,0%</i>	<i>Upper 95,0%</i>
intercept	1,794768	0,286536	6,263678	0,00000012	1,218002	2,371535	1,218002	2,371535
t	0,72331	0,055907	12,9378	0,00000000	0,610776	0,835844	0,610776	0,835844
$\ln Y_{t-1}$	0,016326	0,007842	2,081868	0,04294582	0,000541	0,032112	0,000541	0,032112
Ct	-0,00752	0,159417	-0,04719	<b>0,96256471</b>	-0,32841	0,313368	-0,32841	0,313368

Table 7: Regression analysis for Covid-19

Since  $p(s_2)=0,96>0,05$ , we can't reject null hypothesis. So Covid-19 didn't affect Netflix revenues.

## 3.6 Revenues Forecasting

### 3.6.1 Regression analysis

We will use a sample which consists of 60% of the total observations which in our case is:  $60\% \cdot 51 = 31$  observations. We will perform regression analysis in order to build our model.

$$Y_t = \beta_0 + \zeta_1 Y_{t-1} + K_1 Q_{t1} + K_2 Q_{t2} + K_3 Q_{t3} + \gamma t + e_t,$$

Where,

$\beta_0$ : Coefficient of  $Y_t$

$\zeta_1$ : Coefficient of  $Y_{t-1}$

$Y_{t-1}$ : The logarithm of past observations

$K_{1,2,3}$ : Coefficient of  $Q_{1,2,3}$

$Q_{t1,t2,t3}$ : Seasonal dummy for 1<sup>st</sup>, 2<sup>nd</sup> and 3<sup>rd</sup> quarter (takes the value 1 or 0, which indicates the quarter)

$\gamma$ :slope of the linear trend

$t$ , time trend variable

$e_t$ , is the error term

We will run hypothesis tests for the variables:  $Y_{t-1}$ ,  $Q_{t1}$ ,  $Q_{t2}$ ,  $Q_{t3}$ ,  $t$

Null Hypothesis:  $H_0: \zeta_1=0$  (the variable  $Y_{t-1}$  is insignificant)

Alternative Hypothesis:  $H_1: \zeta_1 \neq 0$  (the variables  $Y_{t-1}$  is significant)

Null Hypothesis:  $H_0: K_i=0$  ( the variable  $Q_{ti}$  is insignificant)

Alternative Hypothesis:  $H_1: K_i \neq 0$  (the variables  $Q_{ti}$  is significant)

Null Hypothesis:  $H_0: \gamma =0$  ( the variable  $t$  is insignificant)

Alternative Hypothesis:  $H_1: \gamma \neq 0$  (the variable  $t$  is significant)

SUMMARY OUTPUT								
Regression Statistics								
Multiple R	0,996456							
R Square	0,992924							
Adjusted R Square	0,991449							
Standard Error	95,83805							
Observations	30							
ANOVA								
	<i>Df</i>	<i>SS</i>	<i>MS</i>	<i>F</i>	<i>Significance F</i>			
Regression	5	30930867	6186173	673,5132	5,66E-25			
Residual	24	220438,4	9184,932					
Total	29	31151305						



	Coefficients	Standard Error	t Stat	P-value	Lower 95%	Upper 95%	Lower 95,0%	Upper 95,0%
intercept	118,6031	58,25149	2,036053	0,052926	-1,62202	238,8283	-1,62202	238,8283
Yt-1	1,102646	0,084343	13,07336	<b>2,08E-12</b>	0,928571	1,276722	0,928571	1,276722
Q1	-54,6779	50,38034	-1,0853	0,288569	-158,658	49,30201	-158,658	49,30201
Q2	-98,5979	50,76173	-1,94237	0,063916	-203,365	6,16918	-203,365	6,16918
Q3	-84,9763	51,27986	-1,65711	0,110514	-190,813	20,86007	-190,813	20,86007
T	-4,4817	9,291791	-0,48233	0,633942	-23,659	14,69562	-23,659	14,69562

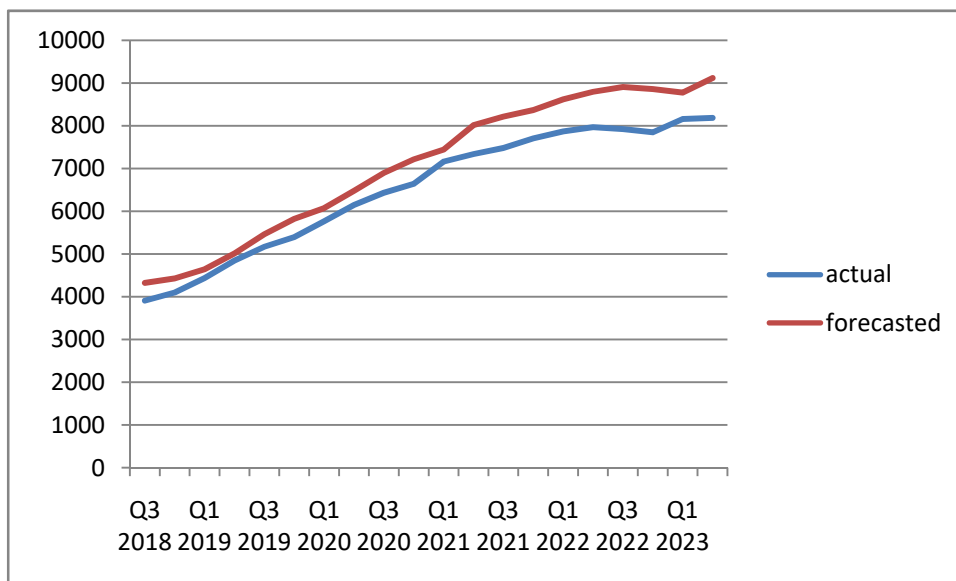
**Table 8: Regression analysis for Netflix streaming revenues**

Since  $p(\zeta_1) = 2,08E-12 < 0,05$ , we reject null hypothesis. The variable is significant for our model for  $\alpha = 0,05$ .

So, the regression model would be:

$$Y_t = 118,6031 + 1,102646Y_{t-1}(M2)$$

We will use the excel function SUMPRODUCT in order to calculate the estimated revenues that emerge from M2 model.



**Figure 15: Actual and estimated revenues with Regression**

quarter	actual	forecasted
Q3 2018	3910,597	4324,5
Q4 2018	4101,684	4430,608
Q1 2019	4440,304	4641,31
Q2 2019	4846,916	5014,688
Q3 2019	5173,028	5463,037
Q4 2019	5398,982	5822,623
Q1 2020	5767,69	6071,771
Q2 2020	6148,29	6478,325
Q3 2020	6435,64	6897,992
Q4 2020	6644,44	7214,838
Q1 2021	7163,28	7445,07
Q2 2021	7341,78	8017,167
Q3 2021	7483,47	8213,99
Q4 2021	7709,32	8370,224
Q1 2022	7867,77	8619,256
Q2 2022	7970,14	8793,971
Q3 2022	7925,59	8906,849
Q4 2022	7852,05	8857,726
Q1 2023	8161,499	8776,637
Q2 2023	8187,299	9117,85

**Table 9: Actual and estimated revenues with Regression**

In [Figure 15] and [Table 9], the actual and estimated revenues are presented. As we notice, it seems that the model overestimates the actual revenues.

At this point, we will make an accuracy evaluation of our model by using accuracy measures we analyze before: MAE and MAPE.

At first we will calculate the error to see at which periods the revenues overestimate or underestimate the actual ones.

The error  $e_t$  is calculated as:  $t = D_t - F_t$ , where  $D_t$  are the actual revenues and  $F_t$  are the forecasted ones.

actual	forecasted	error
3910,597	4324,499739	-413,903
4101,684	4430,608496	-328,924
4440,304	4641,309871	-201,006
4846,916	5014,687965	-167,772
5173,028	5463,037186	-290,009
5398,982	5822,623379	-423,641
5767,690	6071,770723	-304,081
6148,290	6478,325239	-330,035
6435,640	6897,992424	-462,352
6644,440	7214,837841	-570,398
7163,280	7445,07039	-281,79
7341,780	8017,167401	-675,387
7483,470	8213,989768	-730,52
7709,320	8370,223723	-660,904
7867,770	8619,256392	-751,486
7970,140	8793,9707	-823,831
7925,590	8906,848603	-981,259
7852,050	8857,725709	-1005,68
8161,499	8776,6371	-615,138
8187,299	9117,849898	-930,551

**Table 10: Calculated error with regression**

As we notice in the table above, the red color indicates the overestimated revenues. We notice that the model overestimates the actual revenues the whole time and the error increases over time.

The formula for calculating MAE is the average of the absolute calculated different between actual and forecasted revenues. So, by using excel we calculate  $MAE=547,4332$  which is a good one since the value is relative small regarding the revenues values.

The formula for calculating MAPE is  $APE = \frac{1}{n} \sum_{i=1}^n \frac{Y_i - \hat{Y}_i}{Y_i}$ . So by using excel we estimate MAPE to be 8,09% which is a really good one since it is close to 0 and under 10%.

### 3.6.2 EWMA

In order to perform EWMA analysis I will use a sample which contains the 60% of the data, which is 31 observations. I will find the  $\lambda$  which give us the smaller MAPE. After that, I would perform EWMA on the rest 40% of the data and determine if the model is a good fit.

In order to calculate the forecasted revenues I use the following type:

$F_{t+1} = \lambda * F_t + (1 - \lambda) * D_t$ , where  $F_{t+1}$  is the forecast for period  $t+1$ ,  $D_t$  are the sales during period  $t$

I started with  $\lambda=0,01$  which gave  $MAPE=13,39\%$ . After I try with  $\lambda=0,02$  which gave  $MAPE=13,51\%$ . As we notice the MAPE raises. Then I use  $\lambda=0,1$  and  $\lambda=0,2$  which gave  $MAPE=14,47$  and  $MAPE=15,88\%$  respectively. So, the  $\lambda$  which give us the lowest MAPE is 0,01. This result is also confirmed with the use of solver tool of excel. The lower the value of the  $\lambda$ , the more weight it gives to the most recent observations. In our case, the value of the  $\lambda$  is low which means that the model gets affected by sudden changes.

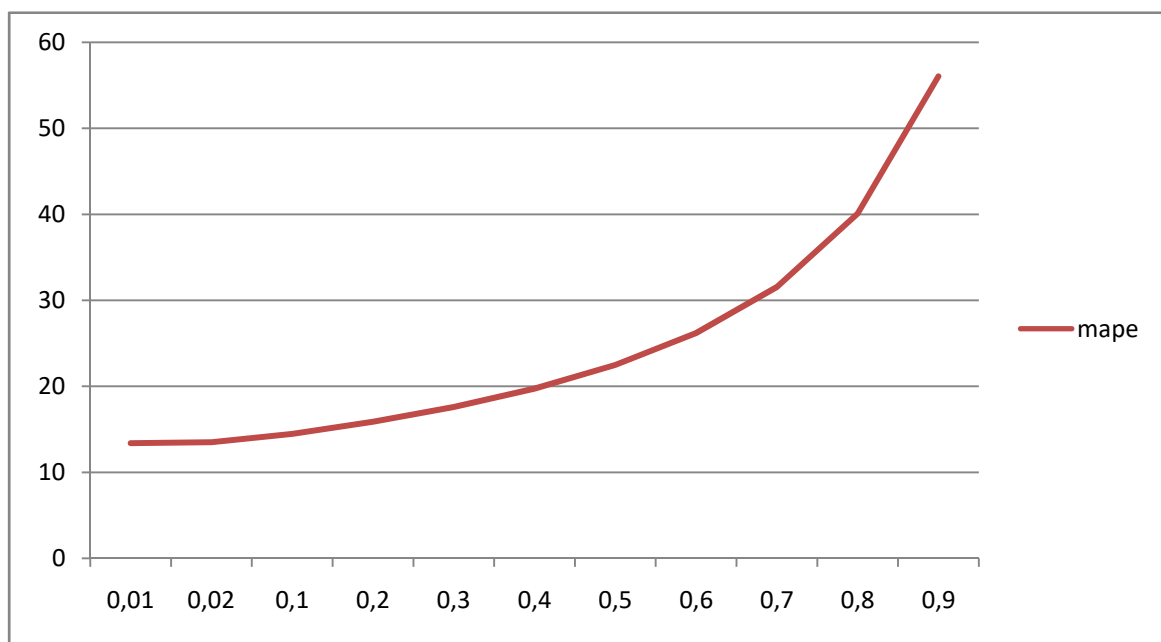


Figure 16: MAPE in relation to  $\lambda$

After we found the  $\lambda$ , we will calculate the forecasted revenues and the performance metrics as well.

Quarter	Actual revenues	Forecasted revenues	Error
Q2 2018	3814,366	3597,859179	
Q3 2018	3910,597	3812,200932	98,39607
Q4 2018	4101,684	3909,613039	192,071
Q1 2019	4440,304	4099,76329	340,5407
Q2 2019	4846,916	4436,898593	410,0174
Q3 2019	5173,028	4842,815826	330,2122

Q4 2019	5398,982	5169,725878	229,2561
Q1 2020	5767,690	5396,689439	371,0006
Q2 2020	6148,290	5763,979994	384,31
Q3 2020	6435,640	6144,4469	291,1931
Q4 2020	6644,440	6432,728069	211,7119
Q1 2021	7163,280	6642,322881	520,9571
Q2 2021	7341,780	7158,070429	183,7096
Q3 2021	7483,470	7339,942904	143,5271
Q4 2021	7709,320	7482,034729	227,2853
Q1 2022	7867,770	7707,047147	160,7229
Q2 2022	7970,140	7866,162771	103,9772
Q3 2022	7925,590	7969,100228	-43,5102
Q4 2022	7852,050	7926,025102	-73,9751
Q1 2023	8161,499	7852,789751	308,7092
Q2 2023	8187,299	8158,411908	28,88709

Table 11: Calculated error with EWMA

With the red color are the underestimated revenues and with blue color are the overestimated ones. At first glance, it seems that there are more underestimated periods than overestimated; however the errors are quite low.

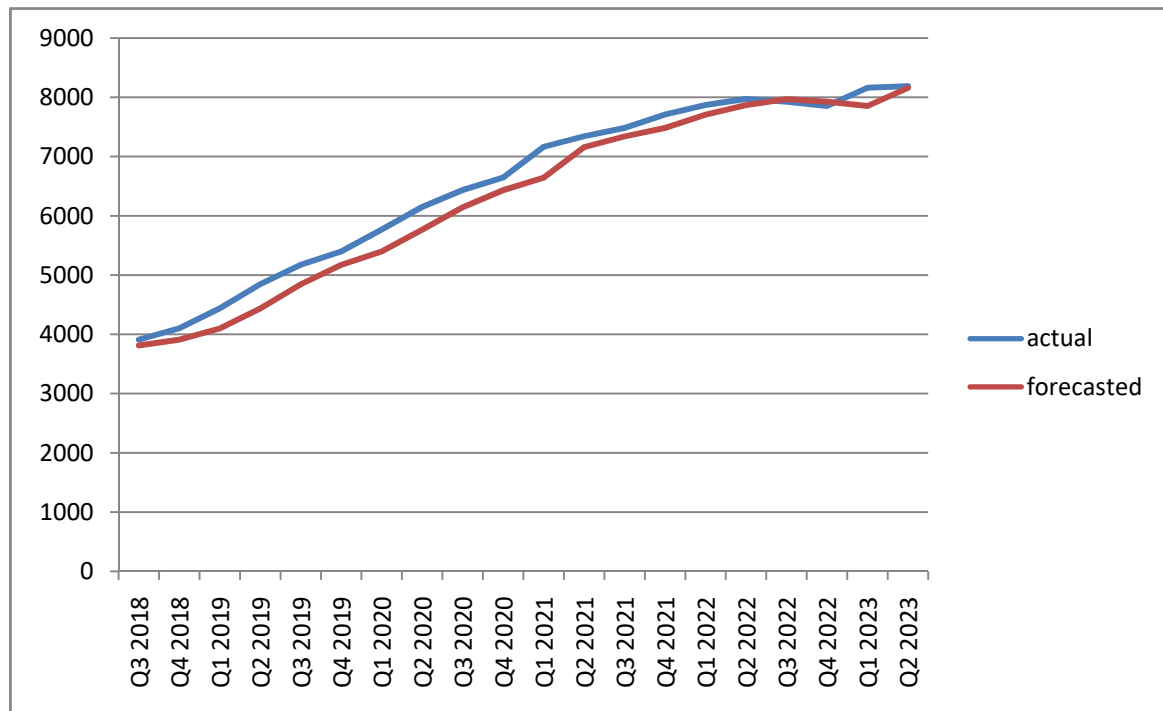


Figure 17: Actual and forecasted revenues using EWMA

From the graph it seems that the forecasted revenues try to reach the actual ones, however most at the times the forecasted ones underestimate actual ones.

By using excel we calculate MAE=232,698 and MAPE=3,9%. Both MAE and MAPE indicate very good model accuracy.

### 3.6.3 SMA

The value of  $\lambda$  of the previous analysis which is quite low indicated that past observations are important for forecasting. So, we will forecast by using a simpler method than EWMA which is SMA, which gives more weight to recent observations. So, for forecast one step ahead we will use a window length of two observations. The formula for calculation SMA would be:  $F_t = (D_{t-1} + D_{t-2}) / 2$

Where  $F_t$ =forecast for period t

$D_{t-1}$  =Revenues for period t-1

$D_{t-2}$  = Revenues for period t-2

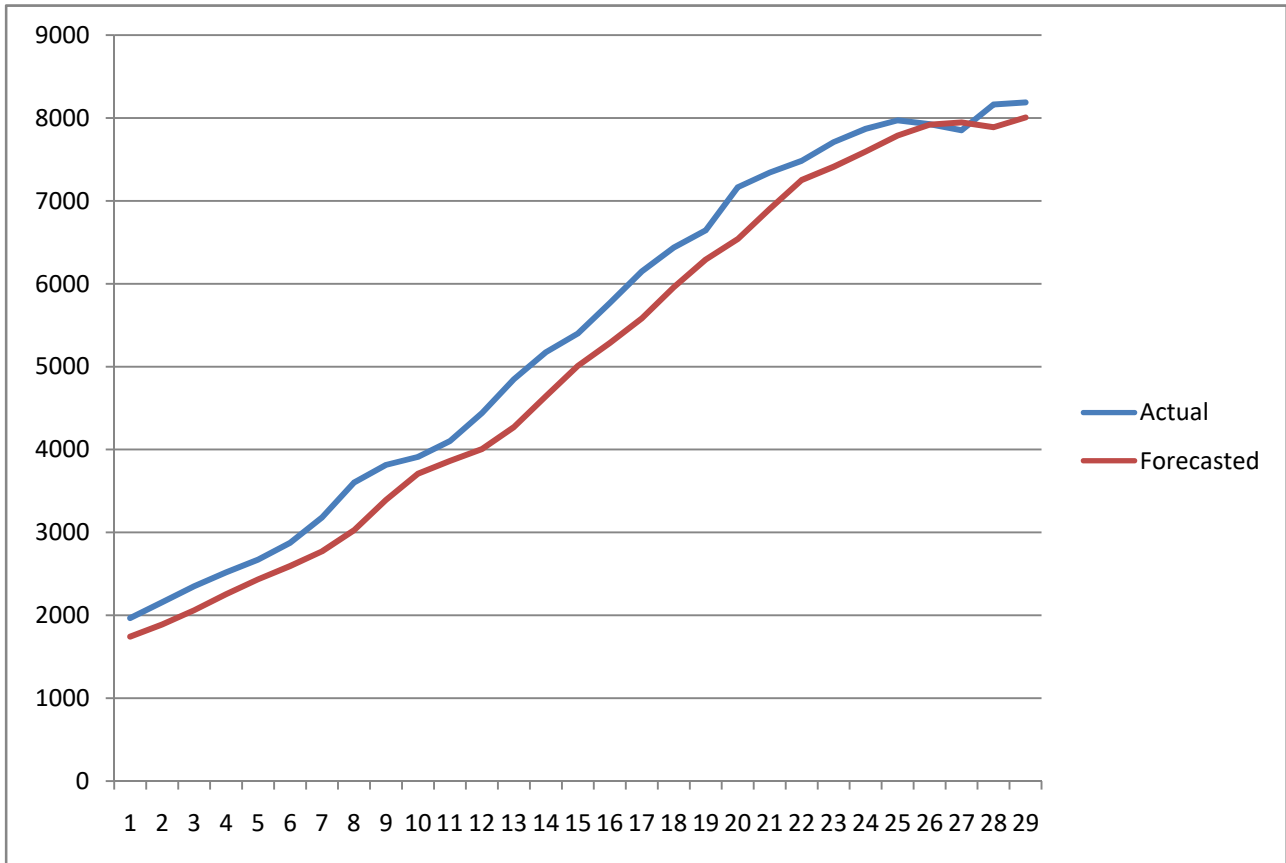
We will use a sample which contains 60% of our data which are 31 observations.

Quarter	Actual	Forecasted	error
Q4 2015	1672,338		
Q1 2016	1812,989		
Q2 2016	1966,472	1742,6635	223,8085
Q3 2016	2157,813	1889,7305	268,0825
Q4 2016	2351,128	2062,1425	288,9855
Q1 2017	2516,241	2254,4705	261,7705
Q2 2017	2670,727	2433,6845	237,0425
Q3 2017	2874,645	2593,484	281,161
Q4 2017	3180,603	2772,686	407,917
Q1 2018	3602,105	3027,624	574,481
Q2 2018	3814,366	3391,354	423,012
Q3 2018	3910,597	3708,2355	202,3615
Q4 2018	4101,684	3862,4815	239,2025
Q1 2019	4440,304	4006,1405	434,1635
Q2 2019	4846,916	4270,994	575,922
Q3 2019	5173,028	4643,61	529,418

Q4 2019	5398,982	5009,972	389,01
Q1 2020	5767,690	5286,005	481,685
Q2 2020	6148,290	5583,336	564,954
Q3 2020	6435,640	5957,99	477,65
Q4 2020	6644,440	6291,965	352,475
Q1 2021	7163,280	6540,04	623,24
Q2 2021	7341,780	6903,86	437,92
Q3 2021	7483,470	7252,53	230,94
Q4 2021	7709,320	7412,625	296,695
Q1 2022	7867,770	7596,395	271,375
Q2 2022	7970,140	7788,545	181,595
Q3 2022	7925,590	7918,955	6,635
Q4 2022	7852,050	7947,865	-95,815
Q1 2023	8161,499	7888,82	272,679
Q2 2023	8187,299	8006,7745	180,5245

**Table 12: Error calculation using SMA**

With the red color are the underestimated revenues and with blue color are the overestimated ones. At first glance, it seems that there are more underestimated periods than overestimated ones.



**Figure 18: Actual and forecasted revenues using SMA**

The forecasted revenues try to follow the actual revenues; however it seems that at most of the times forecasted revenues underestimate actual ones.

By using excel we found that MAE= 338,2938 and MAPE=7,76% which indicates very good model accuracy.



### 3.6.4 Model Comparison

	Regression	EWMA	SMA
MAE	547,4332	<b>232,698</b>	338,2938
MAPE	8,095%	<b>3,9%</b>	7,76%

**Table 13: Model comparison**

As we notice from the table above, the model with best accuracy level is EWMA, followed by SMA and finally by Regression.

The revenues of our sample data revenues were rising with a bigger rhythm at the beginning and that's why the Regression model responds better than the other two models. However, the last observations rise with a slower rhythm and we notice that the error of the Regression model was bigger than in the previous observations. On the opposite, EWMA and SMA underestimate the data during the periods that they raising fast and at during the last observations where the percentage of raise was lower, the error was lower.

## 4 Evaluation, Conclusions and Limitations

### 4.1 Evaluation and Conclusions

At this point of our study, we will focus on the main findings of the analysis we conducted on the previous chapters by evaluate them and reach to some conclusions. We will also, review the questions we addressed at the beginning of the study and answer them based on the findings.

- What are the main statistical features of revenues
- Is there any trend observed?

Regarding the DVD Rental service of Netflix, revenues were rising from the time the service was available to the market till 2011, where started the abrupt fall of revenues. So, there was an upward trend at the beginning till 2011 and then followed a downward trend till the end of the service in 2023. Since the revenues change continuously, the time series is not stationary at all. Also, from the box plots it is verified that there is not sign of seasonality or a cyclical pattern. One possible explanation for the fall of the revenues is the appearance of the Streaming service, which was basically the transition from the old fashioned to a digital form of entertainment.

Regarding the streaming service, revenues were rising fast by the time the service was available to the market and then they increased with a slower rhythm. So, by examine the graph of the revenues we conclude that there is an upward trend. The time series is not stationary since the mean is keep rising and by examining the box plots per quarter, we can to the assumption that there is no sign of seasonality or cyclical pattern.

Netflix announced the end of DVD Rental Service in 2023 since the profits of the service were very low and there were only a few subscribers for this certain service. Also, Netflix would like to invest more on Streaming service, since there is high competition on the field.

- Has the appearance of Disney + and COVID-19 had any significant effect on the revenues of NETFLIX?

Disney+ revenues, by the time the platform was available to the market were high, but as the time passes raised with a lower rhythm and that is obvious from the logarithm revenues diagram. The graph shows that there is an upward trend and no sign of seasonality or cyclical pattern which is also confirmed from the box plots. So, since the mean is rising the time series is not stationery.

Due to sources, Disney+ lost nearly 300.000 subscribers in the US and Canada due a price raise of subscription. Moreover Disney + lost a lot of subscribers due to the loss of the rights to digital streaming of the cricket tournament Indian Premier League.

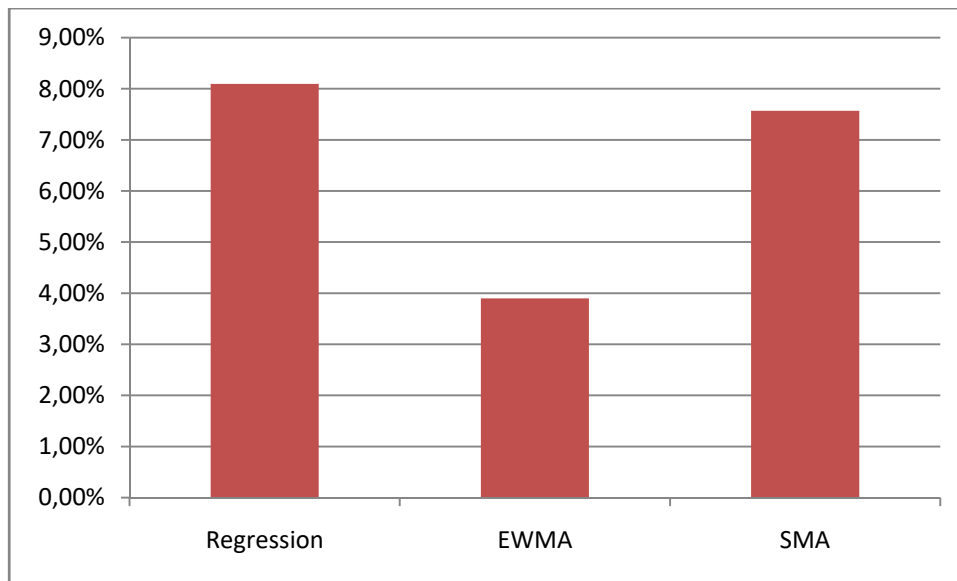
Disney+ is planning to raise the price of its subscription plan in order to raise its revenues. Moreover, it is planning to offer a new service which subscribers would pay one subscription and have access in two platforms: Disney+ and Hulu.

According to our analysis, Disney+ did have an effect on Netflix revenues. Disney launched its new streaming platform in 2020 which received huge acceptance by the public. Its content contains movies and series of Disney, Marvel, Star Wars, etc which are unique choices for families and for fans. Surely, it was very convenient for Disney+ that Netflix during pandemic couldn't provide new Netflix originals content to its subscribers, which make Disney+ more competitive. Finally, the next step Disney+ followed in order to become more competitive is the cooperation with Hulu. From now on, subscribers could choose a new subscription plan offered by Disney+ which includes two streaming platforms in one subscription plan. This means that a subscriber can have access to a huge library of movies and series. So, it remains to see in the future if this cooperation is going to be successful and how will affect Netflix.

Most people would say that COVID-19 should have raised Netflix revenues since people forced to stay at home and had to found ways t entertain themselves. However, from our research, it seems that COVID-19 didn't have a strong effect in Netflix revenues. Of course, there was a raise in revenues, but this percentage is not that remarkable.

So, we could not say that Netflix benefited that much by Covid-19, but instead faced some important issues. First of all, Netflix originals are a characteristic which differ Netflix from other streaming platforms. During pandemic, the filming of movies and series has to be stopped, so Netflix couldn't provide new content to its subscribers, fact which is raise the competition between other streaming platforms which could offer the same service. Moreover, staying at home and spend more hours watching Netflix doesn't mean that Netflix revenues are increased. Finally, after Covid-19 outbreak, Netflix raise of revenues seems to fall, since there wasn't new content available to its subscribers and also faced a strong competition from other streaming platforms.

- Which of the employed forecasting methods fits best the data and at what accuracy rate can revenues be predicted?



**Figure 19: Model comparison**

The EWMA model proved to be the one that best fits the data followed by SMA. As I mentioned before, if we observe closely Regression analysis responded better at the beginning, where the revenues increases faster and the error was smaller in comparison with SMA and EWMA which in the beginning their error was bigger. With  $\lambda=0,01$  which was chosen for EWMA more weight was given to past observations and that's why EWMA keeps up with the boost raise of revenues.

## 4.2 Limitations

One basic limitation in our study is the lack of data for each subscription plan offered by Netflix. The data that are mainly used in our study are the total revenues for each service of Netflix. It would be interesting, if we have access to revenue data for each subscription plan per region in order to widen our research. Moreover, it would be interesting to examine the revenues per country in order to see which country is more profitable. Unfortunately, we didn't have access to that kind of data. So, the lack of sufficient data decreases the field of our research and we had shorter time series. Finally, streaming is a fast growing industry and new streaming platforms enter the market year by year. For that reason, forecasting is very difficult to be conducted with high accuracy and surely it can't be done long term. The companies should watch closely the developments and their competitors in order to be prepared for each circumstance. On the methodology side of things, there are many directions along which this study could be expanded. Apart from simple regression analysis, it would be interesting to apply flexible semi parametric models, such as neural network architectures, to explore a nonlinear functional relationship between our dependent variable and the explanatory ones. However, due to limited data, it is important to implement in parallel a model selection strategy to help us choose the level of model complexity that can be justified by data and avoid over fitting.

## References

Wikipedia. (2024). Netflix. Wikipedia, The Free Encyclopedia

<https://en.wikipedia.org/wiki/Netflix>

Flixpatrol. (2024). *Top Streaming Services by Subscribers.*

<https://flixpatrol.com/streaming-services/subscribers/>

Investopedia. (2023). *How Netflix Is Changing the TV Industry.*

<https://www.investopedia.com/articles/investing/060815/how-netflix-changing-tv-industry.asp>

Digitaltrends. (2023). *The 10 most popular streaming services, ranked by subscriber count.*

<https://www.digitaltrends.com/home-theater/most-popular-streaming-services-by-subscribers/>

VPN wiki. (2024). *Netflix prices in 245 countries (2024 Edition)*

<https://vpnwiki.com/en/netflix-subscription-prices>

Medium. (2023). *MAPE vs MAE: Which Metric is Better?*

<https://medium.com/trusted-data-science-haleon/mape-vs-mae-which-metric-is-better-68dd559cbfb1>

What's on Disney plus. (2023). *Disney Executive Explains why the passed on IPL crickey eights for Disney+ Hotstar*

<https://whatsondisneyplus.com/disney-executive-explains-why-they-passed-on-ipl-cricket-rights-for-disney-hotstar/?fbclid=IwAR2D7JHNI10TC0ANTFu-3OLSo3GWPLZH7xtxxXqJeufN0XUOmXOnGgF8e4k>

CNCB. (2023). *Disney to raise price on ad-free Disney+ to \$13.99 per month starting October.*

[https://www.cnn.com/amp/2023/08/09/disney-to-raise-disney-price-for-ad-free-plan-in-september.html?fbclid=IwAR0cCNNI0uIZ-v0edF0MQy\\_7NgZSf5OvKI9Q3Y7SnF3UYFRWe\\_FP1Sd-1XM](https://www.cnn.com/amp/2023/08/09/disney-to-raise-disney-price-for-ad-free-plan-in-september.html?fbclid=IwAR0cCNNI0uIZ-v0edF0MQy_7NgZSf5OvKI9Q3Y7SnF3UYFRWe_FP1Sd-1XM)

N.S. Thomaidis and G. Dounias (2012), “A comparison of statistical tests for the adequacy of a neural network regression model”, *Quantitative Finance* 12(3), pp. 437-449. (IF: 2.222)

N.S. Thomaidis, and G. Dounias (2011), “On detecting the optimal structure of a neural network model under strong statistical features in errors”, *Journal of Time Series Analysis* 32 (3), pp. 204-222. (IF: 1.366)

**Author's Statement:**

I hereby expressly declare that, according to the article 8 of Law 1559/1986, this dissertation is solely the product of my personal work, does not infringe any intellectual property, personality and personal data rights of third parties, does not contain works/contributions from third parties for which the permission of the authors/beneficiaries is required, is not the product of partial or total plagiarism, and that the sources used are limited to the literature references alone and meet the rules of scientific citations.