## VISVESVARAYA TECHNOLOGICAL UNIVERSITY

"Jnana Sangama", Belagavi – 590 018



An Internship Report on

## "ONE PIECE ARC"

Submitted in partial fulfillment for the award of degree of

**Bachelor of Engineering** 

in

**Computer Science Engineering** 

Submitted By

**Daniel D (4AD18CS017)** 

Internship Carried Out at

AUDAZ VENTURES PVT. LTD.



#### **INTERNAL GUIDE**

ATME College of Engineering Mysuru



**EXTERNAL GUIDE** 

Mr. Vikrant Kumar

L&D Manager

AUDAZ Venture PVT. LTD.

# DEPARTMENT OF COMPUTER SCIENCE AND ENGINEERING ATME COLLEGE OF ENGINEERING

13th KM Stone, Mysuru – Bannur Road, Mysuru – 570028

Phone: 0821 2954081

Website: www.atme.in

2022 - 2023





## DEPARTMENT OF COMPUTER SCIENCE AND ENGINEERING

## **CERTIFICATE**

This is to certify that the internship report titled "ONE PIECE ARC" is carried out by Daniel D bearing USN 4AD18CS017 in partial fulfillment of the requirements for the award of Bachelor of Engineering in Computer Science and Engineering of Visvesvaraya Technological University, Belagavi during the year 2022-2023.

Dept. of CSE
ATME

Signature of HOD

Signature of the External Guide

Mr. Vikrant Kumar

L&D Manager

AUDAZ Ventures

ATME

**Private Limited** 

## **DECLARATION**

I hereby declare that I have completed my four weeks Internship at "AUDAZ VENTURES PVT. LTD.", from 21-08-2022 to 17-09-2022 under the guidance of the internal guide. I have declared that I have worked with full dedication during these four weeks of Internship in partial fulfillment for the award of the degree of Bachelor of Engineering in Computer Science and Engineering from Visvesvaraya Technological University, Belagavi during the year 2022 – 23

Daniel D (4AD18CS017)

### ACKNOWLEDGEMENT

The success and outcome of this internship required a lot of guidance and assistance from many people, and I am extremely privileged to have got this all along the completion of my internship. I thank to all those who have rendered their cherished advice and services towards the completion of the internship.

I wish to express my deep sense of acknowledgement and gratitude to my Internal guide Department of Computer Science and Engineering, for the suggestions and encouragement throughout the making of the internship.

I wish to express my deep sense of acknowledgement and gratitude to my External guide **Mr. Vikrant Kumar,** L&D Manager, Audaz Ventures Private Limited, for the suggestions and encouragement throughout the making of the internship.

I wish to express my deep sense of acknowledgement and gratitude to my Trainers Mr. Azhar, Mr. Mohammed Nawfal and Mr. Mohammed Sadiq for their training and constant support.

I am highly indebted to **Dr. Puttegowda**, Head of Department, Computer science and Engineering, for his kind consents and wholehearted cooperation.

I would like to thank our Principal **Dr. L Basavaraj**, for his encouragement and providing an excellent working environment.

I thank all the lecturers of the dept. for their cooperation and providing with the facilities to carry out the seminar work. I would also express my thanks to all technical and non-technical staff of the Computer Science and Engineering department who have directly or indirectly cooperated with me.

Finally, I would like to express my gratitude to my parents and friends who always stood by me encouraging in all my endeavors.

## **INDEX**

Sl. no.	Particulars	Page No.
1	Executive Summary	1
2	Introduction	2
3	Company Profile	3
4	Objectives and Scope of the study	4
	4.1 Objectives of Study	4
	4.2 Scope of Study	4
5	Theoretical Background	5 - 8
	5.1 What are Data Analytics	5
	5.2 Ways to Use Data Analytics	5
	5.3 Steps Involved in Data Analytics	6
	5.4 Data Analytics Tools	7
	5.5 Data Analytics Applications	8
6	Research Methodology	9
	6.1 Types of Research in Data Analytics	9
7	Analysis and Interpretation of Data	10 - 14
	7.1 Step 1: Extract the Data	10
	7.2 Step 2: Data Cleaning	11 - 12
	7.3 Step 3: Visualization of Data	13 - 14
8	Findings	15
9	Suggestion / Recommendations	15
10	Conclusion	16
11	Limitations	16
12	Bibliography	17

## **EXECUTIVE SUMMARY**

The internship report highlights the major works carried out in terms of academic and non-academic perspectives. The scope of this document is to identify and describe the analysis carried out, project completed, experience gained and focuses on the achievements as an intern.

I was working with AUDAZ VENTURES PVT. LTD. to complete my internship. I found myself rather lucky by getting the chance to work in such an environment that AUDAZ provided and got introduced to some of the new terms, new technologies and languages.

The project that I worked in certainly helped me by increasing my practical knowledge depth. The research and analysis projects were particularly helpful in widening my views regarding data analysis tools and data science. Besides there were some vital lessons which will obviously help in future jobs.

This report intended to describe the analysis of the One Piece Arc. The analysis provides the inferences about the similar patter for the both Manga and Anime in references of Chapters and Episodes respectively for the fans.

The Chapters One to Six provide the introduction to the project and data analytics. The Chapter Seven explains the data used and data processing for analysis. Finally, in the Chapter Eight to ten the analysis and conclusions based on the project analysis.

## **INTRODUCTION**



One Piece is a Manga written and illustrated by Eiichiro Oda and published in Weekly by Shonen Jump magazine since the year 1997, which tells the adventures of Monkey D. Luffy and the Straw Hat Pirates in search of the One Piece treasure.

In today's world the entertainment industry is focusing on more entertaining animations to provide a better viewing experience. As each sector of the market for entertainment is growing, data is building up day by day, we need to keep the record of the data which can be helpful for the analytics and evaluation. Now we don't have data in gigabyte or terabyte but in petabyte and zettabyte and this data cannot be handled with the day-to-day software such as Excel or MATLAB. Therefore, in this report we will be dealing with large data sets with the high-level programming language 'Python'.

The main goal of this project is to aggregate and analyze the data collected from the different data sources available on the internet. This project mainly focuses on the usage of the Python programming language and its data analytics libraries in the field of animation industry. This language has not only it's application in the field of just analyzing the data but also for the prediction of the upcoming scenarios.

The purpose of using this specific language is due to its versatility, vast libraries (Pandas, NumPy, Matplotlib, etc.), speed limitations, and ease of learning. We will be analyzing large housing prices datasets in this project which cannot be easily analyzed in other tools as compared to Python. Python does not have its limitation to only data analytics but also in many other fields such as Artificial intelligence, Machine learning, and many more.

This dataset is an updated version of the One Piece dataset available in the form of CSV file.

## **COMPANY PROFILE**



Audaz Ventures Private Limited

CIN: U80903DL2020PTC365560

Audaz Ventures Private Limited, with the headquarter in New Delhi and corporate office in Bengaluru was established in 2020. It is a service - based company which provides services and solutions to 62 institutes across India. It has its presence in more than 12 cities across India.

Its services include

- Software solutions
- Digital marketing
- SAP software
- ERP software
- Blockchain software
- Placement related training to the Engineering Graduates and many more
- Placement opportunities to students

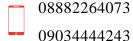
Thanks & Regards

Rahul Oberoi

Director









info@audazlearning.com



C2C/2-100, JANAKPURI, NEW DELHI - 110058

## **OBJECTIVES AND SCOPE OF THE STUDY**

## 4.1 Objectives of the study

- To study the data analytics concepts of data gathering, data cleaning, EDA, and result interpretation.
- To study the Arcs, present in One Piece.
- To study the least and most episodes and chapters present in One Piece.

## 4.2 Scope of Study

- The aim of this study is to analyze the data generated by One Piece.
- The scope of the study is limited to the data available from One Piece as each chapter or episode is just released only one, once a week.

## THEORETICAL BACKGROUND

## 5.1 What are Data Analytics?

Companies around the globe generate vast volumes of data daily, in the form of log files, web servers, transactional data, and various customer-related data. In addition to this, social media websites also generate enormous amounts of data. Companies ideally need to use all their generated data to derive value out of it and make impactful business decisions. Data analytics is used to drive this purpose.

Data analytics is the process of exploring and analyzing large datasets to find hidden patterns, unseen trends, discover correlations, and derive valuable insights to make business predictions. It improves the speed and efficiency of your business. Businesses use many modern tools and technologies to perform data analytics.

## **5.2** Ways to Use Data Analytics

#### Improved Decision Making

Data Analytics eliminates guesswork and manual tasks. Be it choosing the right content, planning marketing campaigns, or developing products. Organizations can use the insights they gain from data analytics to make informed decisions. Thus, leading to better outcomes and customer satisfaction.

#### • Better Customer Service

Data analytics allows you to tailor customer service according to their needs. It also provides personalization and builds stronger relationships with customers. Analyzed data can reveal information about customers' interests, concerns, and more. It helps you give better recommendations for products and services.

## • Efficient Operations

With the help of data analytics, you can streamline your processes, save money, and boost production. With an improved understanding of what your audience wants, you spend lesser time creating ads and content that aren't in line with audience's interests.

#### Effective Marketing

Data analytics gives you valuable insights into how your campaigns are performing. This helps in fine-tuning them for optimal outcomes. Additionally, you can also find potential customers who are most likely to interact with a campaign and convert into leads.

## 5.3 Steps Involved in Data Analytics



There are a few steps that are involved in the data analytics lifecycle. Below are the steps that you can take to solve your problems.

## • STEP 1: Identify

Identifying or understanding the business problems, defining the organizational goals, and planning a lucrative solution is the first step in the analytics process.

#### • STEP 2: Collection

Collection or Data collection, is need to collect transactional business data and customer-related information from the past few years to address the problems your business is facing.

#### • STEP 3: Clean

Clean or data clean, is the data you collect will often be disorderly, messy, and contain unwanted missing values. Such data is not suitable or relevant for performing data analysis. Hence, you need to clean the data to remove unwanted, redundant, and missing values to make it ready for analysis.

#### • STEP 4: Analyze

Analyze or Data exploration and analysis, is you gather the right data, the next vital step is to execute exploratory data analysis. It can be used as data visualization and business intelligence tools, data mining techniques, and predictive modeling to analyze, visualize, and predict future outcomes from this data. Applying these methods can tell you the impact and relationship of a certain feature as compared to other variables.

#### • STEP 5: Interpret the results:

Interpret or Interpret of the results, is the final step is to interpret the results and validate if the outcomes meet your expectations. You can find out hidden patterns and future trends. This helps to gain insights that supports with appropriate data-driven decision making.

## **5.4 Data Analytics Tools**









Here are 7 data analytics tools, including a couple of programming languages that can help you perform analytics better.

- **Python** is an object-oriented open-source programming language. It supports a range of libraries for data manipulation, data visualization, and data modeling.
- **R** is an open-source programming language majorly used for numerical and statistical analysis. It provides a range of libraries for data analysis and visualization.
- Tableau is a simplified data visualization and analytics tool. This helps you create a
  variety of visualizations to present the data interactively, build reports, and dashboards
  to showcase insights and trends.
- **Power BI** is a business intelligence tool that has an easy 'drag and drop functionality. It supports multiple data sources with features that visually appeal to data. Power BI supports features that help you ask questions to your data and get immediate insights.
- QlikView offers interactive analytics with in-memory storage technology to analyze
  vast volumes of data and use data discoveries to support decision making. It provides
  social data discovery and interactive guided analytics. It can manipulate colossal data
  sets instantly with accuracy.
- **Apache Spark** is an open-source data analytics engine that processes data in real-time and carry out sophisticated analytics using SQL queries and machine learning algorithm.

## **5.5** Data Analytics Applications



Data analytics is used in almost every sector of business, below are a few of them:

- Retail helps retailers understand their customer needs and buying habits to predict
  trends, recommend new products, and boost their business. They optimize the supply
  chain, and retail operations at every step of the customer journey.
- **Healthcare** industries analyze patient data to provide lifesaving diagnoses and treatment options. Data analytics help in discovering new drug development methods as well.
- **Manufacturing** sectors can discover new cost-saving opportunities. They can solve complex supply chain issues, labor constraints, and equipment breakdowns.
- **Banking sector** uses analytics to find out probable loan defaulters and customer churn out rate. It also helps in detecting fraudulent transactions immediately.
- **Logistics** companies use data analytics to develop new business models and optimize routes. This, in turn, ensures that the delivery reaches on time in a cost-efficient manner.

## RESEARCH METHODOLOGY

This research is a descriptive analytic, in which we are analyzing Arcs of One Piece. It is based on primary data and inferences derived from it.

## **6.1** Types of Research in Data Analytics

#### • Predictive Analytics

It turns the data into valuable, actionable information. predictive analytics uses data to determine the probable outcome of an event or a likelihood of a situation occurring and holds a variety of statistical techniques from modeling, machine, learning, data mining, and game theory that analyze current and historical facts to make predictions about a future event.

### • Descriptive Analytics

It looks at data and analyze past event for insight as to how to approach future events. It looks at the past performance and understands the performance by mining historical data to understand the cause of success or failure in the past. Almost all management reporting such as sales, marketing, operations, and finance uses this type of analysis. The descriptive model quantifies relationships in data in a way that is often used to classify customers or prospects into groups.

#### • Prescriptive Analytics

It automatically synthesizes big data, mathematical science, business rule, and machine learning to make a prediction and then suggests a decision option to take advantage of the prediction. Prescriptive analytics goes beyond predicting future outcomes by also suggesting action benefit from the predictions and showing the decision maker the implication of each decision option. Prescriptive Analytics not only anticipates what will happen and when to happen but also why it will happen.

#### • Diagnostic Analytics

It generally uses historical data over other data to answer any question or for the solution of any problem. We try to find any dependency and pattern in the historical data of the particular problem.

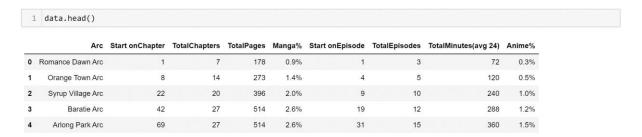
## ANALYSIS AND INTERPRETATION OF DATA

Before analyzing and visualization we need the raw data, and this raw data can be gathered from different open-source data websites available on the internet. This data will be in raw CSV form, it may be individual subunits of data or the complete dataset containing all the groups of files.

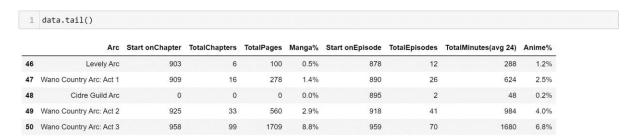
#### 7.1 STEP 1: EXTRACT THE DATA

1 data = pd.read\_csv("OnePieceArcs.csv")

#### Reading the data



### Displays head data



### Displays tail data

- o **Arc:** The name of the Arc
- o **Start on Chapter:** Starting chapter in which the arc was started
- o **TotalChapters:** Total number of chapters drawn In Manga Volume
- o **TotalPages:** Total number of pages used for that particular Arc
- o Manga%: The overall contribution of that particular Arc in One Piece Manga
- o **Start onEpisode:** Starting episode in which the arc was started
- o **TotalEpisode:** Total number of episodes animated by Toei Animation
- o **TotalMinutes(avg 24):** Total number of minutes particular Arc was telecasted in Anime
- o **Anime%:** The overall contribution of that particular Arc in One Piece Anime

#### 7.2 STEP 2: DATA CLEANING

```
1 data.info()
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 51 entries, 0 to 50
Data columns (total 9 columns):
                          Non-Null Count Dtype
# Column
 0 Arc
                           51 non-null
                                           object
    Start onChapter
                           51 non-null
                                           int64
     TotalChapters
                           51 non-null
                                           int64
    TotalPages
                           51 non-null
                                           int64
    Manga%
Start onEpisode
                           51 non-null
                                           object
                           51 non-null
                                           int64
    TotalEpisodes
                           51 non-null
                                           int64
    TotalMinutes(avg 24)
                          51 non-null
                                          int64
                           51 non-null
 8 Anime%
                                          object
dtypes: int64(6), object(3)
memory usage: 3.7+ KB
```

#### **info()** method prints information about the DataFrame

The information contains the number of columns, column labels, column data types, memory usage, range index, and the number of cells in each column (non-null values)

1	data.describe()					
	Start onChapter	TotalChapters	TotalPages	Start onEpisode	TotalEpisodes	TotalMinutes(avg 24)
coun	t 51.000000	51.000000	51.000000	51.000000	51.000000	51.000000
mear	n 268.823529	20.705882	381.490196	354.745098	20.156863	483.764706
sto	d 317.019287	26.133728	471.266268	291.209742	24.345326	584.287826
mir	n 0.000000	0.000000	0.000000	1.000000	2.000000	48.000000
25%	0.000000	0.000000	0.000000	85.000000	4.000000	96.000000
50%	6 106.000000	11.000000	209.000000	313.000000	11.000000	264.000000
75%	519.500000	27.000000	514.000000	549.000000	30.000000	720.000000
max	x 958.000000	102.000000	1830.000000	959.000000	118.000000	2832.000000

**describe()** method returns description of the data in the Data Frame.

If the Data Frame contains numerical data, the description contains this information for each column: count - The number of not-empty values. mean - The average (mean) value. std - The standard deviation.

**columns** displays the data in two-dimensional size-mutable, potentially heterogeneous tabular data structure with labelled axes (rows and columns)

From the data present in the file, we can estimate maximum episodes in the particular arc is,

```
# Maximum Episode before the data merging
max_episode = data[data['TotalEpisodes'] == data['TotalEpisodes'].max()]

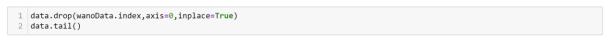
Arc Start onChapter TotalChapters TotalPages Manga% Start onEpisode TotalEpisodes TotalMinutes(avg 24) Anime%

1 Dressrosa Arc 700 102 1830 9.4% 629 118 2832 11.5%
```

Since Wano Country Arc is categorized into 3 parts so, by merging these 3 arcs together we can estimate the maximum episodes in the series as,

	Arc	Start onChapter	TotalChapters	TotalPages	Manga%	Start onEpisode	TotalEpisodes	TotalMinutes(avg 24)	Anime%
47	Wano Country Arc: Act 1	909	16	278	1.4%	890	26	624	2.5%
48	Cidre Guild Arc	0	0	0	0.0%	895	2	48	0.2%
49	Wano Country Arc: Act 2	925	33	560	2.9%	918	41	984	4.0%
50	Wano Country Arc: Act 3	958	99	1709	8.8%	959	70	1680	6.8%
51	Wano Country Arc	909	148	2547	14.0%	890	137	3288	13.3%

Therefore, the data which is unnecessary can be dropped,



	Arc	Start onChapter	TotalChapters	TotalPages	Manga%	Start onEpisode	TotalEpisodes	TotalMinutes(avg 24)	Anime%
44	Marine Rookie Arc	0	0	0	0.0%	780	3	72	0.3%
45	Whole Cake Island Arc	825	78	1349	6.9%	783	95	2280	9.2%
46	Levely Arc	903	6	100	0.5%	878	12	288	1.2%
48	Cidre Guild Arc	0	0	0	0.0%	895	2	48	0.2%
51	Wano Country Arc	909	148	2547	14.0%	890	137	3288	13.3%

Thus, the new data extraction is,

```
# Maximum episode
max_episode = data[data['TotalEpisodes'] == data['TotalEpisodes'].max()]
print(f"{max_episode.iloc[0].values[0]} has highest number of episodes: {max_episode.iloc[0].values[-3]} episodes")
max_episode
```

Wano Country Arc has highest number of episodes: 137 episodes

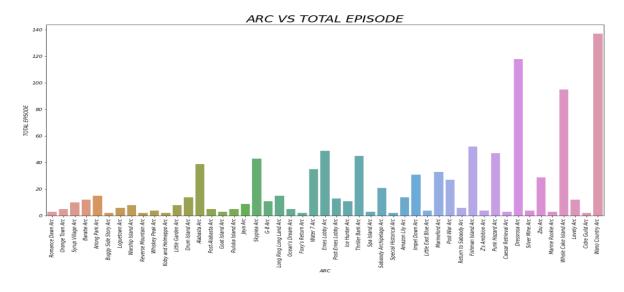
Arc	Start onChapter	TotalChapters	TotalPages	Manga%	Start onEpisode	TotalEpisodes	TotalMinutes(avg 24)	Anime%
1 Wano Country Arc	909	148	2547	14.0%	890	137	3288	13.3%

#### 7.3 STEP 3: VIZUALIZAION OF DATA

#### **BARPLOT**

```
# Barplot can visualization allows you to examine data distribution
sns.barplot(data['Arc'],data['TotalEpisodes'])
plt.xticks(rotation=90)
plt.xlabel("ARC")
plt.ylabel("TOTAL EPISODE")
plt.title("ARC VS TOTAL EPISODE",fontsize=24)
```

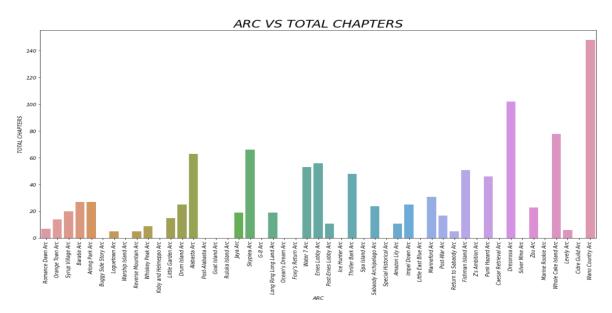
Text(0.5, 1.0, 'ARC VS TOTAL EPISODE')



#### **Used to evaluate Arc vs Total Episode**

```
# Barplot can visualization allows you to examine data distribution
sns.barplot(data['Arc'],data['TotalChapters'])
plt.xticks(rotation=90)
plt.xlabel("ARC")
plt.ylabel("TOTAL CHAPTERS")
plt.title("ARC VS TOTAL CHAPTERS",fontsize=24)
```

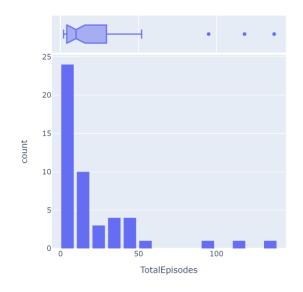
Text(0.5, 1.0, 'ARC VS TOTAL CHAPTERS')

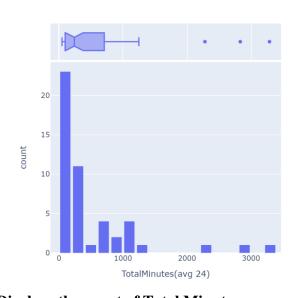


## **Used to evaluate Arc vs Total Chapter**

## HISTOGRAM

```
# Histogram based on Total episodes
ax = px.histogram(data,x='TotalEpisodes',marginal="box",nbins=18)
ax.update_layout(bargap=0.2)
```



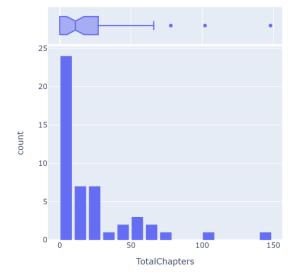


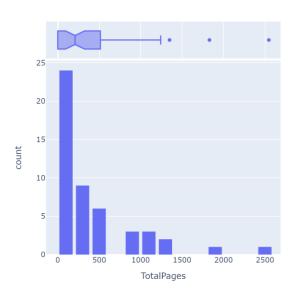
## **Displays the count of Total Episodes**

Displays the count of Total Minutes

```
# Histogram based on Total chapters
ax = px.histogram(data,x='TotalChapters',marginal="box",nbins=18)
ax.update_layout(bargap=0.2)
```







**Displays the count of Total Chapters** 

**Displays the count of Total Pages** 

## **FINDINGS**

Some of the conclusions that we get from this analysis are:

- It shows the total episodes and chapters with respect to anime and manga
- It shows the arcs vs other content

## SUGGESTIONS / RECOMMENDATIONS

The success of any anime depends on its viewers and similarly the success of any manga depends on its readers. Here are few suggestions which is purely based on subjective & objective data analysis.

The availability of a particular content to a particular audience can be increased if author provides with a good plot or arc.

## **CONCLUSION**

The objective of this analysis was to extract useful information for the arcs of One Piece based on the One Piece Arc Data.

This study analyzed the current arcs. The analysis review suggested the longest and shortest arcs. At the same time, it also suggests the fillers which isn't a necessary part for the arc hence it can be ignored. The study used a sample data from whole dataset of One Piece Arc.

## **LIMITATIONS**

In every research undertaken there are certain unavoidable limitations. This research too has the same. This includes the fact that the data is of a certain only for a week, the current trends of data might be not accessible as it is difficult to acquire required information unless the next manga or anime is released.

Another major limitation this analysis is that the data used is specific to a week of the month, and content keeps of coming in weeks.

## **BILBLIOGRAPHY**

- <a href="https://www.kaggle.com/datasets/tarundalal/one-piece-arcs">https://www.kaggle.com/datasets/tarundalal/one-piece-arcs</a>
- <a href="https://www.simplilearn.com/tutorials/data-analytics-tutorial/what-is-data-analytics">https://www.simplilearn.com/tutorials/data-analytics-tutorial/what-is-data-analytics</a>
- <a href="https://matplotlib.org/stable/index.html">https://matplotlib.org/stable/index.html</a>
- <a href="https://pandas.pydata.org/">https://pandas.pydata.org/</a>
- <a href="https://seaborn.pydata.org/">https://seaborn.pydata.org/</a>
- https://www.geeksforgeeks.org/