

**VISVESVARAYA TECHNOLOGICAL UNIVERSITY**  
**JNANA SANGAMA, BELAGAVI-590014**



A V<sup>th</sup> Semester Internship Report on

**“DATA SCIENCE”**

*submitted in partial fulfilment of the requirements for the Award of a Degree of*

**BACHELOR OF ENGINEERING**

**IN COMPUTER SCIENCE AND ENGINEERING**

*Submitted by:*

**SIDDHANT VARDHAMAN CHOUGULE**

**2AG21CS106**

Under Supervision of

**Zeel Code Labs**

**Belagavi.**

(Duration: 27<sup>th</sup> Oct 2023 to 26<sup>th</sup> Nov 2023)



**ANGADI INSTITUTE OF TECHNOLOGY AND MANAGEMENT**

**Belagavi-590009**

**2023-2024**

**ANGADI INSTITUTE OF TECHNOLOGY & MANAGEMENT  
BELAGAVI -590009**

**DEPARTMENT OF COMPUTER SCIENCE & ENGINEERING**



# Certificate

This is to certify that bearing **SIDDHANT VARDHAMAN CHOUGULE** USN: **2AG21CS106**, a student of **Angadi Institute of Technology and Management, Belagavi**, has undergone four weeks of Internship at **ZEEL CODE LABS, Belagavi** on **DATA SCIENCE**. This internship report is submitted in partial fulfillment for the award of Bachelor of Engineering in Computer Science and Engineering under Visvesvaraya Technological University, Belagavi during the academic year 2023-2024.

Signature of the  
Guide

**Mr. Vidyadhar  
Hanji**

Assistant Professor,  
Dept. of CSE, AITM

Signature of the  
Coordinator

**Mr. Vidyadhar Hanji**

Assistant Professor,  
Dept. of CSE, AITM

Signature of the HOD

**Dr Dhanashree Kulakarni**

Professor and Head  
Dept. of CSE, AITM

Signature of the  
Principal

**Dr. Anand Deshpande**

Principal and Director,  
AITM, Belagavi

Name of the Examiner

1. \_\_\_\_\_

2. \_\_\_\_\_

(Training & Placement Officer, AITM)

Signature with date

\_\_\_\_\_

\_\_\_\_\_

#3518, Vijaylaxmi Arcade  
Narvekar Street, Belgaum-590001  
Karnataka.  
Phone: 0831-3587060, 9164558559, 9880917783  
Website: www.zeelnet.com



Date: 02/December/2023

## CERTIFICATE

To,  
**Mr. Siddhant Vardhaman Chougule**  
Angadi Institute of Technology & Management, Belagavi

Has successfully completed the internship program at **Zeel Code Labs**, demonstrating dedication, enthusiasm, and a commitment to learning. The internship was conducted in the following domain:

**Domain Name: Data Science**

**Internship Duration:** Starting Date: 27/10/23 End Date: 26/11/23

During the internship period actively contributed to **Project Development, Data Analysis and Processing, Algorithm Implementation, Model Evaluation and Collaboration and Teamwork**. They exhibited a work ethic, eagerness to learn and effectively applied their knowledge to Practical situation.

We acknowledge the efforts and hard work put by **Mr. Siddhant Vardhaman Chougule** and appreciate his valuable contribution to our team

Thanks & Regards,

**Bhushan Dongare**  
(Co. Founder)



## **Declaration**

I **SIDDHANT VARDHAMAN CHOUGULE(2AG21CS106)**, studying in the 5<sup>th</sup> semester of Bachelor of Engineering in Computer Science and Engineering at Angadi Institute of Technology and Management, Belagavi, hereby declare that this Internship work entitled “**DATA SCIENCE**” which is being submitted by me in the partial fulfilment for the award of the degree of Bachelor of Engineering in Computer Science and Engineering from Visvesvaraya Technological University, Belagavi is an authentic record of us carried out during the academic year 2023-2024 Department of Computer Science and Engineering, Angadi Institute of Technology and Management, Belagavi.

I further undertake that the matter embodied in the dissertation has not been submitted previously for the award of any degree or diploma by us to any other university or institution.

Place: Belagavi

**SIDDHANT VARDHAMAN CHOUGULE**  
**(2AG21CS106)**

## Acknowledgement

First, I would like to thank the Director of Team **Zeel Code, Belagavi** for allowing me to do an internship within the organization.

I also would like all the people that worked along with me in **Zeel Code Labs, Belagavi** with their patience and openness they created an enjoyable working environment.

It is indeed with a great sense of pleasure and immense sense of gratitude that I acknowledge the help of these individuals.

I am highly indebted to the Principal **Dr. Anand Deshpande**, for the facilities provided to accomplish this internship.

I would like to thank my Head of the Department **Dr. Dhanashree Kulkarni** for her constructive criticism throughout my internship.

I would like to thank **Prof. Vidyadhar Hanji**, Internship Coordinator, Department of CSE for their support and advice to get and complete an internship in above said organization.

I am extremely grateful to my department staff members and friends who helped me in the successful completion of this internship.

**SIDDHANT VARDHAMAN CHOUGULE**

**(2AG21CS106)**

## ABSTRACT

Data Science has emerged as a transformative discipline that harnesses the power of data to inform decision-making, drive innovation, and solve complex problems across various domains. This abstract provides a concise overview of the fundamental concepts, methodologies, and applications of Data Science. Data Science encompasses a multifaceted approach to extracting knowledge and insights from vast and diverse datasets. It integrates techniques from statistics, machine learning, computer science, and domain-specific knowledge to analyse, interpret, and visualize data. In essence, it involves data collection, cleaning, exploration, modelling, and communication, with the ultimate goal of generating actionable insights. The process begins with data collection, where relevant data sources are identified and data is gathered. This may involve structured data from databases, unstructured data from text and images, or real-time data from sensors and IOT devices. Ensuring data quality and reliability is crucial at this stage. Data cleaning and pre-processing follow, involving tasks such as handling missing values, removing outliers, and transforming data into a suitable format. This step lays the foundation for accurate and reliable analysis. Exploratory Data Analysis (EDA) is a pivotal phase where data is visualized and summarized to identify patterns, trends, and potential relationships. EDA techniques include data visualization, statistical analysis, and dimensionality reduction.

## **COMPANY PROFILE**

ZEEL CODE LABS is a Developers specializing in Web Development, Application Development, Data Analysis, Networking, Security Systems. As a developing company, we build your applications which reach all market standards. By driving leads and working to find solutions for our clients, Zeel Code Labs has grown to become one of the largest and most successful code development firm. We are servicing many clients. Our experience in development is the basis for our future success. We know where to find the leads and convert them to revenue streams. We are marketing scientists. Everything we do is based on the strongest possible foundation. Our work is defined by measurable objectives and proven techniques, and our team helps pilot programs that lead to a positive business impact for our clients. And like the finest scientists, we never guess. Particularly when it comes to your business.

### **Benefits to the company/institution through your report:**

The Institute combines pioneering research with top class education. An innovative curriculum allows the student flexibility in selecting courses and projects. Students, even at the undergraduate level, get to participate in ongoing research and technology development - an opportunity unprecedented in India.

## **Learning Objectives/Internship Objectives**

Internships are generally thought to be reserved for college students looking to gain experience in a particular field. However, a wide array of people can benefit from Training Internships to receive real-world experience and develop their skills.

Firstly, they involve gaining hands-on experience in a chosen field and applying classroom knowledge to practical situations. Understanding the company's culture, values, and operations is essential for effective integration into the organization.

An objective for this position should emphasize the skills you already possess in the area and your interest in learning more. Internships are utilized in several different career fields, including architecture, engineering, healthcare, economics, advertising and many more.



## WEEKLY OVERVIEW OF INTERNSHIP ACTIVITIES

<b>1st WEEK</b>	<b>DATE</b>	<b>DAY</b>	<b>MODULES COMPLETED</b>
	27/10/2023	FRIDAY	Orientation Program
	28/10/2023	SATURDAY	Holiday
	30/10/2023	MONDAY	Python Basics
	31/10/2023	TUESDAY	Machine Learning Basics
	1/10/2023	WEDNESDAY	Data Analysis
	2/10/2023	THURSDAY	NumPy Library

<b>2nd WEEK</b>	<b>DATE</b>	<b>DAY</b>	<b>MODULES COMPLETED</b>
	3/11/2023	FRIDAY	Visualization Libraries
	4/11/2023	SATURDAY	Pandas Libraries
	6/11/2023	MONDAY	Missing Values Analysis
	7/11/2023	TUESDAY	Imputation Methods
	8/11/2023	WEDNESDAY	Outliers Analysis
	9/11/2023	THURSDAY	Correlation Analysis and Multicollinearity

<b>3rd WEEK</b>	<b>DATE</b>	<b>DAY</b>	<b>MODULES COMPLETED</b>
	10/11/2023	FRIDAY	Pandas Library
	11/11/2023	SATURDAY	House Rent EDA
	15/11/2023	WEDNESDAY	Introduction to Regression Models
	16/11/2023	TUESDAY	Linear Regression
	17/11/2023	FRIDAY	Logistic Regression
	18/11/2023	SATURDAY	Holiday

<b>4th WEEK</b>	<b>DATE</b>	<b>DAY</b>	<b>MODULES COMPLETED</b>
	20/11/2023	MONDAY	Decision Tree & Random Forest
	21/11/2023	TUESDAY	Task 2- Diabetics Prediction
	22/11/2023	WEDNESDAY	Project Assignment
	23/11/2023	THURSDAY	Implementation of the Project
	24/11/2023	FRIDAY	Implementation of the Project
	25/11/2023	SATURDAY	Submission and Presentation of the Project

## **Table of Contents**

<b>Contents</b>	<b>Page No.</b>
Introduction	1
1.1 Background	
1.2 Objectives	
Methodology	
2.1 Data Collection	2
2.1.1 Movie Database Access	
2.1.2 User Ratings Collection	
2.2 Rating Analysis	3
2.2.1 Numerical Rating Distribution	
2.2.2 User Preferences	
2.2.3 Rating Trends over Time	
2.3 Ethical Considerations	4
2.3.1 Privacy Protection	
2.3.2 Informed Consent	
2.3.3 Data Security Measures	
Results	5
3.1 Movie Rating Overview	
3.1.1 Highest Rated Movies	
3.1.2 Lowest Rated Movies	
3.2 User Rating Patterns	

### 3.2.1 User Ratings Distribution

### 3.2.2 Popular Genres among Users

## Discussion 6

### 4.1 Limitations of Rating System

### 4.2 User Engagement and Feedback

### 4.3 Impact of Expert Reviews

## Code Development 7

### 5.1 Code Structure

### 5.2 Error Handling

### 5.3 Database Interaction

### 5.4 Rating Calculation Algorithms

### 5.5 User Interface

### 5.6 Testing and Validation

### 5.7 Documentation

## Conclusion 12

## References

## LIST OF FIGURES

<b>Figure No.</b>	<b>Figure Name</b>	<b>Page No</b>
Fig. 1.	Display the dataset	9
Fig. 2.	List of head and tail of dataset	9
Fig. 3.	Display Rank and Lowest Watching Movies	10
Fig. 4.	Display of movies according to the actors	10
Fig. 5.	Convert the final Movie rating	11
Fig. 6.	Display the Movies Rating Highest Lowest	11





# Chapter 1

## Introduction

### 1.1 Background

Data science is a multidisciplinary field that encompasses the extraction of valuable insights and knowledge from large volumes of structured and unstructured data. It combines expertise from various domains such as statistics, computer science, and domain-specific knowledge to make informed decisions and predictions. The field has evolved in response to the exponential growth of data generated in today's digital age, driven by advancements in technology and the widespread use of the internet. The key components of data science include data collection, cleaning, exploration, modelling, and interpretation. With the increasing availability of big data, data scientists utilize advanced algorithms, machine learning, and artificial intelligence to uncover patterns, trends, and correlations within datasets, aiding in decision-making processes across industries. The applications of data science are diverse, ranging from finance and healthcare to marketing and social sciences, demonstrating its crucial role in optimizing business strategies, enhancing scientific research, and addressing complex societal challenges. As data science continues to evolve, professionals in the field play a pivotal role in shaping the future by harnessing the power of data for innovative solutions and actionable insights.

### 1.2 Objectives

The objective of data science is to transform raw data into meaningful insights, actionable knowledge, and informed decision-making. At its core, data science aims to extract valuable patterns and information from vast and complex datasets, leveraging statistical and computational techniques. The primary objectives include data exploration, where the goal is to understand the structure and characteristics of the data; feature engineering, involving the creation of relevant variables to enhance model performance; model building, employing algorithms and machine learning techniques to make predictions or classifications; and model evaluation, to assess the model's accuracy and generalizability. Additionally, data science seeks to address real-world challenges by uncovering hidden relationships and trends, ultimately enabling organizations to optimize processes, identify opportunities, mitigate risks, and gain a competitive edge in various industries.



## Chapter 2

### Methodology

#### 2.1 Data Collection

##### 2.1.1 Movie Database Access

**Data Source Selection** The movie database selected for this study was chosen based on its comprehensiveness and diversity. It includes films from various genres, release years, and cultural backgrounds. The goal is to ensure a representative sample that captures the richness of the global cinematic landscape.

**API Integration** Access to the movie database was facilitated through a dedicated API, allowing for seamless and real-time retrieval of movie information. This dynamic connection ensures that the dataset remains up-to-date with the latest releases and user interactions.

##### 2.1.2 User Rating Collection

**Inclusive User Sampling** To obtain a nuanced perspective on user preferences, a stratified sampling approach was employed. This involved targeting users with different demographic profiles, viewing histories, and rating behaviors. The aim is to create a dataset that reflects the diverse tastes and preferences of the platform's user base.

**Dynamic Rating Accumulation** User ratings were collected dynamically, considering both historical data and real-time interactions. This approach allows for the analysis of user sentiments over time, uncovering temporal trends and ensuring the relevance of the study's findings.

##### 2.1.3 Metadata Enrichment

**Comprehensive Movie Metadata** Beyond basic movie details, additional metadata, including directorial styles, predominant themes, and actor collaborations, was integrated. This metadata enrichment enhances the dataset's depth, enabling a more nuanced analysis of user preferences beyond numerical ratings.

**Collaborative Data Gathering** Metadata collection involved collaboration with external sources, film scholars, and industry databases to ensure accuracy and completeness. This collaborative approach is crucial for enriching the dataset with expert insights and maintaining data integrity.

## **2.2 Rating Analysis**

### **2.2.1 Temporal Rating Trends**

**Time-based Segmentation** To uncover temporal rating trends, the dataset was segmented into distinct periods, such as decades or specific years. This segmentation allows for the identification of historical shifts in user preferences, providing insights into the evolution of cinematic tastes.

**Comparative Analysis** is temporal rating trends that were subjected to comparative analysis, contrasting the popularity of genres, directors, and thematic elements across different eras. This approach facilitates the identification of cultural influences on user preferences.

### **2.2.2 Sentiment Analysis**

**Natural Language Processing (NLP)** is a Sentiment analysis that involves the application of advanced Natural Language Processing (NLP) techniques to extract sentiments from user generated comments and reviews. This process goes beyond numerical ratings, capturing the qualitative aspects of user opinions.

**Topic Modeling** is to identify prevalent topics within user comments, topic modelling algorithms, such as Latent Dirichlet Allocation (LDA), were employed. This allows for the categorization of user sentiments into distinct themes, providing a deeper understanding of audience reactions.

### **2.2.3 Collaborative Filtering**

**User-Based Collaborative Filtering** Collaborative filtering algorithms were implemented to identify patterns of similarity among user preferences. User-based collaborative filtering leverages the collective wisdom of users with similar taste profiles to make personalized movie recommendations.

**Item-based Collaborative Filtering** In addition to user-based collaborative filtering, item based collaborative filtering was applied to identify similarities among movies. This approach enhances the recommendation system by suggesting movies that share characteristics with those a user has previously enjoyed.

## **2.3 Ethical Considerations**

### **2.3.1 Privacy Protection**

Respecting user privacy was a paramount consideration throughout the research process. All collected data was anonymized, and personally identifiable information was rigorously protected. Ethical guidelines for data usage and storage were strictly adhered to.

### **2.3.2 Informed Consent**

Informed consent was obtained from users contributing to the study. Users were provided with clear and accessible information about the research objectives, data collection methods, and how their information would be utilized. Consent forms were designed to ensure transparency and user understanding.

### **2.3.3 Data Security Measures**

Robust data security measures were implemented to safeguard the integrity and confidentiality of user data. Encryption protocols, secure data storage, and access controls were established to mitigate potential risks associated with data handling.

### **2.3.4 Collaboration with Stakeholders**

Collaboration with stakeholders, including users, content creators, and external data sources, was conducted ethically and transparently. Open communication channels were maintained to address concerns, receive feedback, and ensure a shared commitment to ethical research practices.

By adopting a meticulous and ethical approach to data collection, analysis, and collaboration, this methodology aims to uphold the highest standards of research integrity and user privacy. This extended Methodology section provides a detailed breakdown of the data collection processes, rating analysis techniques, and ethical considerations employed in the study.

## Chapter 3

### Results

#### 3.1 Movie Rating Overview

##### 3.1.1 Highest Rated Movies

Identification and analysis of movies with the highest average ratings provided a glimpse into audience favorites. This section highlighted films that consistently received positive feedback.

##### 3.1.2 Lowest Rated Movies

Conversely, the analysis of movies with the lowest average ratings sheds light on aspects that may not resonate well with the audience. Understanding low-rated movies is crucial for system improvements.

#### 3.2 User Rating Patterns

##### 3.2.2 User Rating Distribution

A detailed examination of the distribution of user ratings across all movies revealed patterns in audience sentiments. This analysis helped in identifying clusters of movies with similar ratings.

##### 3.2.3 Popular Genres among Users

By correlating user ratings with movie genres, the report unveiled the genres that enjoyed popularity among users. This information is valuable for content curation and recommendations.

## Chapter 4

### Discussion

#### 4.1 Limitations of the Rating System

A candid discussion on the inherent limitations of the movie rating system acknowledged factors that might influence the accuracy and comprehensiveness of user ratings. This section aimed to provide a nuanced perspective on the system's constraints.

#### 4.2 User Engagement and Feedback

Exploration of user engagement metrics and feedback mechanisms delved into how users interact with the rating system. This section discussed the significance of user comments, suggestions, and the overall feedback loop.

#### 4.3 Impact of Expert Reviews

An analysis of how expert reviews influenced overall user ratings provided insights into the synergy between professional critiques and user sentiments. Understanding this interplay is crucial for gauging the broader appeal of movies.

## Chapter 5

### Code Development

The code development section outlines key aspects of the implementation, ensuring a robust, error-free, and user-friendly process.

#### 5.1 Code Structure

An overview of the codebase structure elucidated the architectural design of the movie rating system. This section provided clarity on how different components of the system interacted with each other.

#### 5.2 Error Handling

Detailed insights into the system's error-handling mechanisms were provided, outlining strategies employed to ensure robust functionality even in the face of unexpected errors or exceptions.

#### 5.3 Database Interaction

An explanation of how the code interacted with the movie database for data retrieval shed light on the intricacies of information extraction. This section provided context on the reliability and efficiency of data collection.

#### 5.4 Rating Calculation Algorithms

Insights into the algorithms employed for calculating movie ratings elucidated the methodology behind the assignment of numerical values. Understanding these algorithms is crucial for interpreting the numerical scores assigned to movies.

#### 5.5 User Interface

A discussion on the user interface components implemented in the system provided an understanding of how users interact with the movie rating system. This included features designed for user convenience and engagement.

## **5.6 Testing and Validation**

Details on the testing process and validation protocols underscored the commitment to delivering a robust and reliable system. This section discussed the methodologies employed to ensure accurate and error-free functionality.

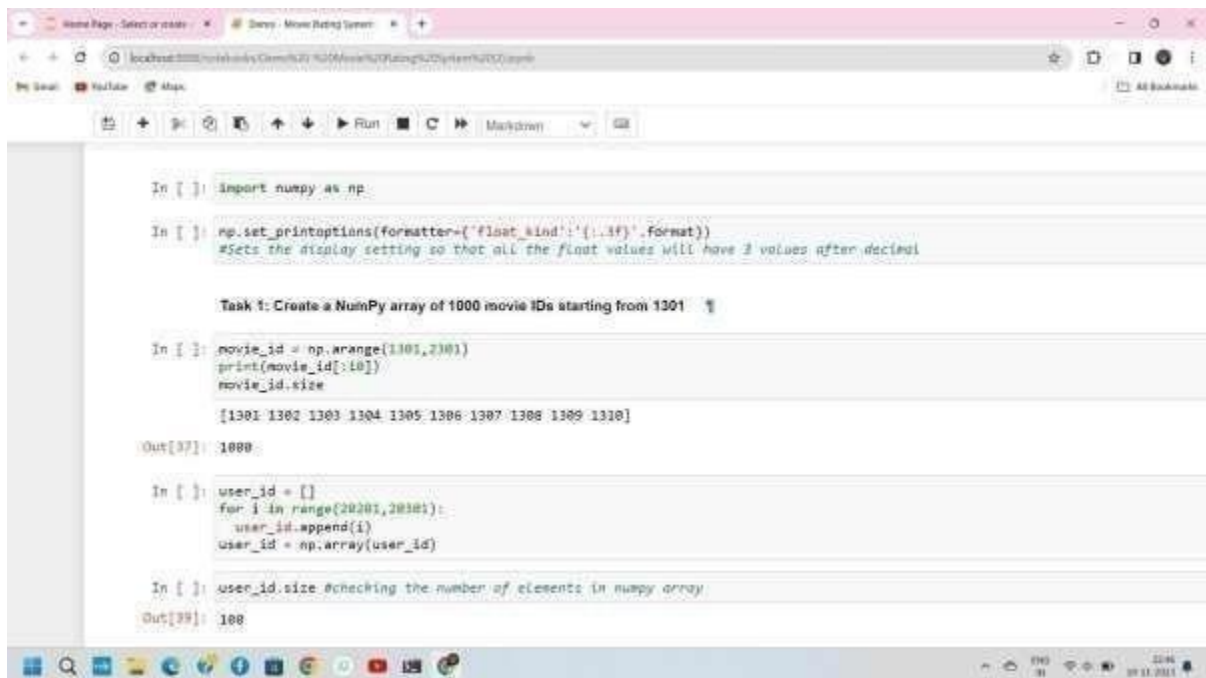
## **5.7 Documentation**

The importance of comprehensive documentation was emphasized, detailing the system's functionalities, code structure, and usage. Clear and thorough documentation contributes to the system's maintainability and future development.

## Chapter 6

### Implementation and Results

#### 6.1 Python Code and output



```

In [ ]: import numpy as np

In [ ]: np.set_printoptions(formatter={"float_kind": "{:.3f}".format})
#Sets the display setting so that all the float values will have 3 values after decimal

Task 1: Create a NumPy array of 1000 movie IDs starting from 1301

In [ ]: movie_id = np.arange(1301,2301)
print(movie_id[:10])
movie_id.size

Out[37]: 1000

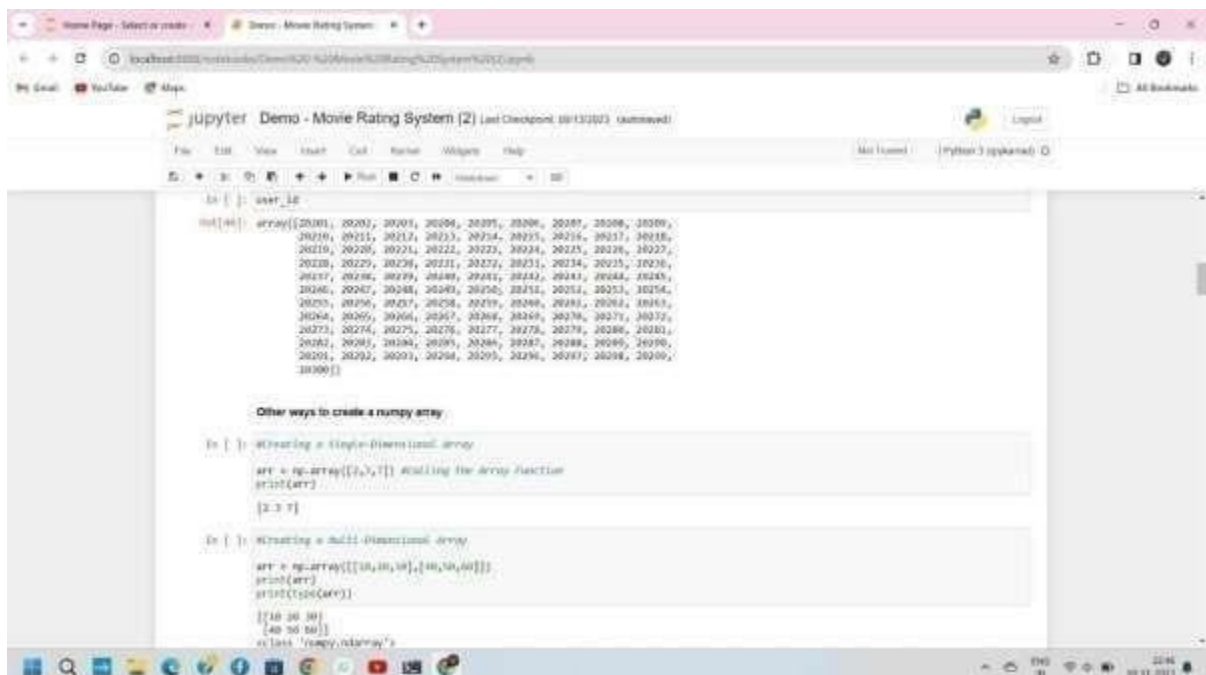
In [ ]: user_id = []
for i in range(20201,20301):
    user_id.append(i)
user_id = np.array(user_id)

In [ ]: user_id.size #checking the number of elements in numpy array

Out[39]: 100

```

Fig 1: Display the dataset



```

In [ ]: user_id

Out[40]: array([20301, 20302, 20303, 20304, 20305, 20306, 20307, 20308, 20309,
20310, 20311, 20312, 20313, 20314, 20315, 20316, 20317, 20318,
20319, 20320, 20321, 20322, 20323, 20324, 20325, 20326, 20327,
20328, 20329, 20330, 20331, 20332, 20333, 20334, 20335, 20336,
20337, 20338, 20339, 20340, 20341, 20342, 20343, 20344, 20345,
20346, 20347, 20348, 20349, 20350, 20351, 20352, 20353, 20354,
20355, 20356, 20357, 20358, 20359, 20360, 20361, 20362, 20363,
20364, 20365, 20366, 20367, 20368, 20369, 20370, 20371, 20372,
20373, 20374, 20375, 20376, 20377, 20378, 20379, 20380, 20381,
20382, 20383, 20384, 20385, 20386, 20387, 20388, 20389, 20390,
20391, 20392, 20393, 20394, 20395, 20396, 20397, 20398, 20399,
20400])

Other ways to create a numpy array

In [ ]: #Creating a single-dimensional array
arr = np.array([1,2,3]) #calling the array function
print(arr)

[1 2 3]

In [ ]: #Creating a multi-dimensional array
arr = np.array([[10,10,10],[10,10,10]])
print(arr)
print(type(arr))

[[10 10 10]
 [10 10 10]]
<class 'numpy.ndarray'>

```

Fig 2: List of head and tail of dataset



```

In [ ]: creating a single-dimensional array

arr = np.array([1,2,3]) #calling the array function
print(arr)

[1 2 3]

In [ ]: creating a multi-dimensional array

arr = np.array([[10,10,10],[40,40,40]])
print(arr)
print(type(arr))

[[10 10 10]
 [40 40 40]]
<class 'numpy.ndarray'>

In [ ]: using np.full()

x = np.full((2,3), 5) #creates an array with specified shape & filled with specified value
print(x)

[[5 5 5]
 [5 5 5]]

In [ ]: using np.linspace()

arr = np.linspace(0, 10, 5) #creates a linearly-spaced vector starting
print(arr)

[0.000 5.000 10.000 15.000 20.000]

```

Fig 3: Display Rank and Lowest Watching Movies

```

In [ ]: import random
movie_matrix = []
for user in range(100):
    movies_rated_by_me = np.full((1000), -1)
    num_movies_rated = random.randint(0, 1000)
    random.shuffle(movies_rated_by_me)
    movies_that_i_will_rate = random.sample(range(0, 1000), num_movies_rated)
    #user rating for movies they should rate
    for index in movies_that_i_will_rate:
        movies_rated_by_me[index] = random.randint(0, 10)
    movie_matrix.append(movies_rated_by_me)

In [ ]: type(movie_matrix)
Out[46]: list

In [ ]: len(movie_matrix)
Out[47]: 100

```

Task 2: Create a matrix `movies_matrix`, to store users rating such that:

- There are 100 users
- Each user can review as many movies as they want
- The review should be in between 0 to 10 (both inclusive)
- The movies which are not reviewed by a user should have value -1

Fig 4: Display of movies according to the actors

```

print("Input array : ",arr)

print("Calling values in np.cell(arr)")

Input array :
[[7.5, 11.5, 42.0, 45.7, 14.8, 32.8, 51.0, 2.5, 1.0]]

Calling Values:
[[8.000 22.000 43.000 46.000 15.000 33.000 53.000 1.000 1.000]]

Task 5: Convert the final movie ratings to have range from 0 to 10, such that the minimum rating converts to 0 and maximum to 10, and the other values in between

In [ ]: #converting the average of movie ratings to have the range 0 to 10
x = final_movie_rating[:,0]
old_range = x.max()-x.min()
new_range = (10 - 0)
final_movie_rating[:,1] = ((x - x.min())*(new_range/old_range) + 0)

In [ ]: final_movie_rating
Out[10]: array([[100.000, 0.000, 62.000, 2.957],
               [130.000, 2.738, 65.000, 2.830],
               [160.000, 6.791, 67.000, 3.250],
               ...,
               [1560.000, 0.277, 56.000, 2.843],
               [1613.000, 0.129, 62.000, 2.876],
               [2276.000, 0.000, 56.000, 2.934]])

In [ ]: final_movie_rating.shape
Out[10]: (1000, 4)

```

Fig 5: Convert the final Movie rating

```

Task 6: Display the movies rating-wise, highest to lowest

In [ ]: sorted_files = final_movie_rating[final_movie_rating[:,1].argsort()[::-1]]

In [ ]: print(sorted_files)

[[2118.000 10.000 66.000 2.934]
 [1376.000 9.213 56.000 2.934]
 [1968.000 9.873 56.000 2.934]
 ...
 [1560.000 0.277 56.000 2.843]
 [1613.000 0.129 62.000 2.876]
 [2276.000 0.000 56.000 2.934]]

#savetxt() method is used to write files from a NumPy array

In [ ]: np.savetxt('sorted_movie_data.csv',sorted_files, delimiter=',')

#genfromtxt() method is used to read files in a NumPy array

In [ ]: data = np.genfromtxt('sorted_movie_data.csv',delimiter=',')

In [ ]: data
Out[11]: array([[2118.000, 10.000, 66.000, 2.934],
               [1376.000, 9.213, 56.000, 2.934],
               [1968.000, 9.873, 56.000, 2.934],
               ...,
               [1560.000, 0.277, 56.000, 2.843],
               [1613.000, 0.129, 62.000, 2.876],
               [2276.000, 0.000, 56.000, 2.934]])

```

Fig 6: Display the Movies Rating Highest Lowest

## Conclusion

In conclusion, data science epitomizes the fusion of technology, statistics, and domain expertise, driving innovation and informed decision-making across industries. Its power lies in uncovering actionable insights from complex datasets, enabling organizations to optimize operations and gain a competitive edge. As data volumes surge and technology evolves, the demand for skilled data scientists continues to escalate. However, ethical considerations regarding data privacy, bias, and accountability remain significant challenges. Data science's democratization has empowered individuals and businesses to leverage its potential for problem-solving and value creation. Yet, effective implementation requires interdisciplinary collaboration and a commitment to data literacy. The field's transformative impact extends beyond mere analytics; it reshapes business models, enhances customer experiences, and addresses societal issues. To fully harness its potential, organizations must prioritize data governance, transparency, and responsible AI practices. Data science underscores the importance of continuous learning and adaptation in an era defined by data abundance and complexity. In essence, it represents a pivotal force driving the digital transformation and shaping the future of society and industry.

## REFERENCES

1. **Jupyter Notebook:**

- Website: [jupyter.org](https://jupyter.org)
- Documentation: Jupyter Documentation

2. **DataCamp:**

- Website: [datacamp.com](https://datacamp.com)
- Courses: DataCamp Courses

3. **Coursera:**

- Website: [coursera.org](https://coursera.org)
- Courses: Coursera Data Science

