

**WiDS ‘22 - ‘23 Final Documentation**

**<** **2-Data analytics using Python>**

**<Aditya>**

|  |  |  |
| --- | --- | --- |
| **Team Member Name** | **Roll Number** | **Email-Id** |
| Sujit Narayan Singh | 22M0603 | 22m0603@iitb.ac.in |
|  |  |  |
|  |  |  |

**Introduction to Problem Statement**

|  |
| --- |
| DATASET-1  India GDP Data visualization and finding insights  DATASET-2  Netflix is one of the most popular media and video streaming platforms. They have over 8000 movies or tv shows available on their platform, as of mid-2021, they have over 200M Subscribers globally. This tabular dataset consists of listings of all the movies and tv shows available on Netflix, along with details such as - cast, directors, ratings, release year, duration, etc. **Interesting Task Ideas**  * Understanding what content is available in different countries * Identifying similar content by matching text-based features * Network analysis of Actors / Directors and find interesting insights. * Does Netflix has more focus on TV Shows than movies in recent years.   DATASET-3  In dataset-3 with the code (on linear regression), fill in the gaps in code. |

**Existing Resources**

|  |
| --- |
| **Python Basics:** (till 1:44:40[modules])  [Data Analysis with Python for Excel Users - Full Course](https://youtu.be/WcDaZ67TVRo)  **NumPy:**  [Python NumPy Tutorial for Beginners](https://youtu.be/QUT1VHiLmmI)  **Pandas:**  <https://www.youtube.com/playlist?list=PL-osiE80TeTsWmV9i9c58mdDCSskIFdDS>    <https://www.hackerearth.com/practice/machine-learning/data-manipulation-visualisation-r-python/tutorial-data-manipulation-numpy-pandas-python/tutorial/>  **Matplotlib:**  [Python Plotting Tutorial w/ Matplotlib & Pandas (Line Graph, Histogram, Pie Chart, Box & Whiskers)](https://www.youtube.com/watch?v=0P7QnIQDBJY&t=58s)  **Seaborn:**  [Seaborn Tutorial : Seaborn Full Course](https://www.youtube.com/watch?v=6GUZXDef2U0)    **Must Do**:  (21-30) **Code along the way**  [Python for Data Analysis - YouTube](https://www.youtube.com/playlist?list=PLiC1doDIe9rCYWmH9wIEYEXXaJ4KAi3jc)    **Documentations:**  <https://numpy.org/doc/stable/user/absolute_beginners.html>  <https://pandas.pydata.org/docs/getting_started/intro_tutorials/index.html>  <https://matplotlib.org/stable/tutorials/introductory/quick_start.html>  <https://seaborn.pydata.org/tutorial.html> |

**Proposed Solution**

|  |
| --- |
|  |

**Methodology & Progress (Mention the work done week-wise)**

|  |
| --- |
| 1. Basics of python, NumPy. 2. Python libraries (Pandas, Matplotlib, Seaborn) & exercises 3. Python exercises, Working with Datasets. 4. Data cleaning and Data visualization, Manipulation of structured data (tables). input/output excel files. 5. More datasets and visualization techniques and getting business insights. 6. Python for data analysis: Statistics |

**Results**

|  |
| --- |
| https://github.com/SujitNarayanSingh/PROJECT |

**Learning Value**

|  |
| --- |
|  |

**Tech-stack Used**

|  |
| --- |
|  |

**Suggestions for others**

|  |
| --- |
|  |

**Contribution by each Team Member**

|  |
| --- |
|  |

**References and Citations**

|  |
| --- |
|  |