

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

**<Statistics and Data Analytics using Python Project UID: 2>**

**<Aditya Prakash>**

**Introduction to Problem Statement**

|  |
| --- |
| Statistics and Data analytics using Python:  The project would mainly focus on the implementation of statistics and analytics concepts using python which provides a good base for learning machine learning algorithms. Starts from learning basic libraries NumPy, pandas matplotlib, etc., then move to the concepts of Sampling and sampling distributions, hypothesis testing, and regressions, etc. then apply these to different datasets in order to draw out meaningful, actionable insights. |

**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)  (21-30) **Code along the way** (required in the final report submission)  [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**

|  |
| --- |
| Used python libraries like NumPy, pandas, matplotlib, and seaborn for analysis of 3+ datasets and get useful information |

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

|  |
| --- |
| Week 1: Learnt NumPy, pandas, and matplotlib and solved one dataset and a few assignments  Week 2: Learnt seaborn, solved datasets 2 and 3  Week 3&4: Sampling and sampling distributions, hypothesis testing, and regressions |

**Results**

|  |
| --- |
| https://github.com/PranitaR12/Data\_Analytics.git |

**Learning Value**

|  |
| --- |
| Learnt python libraries like numPy, pandas, matplotlib and Sampling and sampling distributions, hypothesis testing, and regressions, etc |

**Tech-stack Used**

|  |
| --- |
| Jupyter notebook |