

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

### **2 - Data analytics using Python**

**Aditya Prakash**

|  |  |  |
| --- | --- | --- |
| **Team Member Name** | **Roll Number** | **Email-Id** |
| Merugu Rakesh | 204100006 | 204100006@iitb.ac.in |

**Introduction to Problem Statement**

|  |
| --- |
| Performing linear regression on the train data and plotting. Checking the graphs and correcting for any unexpected higher values. |

**Existing Resources**

|  |
| --- |
| NA |

**Proposed Solution**

|  |
| --- |
| Fit the data and plot and observe for any discrepancies. Run a for loop to find the index of highest value and remove the element and plot the data again. |

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

|  |
| --- |
| Week-5   1. Understanding on Regression fit 2. Introduction to Descriptive statics, Probability distributions and confidence intervals |

**Results**

|  |
| --- |
| 1. Plotting the data   C:\Users\Mr\AppData\Local\Microsoft\Windows\INetCache\Content.MSO\27D44C66.tmp  2. test data  C:\Users\Mr\AppData\Local\Microsoft\Windows\INetCache\Content.MSO\77047D64.tmp   1. Plotting after correcting the highest value   C:\Users\Mr\AppData\Local\Microsoft\Windows\INetCache\Content.MSO\99001D12.tmp  4. test data  C:\Users\Mr\AppData\Local\Microsoft\Windows\INetCache\Content.MSO\58F0EF0.tmp |

**Learning Value**

|  |
| --- |
| **Observed that the data plotting was sensitive the magnitude. Application of previous knowledge to find the index of desired number in the data list/frame** |

**Suggestions for others**

|  |
| --- |
| NA |

**References and Citations**

|  |
| --- |
| 1. Python for Data Analysis: Descriptive Statistics   <https://www.youtube.com/playlist?list=PLiC1doDIe9rCYWmH9wIEYEXXaJ4KAi3jc> |