



## **PROJECT LIST AND RULES** **[DOMAIN: DATA SCIENCE]**

### **PHASE 1 PROJECTS**

**Start Date:** 25 January 2024

**Deadline:** 10 February 2024

1. Create a csv dataset using python , pandas and any public api
2. Clean the dataset replace missing values, remove outliers etc. Dataset :-  
<https://docs.google.com/spreadsheets/d/e/2PACX-1vTSS-TcErkXNk8KB0AlijhitwetxeHD2M3R0HJl2QPMAyFq0fxFX4PFKnzAWLDnratIz67DNL6GsZnV/pub?output=csv>

### **PHASE 2 PROJECTS**

**Start Date:** 10 February 2024

**Deadline:** 25 February 2024

3. Analyze the dataset and create graphs using seaborn and matplotlib.Dataset :-  
<https://docs.google.com/spreadsheets/d/e/2PACX-1vTSS-TcErkXNk8KB0AlijhitwetxeHD2M3R0HJl2QPMAyFq0fxFX4PFKnzAWLDnratIz67DNL6GsZnV/pub?output=csv>

4. Train a simple linear regressing model on dataset and predict the output.

Train dataset :- [https://docs.google.com/spreadsheets/d/e/2PACX-1vRTK2NvcndgPX41Czu6Ft2Ho\\_nE-z50BgTqdzwFW0rsJ2nvyNLe2DoIg1COzUbgw80oaRBjfy5-WtFk/pubhtml](https://docs.google.com/spreadsheets/d/e/2PACX-1vRTK2NvcndgPX41Czu6Ft2Ho_nE-z50BgTqdzwFW0rsJ2nvyNLe2DoIg1COzUbgw80oaRBjfy5-WtFk/pubhtml)

Test dataset :- <https://docs.google.com/spreadsheets/d/e/2PACX-1vRyvZ7lknwiSghK9aen1SaTEYoN3JS40rrGLpcyrsVZy1tB2T4gn6Y3-cdzPUFCPMmmqREWefW3kl4/pubhtml>

## **RULES**

- Maintain a separate GitHub Repository (COB #Domain Name)
- Upload all project source code to that Repository
- Upload the Task Videos to LinkedIn and Tag @CodesOnBytes
- Complete at least one task from each Phase 1 & Phase 2 projects to get eligible to apply for certificate.
- Complete all the given tasks in Phase 1 & Phase 2 projects to get eligible to apply for both certificate and Letter of Recommendation.
- If we discover that your code contains plagiarism, we will fire you by way of the internship.
- Inactiveness during your tenure will lead to terminate from internship.

# THANK YOU