Cloud Counselage

GIFT A CAREER FOUNDATION

Industry-Academia Community

INTERNSHIP PROGRAM

Live Project POC Document

DATA ANALYTICS

Created By:	#Domain Lead Name#	**Approved By:**	Harshada Topale
Created On:	DD-MMM-YYYY	**Approved On:**	DD-MMM-YYYY

Page left blank intentionally

INDEX

Cloud Counselage

General Instructions for using the Live Project POC Document

- This template and the subsequent document created using this template is a confidential document and is the intellectual property of Cloud Counselage Pvt. Ltd. Circulating it outside of the organisation without the consent of Cloud Counselage Pvt. Ltd. is the breach of company policies and will lead to legal actions
- The POC document for the subsequent technology forms the basis of the '**Live Project - Problem Statement'** to be shared with the '**IAC - Internship Program (IP)'** interns
- The POC done should be added as appendices in this document
- The **text between the hashmarks (#) is to be replaced** by relevant text
- Please **remove the yellow highlight on the Text** between the hashmark (#). This is done to help you notice the text to be changed/replaced
- The text in *italics*highlighted in grey is just for reference and should be removed after adding the relevant text

1 PURPOSE

The purpose of this POC Document is to provide the concept and methodology to follow for theimplementation and evaluation of the DATA ANALYTICS domain project for IP Interns.

2 SCOPE

Suitable for interns who are familiar with the fundamentals of Python,PowerBI & Tableau and who aim towards applying these concepts in application of analysing data and drawing insights from them. Moderate programming knowledge related to data cleaning and manipulation of data expected.

3 PROBLEM OBJECTIVES

The Data Analytics Project aims to fulfil the following objectivesfor IP interns

- Getting knowledge of analysing the data using the above mentioned tools and the experience of implementing the same.
- Drawing out insights from the given data and representing them in form charts and othervisualisations.

4 THE PROBLEM STATEMENT

We aim to conduct a comprehensive analysis of our students to gain insights about their academic performance, career outcomes, and factors influencing their success. The institute has collected a dataset containing various attributes for each student.

The objective of this data analysis project is to gain comprehensive insights into the academic performance, career outcomes, and factors influencing the success of the educational institute's students.

5 OPERATING ENVIRONMENT

The interns should use either python, PowerBI or Tableau as the base tools. They have to analyze the data, clean, visualize and perform various operations on the given data in these tools only. The code and output submissions format is standard so development can be done on the platform of choice.

6 DESIGN & IMPLEMENTATION DETAILS

The institute has collected a dataset containing various attributes for each student. These attributes include:

- **First name and last name:** These attributes identify the student individually.
- **Quantity (number of courses completed):** This attribute indicates the number of courses each student has completed during their time at the institute.
- **E-mail (of all students):** This contains the mail-id of all the students.
- **College name:** This attribute specifies the college from which each student graduates.
- **Year of graduation:** This attribute denotes the year in which each student completed their graduation.
- **City:** This attribute captures the city of residence or hometown for each student.

- **GPA (Grade Point Average):** This attribute represents the academic performance of each student based on their GPA.
- **Experience with Python programming (in months):** This attribute indicates the level of experience each student has with Python programming.
- **Family income:** This attribute denotes the income level of each student's family.
- **Expected Salary:** This attribute indicates whether each student is currently employed or unemployed.
- **Leadership skills:** The students possess leadership skills or not.

THE RESULTS FOR THE FOLLOWING QUESTIONS ARE EXPECTED:-

BASIC QUESTIONS:

1. How many students are included in the dataset?
2. What is the average GPA of the students?
3. What is the distribution of students across different graduation years?
4. What is the distribution of student's experience with Python programming?
5. What is the average family income of the student?
6. How does the GPA vary among different colleges? (code using python & show top 5 among them.)
7. Are there any outliers in the quantity (number of courses completed) attribute?
8. What is the average GPA for student from each city?
9. Can we identify any relationship between family income and GPA?

MODERATE QUESTIONS:

10. How many students from various cities.(solve using data visualisation tool).
11. How does the expected salary vary based on factors like GPA, Family income, months of experience in python language?
12. Are there any industry-specific trends in student interest? Which industries tend to attract more student from specific fields of study?
13. Do students who are having leadership positions during their college years tend to have higher GPAs or better expected salary?
14. Is there a correlation between leadership skills and expected salary of the students? (solve using python).
15. How many students are graduating by the end of 2024?
16. Which marketing effects better in gaining attention from the students?
17. Find the total number of students who attended the events related to Data Science?(from all Data Science related courses.)

7 COMPLETION

In order to complete the project, the intern will have to satisfy the following:

- Analyze the Data completely by understanding the necessity of each column.
- Modify the data wherever required to get better results.
- Find the solution to each of the above mentioned question using the particular tool wherever mentioned.
- Present the results in understandable Industry –Standard format.

8 STEPS &TIMELINE

A weekly timeline is devised as follows for the project duration based on the internship duration of 6/8/12 weeks –

1. Obtain the Problem Statement and study it in detail. Clear any doubts or confusion.
2. Evaluate the strategies to use and the requirements for the project.
3. Research and train on tools and procedures as decided upon in the previous week.
4. Commence work on the system.
5. Continue work on the project.
6. Conduct testing on the completed system.
7. Research and add future improvements and suggestions into the report.
8. Compile the report and submit the report and code for evaluation.

9 EVALUATION

For the evaluation following deliverables and weightage is to be considered

Deliverables	**Weightage**
Project Charter	5%
SRS Document	10%
Project Plan	10%
Design: Userflow&Wireframe	20%
Design: Mockup (Only for UI/UX)	50%
Development	40%
Testing (not for UI/UX)	10%
Video	5%

The projectwill be checked for methodology, approach, tools used, techniques and procedures applied,effectiveness of observations, valid improvement suggestions, and completeness. GitLab will beused for automated code evaluation and accuracy checking.