

# DATA SCIENCE

---



# ABOUT COMPANY

---

**CODSOFT** is a vibrant and diverse community that brings together individuals with similar objectives and ultimate goals. Our main focus is on creating opportunities that span various areas, including leadership development, learning, student engagement, and fostering shared interests.

We believe in the power of leadership and its ability to drive positive change. That's why we provide platforms and resources for our community members to develop their leadership skills. Through mentorship programs, workshops, and collaborative projects, we empower individuals to take on leadership roles and make a difference in their respective fields.

# INSTRUCTIONS

- Update your LinkedIn profiles
- For the Data Science internship, you will need to complete at least 3 tasks for successful completion of the internship.
- Maintain a separate GitHub repository(name as CODSOFT for all the tasks and share the link of the GitHub repo in the task submission form(it will be given later through email).
- You can refer to online resources such as Google Search and read tutorials. Watch videos(For Help).

# SUBMISSION

- A TASK SUBMISSION FORM will be shared later through email . Till then please continue your task.
- A video need to be created to showcase your work, a demo of your effort.
- For the Data Science internship, you will need to complete at least 3 tasks for successful completion of the internship.
- The video can be hosted on LinkedIn for proof of your work and to build credibility among your peers. You can tag CODSOFT in such posts.
- Please add #codsoft in each of your task video postings on LinkedIn, Additionally, you can also add hashtags such as #internship #webdevelopment. for more reach and visibility

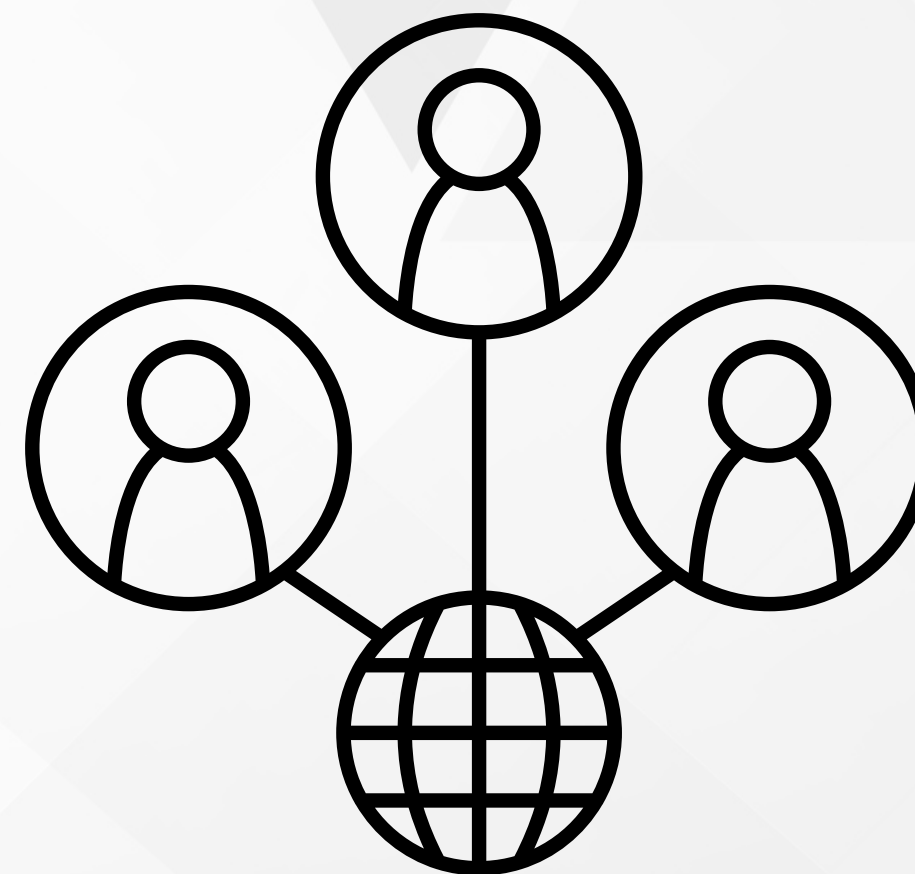
# ABOUT THE INTERNSHIP



**COMPLETION  
CERTIFICATE**



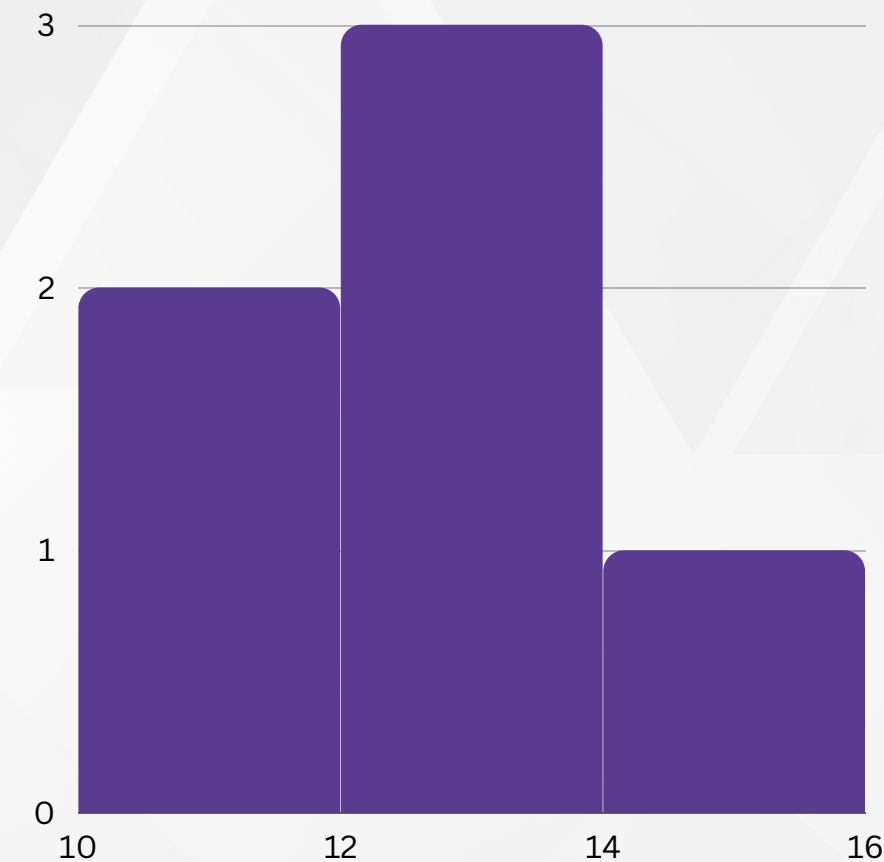
**PLACEMENT  
SUPPORT**



**NETWORK  
OPPORTUNITY**

# DATA SCIENCE

---



**For the Data Science internship, you will need to complete at least 3 tasks for successful completion of the internship.**

# TASK 1

## TITANIC SURVIVAL PREDICTION

- Use the Titanic dataset to build a model that predicts whether a passenger on the Titanic survived or not. This is a classic beginner project with readily available data.
- The dataset typically used for this project contains information about individual passengers, such as their age, gender, ticket class, fare, cabin, and whether or not they survived.

**[DOWNLOAD DATASET FROM HERE](#)**



## TASK 2

### MOVIE RATING PREDICTION WITH PYTHON

- Build a model that predicts the rating of a movie based on features like genre, director, and actors. You can use regression techniques to tackle this problem.
- The goal is to analyze historical movie data and develop a model that accurately estimates the rating given to a movie by users or critics.
- Movie Rating Prediction project enables you to explore data analysis, preprocessing, feature engineering, and machine learning modeling techniques. It provides insights into the factors that influence movie ratings and allows you to build a model that can estimate the ratings of movies accurately.

**DOWNLOAD DATASET FROM [HERE](#)**



# TASK 3

## IRIS FLOWER CLASSIFICATION

- The Iris flower dataset consists of three species: setosa, versicolor, and virginica. These species can be distinguished based on their measurements. Now, imagine that you have the measurements of Iris flowers categorized by their respective species. Your objective is to train a machine learning model that can learn from these measurements and accurately classify the Iris flowers into their respective species.
- Use the Iris dataset to develop a model that can classify iris flowers into different species based on their sepal and petal measurements. This dataset is widely used for introductory classification tasks.

**[DOWNLOAD DATASET FROM HERE](#)**

# TASK 4

## SALES PREDICTION USING PYTHON

- Sales prediction involves forecasting the amount of a product that customers will purchase, taking into account various factors such as advertising expenditure, target audience segmentation, and advertising platform selection.
- In businesses that offer products or services, the role of a Data Scientist is crucial for predicting future sales. They utilize machine learning techniques in Python to analyze and interpret data, allowing them to make informed decisions regarding advertising costs. By leveraging these predictions, businesses can optimize their advertising strategies and maximize sales potential. Let's embark on the journey of sales prediction using machine learning in Python.

**DOWNLOAD DATASET FROM [HERE](#)**

# TASK 5

## CREDIT CARD FRAUD DETECTION

- Build a machine learning model to identify fraudulent credit card transactions.
- Preprocess and normalize the transaction data, handle class imbalance issues, and split the dataset into training and testing sets.
- Train a classification algorithm, such as logistic regression or random forests, to classify transactions as fraudulent or genuine.
- Evaluate the model's performance using metrics like precision, recall, and F1-score, and consider techniques like oversampling or undersampling for improving results.

**[DOWNLOAD DATASET FROM HERE](#)**

# ASK US FOR HELP!

- THE PURPOSE OF THIS INTERNSHIP IS TO LEARN AND GROW
- We have no desire to dictate to you. It is entirely up to you whether you seek guidance or not.
- The given tasks may seem very easy or very difficult. We expect you to approach the tasks with professional diligence and give them the attention they deserve."

# GET SOCIAL WITH US



**[www.codsoft.in](http://www.codsoft.in)**



**[contact@codsoft.in](mailto:contact@codsoft.in)**



**CodSoft**