

Data Science Internship Assignment

LogicLeap Interns – Data Science Projects

Objective:

As a Data Science Intern at LogicLeap, you will work on practical and impactful data analysis problems using real-world datasets. This internship is designed to strengthen your analytical foundations, improve your problem-solving skills, and give you hands-on project experience.

General Instructions:

- ✓ Complete all 3 tasks within the internship period.
- ✓ Submit your work in a well-documented Jupyter Notebook or Python script.
- ✓ Include a short report (1–2 pages) summarizing:
 - ✓ Your approach for each task
 - ✓ Challenges faced

Key learnings and outcomes

Task 1: Data Wrangling & Exploratory Data Analysis (EDA)

Goal: Clean a messy dataset and perform a comprehensive exploratory analysis to uncover insights.

Dataset: Titanic Dataset

<https://www.kaggle.com/competitions/titanic/data>

1. Handle missing values and encode categorical variables.
2. Perform univariate and bivariate analysis.
3. Create visualizations (histograms, box plots, correlation heatmaps) to understand relationships between features and survival.
4. Summarize your key findings about factors influencing survival.

Task 2: Time Series Forecasting

Goal: Build a model to predict future values of a time series.

Dataset: Air Passengers Dataset

<https://www.kaggle.com/datasets/rakannimer/air-passengers>

- 1) Perform time series decomposition (trend, seasonality, residual).
- 2) Use models like ARIMA, SARIMA, or Prophet for forecasting.

- 3) Evaluate your model using metrics like MAE (Mean Absolute Error) or RMSE (Root Mean Squared Error).
- 4) Visualize the forecast against the actual data.

Task 3: A/B Testing Analysis

Goal: Analyze the results of an A/B test to determine if a new change is statistically significant.

Dataset: A/B Test Data (Example)

<https://www.kaggle.com/datasets/zhangluyuan/ab-testing>

1. Formulate null and alternative hypotheses.
2. Perform statistical testing (e.g., Chi-squared test for conversion rates, T-test for continuous metrics).
3. Calculate and interpret the p-value.
4. Provide a business recommendation based on your analysis.

■ Submission Guidelines

- ✓ Code: Submit clean, well-commented Python scripts or Jupyter Notebooks.
- ✓ Report: 1–2 pages covering approach, challenges, and outcomes.
- ✓ Deployment (Optional): Share a link if you created a dashboard (e.g., using Tableau, Power BI).