



Presented By: Certify Team

...

CERTIFY TECHNOLOGY

Innovating the Future with AI, App Development, and Full-Stack Solutions

...



Important Instructions for Level 3 Project (Data Analytics)

Dear Intern,

You have been assigned a Level 3 Data Analytics Task, which involves working on a real-world dataset to extract insights, clean data, perform analysis, and visualize results using appropriate tools. This is a great opportunity to enhance your skills in data handling, Python, SQL, statistics, and dashboard creation.

Important Note:

We have already provided the required dataset for this project. You are strictly instructed to use only the given dataset for all your tasks.

Before starting the analysis, you must perform data cleaning on the dataset.

This includes

- Handling missing or null values
- Removing duplicates
- Converting date formats
- Correcting data types

Only after the data is cleaned properly, proceed to perform the given tasks and answer the questions.

Clean and well-prepared data is mandatory before analysis. Submissions without cleaning steps will not be accepted.





Data Analytics Internship level 1

- Experience: Gained foundational knowledge in data collection, preprocessing, and analysis using statistical techniques.
- Impact: Assisted in analyzing datasets, generating insights, and creating visualizations to support decision-making.
- Skills Gained: Python, Pandas, NumPy, Data Visualization (Matplotlib, Seaborn), SQL, and Basic Statistical Analysis.



Choose any 3 of the following projects:

Task 1 - Handling Missing Data

- Load a dataset (CSV format) containing missing values.
- Identify and handle missing data using methods like imputation or removal.
- Provide visualizations to show the impact of missing data.

Task 2 - Data Transformation & Standardization

- Load a dataset and perform feature scaling (normalization/standardization).
- Convert categorical variables into numerical representations (e.g., one-hot encoding, label encoding).
- Ensure the transformed data is ready for analysis or modelling.





Task 3 - Exploratory Data Analysis (EDA)

- Select a dataset and perform basic exploratory data analysis.
- Generate summary statistics and visualize distributions (histograms, box plots, etc.).
- Identify correlations and trends using scatter plots and heatmaps.

Task 4 - Interactive Dashboard

- Use Python libraries like Plotly or Streamlit to create an interactive dashboard.
- Display key insights from a dataset with visualizations (bar charts, pie charts, line graphs, etc.).
- Ensure the dashboard allows user interaction (e.g., filtering data, selecting parameters).





Data Analytics Internship level 2

- Experience: Developed expertise in data collection, preprocessing, and exploratory data analysis (EDA). Applied statistical techniques to analyze structured and unstructured datasets. Worked with real-world data to extract meaningful insights.
- Impact: Contributed to data-driven decision-making by uncovering trends and insights. Conducted in-depth analysis of key business and social datasets. Created impactful data visualizations to communicate findings effectively.
- Skill Gained : Python, Pandas, NumPy, SQL for data analysis and manipulation. Matplotlib, Seaborn, Tableau, Power BI for data visualization. Experience with data wrangling, statistical modeling, and feature engineering.



Data Processing and Feature Engineering

Task 1 - Data Aggregation and Grouping

- Objective: Gain insights by grouping and summarizing key features in a dataset.
- Load a dataset and perform data aggregation based on key features.
- Use grouping methods to extract meaningful insights. Visualize the aggregated data effectively.
- Dataset:- You Want To use dataset as your wish. You want to mention/upload dataset in google form which you have used





Task 2 - Feature Engineering for Predictive Modeling

- Goal: Enhance predictive model performance by transforming raw data into meaningful features, handling missing values, encoding categorical variables, and optimizing feature selection for improved accuracy and efficiency.
- Select a dataset and create new meaningful features. Handle missing values, outliers, and categorical variables. Prepare the dataset for machine learning.
- Dataset:- You Want To use dataset as your wish. You want to mention/upload dataset in google form which y you have used





Basic Machine Learning Model

Choose one of the following projects:

Task 1 - Regression Analysis

- Goal: To analyze and preprocess data for regression or classification, leveraging data aggregation techniques to improve model performance, extract insights, and visualize key patterns effectively.
- Choose a dataset and build a regression model.
- Train the model and evaluate performance using metrics like RMSE and R². Visualize results and key insights.
- DataSet:- You Want To use dataset as your wish. You want to mention/upload dataset in google form which y you have used





Task 2 - Classification Task

- Goal: To apply data analytics techniques for building and evaluating classification models, leveraging machine learning algorithms to identify patterns, optimize predictions, and derive actionable insights using performance metrics like accuracy, precision, recall, and F1-score.
- Choose a dataset and apply a classification algorithm (e.g., Logistic Regression, Decision Tree, or Random Forest).
- Evaluate the model using accuracy, precision, recall, and F1-score. Present your findings with visualizations.
- DataSet:- You Want To use dataset as your wish. You want to mention/upload dataset in google form which you have used

Data Analytics Internship level 3



- **Experience:** Analyzed real-world datasets using Python, SQL, and Excel. Performed data cleaning, EDA, and created dashboards using Power BI/Tableau.
- **Impact:** Improved data interpretation and storytelling skills. Derived insights and presented findings through visual dashboards and reports.
- **Skill Gained:** Python (Pandas, NumPy), SQL, Excel, Power BI/Tableau, Data Visualization, and Basic Statistics.



Choose One Of The Following Projects:(Any1)

1. RFM Analysis for Customer Segmentation

Segment customers based on Recency, Frequency, and Monetary value.

Instructions::

- Use the transaction dates to compute how recently a customer purchased.
- Count how often each customer bought (Frequency).
- Calculate total spending (Monetary).
- Assign RFM scores and classify customers (e.g., Loyal, At Risk, Churned).

Tools

pandas, datetime, groupby, quantile scoring





2. *Price Range vs Delivery Status*

Analyze how product price affects delivery outcomes.

Instructions

- Create bins for Unit Price (Low, Medium, High).
- Compare delivery statuses (Delivered, Returned, In Transit) by price range.
- Calculate percentage distribution.

Tools

pandas cut(), groupby, crosstab or pivot_table





3. *Product Category vs Sales & Rating*

Identify top-performing product categories and check if sales volume aligns with customer ratings.

Instructions:

- Aggregate total Quantity sold and average Rating per category.
- Calculate correlation between Quantity and Rating.
- Plot bar and scatter charts for insights.

Tools:

pandas groupby, corr(), seaborn/matplotlib





4. Review Text Analysis

Extract insights from customer review text.

Instructions

- Tokenize review text and identify most frequent positive and negative words.
- Calculate average review length.
- Analyze if longer reviews correlate with higher/lower ratings.

Tools

pandas, string functions, text preprocessing



Submission Details:

- Jupyter Notebook (.ipynb) or Python Script (.py)
- File Format: Submission Instructions (Google Form)
- Complete data analytics workflow, including data preprocessing, exploratory data analysis (EDA), feature selection, model building, and evaluation.
- Include relevant data visualizations, graphs, and statistical summaries.
- Screenshot(s) of Output & Visualizations (.png or .jpg)



From logic to leadership—crafting scalable solutions, optimizing intelligence, and redefining the future of tech!

