



## Advanced Certification in Data Analytics- Project

**Case Study: Advanced Analysis of Data Analyst Job Listings in India**

**Objective:** To conduct a comprehensive analysis of data analyst job listings in India using Python, SQL, and Power BI. The study aims to uncover key trends in the job market, perform statistical and predictive analyses, and present actionable insights through dynamic visualizations. This case study will help stakeholders understand the current landscape of data analyst job opportunities and identify factors that influence job ratings and salaries.

**Dataset Link -**

<https://drive.google.com/file/d/10kFHfiwqG5N2yaFQvZZUH5wuYmjWgBmA/view?usp=sharing>

**Part 1: Data Cleaning and Analysis with Python**

**Scenario:**

You are a data analyst at a recruitment firm. Your task is to analyze a dataset of data analyst job listings in India to extract valuable insights. Start by cleaning and exploring the data.

**Tasks:**

**1. Data Cleaning and Preprocessing:**

- Handle missing values in the dataset appropriately.
- Normalize the 'salary' column to extract numerical values for base and max salary if not present.
- Standardize the 'experience' column to ensure a consistent format.

**2. Exploratory Data Analysis (EDA):**

- Perform EDA to find trends in job listings based on location, company, and experience level.
- Visualize the distribution of job ratings and reviews count.

**3. Statistical Analysis:**

- Conduct hypothesis testing to determine if there's a significant difference in ratings between jobs posted on 'Naukri' and 'iimjobs'.
- Perform a correlation analysis between job ratings and reviews count.

#### 4. Machine Learning:

- Build a predictive model to estimate the 'base salary' based on other features in the dataset. Evaluate the model's performance using appropriate metrics.
  - Perform feature selection to identify the most important features influencing the base salary.
- 

### Part 2: Advanced Data Manipulation with SQL

#### Scenario:

You are tasked with querying the company's SQL database to support the Python analysis and provide additional insights. You will write complex SQL queries, create stored procedures, triggers, and views to manage and analyze the data.

#### Tasks:

##### 1. Complex Queries:

- Write a query to find the top 5 companies with the highest average job ratings.
- Use Common Table Expressions (CTEs) to calculate the average 'base salary' and 'max salary' for each location and find locations with an average salary above a certain threshold.

##### 2. Stored Procedures:

- Create a stored procedure to update the 'experience' column by extracting the minimum and maximum years from the 'experience' string and updating 'min exp' and 'max exp' columns.

##### 3. Triggers:

- Write a trigger to automatically update the 'jobListed(days ago)' column based on the current date and the 'postedIn' date when a new job is inserted into the table.

##### 4. Views:

- Create a view to display job listings with complete salary information (base salary and max salary) and filter out jobs with missing salary data.
-

## Part 3: Dynamic Visualizations with Power BI

### Scenario:

The final part of your task is to present your findings through dynamic and interactive visualizations in Power BI. Your visualizations should provide stakeholders with clear insights and enable them to make data-driven decisions.

### Tasks:

#### 1. Data Transformation:

- Import the dataset and perform necessary transformations to handle missing values and standardize data formats.

#### 2. Visualizations:

- Create a dashboard to visualize the distribution of job listings across different locations and companies.
- Develop visualizations to show trends in job ratings and reviews count.

#### 3. Advanced Analysis:

- Use DAX functions to calculate custom metrics, such as the average experience required for jobs by company or location.
- Create interactive visualizations that allow users to filter job listings by rating, experience level, and salary range.

#### 4. Reports and Insights:

- Generate a detailed report summarizing key insights from the dataset, such as the most sought-after locations for data analyst jobs and companies with the highest ratings.