

# Case Study: Construction Progress Analysis

## Objective

This case study will guide you through analyzing a construction project dataset. You will use data analysis techniques to examine project stages, budget utilization, delays, and risk factors in construction projects. By the end, you'll gain insights into effective data handling, visualization, and statistical analysis.

## Scenario

Imagine a company that is managing a large-scale construction project. Stakeholders are interested in understanding the project's timeline, workforce effectiveness, and budget adherence. With data-driven insights, they aim to improve efficiency and proactively address potential issues.

## Tools Used

- **Pandas:** To clean and manipulate the data.
- **Matplotlib & Seaborn:** For visualizations that reveal project progress and resource utilization patterns.
- **Statistics:** To evaluate project milestones and predict possible delays.

## Dataset Overview

- **ProjectID:** Unique identifier for each project
- **ProjectName:** Name of the construction project
- **Stage and Substage:** Current phase of the project
- **StartDate and PlannedEndDate:** Start and planned completion dates
- **ActualEndDate:** The actual date of completion
- **ProgressPercentage:** Percentage of work completed
- **BudgetAllocated and BudgetUtilized:** Budget metrics for project financials
- **TeamAssigned:** The team responsible for the project
- **RiskLevel:** Project's risk classification
- **IssuesReported:** Number of reported issues

## Task 1: Initial Exploration & Cleaning

- Load and explore the dataset using Pandas.
- Check for missing values, data types, and unique values in each column.
- Clean any inconsistencies, such as missing dates or negative values where they shouldn't be.

## Task 2: Calculating Budget Efficiency

- **Objective:** Evaluate budget efficiency across projects.
- **Instructions:**
  - Calculate a **budget efficiency ratio**: [BudgetUtilized/BudgetAllocated]
  - Identify projects that exceeded their budget.
  - Analyze if specific project types tend to have higher budget utilization.

## Task 3: Visualizing Project Progress & Delays

- **Objective:** Visualize project progress by analyzing **ProgressPercentage** and **Delays**.
- **Instructions:**
  - Use a histogram to show the distribution of **ProgressPercentage**.
  - Plot a bar chart comparing **Delays** across different project stages.
  - Create a box plot to understand the spread of delays by **RiskLevel**.

## Task 4: Statistical Summary & Hypothesis Testing

- **Objective:** Perform basic statistical analysis and hypothesis testing.
- **Instructions:**
  - Calculate summary statistics for numerical columns.
  - Compare **BudgetAllocated** and **BudgetUtilized** between **High** and **Low Risk** projects using a t-test.
  - Comment on whether higher-risk projects tend to have a larger variance in budget utilization.

## Task 5: Visualizing Issues & Risk

- **Objective:** Examine relationships between issues reported and risk levels.
- **Instructions:**
  - Create a scatter plot of **IssuesReported** vs. **BudgetUtilized** with color differentiation for **RiskLevel**.
  - Use a heatmap to show correlations among numeric columns, focusing on **ProgressPercentage**, **BudgetUtilized**, **Delays**, and **IssuesReported**.

## Task 6: Project Performance Analysis

- **Objective:** Identify factors impacting project performance.
  - **Instructions:**
    - Calculate the mean **ProgressPercentage** for projects in each **Stage**.
    - Determine the top 3 stages with the highest average **ProgressPercentage**.
    - Create a pie chart to show the proportion of **Status** (e.g., On Track, Delayed) across projects.
- 

## Conclusions

- Summarize insights, such as factors impacting delays, Project Status, budget trends, or underutilized resources.
- Provide actionable recommendations for improved management and cost-efficiency.
- Suggest steps to monitor future progress effectively.

## Save Your Work:

Save all the code syntax and analyses in a **Jupyter Notebook** file for future reference. This file will serve as a comprehensive record of your work and make it easy to review and adjust as needed.

## Make PowerPoint Presentation:

Summarize this analysis by creating a **PowerPoint presentation**.

Highlights:

- The objectives of the analysis
- Key visualizations and findings
- Statistical conclusions and actionable recommendations

This presentation will help you communicate your findings effectively to non-technical stakeholders, illustrating how data-driven insights can impact project decision-making.

---