

QA Analytics: Understanding Defect Trends and Predicting Test Execution Time

Objective

To analyze QA defect and test execution logs to:

1. Explore key trends in defect data (EDA + visualizations).
 2. Identify relationships between defect attributes (severity, module, tester, automation status, etc.).
 3. Predict the **average test execution time** of a test case using regression analysis.
-

Step-by-Step Workflow

1. Data Generation

- Use numpy and pandas to create random data for the above columns.
- Save it as qa_defect_log.csv.

2. Data Exploration

- Use .info(), .describe(), .isnull().sum() to understand missing values.
- Intentionally introduce a few missing values for practice.

3. Handling Missing Values

- Use .fillna() or .dropna() appropriately.
- For numeric: fill with median or mean.
- For categorical: fill with mode.

4. Outlier Detection

- Detect outliers in numeric columns (Execution_Time, Defect_Fix_Time, Lines_of_Code) using:
 - Z-score method
 - IQR method
- Remove or cap the outliers.

5. Feature Engineering

- Convert categorical columns using pd.get_dummies() (or LabelEncoder).
- Normalize numeric columns using StandardScaler, MinMaxScaler, and RobustScaler (to compare effects).

6. EDA & Visualization

- Visualize key relationships:
 - Severity vs Defect_Fix_Time (bar plot)
 - Automation_Covered vs Execution_Time (box plot)
 - Correlation heatmap between numeric columns
 - Distribution plots (histogram + KDE) for Execution_Time and Defect_Fix_Time
 - Scatter plot for Lines_of_Code vs Execution_Time
- Use both matplotlib and seaborn.

7. Model Building (Linear Regression)

- Objective: **Predict Execution_Time** based on defect and test features.
- Train-test split (80/20).
- Fit a `LinearRegression()` model:
- Evaluate using:
 - `r2_score`
 - `mean_squared_error`
 - `root_mean_squared_error`

8. Insights & Interpretation

- Identify which features most strongly affect execution time.
- Example insights:
 - “Execution time increases with lines of code.”
 - “Automated tests show 20% lower average execution time.”
 - “Severity High defects tend to have longer execution and fix times.”