

INTERNSHIP PROJECT PHASE-TESTING

PROJECT 6 – DEVELOPMENT AND TESTING OF YELP DATA ANALYSIS PROJECT ON IBM CLOUD

1. Data Quality Tests

Test Case 1: Null Handling

Test: Check for nulls in all columns.

Result: Nulls found in attributes, handled via default values.

Test Case 2: Category Format

Test: Verify if categories column has consistent format.

Result: Most rows were lists or comma-separated strings. Standardized into list format.

Test Case 3: Duplicate Entries

Test: Check for duplicate business IDs.

Result: No duplicates found.

2. Functional Tests

Test Case 4: Merge Accuracy

Test: Check merged dataset from business_df and checkin_df.

Result: Merge success rate was 93%, i.e., most check-ins had matching business IDs.

Test Case 5: Time Aggregation

Test: Check that all check-ins were parsed into valid datetime formats.

Result: Time parsing successful; used `datetime.strptime`.

3. Visual Output Validation

Test Case 6: Star Rating Distribution Plot

Test: Ensure plot shows logical distribution.

Result: Bell-shaped curve; most businesses between 3 and 4.5 stars.

Test Case 7: Heatmap Readability

Test: Day vs Hour check-in heatmap.

Result: Valid; peak times between 6–9 PM and weekends.

4. Performance Testing (Cloud)

Test Case 8: Notebook Execution Time

Test: Measure runtime on Watson Studio.

Result: Average run time ~5 minutes for full analysis.

Test Case 9: Object Storage Access Speed

Test: Read speed for JSON files.

Result: Upload and read time within acceptable range (<1 min each).

Test Category	Total Tests	Passed	Failed
Data Quality	3	3	0
Functional	2	2	0
Visualization	2	2	0
Performance	2	2	0

Conclusion

The development and testing of the Yelp Data Analysis project on IBM Cloud proved successful. Leveraging Watson Studio and Object Storage enabled streamlined data handling and analysis. The analysis yielded actionable insights on business trends, user behavior, and check-in dynamics. With all tests passed, the platform is validated for further enhancements, including predictive analytics and dashboard integration.