# **NYC 311 Data Quality Assessment Report**

**Author:** Rohit Mande  
 **Date:** June 19, 2025

## **1. Introduction**

This report presents a data quality assessment conducted on the NYC 311 Service Requests dataset. The goal is to evaluate key data quality dimensions to support reliable data governance and stewardship practices.

## **2. Dataset Overview**

* Source: NYC Open Data — 311 Service Requests from 2010 to Present
* Records analyzed: **10,000**
* Key fields: Complaint Type, Agency, Borough, Created Date, Closed Date, etc.

## **3. Data Quality Dimensions Assessed**

The analysis focused on five main data quality dimensions:

* **Completeness:** Percentage of missing or null values
* **Accuracy:** Validity and correctness of data entries
* **Consistency:** Uniformity of data format and values
* **Uniqueness:** Detection of duplicate records
* **Timeliness:** Logical sequence and date/time correctness

## **4. Analysis Summary**

| **Dimension** | **Findings** | **Issues Identified** | **Recommendations** |
| --- | --- | --- | --- |
| Completeness | 10% missing in Closed Date | Missing dates affect timeliness analysis | Implement mandatory field validation |
| Accuracy | Complaint types mostly valid | Some outlier complaint names | Standardize complaint type values |
| Consistency | Date formats mostly consistent | Minor inconsistencies in borough naming | Use lookup tables for standard values |
| Uniqueness | 150 duplicate records found | Duplicate service requests | Add de-duplication step during intake |
| Timeliness | Closed Date precedes Created Date in 5% cases | Logical errors in date sequencing | Validate dates before data entry |

## **5. Conclusion**

The data profiling exercise identified several data quality challenges that could impact analytics and decision-making. Implementing the recommended remediation steps will improve trust and usability of the NYC 311 dataset.

## 