## Title: Ensuring Data Quality and Integrity in the Business Database

## **User Story**

As a database manager,,

I want to to regularly validate and update the business information in our database,

**So that** we maintain high data quality and integrity, ensuring that our business records are always accurate and current.

<u>The Challenge:</u> Over time, business information can change. Companies move, update their contact details, or change their operational status. Our database needs to reflect these changes promptly to remain useful.

## **✓** Acceptance Criteria:

1. **Automated Identification Accuracy:** The system correctly identifies at least 90% of the entries that require updating or verification, as measured against a manually reviewed sample.

## 2. External Data Verification:

- a. The system will use trusted external sources and APIs, like the Secretary of State's database and Google Places, to cross-check the accuracy of our business information.
- b. It will compare our records with these external sources to identify any discrepancies.
- 3. **Alerts and Flags for Review:** When the system finds information that might be outdated or incorrect, it will flag these entries so we can review and verify these entries manually.
- 4. **Continuous Improvement:** The system will learn from each update, becoming more efficient at identifying outdated information over time.

Ensuring Data Quality and Integrity in the Business Database User inputs data Data gets preprocessed and normalized Data gets fed into ML classifier Each record gets a confidence score Output flagged records to file Flagged for manual checking Is record scored low? Exit