Data Engineering Assessment (Firebase-Based Recipe Analytics Pipeline)

Goal:

Assess the candidate’s ability to design and implement a data pipeline using Firebase as the source system. The candidate must use their own recipe as the initial seed data.

Duration:

Designed to be completed within 10 hours.

Requirements:

1. Data Modeling

- Define a complete data model for:

• Recipes

• Users

• User interactions (views, likes, cook attempts)

- Provide a clear schema description or ERD-style diagram.

- Include the candidate’s own recipe as the primary dataset.

2. Firebase Source Data Setup

- Insert the candidate’s recipe into Firebase Firestore.

- Create 15–20 synthetic recipes.

- Create sample user interaction data demonstrating:

• Views

• Likes

• Rating or difficulty usage (optional)

3. ETL / ELT Pipeline

- Export Firestore collections to JSON or CSV.

- Transform the exported data into a set of normalized tables:

• recipe.csv

• ingredients.csv

• steps.csv

• interactions.csv

- Transformation may be implemented in Python or Node.js.

- Output must be clean and free of schema inconsistencies.

4. Data Quality Validation

- Define and apply validation rules such as:

• Required fields present

• Positive numeric fields

• Non-empty arrays

• Valid difficulty values

- Provide a validator script that generates a report indicating:

• Valid records

• Invalid records with reasons

5. Analytics Requirements

Provide at least 10 insights, such as:

- Most common ingredients

- Average preparation time

- Difficulty distribution

- Correlation between prep time and likes

- Most frequently viewed recipes

- Ingredients associated with high engagement

6. Documentation

Provide a README including:

- Explanation of the data model
- Instructions for running the pipeline
- ETL process overview
- Insights summary
- Known constraints or limitations

Deliverables:

- Source files for ETL and validation scripts
- Normalized CSV output
- Analytics summary document
- README file with implementation details
- Optional: Visualization charts (images allowed)

Evaluation Focus:

- Data modeling accuracy
- Pipeline completeness and correctness
- Code clarity and maintainability
- Quality rule implementation
- Depth and relevance of data insights