Project Progress Report: Restaurant Recommendation Engine

Task Overview:
The objective was to build a recommendation engine to predict which restaurants customers are most likely to order from, based on customer location, restaurant information, and order history.

Step 1: Understanding the Problem
- Reviewed the VariableDefinitions.pdf to understand dataset structure and requirements.
- Identified key datasets: train_customers, train_locations, orders, vendors, test_customers, test_locations.

Step 2: Data Analysis & Preprocessing
- Loaded all CSV files using pandas.
- Cleaned data: handled missing values, parsed dates, and ensured data consistency.
- Created a pipeline to store intermediate processed data (using Parquet/CSV).

Step 3: Feature Engineering
- Customer preferences: computed customer-vendor order frequencies.
- Location-specific patterns: calculated vendor popularity per location.
- Global vendor popularity: ranked vendors based on overall order counts.
- Added proximity features using Euclidean distance between customer and vendor masked lat-long.

Step 4: Baseline Recommendation Model
- Created a scoring mechanism combining:
  (a) Customer-vendor frequency (40% weight).
  (b) Location-vendor frequency (25% weight).
  (c) Global vendor popularity (20% weight).
  (d) Inverse distance score (15% weight).
- For each test customer & location, the top vendor was selected based on the combined score.

Step 5: Submission File
- Generated submission CSV file similar to SampleSubmission.csv.
- Stored all results under the submissions/ folder.

Step 6: Code & File Structure
- Organized project into the following structure:
  food-reco/
  ■■■ data/raw/
  ■■■ data/processed/
  ■■■ submissions/
  ■■■ src/ (with config, utils, features, models, pipeline)
- Implemented modular Python scripts for preparing data, training the baseline, and generating predictions.

Step 7: GitHub Repository
- Created GitHub repo: https://github.com/BaluReddy122005/food-reco
- Pushed all code with proper README and requirements.txt.
- Added .gitignore to avoid uploading unnecessary data files.

Step 8: Final Deliverables
- Complete project code on GitHub.
- Submission CSV file attached via email.
- PDF report (this document) summarizing step-by-step progress.

Conclusion:
The task was completed successfully, with a fully working baseline recommendation engine and submission file.