Google POI Confidence - Coding Challenge

At Kuwala, we have two pipelines to get POI information. The first is the OpenStreetMap (OSM) POI pipeline, parsing relevant tags and geo-boundaries from OSM objects. The second pipeline is a scraper for Google Maps. The scraper takes a search string and returns the data for the first corresponding result. Sometimes the result is an exact match, and sometimes it is not. Since it is not feasible to check each result manually, we need to provide a confidence score that indicates how well the Google POI returned from the scraper actually matches the OSM POI the query has been generated on.

The search strings sent to the Google scraper are based on the name of the OSM POI (if available) and its address tags. You find three tables that you need to download from S3.

- osm_poi: https://kuwala-case-studies.s3.eu-central-1.amazonaws.com/backend/google-poi-confidence/osm_poi.csv.gz
- google_poi: https://kuwala-case-studies.s3.eu-central-1.amazonaws.com/backend/google-poi-confidence/google_poi.csv.gz
- google_osm_poi_matching: https://kuwala-case-studies.s3.eu-central-1.amazonaws.com/backend/google-poi-confidence/google_osm_poi_matching.csv.gz

Your task is to calculate the confidence score between 0 and 1, indicating the quality of the result returned from the Google scraper. Add this score as a new column to the google_osm_poi_matching table and save it as a CSV.

Please use a programming language and/or database technology of your choice and upload the source code and/or queries to a GitHub repository.

When you finish the case study, please send us the result table with the confidence score as a CSV and the link to your GitHub repository.

This task should be doable within 1-2 hours.

If you have any questions, please feel free to ask them via email.

We're looking forward to seeing your approach!