Documentation Report: Dakar Restaurant Review Scraper

# 1. Introduction

The purpose of this project was to collect user-generated reviews of restaurants located in Dakar, Senegal, from Google Maps, and prepare a clean dataset suitable for analysis. This involved two main phases: automated web scraping using Selenium and data preprocessing with pandas. The resulting dataset can be used for sentiment analysis, trend detection, customer feedback evaluation, and more.

# 2. Technologies Used

* This project employed the following technologies and libraries:
* • Python 3.10
* • Selenium WebDriver for Chrome automation
* • pandas for data manipulation and cleaning
* • fuzzywuzzy for approximate string matching
* • Google Chrome and Chromedriver
* • CSV for data storage and handling

# 3. Scraping Script: `Scraping\_Maps\_Reviews.py`

This Python script automates a Chrome browser session to visit the Google Maps search results for restaurants in Dakar. The script performs the following major tasks:

* • Launch Chrome browser with a persistent user profile to simulate a real user.
* • Navigate to the Google Maps page filtered for restaurants in Dakar.
* • Scroll the sidebar repeatedly to load all restaurants listed by Google Maps.
* • Collect restaurant names from the sidebar and use fuzzy matching to click each restaurant card.
* • Click the 'Avis' (Reviews) tab to display user reviews.
* • Scroll the review panel multiple times (up to 1000 times) to load more user reviews.
* • Extract the following for each review: restaurant name, restaurant rating, user name, user rating, and review text.
* • Store this data row by row in a file named `dakar\_reviews.csv`.

# 4. Raw Output: `dakar\_reviews.csv`

* This CSV file contains the raw data extracted from Google Maps. It includes:
* • Restaurant — Name of the restaurant
* • Rating — Overall rating of the restaurant
* • User — Name of the reviewer
* • User Rating — Rating given by the user (e.g., '5 étoiles')
* • Review — Text of the review

# 5. Cleaning Script: `Cleaning01.py`

This script cleans and normalizes the raw scraped dataset. Key steps include:

* • Stripping whitespace from column names and values to avoid inconsistencies.
* • Converting user rating text (like '5 étoiles') into numeric format (e.g., 5).
* • Flattening multi-line reviews into single-line entries by removing newline characters.
* • Saving the cleaned data to a new file named `cleaned\_dakar\_reviews01.csv`.

# 6. Final Clean Dataset: `cleaned\_dakar\_reviews01.csv`

The final output is a clean, structured, and standardized CSV file. It retains the original review content but ensures all entries are consistent and ready for downstream analysis tasks such as natural language processing, machine learning model training, and visualization.

# 7. Why We Did This

User reviews offer valuable insights into customer experiences, satisfaction, and opinions. However, data from sources like Google Maps is not readily available for direct analysis due to dynamic content loading. This project enables structured collection of such content, unlocking the potential for local business analysis, competitive benchmarking, and AI-driven recommendation systems.