King Saud University
College of Computer and Information Sciences
Department of Information Technology

IT362: Principles of Data Science

Semester-2, 1447H

Phase 1- loogbook

Topic: Sentiment analysis of reviews of beauty centers

Prepared by

Section No	Name	ID
	Leen Almuqhim	445201272
	Remas Alayiedh	445202140
73059	Lubna Alqifari	445202104
	Aljoury Alreemi	445200074
	Khadijah Alshehri	445202138

Supervised by: Lama Alsudias

Objective of This Phase

The objective of this phase was to collect unstructured customer reviews for beauty salons and spas using the Google Places API in order to build a dataset suitable for sentiment analysis.

Data Collection Method

Data was collected programmatically using the Google Places API.

The script targeted:
• 10 UK cities (London, Manchester, Birmingham, etc.)
• Business types: beauty_salon and spa
• Radius: 4000 meters
• Target places: 700

The process was divided into two phases:

Phase 1 – Collect Place IDs
• Iterated through selected UK city coordinates
• Requested nearby places for each business type
• Handled pagination (up to 3 pages)
• Stored unique place IDs

Phase 2 – Fetch Reviews
• Requested only the "reviews" field to save API quota
• Extracted review text and metadata:
• Rating
• Timestamp
• Relative time description
• Language
• Saved reviews in JSON Lines format

Decisions Made and Rationale

Decision 1: Use Google Places API

Rationale:
Google provides reliable, structured access to real customer reviews with metadata.

Decision 2: Limit pagination to 3 pages

Rationale:
To control API quota usage and avoid excessive requests.

Challenges Faced

1. API Rate Limiting

Issue: Google returned OVER_QUERY_LIMIT errors.
Solution: Implemented exponential backoff and retry logic.

2. Pagination Delay Requirement

Issue: Google requires delay before next_page_token becomes active.
Solution: Added time.sleep() before requesting next page.