Yelp Data-Lake on GCP – Capstone Business Requirements Document (BRD)

# 1. Cover Page

Version / Date: v0.1 (2025-06-05)

Author: Varun Alluri | Advisor: Project Mentor

# 2. Executive Summary

This project aims to deliver actionable insights using dashboards built on Yelp’s public dataset. The insights will support business owners and analysts in understanding customer behavior, identifying top-performing locations, and tracking rating trends across geographies. All tech implementation details are captured in the Design document.

# 3. Problem Statement

Business stakeholders lack easy visibility into trends around reviews, customer sentiment, and geographic performance. This limits their ability to respond quickly to customer feedback or capitalize on high-performing regions.

# 4. Objectives & Key Success Criteria

|  |  |  |
| --- | --- | --- |
| **ID** | **Objective** | **KPI / Acceptance Test** |
| OBJ-01 | Deliver Business Performance Dashboard | Star rating & review count KPIs render correctly |
| OBJ-02 | Deliver Customer Engagement Dashboard | Sentiment KPIs load under 5s |
| OBJ-03 | Build Location-Based & Loyalty Dashboards | Stakeholder feedback ≥ 4/5 |
| OBJ-04 | Keep GCP credits usage ≤ US $30 | Billing report shows spend ≤ US $30 for project month |

# 5. Project Scope

## 5.1 In-Scope

* Dashboards: Business Performance, Customer Engagement, Loyalty, Temporal Trends, Location Insights
* Data: business, review, user, checkin, tip

## 5.2 Out-of-Scope

* Real-time data ingestion
* Photos or video
* Advanced ML or external APIs

# 6. Stakeholders & Roles

|  |  |  |
| --- | --- | --- |
| Role | Name | Responsibility |
| Sponsor | Project Manager (Pratik Kale) | BRD Approval, final signoff |
| Analytics Eng. | Varun Alluri | Dashboard logic & KPIs |
| Data Engineer | Prabhu Varma Cherukuri | Data ingestion, ETL |
| Data Engineer | Vishwajith Varma | Data ingestion, ETL |

# 7. Dashboard Requirements

## 7.1 Dashboard A – Business Performance & Customer Engagement

|  |  |  |  |
| --- | --- | --- | --- |
| **Visual #** | **Chart Type** | **Business Question** | **Granularity / Filters** |
| A-1 | Bar Chart | Avg star rating by category | Category, City, Date Range |
| A-2 | Line Chart | Check-in trends over time | Week, City |
| A-3 | Table | Top 10 by review count | Review Count, Category |
| A-4 | Stacked Bar | Votes (funny/useful/cool) by business | Vote Type, Business ID, Date |
| A-5 | Heatmap | Tip volume by business/day of week | Business, Date |

Default Filters

* Date Range: Last 13 weeks (editable)
* City Selector: Multi-select (top cities)
* Category: Drop-down
* Vote Type: Checkbox or multi-select

## 7.2 Dashboard B – Loyalty, Temporal, and Location Insights

|  |  |  |  |
| --- | --- | --- | --- |
| **Visual #** | **Chart Type** | **Business Question** | **Granularity / Filters** |
| B-1 | Donut Chart | Elite vs. non-elite review split | Elite Status, Category |
| B-2 | Box Plot | Rating spread by user type | Elite vs. Non-Elite, Category |
| B-3 | Line Chart | Review volume over time | Monthly, Quarterly |
| B-4 | Choropleth Map | Star ratings by state | State, Category |
| B-5 | Bubble Map | Check-in volume by city | City |

Default Filters

* Date Range: Last 13 weeks
* City: Single-select
* Category: Multi-select (top 15)
* Elite Status: Toggle

# 8. Data Requirements

|  |  |  |
| --- | --- | --- |
| **Table Needed** | **Required Fields** | **Source File** |
| business | business\_id, name, city, state, categories | business.json |
| review | review\_id, business\_id, stars, date | review.json |
| user | user\_id, yelping\_since, elite | user.json |
| checkin | business\_id, date | checkin.json |
| tip | user\_id, business\_id, compliment\_count, date | tip.json |

Derived Marts

- mart.review\_weekly\_city – weekly counts & avg rating by city

- mart.review\_weekly\_category – weekly counts & avg rating by main\_category

Data quality rules: stars∈1–5, non-null business\_id, ISO date format.

# 9. Non-Functional Requirements

|  |  |
| --- | --- |
| Category | Requirement |
| Performance | Dashboards initial load ≤ 5s (95-percentile) |
| Availability | Scheduled refresh success rate ≥ 99% |
| Usability | Stakeholders complete filter tasks in ≤ 3 clicks |
| Security | RBAC with least privilege; no public write access |
| Cost | Total monthly spend ≤ $30 student credits |

# 10. High-Level Milestones (6 weeks)

# 6-Week Project Milestones

| **Week** | **Deliverable** |
| --- | --- |
| 1 | BRD sign-off, data dictionaries drafted |
| 2 | Draft SQL for marts, curated tables populated on subset |
| 3 | Full dataset loaded, mart.review\_weekly\_city ready |
| 4 | Dashboard A - MVP, mart.review\_weekly\_category ready |
| 5 | Dashboard B - MVP, performance tuning, UAT |
| 6 | Stakeholder feedback fixes, final demo & submission |

# 11. Risks & Mitigations (selected)

|  |  |  |
| --- | --- | --- |
| Risk | Impact | Mitigation |
| Large JSON size | Slows query performance | Pre-aggregate marts; optimize joins |
| Schema drift | Pipeline breaks | Pin dataset version; add validation step |
| Time constraints | Missed deadlines | Lock visuals by Week 5; make extras optional |

# 12. Acceptance

Capstone is accepted when:

* All dashboards render successfully with live data and meet KPI load-time targets.
* Stakeholder UAT checklist (filter tests, metric verification) passes ≥ 90%.
* Billing report confirms spend within credit limit.  
    
  (Pipeline design, service selection, and cost allocations are covered in the Design document.)