Documentation: Comprehensive Data Analysis and Visualization of IAB-Registered Companies

Project Overview

This project aims to transform the unstructured data of architectural firms registered with the Institute of Architects Bangladesh (IAB) into meaningful, data-driven insights. The final outcomes include cleaned and structured data, geographic and temporal analysis, professionalism scoring, predictive fee analysis, and a dynamic dashboard for interactive exploration.

Stage-1: Project Roadmap & Core Objectives

Stage	Core Objectives
1	Define objectives and plan the workflow
2	Extract unstructured data from PDF into Excel
3	Clean and enrich the dataset
4	Analyze using Excel Pivot Tables and Python
5	Build an interactive dashboard in Looker Studio
6	Prepare the final report and documentation
7	Project Deliverables
8	Key Learnings of the Project

Tools by Phase

Phase	Tools Used
Data Extraction	Adobe Acrobat, Manual Transcription
Data Cleaning/ Tabulation	Microsoft Excel, Google Sheets
Data Analysis	Excel Pivot Tables, Python (Google Colab)
Visualization	Google Looker Studio
Reporting & Documentation	Google Docs (converted to PDFs)

Stage-2 & 3: Data Pipeline

Step 1: Data Aggregation & Extraction

- *Source:* IAB Raw Data (unstructured, text-based, PDF format)
- *Method:* Data was manually transcribed and organized into structured columns: Firm Name, Year, Address, Division, Thana, Website, Email, Google Rating, Staff Count, Geo-coordinates, etc.

Step 2: Data Cleaning & Processing

Issue	Action	Formula/Tool
Whitespace Errors/ Extra Spaces	Trimmed extra spaces	=TRIM()
Inconsistent Year Format	Converted to YYYY	=YEAR() or =TEXT()
Missing Websites/ Emails	Tagged as “Unregistered”	=IF(ISBLANK(), "Unregistered", X)
Fragmented addresses	Split into Thana and Division	Text-to-Columns, =MID(), =LEFT()

Duplicate entries	Removed	=UNIQUE() + manual review
Missing ratings / coordinates	Added manually	Google Maps and Google Business Profiles
Missing geo-coordinates	Searched manually via Google Maps	Google Maps
Inconsistent firm names	Standardized naming variations	=SUBSTITUTE()
Missing contact/web info	Filled via targeted online search	Manual Google Search
Unstructured raw data	Transcribed and organized into Excel tables	Manual + Excel formatting

Step 3: Data Enrichment

- Manually verified missing websites and emails
- Collected geo-coordinates from Google Maps
- Researched Google ratings from business listings

Stage-4: Analysis Framework

Key Metrics and Predictions Logic

Metric	Formula / Logic	Insight / Purpose
Total Firms	=COUNTA()	Counted 142 registered firms
Average Rating	=AVERAGE()	Overall average rating was 4.3
Website Presence	=COUNTIF (Status="Registered")	63% of firms had active websites

Experience Buckets	Year-based grouping	Categorized into 5-year intervals
Project Estimate	Years × 3 projects per year	Used for service volume projection
Professionalism Score	Rating ≥ 4.7, website present, ≥ 5 architects	Logic-based scoring
Fee Prediction (High)	Rating ≥ 4.8, website present, ≥ 8 members	Indicator for higher service fees
Fee Prediction (Low)	No website, rating < 4.0, ≤ 3 members	Indicator for lower service fees

Analysis Conducted

- Establishment trends over time
- Website and email adoption rates
- Google rating distribution
- Regional breakdown by Division and Thana
- Top firms per region based on experience and rating
- Predictive scoring for professionalism and service tiers

Stage-5: Dashboard Development

Platform

- Google Looker Studio

Data Source

- Connected directly to the cleaned dataset in Google Sheets

Dashboard Features

- **Filters:** IAB ID, Company Name, Lab ID, Division, Thana, Postcode, Contact, Year Group, Website Status

- **Visuals:**
  - Bar charts (Firms by year, Ratings, Top Thanas)
  - Pie/Donut charts (Division share, Rating availability)
  - Geo map (Firm locations in Dhaka & Chittagong)
  - KPI cards (Company count, Architect stats)
  - Funnel chart (Member designations)
  - Paginated email ID table
- **Interactivity:**
  - Filters update charts and map in real time
  - Hover and pagination for deeper exploration
- **Live Dashboard Link:**
  Access the Dashboard

Stage-6: Report Writing

Tools Used

- Google Docs / Microsoft Word for documentation
- Visuals and charts exported from Google Colab and Looker Studio
- Embedded links to dashboard and code resources

Report Structure

1. Introduction and Objectives
2. Data and Methodology
3. Key Findings
4. Visuals: Charts, Graphs, Dashboard
5. Logical Scoring and Assumptions
6. Challenges & Solutions
7. Learning Outcomes & Future Work
8. Conclusion
9. Appendices & Deliverables (Dashboard link, QR code, code notebook, etc.)
10. References

Stage-7: Project Deliverables

- Static Descriptive Report & Documentation
- Cleaned Dataset
- Interactive Dashboard
- Google Colab & Codebase
- Click Here to Access the Full Project on GitHub

Stage-8: Key Learnings of the Project

- **Mastery of Data Cleaning:** Gained hands-on experience in cleaning, structuring, and enriching real-world, unstructured data, accounting for nearly 80% of the total project effort.
- **Effective Use of Logical Scoring:** Developed predictive insights, such as professionalism and fee estimates, using logic-based criteria, without relying on complex machine learning models.
- **Dashboard Design Expertise:** Strengthened skills in creating dynamic, user-friendly dashboards using Google Looker Studio, with a focus on interactivity, filters, and tooltips for better data accessibility.
- **Professional Reporting & Storytelling:** Enhanced ability to communicate insights clearly through well-structured documentation, visual storytelling, and business-centric reporting.