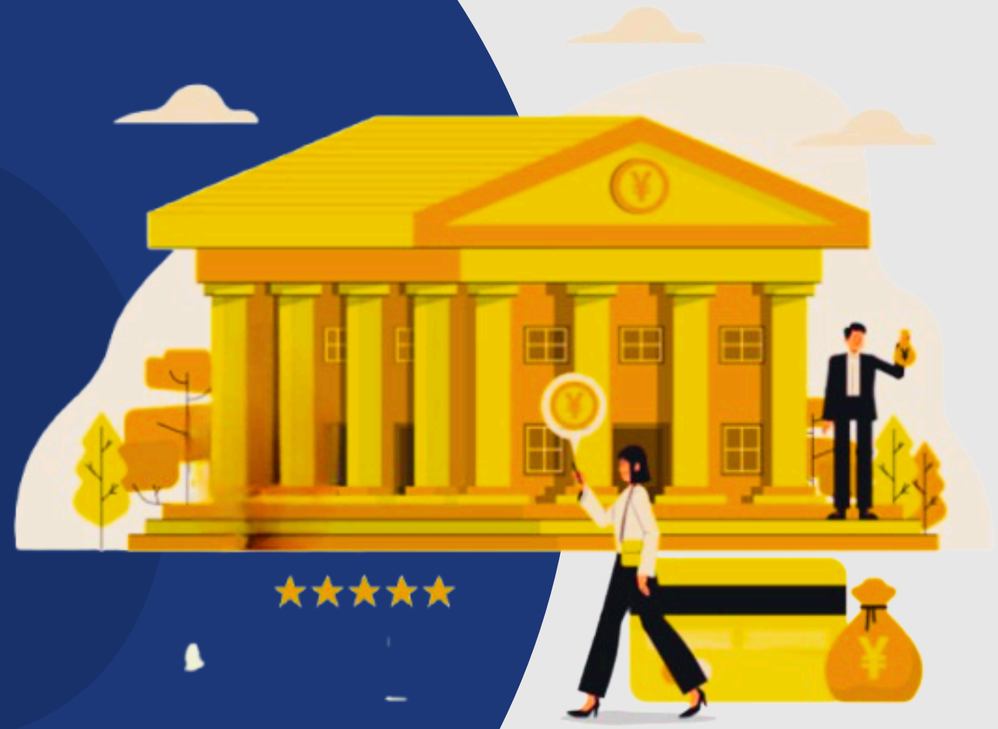
Supervised By
Karim EL Bakli

INDIAN BANKS CUSTOMERS REVIEWS ANALYSIS



Group

1- Ahmed Moawad
2- Amal Adel
3- Mohamed Gad
4- Reem Ashraf
5- Sayed Alaa
6-Youssef Saber

Outline

1. Project Overview
2. Data Preparation and Transformation
    - Cleaning Steps
    - Data Enrichment and Transformation
3. Data Modeling
4. Data Analysis & Key Performance Measures
5. Dashboards
6. Executive Summary
    - Key Insights

Project Overview

This project analyzes customer reviews for major Indian banks to understand satisfaction levels, identify key service issues, and track trends over time. Using a structured data model and AI-based text analysis, the insights are presented in an interactive Power BI dashboard designed to support executive decision-making and improve customer experience.

Objective

To transform raw bank review data into actionable business intelligence using a comprehensive data model and visualization.

Dataset Overview

(Scope and Volume)

The analysis spans a large dataset of customer feedback, providing wide coverage across the financial sector

- Columns Names:

Author, Date, Address, Bank, Rating, Review title by user, Review, Bank image, Rating title by user, Useful count

- Number of Columns: 10 Columns (7 text & 3 numerical)
- Total Records: 3K customer reviews.
- Entities Analyzed: 12 distinct banks.
- Geographic Coverage: Feedback is collected from 325 different cities.

Data Preparation and Transformation

1. Cleaning Steps

a Missing Value Handling

Missing entries for Author and Bank name were replaced with placeholders "Unknown user," & "Unknown bank".

Nulls in Useful Count were set to 0, and blank review titles were set to "Untitled Review."

b Data Type Validation & Standardization

Extra white spaces were trimmed, and casing was standardized (capitalizing each word) for Author and Address for uniformity.

Column headers were standardized by capitalizing and replacing underscores with space. Key columns were relabeled for clarity, renaming Author to Reviewer and Date to Review Date.

The raw text-based date format was converted to a consistent numeric date, which is essential for accurate time-series analysis and efficient filtering operations

Data Preparation and Transformation

2. Data Enrichment and Transformation

a Rating Categorization

Two features were created to categorize the numerical rating:

Service Quality: The raw ratings were grouped into three levels of service performance:

Ratings (0 – 2.5): Bad Service
Ratings (2.5 – 3.5): Good Service
Ratings (3.5 – 5): Excellent Service

Rating Category: ratings were also grouped into predefined categories:

Ratings (4–5): Positive
Rating (3): Neutral
Ratings (1–2): Negative

Data Preparation and Transformation

2. Data Enrichment and Transformation

b Location Standardization

A new City column was derived from the Address field to ensure all locations map to valid cities in India, improving the accuracy of location-based insights. An AI-assisted tool applied the following mapping rules:

Category	Recommended Action (Unify as Indian Cities)
Valid Cities	Keep all as valid city entries.
Districts / Localities / Suburbs (not cities, but administrative regions)	Replace with nearest major city
States (not cities, but Union Territories)	Replace with the capital
Variants (Cities Changed its name or duplicates)	Keep only the official/modern city name for consistency

Data Preparation and Transformation

2. Data Enrichment and Transformation

c Review Metrics and Scoring

- **Review Word Count**

Measures the total number of words in each review.
Used to understand how detailed or concise customer feedback is.

- **Influential Review (Binary Indicator)**

Identifies reviews with exceptionally high engagement.
  - Yes: Useful Count is above the 75th percentile.
  - No: Useful Count is at or below the threshold.

Helps highlight reviews that strongly resonate with customers

- **Bank Composite Score**

An overall performance index created using a weighted average of key metrics:
  - Rating – 60%
  - Engagement – 20%
  - Share of Voice – 20%

Provides a holistic score to compare banks across multiple dimensions.

Data Preparation and Transformation

2. Data Enrichment and Transformation

d Text Analysis

An AI-assisted tool generated a Python script to analyze all text-based fields. The script produced an Excel file containing three sheets, each designed for a specific analytical purpose.

Sheet Name	Source Column	Coulmns	Data Model Connectivity
Top Words per Category	**Rating Title by User** • The titles are concise and focused. • They provide clear, indicative keywords. • Ideal for isolating terms that strongly correlate with each rating level.	• Word • Count • Rating Category(Positive, Negative & Neutral)	• Connected to DimRatingCategory. • Joins indirectly to FactReviews via a minor snowflake structure. • Supports filtering and dimensional analysis.
Words Summary by Rating		• Rating Category(Positive, Negative & Neutral) • Most/Least Frequent Word • Most/Least Frequent Word Count	• Key fields merged directly into DimRatingCategory. • Provides immediate access to top and bottom words for each category.
Word Cloud Data	**Review** • Full review text offers richer vocabulary. • Ensures the word cloud reflects broad customer feedback trends. • Ideal for visuals that require large text volume.	• Word/Phrase • Occurrence Count	• Contains the top 50 most frequent words/phrases across the entire Reviews and their Occurrence Count, • Used exclusively for visual word cloud generation.

** To ensure meaningful results, the script applied multiple exclusion layers beyond standard stop words as Banks names & Generic Noise Terms (e.g., customer, service, account, review)

Data Modeling

The data model follows dimensional modeling best practices using a Hybrid Starflake Schema. This structure maintains the speed of a Star Schema while incorporating targeted normalization to support advanced text-analysis outputs.

Schema Identification

[Hybrid Schema (Starflake) Structure]

- **Star Schema Core**

The central FactReviews table connects directly to major dimensions (e.g., DimBank, DimDate), ensuring fast query performance and simpler reporting.

- **Snowflake Extensions**

Light normalization is added where analytically necessary.

  - DimRatingCategories → Top Words per Category

This supports enriched text-analysis while keeping the model efficient.

Data Modeling

1. Fact & Dimension Tables

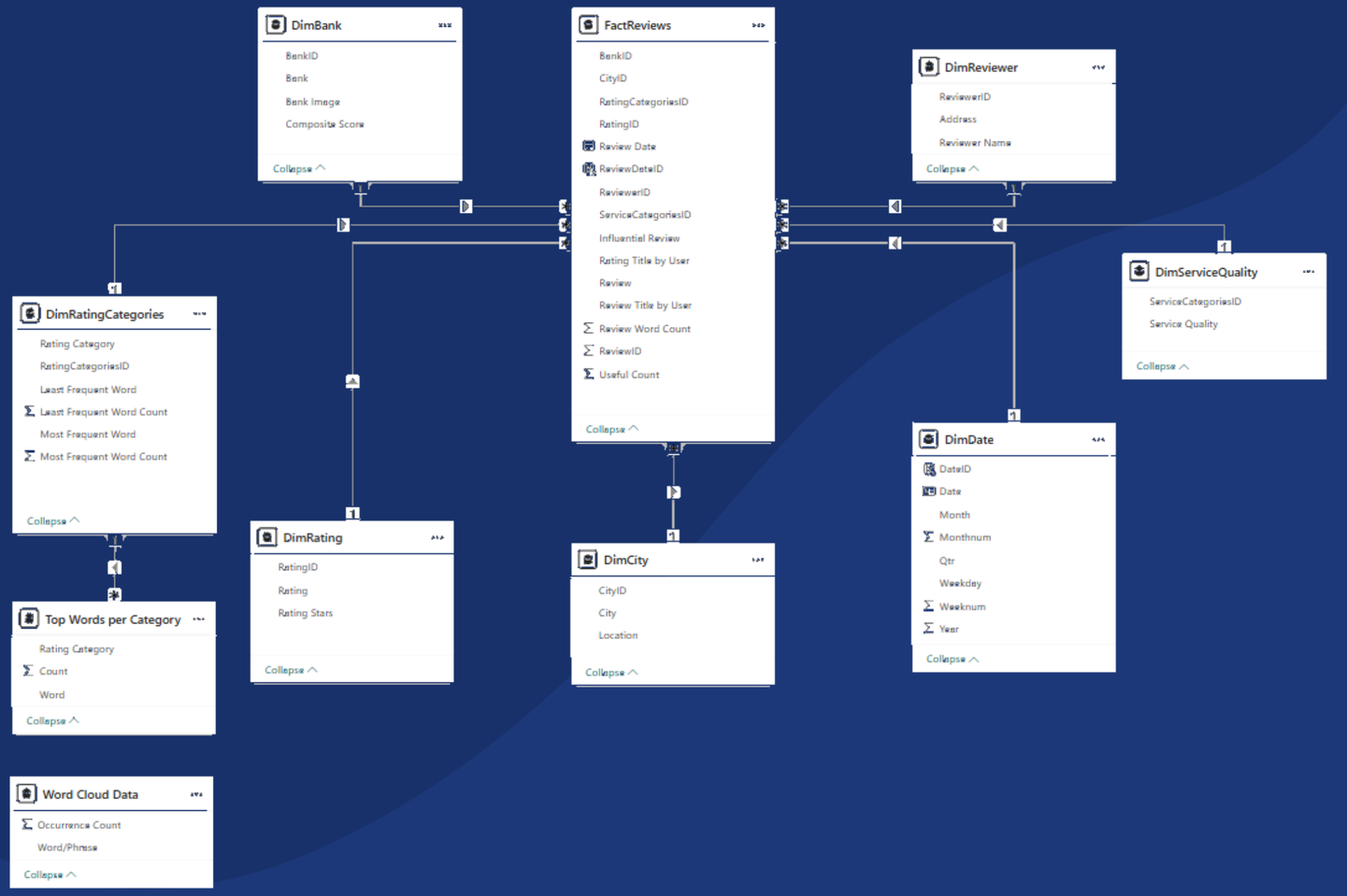
Table Name	Type	Description
FactReviews	Fact	One row per review. Contains foreign keys to all dimensions plus measures like *Review Word Count* and *Useful Count*.
DimDate	Dimension	Time attributes (Year, Month, Qtr, Weekday) enabling time-series analysis.
DimBank	Dimension	Bank attributes including *BankID* and the computed *Composite Score*.
DimReviewer	Dimension	Details about the reviewer (e.g., Name, Address).
DimCity	Dimension	Standardized city values derived from the Address field.
DimRating	Dimension	Rating stars from 1 to 5.
DimServiceQuality	Dimension	Categorizes raw ratings into *Bad, Good, Excellent*.
DimRatingCategories	Dimension	Sentiment-based categories (*Positive, Neutral, Negative*) enriched with merged data from *Words Summary by Rating*.

Data Modeling

2. Analysis / Lookup Tables

Table Name	Type	Description
Top Words per Category	Snowflake Lookup (linked via DimRatingCategories)	Contains top words and their counts per sentiment category, enabling detailed keyword filtering.
Word Cloud Data	Analysis/Lookup	Top 50 most frequent words/phrases from all Reviews, used for generating Word Cloud visuals only.

Data Modeling



Data Analysis & Key Performance Measures

- **Descriptive Metrics**

Track overall review volume, number of banks/cities, average rating, and review word length.

- **Customer Engagement Metrics**

Capture usefulness (total/average useful votes) and identify high-impact reviews.

- **Bank Performance Metrics**

Include normalized Composite Score, Share of Voice %, and sentiment/engagement distribution.

- **Text Analysis Metrics**

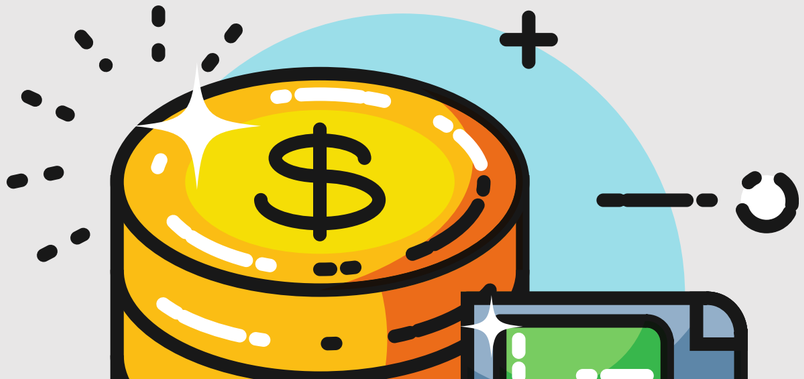
Word cloud vocabulary coverage, sentiment-based keyword trends, and top indicative terms across rating categories.

- **Time Intelligence Metrics**

Month-over-Month (MoM) and Year-over-Year (YoY) trends for short- and long-term performance insights.

Dashboards

Dashboards enable multi-dimensional analysis of customer feedback, providing actionable insights across four main areas.



- Overview & Descriptive Analysis
- Bank Performance & Comparative Analysis
- Review Insights & Thematic Analysis
- Time-Based Trend Analysis

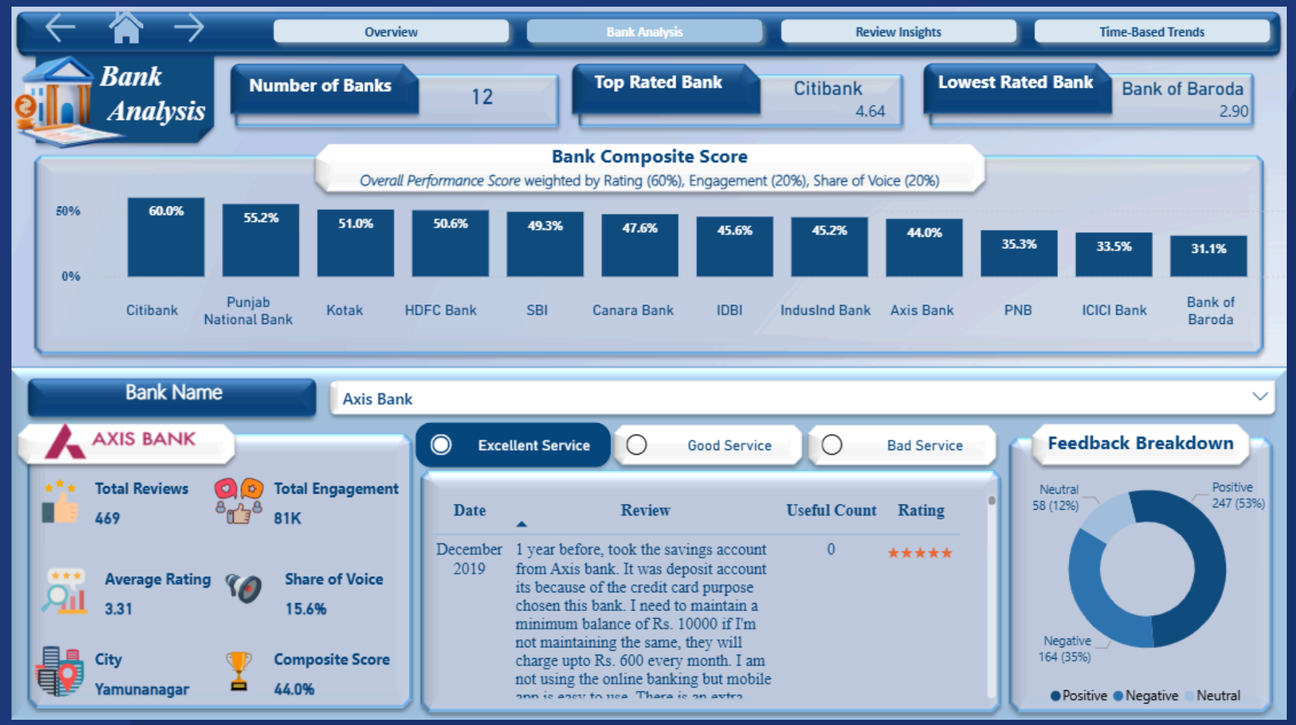
1

Overview & Descriptive Analysis

This initial page provides a high-level, single-view snapshot of the entire customer feedback landscape. It focuses on foundational metrics (KPIs) like Total Reviews, Average Rating, and geographic distribution, and uses the Ratings vs. Reviews Scatter Plot and the Service Quality Split to segment banks based on volume and quality.

2 Bank Performance & Comparative Analysis

This dedicated page shifts the focus to deep, comparative evaluation by ranking all banks based on the proprietary Bank Composite Score (a weighted average of Rating, Engagement, and Share of Voice). It offers drill-down capabilities, allowing users to select a specific bank to view its key metrics summary and filter individual review records by Service Quality.



3 Bank Performance & Comparative Analysis

This analytical page integrates AI-generated text mining outputs, focusing on qualitative insights. It dynamically displays the Most Frequent Words and Rating categories specific KPIs, allowing analysts to filter by Positive, Neutral, or Negative categories to uncover the root cause vocabulary driving customer satisfaction or dissatisfaction.

4 Bank Performance & Comparative Analysis

This final page is dedicated to Time Intelligence, featuring trend visuals that track Review Volume and Engagement across different sentiment groups over time. A crucial Dynamic MoM/YoY Tooltip provides instant change analysis (Month-over-Month or Year-over-Year), helping to monitor performance stability and detect critical trends or volatility in real-time.

Executive Summary

This analysis applied a Hybrid Schema data model and AI-driven text mining to transform 3,000 reviews from 325 cities into clear, actionable intelligence. The following insights highlight the strongest performers, critical service gaps, dominant customer themes, and key temporal dynamics shaping customer experience across the sector.

Key Insights

- Citibank leads the sector with the highest Rating (4.64) and Composite Score (60%), with all categorized reviews classified as Excellent Service.
- Bank of Baroda and PNB show significant service gaps, driven by low ratings and a high share of Bad Service reviews.
- Positive sentiment dominates (54.7%), with “Blown Away” as the top positive keyword and longer, more descriptive reviews from satisfied customers.
- Negative sentiment centers on service delays and staff issues, with “Disappointed” as the most frequent negative term.
- Neutral reviews highlight improvement opportunities, with recurring “Okay”/“Decent Experience” themes indicating easy wins for customer conversion.
- Seasonal review peaks in January and December signal key periods for resource planning and service readiness.
- Negative reviews produce outsized engagement, reaching up to 44K interactions in 2023—requiring real-time monitoring and fast response frameworks.

Thank You