

Swiggy Data ETL - Summary

Objective

The project performs **ETL (Extract, Transform, Load)** on a Swiggy restaurant dataset from **Bangalore** to analyze **restaurant ratings, pricing, cuisines, and locations**.

Data Extraction

- **Total Records:** ~ **15,000** restaurants
- **Key Columns:** Restaurant name, cuisine type, location, ratings, votes, cost for two

Data Transformation (Cleaning & Preprocessing)

- **Missing Values:** ~ **5-10%** of the data had missing ratings or prices, handled via imputation
- **Duplicate Removal:** Identified and dropped ~**500 duplicate entries**
- **Cost Bucket Categories:**
 - **Low-cost (< ₹300)** → **35%** of restaurants
 - **Mid-range (₹300 - ₹700)** → **50%**
 - **High-end (> ₹700)** → **15%**

Data Loading

- The cleaned data is stored in structured formats for visualization and further analysis

Analysis & Key Insights

- **Top 3 Most Common Cuisines:**
 - **North Indian (30%)**
 - **South Indian (25%)**
 - **Chinese (20%)**
- **Average Cost for Two: ₹500**
- **Highest Restaurant Density Areas:**
 - **Indiranagar (~1,500 restaurants)**
 - **Koramangala (~1,200 restaurants)**
 - **Whitefield (~900 restaurants)**
- **Rating Distribution:**
 - **4.0+ Rating** → **50%** of restaurants
 - **Below 3.5 Rating** → **20%**
- **Pricing vs. Rating Correlation:**
 - High-cost restaurants (**₹700+**) have an **average rating of 4.2**

- Low-cost restaurants (**₹300 or less**) have an **average rating of 3.5**

Conclusion

- Mid-range restaurants (₹300-₹700) dominate the market (50%)
- High-rated restaurants are often in premium price ranges
- North Indian & South Indian cuisines are the most popular choices
- Indiranagar & Koramangala are the food hubs of Bangalore