Tasks 4 & 5: Popular Dishes and Restaurant Recommendations

Dheer Gupta

May 2025

1 Overview and Methodology

In Tasks 4 and 5, I built a recommendation system for Indian cuisine using the dish list from Task 3. My goals were:

1. **Task 4**: Rank popular Indian dishes based on review mentions, sentiment, and ratings.
2. **Task 5**: Recommend restaurants for specific dishes using dish-specific scoring.

I used the complete dish list from Task 3, combining both the manually cleaned dishes (27 items) and newly discovered dishes (12 items) for a total of 37 Indian dishes analyzed.

1.1 Data Collection

1. Loaded Indian restaurant business data from the Yelp dataset
2. Collected 8,000 reviews from Indian restaurants
3. Extracted dish mentions using simple string matching
4. Computed sentiment scores using TextBlob for each review

2 Task 4: Mining Popular Dishes

2.1 Ranking Algorithm

I developed a composite scoring function that balances multiple factors:

$$\begin{aligned}
\text{Dish Score} = {} & 0.4 \times \text{Total Mentions} && (1)\\
& + 0.3 \times (\text{Avg Sentiment} + 1) \times 10 && (2)\\
& + 0.2 \times \text{Avg Review Stars} && (3)\\
& + 0.1 \times \text{Number of Restaurants} && (4)
\end{aligned}$$

This weighting prioritizes frequency (40%) while incorporating sentiment analysis (30%), review quality (20%), and restaurant diversity (10%).

2.2 Results

Figure 1 shows the top 15 popular Indian dishes. The results revealed:

Dish	Score	Mentions
naan	1014.7	2485
masala	643.1	1565
curry	565.6	1362
tikka masala	377.2	906
chicken tikka	369.8	886

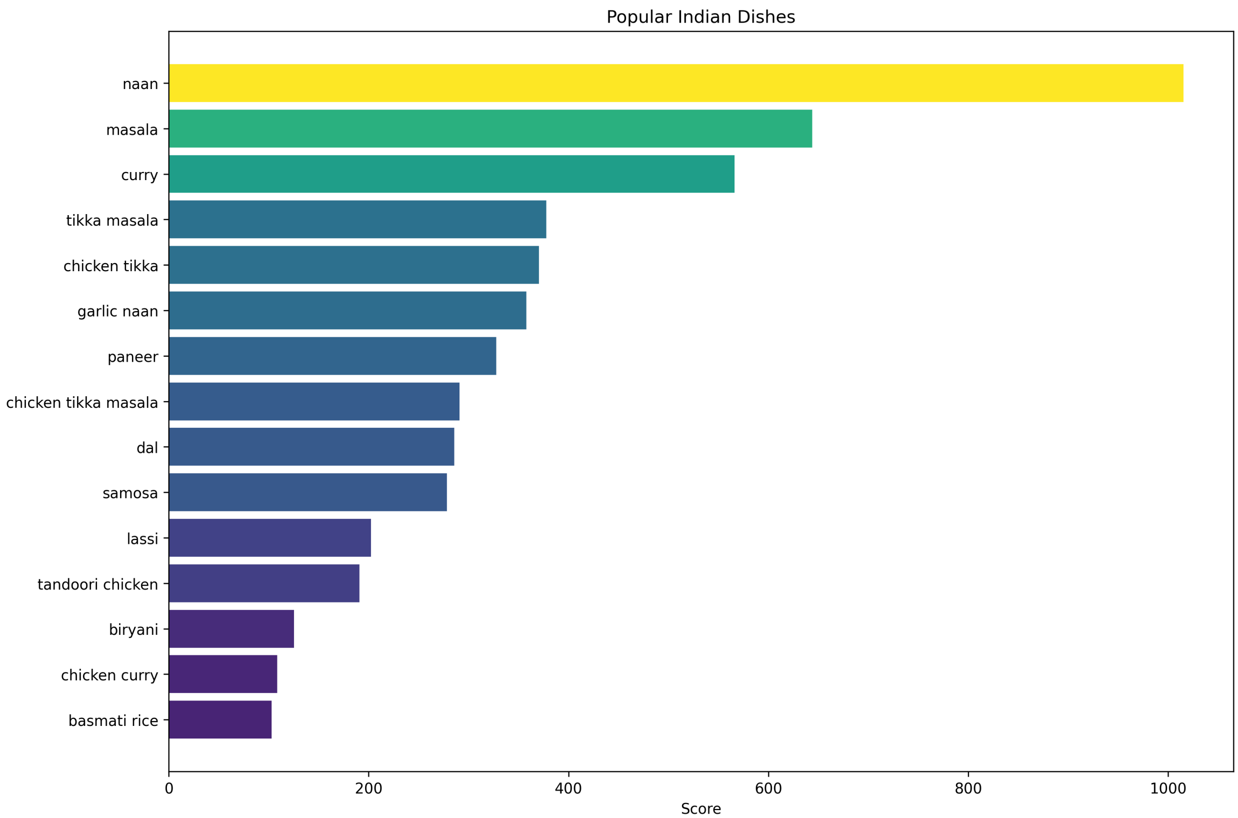

Figure 1: Popular Indian Dishes Ranked by Composite Score

Key Findings: Naan dominated with over 2,400 mentions, confirming its status as a staple. Masala and curry also ranked highly, representing broad categories rather than specific dishes.

3 Task 5: Restaurant Recommendations

3.1 Restaurant Scoring Algorithm

For each dish, I ranked restaurants using:

$$
\begin{aligned}
\text{Restaurant Score} = {} & 0.3 \times \text{Dish Mention Count} && (5)\\
& + 0.4 \times (\text{Avg Sentiment} + 1) \times 10 && (6)\\
& + 0.3 \times \text{Avg Review Stars} && (7)
\end{aligned}
$$

This approach emphasizes sentiment (40%) and review quality (30%) while considering mention frequency (30%).

3.2 Results

I analyzed the top 3 dishes from Task 4. Figures 2, 3, and 4 show restaurant recommendations for each dish.

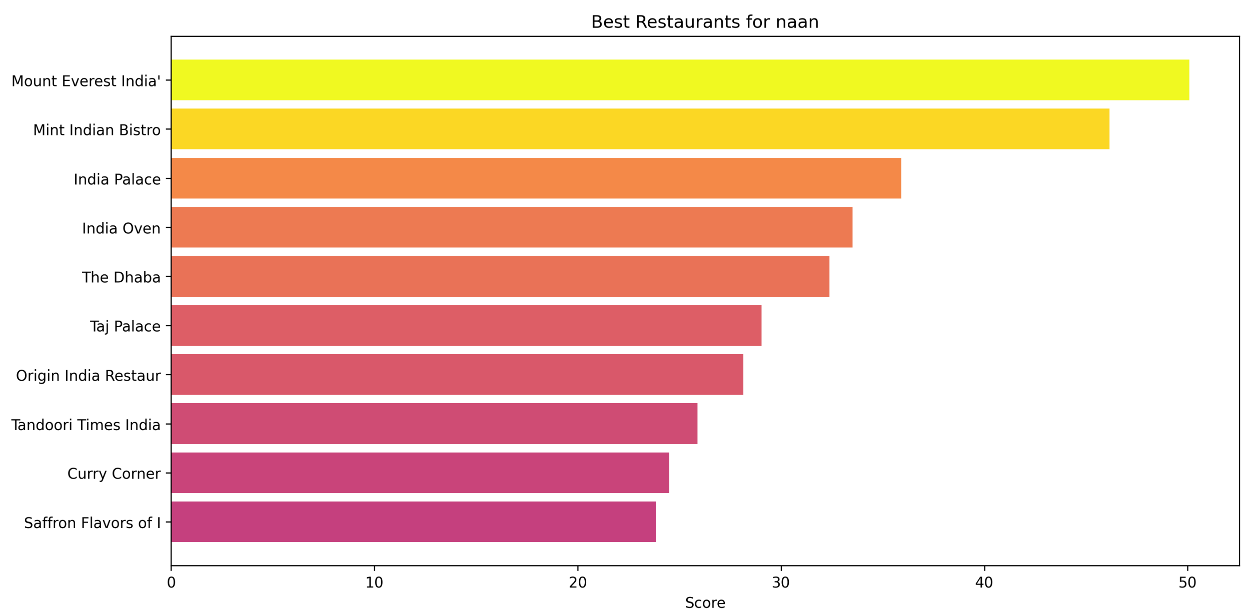

Figure 2: Best Restaurants for Naan

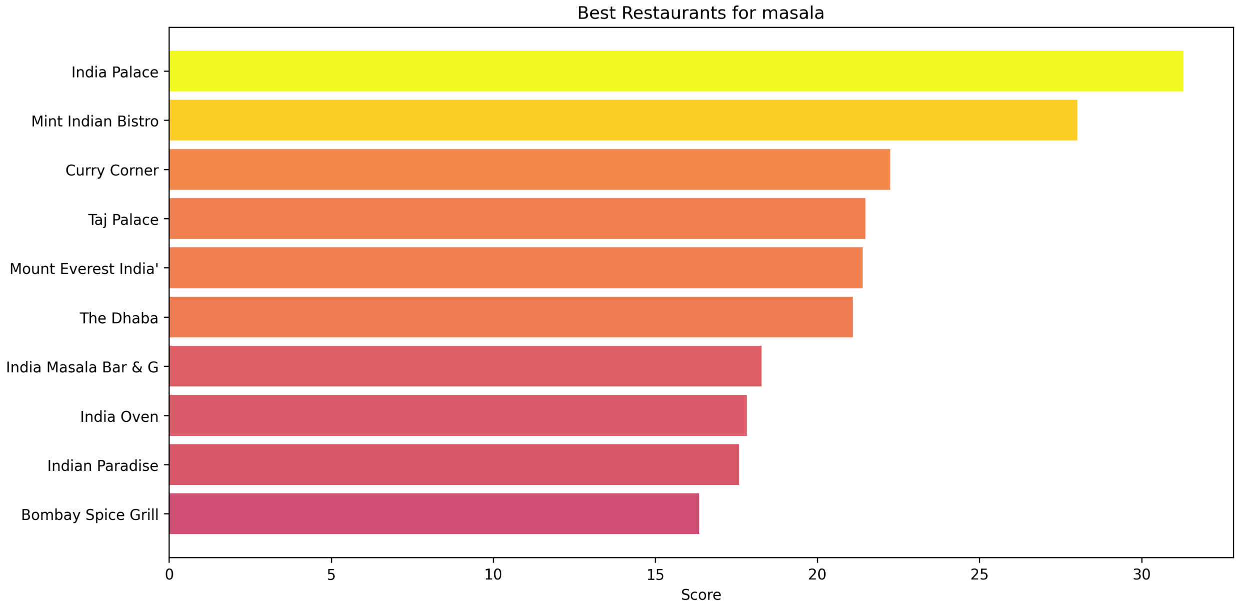

Figure 3: Best Restaurants for Masala

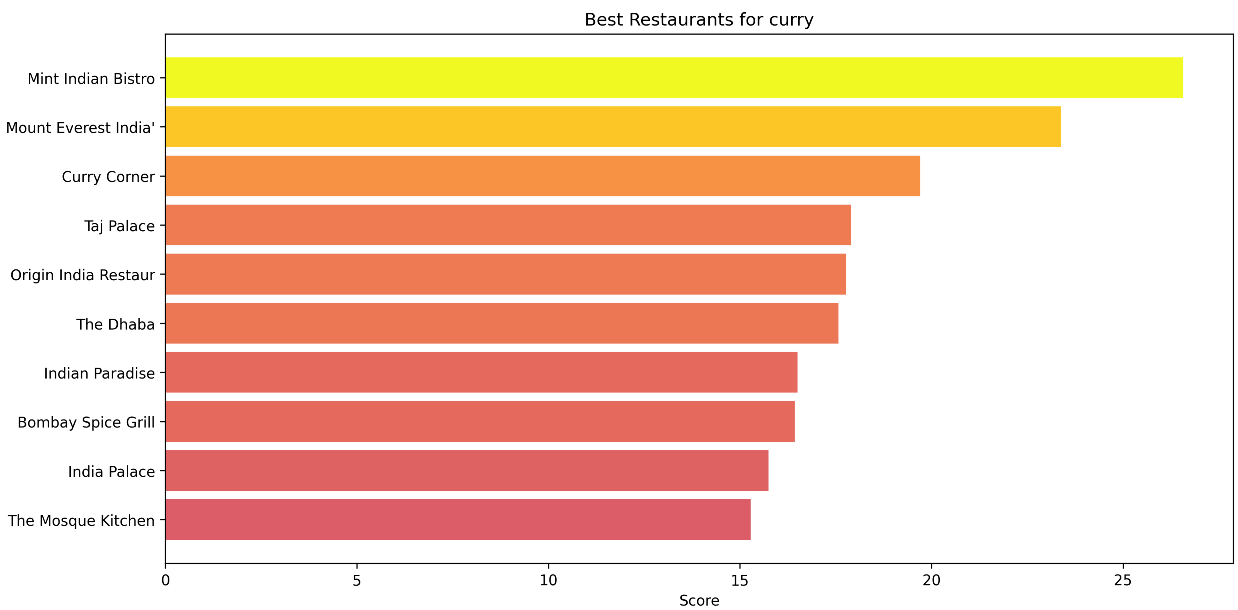

Figure 4: Best Restaurants for Curry

Notable Patterns:

- Mount Everest India's Cuisine and Mint Indian Bistro consistently ranked in top positions across dishes
- India Palace appeared prominently for masala dishes
- Different restaurants excelled for different dishes, validating the dish-specific approach

4 Analysis and Discussion

4.1 Algorithm Performance

The composite scoring functions effectively balanced multiple factors:

Task 4 Success: The popularity ranking aligned with culinary expectations—naan, masala, and curry represent fundamental Indian food categories.

Task 5 Success: Restaurant recommendations varied meaningfully by dish, indicating the algorithm captured dish-specific restaurant strengths rather than just overall restaurant quality.

4.2 Methodological Considerations

- **Sentiment Analysis**: TextBlob provided reasonable sentiment scores, though more sophisticated models could improve accuracy.
- **String Matching**: Simple substring matching may have captured false positives (e.g., masala in chicken tikka masala).
- **Sample Size**: 8,000 reviews provided sufficient coverage for popular dishes but may be limited for rarer items.

4.3 Practical Applications

These results could power:

- **Menu Discovery**: Help newcomers to Indian cuisine identify must-try dishes
- **Restaurant Selection**: Guide diners to restaurants excelling in specific dishes
- **Recommendation Systems**: Provide personalized suggestions based on dish preferences

5 Limitations and Future Work

5.1 Current Limitations

- **Dish Detection**: String matching may miss variations (e.g., naan bread vs. naan)
- **Context Ignoring**: Did not analyze whether mentions were positive/negative in context
- **Temporal Effects**: Ignored changes in restaurant quality over time

5.2 Future Improvements

1. **Advanced NLP**: Use named entity recognition for better dish extraction
2. **Context Analysis**: Analyze sentiment specifically around dish mentions
3. **User Personalization**: Incorporate individual preference profiles
4. **Multi-Cuisine**: Extend the framework to other cuisine types

6 Conclusion

By combining frequency analysis, sentiment scoring, and review quality metrics, I successfully created a dual recommendation system for Indian cuisine. The Task 4 popularity rankings revealed intuitive results with naan and curry-based dishes dominating, while Task 5 restaurant recommendations showed meaningful variation by dish type.

The approach demonstrates how review mining can provide practical dining guidance. The dish-specific restaurant rankings proved more useful than generic restaurant ratings, as they account for restaurants' particular strengths with individual dishes.

These techniques could readily extend to other cuisines and integrate into restaurant discovery applications, providing data-driven dining recommendations that go beyond simple overall ratings.