Gastronomix

The Recipe for Restaurant Success

A dynamic and competitive industry where success depends on spotting opportunities, overcoming challenges, and adapting to constant change.

Presented by Group 10

Getting Started

Gastronomix

Why the Restaurant Industry Remains Promising

Source: WebstaurantStore - Restaurant Industry Statistics (2024)

01

High-Income Customer Base

- Households earning over $100,000 contribute to 60% of restaurant spending.
- Households earning under $50,000 contribute 20%.

02

Strong Business Revenues

- 78% of foodservice businesses earn between $100,000 and $5 million annually.

03

Frequent Dining Habits

- Average American eats out 5–6 times per week.

04

Annual Spending on Dining

- Americans spend over $2500 each year dining out.

Gastronomix

Presentation Overview

Understanding the Restaurant Industry

- Analyzing key factors that affect restaurant revenue,
- Identifying patterns or groupings among restaurants,
- Providing recommendations to boost revenue.

Introducing GastronoMix

- Predict revenue for your restaurant startup ideas or existing restaurants,
- Get actionable insights to increase your revenue.

Behind the Scenes

- Models we used for predictions.
- Evaluation of model accuracy to ensure trustworthy results.

Gastronomix

Dataset Overview

17
Variables

Location, Cuisine
Parking Availability

Target -
Annual Revenue

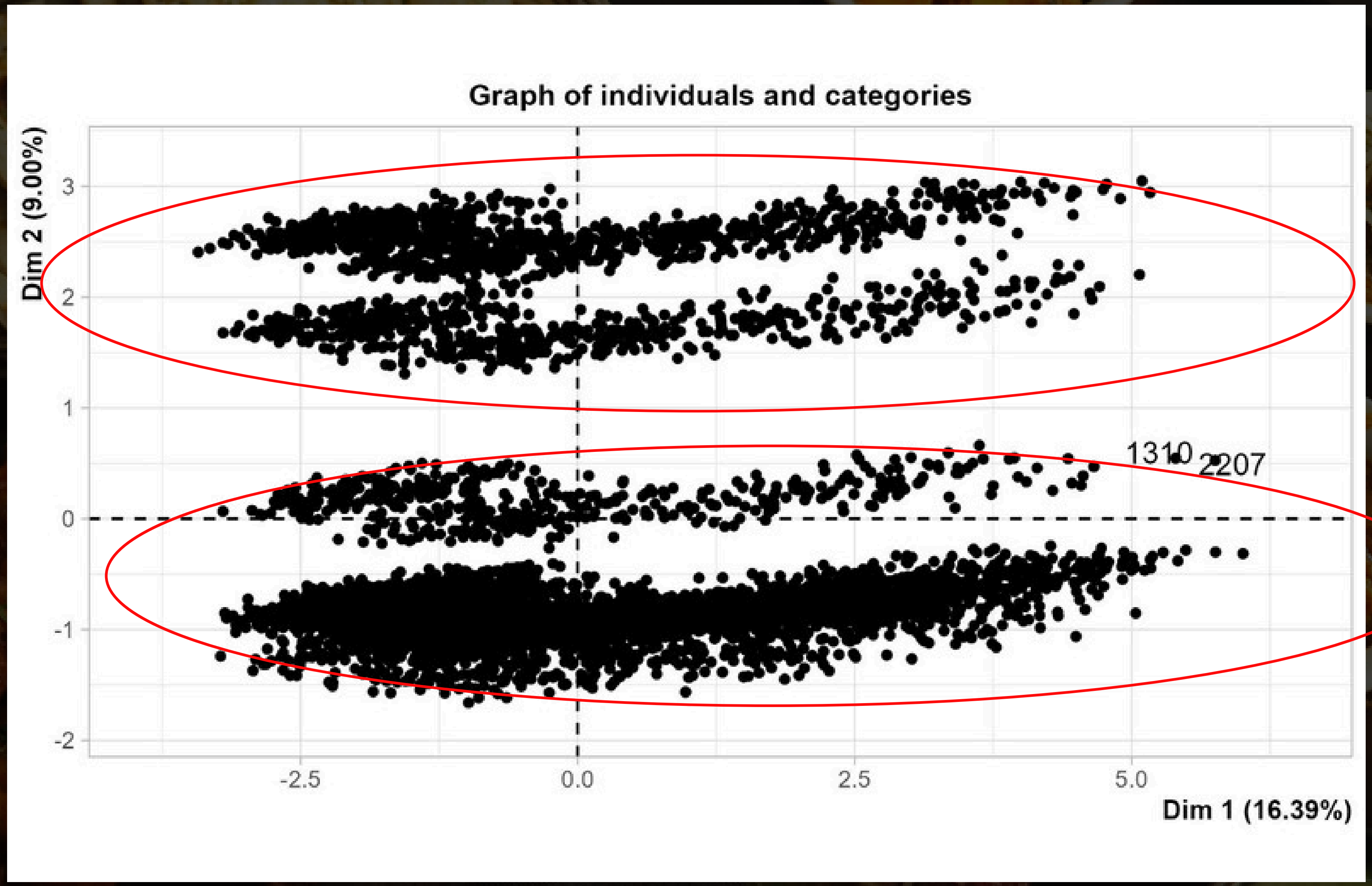
8368
Restaurants

Rating, Seating Capacity, Average Meal Price, Marketing Budget, Chef Experience , Number of Reviews, Average Review Length, Ambience Score, Service Quality Score, Reservations,Followers

- No missing values
- No duplicate records
- 59 outliers detected using Isolation Forest

Feature Engineering

- Review Quality Score= Rating * Average Review Length
- Average Meal Price categorized into:
  - Low (<$30)
  - Medium ($30—$60)
  - High (>$60)



Can we discover groups of restaurants that follow a similar path to success and use that to boost their revenue?

Gastronomix

Cluster Analysis

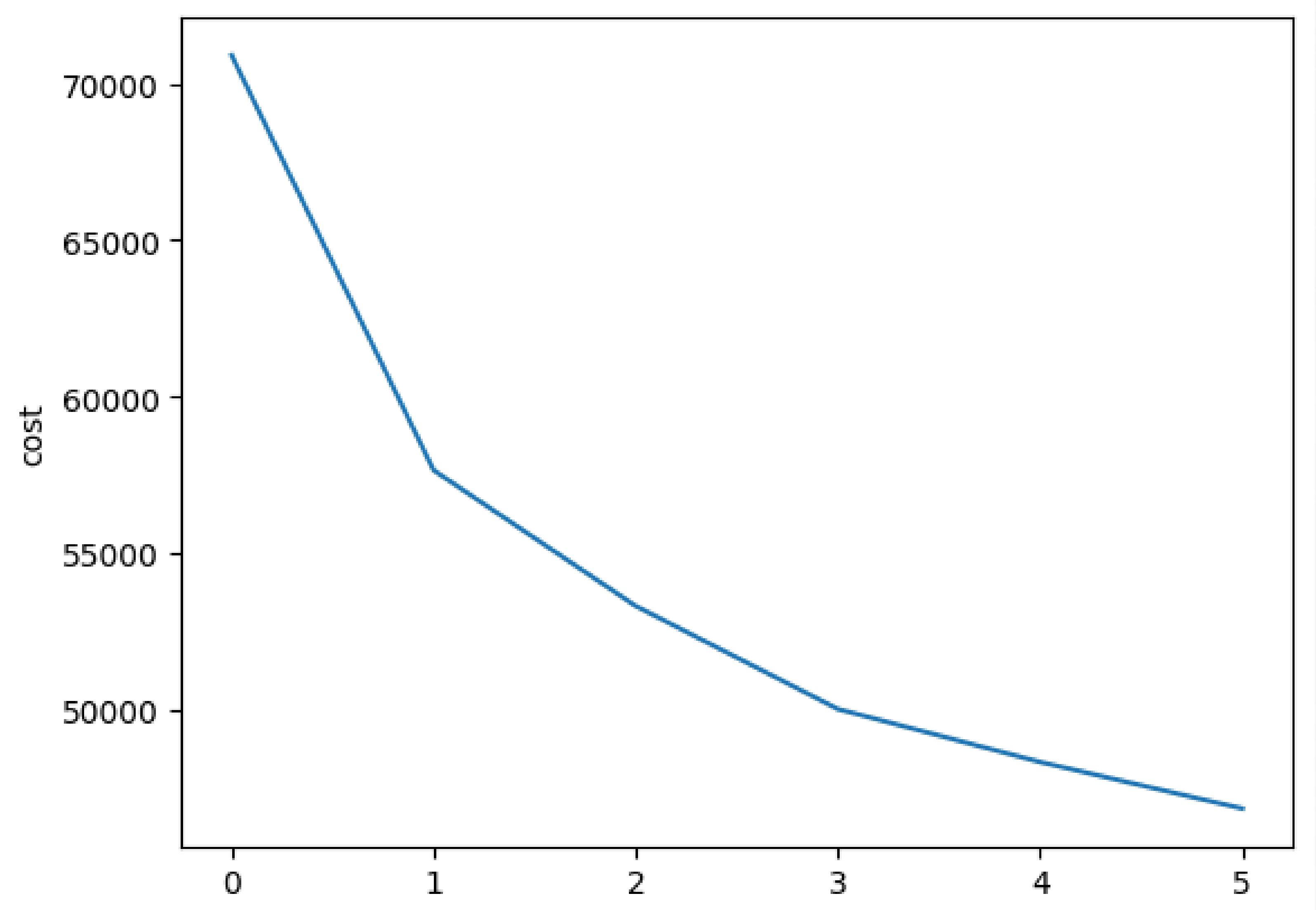
Since we have mixed data,We performed Kprotype Clustering

- Data naturally divides into 2 clusters, but separation is weak (Average Silhouette Score = 0.5032).

Cluster 01
29.48%

Cluster 02
70.52%

Variable	Cramer's V
Cuisine	0.0308
Location	0.8652
Meal Price Category	0.0103
Parking Availability	0.0057



KneeLocator class will detect elbow at k = 2 clusters

Cramér's V = 0.87

Clusters strongly align with location

Location-Based Cluster Analysis

Urban vs. Non-Urban Restaurant Segmentation

Cluster 01

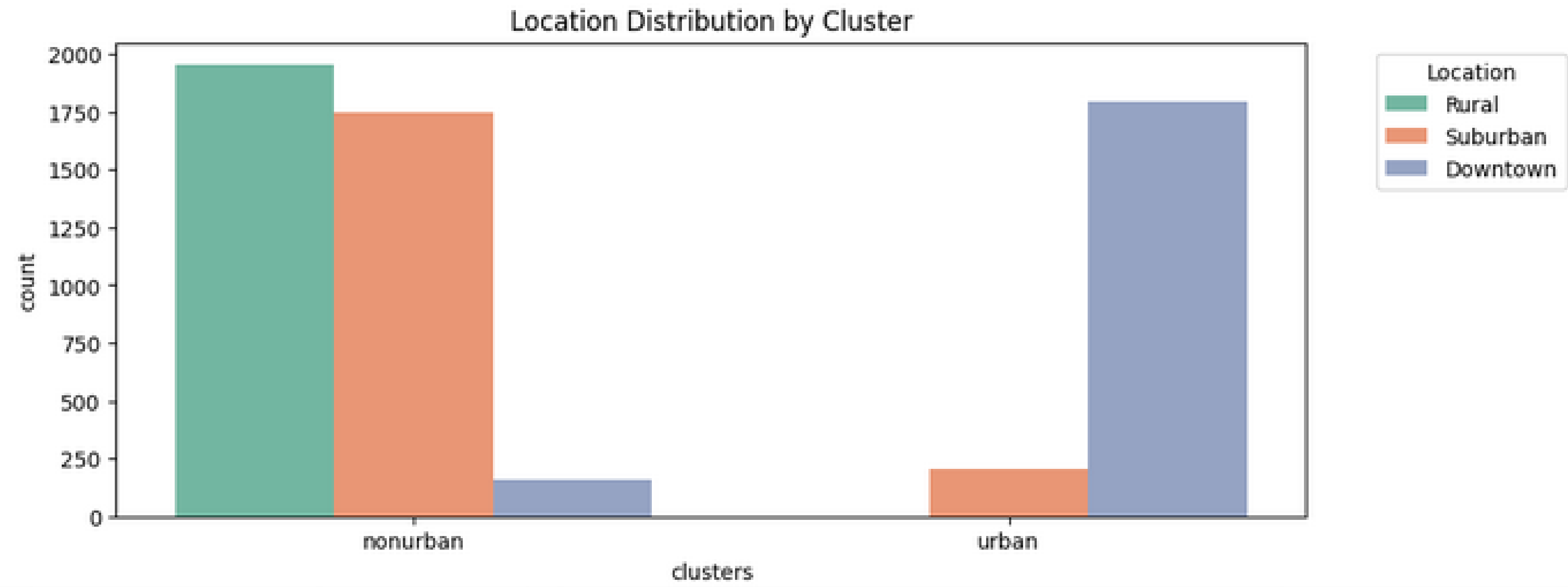
Rural (50.6%)
Suburban (45.6%)
Downtown (3.9%)

→ **Nonurban**

Cluster 02

Rural (0%)
Suburban (9.6%)
Downtown (90.4%)

→ **Urban**



Cluster Analysis – Variable Significance

Identifying Key Differentiators Across Clusters

Variable	P Value	Significant ✅ Not ❌
Rating	0.0000	✅
Seating Capacity	0.0000	✅
Marketing Budget	0.0000	✅
Social Media Followers	0.0000	✅
Chef Experience	0.0883	❌
No of Reviews	0.9224	❌

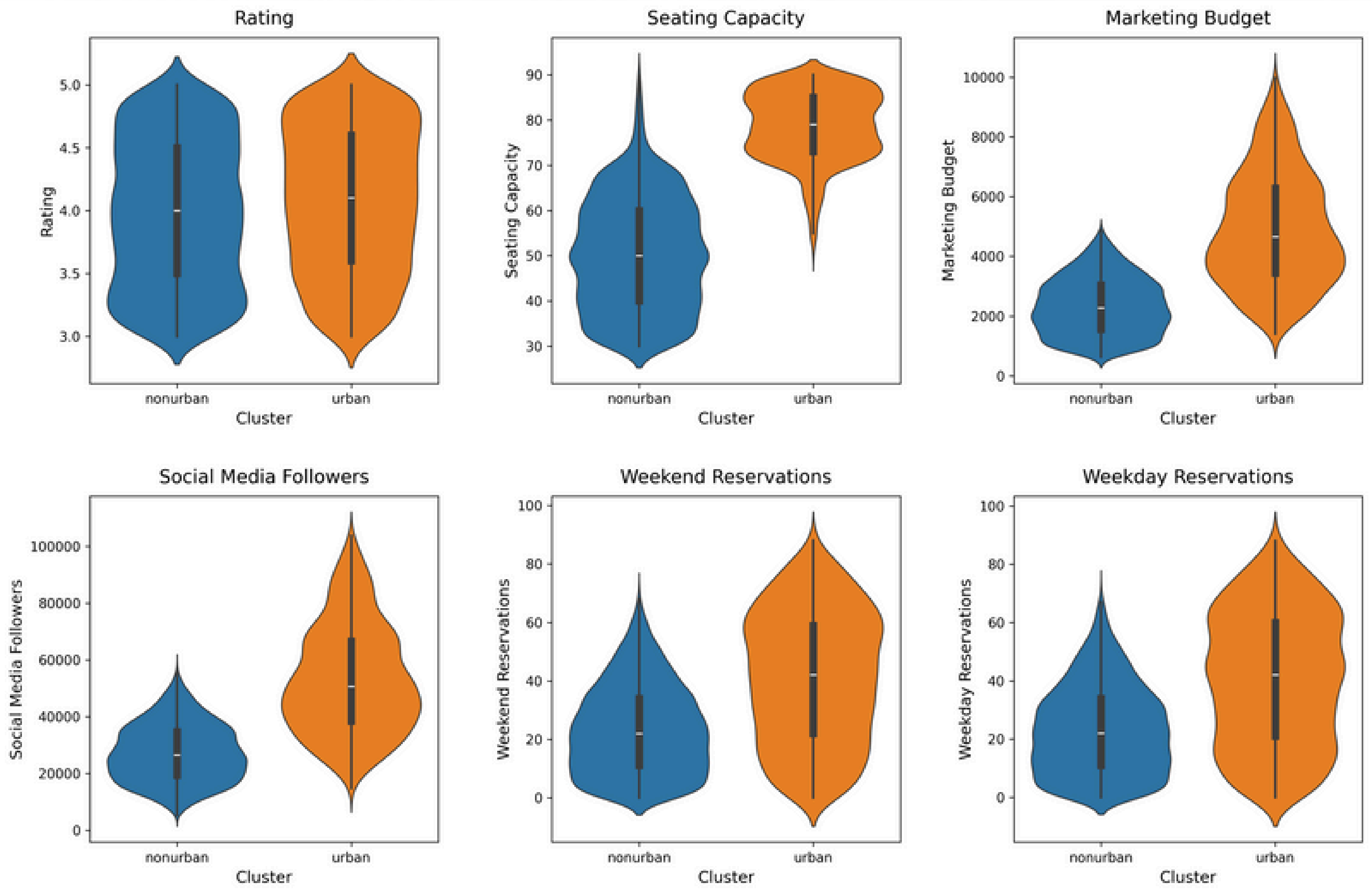
Variable	P Value	Significant ✅ Not ❌
Ambience Score	0.4744	❌
Service Quality	0.6324	❌
Weekend Resevation	0.0000	✅
Weekday Reservation	0.0000	✅
Review Quality	0.6368	❌

- The clusters are primarily differentiated by Rating, Seating Capacity, Marketing Budget, Social Media Followers, and Reservation rates (Weekend/Weekday).

Cluster Profiling – Urban vs. Nonurban

Key Differentiators Based on Significant Variables

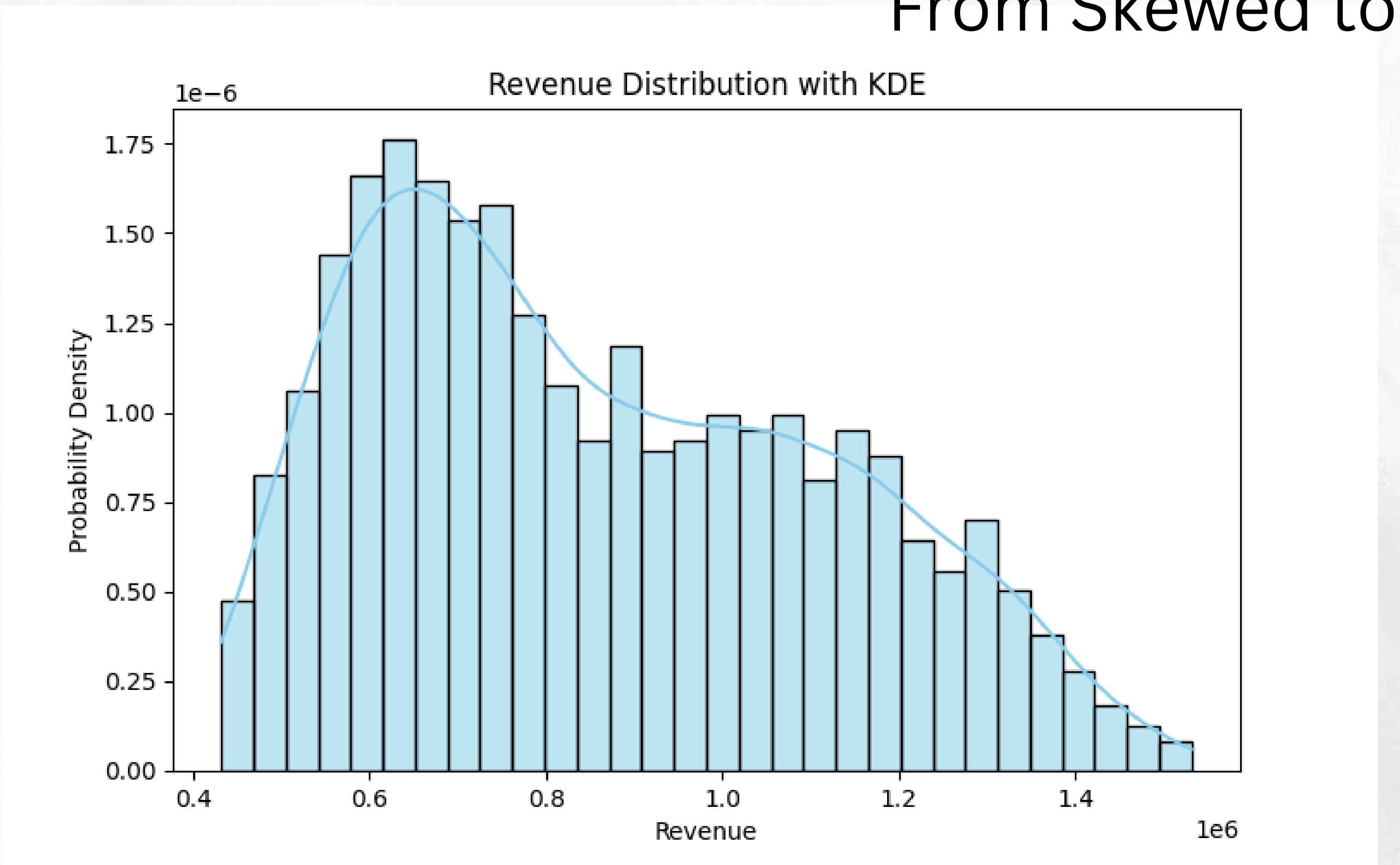
Urban clusters show higher ratings, larger social media followings, greater seating capacity, increased marketing budgets, and more reservations compared to Nonurban areas.

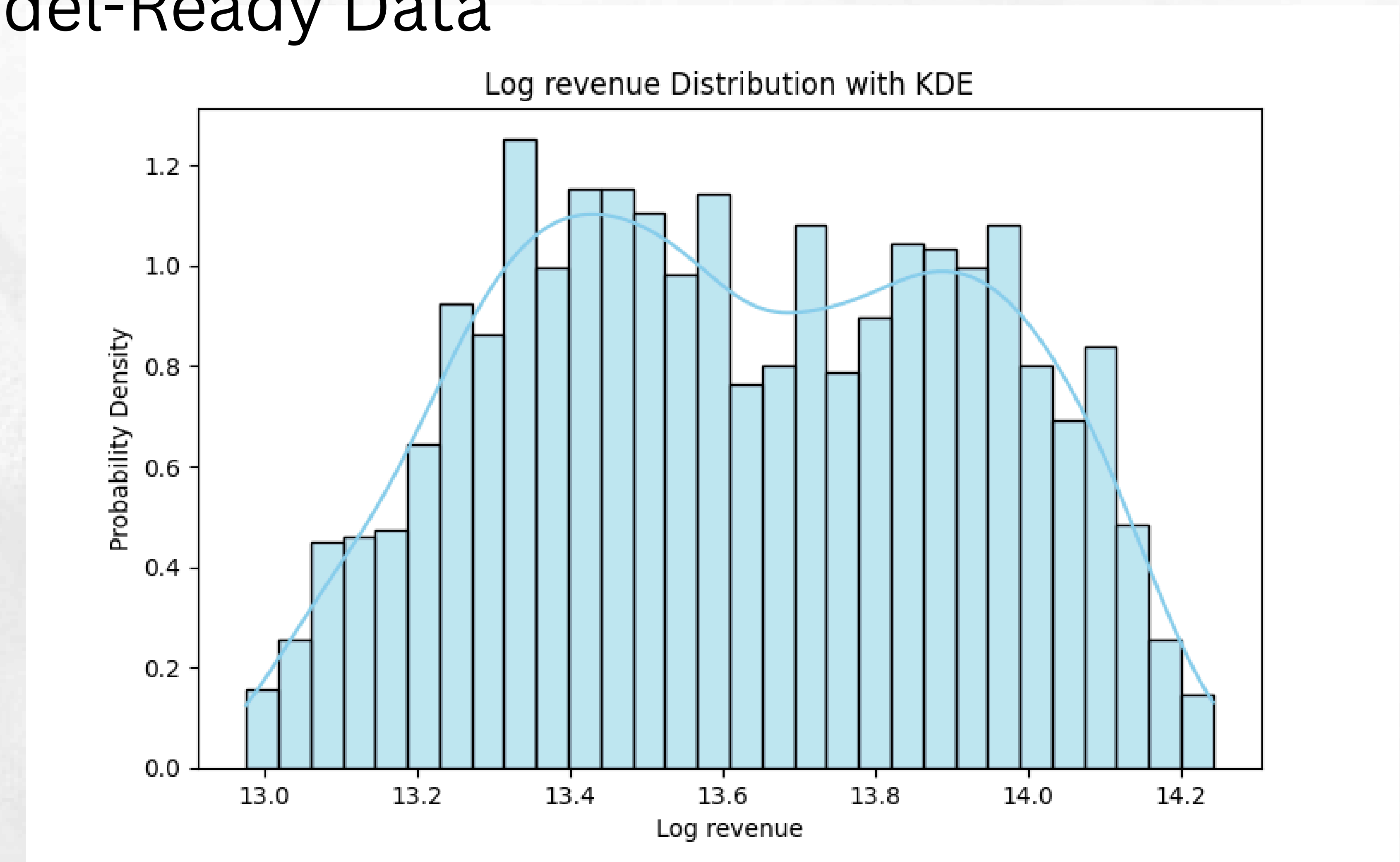


Urban Restaurants

Fixing Revenue Distribution with Log Transform

From Skewed to Model-Ready Data





🔴 Severe right-skew (most data crammed left)

📏 Huge range (0.4x10^6 to 1.66x10^6)

🔵 Near-normal distribution (symmetric curve)

✂️ Compressed range (13.0 to 14.2) → better for modeling

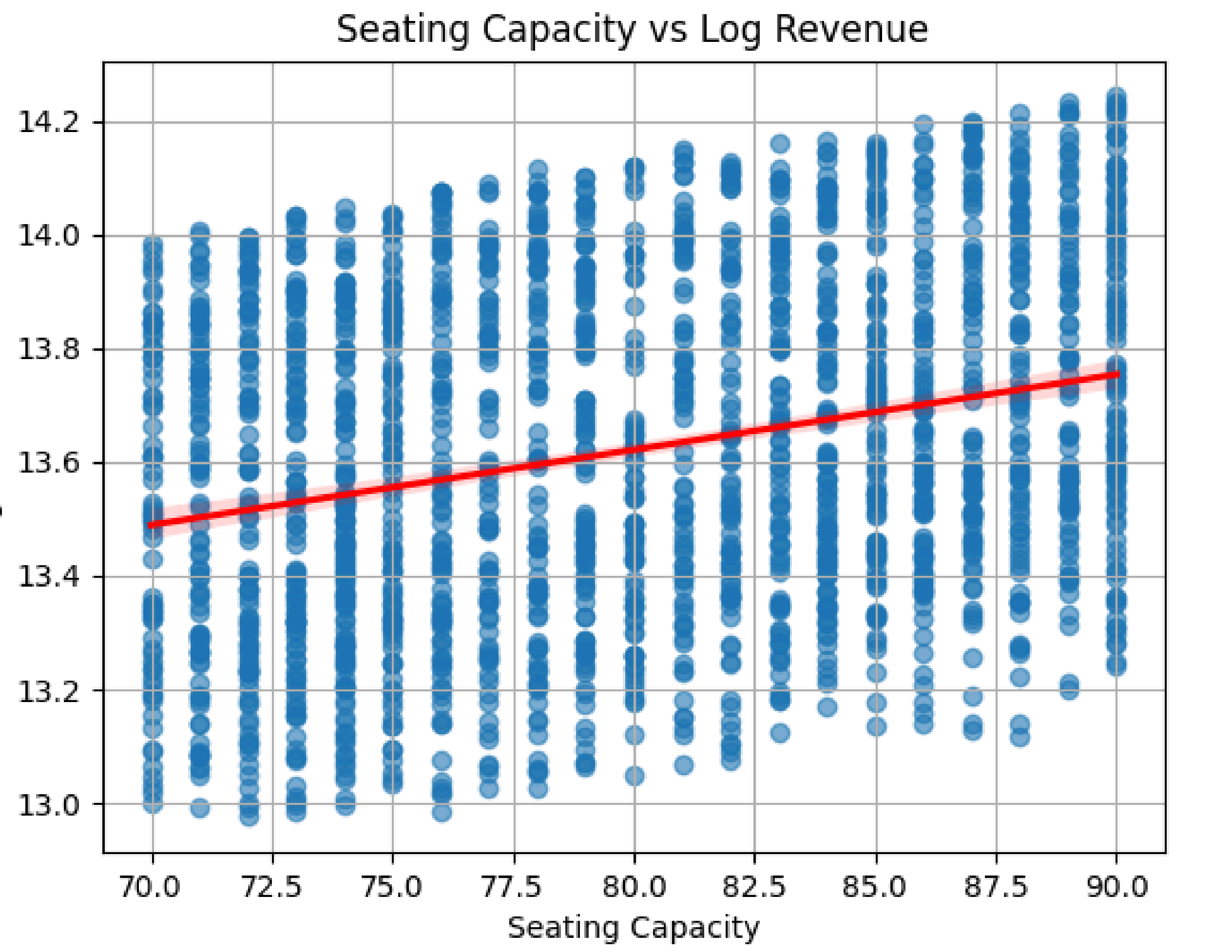
Why It Matters

✓ Enables reliable statistical tests

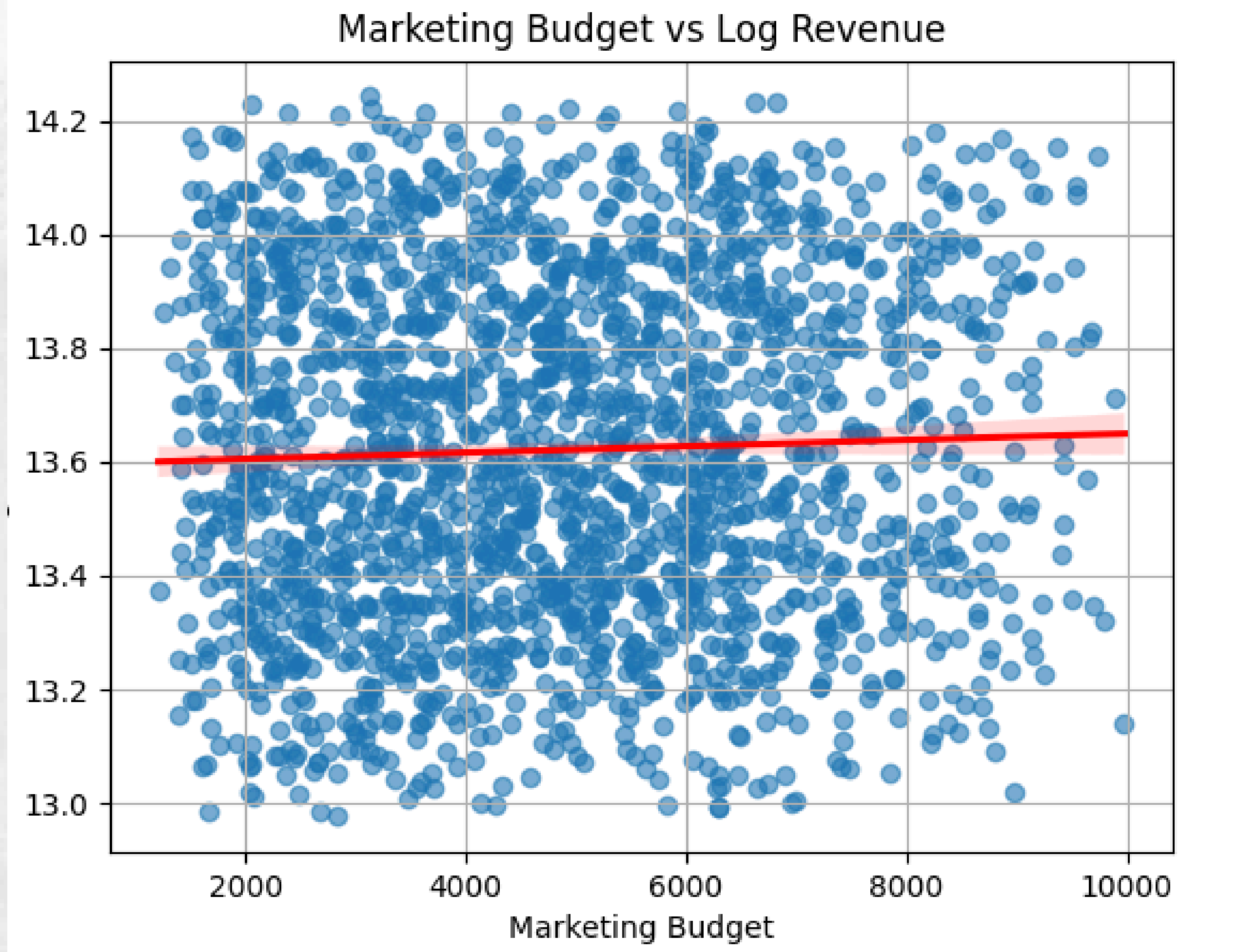
✓ Reduces outlier impact

✓ Improves linear model performance

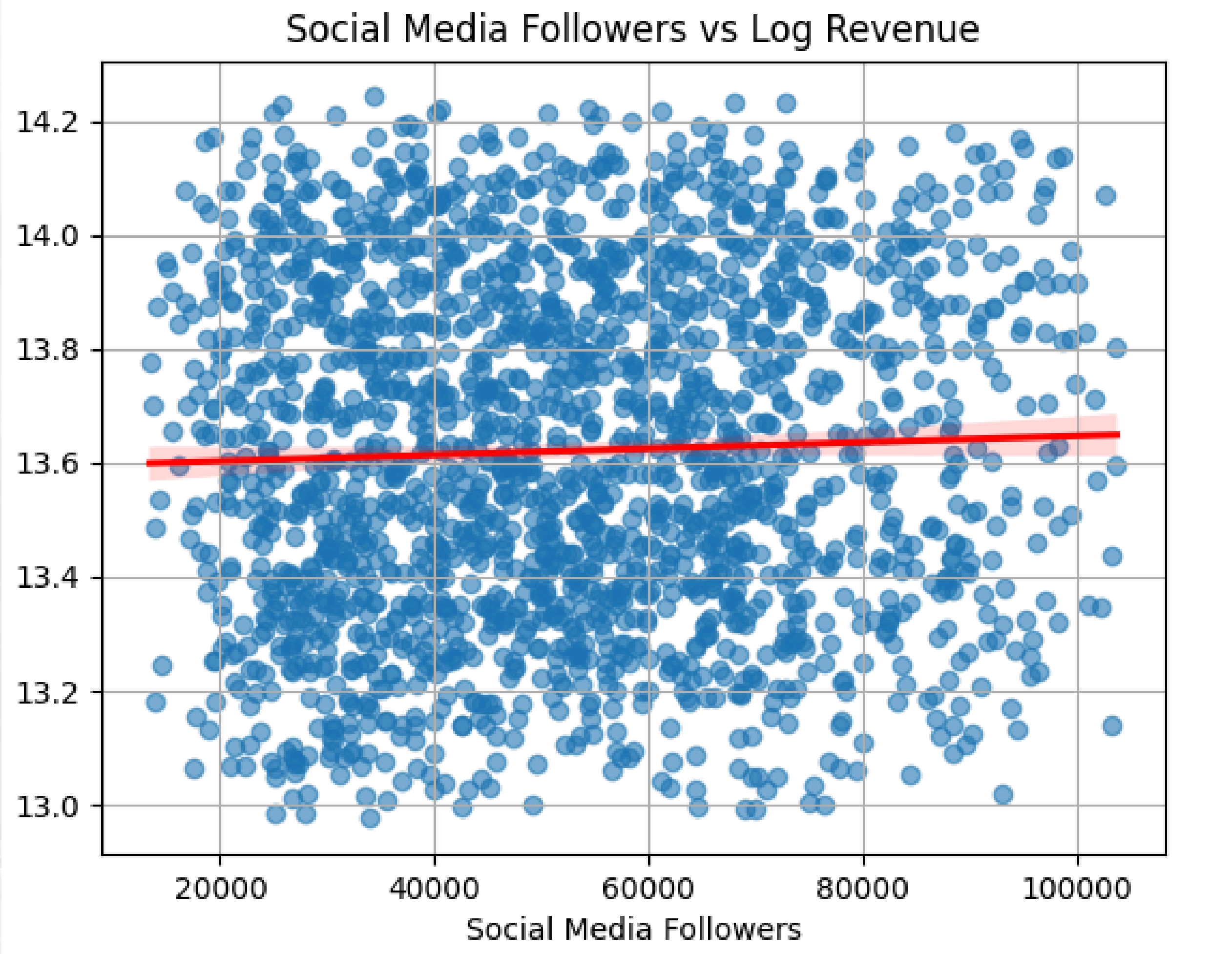
Key Drivers of Revenue in Urban Restaurants



Shows a moderate positive relationship with revenue.



Minimal visible effect on revenue.

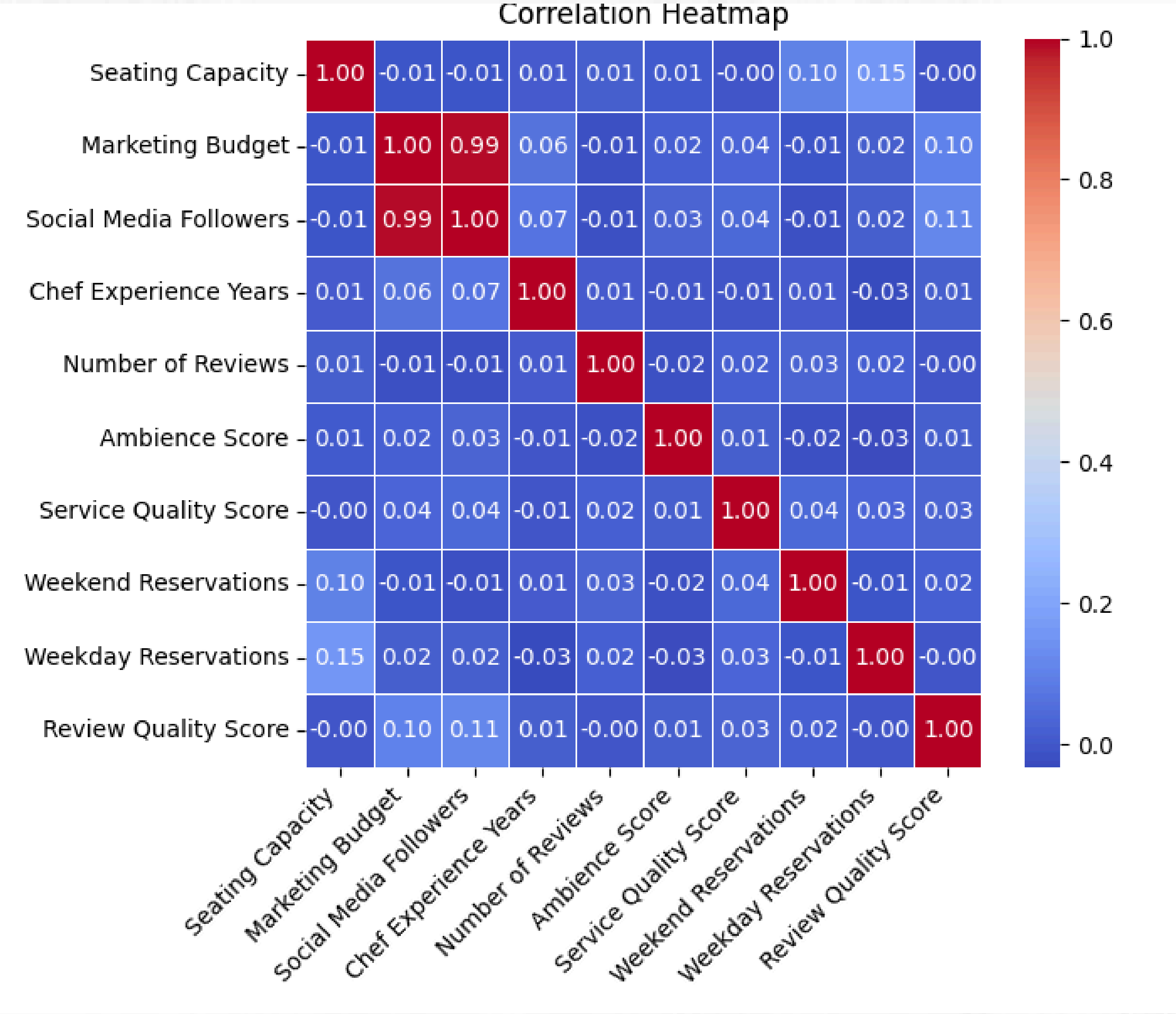


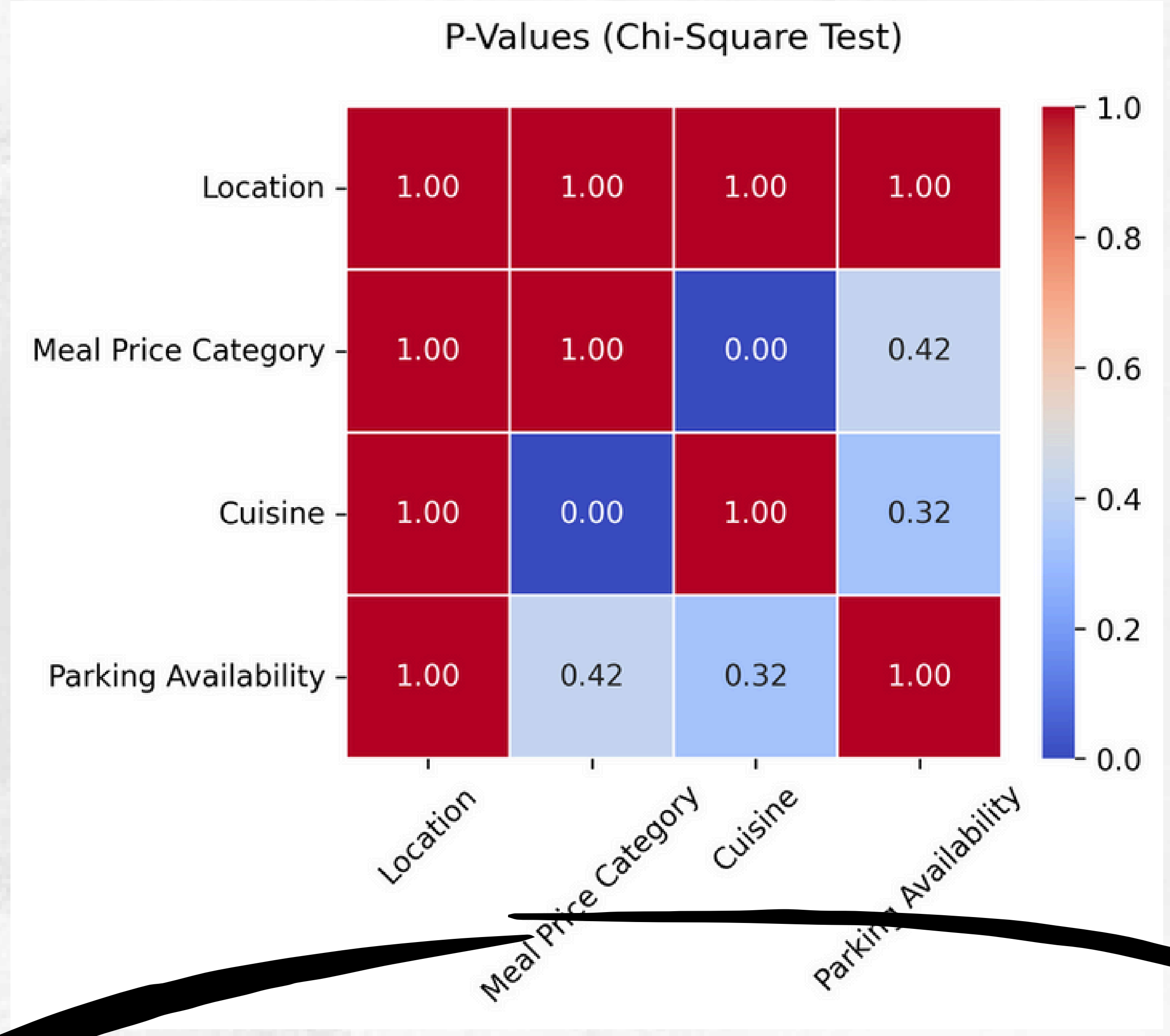
Weak upward trend, low impact.

seating capacity has the most consistent positive association with revenue, while marketing and social media show weak effects.

Correlation & Multicollinearity Analysis

Identifying Key Relationships in Urban Restaurants





Strong Positive Correlation (>0.5):

- Marketing Budget ↔ Social Media Followers (0.99)

Since $p < 0.05$

- Strong Dependence Between Cuisine & Meal Price

🚨 Multicollinearity Alert

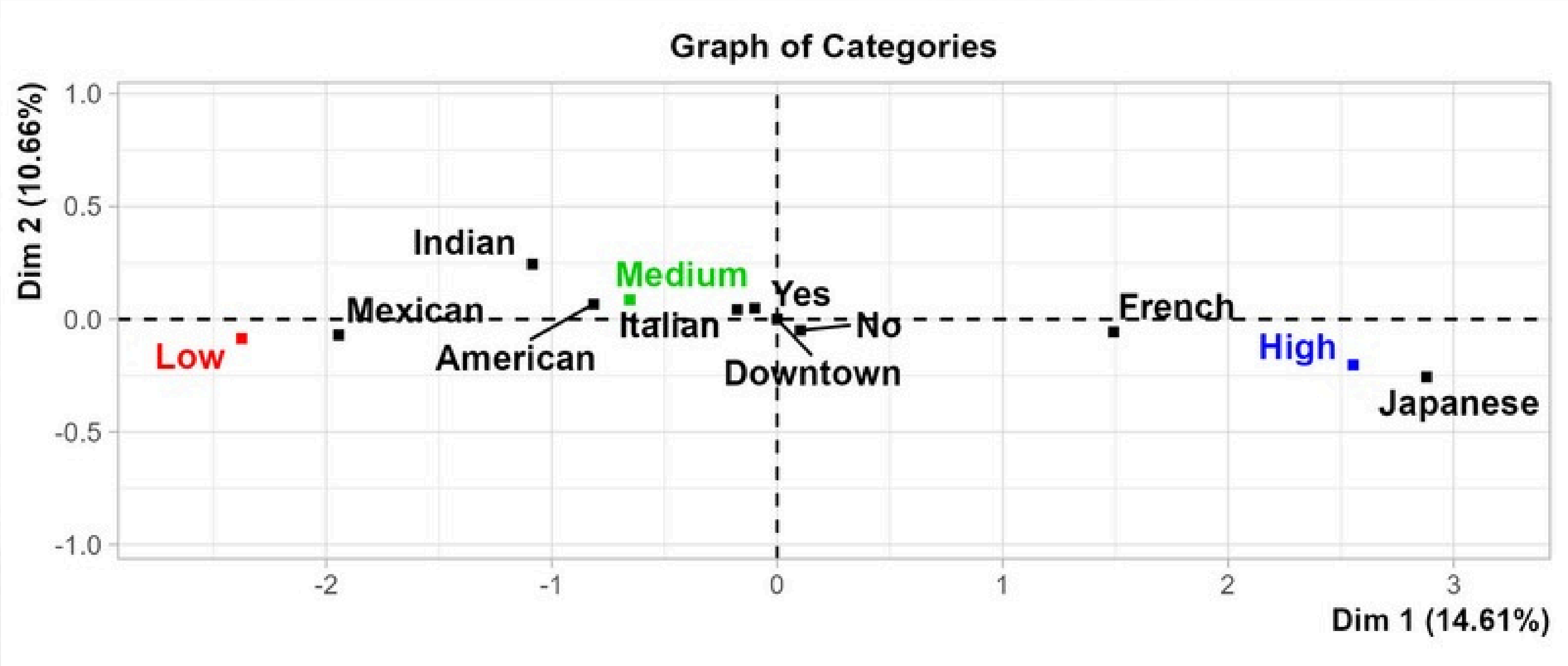
The Power of Cuisine in Urban Restaurant Pricing

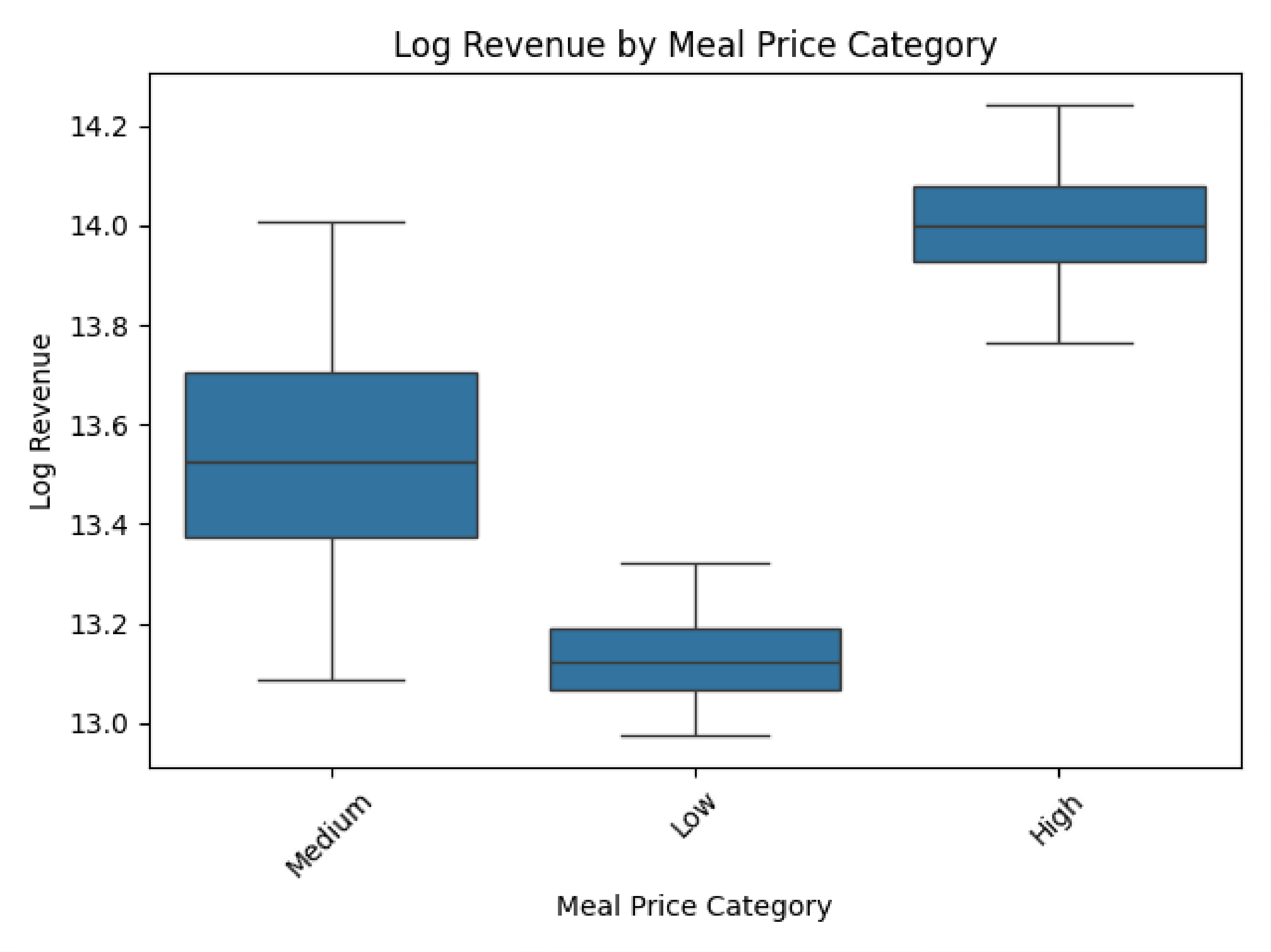
- Cuisine type has a clear association with pricing category.

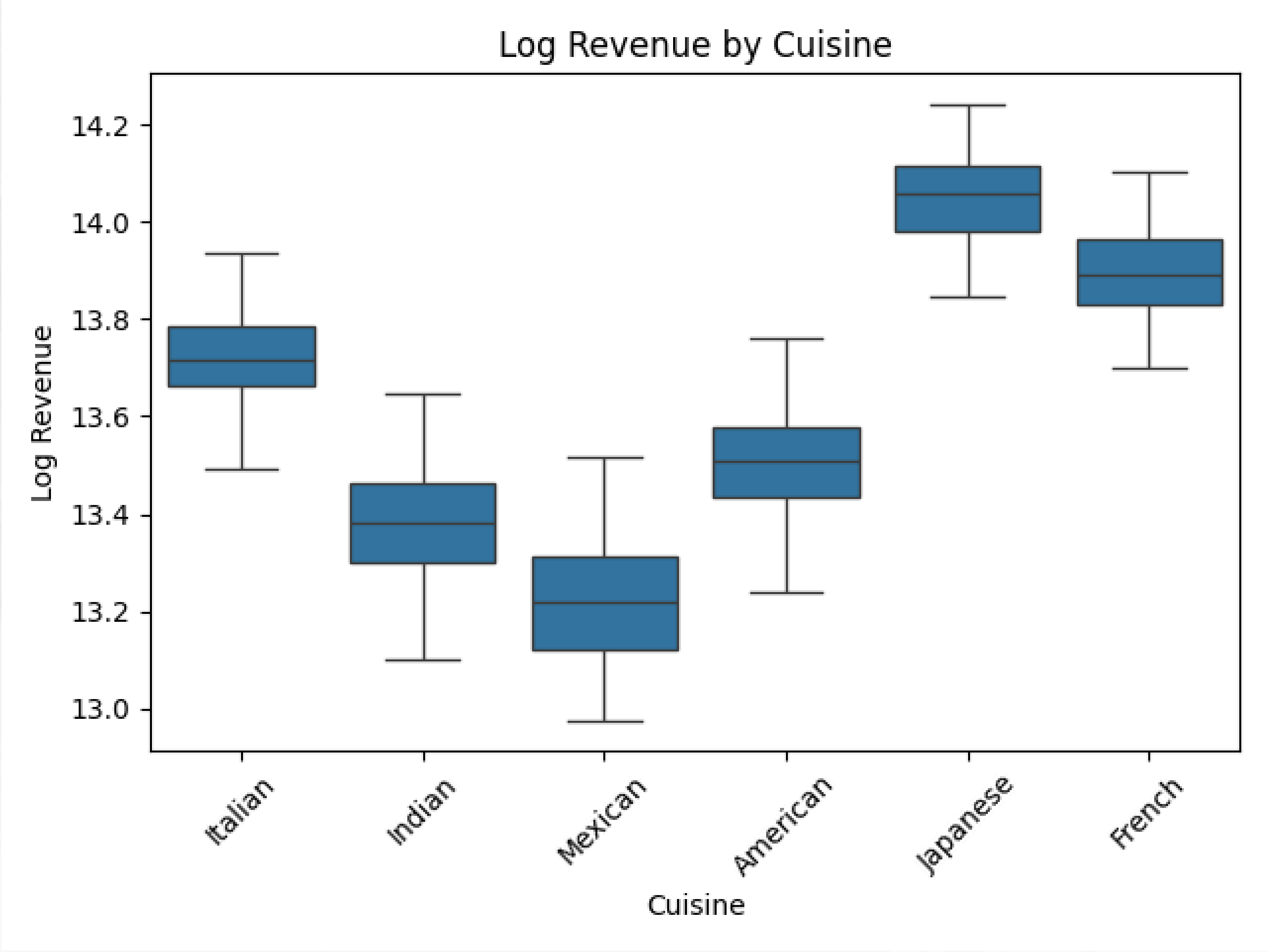
High Price ➤ Japanese , French

Medium Price ➤ Indian,Amaerican,Italian

Low Price ➤ Mexican





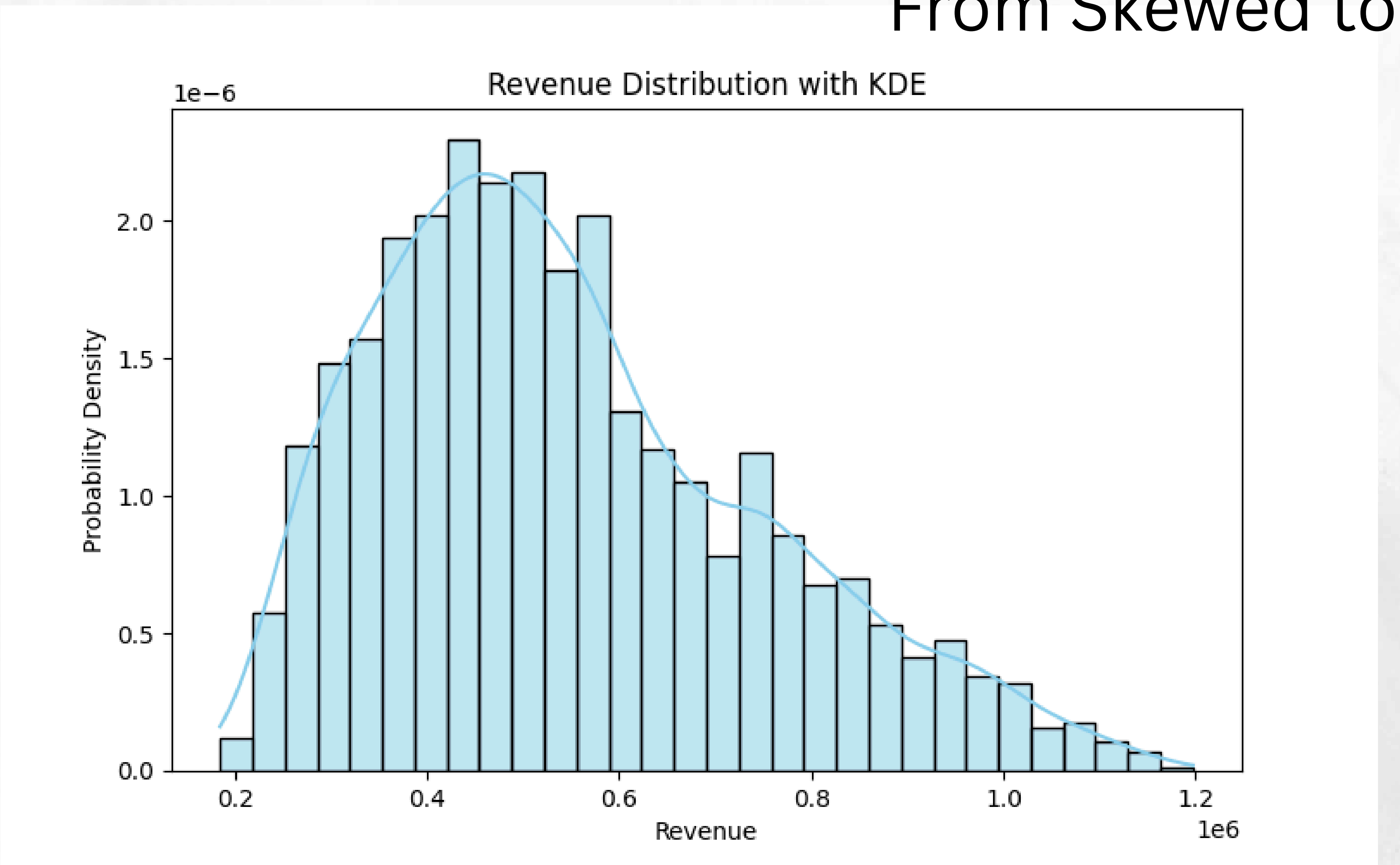


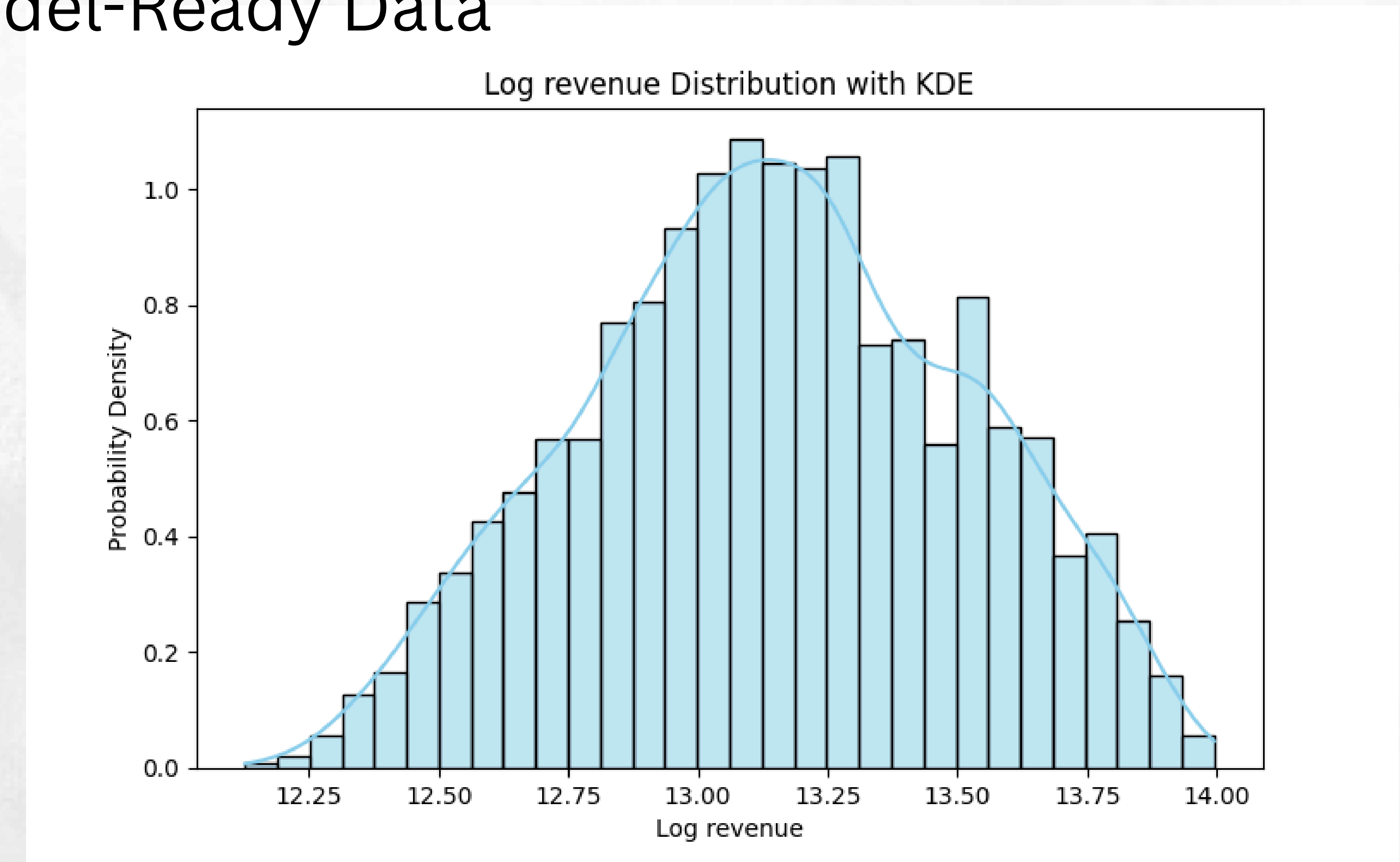
- Japanese and French restaurants typically have high meal prices and correlate with higher revenues

Nonurban Restaurants

Fixing Revenue Distribution with Log Transform

From Skewed to Model-Ready Data





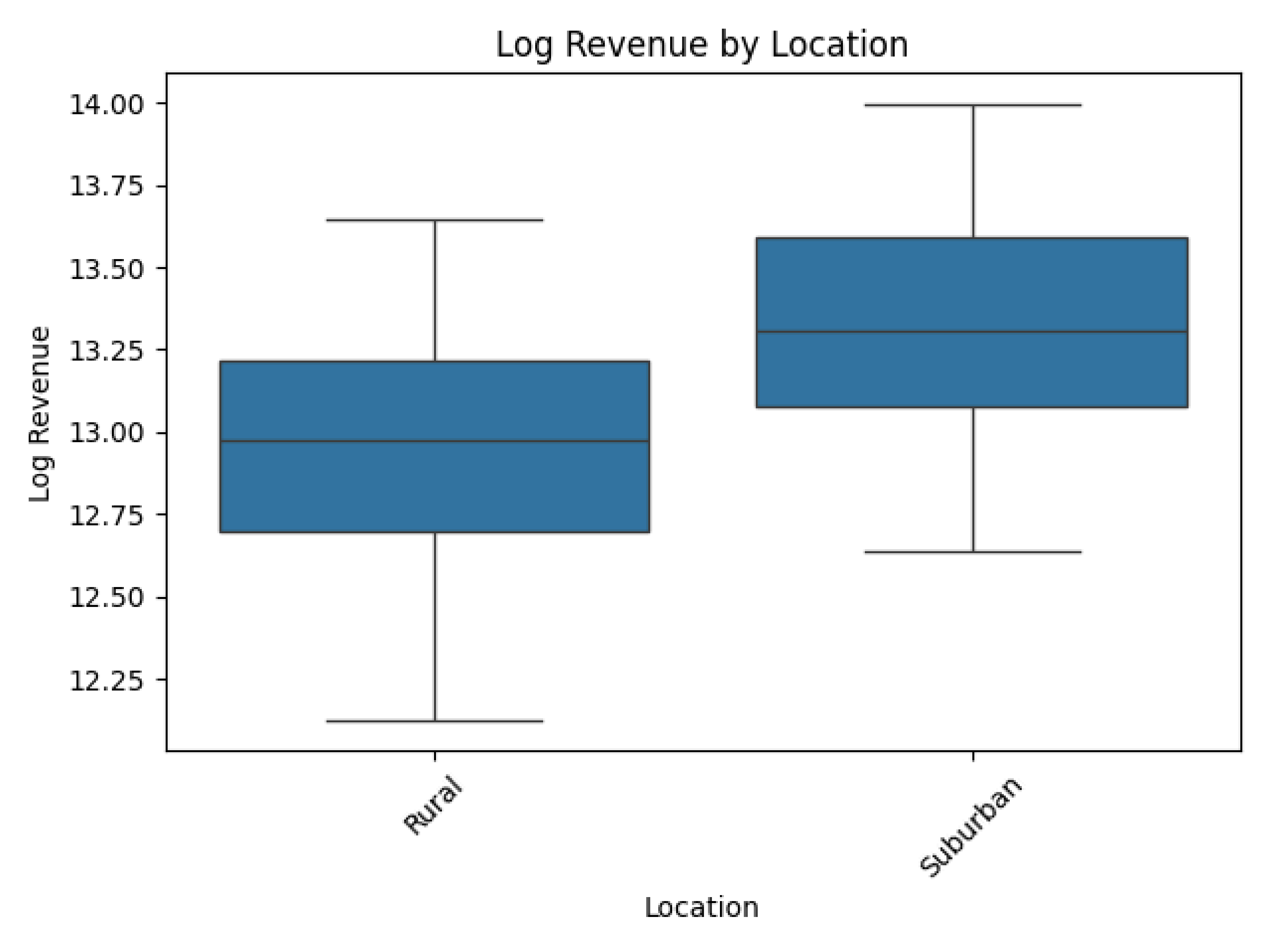
🔴 Severe right-skew (most data crammed left)
📏 Huge range (0.4x10^6 to 1.66x10^6)

🔵 Near-normal distribution (symmetric curve)
✂️ Compressed range (13.0 to 14.2) → better for modeling

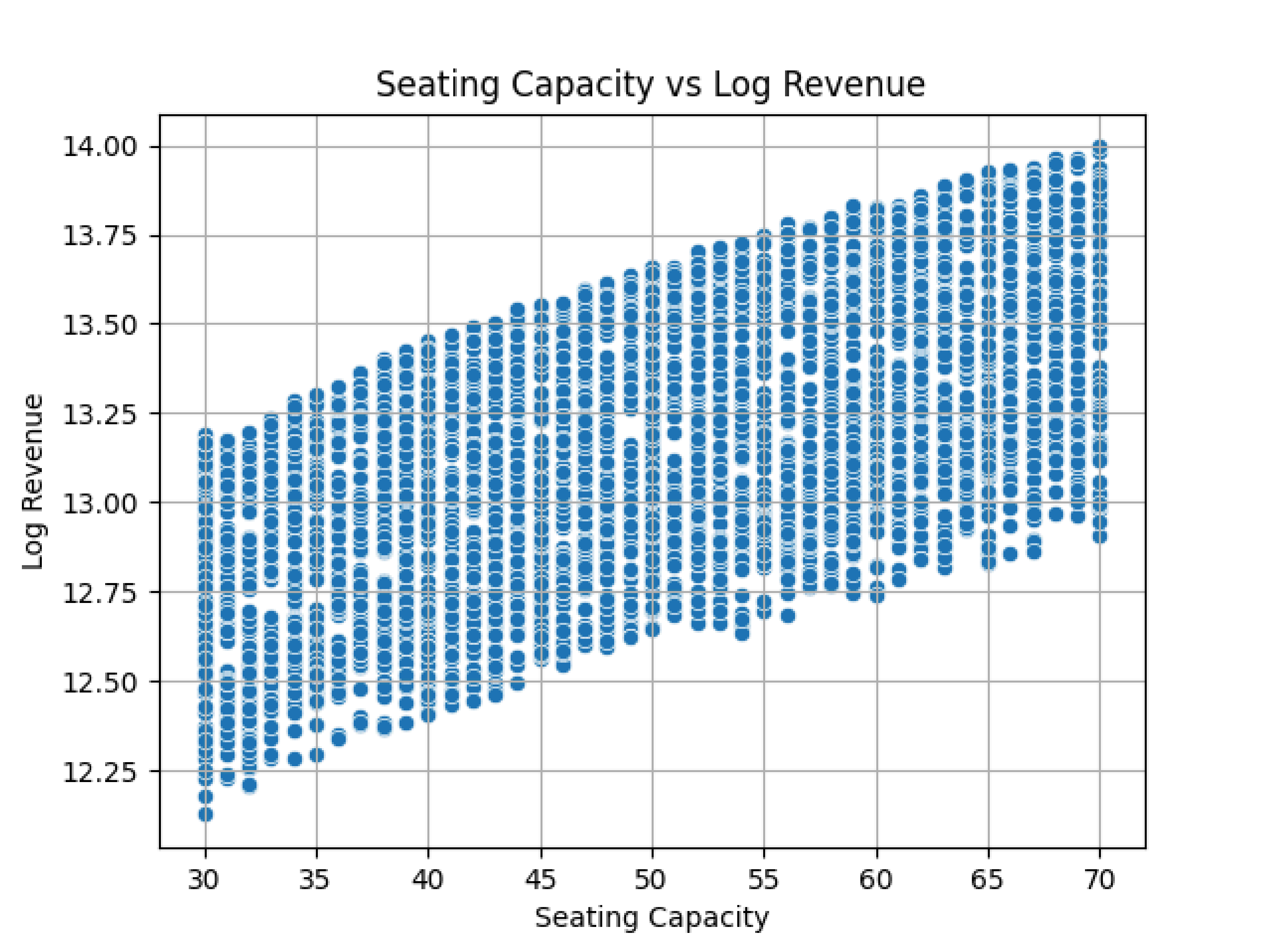
Why It Matters

✓ Enables reliable statistical tests
✓ Reduces outlier impact
✓ Improves linear model performance

Key Drivers of Revenue in Nonurban Restaurants



Suburban restaurants tend to have higher log revenue than rural ones

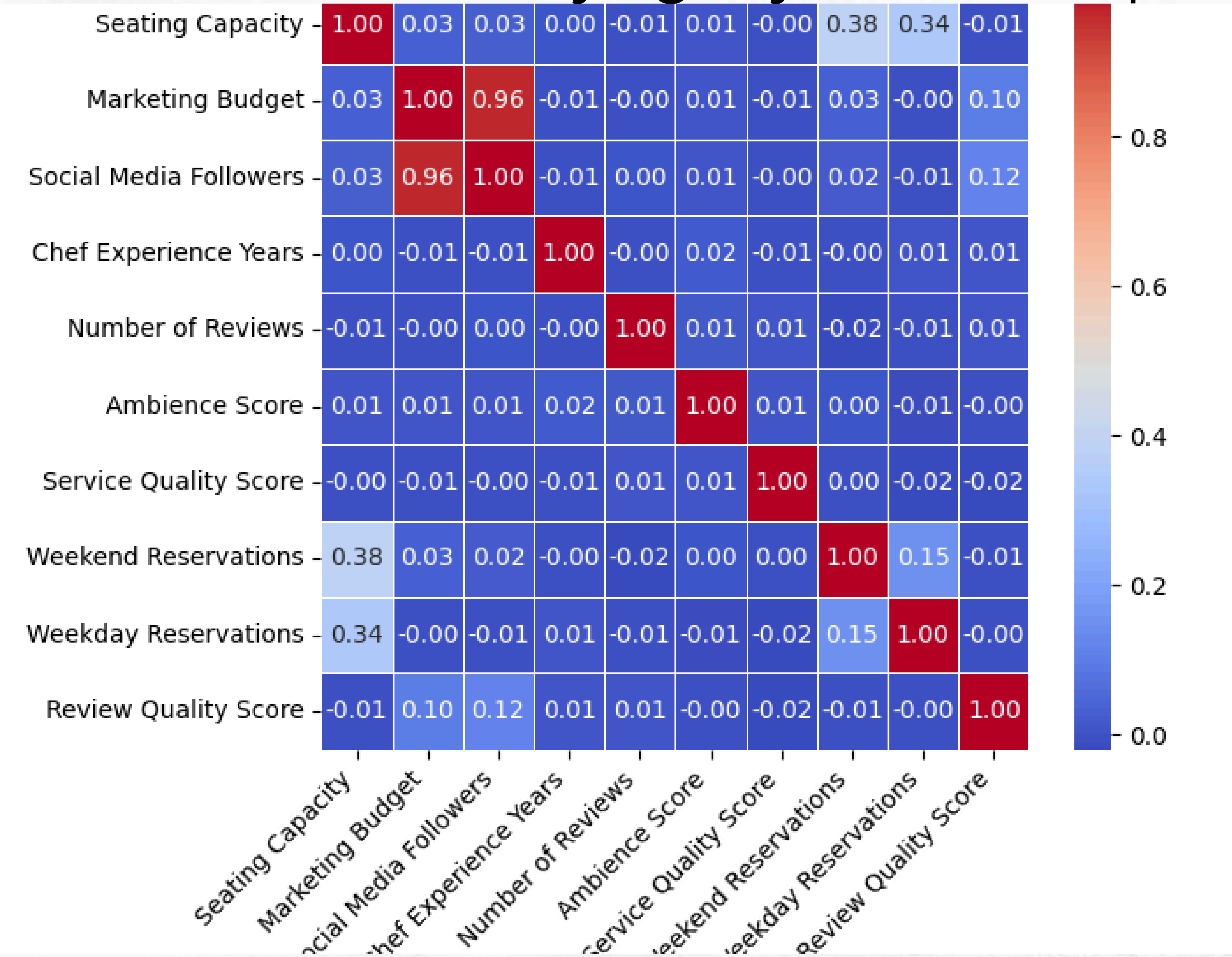


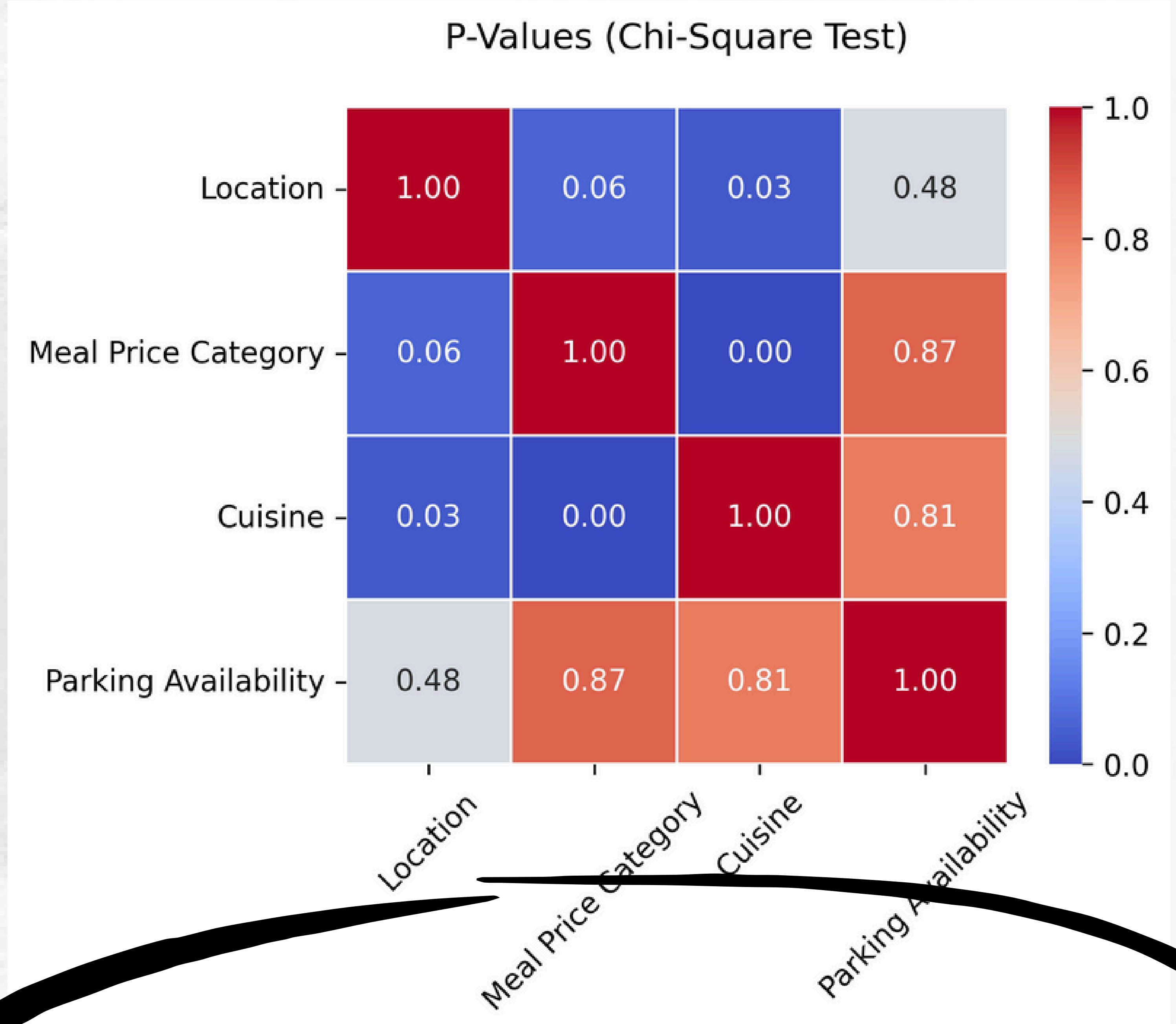
Seating capacity shows a strong positive association with revenue

Seating capacity is a key revenue driver across nonurban locations, with suburban restaurants consistently outperforming rural ones.

Correlation & Multicollinearity Analysis

Identifying Key Relationships in Nonurban Restaurant Data





Strong Positive Correlation (>0.5):

- Marketing Budget ↔ Social Media Followers (0.96)

Since $p < 0.05$

- Strong Dependence Between Cuisine & Meal Price

🚨 Multicollinearity Alert

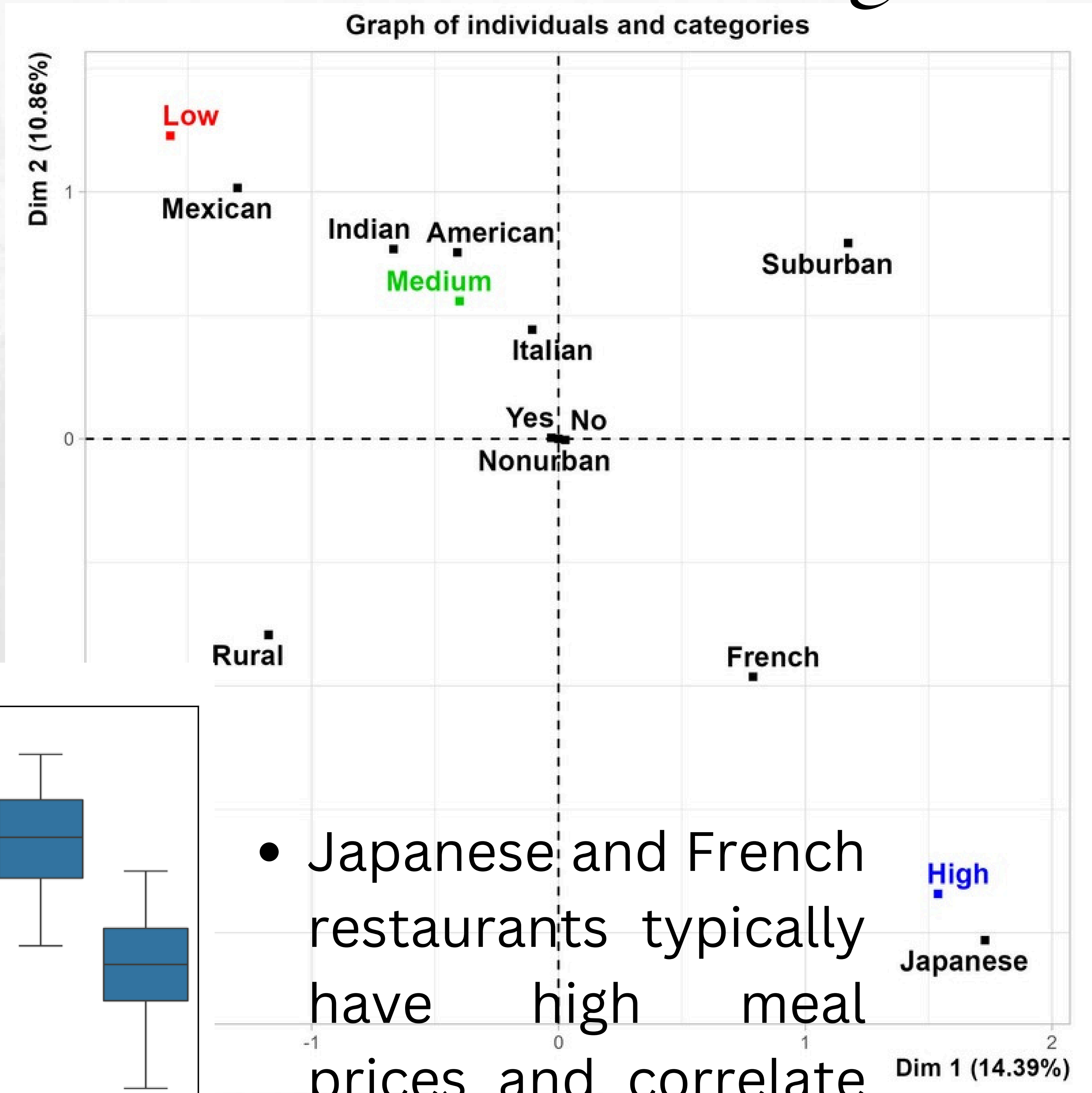
The Power of Cuisine in Nonurban Restaurant Pricing

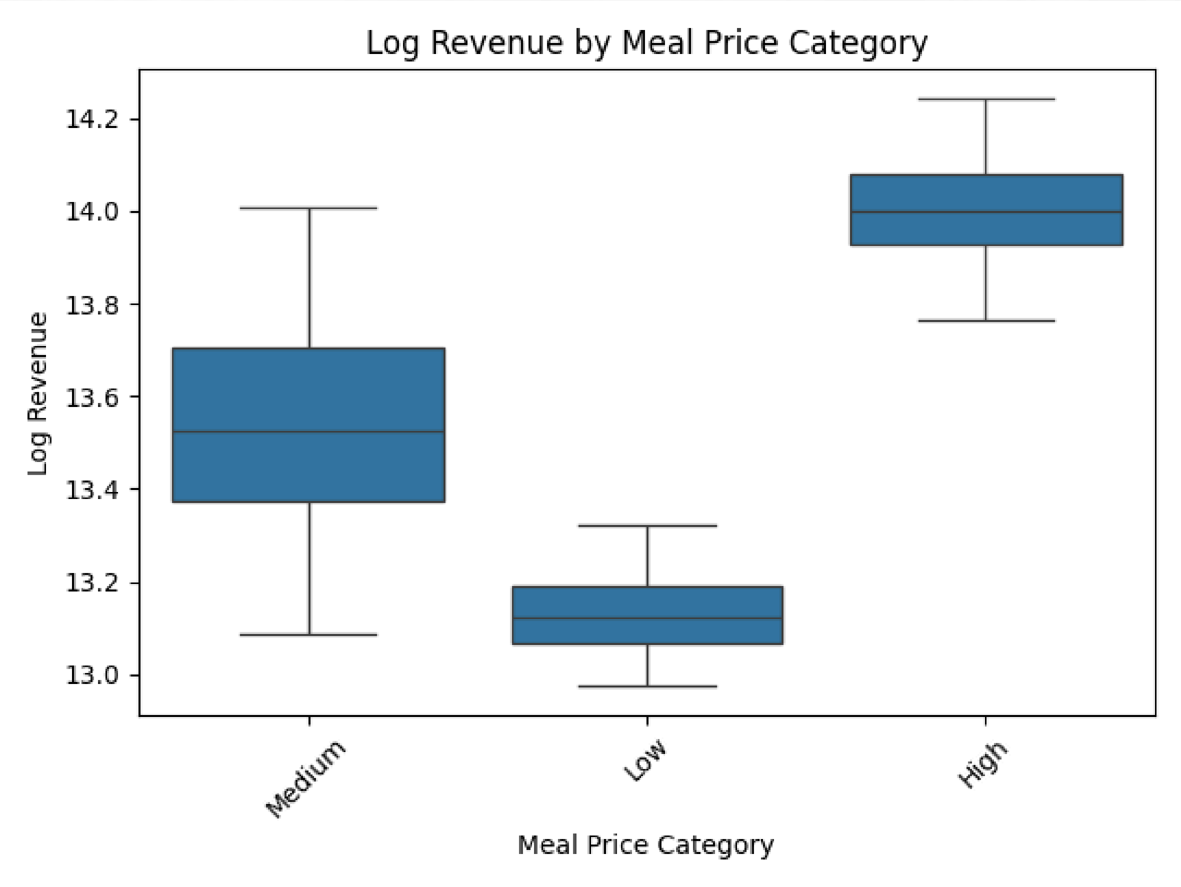
- Cuisine type has a clear association with pricing category.

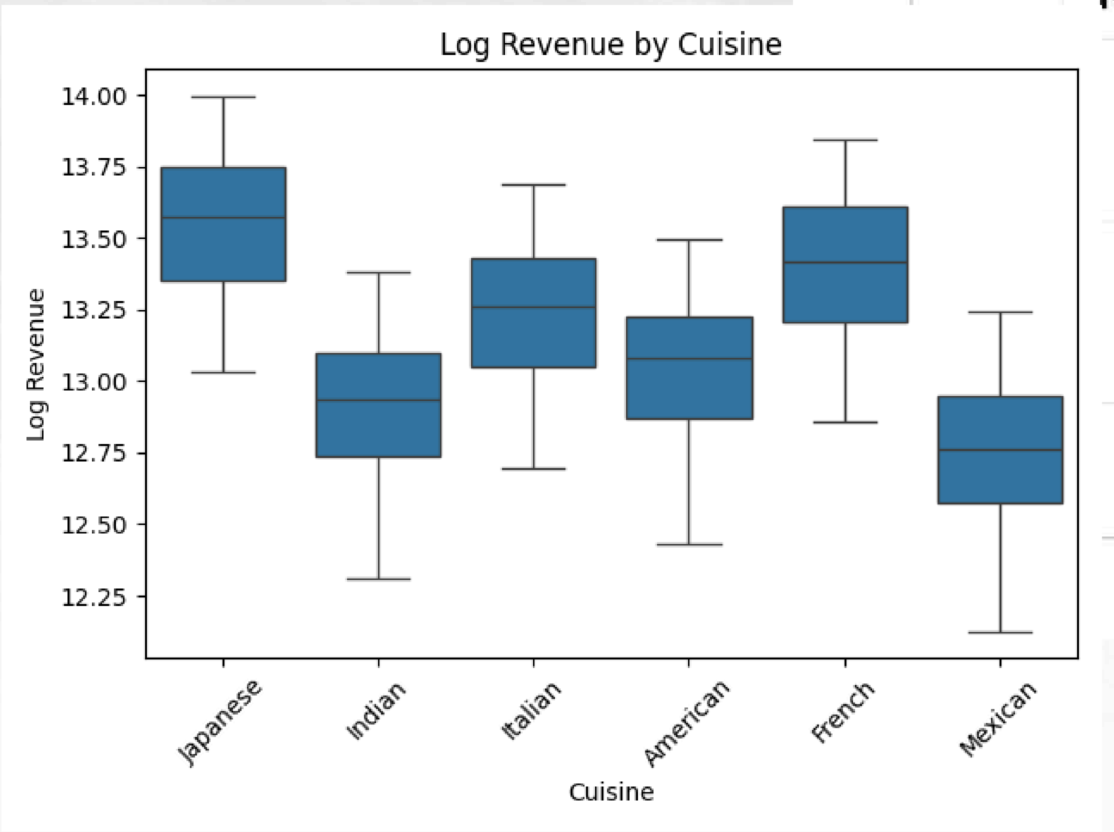
High Price → Japanese , French

Medium Price → Indian,Amaerican,Italian

Low Price → Mexican



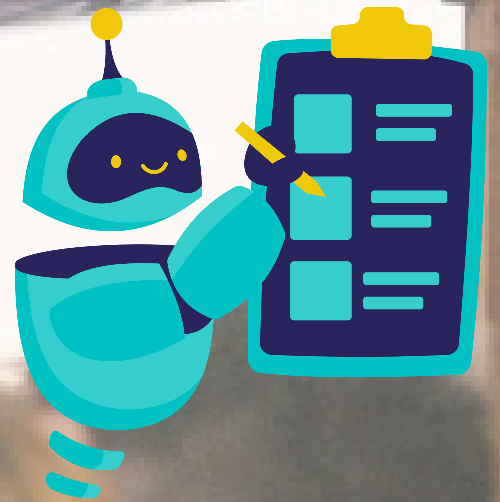


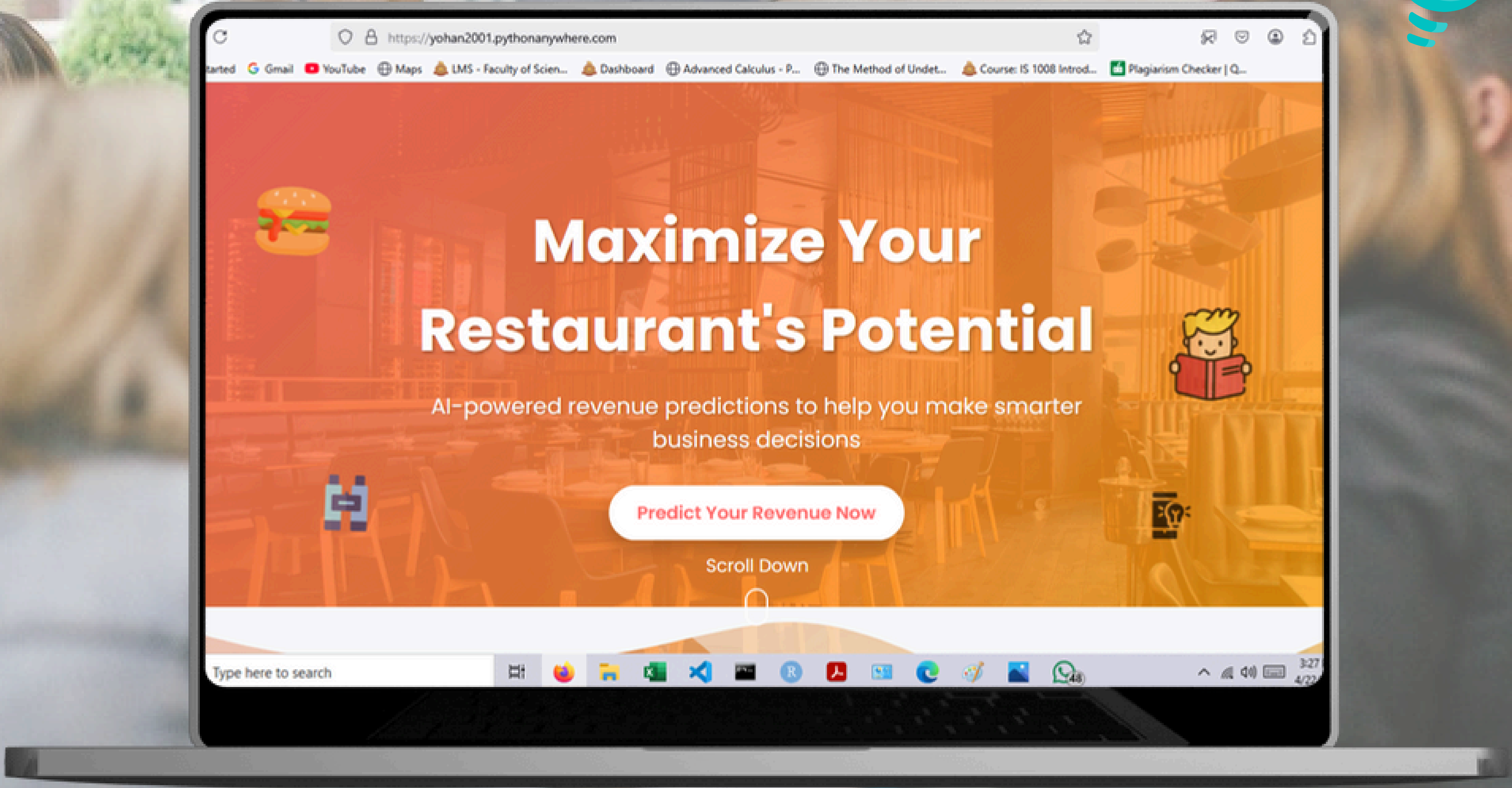


- Japanese and French restaurants typically have high meal prices and correlate with higher revenues

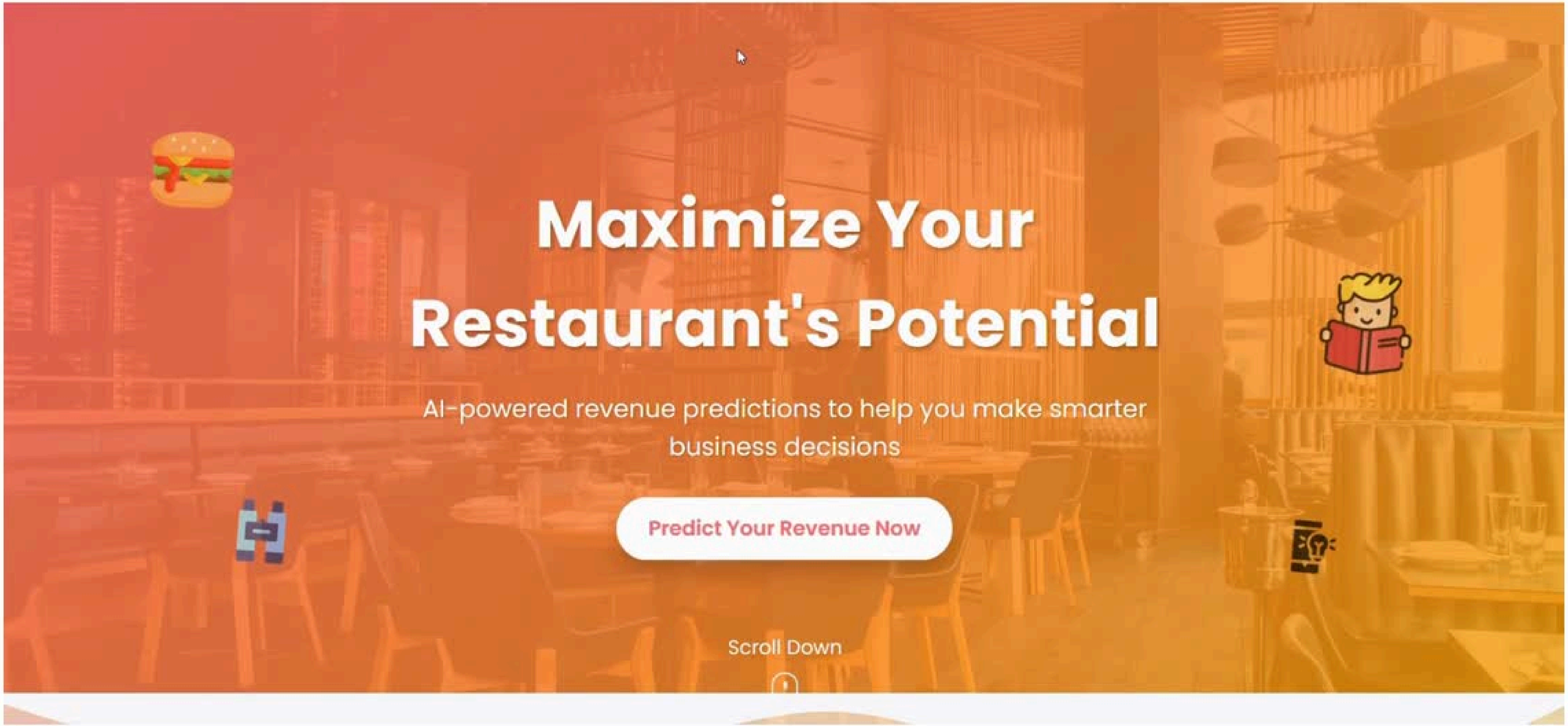
Smart Restaurant Advisor

Boost Your Revenue with AI





Try Now: https://yohan2001.pythonanywhere.com/


Maximize Your
Restaurant's Potential
AI-powered revenue predictions to help you make smarter business decisions
Predict Your Revenue Now
Scroll Down

Why Trust Gastronomix?

How accurately it predict restaurant revenue?
What is behind the Gastronomix ?

🎯 Model Selection Journey

To build an accurate and interpretable revenue prediction model, we started with some standard techniques and improved from there.

- Separate models were trained for:
  - Urban Restaurants
  - Non-Urban Restaurants

Step 1

Tried Multiple Linear Regression (MLR)

- Explored linear relationships in the data.
- Faced issues with multicollinearity due to strong correlations between Social Media Followers & Marketing Budget.
- Violated MLR assumptions of linear independence.

Step 2

Switched to ElasticNet Regression

- Combines Lasso & Ridge penalties to handle multicollinearity.
- Automatically selects important features by shrinking less useful ones to zero.
- Better fit for our dataset with many variables.

Step 3

Tried Ensemble Models

- We experimented with Random Forest and XGBoost.
- But despite being powerful, they did not outperform ElasticNet in accuracy or interpretability.

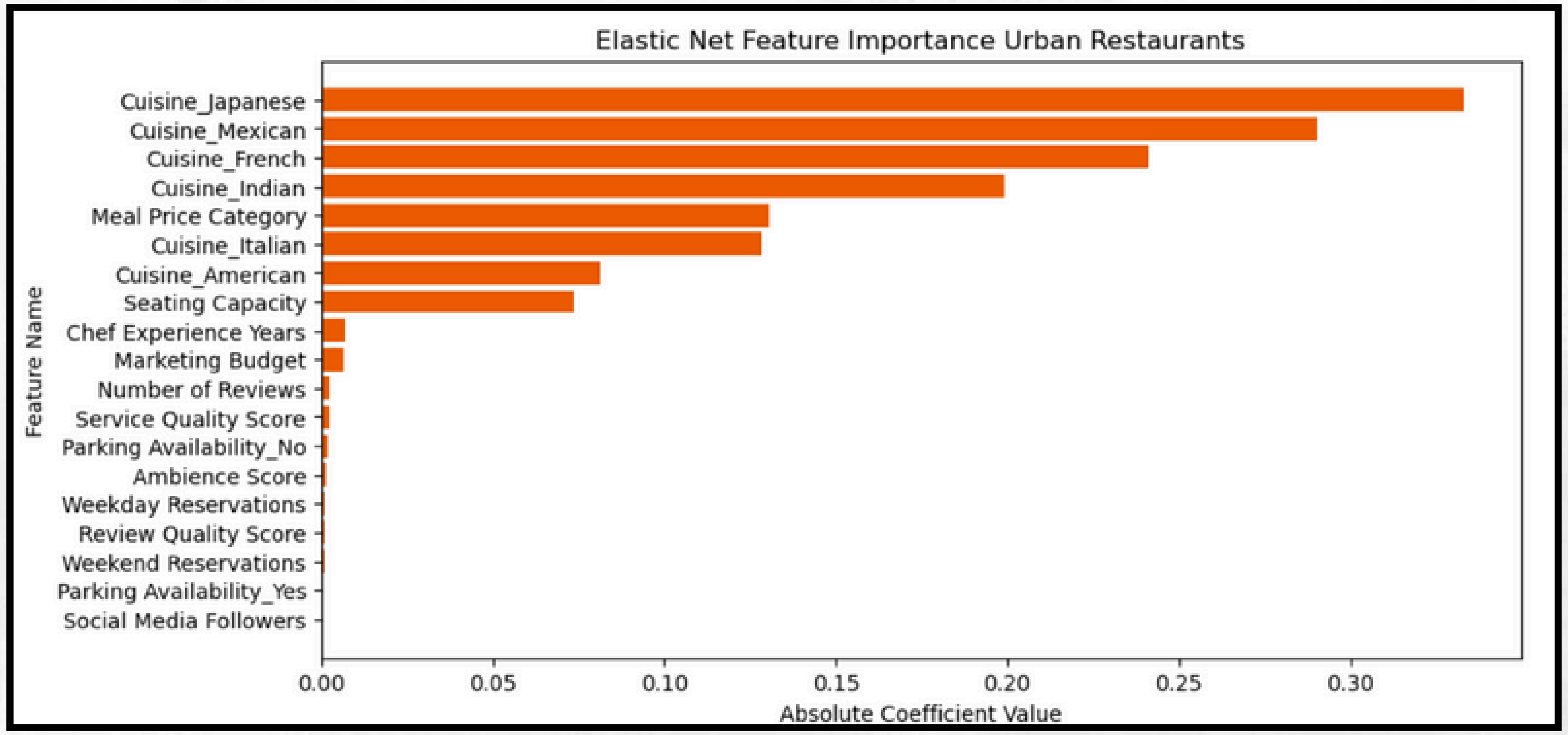
Gastronomix

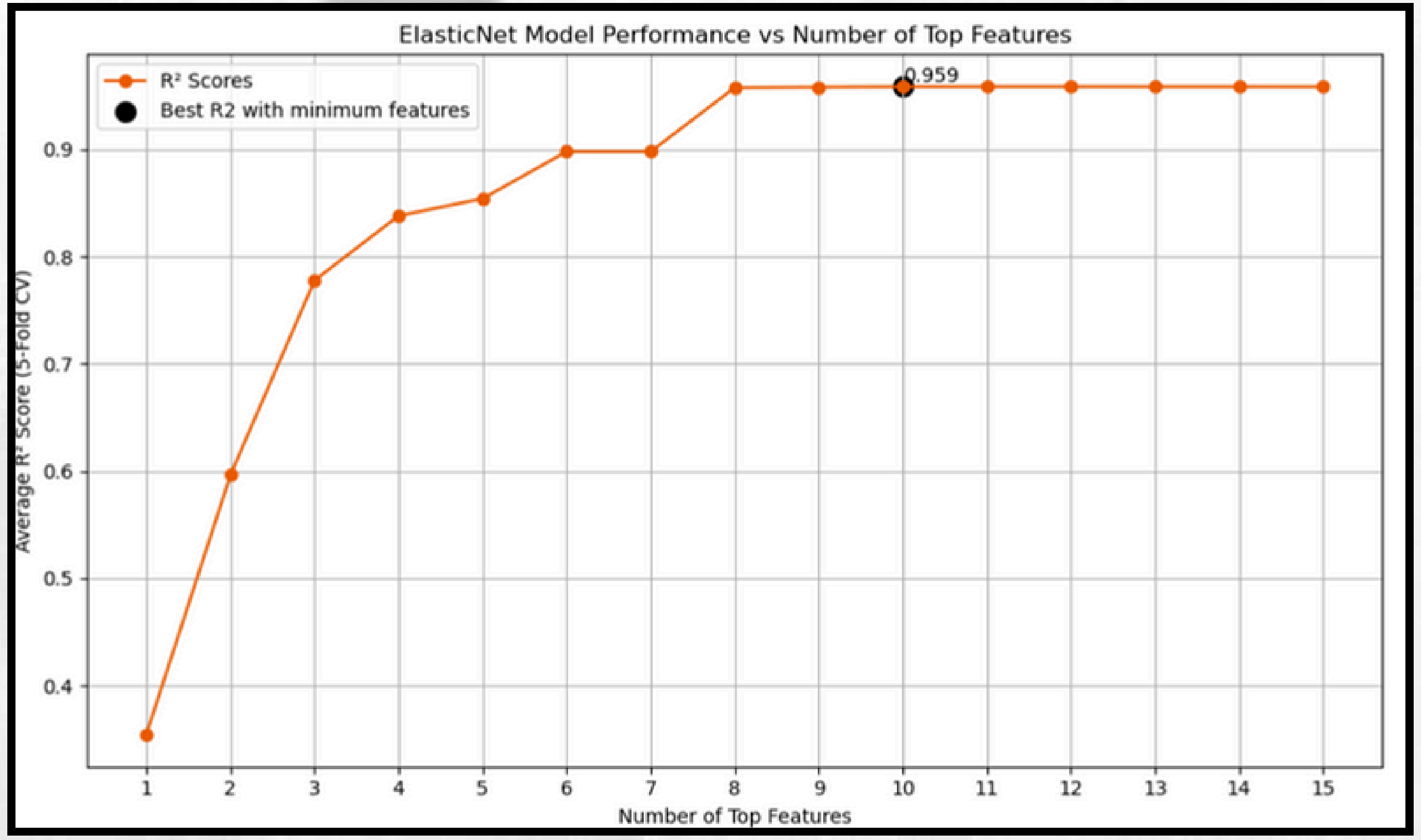
Urban Restaurants

	Train Set		Test Set	
	R2	MSE	R2	MSE
MLR	0.9591	0.0037	0.9568	0.0038
Elastic Net	0.9591	0.0037	0.9575	0.0038
Random Forest	0.9807	0.0017	0.9559	0.0039
XG Boost	0.9657	0.0031	0.9577	0.0037

Gastronomix

Urban Restaurants





Urban Restaurants top predictors

- Cuisine Type
- Meal Price category
- Seating capacity
- Chef Experience years
- Marketing Budget

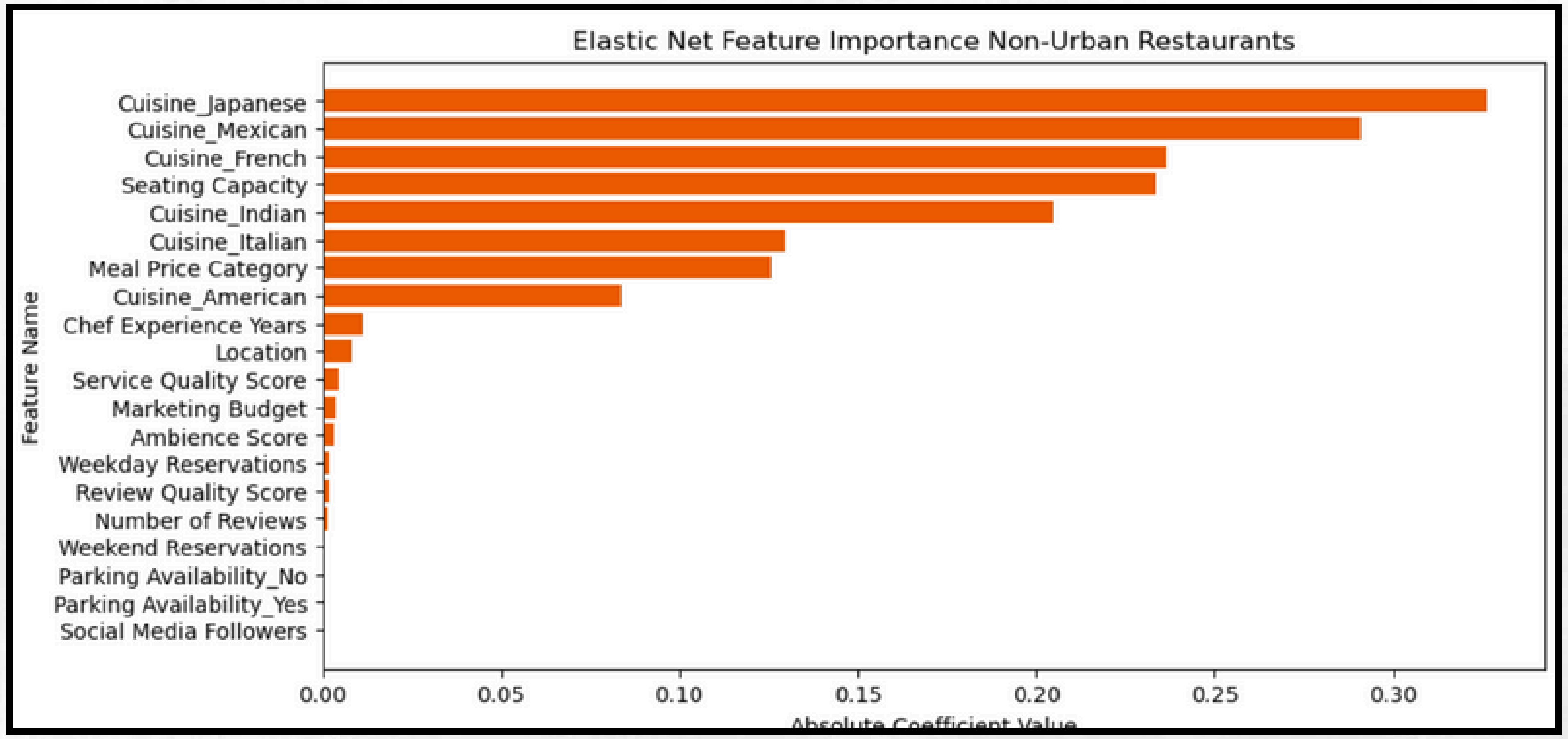
Next we refit the elastic net model with important variables

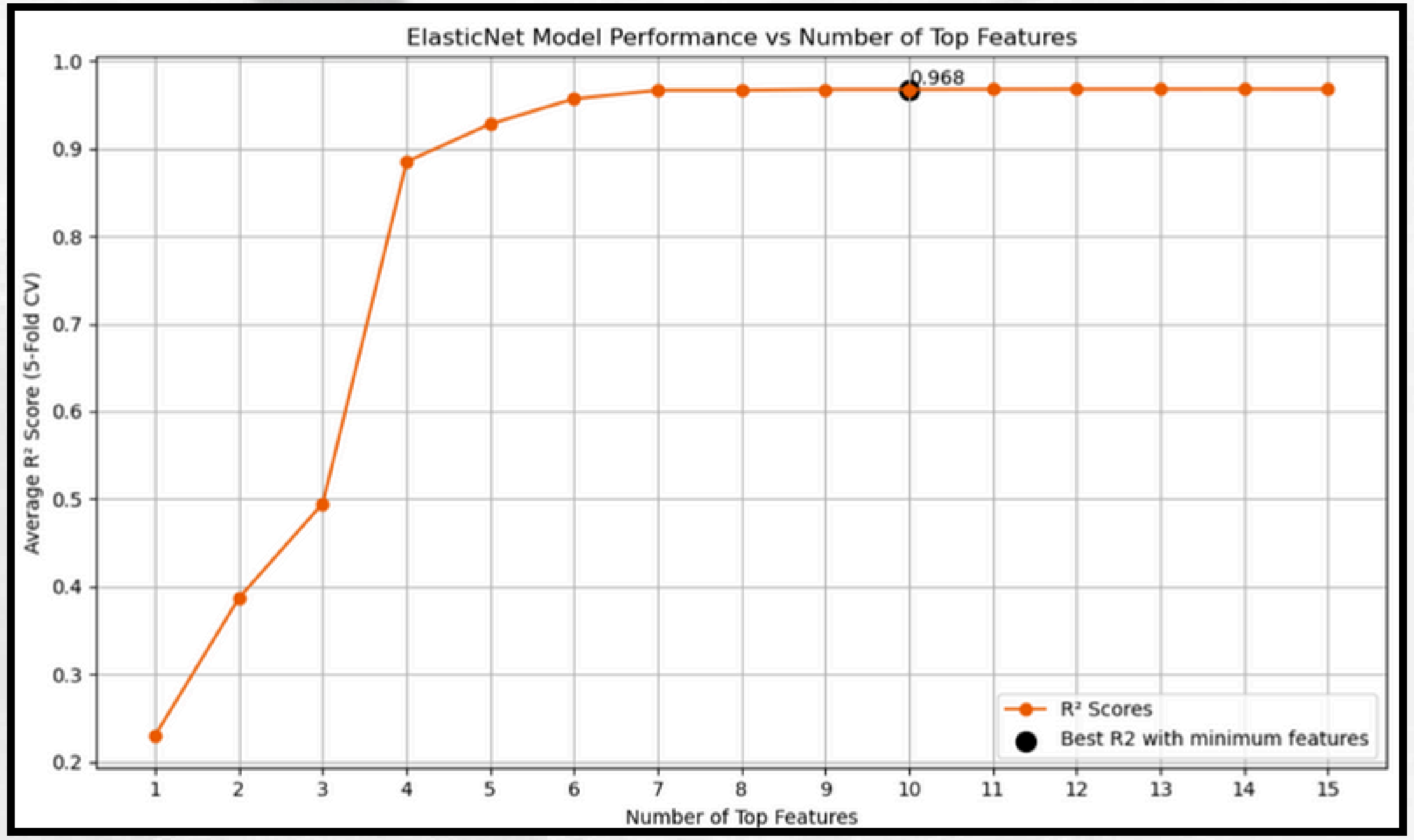
Non-Urban Restaurants

	Train Set		Test Set	
	R2	MSE	R2	MSE
MLR	0.9681	0.0042	0.9647	0.0045
Elastic Net	0.9682	0.0042	0.9647	0.0045
Random Forest	0.9852	0.0019	0.9677	0.0041
XG Boost	0.9797	0.0027	0.9687	0.0040

Gastronomix

Non-Urban Restaurants





Non-Urban Restaurants top predictors

- Cuisine Type
- Meal Price category
- Seating capacity
- Chef Experience years

Next we refit the elastic net model with important variables

- **Performance of reduced model is good.**
- **So in our website we used reduced models for predictions**

	Train Set		Test Set	
	R2	MSE	R2	MSE
Urban Restaurants	0.9589	0.0037	0.9572	0.0038
Non-urban Restaurants	0.9679	0.0043	0.9647	0.0045

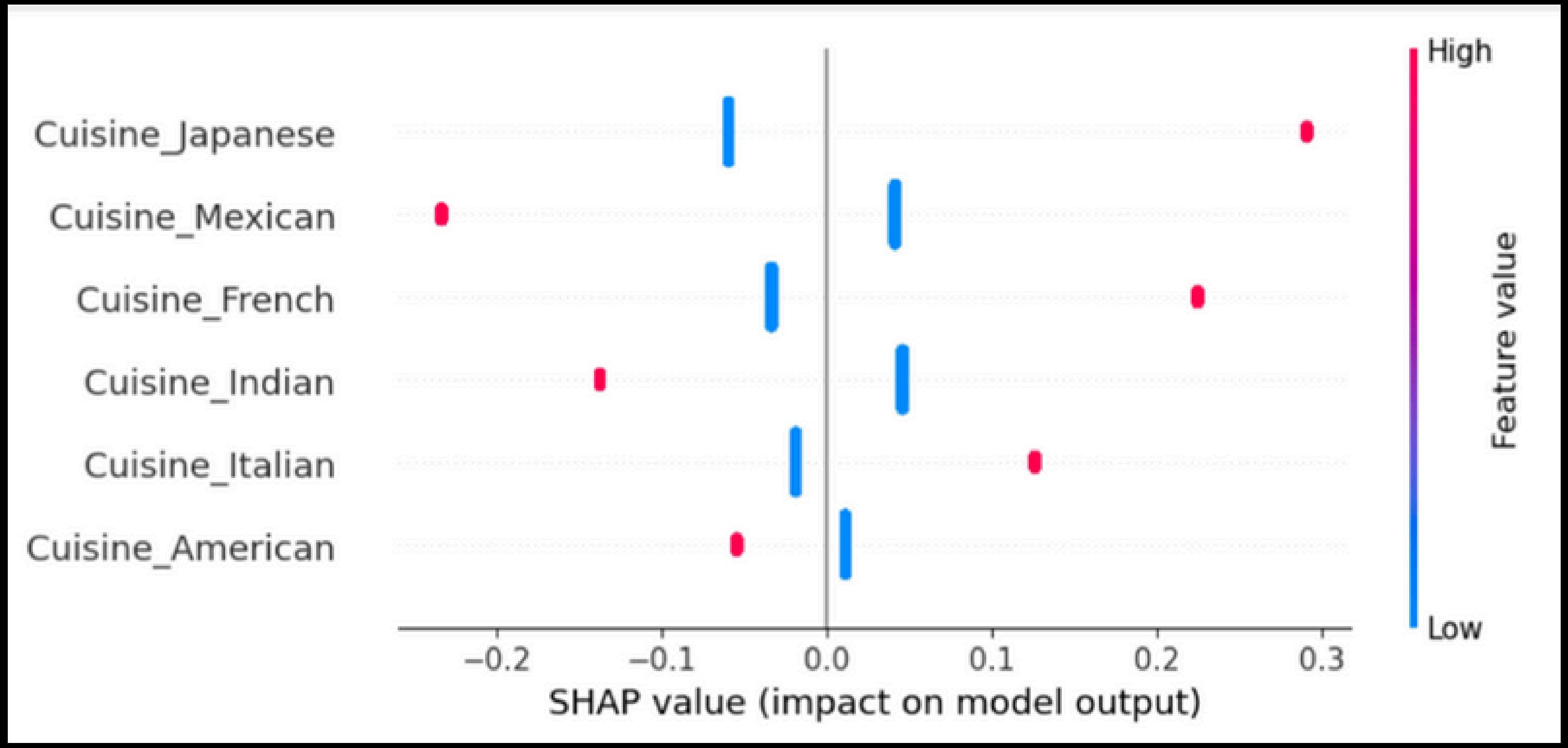
The model explains more than 95% of variance

That's Why you can trust Gastronomix

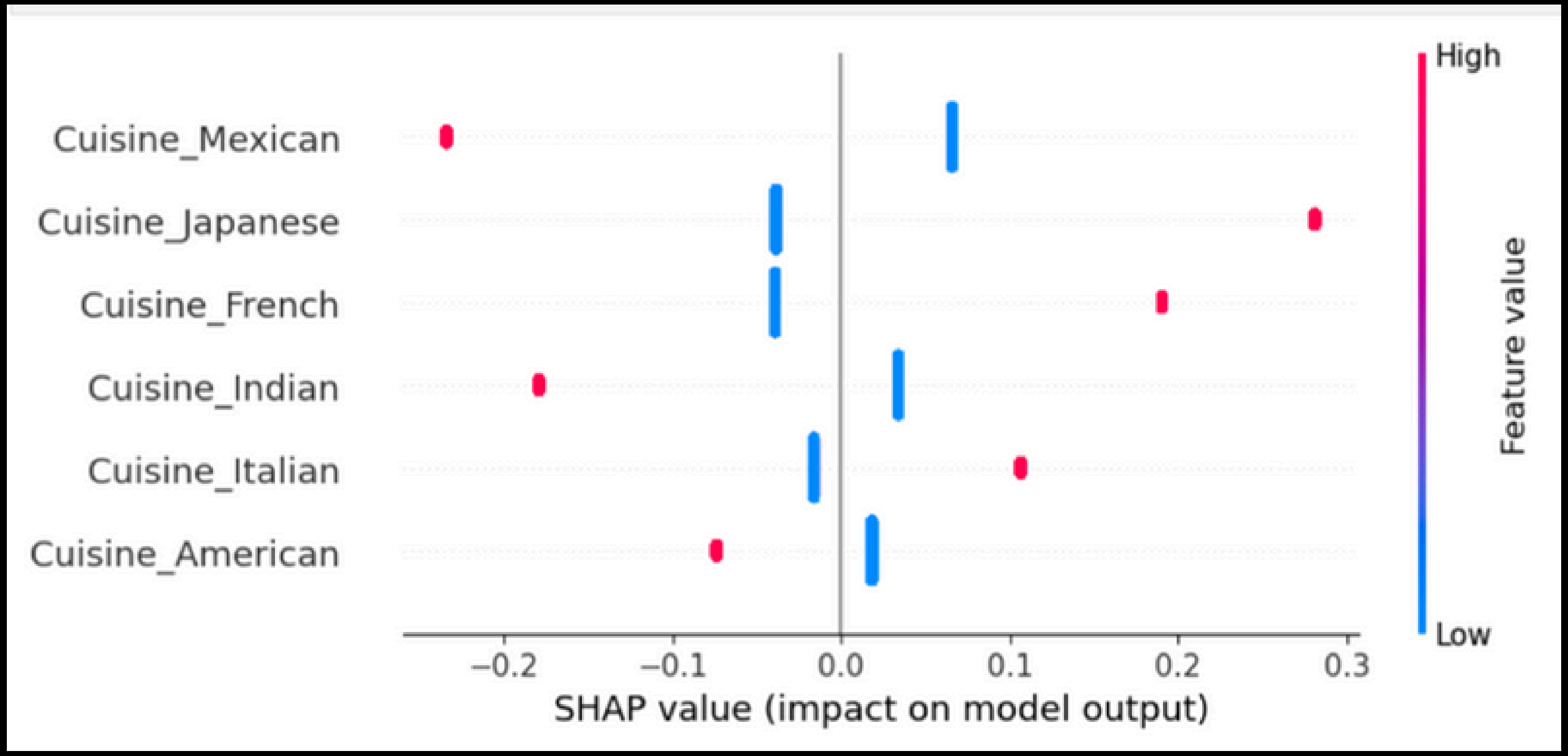
Gastronomix

📊 Cuisine Type: How It Affects Restaurant Revenue (Using SHAP)

Urban Retaurants



Non- Urban Retaurants



Gastronomix

Why Do Different Cuisines Impact Revenue Differently?

👉 Higher meal price → higher revenue impact
👉 Lower meal price → lower revenue impact

Urban Retaurants

Cuisine Type	High (%)	Medium (%)	Low (%)
American	-	100.00%	–
French	58.40%	41.60%	–
Indian	–	99.70%	0.30%
Italian	–	100.00%	–
Japanese	100.00%	–	–
Mexican	–	50.80%	49.20%

Non- Urban Retaurants

Cuisine Type	High (%)	Medium (%)	Low (%)
American	–	100.00%	–
French	53.20%	46.80%	–
Indian	–	100.00%	–
Italian	–	100.00%	–
Japanese	100.00%	–	–
Mexican	–	53.00%	47.00%

Cuisines that are usually priced higher also tend to contribute more positively to restaurant revenue.

📊 Real-world Data Confirms It

What Americans are willing to pay for "ethnic" restaurant food

Type of restaurant	Price rank in 1986	Price rank in 2014
French	1	1
Japanese	6	2
American	5	3
Continental	3	4
Italian	2	5
Spanish	7	6
Greek	12	7
Korean	13	8
Indian	8	9
Mexican	9	10
Southern	4	11
Chinese	11	12
Vietnamese	10	13
Thai	14	14

Gastronomix

Why Do Some Cuisines Cost More Than Others?

💎 High-Priced Cuisines

Japanese

➤ Uses rare, seasonal seafood
➤ Requires skilled chefs and artistic presentation
➤ Seen as premium and elegant.

French

➤ Rich in butter, wine, and fine ingredients
➤ Complex recipes and gourmet reputation
➤ Often associated with luxury dining

🌿 More Affordable Cuisines

Indian

➤ Uses grains, spices, and vegetables
➤ Typically served in casual, shared dining
➤ Known for flavorful and filling meals

Italian

➤ Pasta, pizza, and fresh ingredients
➤ Found in both upscale and casual settings
➤ Loved for its simplicity and availability

- https://www.ncesc.com/geographic-pedia/is-food-more-expensive-in-france/
- https://www.chefsresource.com/faq/is-food-in-japan-expensive/
- https://fooddrinktalk.com/what-the-average-food-cost-in-restaurants/

Gastronomix

Success Starts with Smart Choices

It's not just about cooking great food - it's about offering the right food in the right place.

Limitations & Future Direction

📌 Current Limitation

- In our model, location is categorized only as:
  - Rural
  - Downtown
  - Suburban
- It does not capture broader geographic variations like:
  - Different cities (e.g., Tokyo vs. Sydney)
  - Different countries (e.g., France vs. USA)

🔍 Why This Matters

- Food prices, cultural preferences, and cuisine perceptions can vary widely between countries and cities.

- For example, Japanese food may be affordable in Japan, but considered premium-priced in Australia.

📈 Future Research Suggestion

- Future studies should consider more detailed geographic features (like specific countries or cities)to understand regional pricing dynamics of different cuisine types more accurately.

Gastronomix

Thank You !!!

We appreciate your time and engagement throughout this presentation. Your interest continuously fuels our passion to deliver excellence every step forward.

+123-456-7890

gastronomix@gmail.com

gastronomix.pythonwhere.com

References

- https://youtu.be/X6ehVmIV2Go
- https://datascientistsdiary.com/shap-values-for-categorical-features/
- https://scikit-learn.org/stable/modules/partial_dependence.html
- https://www.ncesc.com/geographic-pedia/is-food-more-expensive-in-france/
- https://www.chefsresource.com/faq/is-food-in-japan-expensive/
- https://www.youtube.com/watch?v=bkXe-eDGjPY&list=PL495mke12zYBQjqBy-wUYh2LCAWSj4Ayn
- https://www.geeksforgeeks.org/flask-tutorial/
- https://codinginfinite.com/find-the-best-k-in-k-prototypes-clustering/
- https://chat.deepseek.com/
- Restaurant Industry Statistics
- https://www.canva.com/help/creating-and-editing-videos/

+123-456-7890

gastronomix@gmail.com

gastronomix.pythonwhere.com