

The Global Gourmet

Operational Excellence Case Study Competition

Ingenium — IIT Indore

Competition Overview

Global Gourmet is a multi-billion-dollar food-tech aggregator operating across hundreds of cities worldwide. Despite rapid expansion, the platform faces persistent inefficiencies in last-mile delivery, leading to inaccurate delivery estimates and customer dissatisfaction.

In this competition, participants act as the Lead Data Science Taskforce and are required to design an end-to-end, multi-stage predictive system integrating logistics, natural language processing, and economic optimization to improve platform performance and customer retention.

Phase I: Data Synthesis & Architectural Integrity

Participants are provided with a heterogeneous data ecosystem reflecting real-world enterprise complexity.

Available Data Sources:

- **Transactional Logs (SQL):** order_id, user_id, timestamp_placed, total_amount
- **Restaurant Metadata (JSON):** geographic coordinates, preparation times, cuisine categories
- **Courier Telemetry (Text/CSV):** high-frequency GPS pings from drivers

Task:

Design a robust ETL (Extract, Transform, Load) pipeline capable of:

- Resolving inconsistent timestamps across time zones
- Handling duplicate entries
- Mapping courier telemetry to order lifecycles with timestamp drift up to 180 seconds

The final output must be a unified **Golden Table** representing the complete lifecycle of each delivery.

Phase II: The Spatio-Temporal ETA Engine

Participants must develop a high-fidelity regression model to predict total delivery duration.

Feature Engineering Requirements:

- Kitchen Lag: Difference between order placement and pickup time
- Travel Decay: Impact of peak-hour congestion

External Context Integration:

- Weather data
- Local events or environmental disruptions

Success Metric:

- Minimize Mean Absolute Error (MAE)
- Special emphasis on reducing extreme lateness (outliers beyond 2 standard deviations)

Phase III: Linguistic Intelligence & Topic Mining

The platform receives a large volume of unstructured, text-based customer reviews daily.

Tasks:

- Build a sentiment classifier categorizing reviews as Satisfied, Neutral, or Dissatisfied
- Apply topic modeling techniques (LDA or BERTopic) on dissatisfied reviews to identify primary grievances:
 - Cold Food
 - Missing Items
 - Driver Professionalism
 - Late Delivery
- Quantify the causal relationship between ETA prediction error and the probability of receiving a 1-star review

Phase IV: Churn Propensity & Predictive Retention

Participants must build a binary classification model to identify users at risk of leaving the platform.

Feature Set Includes:

- Historical delivery lateness
- Frequency of negative review topics
- Changes in ordering behavior over the last 90 days

Primary Constraint:

The model should prioritize Recall, ensuring that at-risk users are not missed.

Phase V: Economic Optimization & Surge Logic

The final phase focuses on prescriptive analytics through dynamic pricing.

Objective:

- Design a dynamic surge pricing algorithm to balance demand and supply
- Maximize total platform revenue using price elasticity of demand
- Ensure churn propensity does not increase by more than 2% for any user segment

Dataset References

Participants may use the following datasets as references:

- Core Reviews Dataset: <https://www.yelp.com/dataset>
- Logistics Simulation: <https://www.nyc.gov/site/tlc/about/tlc-trip-record-data.page>
- Weather Data: <https://open-meteo.com/>

Participants are encouraged to use alternative datasets with greater complexity and demonstrate creativity in data sourcing, provided sufficient technical depth is maintained.

Evaluation Scorecard

Criterion	Weightage	Description
ETL Resilience	20%	Data cleaning robustness and entity resolution logic
Predictive Accuracy	30%	Performance on hidden test sets for ETA and churn models
Explainability	25%	Use of SHAP or LIME to justify predictions
Business Logic	25%	Mathematical rigor of surge pricing strategy

Rules are subject to change at the discretion of the organisers. Further instructions will be communicated to registered participants.