Stint Data Science Technical Task: Restaurant Demand Forecasting

Background

At Stint, accurate demand forecasting is at the core of our business. Our ability to predict when and where staffing is needed enables us to provide our partners with high-quality staffing recommendations, ensuring they have the right people at the right time. For senior roles, we need data scientists who can handle complex real-world data, implement sophisticated techniques, and maintain a sharp focus on business value.

The Challenge

You have been provided with 4 years of historical restaurant demand data from five different restaurants, along with external factors that influence demand. Each restaurant has unique characteristics, customer patterns, and business models. Your task is to build an advanced forecasting system that not only predicts demand accurately but also accounts for the asymmetric costs of forecasting errors - understaffing is far more costly than overstaffing.

Time Expected: 2 hours

Dataset Overview

The dataset (`restaurant_demand_*.csv`) contains:

Core Demand Signals (30-minute intervals over 4 years):

- Customer counts per period
- Sales by menu category (starter, main_meal, soft_drink, complex_drink, dessert, child_meal)
- Total sales revenue

External Factors:

- Weather data: temperature (Celsius), precipitation (mm)
- Economic indicator (normalised economic conditions)
- Competitor promotions (0-1 scale indicating promotional intensity)
- Social media trends (viral event impact scores)
- Local events (impact scores for concerts, sports, etc.)

Business Context:

- Restaurant types: casual bistro, seafood, family restaurant, fine dining, fast casual
- Operational constraints: capacity limits that change over time
- Reputation scores (1-5 scale, evolving over time)
- Temporal features: timestamp, day_of_week, hour, minute

Part 1: Business-Driven Analysis *(30 minutes)*

Perform an EDA on the data set - things you might consider:

- Identify the primary drivers of demand for each restaurant type
- Quantify how external factors impact demand (e.g., "Temperature above 25°C increases demand by X%")
- Find actionable insights (e.g., "Family restaurants show 40% higher demand on weekends")
- Derive and characterize peak demand periods from the data
- Identify the hardest-to-predict periods and explain why
- Analyse how forecast difficulty varies by restaurant type
- Quantify the business impact of demand volatility during peak periods

Deliverable: 3-4 visualisations that reveal insights leading to clear next steps

Part 2: Forecasting Solution *(60 minutes)*

Implement an advanced model with custom asymmetric loss function - things you might consider:

- Penalises understaffing more heavily than overstaffing
- Define and justify your penalty ratio (e.g., 2:1, 3:1)
- Can be adjusted based on demand levels

Choose ONE modeling approach:

- **Option A: Machine Learning with Custom Objective** - XGBoost/LightGBM with your custom loss function, feature engineering using all external factors, SHAP analysis for interpretability
- **Option B: Deep Learning with Custom Loss** - Neural network (LSTM/GRU) with asymmetric loss, attention mechanisms for external factors, uncertainty quantification
- **Option C: Probabilistic Approach** - Models that output full distributions, quantile regression or probabilistic neural networks, showing how different percentiles map to staffing strategies

Required for all models:

- Forecast horizons: Next 24 hours (48 periods) and next 7 days
- Uncertainty quantification: Provide 80% and 95% prediction intervals
- Handle capacity constraints and missing data appropriately

Deliverables: Build a forecast model using one of the options above using asymmetric loss

Part 3: Model Evaluation & Peak Performance Analysis *(30 minutes)*

Evaluate your model using multiple metrics (MAPE, RMSE, and your custom business metric). Show how the asymmetric loss affects predictions - things you might consider:

Peak Period Performance Deep Dive

- Define peak periods from the data (e.g., top 20% demand periods)
- Separate accuracy metrics for peak vs off-peak periods
- Analyse error distribution during different peak types (lunch rush, dinner rush, weekends, event-driven)
- Evaluate peak detection accuracy and peak magnitude prediction accuracy

Business Impact of Asymmetric Loss

- Quantify the reduction in understaffing incidents
- Calculate the cost tradeoff of your approach

Model Confidence & Limitations

- When is the model most/least reliable?
- How do external factors affect prediction uncertainty?
- What scenarios would cause the model to fail?

Deliverables

1. **Code**: Well-documented Jupyter notebook(s) containing your analysis, modeling, and evaluation
2. **Walkthrough**: however you see fit to present your approach and findings in the next interview

---

We're not looking for the most complex solution, but rather the most valuable one for our customers.

Good luck!