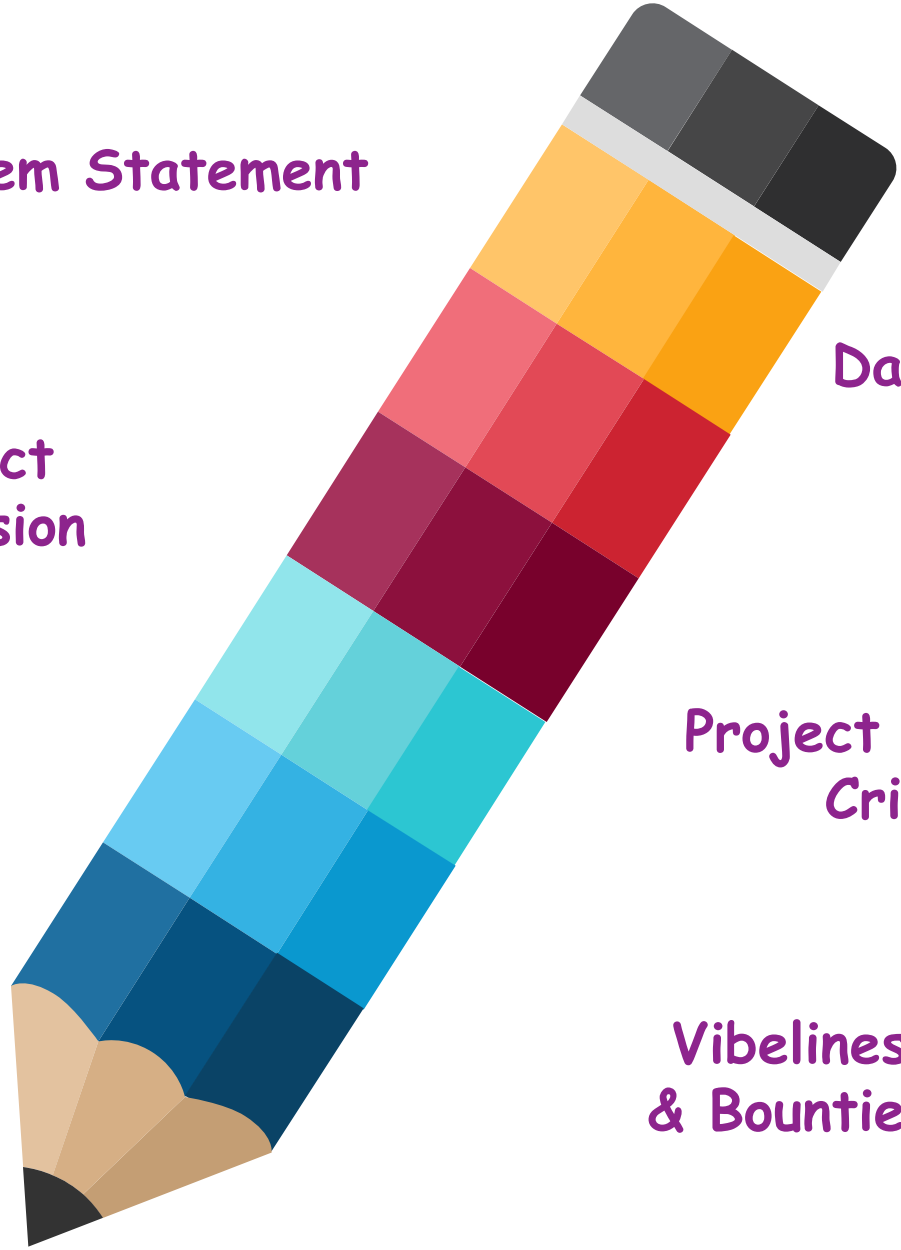
Launching a New HORECA Venture

Problem Statement

Artifact Submission

Recommendation



Data Description

Project Evaluation Criteria

Vibelines & Bounties

Problem Statement

A new Food Entrepreneur, let's call them "Chef Innovator," is planning to launch a new venture in the highly competitive HORECA space. They have access to a rich Zomato dataset (which you are analyzing) and need to define their strategy based on data-driven insights to maximize their chances of success and profitability.
Your role, as a Data Scientist, is to provide the critical market intelligence by analyzing the existing landscape. Chef Innovator is primarily concerned with three high-level decisions:

- Category (Cuisine)
- Market (City/Location)
- Channel (Dining vs. Delivery).

The Entrepreneur's Key Questions (Challenge Objectives)

Your analysis should address the following three strategic questions by using the provided Zomato dataset:

1. Category: Identifying the Next Winning Cuisine

- **Goal:** Determine the optimal cuisine category to enter.
- **Analysis Required:** Analyze the distribution of cuisines and their performance metrics (like Average_Rating, Restaurant_Popularity, and Total_Votes).
- **Key Insight:** Which **cuisine category** shows a high average rating and strong demand, but perhaps has a lower saturation (fewer restaurants/low average price), indicating a potential gap in the market for a premium/high-quality offering?

2. Market: Pinpointing the Best City/Location

- **Goal:** Determine which metropolitan market offers the best growth opportunity.
- **Analysis Required:** Compare performance metrics across different City and Place_Name columns.
- **Key Insight:** Which **City or area** exhibits the highest potential profitability, possibly indicated by a favorable ratio of Avg_Price_City to Avg_Rating_City, or a low count of Is_Highly_Rated restaurants despite high demand, suggesting an underserved market?

3. Channel: Defining the Service Model (Delivery Focus vs. Dining Focus)

- **Goal:** Decide whether the new venture should prioritize a fine-dining experience or an optimized delivery model.
- **Analysis Required:** Compare the spread and correlation between Dining_Rating vs. Delivery_Rating and Dining_Votes vs. Delivery_Votes.
- **Key Insight:** Are there specific categories or markets where **delivery performance** significantly exceeds dining performance (or vice versa)? This will suggest the most profitable **channel strategy**—a streamlined kitchen for delivery or a premium physical location for dining.

Data Description

Column Name	Description
Restaurant_Name	Name of the restaurant listed on Zomato.
Dining_Rating	User rating for the dine-in experience (0.0 to 5.0).
Delivery_Rating	User rating for the delivery experience (0.0 to 5.0).
Dining_Votes	Number of votes received for dine-in service.
Delivery_Votes	Number of votes received for delivery service.
Cuisine	Type of cuisine served (e.g., Fast Food, Chinese).
Place_Name	Local area or neighborhood of the restaurant.
City	City in which the restaurant is located.
Item_Name	Name of the menu item listed.
Best_Seller	Bestseller status (e.g., BESTSELLER, MUST TRY, NONE).
Votes	Combined total votes received.
Prices	Price of the menu item in INR.
Average_Rating	Mean rating calculated from available sources.
Total_Votes	Sum of all types of votes.
Price_per_Vote	Ratio of price to total votes (used to evaluate value for money).
Log_Price	Log-transformed price to reduce skewness in analysis.
Is_Bestseller	Binary flag indicating if item is marked as a bestseller.
Restaurant_Popularity	Number of items listed by the restaurant in the dataset.
Avg_Rating_Restaurant	Average rating of all items from the same restaurant.
Avg_Price_Restaurant	Average price of all items from the same restaurant.
Avg_Rating_Cuisine	Average rating across all restaurants serving the same cuisine.
Avg_Price_Cuisine	Average price across all restaurants serving the same cuisine.
Avg_Rating_City	Average rating across all restaurants in the same city.
Avg_Price_City	Average price of menu items in the same city.
Is_Highly_Rated	Binary flag for ratings ≥ 4.0.
Is_Expensive	Binary flag for prices above city's average.

Artifact Submission

Your submission must include the following five artifacts, all packaged within a single GitHub repository.

1. Jupyter Notebook (.ipynb) This is the core of your submission. Your Jupyter Notebook should be a complete, well-documented narrative of your data analysis journey. It must include:

- **Detailed Explanations:** Use Markdown cells to explain your thought process, the questions you are trying to answer, and the insights you've uncovered.
- **Clean Code:** The code should be well-structured, easy to read, and free of unnecessary clutter.
- **Comprehensive Comments:** Use comments to explain complex logic and the purpose of different code blocks.
- **Key Visualizations:** All visualizations should be clear, properly labeled, and directly support your findings.

2. Presentation (.pptx or .pdf)

Create a compelling presentation that summarizes your team's analysis and key findings. This presentation should serve as your final pitch. It must include:

- **Executive Summary:** A concise overview of your findings.
- **Key Insights:** The most important takeaways from your analysis.
- **Data-Driven Recommendations:** Actionable steps that can be taken based on your insights.
- **Supporting Visualizations:** A selection of your best visualizations to illustrate your points.

3. README File (.md)

The README file is the first thing we'll look at. It should serve as a quick guide to your project and provide essential details. It must include:

- **Project Title :**
- **Brief Problem Statement:** A summary of the project and your approach.
- **Summary of Findings:** A bullet-point summary of your most significant insights.

4. Attached Dataset

Please include the original dataset (.csv or other format) within your repository. This ensures the judges can reproduce your analysis without any issues.

5. GitHub Repository

Your final submission will be your GitHub repository. The repository name **must follow this exact format:**
HORECA_ProjectName_TMP

Challenge Evaluation Criteria

Criteria Name	Criteria weight
Data Understanding and Initial Insight	20%
Visualization and Pattern Identification	20%
Statistical Analysis and Hypothesis Generation	20%
Business Recommendation	30%
Coding guidelines and standards	10%

Project Outcomes - Strategic Area 1: Category Selection (Cuisine)

Task	Required Metrics & Analysis	Target Insight
Performance vs. Popularity	Create a matrix or visualization that compares: • **Demand (Volume):** Votes, Delivery_Votes, Total_Votes • **Quality (Rating):** Average_Rating, Dining_Rating, Delivery_Rating	Identify top-performing (high rating/high demand) cuisines.
Market Saturation	Group the data by Cuisine and calculate the **Count of Restaurants**. Compare this count against the average Log_Price and Avg_Rating_Cuisine.	Locate high-demand, high-rating cuisines that currently have a *lower number* of established competitors, suggesting a blue ocean opportunity.
Price Point Analysis	Analyze the spread and mean of Price_per_Vote and Log_Price within the top 5 cuisines.	Define the optimal pricing strategy (premium, mid-range, budget) for the chosen cuisine.

Project Outcomes - Strategic Area 2: Market Selection (City & Place)

Task	Required Metrics & Analysis	Target Insight
City-Level Viability	Aggregate data by City. Calculate the overall average Average_Rating and median Total_Votes for each city. Filter to include only cities above a minimum vote threshold.	Identify the city with the best *overall* market health and consumer engagement.
Micro-Market Analysis	Within the top-performing city, analyze the Place_Name distribution. Compare the density of Restaurant_Name (saturation) against the mean Restaurant_Popularity.	Pinpoint the specific **neighborhood/place** that offers high consumer interest (high popularity) relative to the existing restaurant density.
Risk Assessment	Use the Is_Highly_Rated (Count) and Is_Expensive (Count) columns to characterize the competition in the top locations.	Understand the intensity of high-quality/high-price competition in the target market.

Project Outcomes - Strategic Area 3: Channel Selection (Dining vs. Delivery)

Task	Required Metrics & Analysis	Target Insight
Channel Disparity	Calculate the correlation coefficient between (Dining_Rating, Delivery_Rating) and (Dining_Votes, Delivery_Votes).	Determine if consumers judge dining and delivery services similarly or if one channel is consistently underperforming/overperforming.
Optimal Channel Focus	For the recommended **Cuisine** and **Market**, plot Dining_Rating vs. Delivery_Rating for all restaurants. Segment the scatter plot based on Is_Bestseller status.	Recommend a clear primary channel: **Delivery-First:** If Delivery Ratings/Votes dominate. **Dining-First:** If Dining Ratings/Votes dominate.

Lets Go