Ukrainian Catholic University

Faculty of Applied Sciences

Business Analytics Program

Analyzing Factors Affecting Cafe Check Amount

Econometrics final report

Authors:
Sofiia Bazylevych
Solomiia Shopiak
Tetiana Shvets

4 May 2025

1 Topic and motivation

The project focuses on analyzing the factors that influence the total check amount in a cafe setting. Our primary goal is to identify set of recommendations/propositions/offers that the cafe might offer for clients to increase average customer spending.

This research is specifically related to and is based on information from a cafe in Lviv. This research is valuable for cafe owners and managers. Understanding the buying behaviour of consumers in this establishment will allow them to make data-driven decisions about the menu offers for customers. In this way, businesses can improve their revenue streams and customer satisfaction.

Additionally, the cafe will be able to improve its procurement plans based on the conclusions about recommended menu offerings and descriptive data analysis, ensuring that products are available at peak times.

2 The aim and the tasks

Aim: identify set of promotions that the cafe might offer for clients to increase average customer spending

The primary goal of our research is to analyze the factors that influence the total check amount in a cafe. By examining the relationship between variables such as dish categories, time of day, day of the week, the objective is to uncover insights about customer spending behavior and spending patterns in the cafe.

Our main research questions:

- What is the impact of time of day and day of week or month on spending patterns? Do these factors influence popularity of the category?
- What categories of dishes are likely to be ordered together?

Objectives and Problem Description.
Our research aims to address key issues in cafe revenue management by:

- Using exploratory data analysis to identify how different factors (product categories, visit duration, and time of visit) influence customer spending, allowing for improved menu design and pricing strategies.
- Using econometric modeling techniques to derive valuable insights such as promotion, that enhance inventory management, minimize waste, and optimize procurement planning.

3 Data Analysis

Our dataset consists of cafe transaction records throughout the year, which is cross-sectional data with the following key columns:

- **Receipt number**: Unique identifier.
- **Opened, Closed**: Date and time of start and finish of customer service.
- **Guests amount**: Number of visitors per check.
- **Product name**: Dishes that the order consists of (bag of words - needs preprocessing).
- **Product price**: Price for one unit of a specific dish in UAH.
- **Product amount**: Amount of units of a specific dish in a certain order.
- **Sum of products**: Total price for units of a specific dish in a certain order in UAH.
- **Check amount**: Price for the order in UAH.
- **Profit**: Revenue per order without product expenditures.

3.1 Data preprocessing

The product categories in our analysis are based on the restaurant menu. We combined categories that were too detailed into larger categories (for example, tea and coffee into one category - hot drinks). So, we work with the following categories from the menu: Alcohol, Cold drinks, Complements, Desserts, Grill, Hot drinks, Main courses, Salads, Side dishes, Soups, Starters. The presence of each category of dishes in the order is presented as 0 or 1.

Our dataset contains both in-cafe and delivery orders data. However, we consider only the in-cafe data, since sample size of delivery data is not big enough.

3.2 EDA & Visualizations

All figures are attached in the Appendix.

From figure 1, as expected, we can conclude that among longer visits there are more of big-sized checks. However, the mean value of check size doesn't vary a lot among visits of different duration.

After analysis we discovered that this restaurant is practically not visited by large groups of people. The most common number of guests is 1-2. From figure 2 we can observe that larger groups tend to spend more.

Figure 3 shows that on weekends guests of this restaurant tend to spend more.

Figures 4 depict the difference in customers' spending and visiting depending on time of day. We observe that visits during noon are the most frequent and that on average the check amount is higher compared to others. Such features can be caused by the cafe's location as it is situated relatively close to city center and is right between two universities (We assume that students often come here for lunch break).

Figure 5 visualizes the total number of items sold per category across morning, lunch, and evening periods. The most notable observation is the consistent dominance of main courses, which are most frequently ordered during lunch (1539) and evening (1170). Other categories such as hot drinks, soup, and starters also show strong performance during morning and lunch, suggesting they are popular choices to start the day or accompany lunch.

Figure 6 shows weekly demand patterns for each food category. Main courses again dominate consistently, peaking on Sundays (664), followed by Saturday (595) and Friday (573), confirming strong demand during weekends. Other notable patterns include: Cold drinks and alcohol peak during weekends, especially Sunday. Soup and starters gradually increase toward the weekend.

Figure 7 breaks down the average number of items ordered per hour, segmented by day of the week. Each subplot gives a detailed view of category trends throughout the day: Main course dominates every day, especially during lunch (12–15) and dinner (18–21). Hot drinks spike in the morning, confirming breakfast-time interest. Cold drinks, desserts, and alcohol are more prominent in the evening, especially on Friday–Sunday. Soup and starters consistently appear mid-day and early evening.

Figure 8 illustrates the total monthly profit generated by orders that contain each food category throughout the year. Main courses consistently lead in revenue contribution, with clear peaks in February, August, and December. Other strong contributors include cold drinks, soups, and alcohol, which show seasonal fluctuations but remain profitable across months. These trends can support procurement planning and reveal high-margin categories worth promoting during specific periods. Figure 9 shows the monthly count of orders that included each category. While main courses again dominate in frequency, cold drinks, soups, hot drinks, and starters also show strong customer demand, especially in warmer months.

Key observation: From Fig. 5, 6, 7, 8, 9 we see that the most popular category is main_course. Consequently, main_course seems to lead to higher customer spending. So an **additional research question** arises:

How do the time of day (morning, lunch, evening) and the type of dish interact to influence the main dish purchase?

4 Methodology

4.1 Binary Logistic Regression

Based on EDA observations we would like to estimate what factors influence the probability of ordering main_course.
To do that we apply a **binary logistic regression model**:

$$P(\text{Main Course}_i = 1) = \frac{1}{1 + \exp(-(\beta_0 + \beta_1 X_{1i} + \beta_2 X_{2i} + \dots))}$$

Dependent variable:

- `main_course` (binary): Indicates whether an order includes a main course item. In terms of logistic regression shows the likelihood that a main course is included, given the explanatory variables.

Independent variables:

- Product Categories (different food types, e.g., 'alcohol", 'cold_drink", 'grill", "salad", "dessert", "beverages").
- Time of Visit (morning, lunch, evening).
- Visit Duration (time between order opening and closing).
- Day of the Week (weekday vs. weekend).
- Number of Guests (how group size affects spending).

4.2 Interaction Terms

Interaction terms are included to explore whether the relationship between the predictors and the probability of ordering a main course varies across different conditions. These are specified as multiplicative terms between two dummy or continuous variables.

Key interactions included in the model:

- **Grill × Salad**: Indicates whether customers who choose grilled items are more likely to also prefer lighter sides such as salads.
- **Alcohol × Starter/Side Dish**: Tests whether alcohol consumption is associated with ordering small plates like starters or sides.
- **Guest Count × Starter/Side Dish**: Checks whether larger groups are more likely to order a variety of small dishes.
- **Weekend × Starter/Soup**: Explores if weekend visits are associated with higher likelihood of ordering starters or soups.
- **Morning × Soup**: Investigates whether soups are more common in the morning compared to other times.

The inclusion of these terms allows us to evaluate not just main effects, but context-dependent preferences that may guide menu design or promotional strategies.

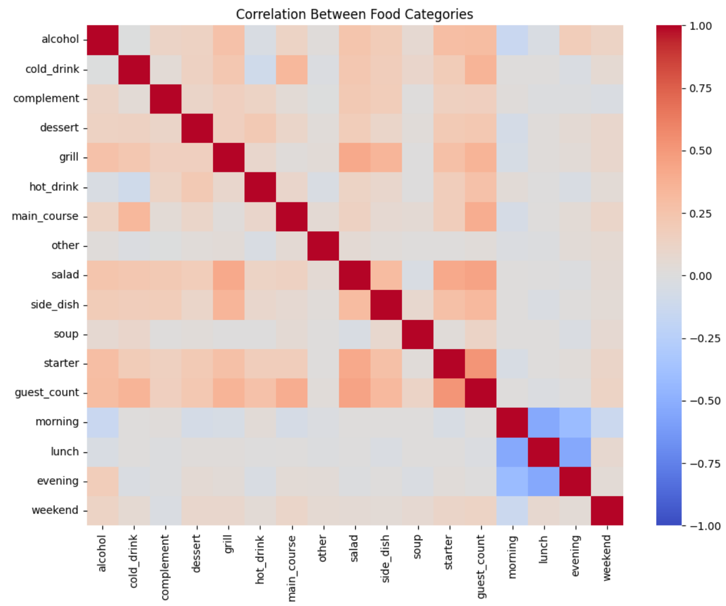
4.3 Model Selection: Backward Elimination

We apply a backward stepwise selection procedure based on p-values to refine the model. Starting with a full model including all main effects and interaction terms, we iteratively remove predictors with the highest p-values above a threshold (0.05) until only statistically significant predictors remain.

4.4 Logistic Model Assumption Testing and Model Diagnostics

Logistic regression, unlike general linear models, does not require several core assumptions of ordinary least squares (OLS) regression such as homoscedasticity, normality of residuals, linear relationship between dependent and independent variables. However, logistic regression still relies on a set of its own assumptions, which are important to validate before interpreting the model results. [1]

- **Sample size**: Logistic regression performs best with a large sample size. A common rule of thumb is to have at least 10 events per predictor for the least frequent outcome. Given the proportion of main course orders and number of predictors in our model, the required minimum sample size would be 245. Our sample size is 4057, meaning it is sufficient to provide stable estimates.
- **Multicollinearity**: Although logistic regression is robust to some multicollinearity, we evaluate variance inflation factors (VIFs) and examine correlation matrices between predictors to ensure that no problematic multicollinearity exists.



VIF	feature
10.850455	const
2.493291	grill
1.651754	salad
1.838080	alcohol
4.796567	starter
4.109502	side_dish
1.138406	cold_drink
3.270495	guest_count
2.007806	weekend
3.631646	soup
2.328096	lunch
2.356279	evening
1.824166	duration
2.858326	grill_salad
2.831708	alcohol_starter
2.550969	alcohol_side_dish
9.088245	guest_count_starter
8.186055	guest_count_side_dish
2.140448	weekend_starter
2.160653	weekend_soup
5.110578	duration_starter
5.011732	duration_side_dish
3.347375	lunch_soup
2.667524	evening_soup
2.765569	grill_side_dish
1.075438	dessert_hot_drink

Figure 10: Correlation matrix

Figure 11: Variance inflation factor values

After analyzing Figure 10 we can conclude that the matrix shows mostly light blue (low correlation values), meaning that most independent variables have weak linear relationships with each other. This is a good sign for avoiding multicollinearity. We can observe some slight red squares indicating weak positive correlation between grill & side dish or salad, cold_drink & main_course, which is intuitively expected and understandable. Moreover we can notice slightly darker blue indicating weak negative correlation between morning & lunch, lunch & evening. But one of these dummy variables won't be included into model and the correlation is relatively not high. Meaning that even though we have slightly big correlation values they **are not** high enough to claim that we have multicollinearity in our case.

To further confirm our findings we also calculate **VIF**s:

From Figure 11 we can see that none of the variables have VIF value bigger than 10 (which is considered to be a value that indicates that multicollinearity is present). Although there are some variables that might be a concern (like guest_count_starter and guest_count_side_dish), we will see that these will be eliminated in the process of model selection.

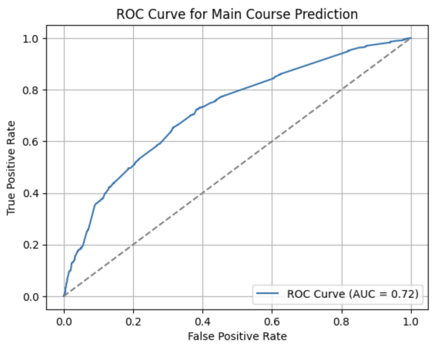
Thus, considering the results of correlation matrix and VIF values we conclude that no multicollineariry assumption is not violated.

- **Model Fit**: We assess model quality using pseudo-R^2 measures and likelihood ratio tests.

```
                           Logit Regression Results
==============================================================================
Dep. Variable:            main_course   No. Observations:                 4057
Model:                          Logit   Df Residuals:                     4046
Method:                           MLE   Df Model:                           10
Date:                Thu, 24 Apr 2025   Pseudo R-squ.:                  0.1083
Time:                        09:30:19   Log-Likelihood:                -2444.3
converged:                       True   LL-Null:                       -2741.1
Covariance Type:            nonrobust   LLR p-value:                 4.411e-121
==============================================================================
```

As we can see the estimated model is jointly significant, as evidenced by the extremely small p-value for the likelihood ratio test ($p < 0.001$), and the pseudo $R^2 = 0.1083$ indicates a reasonable explanatory power for cross-sectional logit models in consumer behavior.

- **Discrimination**: The ability of the model to distinguish between 0 and 1 outcomes is evaluated using classification accuracy, confusion matrices, and ROC curves.



The AUC value of 0.72 indicates that the model has a fair ability to distinguish between orders that include a main course and those that do not.

4.5 Hypotheses

The goal of this model is to test several hypotheses regarding customer ordering behavior. Specifically, we want to examine whether:

- Contextual effects, such as time of day or group size, influence the probability of ordering a main course.
- The presence of both grilled dishes and salad further decreases the probability of ordering a main course.
- Orders that include alcohol are more likely to include a main course.

- Orders that include soup are more likely to include a main course.

The binary logistic regression framework, along with interaction terms, enables us to rigorously evaluate these hypotheses and provide actionable insights to the business.

5 Results

By this time, we have considered the logit model as described in 4 Methodology.

```
                           Logit Regression Results
==============================================================================
Dep. Variable:            main_course   No. Observations:                 4057
Model:                          Logit   Df Residuals:                     4046
Method:                           MLE   Df Model:                           10
Date:                Thu, 24 Apr 2025   Pseudo R-squ.:                  0.1083
Time:                        09:30:19   Log-Likelihood:                -2444.3
converged:                       True   LL-Null:                       -2741.1
Covariance Type:            nonrobust   LLR p-value:                 4.411e-121
=========================================================================================
                            coef    std err          z      P>|z|      [0.025      0.975]
-----------------------------------------------------------------------------------------
const                    -1.0442      0.089    -11.798      0.000      -1.218      -0.871
grill                    -0.6002      0.140     -4.285      0.000      -0.875      -0.326
salad                     0.6962      0.123      5.665      0.000       0.455       0.937
alcohol                  -0.2611      0.094     -2.789      0.005      -0.445      -0.078
cold_drink                0.9052      0.076     11.962      0.000       0.757       1.053
guest_count               0.6521      0.051     12.803      0.000       0.552       0.752
lunch                     0.1929      0.072      2.662      0.008       0.051       0.335
grill_salad              -1.1896      0.219     -5.443      0.000      -1.618      -0.761
alcohol_starter           0.3557      0.137      2.588      0.010       0.086       0.625
guest_count_side_dish    -0.2620      0.046     -5.645      0.000      -0.353      -0.171
evening_soup              0.4173      0.128      3.254      0.001       0.166       0.669
=========================================================================================
```

Figure 10: Logit model output

In the process of model selection we are left with statistically significant variables, that allow for meaningful interpretation:

Interpretation of Coefficients

- **Constant (Intercept)**: Coef = -1.0442, Odds Ratio = 0.352. The baseline odds of ordering a main course are low when all other predictors equal zero.
- **Grill**: Coef = -0.6002, Odds Ratio = 0.549. If a guest ordered a grilled item, the odds of also ordering a main course decrease by 45.1%, suggesting grilled items may substitute for mains.
- **Salad**: Coef = 0.6962, Odds Ratio = 2.006. If a guest ordered a salad, the odds of ordering a main course are approximately twice as high, indicating that salads may complement rather than replace main courses.
- **Alcohol**: Coef = -0.2611, Odds Ratio = 0.770. Ordering alcohol decreases the odds of selecting a main course by 23%, suggesting alcohol may be paired more with lighter or shared meals.
- **Cold Drink**: Coef = 0.9052, Odds Ratio = 2.472. Cold drinks increase the odds of ordering a main course by 147%, possibly reflecting bundled or combo meal behavior.
- **Guest Count**: Coef = 0.6521, Odds Ratio = 1.920. Each additional guest nearly doubles the odds of ordering a main course, reinforcing the opportunity for group-based promotions.
- **Lunch**: Coef = 0.1929, Odds Ratio = 1.213. Lunch orders are 21.3% more likely to include a main course, supporting lunch-focused marketing or deals.
- **Grill + Salad (Interaction)**: Coef = -1.1896, Odds Ratio = 0.304. The combination of grilled items and salads reduces the odds of ordering a main course by about 70%, indicating this pairing may act as a standalone meal.
- **Alcohol + Starter (Interaction)**: Coef = 0.3557, Odds Ratio = 1.427. When alcohol is ordered along with a starter, the odds of a main course increase by 42.7%, suggesting a more complete or celebratory meal experience.

- **Guest Count + Side Dish (Interaction)**: Coef = -0.2620, Odds Ratio = 0.770. For each additional guest who orders a side dish, the odds of ordering a main course decrease by 23%, which may point to side dishes replacing main courses in group orders.

- **Evening Soup**:Coef = 0.4173, Odds Ratio = 1.518. Ordering soup in the evening increases the odds of including a main course by 51.8%, suggesting dinner-time preferences for fuller meals.

Variables Removed from the Model

Insignificant variables were excluded from the model, as they do not have a statistically meaningful impact on the likelihood of ordering a main course within our sample.

These include: **Side Dishes, Starters, Dessert with Hot Drinks, Soup, Duration, Weekend Indicators, and Most Interactions with Duration**.

The lack of statistical significance suggests these variables do not provide additional explanatory power in predicting main course orders, at least within this model structure. This does not imply these features are unimportant in general, but that their effects may be better captured through other variables or combinations, or are simply negligible in this business context.

6 Conclusions

Our research confirms that customer ordering behavior in the cafe is significantly influenced by both individual dish categories and the context in which the order is made — including the time of day, guest count, and combinations of items. These findings enable the development of targeted promotional strategies and procurement recommendations to optimize customer spending and operational efficiency.

6.1 Key insights

We summarize the key insights below, structured by each research question:

1 What is the impact of time of day, day of week, month on spending patterns? Do these factors influence popularity of the category?

Based on the data analysis presented in Section 3.2 of our report, we identify clear temporal patterns in customer spending behavior:

1.1 Main courses dominate during lunch and evening hours, and on weekends, confirming them as key targets for timed promotions and inventory planning.

1.2 Hot drinks and soups perform strongly in the morning and midday, suggesting suitable pairings for breakfast and lunch sets.

1.3 Alcohol and cold drinks see peak demand during evenings and weekends, highlighting them as strong complements for main dishes during those times.

2 What categories of dishes are likely to be ordered together?

Based on the correlation analysis between dish categories presented in Section 4.4 of our project, we point out stable patterns of recurrence of certain categories of dishes:

2.1 Main courses are regularly ordered with cold drinks and starters.

2.2 Salads tend to be ordered together with grill items, side dishes, and starters.

2.3 Starters are frequently accompanied by alcohol, cold drinks, salads, and main courses.

2.4 Grill items are commonly ordered with side dishes, salads, and starters.

2.5 Cold drinks often appear alongside main courses and starters.

2.6 Alcohol is frequently ordered together with starters and grill items.

These combinations indicate a recurring ordering behavior that reflects the overall order profile.

3 How do the time of day (morning, lunch, evening) and the type of dish interact to influence the main dish purchase?

Based on the results of the logistic regression model presented in Section 5 of the report, we observe the following interaction patterns between time of day, dish type, and the likelihood of ordering a main course:

3.1 Salads, soups, and a higher guest count significantly increase the probability of ordering a main course.

3.2 Grilled dishes, particularly when ordered together with salads, are associated with a decreased probability of main course selection.

3.3 Alcohol alone is negatively associated with main course orders; however, when ordered in combination with a starter, it increases the likelihood of a main course being ordered.

3.4 Soup ordered during evening hours is strongly associated with an increased probability of main course selection.

6.2 Recommendations

The insights formed the basis for a set of practical recommendations aimed at enhancing the cafe's menu strategy.

In particular, we identified recurring ordering patterns that can be leveraged through well-timed combo offers. The combinations outlined in Table 1 are tailored to reflect the most frequent and profitable pairings observed in the data. Each combo is supported by specific findings from our analysis.

Combo Components	**Recommended Time**	**Based on Finding(s)**	**Business Benefit**
Starter + Main course + Cold drink	Lunch (12:00–16:00)	1.1, 2.1, 2.3, 2.5	This combo builds on the already high demand for main dishes during lunch and enhances profitability by encouraging customers to add a starter and a cold drink. Both additions typically have high margins and low preparation costs, making this a lucrative upsell strategy without significantly increasing perceived cost for the customer.
Salad + Grill	Lunch (12:00–16:00)	2.2, 2.4	Grill items are the most expensive dishes on the menu and are commonly ordered by larger groups. Pairing them with a salad creates a balanced, appealing combo that can boost the average order value while catering to customers who are willing to spend more during social or group dining occasions.
Soup + Main course	Evening (16:00–22:00)	1.2, 3.4	Soups not only remain popular in the evening but are also statistically linked to a higher likelihood of ordering a main course. This pairing subtly nudges undecided customers toward choosing a main dish, while offering a satisfying and well-structured meal experience.
Main course + Cold drink	Evening (16:00–22:00)	1.1, 1.3, 2.1, 2.5	Given the evening popularity of main dishes, combining them with cold drinks provides an easy and effective upsell opportunity. Cold drinks have high markups, and this combo leverages existing ordering patterns to maximize revenue with minimal operational complexity.
Alcohol + Starter	Evening (16:00–22:00)	1.3, 2.3, 2.6	Starters served with alcohol tend to increase the likelihood of customers proceeding to order a main course, as identified in our model. This combination is particularly effective on weekends, when group dining and alcohol consumption are more common. It serves as both a profitable pairing and a behavioral trigger for larger orders.

Table 1: Recommended combos based on behavioral patterns and logistic regression findings

By combining descriptive analytics with regression-based modeling, this study provides actionable strategies for menu design, targeted promotions, and procurement planning, all of which can help the cafe increase revenue while improving customer satisfaction and operational flow.

Literature

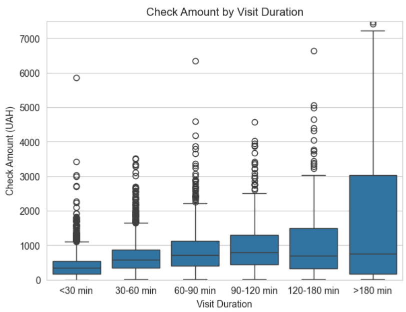
- Provost, F., & Fawcett, T. (2013). Data science for business: What you need to know about data mining and data-analytic thinking.
- Kimes, S.E.(2004). Restaurant revenue management: Implementation at Chevys Arrowhead.
- Wooldridge, J. M. (2015). "Introductory Econometrics: A Modern Approach."

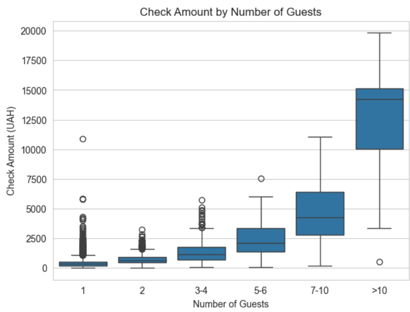
References

[1] Logistic and Linear Regression Assumptions: Violation Recognition and Control, https://www.researchgate.net/publication/341354759_Logistic_and_Linear_Regression_Assumptions_Violation_Recognition_and_Control

Appendix

Link to Github repository: https://github.com/pythotonix/Econometrics-Final-Project.git





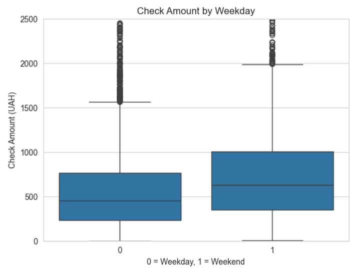


Figure 1: Visit Duration & Check Amount 2: Number of Guests & Check Amount 3: Weekday & Weekend comparison

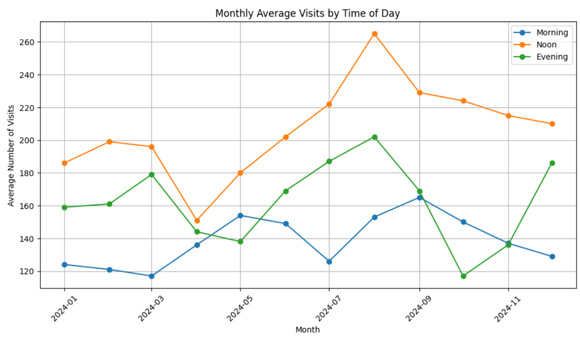
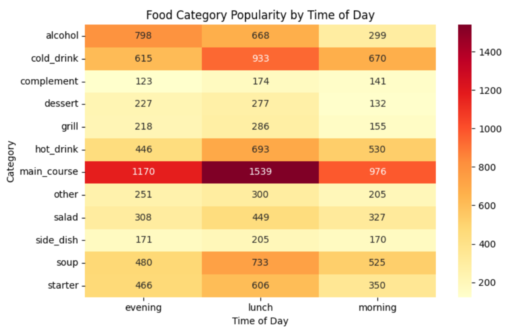


Figure 4: Average Visits Amount by Time of Day



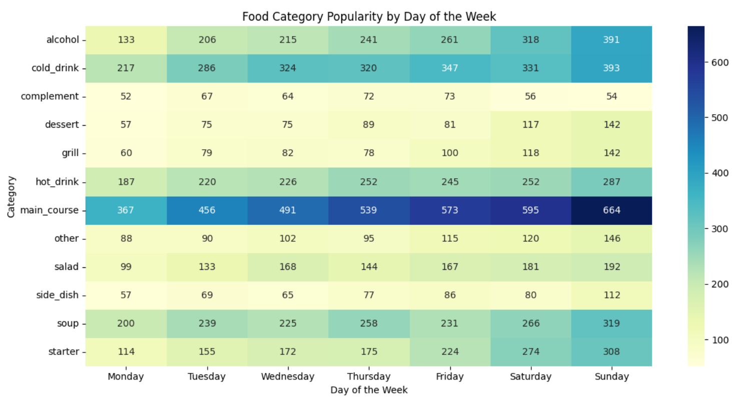


Figure 5: Food Category Popularity by Time of Day Figure 6: Food Category Popularity by Day of the Week

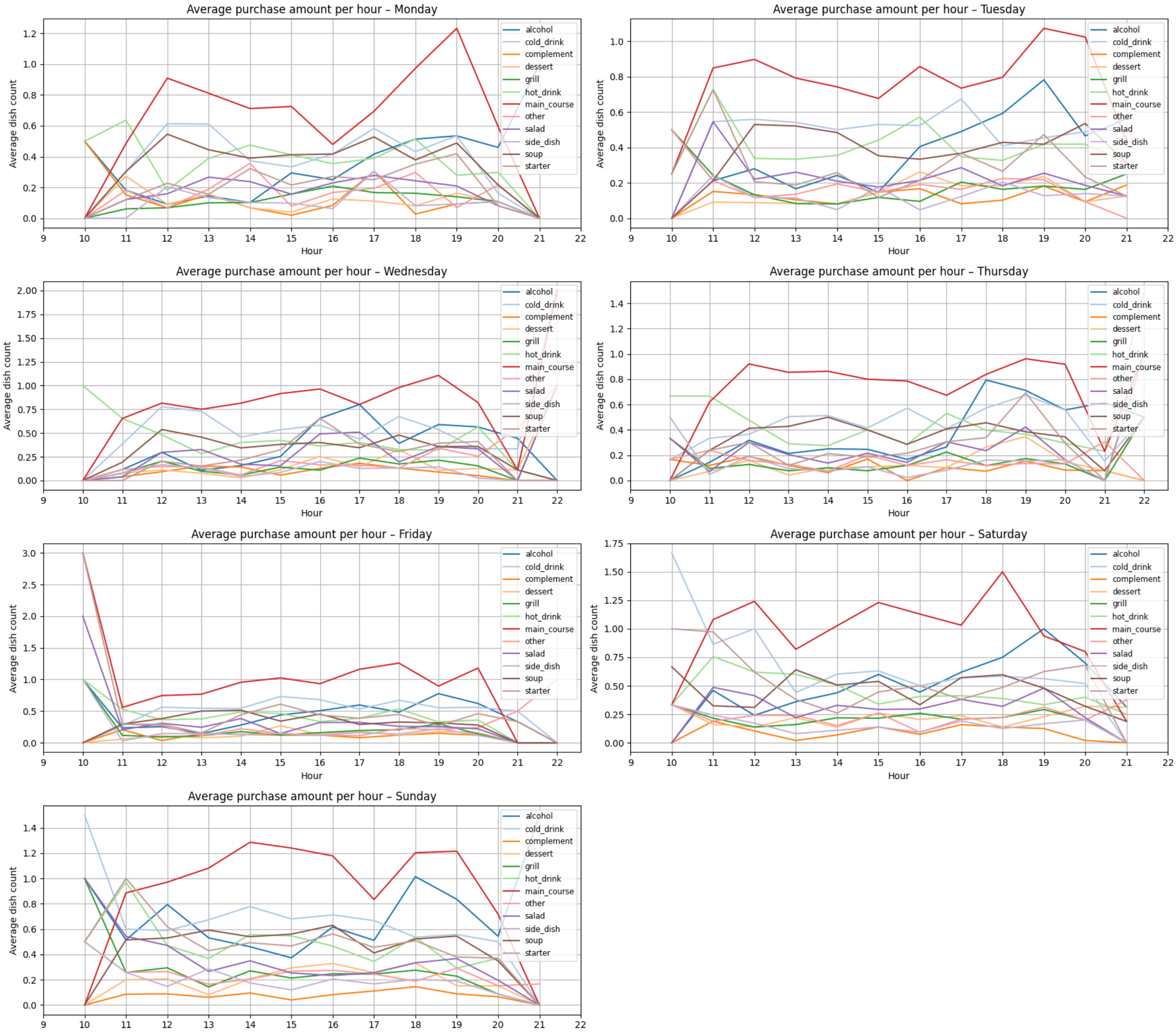


Figure 7: Food Category Popularity by Day of the Week

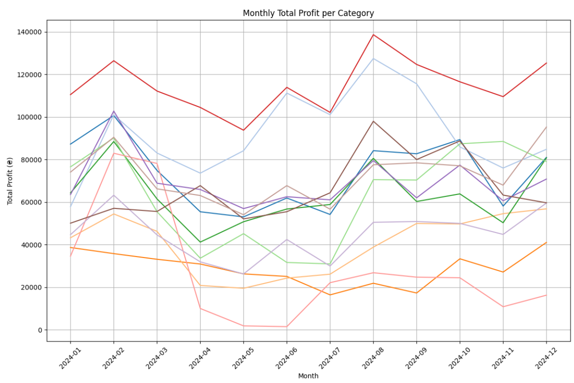


Figure 8: Food Category Profitability by Month

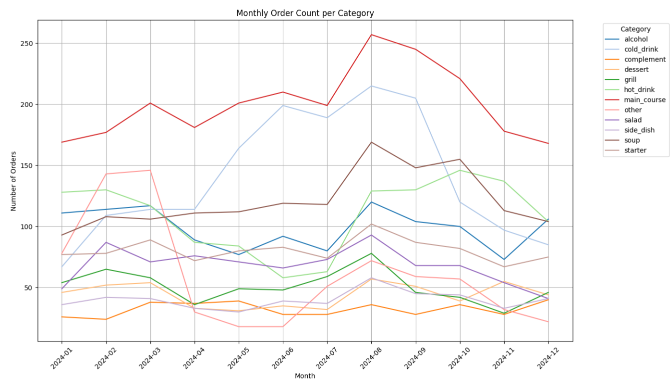


Figure 9: Food Category Popularity by Month