What are the best times to raise prices on Instacart?

I. Introduction

Following the years of the Pandemic, grocery delivery and pickup services such as Instacart and DoorDash have grown dramatically. More specifically, Instacart has grown by impressive margins, surging from 3.3 million annual users in 2017 to 14.4 million as of last year, and this year is projected to reach 14.9 million annual users. However with all these users, comes an increase in orders: roughly 9 per second, or approximately 260 million per year (1).

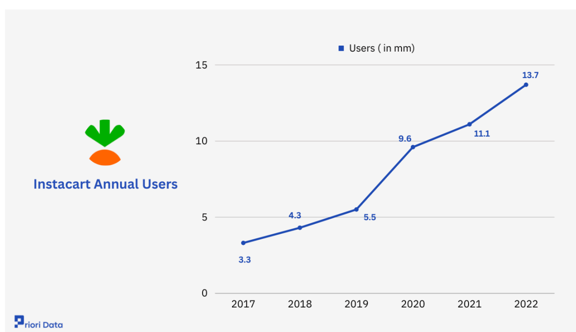

Figure 1. A graphical representation of the number of users on Instacart by year (1).

To increase profits, Instacart uses a process called "surge pricing" where prices on different products are raised at different times to match demand or supply. We wonder, is this idea backed by statistical evidence, or is it unnecessary? To answer our question, we conducted an observational study using publicly released data from Instacart to determine if there was a difference in consumption patterns at different times.

II. Statistical Question

Among Instacart users who have placed at least four orders, is there a statistically significant difference in the average time of day when orders are placed across different departments?

III. Data Collection

The data used in this study comes from the publicly available Instacart Market Basket Analysis dataset from 2017 (2), which was originally published for a Kaggle competition. The data includes over three million orders from more than 200,000 customers who placed at least four orders on Instacart. Although the dataset has many insightful details, the sampling methodology has not been fully disclosed by Instacart. Therefore, we will proceed with caution when generalizing the results to all Instacart users. However, given the size of the dataset and its use in a data science competition, it’s most likely that the data was sampled to represent all of Instacart’s users on their servers.

Due to the size of the dataset, it will not fully be displayed here. Instead, we can provide a preview of the first ten rows of orders made by the customers:

First 10 Rows of Order Products

department	order_hour_of_day
beverages	7
beverages	12
beverages	7
beverages	15
beverages	7
beverages	9
snacks	14
dairy eggs	16
beverages	8
produce	10

Figure 2. First ten lines of the dataset visualized in Matplotlib.

IV. Data Description

The dataset includes many categories to analyze buying patterns, but the ones we will focus on will include the hour of the day in which a product was purchased and the 19 different departments into which the ~50 thousand products can be categorized (We exclude the 'missing' and 'other' departments). In the context of our study, a department refers to one of Instacart's predefined product categories that group similar items, like produce, frozen, or bakery.

To answer the statistical question, we will run a Kruskal-Wallis test, which is a non-parametric statistical test to compare the distributions of a continuous variable (the hour of the day when an order is placed in this case) across at least three independent groups. Each group will be represented by a department and its mean purchase hour and variance will be calculated. The reason we cannot run an ANOVA (Analysis of Variances) test is that the variances change across departments. Then, using Python, we will run the Kruskal-Wallis test using Scipy.stats and graph the distributions of various departments by hour using Matplotlib. All code used is referenced on GitHub (3).

V. Statistical Test Conducted

Null hypothesis: The distribution of order times in hours of day is the same across all departments.

Alternative hypothesis: At least one department has a different distribution of order times compared to the others.

Significance level: $\alpha = 0.01$

Before running the test, we need to check the assumptions for Kruskall-Wallis.

Independence: ~200 thousand sampled users without replacement is less than 10 percent of the total 3.3 million users on Instacart at the time. ✔

Ordinal or continuous scale: The variable representing the order hour of day is continuous as it ranges from 0 to 23 and can be ranked as well, making it ordinal. ✔

Similar Distribution shapes: All of the departments are slightly skewed left to a similar degree, and we can proceed as the test only requires the distributions to be broadly similar (even though there are small variations between separate departments). ✔

Sample Size: Every group (department) has a sample size of at least five (the lowest sample size is over three thousand).✔

Put simply, the Kruskal-Wallis test ranks all the observations from every department and finds the average rank of the respective department. In this case, the average rank represents the average hour in which a product is bought per department, as the hours range from 0 to 23. The rank is found by first ranking all the hours in the dataset from lowest to highest, irrespective of department. Then, each order is ranked based on the time of day. For example, buying beverages at 8 AM will have a lower rank than buying beverages at 6 PM, and if multiple orders happen during the same hour, they are given the average of the ranks they occupy. Finally, after all, three million orders are ranked in this way because every order has a corresponding department, the ranks are grouped by department and averaged to get the following table:

Average Rank of Order Hour by Department

Department	Average Rank
beverages	1577510.6
breakfast	1577852.4
bulk	1577888.0
dairy eggs	1581904.9
household	1584677.5
snacks	1593263.1
produce	1593880.1
bakery	1602431.2
pantry	1615193.3
babies	1617963.8
canned goods	1621417.8
meat seafood	1629039.4
personal care	1630280.6
deli	1635335.8
international	1644993.4
dry goods pasta	1661384.0
alcohol	1685209.2
pets	1700821.8
frozen	1701983.2

Figure 3. Average rank calculated in Python.

Then, we determine how much the ranks differ between groups by finding the Krustal-Wallis test statistic H, which gives us the degree to which rank distributions diverge from the expected equal distribution.

Number of cases in group i

Mean rank sum in group i

Total sample size

Expected value of the rankings

$$H = \frac{n-1}{n} \cdot \sum_{i=1}^{k} \frac{n_i \cdot (\bar{R} - E_R)^2}{\sigma^2}$$

Rank variance

Figure 4. Krustal-Wallis Test Statistic (4).

- In our test, $n = 3,214,875$, which is the total number of orders

- $k = 19$ is the number of departments
- R, the mean rank sum is taken from Figure 3 for each respective department
- E_R, the expected average rank under the null hypothesis is

$\frac{n+1}{2} = \frac{3,214,875+1}{2} = 1,607,438$

- $\sigma^2 = \frac{n^2-1}{12} = \frac{3,214,875^2-1}{12} \approx 8.61 \times 10^{11}$

Plugging all these values into the formula (or implementing it using Python as it was done here), we get the H statistics to equal 3869.023.

Normalized Department Contributions to Kruskal–Wallis H-statistic

Department	Contribution to H
frozen	2114.53
dairy eggs	316.28
beverages	294.22
alcohol	279.94
dry goods pasta	210.44
pets	139.45
deli	112.62
produce	74.62
meat seafood	59.56
personal care	47.29
international	42.12
breakfast	39.60
household	30.77
pantry	30.09
canned goods	26.92
snacks	21.49
babies	7.12
bulk	2.54
bakery	0.00

Figure 5. Each Department's contribution to the H-Statistic implemented in Python.

Finally, to find the p-value, we can use Python's scipy.stats.kruskal function, which computes it for us to get a p-value of approximately 0 (less than 10^{-308}). Alternatively, we can use the chi-squared distribution as well with $df = k - 1 = 18$ degrees of freedom as the

Kruskal-Wallis test statistic approximately follows a chi-squared distribution when the sample size is large, also giving us a p-value of approximately 0.

VI. Data Analysis

The Kruskal-Wallis test revealed a p-value of 0, which is well below the significance level of $\alpha = 0.01$, meaning that we reject the null hypothesis. Therefore, there is strong statistical evidence that at least one department has a different distribution of order times compared to the others. The mean hour at which departments are bought can better help us understand the data:

Mean Hour of Day Orders Are Placed (Overall and by Department)

Department	Mean Order Hour
Overall	13.44
beverages	13.33
breakfast	13.36
dairy eggs	13.37
bulk	13.37
household	13.37
snacks	13.41
produce	13.41
bakery	13.44
pantry	13.50
canned goods	13.52
meat seafood	13.54
babies	13.55
personal care	13.55
deli	13.58
international	13.60
dry goods pasta	13.67
alcohol	13.75
pets	13.82
frozen	13.84

Figure 6. Mean order hour for every department.

Referencing Figures 5 and 6, we can observe that the frozen department has the highest contribution to the test statistic and the highest mean hour order. A possible explanation for this could be due to late-night impulse buying or cravings. Frozen pizza and ice cream, both being part of the frozen department, seem to consistently be purchased later than other categories:

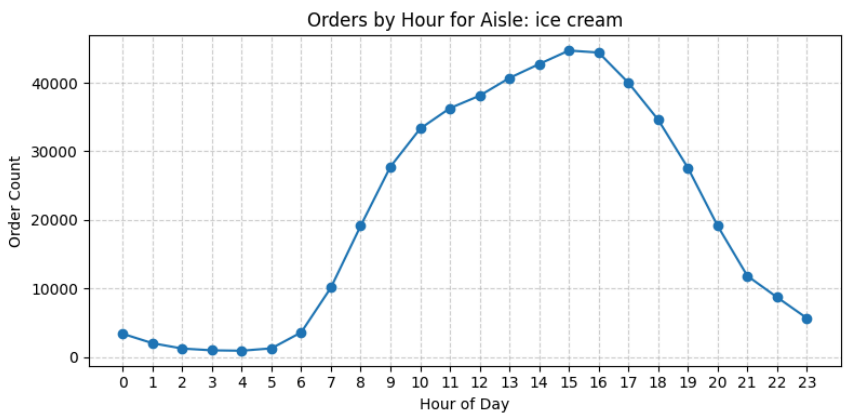

Figure 7. Order by Hour for Ice Cream.

On the flip side, beverages also have a relatively high contribution to the test statistic, however they are usually bought in the mornings:

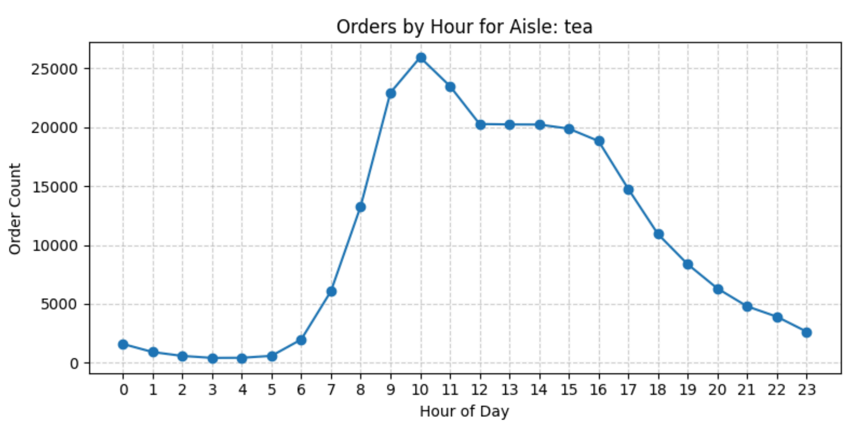

Figure 8. Mean order hour for tea (visualized in Python using Matplotlib).

A possible explanation for tea (and other beverages) being bought so early may be due to consumers habitually buying in the morning to wake themselves up.

Lastly, an explanation as to why all distributions across departments tend to dip at the 12-hour mark may be due to most people having lunch at around noon and customers being away from their phone or computer; not actively shopping.

VII. Conclusion and Implications

Overall, the findings provide actionable insights into consumer purchasing behavior. Dynamic pricing and personalized marketing can help e-commerce companies maximize their profits. For example, Instacart may raise the price of frozen foods like ice cream during peak hours in the afternoon to maximize profits or make sure that these foods are stocked in time. The peak purchase window for beverages seems to be more narrow, meaning that a small price increase of tea or coffee in the morning may be absorbed due to high consumer demand. Personalized coupons and discounts may be offered mid-day when demand falls slightly.

Regarding the potential sources of error in this study, we must be cautious when generalizing the result to the entire Instacart user base. The dataset is from 2017, and unfortunately, Instacart has not disclosed the methodology by which the data was sampled, meaning that we can only infer that it is representative of the population. Additionally, the user base of Instacart has grown more than four-fold, and post-pandemic shopping habits may have changed the way in which consumers purchase products today. If we had the ability to run the experiment again, we would likely make sure to get more recent data and use a simple random sample to gather it. Other factors such as user demographics may also impact purchasing decisions, which would be helpful in market basket analysis.

Works Cited

1. Larson, S. (2025, May 6). *Instacart Revenue, Valuation & Stats 2025*. Priori Data. https://prioridata.com/data/instacart-stats/
2. H, M. Y. (2022, January 25). *Instacart Online Grocery Basket Analysis Dataset*. Kaggle. https://www.kaggle.com/datasets/yasserh/instacart-online-grocery-basket-analysis-dataset
3. Andrey199123. (2025, May 27). *Andrey199123/ASA-statistics-research-project: Statistical analysis of Instacart Order Behavior by department and time of day using the kruskal–wallis test to identify optimal pricing windows.* GitHub. https://github.com/Andrey199123/ASA-Statistics-Research-Project
4. *T-test, Chi-square, ANOVA, regression, correlation...* Datatab. (n.d.). https://datatab.net/tutorial/kruskal-wallis-test