

Strategies for optimizing electronic tips service profit

Ekaterina Dmitrievna Lapina^{1,*}, Viacheslav Igorevich Gorikhovskii^{2,*}

¹ Graduate School of Management, Saint Petersburg State University, 198504 Saint-Petersburg, Russia

² Faculty of Mathematics and Mechanics, Saint Petersburg State University, 198504 Saint-Petersburg, Russia

* **Corresponding authors:** Ekaterina Dmitrievna Lapina, cat.lap2018@yandex.ru; Viacheslav Igorevich Gorikhovskii, gorihovskyyvacheslav@gmail.com

ARTICLE INFO

Received: 16 June 2023
Accepted: 31 July 2023
Available online: 9 August 2023

doi: 10.59400/jam.v1i2.105

Copyright © 2023 Author(s).

Journal of AppliedMath is published by Academic Publishing Pte. Ltd. This article is licensed under the Creative Commons Attribution License (CC BY 4.0).
<https://creativecommons.org/licenses/by/4.0/>

ABSTRACT: Electronic tip systems have become very popular in the era of cashless payments. With the widespread use of such services, the problem of maximizing profits has arisen, which concerns both establishments using electronic tip services and the services themselves that provide such services. Identifying factors that influence guests' economic behavior when leaving tips will allow for the creation of an optimal strategy to increase the efficiency of the system. This study used data on the profits of the electronic tip service in public catering establishments. A simulation model was created to evaluate and compare the effectiveness of different strategies; a methodology for finding the optimal strategy was described; and a clustering of establishments was performed using analysis of variance to customize the optimal strategy.

KEYWORDS: electronic tips; profit maximization; analysis of variance; simulation model; optimal strategy

1. Introduction

In the last ten years, the share of non-cash payments in various transactions has increased day by day. That is why the issue of an economic phenomenon widespread in the sphere of service tips has risen. Electronic tip systems have become popular in recent years.

The problem of increasing electronic tip service profit can be solved by finding factors that influence the amount of tip left and finding optimization strategies. Since the tip culture is most developed in the catering industry, this industry is of the greatest interest. Learning optimization strategies is necessary for several reasons:

- An optimization strategy will allow businesses to increase revenues, improve service quality, and improve customer satisfaction.
- For electronic tip services, the optimal strategy will maximize profits.

In the article, the profit of the electronic tip services in the catering sector is analyzed. In restaurants or cafes connected to the service, guests are offered to scan the QR code and go to the page of electronic tips. In the "Tips" section, the guest is offered a set of five buttons. Four of them have buttons with the values of either the percentage of the check with a pre-calculated amount of tips left in rubles (₽) or buttons with specific amounts (the set of buttons offered may vary depending on the size of the check). The fifth button allows users to manually enter the desired tip amount. The main part of the work is a study of the effect of the set of buttons offered to the guest on the number of tips left.

Due to the fact that electronic tips are just beginning to gain popularity, this area has not yet been sufficiently studied due to trade secrets. Although there are several approaches to solving the problem of profit maximization for electronic tip systems, it is quite difficult to compare the effectiveness of the approaches with each other, and there is no tool to determine the most effective of them.

The main goal of the work is to determine the factors influence the profit, create a tool to estimate the effectiveness of the strategy of offering sets of buttons and find strategies that will increase profit.

For a more accurate analysis, the data were divided into groups according to the amount of the check. The groups were formed based on the results of the analysis of variance, which allows us to identify the presence of statistically significant differences in the average percentage of tips from the check amount in the groups under consideration^[1]. Post hoc tests allowed us to determine the exact boundaries of the intervals between the groups^[2].

Simulation modeling was performed to determine the optimal set of buttons for each range. Simulation modeling is needed to create virtual models of real systems and processes that allow analyzing their behavior and predicting results under various conditions^[3]. Based on the obtained sets, an optimization strategy was created.

A more precise division of data into groups was proposed: clusters were created on each of the ranges of the check amount based on the region, the price segment of the restaurant, and the time of the transaction. The customized strategy is based on finding optimal sets for each of the clusters separately.

2. Materials and methods

For processing and analysis, a database containing information about all transactions carried out in establishments using the service “Netmonet” since July 2022 is used. The database contains information about transactions when sets of buttons were offered randomly and about transactions when data was divided into groups by the amount of the check and a fixed set was offered for each group. The database contains:

- the number of tips left;
- amount of the check;
- time of the transaction;
- region where the place is located;
- price segment of the place;
- offered a set of buttons.

Of interest for the study are those transactions for which the total amount of the check is known, because the key role in the analysis and building an optimal strategy is to determine the percentage of tips left of the total amount of the check. It should also be noted that the database contains only “completed” transactions, that is, if the user scanned the code and got to the page for leaving the electronic tips, but the tips were not left, then there is no information about it in the database.

2.1. Analysis of variance

Simulation modeling and analysis of variance were used for a detailed analysis of service profits. Analysis of variance (or ANOVA) and further post hoc tests were performed to identify the influence of various factors on the number of tips left. Based on these results, an optimal strategy was constructed. ANOVA test hypotheses for n groups:

$$H_0: \mu_1 = \mu_2 = \dots = \mu_n$$

$$H_1: \exists i, j \in \{1, \dots, n\}: \mu_i \neq \mu_j$$

where μ_i mean of i -th group^[1].

All transactions were divided into 11 groups by check amount. The rationale for the separation intervals is given below.

An analysis of variance was used to see if there is a statistically significant difference between the mean values of the percentage of check on the different intervals of the check amount.

To perform the analysis of variance, the following conditions must be checked:

- 1) Quantitative data type.
- 2) Independent samples.
- 3) Normal distribution of the studied trait in the population from which the samples were taken.
- 4) Equality of variance of the studied trait in the populations from which the samples were taken.
- 5) Independent observations in each of the samples.

The average tip percentage of the check amount is a quantitative value; the groups are independent because one transaction can only be attributed to one of the intervals of the check amount. Random samples of size 1000 from each interval were used to test statistical hypotheses.

Levene's test is used to test that k samples have equal variance. Levene's test hypotheses:

$$H_0: \sigma_1^2 = \sigma_2^2 = \dots = \sigma_k^2$$
$$H_1: \exists i, j \in \{1, \dots, k\}: \sigma_i^2 \neq \sigma_j^2$$

where σ_i^2 variance of i -th group. Formulas for calculating the statistics of the Levene's test^[4].

The Levene's test with a significance level of $\alpha = 0.01$ was used.

The result statistic is 245.531, and the p -value is 0.00. According to the p -value, it was concluded that there are differences in the value of variance in the groups under consideration. However, the results of the analysis of histograms revealed that there is a shift of the distribution to the left and the distribution of the data is not normal, so a modified version of the Welch test for trimmed means and Winsorized variances was used^[5]. The test was also conducted with a significance level of $\alpha = 0.01$.

The result statistic is 146.872, and the p -value is 0.00.

According to the results of the test, it can be concluded that there are statistically significant differences between the mean values of the groups.

Post hoc tests are an integral part of analysis of variance. Using analysis of variance to test for the equality of multiple group means will help determine if there are statistically significant differences between the group means. However, the results of such analysis do not determine which differences between pairs of means are significant. Post hoc tests are used to examine the groups between which a difference emerged from the results of the analysis of variance. When multiple pairwise comparisons are made, the probability of type I error increases; post hoc tests solve this problem^[5]. The Games-Howell method is one of the methods of post hoc analysis applicable in cases where the assumption of equality of variance is violated. This method controls for type I error and maintains a given level of significance even with different sample sizes. The test does not assume equality of variance in the groups under consideration, equal sample sizes, and is robust to data not obeying the normal distribution law in cases of large sample size^[6]. The test was conducted with a significance level of $\alpha = 0.01$. The post hoc Games-Howell test showed that there were significant differences in 47 out of 55 range pairs.

Based on the results of the statistical tests, it was concluded that the average percentages of tips at different intervals of the check amount have statistically significant differences, so the simulation will be run on 11 different intervals of the check amount.

2.2. Simulation modeling

Simulation modeling is an important step in research. It involves creating an effective simulation model that is as accurate as necessary to simulate the behavior of the underlying system and reproduce useful observations for later analysis.

Monte Carlo simulation is a very useful mathematical technique for analyzing uncertain scenarios and probabilistic analyses of various situations^[7]. In Monte Carlo simulation, we define a statistical distribution that we can use as a source for each of the input parameters. We then take random samples from each distribution, which then represent the values of the input variables. For each set of input parameters, we obtain a set of output parameters. The value of each output parameter represents one particular scenario in the simulation. We collect such output values from several simulation runs. Finally, we perform a statistical analysis of the values of the output parameters to decide on further actions.

In the case of the electronic tips service, simulation modeling will allow us to estimate the expected profits, provided that certain sets of buttons are offered.

Preliminary preparation is required for modeling.

- Selection transactions from the database, where the investigated set of buttons was proposed;
- Calculation: the probability of selecting a button from a set or manual input. As a probability, it was the frequency probability—the ratio of the number of button selections (or manual input selections, respectively) to the total number of transactions;
- On the condition that the button is selected from the set, calculate the probability of the user selecting each specific button out of the four buttons (as a ratio of the number of selections of the first button, second, third, and fourth buttons from the set to the total number of selections of the button from the set, not manual input);
- Modeling check amounts with the same distribution as the original set of check amounts on each interval. At each of the 11 intervals of the check amount, all transactions were divided into groups in increments of 50 ₺. The probability of the check amount falling into each of the small ranges is found as the ratio of the number of check amounts falling into this range to the total number of check amounts in the interval under consideration. With further generation of the check amounts, they fell into each narrow range of 50 ₺ with given probabilities.

Three different models were used for different types of offered sets: a model for a set of buttons offering specific amounts; a model for a set of buttons with a percentage of the check; and a model for the case where a set is offered at random.

The principle of the models for sets with specific amounts and for sets with percentages is:

- A set of buttons, the calculated probabilities of selecting some of those buttons, or manually input, and a set of tips that have been left manually are input.
- The first step of the model simulates the selection of a button from the set or manual input with the given probabilities.
- If manual input was chosen, we randomly selected the size of the tip left from the set of tip values given as input.

- If the button from the set was selected, with a given probability, we would choose which button of the four was chosen.
- For all values from the set of check amounts, simulate the tip amount using the procedure described above and determine what percentage of the check amount it represents. The result of the simulation is the average percentage of tips from the check amount.

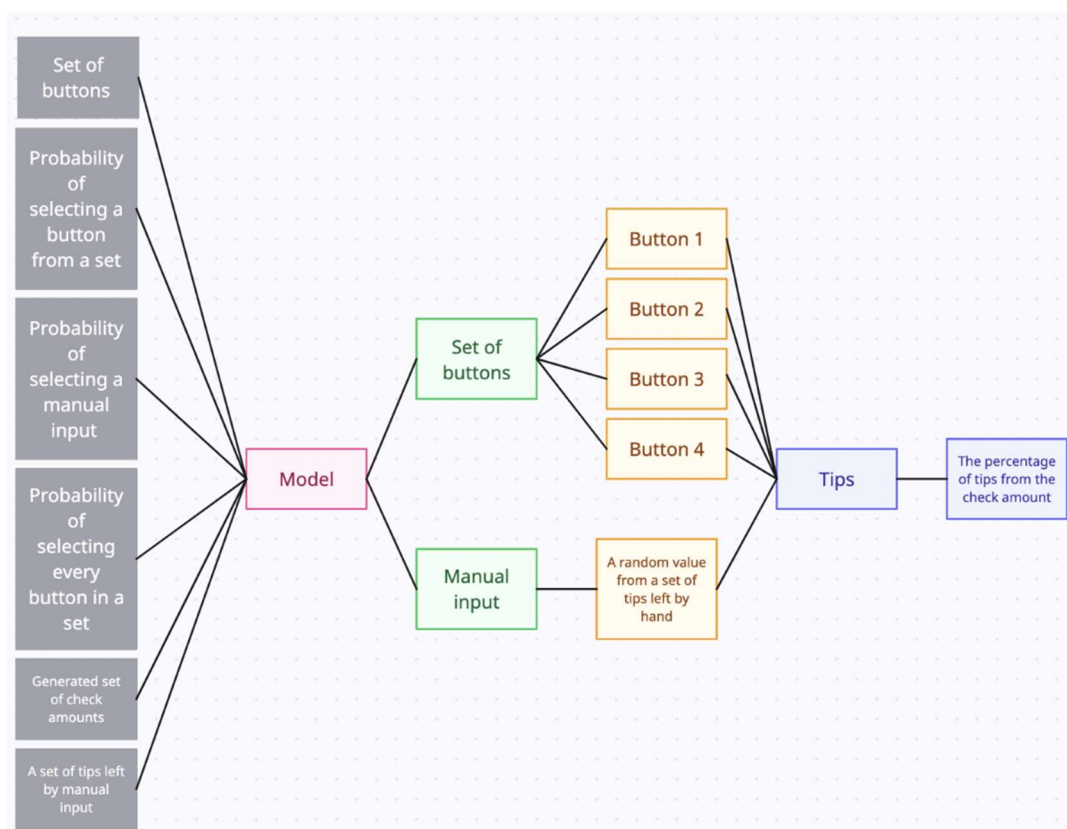


Figure 1. Scheme of simulation modeling.

The operating principle of the simulation model with a random suggestion of buttons is shown in **Figure 1** and is described in the Appendix.

10,000 simulations are performed for each of the cases in question to obtain results accurate to a tenth of a percent.

According to the simulation results presented in **Table 1**, the difference between the same button sets with a random and deterministic approach (when only one set was offered in each range) was within one percent. Thus, the validation of simulation results on data obtained in various ways showed that simulation modeling can really be used to find the optimal strategy.

Table 1. The average percentage of the tip from the check received after simulation on each of the groups of the check amount with different approaches.

	1	2	3	4	5	6	7	8	9	10	11
Random approach	0.192	0.117	0.111	0.104	0.094	0.095	0.089	0.090	0.085	0.081	0.088
Deterministic approach	0.198	0.117	0.111	0.103	0.094	0.095	0.089	0.087	0.085	0.081	0.086

Offering specific amounts yields a lower predicted average percentage of the check than offering buttons with percentages.

At each of the considered intervals of the check amount with the created tool, the optimal set of buttons was found.

Thus, after the simulations, it was concluded that the proposed set of buttons affects the amount of this left, so the optimal strategy will directly depend on the set of buttons offered.

2.3. Customization

It was hypothesized that not only check amounts affect the number of tips left but also other factors such as the region where the restaurant is located, the price segment the establishment belongs to, and the timing of the transaction.

If this hypothesis is correct, then it would be possible to create an individual strategy determined not only by the amount of the check but also by other factors affecting the behavior of visitors.

Cluster analysis

To make sure that the factors listed above really affect the number of tips left, we took a sample of all transactions for one of the months when the data was collected randomly and conducted a variance analysis of these data to see if there is a statistically significant difference between regions and price segments in the average percentage of tips. The analysis of variance was carried out for each of the considered ranges of check amounts. First, regions were compared, and if statistically significant differences were found, post hoc tests were conducted, according to the results of which the data were divided into groups by region. Then, an ANOVA was conducted within each region to identify statistically significant differences between price segments. In general, the clustering was performed according to these two criteria.

Before applying the analysis of variance, an equality of variance test was performed. The ratio of the maximum variance among the groups to the minimum variance was calculated; if it was less than 4, the variances were assumed to be equal. Levene's test for equality of variance was not performed since the sample sizes were too large. Next, histograms of the average percentage of the check on the groups were plotted, from which it was clear that the data are always strongly skewed to the left and have a large "tail" on the right, as seen in **Figure 2**.

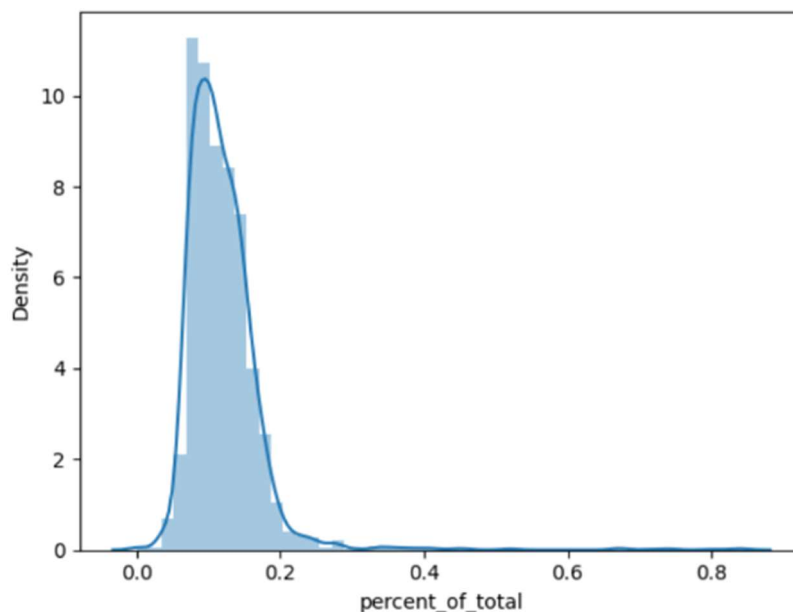


Figure 2. Example of a distribution of the percentage of tips from the check amount.

If the variances were equal, we performed the classical ANOVA test, which is robust to non-normally distributed data. If we took the variances to be unequal, we performed Welch’s ANOVA test for trimmed means and Winsorized variances, which is robust to such situations^[8]. Based on the results of the analysis of variance in the case of statistically significant differences, the post hoc Games-Howell test was conducted, which is robust to the case of non-normally distributed data with large sample sizes and unequal sample sizes^[6].

Based on the results of the analysis of variance, clustering by region and price segment was carried out. The clustering has a tree structure. An example of clustering is presented in **Figure 3**. Within each range, the regions to be considered together or, conversely, separately were first identified. Next, each of the resulting groups of regions was divided into groups by price segments. Thus, for each cluster, which is the path from the root of the tree to the leaf, its own optimization strategy can be developed by further simulation modeling.

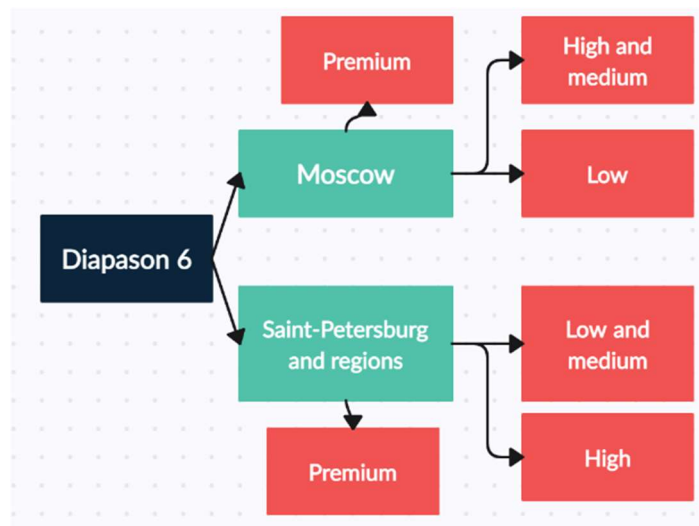


Figure 3. Example of clusterization in one of the check sum groups.

3. Optimization strategies

The analysis of variance showed that many factors influence the amount of profit. One of the key results of the study is to identify the impact of the offered button sets on the size of the estimated profit. Thus, it is important to track trends in the effectiveness of the sets over time (for example, such a fact as seasonality should be taken into account), so it is proposed to show the button sets randomly in a certain percentage of cases.

In the course of the work, two optimization strategies were proposed: the first is based on finding the optimal set in each of the suggested ranges of the check amount, and the second is based on finding the optimal set in the found clusters. For each of these cases, data collection should involve some percentage of buttons offered at random.

In both strategies, in most cases, the sets that performed best in the simulations will be offered equally; in other cases, the set is chosen at random from all the sets ever offered in this interval when a set of buttons was offered randomly.

Software code was written to calculate the maximum possible percentage of buttons that could be offered at random. To do this, a simulation was used, which is a combination of the simulation for offering percentages (or specific amounts) described in the previous section and the simulation for offering

random sets. The marginal value of the interest for all offered sets, at which the expectation of profit with random data coincides with the expectation of profit when the optimal set of buttons is offered, is computed.

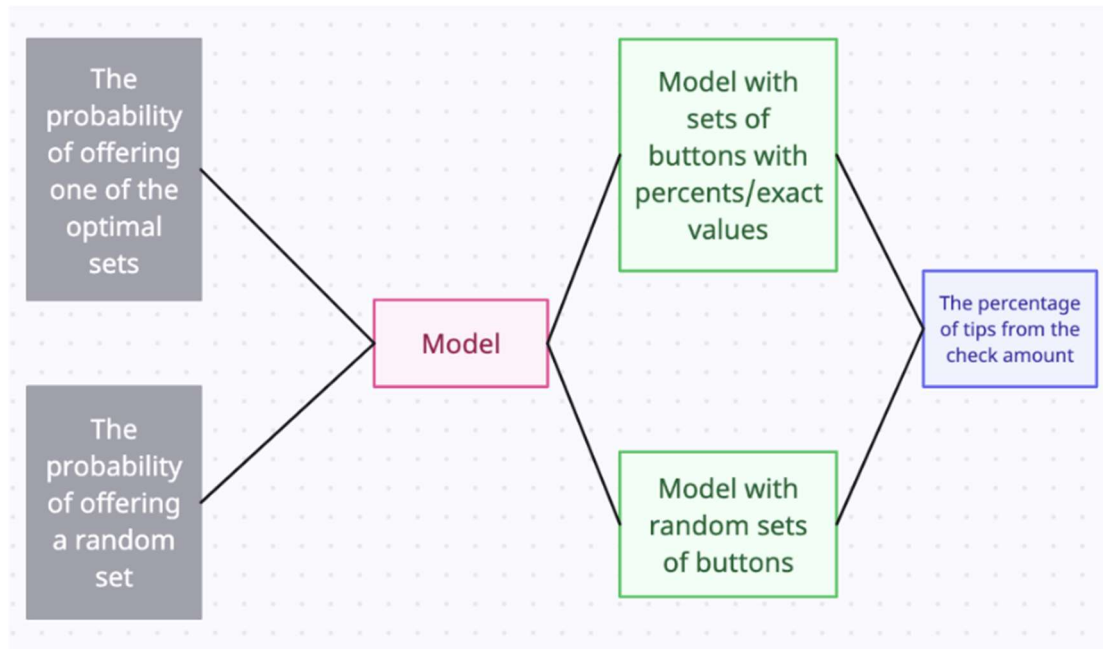


Figure 4. Scheme of finding marginal percents.

For each of the considered ranges of the check amount, a limit value was calculated, as the optimal sets here are those that were tested during the work and gave the best result.

When using a customized strategy, the marginal percentages are found separately for each narrow cluster. The scheme for calculating the marginal percentages is shown in **Figure 4**.

The marginal percentages were calculated for each range, and the marginal percentages were obtained for clusters by region and price.

Using such a method will allow you to collect statistics and subsequently build a customized strategy.

To determine the optimal set of buttons in each range or cluster, simulations on all the sets that have been offered so often that there is enough historical data to perform a comparative analysis between them were run.

To improve the customized strategy, another level of clustering can be added—time of transaction. The time of day was divided into 4 parts: breakfast, lunch, dinner, and night.

However, the data in narrow clusters was not always sufficient to test the effectiveness of the strategy, so where there is insufficient data, it is proposed to show the button sets that are optimal for the broader cluster.

Thus, in clusters built by region, segment, and time, the optimal button sets are selected. If there is one set, it is offered in its cluster in most cases, except for the percentage of cases found earlier. If there is more than one set, they are offered with equal probability; in other cases, random sets are offered. In the case of a data shortage, the strategy for the narrow cluster is the same as the strategy found for the wider cluster.

4. Results

Transactions were divided into groups according to the check amount. It is justified by the analysis of variance. The division into groups was done in order to obtain more accurate information about the results of the offer of sets of buttons, since the same set shows different results at different intervals of the check amount.

A simulation tool has been developed to estimate the expected average percentage of tips from the check amount. The program code can be found at the following link: <https://github.com/Algebravsredu/electronic-tips-profit-optimization/tree/main/anova>. Simulation modeling allowed us to find the optimal sets of buttons on each of the ranges under consideration.

Two optimization options were proposed: the search for optimal sets of buttons on each range of checks and a customized strategy.

To create a customized strategy, the data was divided into clusters based on the characteristics of the region, price segment, and time of the transaction. The division was made on the basis of variance analysis and post-hoc tests. The Python programming code for clustering can be found at <https://github.com/Algebravsredu/electronic-tips-profit-optimization/tree/main/anova>.

It was concluded that it is necessary to monitor the effectiveness of the sets over time, so in optimal strategies, in some percentage of cases, button sets should be offered randomly. However, the percentage of cases when the sets are offered randomly should not affect the expected amount of profit. Therefore, a tool has been developed that allows you to find the maximum percentage of sets that can be offered randomly. The Python program code can be found at the link <https://github.com/Algebravsredu/electronic-tips-profit-optimization/tree/main/simulation>. Such marginal percentages are found in the case of any of the optimization strategies, either for each range of the check amount or for each cluster.

5. Conclusion

In the paper, a methodology for finding strategies for optimizing the profit of the electronic tip service is described, and a source code that can be used by readers is also provided.

Author contributions

Conceptualization, EDL and VIG; methodology, VIG; software, EDL; validation, EDL and VIG; formal analysis, EDL and VIG; investigation, EDL; resources, EDL and VIG; data curation, EDL and VIG; writing—original draft preparation, EDL; writing—review and editing, EDL and VIG; visualization, EDL; supervision, VIG; project administration, VIG; funding acquisition, VIG. All authors have read and agreed to the published version of the manuscript.

Acknowledgments

This work was supported by St. Petersburg State University (project ID 94034084).

Conflict of interest

The authors declare no conflict of interest.

References

1. Schmider E, Ziegler M, Danay E, et al. Is it really robust? Reinvestigating the robustness of ANOVA against violations of the normal distribution assumption. *Methodology: European Journal of Research Methods for the Behavioral and Social Sciences* 2010; 6(4): 147–151. doi: 10.1027/1614-2241/a000016
2. Post-hoc testing. Available online: http://bayes.acs.unt.edu:8083/BayesContent/class/Jon/ISSS_SC/Module009/iss_m91_onewayanova/node7.html (accessed on 12 June 2023).
3. Grace-Martin K. The problem with using tests for statistical assumptions. Available online: <https://www.theanalysisfactor.com/the-problem-with-tests-for-statistical-assumptions/> (accessed on 12 June 2023).
4. Brown MB, Forsythe AB. Robust tests for the equality of variances. *Journal of the American Statistical Association* 1974; 69(346): 364–367. doi: 10.2307/2285659
5. Frost J. Using post hoc tests with ANOVA. Available online: <https://statisticsbyjim.com/anova/post-hoc-tests-anova/> (accessed on 12 June 2023).
6. Games PA, Keselman HJ, Clinch JJ. Tests for homogeneity of variance in factorial designs. *Psychological Bulletin* 1979; 86(5): 978–984. doi: 10.1037/0033-2909.86.5.978
7. Raychaudhuri S. Introduction to Monte Carlo simulation. In: Proceedings of the 2008 Winter Simulation Conference; 7–10 December 2008; Miami, FL, USA. pp. 91–100.
8. Cribbie RA, Fiksenbaum L, Keselman HJ, Wilcox RR. Effect of non-normality on test statistics for one-way independent groups designs. *British Journal of Mathematical and Statistical Psychology* 2012. 65(1): 56–73. doi: 10.1111/j.2044-8317.2011.02014.x

Appendix

Preliminary preparation is required for modeling.

- On each of the ranges of check sums, all the sets of buttons ever offered are considered. The five most interesting study sets are considered separately:
- For each of these five sets, we calculate the frequency probability of the proposal for this set. About the other sets, we assume that they are offered with equal probability.
- For each of these five sets, we find the probability of pressing a button from the set in them or choosing manual input. For the remaining sets, the total probability of pressing the button from the set and the probability of manual input are considered.
- For each of these five sets, we find the probabilities to click on each specific button out of four. For the other sets, in general, there is a probability of pressing the first button, the second button, etc.

The simulation process:

- 1) Possible sets of buttons, calculated probabilities of offering a set and probabilities of choosing one of these buttons or manual input, and a set of tips that were left manually are submitted to the input.
- 2) At the first step of the model, the selection of a set with specified probabilities is simulated.
- 3) At the second step of the model operation, the choice of pressing a button from a set or manual input with specified probabilities is simulated.
- 4) If manual input was selected, then randomly select the size of the tip left from the input set of tip values.
- 5) If the button was pressed, then, with a given probability, we would choose which of the four buttons was pressed.
- 6) For all values from the set of check amounts, we model the tip amount according to the procedure described above and determine what percentage of the check amount they make up. The result of the simulation is the average percentage of tips from the amount of the check.