STM JOURNALS

Journal of
Computer Technology & Applications

https://journals.stmjournals.com/jocta

ISSN: 2229-6964 (Online)
ISSN: 2347-7229 (Print)
Volume 15, Issue 3, 2024
September–December
DOI (Journal): 10.37591/JoCTA

Review JoCTA

User App Segmentation for Better Understanding of Reviews and Password Resets Using Regression

Parth Agrawal[1,*], Lishiv Sharma[2], Alaap Varma[3]

Abstract

In today's digital landscape, understanding users' needs and behaviors is key to improving app experiences. This paper focuses on how we can better understand user reviews and password resets in mobile apps through a method called user segmentation. By dividing users into groups based on their feedback and password reset patterns, we aim to uncover insights that can enhance app design and security. Using a mix of user reviews and password reset data from a diverse set of users, we analyze patterns using regression. This statistical method helps us identify different types of users and their preferences. Our findings reveal valuable insights into user sentiments and the likelihood of password resets. Ultimately, this research highlights the importance of tailoring app experiences to different user groups. By gaining insights into users' needs and behaviors, we can develop applications that are more user-friendly and secure. This can be utilized for attaining market insights for any penetration in comparison to other existing clients and competitors. The method is a way out by just reading out the app's data to form and display the required information. It also enables the institution to identify where they can target the audience they want to catch, if not they can amend changes to the same. Ensuring unbiased numeric values showcasing up reviews.

Keywords: User reviews, App segmentation, password resets, multinomial logistic regression, SPSS

INTRODUCTION

In the technological era, apps on mobile platforms are becoming the lifeline of our lives, putting convenience and connectivity within the reach of a single touch. However, progress will not only depend on their functionality but also on the way they cater to different needs and taste preferences. It is vital for app developers and businesses to build better user experiences, and app performance optimization is gaining a deeper understanding of users' behavior and preferences [1]. User segmentation, a generally known strategy from user experience design and marketing for grouping users into subsets with similar interests, habits, or mental models, provides a scheme for classifying users [2]. Segmenting users enables developers to focus their strategies in advance to meet the needs of different user groups, ultimately enhancing user satisfaction and engagement [3]. This study takes us into the field of user segmentation, specifically highlighting its application in the mobile application sector. Specifically, our goal is to examine in depth two of the app's key areas that influence engagement and performance of the application, including user segmentation in the case of user reviews and password resets, and derive from user satisfaction [4]. The most useful part of user reviews is that they are a very good source of user feedback for app developers, granting the knowledge of user sentiment, preference, and what the users find inconvenient [5].

Nevertheless, realizing and interpreting such a large number of reviews and the difference in their verdicts is trickier. Yes, password resets like these are very relevant for app security, but they still give

*__Author for Correspondence__
Parth Agrawal
E-mail: pa.8306408819@gmail.com

[1-3]Student, School of Technology, Management & Engineering, NMIMS University, Navi Mumbai, Maharashtra, India

Received Date: June 28, 2024
Accepted Date: August 01, 2024
Published Date: September 12, 2024

Citation: Parth Agrawal, Lishiv Sharma, Alaap Varma. User App Segmentation for Better Understanding of Reviews and Password Resets Using Regression. Journal of Computer Technology & Applications. 2024; 15(3): 1–9p.

us knowledge about user behavior and engagement [6]. Using logistic regression, a statistical method for the selection of model and segmenting model, the relational media aims to assign meaning to the review sentiments and password reset behaviors of the various segments of the users. Hence, by searching for trends and relationships present in the data, one can get to know and label specific groups of users with unique traits and common interests [7].

LITERATURE REVIEW

User segmentation is an inevitable part of marketing and data analysis tools that modern marketers need to build a platform for communicating customers' needs effectively [8]. In the present-day world, with enormous volumes of data and constantly progressing cutting-edge technology, researchers and practitioners remain curious about devising new methodologies and applications that can segment the audience into various spheres of life. Machine learning techniques, [9] as recent engines helping in the profound analysis of massive amounts of data and the identification of hidden associations and trends in it, are also good in user segmentation.

The report [10] focuses on consumer segmentation with the help of machine learning methods, such as k-means and hierarchical clustering. This indicates that these methods are efficient in separating common groups of consumers, which can be differentiated using various attributes. The following paper [11] investigates market segmentation approaches in e-commerce featuring supervised and unsupervised methodologies to demonstrate how intelligence from machine learning (ML) can be precise in identifying lifestyle indices, opinions, behaviors, and interests of targeted marketing subjects.

The utilization of such tools allows businesses to comprehend customers' preferences as well as the main patterns in their behavior and can additionally help companies create individual connections and personalized encounters [12], which will ultimately benefit customer satisfaction and loyalty. Psychographic and demographic segments had their right proportion, considering that they were the most popular in mobile fintech services.

An example, such as the work of Birtolo et al. (2013) [13], which examines the role of psychographics and demographics in customer profiling within the mobile banking sector, thereby providing insights into the complex inclinations and reactions of consumers [14].

By examining factors such as lifestyle, personality traits, and interests that coincide with demographic characteristics, marketers can develop a more comprehensive understanding of consumers' sensibilities and decision-making mechanisms, thus catering to a more apt fashion of marketing operations and products. Furthermore, user segmentations of online music services using fuzzy clustering techniques [15] imply meaningful information on user preferences and engagement patterns in a virtual music environment [16].

By gaining insight into user segmentation from the perspective of the targeted industry, companies can customize their strategies to match the application of specific users who are an audience of the business and the availability of the brand in the market [17]. Similarly, analysts such as [18] profoundly study subject-related segmentation and usage profiling, especially focusing on the relevance of regular segmentation attitude changes with respect to user dynamics for personalized treatment and engagement strategies [19]. Such a "just-in-the-moment" approach allows brands to explain these changes to their audience promptly and effectively, thus improving the effectiveness of their marketing efforts.

In addition, comprehensive surveys, such as [20], deliver the diversity of analytical tools used for customer segmentation for E-commerce personalized targeting, which emphasizes the dependence of tailor-made marketing strategies on clientele engagement and retention [21].

By implementing the power of the dynamic segmentation approach, businesses can maximize their marketing programs by providing more relevant content, [22] thereby tightening the bond with their audience so that they can grow their businesses to successful levels. Frontier sets of data analytics tools and methods that drive customer segmentation, powered by a sophisticated study of user actions and interaction patterns, facilitate segmentation.

For example, [23] highlighted the ability of marketing tools based on data analytics to comprehend and engage the target audience. Therefore, this study confirms the significance of data-informed decisions in marketing planning.

In addition to more precise profiling and advice to customers from business, machine learning, natural language processing, and predictive modeling are sophisticated analytics that can help businesses dig deep into big data to obtain insightful information [24].

The data-driven approach may assist companies in optimizing their marketing resources and achieving maximum return on investment (ROI), increasing the number of satisfied customers, and overcoming competition in the market—be motivational and increase users' experiences by stimulating them. Amplifying and highlighting the voices of activists, politicians, and healthcare workers are crucial for raising awareness and triggering actions.

The makeup retail business is set in many categories. These experts and professionals, through the utilization of the latest technologies and analytical platforms, can efficiently parse large and large datasets to provide a fertile ground for segmentation and target inclusion; hence, the desired and specific results are achieved in a short time [25]. Technology evolvement at a high speed and the growing users' expectations led to companies' segmentation methods constantly evolving to satisfy both users' needs and expectations.

Nevertheless, these organizations must ensure that the methods they apply are flexible and capable of integrating new techniques [26]. The application of the newest analytics, purposeful segmentation, as well as personalized interaction methodology, leads to the advantage of the companies since the latter can benefit from the shift in the market environment and create bonds with the audiences that lower their attrition rate and lead to loyalty that organically works towards the sustainable growth of the concerned company in attaining the best response [27].

RESEARCH METHODOLOGY

The methodology used to amend the data into the desired format of attaining insights revolves around considering all the factors, such as data type, data consistency, data manipulation values, and the introduction of false inputs to solve inconsistency. Regression methods were used to identify the necessary columns.

Binary Regression

We need to construct a binary logistic regression model that can forecast whether a user will post a review according to his/her behavior and personality. To achieve this, we used a dataset of 10,000 observations so that it could handle missing data (if it existed) and categorical variables that were identified using statistical encoding). Microdata, such as website clicks, purchase history, and length of time on a site, as well as non-numerical activities, such as site content consumption. Major attributes such as age, gender, IP address, and no. of purchases and time spent on the site are gone, and useless ones, such as user ID and location factors. The split into training (70%) and testing (30%) was performed. The default parameters of the logistic regression model were used during initialization. Training is performed on the training set, followed by the evaluation process using measures, including precision, accuracy, recall, and F1 score on the test data. Among the approaches to enhance sample performance, grid search, and random search are some of the tools that we use. The logistic regression model coefficients are obtained, and the consequences of each characteristic on the probability of the user accepting to submit a review are interpreted. In addition, there is an option for the model to forecast new data by monitoring and updating after a certain performance.

Multinomial Regression

Multinomial logistic regression models for predicting customer intentions in case of targeting their reviews contain attributes such as consumer age, gender, purchase history (frequent purchases, amount spent), and user behavior metrics (time spent on the website, number of interactions). These aspects can be considered in the model, and the interactions between them can be used to capture nonlinearities. Regularized methods that employ L1 or L2 are very useful for handling redundancy, but feature selection tools, such as recursive feature extraction or importance enhancement, can be used to reveal the top predictors. In addition, the method is being improved by adding analytical predictions derived not only from literature data to reviews, which could raise the accuracy rate of the model. Techniques to check the accuracy and generalization of the model with the help of cross-validation can be leveraged along with the contributed use cases, such as dependent variable significance and trend analysis, that increase the interpretability of the model.

COMPARATIVE ANALYSIS

Based on the applied regression techniques, further data are then used to compare with the other existing factors, which allows a deep study for considerate analysis. Presenting the findings as a declarative statement highlights the flaws and correct behavior of the applied model of analysis in Figure 1.

Rating Distribution

The ratings given were divided in the current scenario with a higher value in the 6–10 range, providing positive insight into the same.

Model Fit

- Likelihood ratio tests suggest that the overall model is statistically significant ($p<0.001$), indicating that the included predictors collectively contribute to predicting a higher value of ratings in the given solution (Figure 2).
- Goodness-of-fit tests (Pearson and Deviance) are significant, but the extremely high p-value for deviance suggests that the model fit is not significantly better than that of an intercept-only model. Further model refinement may be considered to improve fitness (Figure 3).

Pseudo R-Square

Pseudo R-Square values (Cox and Snell, Nagelkerke, and McFadden) provide insights into the proportion of variance explained by the model in Figure 4. A higher R-squared value suggests better explanatory power.

Case Processing Summary

		N	Marginal Percentage
Ratings	0	38	3.8%
	1	32	3.2%
	2	42	4.2%
	3	32	3.2%
	4	35	3.5%
	5	141	14.1%
	6	131	13.1%
	7	136	13.6%
	8	133	13.3%
	9	142	14.2%
	10	137	13.7%

Figure 1. Case processing summary.

ISSN: 2229-6964 (Online), ISSN: 2347-7229 (Print)

Model Fitting Information

Model	Model Fitting Criteria	Likelihood Ratio Tests		
	-2 Log Likelihood	Chi-Square	df	Sig.
Intercept Only	4451.330			
Final	3964.650	486.680	10	<.001

Figure 2. Model fitting information.

Goodness-of-Fit

	Chi-Square	df	Sig.
Pearson	12462.720	9970	<.001
Deviance	3964.650	9970	1.000

Figure 3. Goodness of Fitting test.

Pseudo R-Square

Cox and Snell	.386
Nagelkerke	.390
McFadden	.109

Figure 4. Pseudo R-Square test results.

Predictor's Contribution

- Each predictor's contribution is assessed through likelihood ratio tests.
- AverageScreentime is a significant predictor in the rating section.
- AverageSpentonApp is a significant predictor in the rating as well.
- LastVisitedMinutes is a significant predictor in the higher rating system.
- NewPasswordRequest is a non-significant predictor in the given higher rating section.

Odds Ratios and Coefficients

We can infer that all the effects are statistically significant at the 0.05 level, meaning that we can reject the null hypothesis and conclude that the effects are different from zero. For example, the chi-square statistics for the "Intercept" effect is 4033.086, the df is 68.436, and the Sig is less than 0.001. This means that there is strong evidence to reject the null hypothesis that the interception is equal to zero Figure 5.

For the given ratio test, our considered hypothesis tests are as follows:

- *H0:* This indicates that the values of average screen time, average spent on the app, and Last Visited Minutes do not affect the overall higher rating of any app in segmenting the user.
- *H1:* The alternate hypothesis states that it is dependent on the following statements: This revision clarifies that the alternative hypothesis is being presented and that it depends on certain statements. The word "dependent" was changed to "depends" for clearer phrasing.

Classification Table

The model's predictive accuracy is summarized in the classification Table 1.

With Scarcity as the rating number, the average prediction percentage for the models is given in Table 2.

Accuracy = (True positive+True negatives)/Total observations

Therefore, in this case, it is given by 247/4 = 61.75%

Likelihood Ratio Tests

Effect	Model Fitting Criteria -2 Log Likelihood of Reduced Model	Likelihood Ratio Tests Chi-Square	df	Sig.
Intercept	4033.086	68.436	10	<.001
AverageScreenTime * AverageSpentonAppINR * LastVisitedMinutes	4451.330	486.680	10	<.001

The chi-square statistic is the difference in -2 log-likelihoods between the final model and a reduced model. The reduced model is formed by omitting an effect from the final model. The null hypothesis is that all parameters of that effect are 0.

Figure 5. Likelihood ratio test results.

Table 1. Accuracy predictive analysis table.

Observed	0	1	2	3	4	5	6	7	8	9	10	Percent Correct
0	0	13	17	0	0	8	0	0	0	0	0	0.00%
1	0	18	11	0	0	3	0	0	0	0	0	56.30%
2	0	15	17	0	0	10	0	0	0	0	0	40.50%
3	0	15	7	0	0	10	0	0	0	0	0	0.00%
4	0	20	10	0	0	5	0	0	0	0	0	0.00%
5	0	15	17	0	0	65	0	0	0	39	5	46.10%
6	0	2	2	0	0	0	67	0	0	48	5	0.00%
7	0	2	9	0	0	0	67	0	0	52	6	0.00%
8	0	0	12	0	0	69	0	0	0	50	2	0.00%
9	0	0	9	0	0	72	0	0	0	55	6	38.70%
10	0	0	7	0	0	67	0	0	0	57	6	4.40%
Overall Percentage	0.00%	9.80%	11.80%	0.00%	0.00%	45.20%	0.00%	0.00%	0.00%	30.10%	3.00%	16.10%

Table 2. Error perception of different factors.

Ratings		B	Std. Error	Wald	df	Sig.	Exp(B)	Lower Bound	Upper Bound
0	Intercept	1.33	0.271	24.155	1	<.001	3.78	2.226	6.419
	AverageScreenTime * AverageSpentonAppINR * LastVisitedMinutes	0	0	28.916	1	<.001	1	1	1
1	Intercept	1.518	0.29	27.401	1	<.001	4.562	2.596	8.02
	AverageScreenTime * AverageSpentonAppINR * LastVisitedMinutes	0	0	23.112	1	<.001	1	1	1
2	Intercept	1.299	0.261	24.712	1	<.001	3.665	2.174	6.179
	AverageScreenTime * AverageSpentonAppINR * LastVisitedMinutes	0	0	31.815	1	<.001	1	1	1
3	Intercept	1.015	0.284	12.733	1	<.001	2.76	1.623	4.693
	AverageScreenTime * AverageSpentonAppINR * LastVisitedMinutes	0	0	24.835	1	<.001	1	1	1

4	Intercept	1.2	0.277	18.758	1	<.001	3.32	1.938	5.688
	AverageScreenTime * AverageSpentonAppINR * LastVisitedMinutes	0	0	26.918	1	<.001	1	1	1
5	Intercept	0.356	0.172	4.293	1	0.038	1.428	1.02	2
	AverageScreenTime * AverageSpentonAppINR * LastVisitedMinutes	0	0	6.902	1	0.009	1	1	1
6	Intercept	0.028	0.176	0.026	1	0.873	1.028	0.728	1.451
	AverageScreenTime * AverageSpentonAppINR * LastVisitedMinutes	0	0	0.329	1	0.566	1	1	1
7	Intercept	0.017	0.175	0.009	1	0.924	1.017	0.727	1.423
	AverageScreenTime * AverageSpentonAppINR * LastVisitedMinutes	0	0	0.036	1	0.85	1	1	1
8	Intercept	0.142	0.175	0.662	1	0.416	1.153	0.826	1.61
	AverageScreenTime * AverageSpentonAppINR * LastVisitedMinutes	0	0	1.858	1	0.173	1	1	1
9	Intercept	0.053	0.173	0.093	1	0.76	1.054	0.762	1.458
	AverageScreenTime * AverageSpentonAppINR * LastVisitedMinutes	0	0	0.019	1	0.892	1	1	1

CONCLUSION

Based on the results of the logical-level analysis, this study concludes that user segmentation plays an important role in understanding user behavior and preferences in a mobile application environment. By segmenting users based on review intent and password-setting behavior, a score of 61.75% indicates that the given relation between the ratings of the app and the app installed minutes varies within a given range of a median of five stars. We can identify unique user segments with unique characteristics and preferences. These features offer valuable insights for app developers and businesses that aim to enhance user experience and optimize app performance. By aligning strategies to meet the specific needs of different user groups, developers can improve user satisfaction and integration, ultimately leading to mobile app success in a highly competitive digital environment. Additionally, this study highlights the importance of using data-driven approaches, such as binary and multinomial regression, to extract insights from user feedback and behavioral data. Continuous exploration and improvement of user segmentation technology enables app developers to meet the evolving needs and desires of users, driving continuous improvement and innovation in the mobile app industry. Multinomial regression allows the parameters to be considered simultaneously for the calculative value considered.

Acknowledgment

A successful project of high-level technical, patience, and sustainability under the immense support of the guides is an attribute to teamwork and we would like to thank the following individuals: Our beloved mentor Prof. Aditya Kasar for his constant interest and involvement in our project, from the very conception of an idea, both theoretically and practically, which was instrumental in guidance to us in the Field of our Domain. To show the transition of our project idea from the perspective of the physical world, we express our gratitude to Prof. Aditya Kasar for the immense support, without which the idea would never have been enhanced into a useful project. Special thanks to our college SVKM's Narsee Monjee Institute of Management Studies (Deemed-to-be University), Navi Mumbai, India, for providing us with this opportunity to learn and grow.

REFERENCES

1. Lee Y, Cho S. User segmentation via interpretable user representation and relative similarity-based segmentation method. Multimedia Systems. 2021;27(10):61–72.
2. Roux M. A comparative study of divisive and agglomerative hierarchical clustering algorithms. J Classif. 2018;35:345–366. DOI: 10.1007/s00357-018-9259-9.
3. Chan CCH. Intelligent value-based customer segmentation method for campaign management: A case study of automobile retailer. Expert Syst Appl. 2008;34:2754–2762. DOI: 10.1016/j.eswa.2007.05.043.
4. Christy AJ, Umamakeswari A, Priyatharsini L, Neyaa A. RFM ranking – An effective approach to customer segmentation. J King Saud Univ Comput Inf Sci. 2021;33:1251–1257. DOI: 10.1016/j.jksuci.2018.09.004.
5. Chen X, Fang Y, Yang M, Nie F, Zhao Z, Huang JZ. Purtreeclust: A clustering algorithm for customer segmentation from massive customer transaction data. IEEE Trans Knowl Data Eng. 2018;30:559–572. DOI: 10.1109/TKDE.2017.2763620.
6. Böttcher M, Spott M, Nauck D, Kruse R. Mining changing customer segments in dynamic markets. Expert Syst Appl. 2009;36:155–164. DOI: 10.1016/j.eswa.2007.09.006.
7. Dixit S, Bhamare A, Darpel A. Customer segmentation using machine learning. 2021.
8. Bian J, Dong A, He X, Reddy S, Chang Y. User action interpretation for online content optimization. IEEE Trans Knowl Data Eng. 2013;25:2161–2174. DOI: 10.1109/TKDE.2012.130.
9. Banduni AM, Ilavendhan A. Customer segmentation using machine learning. Int J Innov Res Technol. 2020;7(2):116–22.
10. Ben Ayed A, Ben Halima M, Alimi AM. Survey on clustering methods: Towards fuzzy clustering for big data. In: Proceedings of the 6th International Conference of Soft Computing and Pattern Recognition (SoCPaR). 2014. p. 331–336. DOI: 10.1109/SOCPAR.2014.7008028.
11. Burri M, Schär R. The reform of the EU data protection framework: Outlining key changes and assessing their fitness for a data-driven economy. J Inf Policy. 2016;6:479–511. DOI: 10.5325/jinfopoli.6.2016.0479.
12. Koziel AM, Shen CW. Psychographic and demographic segmentation and customer profiling in mobile fintech services. Kybernetes. 2023. DOI: 10.1108/K-07-2023-1251.
13. Birtolo C, Diessa V, De Chiara D, Ritrovato P. Customer churn detection system: Identifying customers who wish to leave a merchant. In: Proceedings of the International Conference on Industrial Engineering and Other Applications of Applied Intelligent Systems. 2013. p. 411–420.
14. Tiwari R, Saxena MK, Mehendiratta P, Vatsa K, Srivastava S, Gera R. Market segmentation using supervised and unsupervised learning techniques for e-commerce applications. J Intell Fuzzy Syst. 2018;35:5353–5363. DOI: 10.3233/JIFS-169818.
15. Chalupa S, Petricek M. Understanding customer’s online booking intentions using hotel big data analysis. J Vacat Mark. 2024;30:110–122. DOI: 10.1177/13567667221122107.
16. Bezdek JC, Ehrlich R, Full W. FCM: The fuzzy c-means clustering algorithm. Comput Geosci. 1984;10:191–203. DOI: 10.1016/0098-3004(84)90020-7.
17. Barman D, Chowdhury N. A novel approach for the customer segmentation using clustering through self-organizing map. Int J Bus Anal. 2019;6:23–45. DOI: 10.4018/IJBAN.2019040102.
18. Golmes MA. A review on customer segmentation methods for personalized customer targeting in e-commerce use cases. 2023.
19. Chan CCH, Cheng CB, Hsien WC. Pricing and promotion strategies of an online shop based on customer segmentation and multiple objective decision making. Expert Syst Appl. 2011;38:14585–14591. DOI: 10.1016/j.eswa.2011.05.024.
20. Chang HC, Tsai HP. Group RFM analysis as a novel framework to discover better customer consumption behavior. Expert Syst Appl. 2011;38:14499–14513. DOI: 10.1016/j.eswa.2011.05.034.
21. Chan CCH, Hwang YR, Wu HC. Marketing segmentation using the particle swarm optimization algorithm: A case study. J Ambient Intell Human Comput. 2016;7:855–863. DOI: 10.1007/s12652-016-0389-9.

22. Stocchi L, Pourazad N, Michaelidou N, Tanusondjaja A, Harrigan P. Marketing research on mobile apps: Past, present and future. J Acad Mark Sci. 2022;50:195–225. DOI: 10.1007/s11747-021-00815-w. PubMed: 34776554.
23. Brito PQ, Soares C, Almeida S, Monte A, Byvoet M. Customer segmentation in a large database of an online customized fashion business. Robot Comput Integr Manuf. 2015;36:93–100. DOI: 10.1016/j.rcim.2014.12.014.
24. Kasem MS, Hamada M, Taj-Eddin I. Customer profiling, segmentation, and sales prediction using AI in direct marketing. Neural Comput Appl. 2024;36:4995–5005. DOI: 10.1007/s00521-023-09339-6.
25. Chen X, Sun W, Wang B, Li Z, Wang X, Ye Y. Spectral clustering of customer transaction data with a two-level subspace weighting method. IEEE Trans Cybern. 2019;49:3230–3241. DOI: 10.1109/TCYB.2018.2836804. PubMed: 29994344.
26. Pitka T, Bucko J. Segmenting customers with data analytics tools: Understanding and engaging target audiences. Acta Inform Pragensia. 2023;12:357–378. DOI: 10.18267/j.aip.220.
27. Bellini P, Palesi LAI, Nesi P, Pantaleo G. Multi clustering recommendation system for fashion retail. Multimed Tools Appl. 2023;82:9989–10016. DOI: 10.1007/s11042-021-11837-5. PubMed: 35043044.

