JOURNAL OF EMERGING TECHNOLOGY AND DIGITAL TRANSFORMATION

ONLINE ISSN
3006-9726

PRINT ISSN
3006-9718

VOLUME . 4 ISSUE . 3 (2025)

Customer Segmentation and the Effectiveness of Personalized Service Offers in the GSM Sector in Pakistan Using Clustering Techniques

Muhammad Talha Ahmed Farid
Department of Computer, Science SZABIST Karachi, Pakistan
Dr. Khalid Rasheed*
Department of Computer, Science SZABIST Karachi, Pakistan

Email: khalid.rasheed@szabist.edu.pk

Corresponding Author: Dr. Khalid Rasheed (Email: khalid.rasheed@szabist.edu.pk)

Received	Approved	Published
21th October 2025	12th June 2025	15th June 2025

Abstract:

In an intensely competitive GSM telecom industry, customer retention and provision of customized services are vital. This research employs K-Means clustering and Principal Component Analysis (PCA) to classify mobile subscribers in Pakistan according to their usage of voice, SMS, and data services. On the basis of anonymized customers' data from a leading telecom operator, trends were examined with regard to demographic and location variables. The clusters that were produced showed clear patterns of user behavior, which were matched with appropriate commercial propositions to enable personalized delivery of services. The findings emphasize the constraints on generalized offers and illustrate how segmentation using data can drive both customer satisfaction and business performance. The methodology that has been used can be applied by telecom companies in other emerging economies to enable focused marketing and product planning.

Keywords: *Customer Segmentation, GSM Sector, Clustering, K-Means, Principal Component Analysis (PCA), Telecom Analytics, Personalized Offers, Usage Behavior, Machine Learning, Pakistan*

ONLINE ISSN
3006-9726
PRINT ISSN
3006-9718
VOLUME . 4 ISSUE . 3 (2025)

I. Introduction

The telecommunications region has evolved into a essential enabler of virtual communication, particularly inside GSM networks that dominate worldwide connectivity. In markets like Pakistan, cell utilization is developing unexpectedly, however consumer behavior varies substantially throughout age, gender, vicinity, and affordability. Traditional consumer segmentation—based solely on plan kind or revenue—fails to seize this complexity, proscribing operators' potential to supply relevant, personalized offerings.

This observe addresses that hole by means of making use of unsupervised gadget gaining knowledge of strategies, specifically K-Means clustering and Principal Component Analysis (PCA), to segment customers based on real utilization conduct across voice, SMS, and mobile facts. Using anonymized subscriber information from a prime GSM operator, the studies explore behavioral styles throughout demographic and geographic dimensions. The primary goal is to identify meaningful customer clusters and endorse personalized telecom offers tailored to the usage wishes of every segment. By focusing on what customers devour—and how—they may be grouped into actionable clusters that tell centered campaigns, product layout, and pricing techniques.

The method gives a scalable framework for other emerging telecom markets looking for to enhance service effectiveness and purchaser pleasure. This conduct-pushed segmentation highlights how facts analytics can bridge the space among service transport and real user desires, in the long run using each consumer loyalty and enterprise overall performance.

II. LITERATURE REVIEW

The application of data analytics in the telecommunication industry has been a focus of substantial academic and business interest during the last two decades. Academics have researched different aspects of customer segmentation, with emphasis on discovering relevant behavioral patterns that can be used for strategic planning. The most dominant methodologies employed are clustering techniques, predictive modeling, and association rule mining, supplemented by dimensionality reduction methods like Principal Component Analysis (PCA). In one of the earliest works, Wedel and Kamakura (2000) highlighted the need for applying behavioral data to segment markets for improved customer targeting and retention strategies. Likewise, Xu and Walton (2005) illustrated the superiority of unsupervised learning techniques such as K-Means in applying to telecom user segmentation from usage patterns. Their research set the stage for cluster-based personalization, illustrating that behavioral clustering performed better than demographic-only models.

Later work has focused even more on the hybrid application of demographic and behavioral information. Zhang et al. (2018), for example, used K-Means clustering on mobile usage logs and segmented groups of users that strongly differed in terms of both service tastes and profitability. Their evidence suggested that refined segmentation would result in improved offer design and more efficient churn prevention. In Pakistan, with the GSM sector having a monopoly in the mobile segment, less academic research has been undertaken in spite of the massive and diversified subscriber base. Studies by the Pakistan Telecommunication Authority (PTA) indicate that though smartphone penetration and data usage have increased at an exponential rate, operators continue to depend on blanket-level offerings in place of personalized services.

PCA, the literature also recommends, assists in simplifying and visualizing high-dimensional telecom data sets in a manner that makes clusters more understandable. Jain and Dubes (2002) pointed out how PCA could enhance K- Means by eliminating feature redundancy as well as enhancing compactness of clusters—both being instrumental in customer behavior modeling. This research extends these findings by not only segmenting the user base through K-Means clustering and PCA but also associating these segments with actionable offer recommendations. The added geographic and demographic filtering component (age, gender, and tehsil) makes this research especially of value for emerging markets such as Pakistan, where regional variations can dramatically affect consumption behavior.

III. RESEARCH METHODOLOGY

This study takes a quantitative, data-intensive approach to investigate the ways in which clustering algorithms can be used to personalize telecom service offerings in Pakistan's GSM market. The research process is focused on the conversion of raw customer usage data into strategic information through machine learning-based segmentation and then onto service recommendation modeling.

A. Research Design

The research is designed based on the concepts of exploratory data analysis (EDA) and unsupervised machine learning. The method is observational and non-experimental in nature, utilizing already available anonymized data without immediate manipulation or intervention.

B. Data Source and Scope

The dataset used in this study was made available by one of the top GSM operators in Pakistan. It is comprised of anonymized records showing individual customer demographics (gender, age, and tehsil) and in-depth service usage metrics like on-net/off-net call durations, SMS volumes, and mobile data usage (social, off-peak, and generic data).

The scope of research is restricted to Karachi districts' customers only and aims at determining behavioral clusters to facilitate offer personalization.

C. Tools and Technologies

The analysis was performed with the help of Python utilizing the following libraries and tools:

- Data manipulation using Pandas and NumPy
- Visualizations using Matplotlib and Seaborn

- Clustering (K-Means), PCA, and data scaling using Scikit-learn

Geographic visualizations (heatmaps) with Folium Reporting results export to Excel and PowerPoint

D. Methodological Assumptions

The given data is a representative sample of the larger GSM subscriber base in Pakistan.

- Usage trends (voice, SMS, data) are taken to represent user choice and can be good clustering features.
- Demographic segmentation (age, gender, location) is taken to impact consumption behavior and provide responsiveness.

E. Validation Strategy

The number of clusters was confirmed with the Elbow Method to obtain the optimal segmentation level without overfitting. Furthermore, Principal Component Analysis (PCA) was utilized for visualization purposes and to ensure that clusters established in reduced-dimensional space-maintained interpretability.

F. Ethical Implications

Anonymized, non-identifiable customer information alone was utilized in this research. No personally identifiable information (PII) was viewed or analyzed. The study is consistent with ethical data handling and telecommunications data handling standards.

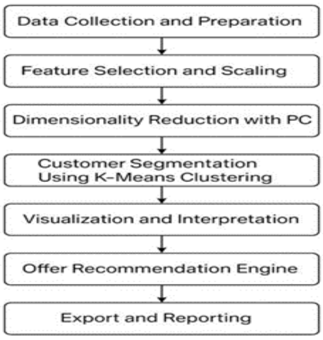

Data Collection and Preparation
Feature Selection and Scaling
Dimensionality Reduction with PC
Customer Segmentation Using K-Means Clustering
Visualization and Interpretation
Offer Recommendation Engine
Export and Reporting

Fig. 1. Analysis Methodology flowchart

IV. PROPOSED RESEARCH METHODLOGY

The analysis here adopts a rigorous data science process intended to derive useful patterns out of raw telecom usage data and actionable customer segments for offer recommendations for personalized services. The process combines several steps from data preprocessing to offer mapping and cluster validation, as outlined below:

A. Data Collection and Preparation

Anonymized telecom usage information was obtained from a Pakistan GSM operator. The data contained customer demographic information (age, gender, and tehsil) along with usage statistics like:

- On-net and off-net minutes
- Generic and social data volume
- Off-peak data usage
- SMS count
- Total minutes and GPRS volume

ONLINE ISSN
3006-9726

PRINT ISSN
3006-9718

VOLUME . 4 ISSUE . 3 (2025)

- Active users by service type (voice/data)

The data was cleaned to deal with missing values, normalize formats, and cast data types. Age values were bucketed into significant brackets (e.g., 18–26, 27–35), and numeric fields were summed up for uniformity.

B. Feature Selection and Scaling

Key usage features were chosen considering customer behavior relevance. These were call minutes, data usage, and SMS consumption. As these are differing scales, Standard Scaler was used to normalize them so that equal distances can be measured by clustering algorithms.

C. Dimensionality Reduction with PCA

For easier visualization and noise reduction, Principal Component Analysis (PCA) was used. PCA reduces the data into two principal components while maintaining variance and makes it simpler to interpret clusters as well as identify specific user patterns.

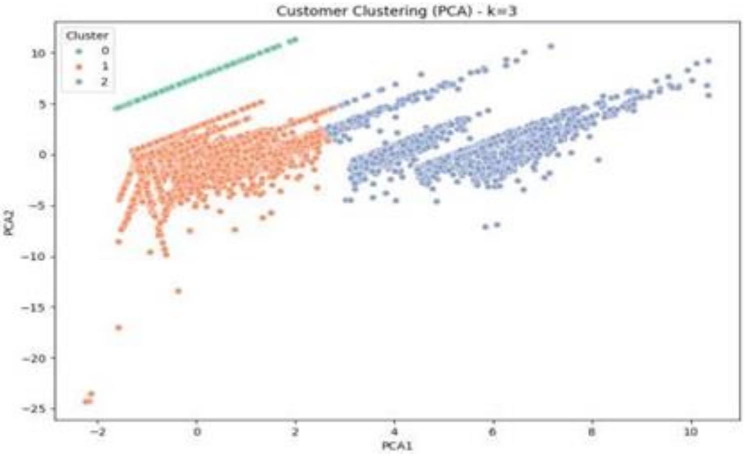


Fig. 2. PCA scatter plot showing user distribution in 2D space

D. Customer Segmentation Using K-Means Clustering

The K-Means algorithm was used to cluster the users based on their usage patterns. The best value of clusters (k=3) was chosen via the Elbow Method, which compares the within-cluster sum of squares (WCSS) for different values of k. A resulting cluster is a set of similar users based on their service usage pattern, e.g., high data users, heavy voice users, or low-usage subscribers.

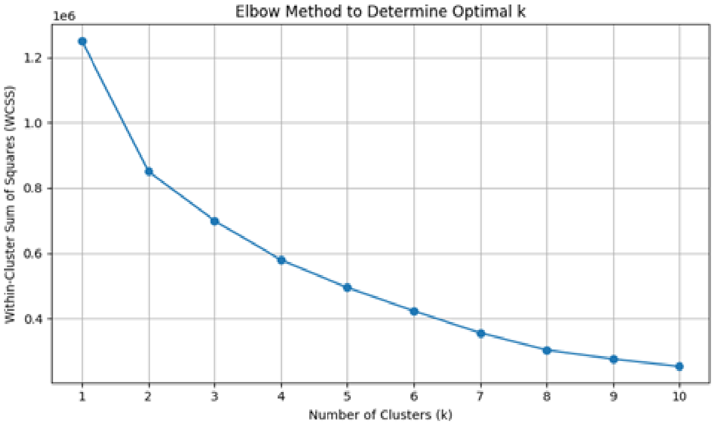


Fig. 3. Elbow Method graph showing WCSS against different values of k.

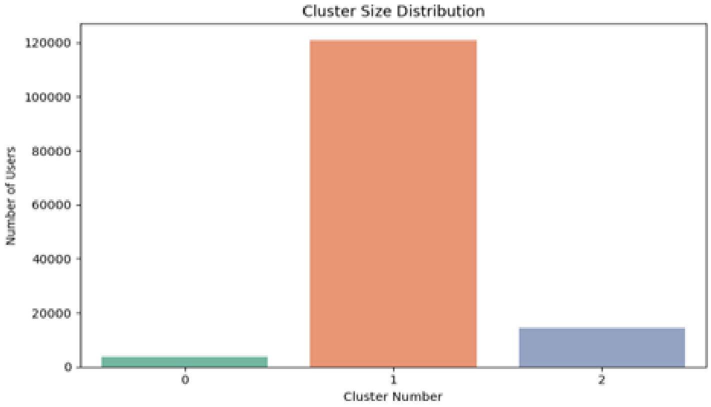


Fig. 4. Cluster summary bar chart showing size of each cluster.

E. Visualization and Interpretation

Cluster assignments were represented with PCA scatter plots, providing easy visual understanding of customer distribution. Heatmaps and bar graphs were similarly constructed to provide geographic visualization of service consumption as well as by offer type.

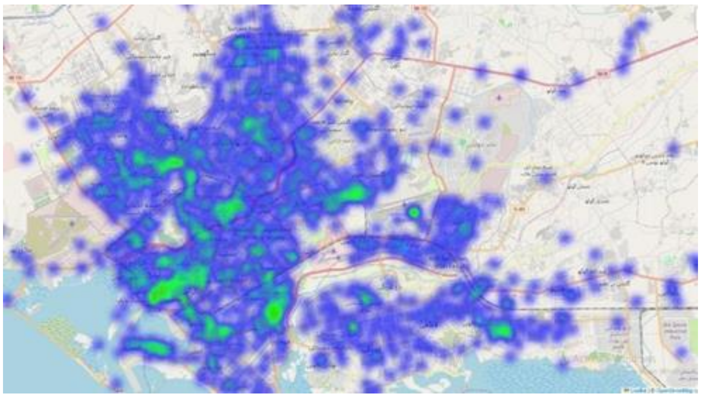


Fig. 5. Heatmap Data Usage by Location (Green scale) & Voice Minutes by Location (Blue scale).

F. Offer Recommendation Engine

With the help of the cluster data, usage patterns were examined at the intersection of gender, age group, and tehsil. Within each segment, the best-fitting offers were presented based on representative usage within the cluster. Offers were not copied blindly; rather, proposed resources (e.g., minutes, SMS, data) were re-weighted based on cluster-level usage patterns rather than merely duplicating previous purchases.

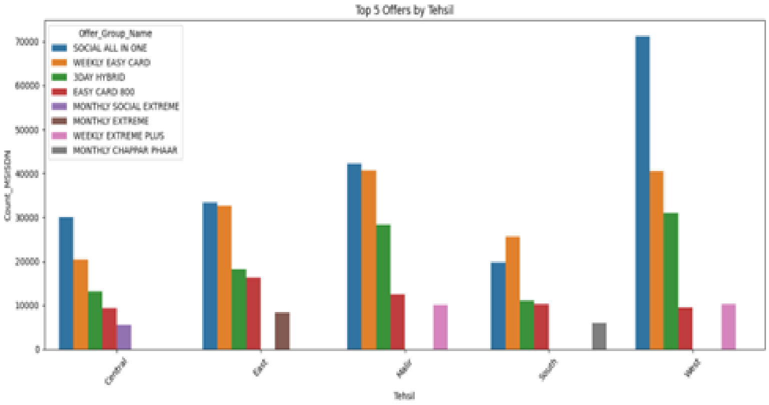
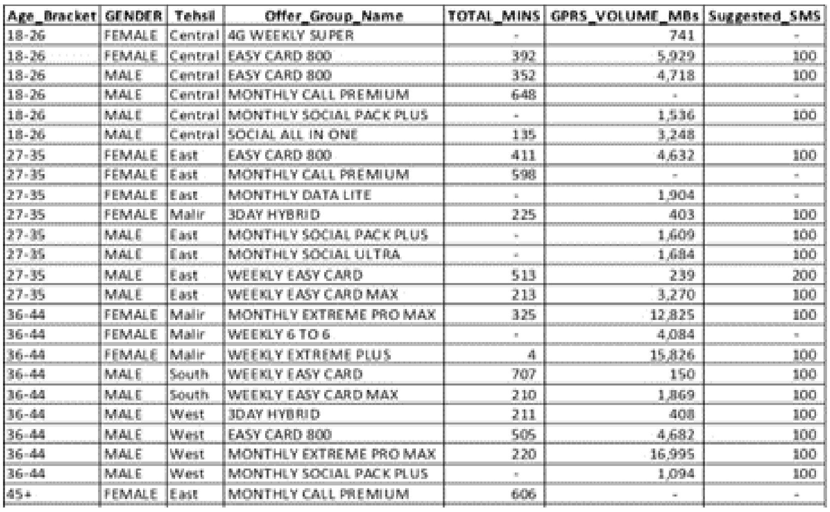


Fig. 6. Bar chart of Top 5 Offers by Tehsil.

G. Export and Reporting

Complete list of recommended offers with PCA and cluster analysis information

- Top 1 and top 3 offer recommendations by age, gender, and location
- Rounded usage estimates for planning operations
- Visual artifacts including PCA charts, elbow plots, bar graphs, and geographic heatmaps



Age_Bracket	GENDER	Tehsil	Offer_Group_Name	TOTAL_MINS	GPRS_VOLUME_MBs	Suggested_SMS
18-26	FEMALE	Central	4G WEEKLY SUPER	-	741	-
18-26	FEMALE	Central	EASY CARD 800	392	5,929	100
18-26	MALE	Central	EASY CARD 800	352	4,718	100
18-26	MALE	Central	MONTHLY CALL PREMIUM	648	-	-
18-26	MALE	Central	MONTHLY SOCIAL PACK PLUS	-	1,536	100
18-26	MALE	Central	SOCIAL ALL IN ONE	135	3,248	
27-35	FEMALE	East	EASY CARD 800	411	4,632	100
27-35	FEMALE	East	MONTHLY CALL PREMIUM	598	-	-
27-35	FEMALE	East	MONTHLY DATA LITE	-	1,904	-
27-35	FEMALE	Malir	3DAY HYBRID	225	403	100
27-35	MALE	East	MONTHLY SOCIAL PACK PLUS	-	1,609	100
27-35	MALE	East	MONTHLY SOCIAL ULTRA	-	1,684	100
27-35	MALE	East	WEEKLY EASY CARD	513	239	200
27-35	MALE	East	WEEKLY EASY CARD MAX	213	3,270	100
36-44	FEMALE	Malir	MONTHLY EXTREME PRO MAX	325	12,825	100
36-44	FEMALE	Malir	WEEKLY 6 TO 6	-	4,084	-
36-44	FEMALE	Malir	WEEKLY EXTREME PLUS	4	15,826	100
36-44	MALE	South	WEEKLY EASY CARD	707	150	100
36-44	MALE	South	WEEKLY EASY CARD MAX	210	1,869	100
36-44	MALE	West	3DAY HYBRID	211	408	100
36-44	MALE	West	EASY CARD 800	505	4,682	100
36-44	MALE	West	MONTHLY EXTREME PRO MAX	220	16,995	100
36-44	MALE	West	MONTHLY SOCIAL PACK PLUS	-	1,094	100
45+	FEMALE	East	MONTHLY CALL PREMIUM	606	-	-

Fig. 7. Sample of Suggested Offers.

V. DISCUSSION OF THE RESARCH RESULTS

This section addresses the research findings in the context of the initial objectives, extracts strategic implications for stakeholders. Additionally, contrasts the findings with existing work, and defines the study's limitations.

A. Alignment with Research Objectives

This study aimed to segment GSM clients based on carrier usage and suggest centered offers. The utility of K-Means clustering and PCA efficaciously grouped users into wonderful behavior-based clusters, such as excessive- facts/low-voice users, voice-heavy customers, and mild- carrier users. These clusters, while mixed with demographic (age, gender) and geographic (tehsil) records, enabled specific offer recommendations—accomplishing the research goal of actionable segmentation.

B. Strategic Implications

The findings have strong implications for telecom operators. Unlike widely wide-spread nationwide gives, cluster-based pointers spotlight clean behavioral differences. Younger users in urban regions desire social records bundles, while middle-aged customers display hybrid utilization. Dormant customers may be re-engaged with low-price offers. This focused technique supports data-driven marketing, nearby marketing campaign design, and product optimization, moving providers toward customized carrier transport.

C. Comparison with Previous Studies

Previous studies, such as that via Xu & Walton (2005) and Zhang et al. (2018), helps usage-based totally clustering over demographics by myself. This study extends those findings in a Pakistani context by means of such as geographic segmentation and linking clusters directly to industrial offers—bridging the space among evaluation and implementation.

D. Study Limitation

Although the findings are encouraging, the study also has some limitations:

1) **Data Scope and Duration:** The research was conducted on one snapshot of data. Seasonal trends or customer lifecycle (e.g., onboarding, churn, reactivation) longitudinal data would provide healthier insights
2) **No Churn Prediction:** The existing model is concentrated on segmentation and recommendation but lacks churn probability, an important KPI for retention-driven strategies.
3) **Static Offer Matching:** While usage patterns were employed to make inference-based recommended offers, actual real-time offer optimization or A/B testing for conversion rate was not performed.
4) **No Real-Time Network Metrics:** Quality of service metrics such as dropped calls, data rate, or signal quality were not provided, which could impact customer desire and offer take-up.
5) **Anonymization Constraints:** Because of data privacy needs, some user-level information that could contribute towards improved targeting (e.g., handset, billing) was not available.

In spite of such constraints, the research forms a worthy point of departure for telecom operators who would like to include machine learning in their customer interaction models.

VI. CONCLUSION

This research focused on solving a core issue for the Pakistani GSM telecommunication market—how to segment mobile customers correctly and then propose suitable service offerings based on data-driven methodologies. Employing unsupervised learning techniques such as K-Means clustering along with Principal Component Analysis (PCA), this research could segment significant customer usage behaviors into meaningful segments as per voice, SMS, and data usage practices. These clusters were further divided by gender, age group, and tehsil, enabling in-depth analysis of demographic and geographical variation. The results categorically demonstrate that users don't consume services in a similar fashion. There were certain segments with heavy mobile data usage but zero voice or SMS usage, and others were balanced or voice-centric. These differences in behavior would have been ignored by generic, non-personalized marketing. The synergy of a cluster-based recommendation engine enabled more tailored offers—offering plans closely matching the needs of each segment. Plots such as PCA plots, elbow plots, bar plots, and usage heatmaps made cluster behavior easier to interpret, enabled geographic contextualization, and improved results communication to non-technical audiences. In addition, the final suggested offer datasets, which include rounded estimates of resources and segment-wise information, offer a ready-to-be-implemented output easily utilized by marketing teams or operations planners in telecommunication firms. This research not only confirms the applicability of clustering techniques to telecom analytics but also presents an extendible model for other emerging economies with demographic heterogeneity and low ARPU (Average Revenue Per User) where personalization is a business necessity.

VII. FUTURE WORK

While this study affords a whole framework for client segmentation and provide personalization the use of clustering strategies, several areas offer opportunities for enhancement:

1) Temporal Behavior Analysis: Using time-series records might allow lifecycle segmentation (onboarding, increase, or dormant users) and help dynamic focused on techniques.
2) Real-Time Deployment: Embedding the model in CRM structures for live segmentation using person hobby and recharge statistics would enable adaptive, scalable personalization.
3) Feedback Loop Mechanism: Incorporating user reaction facts (offer reputation, recharges, opt-outs) can enhance version accuracy and personalization thru reinforcement learning.
4) Expanded Feature Set: Adding data which includes handset type, recharge behavior, and complaint history could enrich segmentation and boom business applicability.
5) Algorithmic Comparison: Testing superior clustering techniques like DBSCAN, GMM, or hierarchical clustering ought to yield higher-becoming segments.
6) Enhanced Demographics: Including psychographic data like profits or life-

style, wherein permissible, would allow deeper behavioral insights.

7) Network-Aware Segmentation: Integrating geographic and tower-level network facts might guide region-primarily based optimization and community planning.

8) Churn Prediction Integration: Merging segmentation with churn fashions can help target high-threat users with retention gives, boosting purchaser lifetime price.

Together, those guidelines can raise the model into a scalable employer-grade machine tailored for rising telecom markets.

TABLE 1 CLUSTER-BASED SUMMARY OF GSM USER BEHAVIOR AND PERSONALIZED OFFERS

Cluster Profiles, Usage Patterns, and Offer Recommendations			
Cluster Type	**Usage Summary**	**Demographics**	**Offer Recommendation**
Cluster 1: Heavy Data Users	High Data (avg. > 4 GB), Low Voice, Moderate SMS	Age 18–26, Urban, Male Dominant	Social Data Pack, Weekly Ultra
Cluster 2: Voice- Centric Users	High Voice (>1000 min), Low Data, Low SMS	Age 27–44, Mixed Gender, Semi- urban	Hybrid Voice+ 1GB Bundle
Cluster 3: Low Usage Segment	Low across all metrics	Age 35+, Rural, Female-leaning	Lite Pack, Re- engagement Offer

REFERENCES

[1] Gao, G., & Li, M. (2018). Telecom Customer Segmentation Based on Principal Component Analysis and K-means Clustering. 4th International Conference on Science and Social Research (ICSSR), 1– 4.

[2] Hao, Z., & Jiang, H. (2016). Telecom Customer Segmentation Based on Principal Component Analysis and K-means Clustering Algorithm. Journal of Convergence Information Technology, 11(11), 194–201.

[3] Cai, Q., Luo, Y., Xi, H., & Zhu, G. P. (2012). Telecom Customer Segmentation Based on Cluster Analysis. International Conference on Computer Science and Information Processing (CSIP).

[4] Qamar, A. M. (2013). Customer Segmentation and Analysis of a Mobile Telecommunication Company of Pakistan using Two Phase Clustering Algorithm. Eighth International Conference on Digital Information Management (ICDIM).

[5] Reza, M. M. (2018). Segmentation of Mobile Customers using Data Mining Techniques. IJERT.

[6] Abdulhafedh, A. (2021). Incorporating K-Means, Hierarchical Clustering and PCA in Customer Segmentation. Journal of City and Development.

[7] Wedel, M., & Kamakura, W. A. (2000). Market Segmentation: Conceptual and Methodological Foundations. Springer. (Foundational theoretical reference on behavioral segmentation.)

[8] Xu, R., & Walton, J. (2005). Use of Unsupervised Learning in Telecom User Segmentation. (Demonstrated PCA + K-Means usage in telecom.)

[9] Liu, Y., & Zhu, H. (2015). Telecom Customer Segmentation Based on PCA and K-means Clustering Analysis. International Conference on Intelligent Transportation, Big Data & Smart City, 283–285.

[10] Z. Sun & Hu, Y. (2014). Telecom Customer Segmentation Based on PCA and K-means Clustering. Sixth International Symposium on Parallel Architectures, Algorithms and Programming, 282–286.

[11] Tang, X., Cheng, C. (2021). Research and Application of Precision Marketing Algorithms for Telecom Credit Data, LN in Electrical Engineering.

[12] Phua, C., Cao, H., Gomes, J. B., & Nguyen, M. (2012). Predicting Near-Future Churners and Win-Backs in the Telecommunications Industry. arXiv.

[13] Pothuvaichit's et al. (2020). A comparative dimensionality reduction study in telecom customer segmentation using deep learning and PCA.Journal of Big Data

[14] Masood, Qamar, et al. (2013). Customer segmentation and analysis of a mobile telecommunication company of Pakistan using two-phase clustering algorithm. ICDIM

[15] Masood et al. (2013). Comparison of two step clustering and K-Means for Pakistani telecom data. DLINE Journal

[16] AlKhairyat et al. (2020). Dimensionality reduction + Autoencoder vs PCA for telecom clustering. Journal of Big Data

[17] Lewlisa Saha et al. (2022). Machine learning model for personalized tariff plan in telecom. IJACSA

[18] Ufeli, Sattar, Hasan, Mahmood (2025). FAMD with clustering for customer segmentation. Information (MDPI)

[19] Embalo (2018). Clustering vs SOM for customer segmentation among mobile providers. University of Nairobi

[20] Qamar et al. (2013). Two-phase clustering including RFM model in Pakistan telecom. NUST Thesis

[21] Leng et al. (2015). Behavior modeling & clustering using topic models in wireless network users. arXiv

[22] Embalo Calvins (2018). Customer behavior segmentation with K-Means vs SOM. UoN Research

[23] Ikram et al. (2023). Optimized deployment of telecom field teams via K-Means in Pakistan. JDSS

[24] Upadhyay (2016). Customer profiling and segmentation using data mining. Journals

[25] Arumawadu et al. (2015). Mining telecom customer profitability via K-Means. JDAIP

[26] Manzano et al. (2020). Benefits and limitations of K-Means in mobile segmentation. rechargement analysis conferences

[27] Mulyawan et al. (2019). Clustering approach in churn prediction systems. IJIKM