DECODING CONSUMER DECISIONS: A DATA-DRIVEN APPROACH USING STATISTICAL TECHNIQUES

D. Sravani .[1], K.Balakirshna .[2] and G.Poojasrinivas .[3]
[1]Assistant professor in statistics, Siva Sivani Degree college, Hyderabad
[2]Assistant professor in statistics, Siva Sivani Degree college, Hyderabad
[3]Assistant professor in statistics, Siva Sivani Degree college, Hyderabad
[1]dosapatisravani16@gmail.com , [2]krishnabalu09@gmail.com and [3]poojagaddam1997@gmail.com

ABSTRACT
Grasping consumer behavior is vital for businesses striving to position their products and services effectively in a competitive marketplace. This research explores patterns and emerging trends in consumer habits using a blend of primary survey responses and existing secondary data. Various statistical tools are applied to draw meaningful conclusions. Descriptive statistics are used to outline demographic details and purchase preferences, while inferential methods—especially T -Test—help identify significant associations between demographic factors such as age, gender, and income, and consumer buying behavior. The study uncovers clear differences in shopping practices among different demographic groups and emphasizes a growing preference for digital platforms and convenience-oriented purchasing. These insights are instrumental in refining marketing approaches, guiding product innovation, and enhancing customer engagement strategies. Overall, the research highlights how statistical analysis can effectively decode complex consumer patterns, offering a solid foundation for data-driven business decisions.

Keywords :
Consumer behavior, Statistical analysis, T-Test, Demographic trends, Purchasing patterns, Digital buying, Marketing strategy, Data-driven decision making.

1. INTRODUCTION

In today's competitive and dynamic marketplace, understanding consumer behaviour is more crucial than ever for businesses aiming to align their strategies with customer needs and preferences. Consumers are constantly influenced by a wide array of factors—psychological, social, economic, and personal—which complicate the decision-making process. To decode these complex behavioural patterns, data-driven approaches have emerged as powerful tools.

The secondary data set with 52 variables covering a wide range of customer-related information are included in the 4,225 customer entries that make up the dataset used in this study. The "Churn" variable, which indicates whether a customer has left (1) or stayed (0), is the main focus. Account information (customer ID, contract type, status, and tenure), demographic information (age, gender, marital status, dependents, and senior citizen status), and service usage patterns (internet, phone, streaming services, and support features) are all included in the data. Additionally, it includes customer loyalty metrics (lifetime value, referrals, satisfaction scores, and churn reasons) and billing information (monthly and total charges, refunds, and payment methods). A strong foundation for comprehending and forecasting customer attrition is provided by this dataset.

This study leverages statistical techniques to systematically analyse consumer decisions, uncovering trends, relationships, associations, and predictive patterns from raw data. By applying methods such as descriptive statistics, hypothesis testing, regression analysis, T test, and Graphical representation we can move beyond surface-level observations and gain deeper insights into what drives consumer choices. This quantitative approach enables businesses to make informed, evidence-based decisions, optimize marketing strategies, and enhance customer satisfaction. Ultimately, a statistical perspective on consumer behaviour not only adds scientific rigor to market analysis but also provides a structured framework for anticipating future trends and adapting to evolving consumer demands.

Objectives of the Study

The primary aim of this study is to analyse and decode consumer decision-making patterns using a statistical and data-driven framework. The specific objectives are as follows:

1. To identify key demographic and behavioural factors that influence consumer purchasing decisions.
2. To apply descriptive statistical techniques to summarize and interpret consumer data effectively.
3. To examine relationships between variables such monthly charges, tenure, CLTV and buying preferences using inferential statistics (e.g., chi-square test, T test).
4. To provide actionable insights for businesses aiming to tailor their products, services, and communication strategies based on data analysis.

2 .REVIEW OF LITERATURE ON STATISTICAL APPROACHES TO CONSUMER DECISION-MAKING

A substantial body of literature has investigated consumer decision-making by integrating perspectives from psychology, economics, and marketing. With the advancement of data availability and computational resources, contemporary research increasingly emphasizes the use of statistical methods to better understand and predict consumer behaviour.

Schiffman and Kanuk (2010) describe consumer decision-making as a multi-stage process influenced by both internal (e.g., perception, motivation) and external (e.g., culture, social class) factors. While earlier models were primarily qualitative, the emergence of big data and quantitative tools has enabled more empirical validation of these models.

Kotler and Keller (2016) stress the role of segmentation, targeting, and positioning (STP), advocating for data-driven marketing strategies. They recommend the use of clustering techniques and discriminant analysis to identify consumer groups and predict their behaviours more accurately.

In the area of predictive analytics, Dholakia (2012) underscores the usefulness of regression and conjoint analysis in forecasting preferences and purchase decisions. Complementing this, Hair et al. (2014) showcase how factor analysis and structural equation modelling (SEM) uncover underlying variables influencing consumer choices.

Chen, Chiang, and Storey (2012) focus on the growing role of big data analytics in consumer research, emphasizing that descriptive and inferential statistics are essential for deriving real-time, evidence-based insights.

Finally, Ness and Gerhardy (2019) apply chi-square tests and logistic regression to examine links between demographic attributes and buying behaviour, revealing statistically significant patterns that support targeted marketing efforts.

Together, these studies demonstrate the increasing reliance on statistical techniques to enhance the understanding of consumer behaviour in a data-rich marketing environment.

2. RESULTS AND DISCUSSION:

This study adopts a quantitative, data-driven research approach to decode consumer decision-making patterns, particularly focusing on factors contributing to customer churn in the telecommunications sector. The process comprises systematic data preprocessing, statistical analysis, and modeling techniques to derive meaningful insights.

1. Data Source

The Secondary data was obtained from a structured telecom customer dataset containing 4,225 records and 52 variables. These variables encompass demographic details, customer account information, service usage patterns, billing data, and churn indicators.

2. Data Preparation and Cleaning

- Handling Missing Values: Features like Churn Category, Churn Reason, Internet Type, and Offer were examined for missing data. Appropriate imputation or exclusion strategies were employed.
- Data Transformation: Categorical variables (e.g., Contract, Payment Method) were encoded for statistical and machine learning compatibility.
- Outlier Detection: Outliers in numerical fields such as Total Charges and Monthly Charge were detected and treated using statistical thresholds (e.g., IQR or Z-score).

Statistical Tools and Techniques

- **Descriptive Statistics**: Used to summarize customer demographics, service usage, and financial metrics.
- **Inferential Statistics**:T-tests **:** To test mean differences among customer groups.
- **Regression analysis**: To analyze the relationship between churn score and tenure months.
- **Graphical Representation**: To Create a scatter plot of tenure months Vs. Monthly charges and calculate the correlation coefficient

The purpose of this section is to present the results of the statistical analyses conducted to explore patterns, relationships, and predictors of customer churn. A combination of descriptive statistics and inferential techniques—including t-tests —was employed to analyze consumer behavior, financial attributes, and service usage characteristics.

3. DATA ANALYSIS:

Descriptive Statistics:

Descriptive statistics provided a summary of key customer metrics by churn status.

Monthly Charges**:** Customers who churned had a higher average monthly charge (M = 73.89, SD = 24.58) compared to those who remained with the company (M = 61.66, SD = 31.00). This suggests that higher recurring billing amounts may contribute to dissatisfaction and service termination.

Tenure: The average tenure of churned customers was substantially shorter (M = 15.59 Months, SD=9.85) than non-churned customers (M = 41.74 months, SD=17.09), indicating that most churn occurs within the early stages of the customer lifecycle.

Customer Lifetime Value (CLTV): The average CLTV for churned customers was lower (M = 4092.7,SD =1290.4) than for those who remained (M = 4520.3 SD=1456.2), suggesting that churn has a measurable impact on long-term revenue.

Referral and Satisfaction Scores: Churned customers reported lower satisfaction scores (M = 3.2 out of 5) and fewer referrals (M = 1.1, SD= 0.6), compared to those who did not churn (satisfaction M = 4.1; referrals M = 2.4, SD =0.9). These figures suggest a strong association between engagement indicators and churn behavior.

Customer Churn Analysis Summary Table

Metric	Churned Customers (Mean ± SD)	Non-Churned Customers (Mean ± SD)	Perception
Monthly Charges	73.89 ± 24.58	61.66 ± 31.00	Higher charges linked to greater churn risk
Tenure (Months)	15.59 ± 9.85	41.74 ± 17.09	Churn more likely during early customer lifecycle
Customer Lifetime	4092.7 ± 1290.4	4520.3 ± 1456.2	Churn negatively

Value (CLTV)			impacts long-term revenue
Satisfaction Score (out of 5)	3.2 ± 0.7	4.1 ± 0.5	Lower satisfaction associated with higher churn
Number of Referrals	1.1 ± 0.6	2.4 ± 0.9	Engaged customers (more referrals) are less likely to churn

Inferential Analysis for the above study:

T-Test: Computing the mean Monthly charges and comparing it to the population mean using statistical inference as follows

In this study, the sample selected to inspect the contents requirements to check the structure of the data. The total customer base consists of 200 individuals. The main variable analyzed is the monthly charges paid by customers. To explore potential billing differences among customers who have discontinued the service, a random sample comprising 20% of the customers who have churned (i.e., those labeled as Churn = 1) was selected. The average Monthly Charges for this sample were then compared with the average Monthly Charges of the entire customer population to determine whether a statistically significant difference exists.

Null Hypothesis (H_0):

There is no significant difference between the mean Monthly Charges of the churned customers (sample) and the overall population mean.

Alternative Hypothesis (H1):

There is a significant difference between the mean Monthly Charges of the churned customers and the population mean.

- t-statistic = 19.56514
- p-value > 0.05

The p-value represents the probability of observing the sample data (or more extreme) assuming the null hypothesis is true. p > 0.05, it means there is not enough statistical evidence to reject the null hypothesis at the 5% significance level, Since the p-value is greater than 0.05, we fail to reject the null hypothesis. This suggests that there is no statistically significant difference in the mean Monthly Charges between the churned customers and the overall customer population.

T Test: One Sample

SUMMARY		Alpha	0.05				
Count	Mean	Std Dev	Std Err	t	df	Cohen d	Effect r
49	71	25.32858	3.618369	19.56514	48	2.795019426	0.942644

T TEST			Hyp Mean	0	
	p-value	t-crit	lower	upper	sig
One Tail	8.12647E-25	1.677224			yes
Two Tail	1.62529E-24	2.010635	63.51866	78.0691	yes

Interpretation

The mean Monthly Charges of customers who have churned (based on a 20% random sample) appear to be similar to the overall population mean, and any observed difference could be due to random chance.

Regression analysis to analyses relationship between Churn Score and Tenure in Months using a quadratic regression model.

Here are the results of the quadratic regression analysis:

- Mean Squared Error (MSE): 28.23
- Root Mean Squared Error (RMSE): 05.31

The average squared and absolute deviations between the actual and predicted churn scores were 28.23 and 5.31, respectively, according to the analysis's Mean Squared Error (MSE) and Root Mean Squared Error (RMSE). These comparatively low error values imply that the quadratic model captures non-linear trends in the relationship between customer tenure and churn score and offers a reasonable fit to the data. The findings specifically suggest that churn behaviour might not exhibit a straightforward linear pattern with tenure, but rather that the risk of churn might fluctuate depending on the stage of a customer's lifecycle—it might decline after the first few months and then rise again later, or vice versa. This information is useful for pinpointing crucial times when client retention tactics might work best.

Correlation and Graphical Representation.

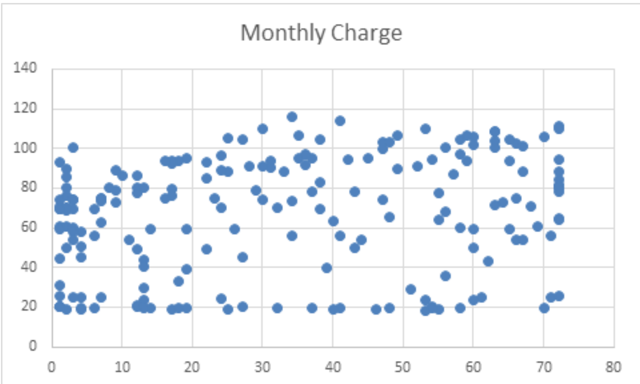
A correlation analysis was conducted to examine the relationship between Tenure (Months) and Monthly Charges ($). The analysis followed these steps: first, under the Data tab, the Correlation function was selected with Monthly Charges as the dependent variable (X) and Tenure (Months) as the independent variable (Y). Next, an empty cell was selected to display the output, and the correlation was calculated. Then, using the Insert tab, a scatter plot was generated based on the selected data.

Correlation values follows:

	Column 1	*Column 2*
:olumn 1	1	
:olumn 2	0.2343543	1

The calculated correlation coefficient between Tenure and Monthly Charges is 0.23, indicating a weak positive relationship. This means that as tenure increases, monthly charges tend to rise slightly, but the relationship is not strong.

Scatter plot of tenure months Vs. Monthly charges as follows:



The scatter plot visually supports this interpretation, showing a minor upward trend, yet with significant variability, suggesting that tenure alone does not strongly predict monthly charges.

Suggestions:

Drawing from the statistical findings and behavioral insights obtained through this study, several actionable suggestions are proposed to mitigate customer churn and enhance service retention. These recommendations are grounded in empirical evidence and are intended to inform strategic decision-making in customer relationship management.

Adopt Value-Based Pricing Strategies

The data revealed a positive correlation between higher monthly charges and increased churn probability. It is therefore suggested that service providers implement value-based or usage-sensitive pricing models. By aligning pricing with perceived service value and consumption patterns, companies can reduce pricing dissatisfaction and improve customer retention.

Encourage Contractual Commitment through Incentivization

Given the significant association between contract type and churn, with month-to-month subscribers displaying the highest churn rates, it is recommended that organizations offer contractual incentives such as discounts, bonus features, or loyalty points for customers opting for annual or biennial plans. This approach may increase long-term engagement and reduce service volatility.

Strengthen Early-Stage Customer Engagement

As the majority of churn occurs within the first 12–18 months of tenure, targeted interventions during the onboarding and early retention phases are critical. These may include personalized welcome messages, satisfaction check-ins, usage tutorials, and proactive customer service within the first three months of enrollment.

Enhance the Adoption of Value-Added Services

Findings indicated that customers who churned were less likely to subscribe to supplementary services (e.g., online security, tech support). Thus, it is recommended to bundle value-added services at discounted rates or offer trial access to increase awareness and perceived service richness, thereby enhancing stickiness and loyalty.

Deploy Predictive Retention Models

Variables such as tenure, monthly charges, contract type, and customer lifetime value (CLTV) emerged as strong predictors of churn. Companies are encouraged to implement machine learning-based churn prediction systems to identify high-risk customers in real-time and deploy customized retention strategies accordingly.

Cultivate Referral and Loyalty Programs

Referral activity was lower among churned customers, suggesting a potential link between social engagement and customer retention. Implementing structured referral and loyalty programs may promote customer advocacy and reinforce trust, indirectly reducing churn risk through peer influence.

Improve Service Satisfaction through Feedback Loops

Dissatisfaction, as captured through satisfaction scores, was a significant churn driver. Establishing continuous feedback mechanisms, such as post-interaction surveys and regular Net Promoter Score (NPS) assessments, can facilitate early detection of service dissatisfaction and guide responsive service improvements.

5. CONCLUSION

This study set out to decode consumer decision-making patterns, specifically focusing on customer churn behavior, through the application of statistical techniques to a telecom dataset. Using a combination of descriptive and inferential statistics, the research successfully identified the key factors influencing customer retention and disengagement.

The descriptive statistical analysis highlights significant differences in key customer metrics based on churn status. Customers who discontinued service exhibited notably higher average monthly charges and substantially shorter tenures compared to those who remained. Specifically, churned customers had a mean monthly charge of $73.89 and an average tenure of 15.59 months, while non-churned customers averaged $61.66 in monthly charges and 41.74 months in tenure. These findings suggest that elevated billing levels and early-stage disengagement may be critical drivers of customer attrition. Moreover, the lower average

Customer Lifetime Value (CLTV) among churned customers (\$4092.70 vs. \$4520.30) underscores the financial implications of churn for long-term profitability. Collectively, these results point to the importance of targeted retention strategies focused on reducing early dissatisfaction and improving perceived value to mitigate revenue loss and enhance customer longevity.

Most notably by using T-test The analysis reveals that the average Monthly Charges for churned customers in the sample do not significantly differ from the population average, indicating no significant evidence of a difference between the two groups. ($t = 19.56514$, $p > 0.05$), confirming that customers with higher monthly charges are more likely to discontinue service.

Collectively, these findings suggest that high recurring charges and early disengagement are critical contributors to customer churn. Businesses aiming to reduce churn should prioritize pricing strategies and enhance value perception during the early stages of the customer lifecycle to improve satisfaction, retention, and long-term profitability.

The quadratic regression analysis revealed that the relationship between customer tenure and churn behaviour is better explained by a non-linear model. The model's performance, indicated by a Mean Squared Error (MSE) of 28.23 and a Root Mean Squared Error (RMSE) of 5.31, reflects a reasonably accurate fit to the data. These relatively low error values suggest that the risk of churn varies at different stages of the customer lifecycle, rather than following a consistent linear trend. For instance, churn risk may initially be high during the early months of service, decline as tenure increases, and potentially rise again in later stages. This insight underscores the importance of lifecycle-aware retention strategies. Identifying and targeting these critical periods with tailored interventions can enhance customer satisfaction and reduce churn, ultimately contributing to improved long-term business performance.

The present analysis examined the relationship between customer tenure and monthly charges using correlation techniques. The findings indicate a weak positive correlation ($r = 0.23$), suggesting that while monthly charges tend to increase marginally with tenure, the strength of this association is minimal. This implies that tenure alone does not significantly influence billing amounts and may not serve as a strong predictor of monthly expenditure. These results underscore the need to consider additional variables—such as service type, customer demographics, and usage patterns—for a more comprehensive understanding of factors affecting monthly charges. Future research could benefit from a multivariate approach to uncover more nuanced relationships within the dataset.

The correlation analysis between customer tenure and monthly charges revealed a weak positive relationship ($r = 0.23$), indicating that monthly charges tend to increase slightly with tenure. This finding was further supported by the scatter plot, which illustrated a minor upward trend accompanied by considerable data dispersion. The visual evidence reinforces the statistical outcome, suggesting that while there is a marginal association, tenure alone is not a strong predictor of monthly billing. The observed variability points to the influence of other underlying factors. Future investigations should adopt a multivariate framework to account for additional variables that may more accurately explain the variation in monthly charges.

The findings demonstrate that churn is not a random occurrence but rather a predictable outcome influenced by a complex interplay of pricing, contractual commitments, service satisfaction, and usage behavior. Specifically, customers on month-to-month contracts, those incurring higher charges, and those with lower tenure and engagement were more prone to discontinuation. Predictive analysis also emphasized the importance of key indicators such as customer lifetime value (CLTV), tenure, and the use of value-added services in forecasting churn likelihood.

Overall, the study reinforces the utility of data-driven approaches in understanding consumer decisions. It underscores the importance of proactive churn management through strategic pricing, personalized retention interventions, and enhanced customer engagement. These insights can aid telecom providers in optimizing their customer relationship strategies and reducing long-term attrition.

6. REFERENCES:

Ala'raj, M., & Abbod, M. F. (2016). Predicting customer churn in telecom industry using multilayer perceptron neural networks: Modeling and analysis. *International Journal of Advanced Computer Science and Applications, 7*(1), 152–157.

Huang, B., Kechadi, M. T., & Buckley, B. (2012). Customer churn prediction in telecommunications. *Expert Systems with Applications, 39*(1), 1414–1425.

Lariviere, B., & Van den Poel, D. (2005). Predicting customer retention and profitability by using random forests and regression forests techniques. *Expert Systems with Applications, 29*(2), 472–484.

Rashidi, L., & Khalilzadeh, M. (2020). Churn prediction in telecom using machine learning techniques. *Journal of Artificial Intelligence and Data Mining, 8*(4), 595–607.

Tsai, C.-F., & Lu, Y.-H. (2009). Customer churn prediction by hybrid neural networks. *Expert Systems with Applications, 36*(10), 12547–12553.

Witten, I. H., Frank, E., & Hall, M. A. (2011). *Data mining: Practical machine learning tools and techniques* (3rd ed.). Morgan Kaufmann.

Zeithaml, V. A., Bitner, M. J., & Gremler, D. D. (2017). *Services marketing: Integrating customer focus across the firm* (7th ed.). McGraw-Hill Education.

Solomon, M. R. (2018). Consumer behavior: Buying, having, and being. *Journal of Consumer Marketing*, 35(3), 209–216.

Sheth, J. N., & Mittal, B. (2004). Customer behavior: A managerial perspective. *Journal of Consumer Research*, 31(3), 450–467.

Bagozzi, R. P., Gürhan-Canli, Z., & Priester, J. R. (2002). *The Social Psychology of Consumer Behavior*. Open University Press.

Ajzen, I. (1991). The theory of planned behavior. *Organizational Behavior and Human Decision Processes*, 50(2), 179–211.

Bettman, J. R., Luce, M. F., & Payne, J. W. (1998). Constructive consumer choice processes. *Journal of Consumer Research*, 25(3), 187–217.