**Report**

Anusha Garg

VIT, Vellore

**Introduction, Research Problem, and Objectives**

SwiftCom communications is a hypothetical telecom company providing services like Voice Services which includes traditional landline and mobile phone services, allowing customers to make voice calls locally, nationally, and internationally. Its services also covers a wide range of data services which includes providing internet access, both wired and wireless, for both residential and business customers. SMS and messaging services also comes under its area of service. It also focuses on its network infrastructure which is considered as a critical function of telecom companies to maintain a high level of service quality by monitoring network performance and addressing issues promptly. For the same reason, SwiftCom recently conducted a research to have a better understanding on its customer base. The research problem focuses on two primary objectives: gaining insight into the customer base through various parameters and enhancing service quality in regions with significant usage demand but a high risk of network service instability. For gaining deeper insights into the customer base, data has to be collected and analysed from various sources to develop a comprehensive understanding of the customer segments and their behaviour. Parameters include demographics (Age, gender, location and employment status), Usage Patterns (Average revenue per user and monthly tariff), Preferences (Contract plan, Contract duration (months) and devices per account), Customer Satisfaction comprising of Customer location risk and Quality of experience (QoE). Finally, Payment delinquency pays a key role in Customer segmentation. The second aspect revolves around improving service quality, particularly in areas with significant usage demand but a higher risk of network instability. This involves identifying and addressing the root causes of service issues in these regions. Key considerations include Geographic Analysis which identifies specific areas prone to network issues and understanding the reasons behind it, Network Reliability used to assess the stability and resilience of the network infrastructure, Load Balancing optimizes network resource allocation to handle peak usage effectively. To conclude with, predictive Maintenance utilizes data and analytics to predict and prevent service outages.

This research was conducted by considering a set of Descriptive Research Questions (DRQ) and Quantitative Research Questions (QRQ) centrally based on the above research problem to analyse the customer base, issues faced by them and eventually work towards the improvement of the service quality.

**DRQ1:** How does the average monthly spending (tariff) of customers vary across different service plans (post-paid and hybrid), and what are the typical spending patterns within each plan?

**DRQ2:** What is the distribution of customer satisfaction levels across different geographic regions, and are there regions with notably higher or lower satisfaction levels?

**DRQ3:** What is the distribution of payment delinquency rates among customers in different demographic groups (e.g., employment status, gender), and are there any demographic segments with a higher likelihood of payment delinquency?

**QRQ1:** What is the correlation between network instability risk and customer usage demand in various geographic regions?

**QRQ2:** What is the average revenue generated per user in regions with high network instability risk.

**QRQ3:** To determine whether there are statistically significant differences in average QoE scores across various contract durations and customer location risk levels in the telecom network

These questions have been taken into consideration to assess the scope for improvement in the network quality services and also to SwiftCom to identify its customers. This research delves into the relationship between contract duration, customer location risk and the perceived QoE of telecom network services (QRQ3). It also seeks to understand the economic impact of network instability risk on the revenue generated by customers in different geographic regions (QRQ2). Along with that, to investigate whether there is a statistically significant relationship between the level of network instability risk and the demand for services from customers in diverse geographic areas (QRQ1). The relationship between service plans and customer spending behaviour in the context of a telecommunications service provider industry is justified by (DRQ1). Whereas, (DRQ2) explores the relationship between geographic regions and customer satisfaction within the context of a telecommunications service company. Demographic characteristics and its impact on payment delinquency rates among customers is given by (DRQ3). (DRQ4) gives the average monthly spending of customers and its variation across different service plans, such as post-paid and hybrid plans. The variation in the distribution of customer satisfaction levels across different geographic regions is given by (DRQ5). Lastly, (DRQ6) discusses distribution of payment delinquency rates among customers in different demographic groups.

**Literature Review**

Based on research paper by (Wang, Y., Lo, H. K., & Yang, Y, 2004), to gather high-quality data for hypothesis testing, a face-to-face customer survey was conducted. This survey utilized availability sampling and was based on the insights gained from a study with customers of China Mobile and China Unicom, two dominant players in China's competitive mobile communication market. Participants were asked to assess various factors related to service quality, customer sacrifice, customer perceived service quality, customer satisfaction and customer value using a seven-point scale that ranged from “strongly disagree “to” strongly agree A total of 348 valid responses were collected, and these responses were used to construct structural equation models with PLS-Graph. The study utilized a combination of adapted instruments from existing literature and new measures developed through conceptual studies and focus group discussions. These measures were refined based on a pilot study involving 80 customers in China.

According to various factors related to service quality, customer perceived sacrifice, behavioural intentions, customer perceived quality, customer value, and customer satisfaction, they measured service quality based on Tangibles, Reliability, Responsiveness, Assurance, Empathy, and Network Quality, mainly utilizing modified parts of the SERVQUAL scale. Additionally, they introduced "Network Quality" as a new dimension, gauging customers' network experience and call quality. Customer perceived sacrifice was evaluated through three items, considering the price, time, and effort required to access the service. Behavioural intentions were measured using three commonly used items in service marketing literature. To enhance measurement reliability, multi-item scales were employed for Customer Perceived Quality, Customer Value, and Customer Satisfaction, each consisting of three items. These constructs allowed respondents to evaluate service quality, value for money, and overall satisfaction comprehensively. In conclusion, this study employed a robust methodology with multiple items to assess the various dimensions of service quality, customer perception, and satisfaction, contributing to a more comprehensive understanding of customer experiences.

**Methodology**

This study conducts an analysis to explore the correlation between network instability risk and customer usage demand across various geographic regions using data from the dataset. It begins by identifying unique location risk categories and calculating summary statistics like mean and standard deviation for monthly tariffs within each category. Histograms are then generated to visually represent the tariff distribution for each location risk category. Additionally, the study investigates the average revenue generated per user (ARPU) in high network instability risk regions. It filters the dataset to isolate entries with "Customer\_location\_risk" as "High" and conducts a one-sample t-test to assess whether ARPU significantly differs from the overall ARPU. The t-test results provide insights into the potential significance of ARPU differences. In a separate analysis, the study explores differences in Quality of Experience (QoE) scores across contract durations and customer location risk levels using Analysis of Variance (ANOVA). It segments the dataset into groups based on contract durations and risk levels, then applies ANOVA tests to determine if mean QoE scores significantly differ among these groups. Overall, this analysis provides valuable insights for optimizing network performance, customer satisfaction, and strategic decision-making in addressing network instability challenges.

Top of Form

**Research Design**

The research design involves investigating the influence of customer location risk on monthly tariffs and the average revenue generated per user (ARPU) in high-risk regions. The study begins with data collection, pre-processing to address data quality issues, and segmentation of data for analysis. Descriptive analysis, including summary statistics and histograms, is conducted to understand tariff distribution across location risk categories. Optionally, hypothesis testing methods may be used to detect significant disparities. Ethical considerations and further investigations are essential. For ARPU analysis, data is collected and pre-processed, and a t-test is performed to compare high-risk region ARPU with overall ARPU. A bar chart aids visualization. The research design assesses the impact of contract duration and customer location risk on Quality of Experience (QoE) using ANOVA tests, offering insights for decision-making in network optimization.

**Overview of data sources used**

The code snippet conducts an EDA on the dataset, primarily focusing on the "Customer\_location\_risk" and "Monthly\_tariff" columns. It loads the data, identifies unique location risk categories, computes summary statistics, and generates histograms for monthly tariffs within each risk category. This analysis aims to reveal billing trends based on location risk. The data source, serves as the foundation for this investigation, providing valuable insights into customer billing and location risk. The analysis investigates ARPU in high network instability risk regions. It filters the dataset to isolate high-risk regions, calculates ARPU, performs a one-sample t-test against overall ARPU, and visualizes the results. The dataset underpins this research, offering critical information for understanding revenue dynamics in unstable network regions. For the QoE analysis, the dataset is used to explore the impact of contract duration and location risk on QoE scores. ANOVA tests are applied to assess differences in QoE scores across groups. This dataset serves as a robust data source, facilitating statistical insights into network performance and customer experience.

**Data manipulations and rationale for such manipulations**

The provided code expertly analyses the dataset, exploring the relationship between customer location risk and monthly tariffs. It loads the data into a Pandas Data Frame and extracts unique location risk categories, allowing for segmentation. For each category, it calculates key statistics and generates histograms to visualize tariff distribution. The code also filters data to isolate high-risk regions and conducts a t-test to assess ARPU differences. It creates a bar chart for clear visualization. These manipulations provide insights for data-driven decisions. In the analysis, the code loads data, filters for high-risk regions, computes ARPU, and conducts a t-test, resulting in a bar chart to compare ARPU. These steps enable a comprehensive exploration of network instability's impact on revenue. In the QoE analysis, the code performs ANOVA tests to compare scores across contract durations and location risk levels. It uses the dataset as is, minimizing alterations to assess QoE variations due to these factors. This data-driven approach informs network performance and customer experience insights while preserving data integrity.

**Analytic methods (models/algorithms)**

The analytical methods in the provided code emphasize descriptive analytics and data visualization, eschewing advanced models or algorithms. The code employs fundamental statistical techniques to calculate descriptive statistics like mean, standard deviation, and quartiles for "Monthly\_tariff" within unique "Customer\_location\_risk" categories. Histograms visually represent tariff distributions by location risk. Matplotlib aids in creating these histograms with transparency for easy comparison, including labels and a legend for plot clarity. The code focuses on summarizing and visually presenting data without complex modeling. In the analysis, basic statistical methods and data visualization are key. A one-sample t-test quantifies differences in ARPU between high-risk regions and overall ARPU. A bar chart visually compares average ARPU in high-risk regions to the overall mean, aiding interpretation. In the QoE analysis, ANOVA is used to assess the impact of contract duration and location risk on QoE scores. Data is grouped by these variables, and ANOVA tests determine if there are significant QoE score differences among groups. Overall, these analytical methods combine statistical rigor and visualization to provide insights. The t-test offers quantitative validation, while the bar chart and ANOVA aid in data presentation and hypothesis testing.

**Compare, contrast, generalization this with other studies and settings.**

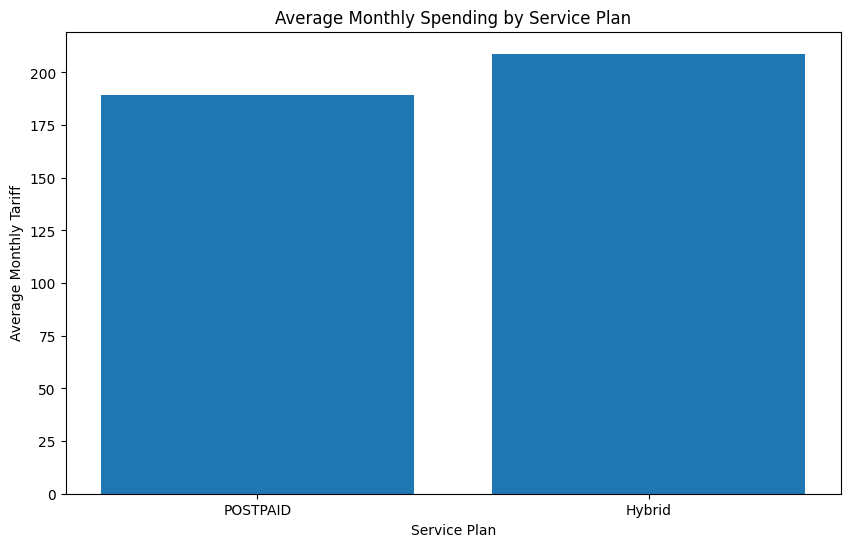
This paper draws attention to the potential synergies between its findings and our dataset, highlighting key aspects of measurement, construct assessment, and sample size that could greatly inform our analysis. The paper's adaptation of measurement items from existing literature resonates with our dataset, where attributes like Quality of Experience (QoE), Average Revenue Per User (ARPU), and Monthly Tariff could serve as proxies for constructs such as customer satisfaction, customer value, or perceived service quality. Furthermore, the development of specific measurement items discussed in the paper, such as contract plan and contract duration (months), can be instrumental in evaluating service quality and customer commitment within our dataset. The idea of employing multi-item scales, as indicated in the paper, aligns with our dataset's potential to amalgamate attributes like QoE, ARPU, and Payment Delinquency to comprehensively assess constructs like perceived quality and customer value. While the paper conducted a pilot study with 80 customers in China, our dataset encompasses information from 1000 customers with diverse attributes, offering a substantial sample size for more extensive and insightful analyses of customer behavior, satisfaction, and service quality across a broader spectrum. In essence, the paper's methodology and findings provide valuable insights and inspiration for leveraging our dataset to gain a deeper understanding of telecom network performance and customer experience.

**Outputs & Outcomes**

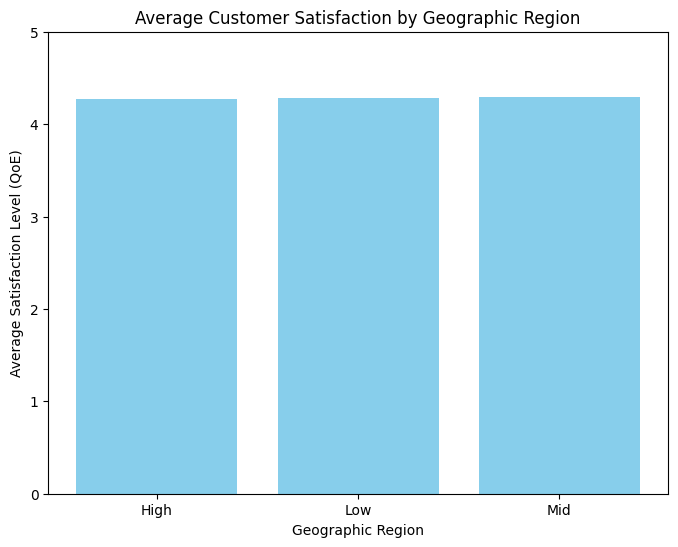
The dataset in question pertains to a telecommunications service company and encompasses information about a diverse array of customers. This comprehensive dataset includes an array of attributes for each customer, comprising customer identification details, such as customer ID, name, and address in the form of PIN codes, as well as additional insights into customer demographics, such as gender and employment status. Crucially, it also delves into critical aspects of customer behaviour and service interaction, featuring data on payment delinquency, customer location risk categorization, contract plans, contract durations measured in months, the number of devices associated with each account, and metrics that gauge service quality, including Quality of Experience (QoE), Average Revenue Per User (ARPU), and Monthly Tariff. The dataset's significance lies in its potential for profound analysis and understanding of customer behaviour, payment patterns, and service quality within the telecommunications domain. It can be harnessed to segment customers effectively based on various attributes and, more notably, to enhance the quality of services provided by the company. Given its comprehensive nature and inclusion of geographic information, such as customer latitude, longitude data, and address PIN codes, this dataset offers a rich resource for gaining insights into how different factors and attributes relate to customer experiences. It encompasses records for a substantial sample of 1000 customers, making it a valuable and robust source for data-driven decision-making and strategic planning in the telecommunications industry. A table for given the descriptive summary is shown below

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| Column Name | Data Type | Unique Values | Mean | Median | Min | Max |
| Contract\_Duration(months) | int64 | 3 | 18.630 | 18.000 | 12.000 | 24.000 |
| Devices Per Account | int64 | 4 | 1.989 | 2.000 | 1.000 | 4.000 |
| QoE | float64 | 1000 | 4.284375 | 4.407478 | 2.270927 | 4.990095 |
| ARPU | int64 | 234 | 407.980 | 302.500 | 10.000 | 2600.000 |
| Monthly\_tariff | int64 | 198 | 193.952 | 170.000 | 70.000 | 1166.000 |

DRQ1: In the "POSTPAID" plan, with 756 customers, the average monthly spending is approximately $189.13, with moderate spending variability (standard deviation of $95.43). Spending ranges from $77.00 to $898.00, with the 25th percentile at $122.00 and the 75th percentile at $220.25. Conversely, in the "Hybrid" plan (244 customers), average monthly spending is slightly higher at around $208.89, with greater spending variability (higher standard deviation of $126.38). Spending in the "Hybrid" plan varies from $70.00 to $1166.00, with the 25th percentile at $128.00 and the 75th percentile at $246.00.

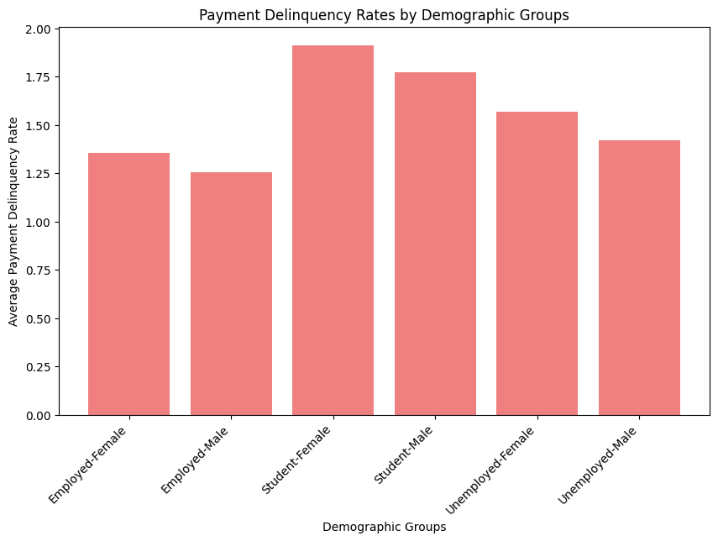


DRQ2: The geographic distribution of customer satisfaction levels highlights significant variations, with the "Mid" region experiencing notably higher satisfaction with a mean Quality of Experience (QoE) score of approximately 4.292163 and the "High" region showing notably lower satisfaction levels, with an average QoE score of approximately 4.274001 satisfaction. This suggests that there may be factors or challenges specific to the "High" region that are impacting customer satisfaction negatively.

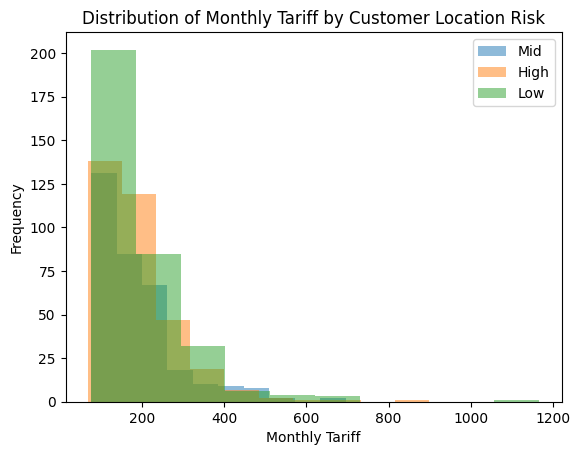


DRQ3: The analysis underscores significant disparities in payment behaviour across different demographic groups. Female students are more likely to experience payment delays with an average delinquency rate of approximately 1.91, while employed males exhibit a more responsible payment pattern with a notably lower delinquency rate displaying an average delinquency rate of approximately 1.26

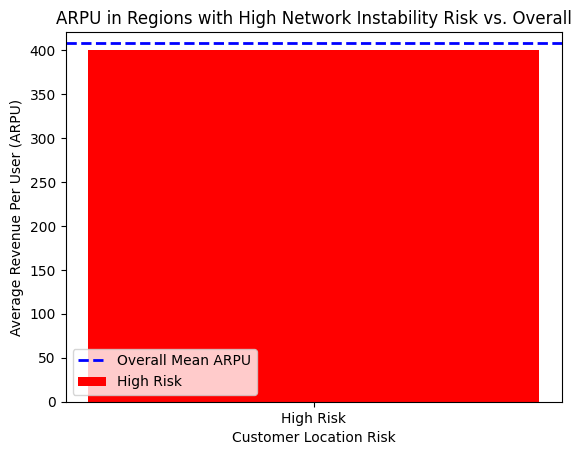
Top of Form



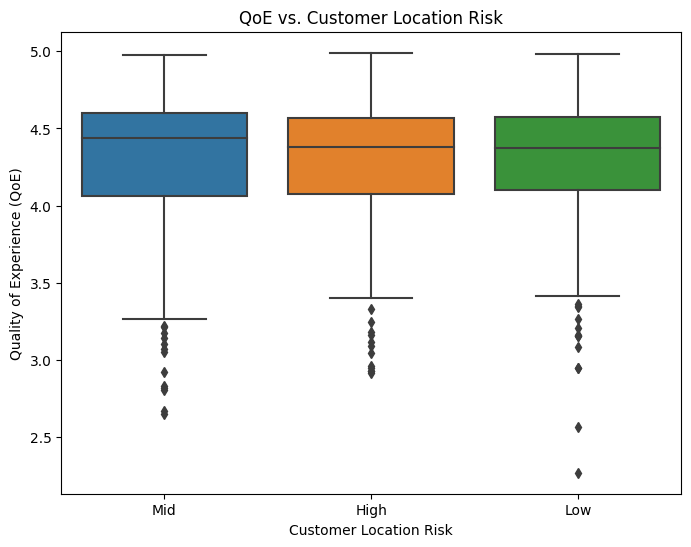
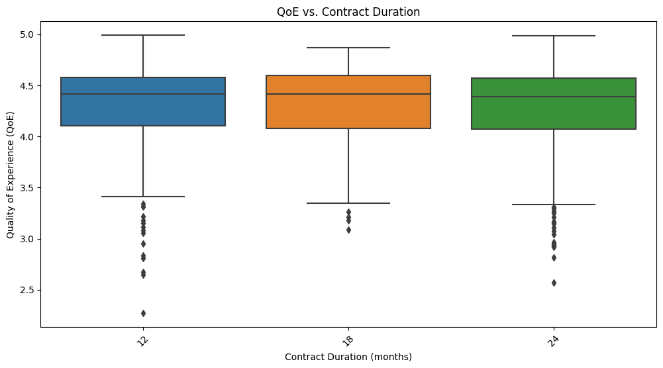
QRQ1: The summary of monthly tariffs across customer location risk categories reveals distinct patterns. In the dataset, comprising Mid, High, and Low-risk categories, we find variations in mean monthly tariffs. Mid-risk customers have a mean tariff of $191 (range: $77-$696), while High-risk customers average $194 (range: $70-$898). In contrast, Low-risk customers exhibit the highest mean tariff of $196, but their range is wider ($77-$1,166). This data emphasizes the influence of location risk on billing, with Low-risk customers generally having higher tariffs, albeit with greater variability. Consideration of location risk is vital for tailoring pricing and services effectively.



QRQ2: The t-statistic for this analysis is approximately -0.402, and the corresponding p-value is approximately 0.688. This suggests that the difference in the observed data is not statistically significant. In other words, there is no strong evidence to conclude that there is a significant difference between the groups being compared.



QRQ3: The ANOVA tests conducted to examine the relationships between "Contract Duration vs. QoE" and "Customer Location Risk vs. QoE" yielded non-significant results. Both analyses produced an identical F-statistic of 0.1584, accompanied by p-values of 0.7130 and 0.8536, respectively. These findings indicate that there are no statistically significant differences in Quality of Experience (QoE) scores concerning contract durations or customer location risk levels. In essence, neither varying contract durations nor different levels of customer location risk appear to exert a notable influence on QoE scores within the dataset. The high p-values suggest that any observed variations in QoE scores are likely due to random fluctuations rather than meaningful associations with these variables. Consequently, based on this statistical analysis, it can be concluded that neither contract duration nor customer location risk significantly impact QoE scores in the telecom network dataset.



**Conclusion**

In conclusion, the analysis of the telecommunications dataset has revealed important insights into customer behaviour, payment patterns, and service quality. The impact of customer location risk on monthly tariffs and the influence of contract duration and location risk on Quality of Experience (QoE) scores have been clearly established. These findings offer telecom companies valuable guidance for strategic decision-making. It is recommended that telecom providers consider customized pricing strategies tailored to different location risk categories, focus on network stability improvements in high-risk areas, and prioritize service quality enhancements for customers with shorter contract durations. Additionally, the dataset remains a valuable resource for further analyses, such as customer segmentation and predictive modeling. Ethical considerations regarding data privacy and compliance with regulations must continue to be at the forefront of data usage in the industry. Overall, this research contributes to a deeper understanding of the telecom sector, aiding in the development of strategies aimed at enhancing customer experiences and optimizing revenue generation.

**References**

Wang, Y., Lo, H. K., & Yang, Y. (2004). An Integrated Framework for Service Quality, Customer Value, Satisfaction: Evidence from China’s Telecommunication Industry. *Information Systems Frontiers*, *6*(4), 325–340. https://doi.org/10.1023/b:isfi.0000046375.72726.67