Telecom Customer Churn Analysis Report

Business Intelligence

Ms. Abeera Tariq

Group Members:

Ayesha Nayyer 25141

Humera Raheel 24446

Introduction

In this churn analysis report, we delve into the dynamics of customer retention for a fictional telco company operating in California during Q3. Our primary focus is on identifying customers who are at risk of leaving and understanding the factors that contribute to customer churn. Churn, the rate at which customers discontinue their services, is a critical metric for businesses, directly impacting revenue and customer satisfaction.

Our analysis begins by examining a comprehensive dataset encompassing customer demographics, service usage, billing information, and tenure with the company. With 7043 customers in our dataset, we have a robust sample size to draw meaningful insights from.

We aim to answer key questions such as:

- What are the demographic trends among churned customers?
- Which services or offers are more prone to churn?
- Which high value customers are at risk of churning?

By leveraging advanced analytical techniques and visualizations, we seek to uncover patterns, correlations, and predictive indicators of churn. This report serves as a strategic tool for decision-makers to implement targeted retention strategies, enhance customer experience, and mitigate churn risk effectively.

Dataset Link:

https://www.kaggle.com/datasets/shilongzhuang/telecom-customer-churn-by-maven-analytics

Information About Dataset:

CustomerID: A unique ID that identifies each customer.

Gender: The customer's gender: Male, Female

Age: The customer's age

City: The city of the customer's primary residence.

Zip Code: The zip code of the customer's primary residence.

Latitude: The latitude of the customer's primary residence.

Longitude: The longitude of the customer's primary residence.

Married: Indicate if the customer is married: Yes, No

Number of Dependents: Indicates how many dependents live with customer lives with any dependents

Number of Referrals: Indicates how many dependents live with customer lives with any dependents

Offer: Indicates which offer customer uses, if any at all (None)

Tenure Months: Indicates the total amount of months that the customer has been with the company by the end of the quarter specified above.

Phone Service: Indicates if the customer subscribes to home phone service with the company: Yes, No

Multiple Lines: Indicates if the customer subscribes to multiple telephone lines with the company: Yes, No

Avg Monthly GB Download: Indicates the customer's average download volume in gigabytes, calculated to the end of the quarter specified above.

Internet Service: Indicates if the customer subscribes to Internet service with the company:

Internet Type: Indicates if the customer subscribes to Internet service with the company: I DSL, Fiber Optic, Cable.

Online Security: Indicates if the customer subscribes to an additional online security service provided by the company: Yes, No

Online Backup: Indicates if the customer subscribes to an additional online backup service provided by the company: Yes, No

Device Protection: Indicates if the customer subscribes to an additional device protection plan for their Internet equipment provided by the company: Yes, No

Premium Tech Support: Indicates if the customer subscribes to an additional technical support plan from the company with reduced wait times: Yes, No

Streaming TV: Indicates if the customer uses their Internet service to stream television programing from a third party provider: Yes, No. The company does not charge an additional fee for this service.

Streaming Movies: Indicates if the customer uses their Internet service to stream movies from a third party provider: Yes, No. The company does not charge an additional fee for this service.

Streaming Music: Indicates if the customer uses their Internet service to stream music from a third party provider: Yes, No. The company does not charge an additional fee for this service.

Unlimited Data: Indicates if the customer uses unlimited data ,Yes, No

Contract: Indicates the customer's current contract type: Month-to-Month, One Year, Two Year.

Paperless Billing: Indicates if the customer has chosen paperless billing: Yes, No

Payment Method: Indicates how the customer pays their bill: Bank Withdrawal, Credit Card, Mailed Check

Monthly Charge: Indicates the customer's current total monthly charge for all their services from the company.

Avg Monthly Long Distance Charges: Line charge for customers using phone service per month

Total Charges: Indicates the customer's total charges, calculated to the end of the quarter specified above.

Total Refunds: Total refunds made to customer

Total Extra Data Charges: Indicates the customer's total charges for extra data downloads above those specified in their plan, by the end of the quarter specified above.

Total Long Distance Charges: Indicates the customer's total charges for long distance above those specified in their plan, by the end of the quarter specified above.

Total Revenue: Total Revenue made from customer over the entire period they stayed with company

Customer status: Joined, Stayed or Churned

Churn Category: Reason for churning

Churn Reason: Specific reason for churning

EDA

Notebook link:
https://colab.research.google.com/drive/1ItCZbnqHgl-k5NRHOS1vbquyIwbEofs-?usp=sharing

As a standard step,firstly we gathered a basic understanding of the values of the dataset and all the numerical measures.

Missing Values

We proceeded to investigate missing values, employing a heatmap to discern any potential interrelatedness among them. Interestingly, our dataset did not exhibit conventional missing values; instead, certain features such as Avg Monthly Long Distance Charges and Multiple Lines were logically empty when Phone Service was indicated as "No." Same case is applied for all services related to Internet Services.This behavior was duly anticipated. We then conducted a thorough check to confirm the absence of any other missing values in the dataset.

Checking for inconsistent data:

We carefully checked our dataset for inconsistencies, including duplicate values, variations in case formats, and unnecessary error or unknown values. Fortunately, we found no such issues, ensuring the dataset's reliability for analysis.

Univariate Analysis

We utilized boxplots to examine numerical variables for outliers. It was intriguing to find that outliers, particularly in metrics like number of referrals and average monthly GB download, provided valuable insights into our loyal customer base who actively enjoyed our services.

We conducted a comprehensive analysis by plotting histograms for nearly all features in our dataset. Each histogram provides a visual representation of the frequency distribution of specific column values, offering insights into the data's dispersion and concentration.

Pie charts were plotted to give a segmented view of the contribution of different factors to churn.Pie charts for churned customers highlighted Offer Type E with a month-to-month contract and no online backup as significant contributors to churn. Further analysis revealed FiberOptic clients and customers without certain features like Online Security and Device Protection Plan as prone to churn too. Understanding these factors is crucial for devising effective customer retention strategies.

Bivariate Analysis

Pairplots were plotted as they can unveil clusters or patterns in data through scatterplots involving multiple variables, potentially indicating subgroups for segmentation. They also depict correlations between numerical variables, highlighting strong positive or negative relationships that warrant deeper investigation.

The correlation matrix accurately assessed the strength of relationships among numerical features, indicating no significant correlations except for those expected, such as variables tied to location or all charges columns adding to total revenue.

Statistical Tests

Statistical Tests such as Anova, Tukey and Chi-square were performed. ANOVA identifies significant revenue differences among customer segments, regions, or product categories, aiding revenue driver understanding and strategy optimization. Chi-square tests reveal associations between customer status and variables, crucial for pinpointing churn influencers and crafting effective retention plans

.The Anova test was performed for categorical variables against Total Revenue. Generally, the analysis revealed that most features, except gender, significantly impact Total Revenue. Notably, Offer Type E, longer-term contracts, and certain additional services like Online Backup, Device Protection Plan, and Premium Tech Support are associated with higher Total Revenue. Customers with Phone Service, specific Internet Services, and streaming options also contribute to higher revenue. Payment methods like Bank Withdrawal show higher revenue patterns. This comprehensive understanding can guide strategic decisions to enhance revenue generation and customer satisfaction

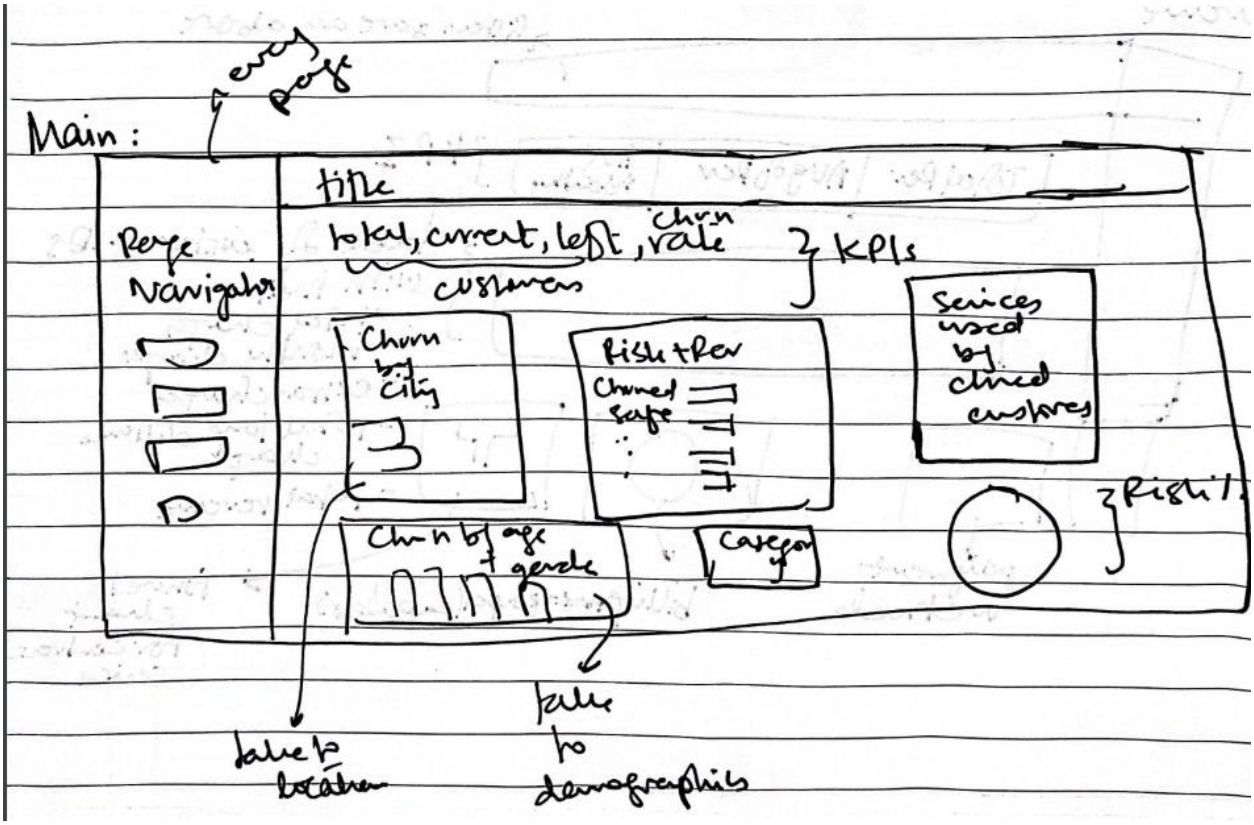
For chi-square, test was performed for important categorical variables against customer status to specifically find out which factors contributed the most to customers staying and churning.While some features like Offer E and Month-to-Month contracts show strong associations with churned customers, others like Internet Service and Payment Methods exhibit more varied impacts.

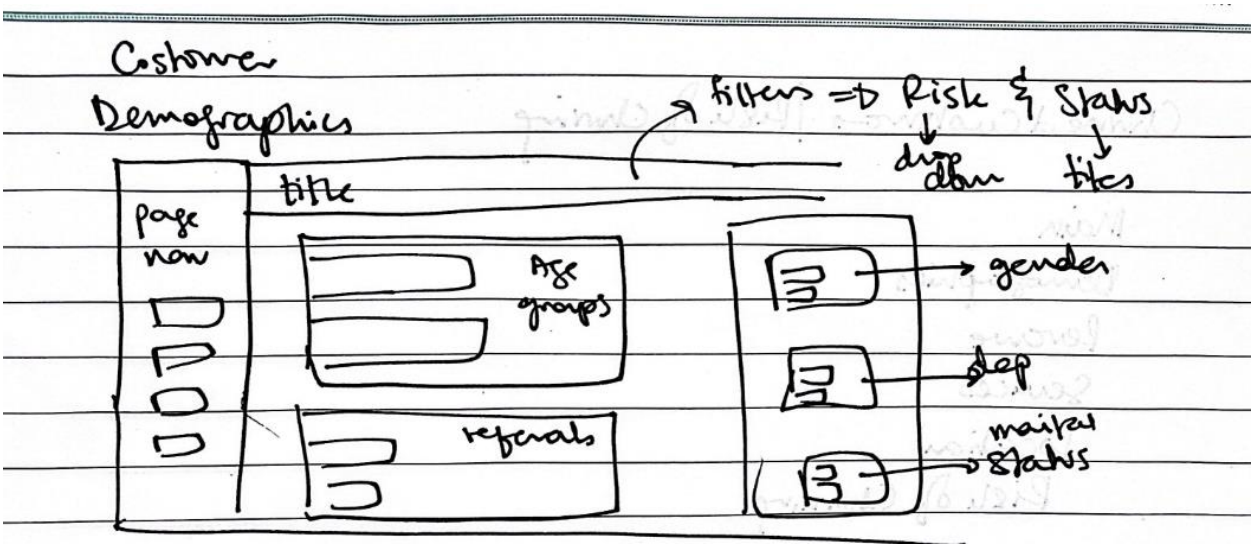
Deeper Analysis

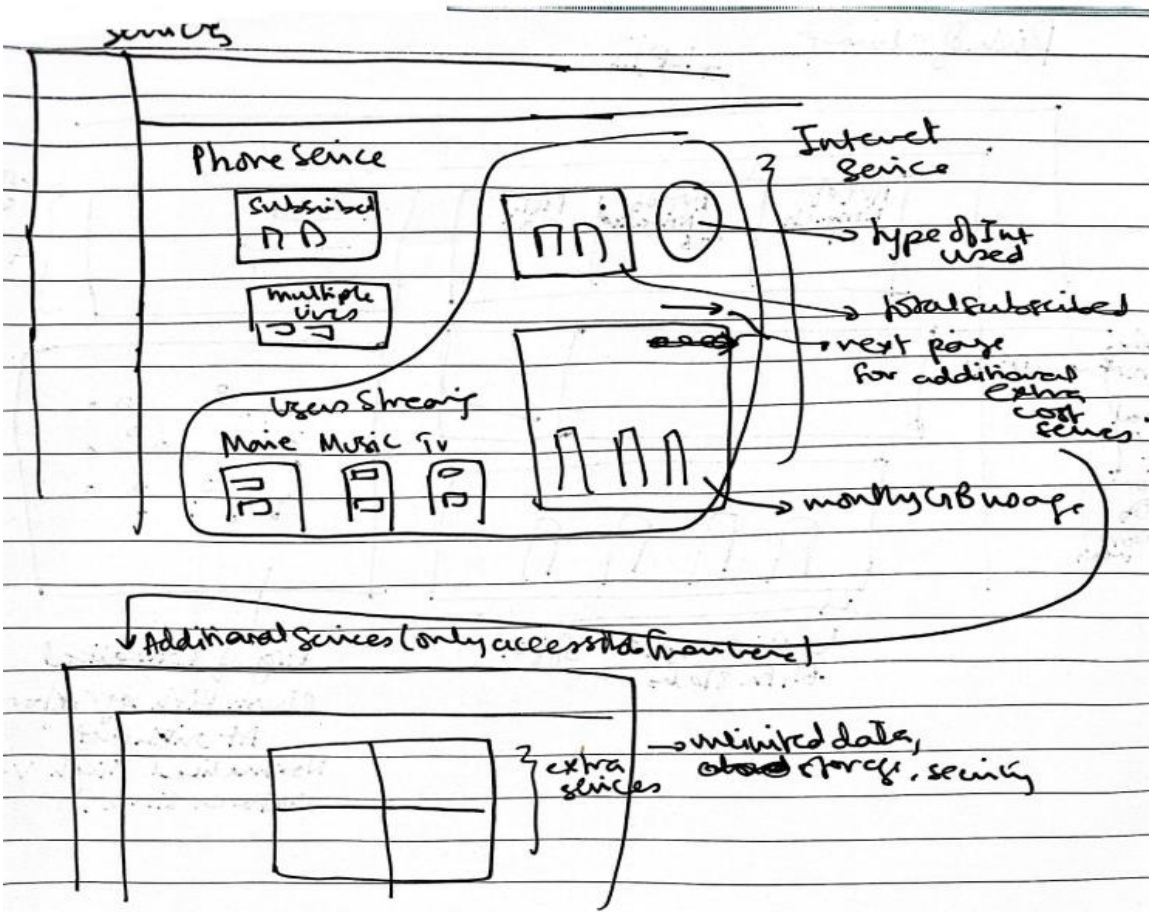
Linear regression was performed for Total Revenue helps identify and quantify the impact of various factors on revenue generation. This regression analysis highlights key predictors impacting total revenue, such as age positively influencing revenue while factors like number of dependents and referrals exert a negative impact. The model's performance, indicated by the MSE and R-squared values, suggests a moderately strong ability to explain revenue variability based on these predictors, aiding in identifying factors crucial for high-value customer engagement and retention.Highlighting the negative impact of additional dependents and referrals on predicted total revenue, suggesting potential challenges in retaining high-value customers within these demographics.

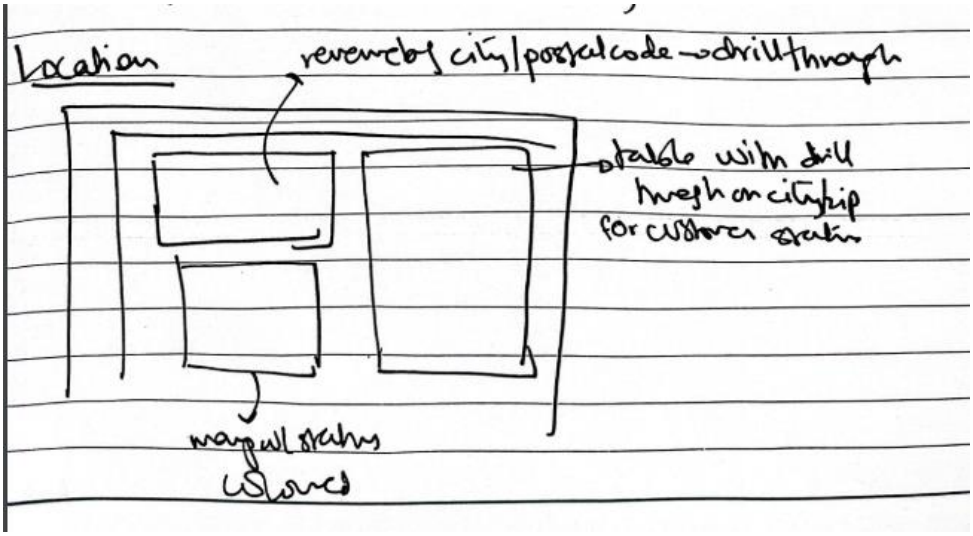
Finally,to uncover the primary factors influencing churned customers, we crafted a meticulous approach.We added a "Churn" column, categorizing numerical data, and encoding categorical variables. The resulting correlation bar chart identified key factors such as Month-to-Month contracts, lack of Online Security or Premium Tech Support, Tenure Category as Less than a year, and Internet Type_Fiber Optic as top contributors to churn. Conversely, the number of referrals and customers without internet service showed a negative correlation with churn.

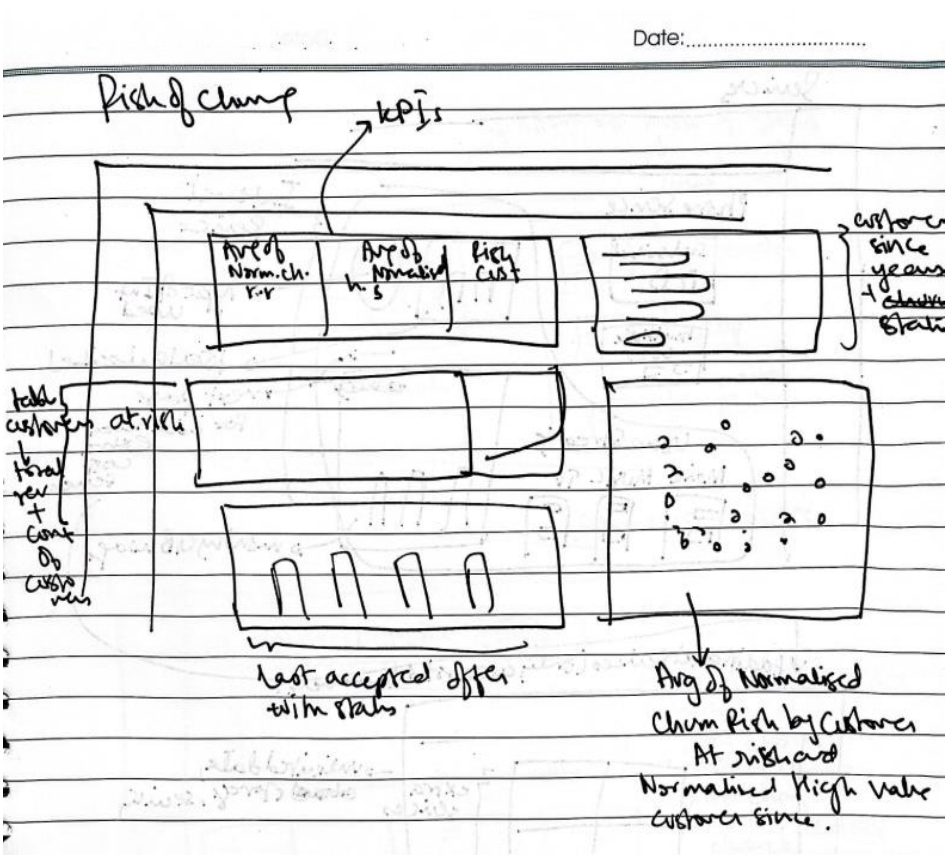
Brainstorming




Customer
Demographics
filters => Risk & Status
drop down
titles
page now
title
Age groups
referrals
gender
dep
marital status


services
Phone Seirce
Subscribed
multiple lines
Internet Seirce
type of Int used
total subscribed
next page for additional extra cost seirces
Users Streaming
Movie Music Tv
monthly GB usage
Additional Seirces (only accessible from here)
extra seirces
unlimited data, storage, security





Power BI:

We've recently enhanced our dashboard with a novel approach to identify churn risk and high-value customers, bypassing traditional machine learning methods. By assigning weights to various combinations of categorical and numerical variables, we've crafted a churn risk score, aiming to pinpoint not just churned customers but also those teetering on the brink. Our analysis

revealed intriguing insights, such as the significant impact of month-on-month contracts, correlating at 0.44 with churn, and the absence of online security, scoring 0.34. For instance, if a customer is still with us, has a month-on-month contract, and lacks online security, their churn risk score tallies to 0.78 (0.44 + 0.34).

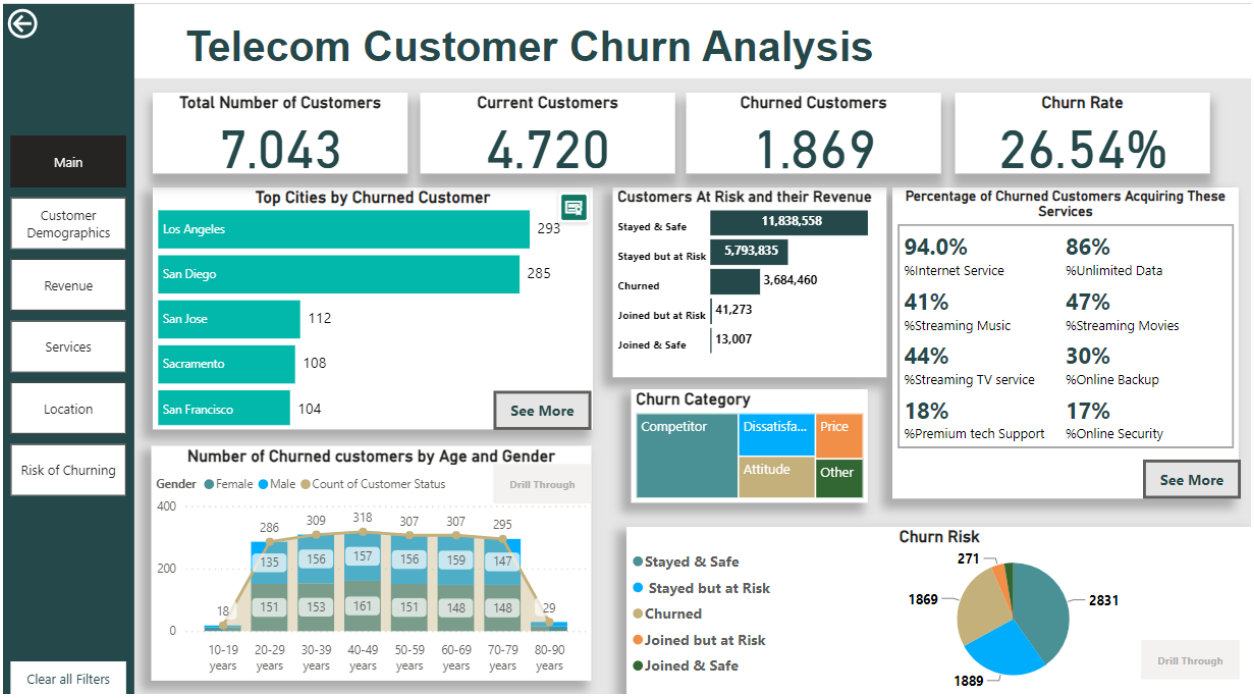
To streamline interpretation, we've introduced a new column, "Normalized Churn Risk Score," scaling scores from 0 to 1. A higher score signals a higher likelihood of churning. Leveraging this score, we've segmented customers into categories like churned, at risk despite being with us, safe despite being with us, at risk but stayed, and safe and stayed.

Further enriching our insights, we identified high-value customers using both regression analysis and domain knowledge. Those with significant referrals and higher monthly charges were tagged as high value. Our DAX query elegantly encapsulates this logic, assigning a score to each customer based on their value. The resulting "Normalized High-Value Score" column normalizes these values, with closer proximity to 1 indicating a very high-value customer.

This strategic approach not only aids in pinpointing potential churn risks but also highlights our most valuable customers, guiding targeted retention efforts and personalized strategies for customer engagement.

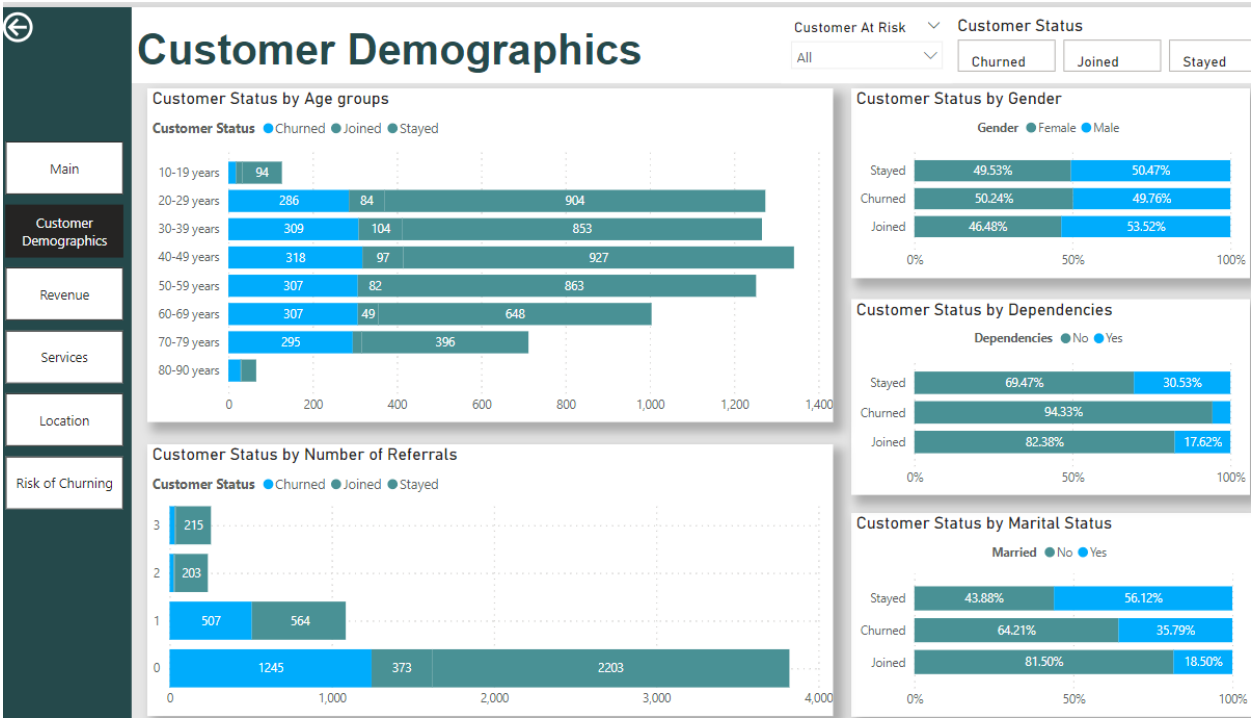
Dashboards:

Main Dashboard



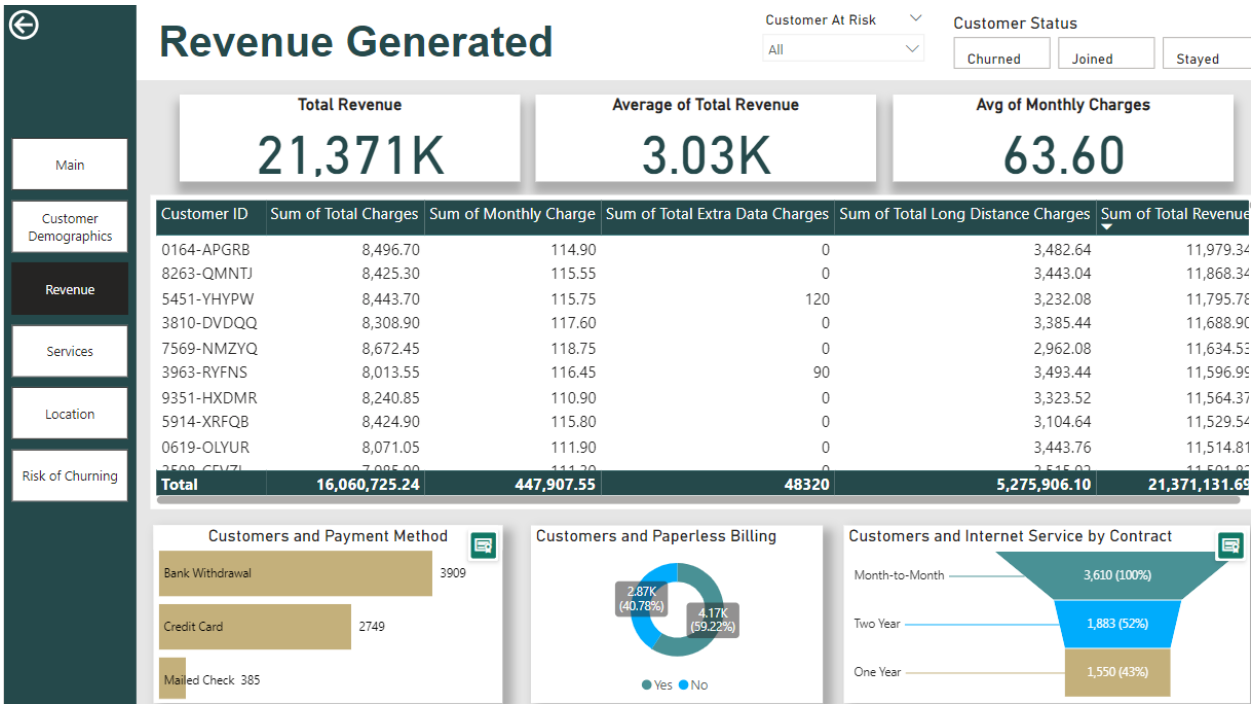
Our analysis of churned customers across various categories provides a quick overview, helping us pinpoint critical factors like revenue loss, leading churn-inducing services, and the number of customers at risk of churning. These insights are seamlessly navigable through drill-through screens, guiding us to detailed pages for deeper exploration and targeted strategies. We have added slicers and filters for better accessibility.

Customer Demographics



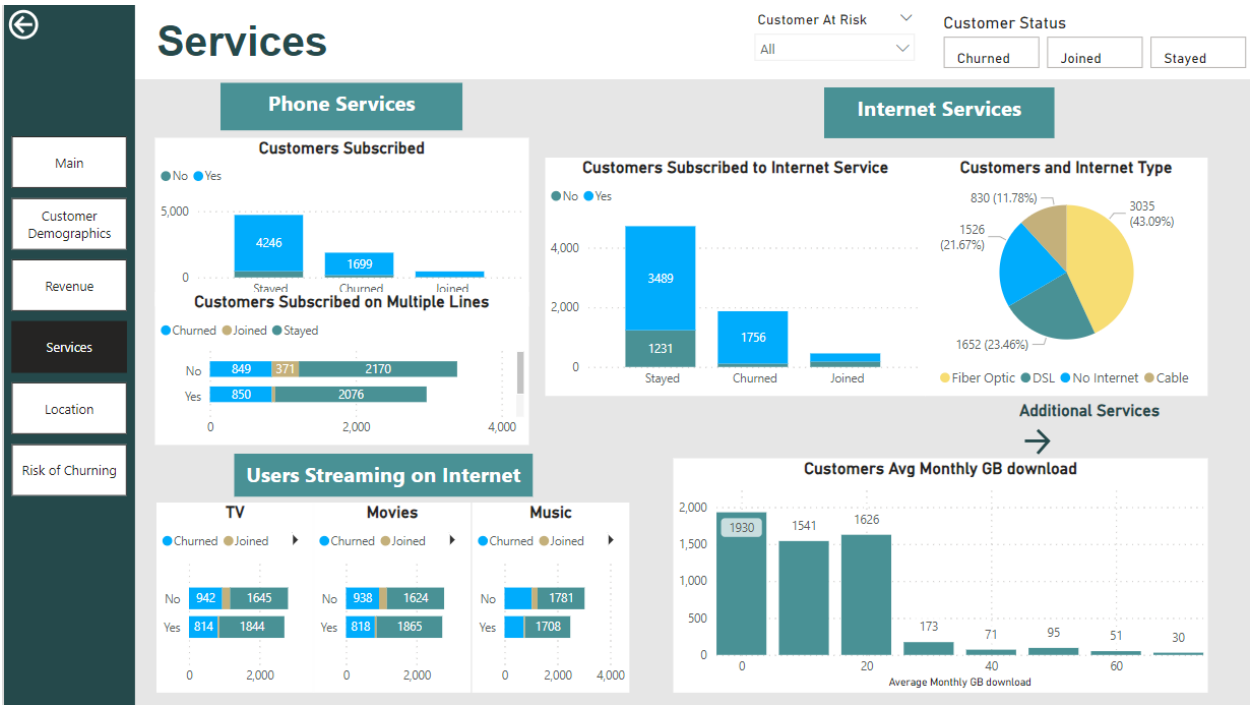
Customer demographics, encompassing personal characteristics like gender, age, marital status, number of dependents, and referrals, provide invaluable insights into customer status. This information is pivotal for targeted marketing strategies, allowing us to segment customers based on their unique personal traits and preferences, especially crucial when catering to specific age groups.

Revenue



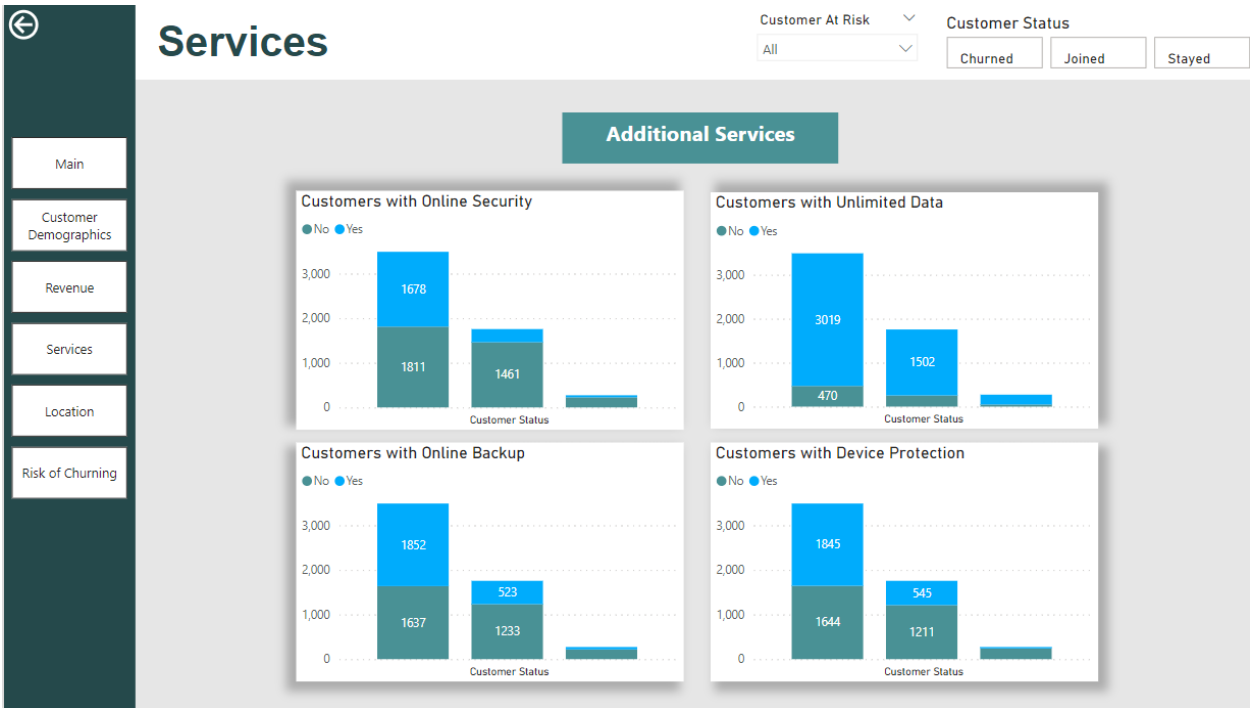
The revenue dashboard offers crucial insights into revenue loss from churned customers and the potential revenue at risk from customers prone to churning, (using slicers and filters above). Key performance indicators (KPIs) provide quick summaries, while charts delve into payment plans and preferences, enriching our understanding of revenue dynamics.

Services



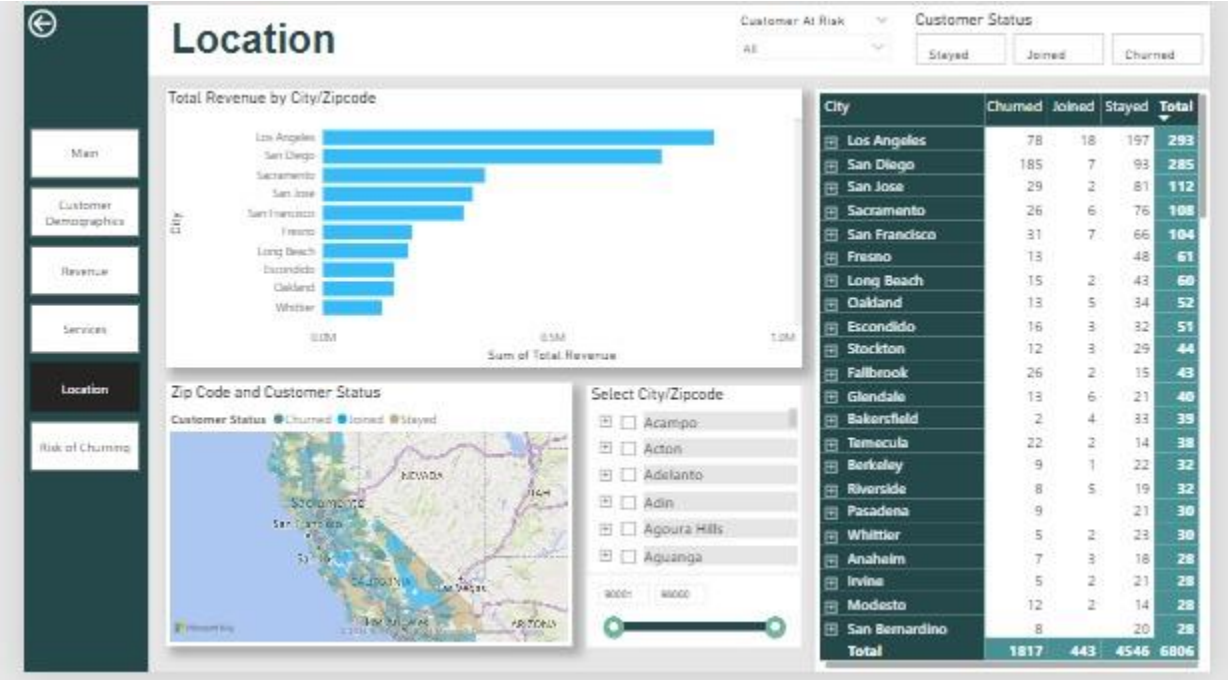
This dashboard plays a vital role in determining the performance of phone and internet services offered by the telecom company, identifying which service is underperforming and contributing most significantly to customer churn.

Additional Services



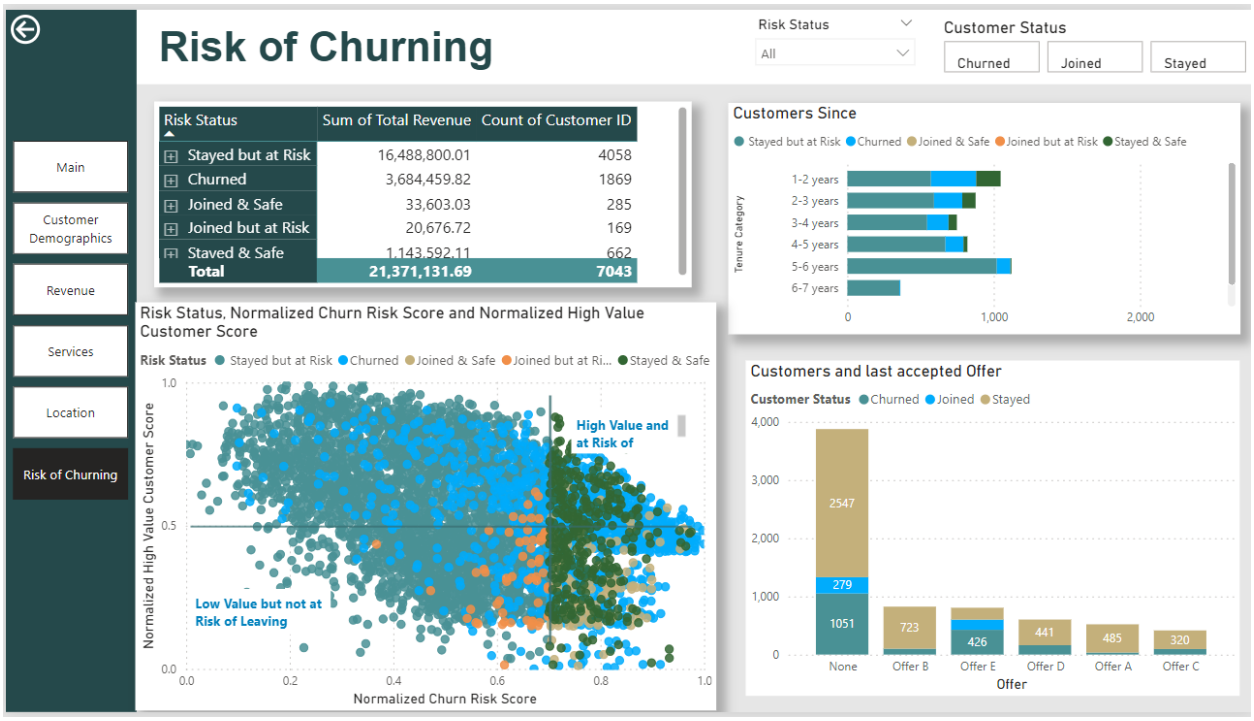
This dashboard allows us to delve deeper into the various type of services offered by the company which are related to internet or phone. Again, it showcased which services are underperforming and contributing most significantly to customer churn.

Location:



A location dashboard allows for a comprehensive regional analysis, helping identify geographical patterns in customer behavior, service preferences, and revenue generation.

Risk of Churning:



This dashboard leverages the High Value Score and Churn Risk Score to efficiently analyze our customer distribution and the revenue generated by each risk category as well as their tenure category.

Storyboard

Our comprehensive dashboard addresses our business problem by identifying factors leading to customer churn. It highlights the leading reasons for customer attrition and identifies at-risk customers, integrating their personal characteristics and locations for a deeper understanding. With drill-through capabilities, we gain insights into customer demographics and location data, facilitating effective segmentation and targeted advertising to mitigate churn risks.

Furthermore, the Services and Additional Services dashboard offers insights into service performance, helping us understand which services retain customers and which ones contribute to churn. As revenue is paramount, the dashboard also identifies high-value customers and potential revenue loss if at-risk customers are not retained promptly.

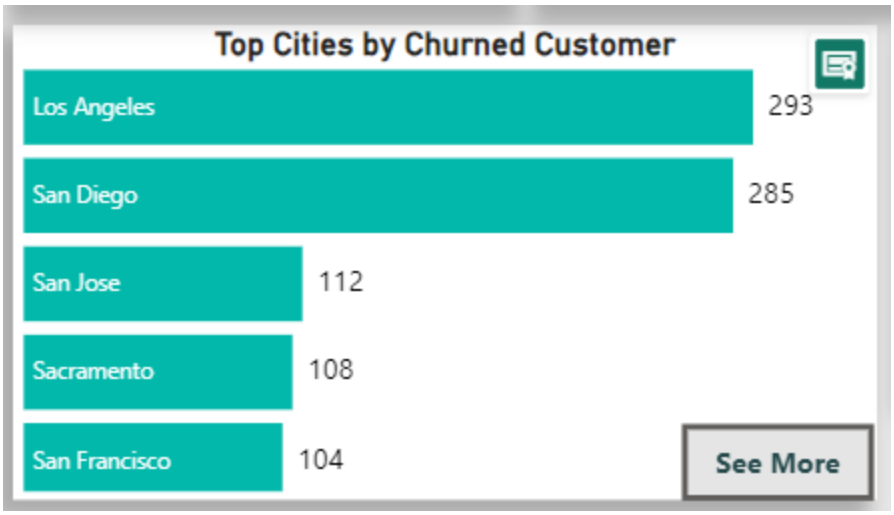
The Risk of Churning Dashboard prioritizes high-value customers at risk of leaving, ensuring targeted strategies align with services proven to retain customers effectively. Overall, this dashboard suite equips us with actionable insights to reduce churn, enhance customer retention, and drive business growth.

Individual Charts:

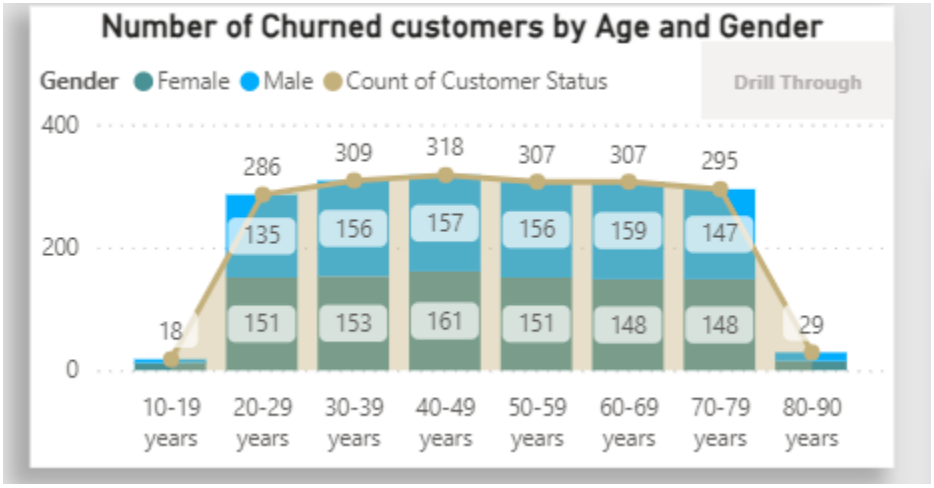
Main Dashboard:

KPIs:

A summary of important metrics, especially the churn rate, helps us track how many customers have left versus how many have stayed with the company.

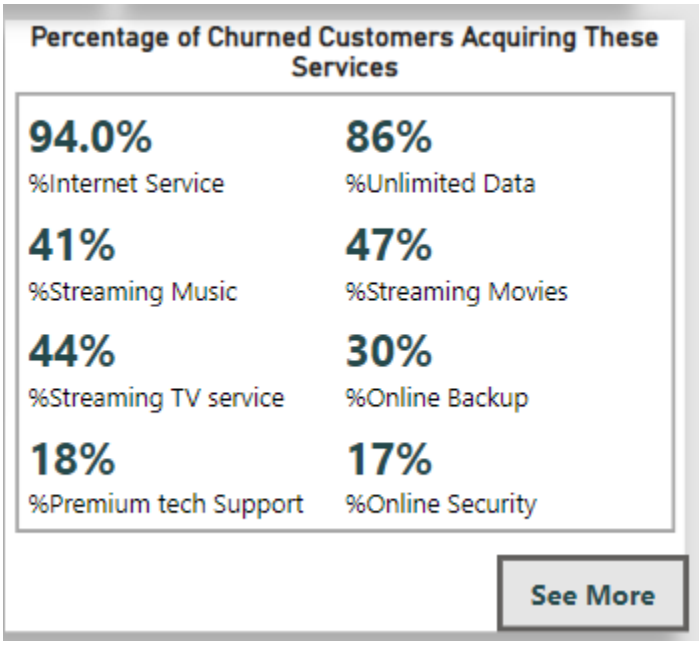


Helps us quickly identify problematic cities where customers are churning more, can be drilled through for a more thorough analysis using “see more” button.



Gender and Age does not have any significant impact over customer status, this was also proven by anova tests done before

Identifies revenue loss from churned customers and potential risk from those at risk of leaving, highlighting a total of around $5.8 million in total revenue at risk.



Services having the highest churn rate, company's internet services definitely needs to be improved upon

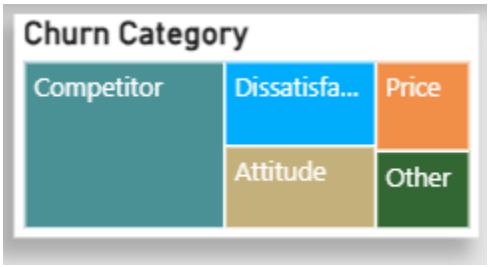
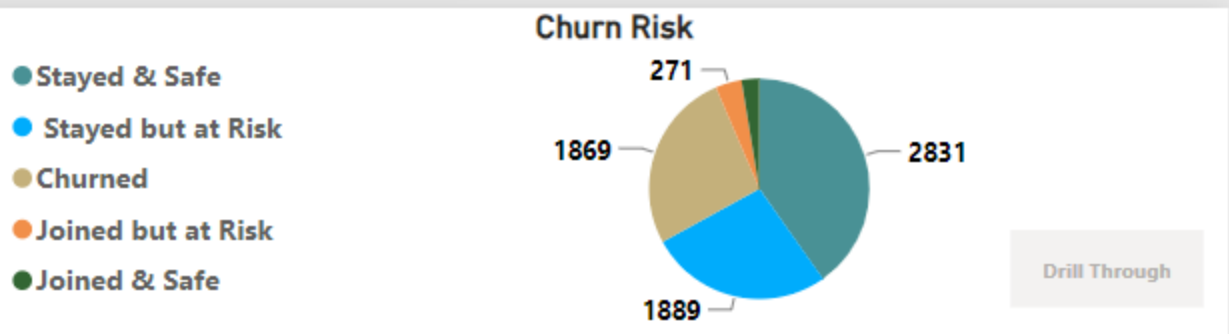
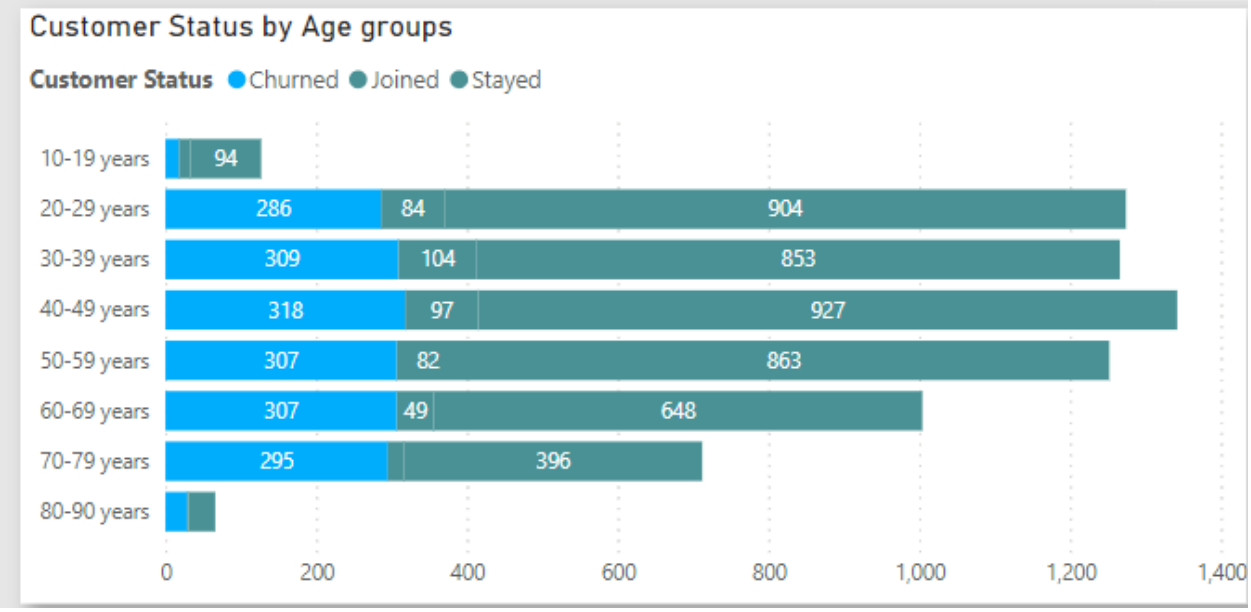


Chart shows competitor performance as the leading cause of churn, followed by dissatisfaction so we need to compare and improve our services to retain customers effectively.

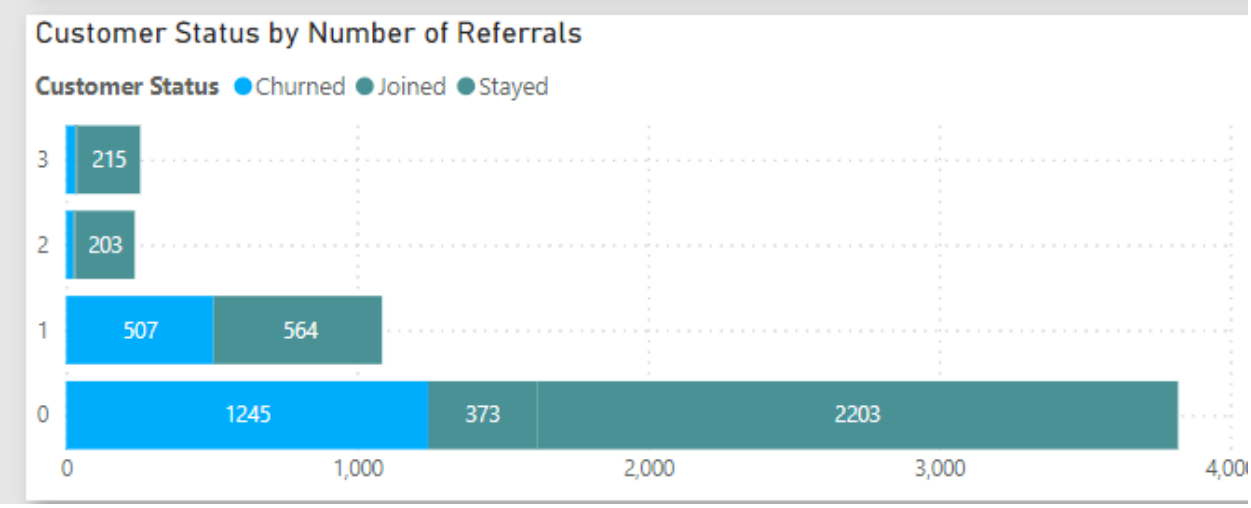


Quantify the count of customers: at risk of leaving, those who have left, and those who have stayed without risk of leaving.

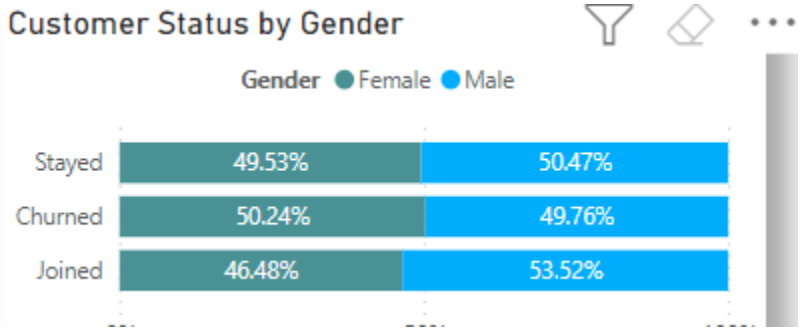
Customer Demographics:



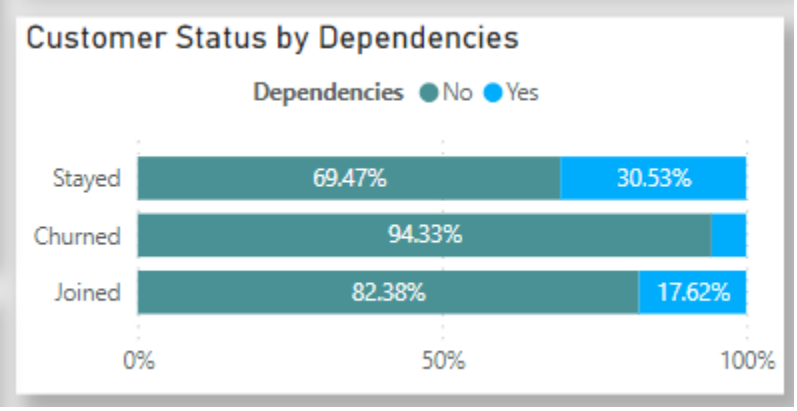
While all age groups show a tendency to churn, customers aged 50+ exhibit a little higher probability; utilize drill-through analysis helps uncover the specific reasons driving this trend.



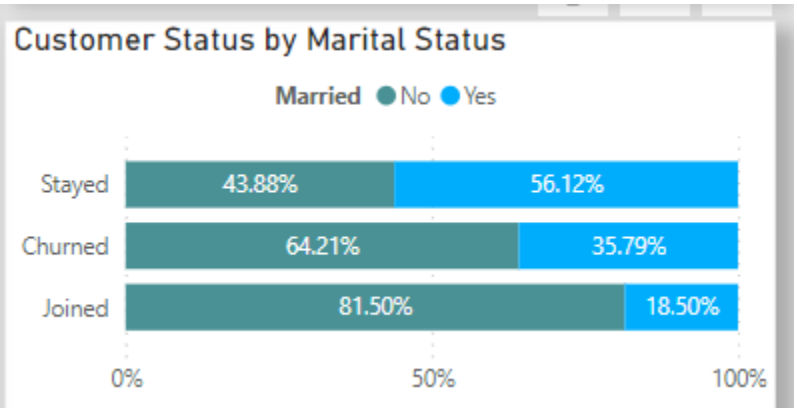
High value customers can be identified as they have higher number of referrals and are more prone to staying with company.We also need to improve our services so more people can refer our company to others.



Gender doesn’t heaven impact on churn behaviour as the total split for all customers was almost 50/50.

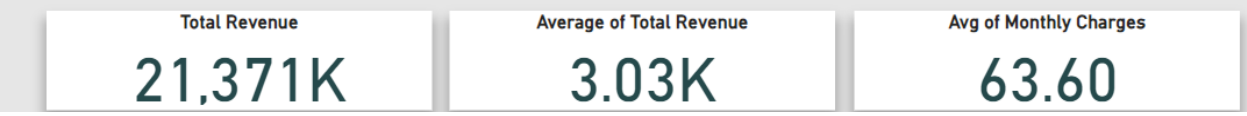


Number of dependents does not have a direct impact on churn behaviour as a large proportion of our customers don’t have any dependants, this is why they have a huge probability of churning or staying



EDA revealed a nearly equal split between married and unmarried customers, indicating that marital status has minimal impact on churn behavior. The chart illustrates a higher probability of both staying and churning regardless of marital status.

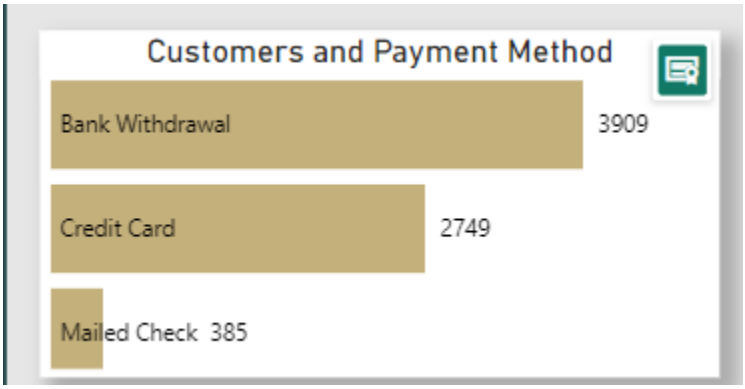
Revenue



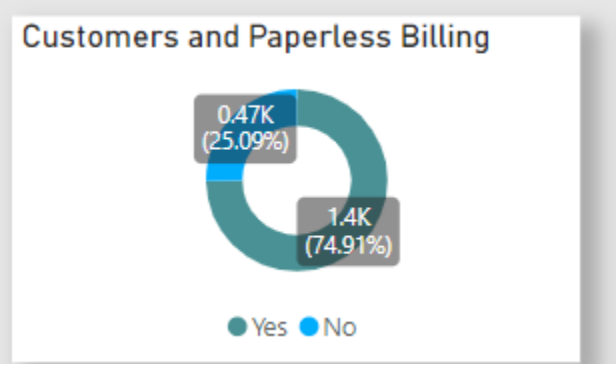
Important metrics such as total revenue collected from customer over the quarter and average monthly charge per customer.

Customer ID	Sum of Total Charges	Sum of Monthly Charge	Sum of Total Extra Data Charges	Sum of Total Long Distance Charges	Sum of Total Revenue
0164-APGRB	8,496.70	114.90	0	3,482.64	11,979.34
8263-QMNTJ	8,425.30	115.55	0	3,443.04	11,868.34
5451-YHYPW	8,443.70	115.75	120	3,232.08	11,795.78
3810-DVDQQ	8,308.90	117.60	0	3,385.44	11,688.90
7569-NMZYQ	8,672.45	118.75	0	2,962.08	11,634.53
3963-RYFNS	8,013.55	116.45	90	3,493.44	11,596.99
9351-HXDMR	8,240.85	110.90	0	3,323.52	11,564.37
5914-XRFQB	8,424.90	115.80	0	3,104.64	11,529.54
0619-OLYUR	8,071.05	111.90	0	3,443.76	11,514.81
Total	**16,060,725.24**	**447,907.55**	**48320**	**5,275,906.10**	**21,371,131.69**

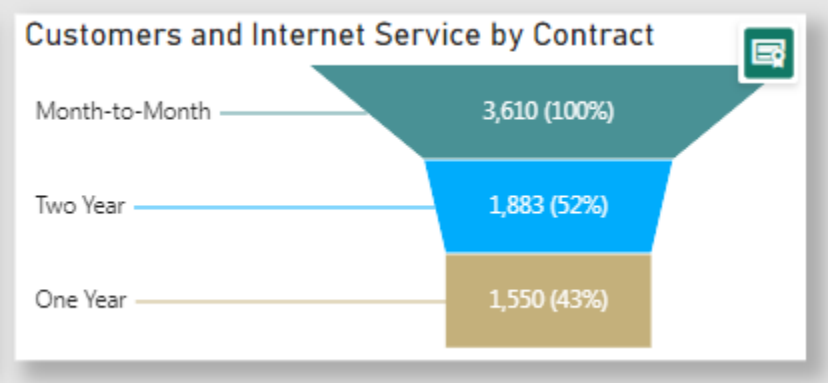
Use filters and slicers to review each customer's charges, helping create targeted marketing plans based on specific needs.



Customers preferred payment method is bank withdrawal, including churned customers which means payment method cannot be directly attributed to churn status.

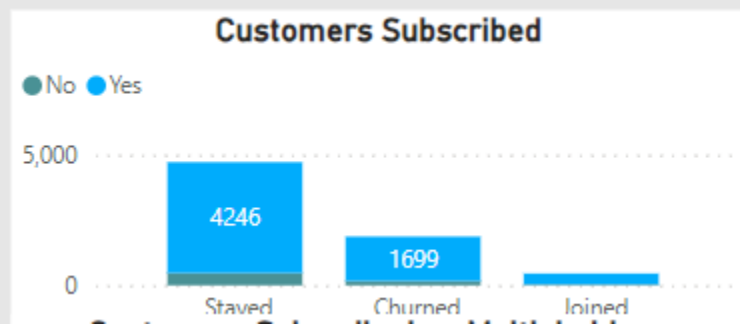


Slicers revealed that a notable portion of churned customers opted for paperless billing, despite the initial split being 58:42 for all customers. This suggests that paperless billing might have a slight impact on the churn rate.

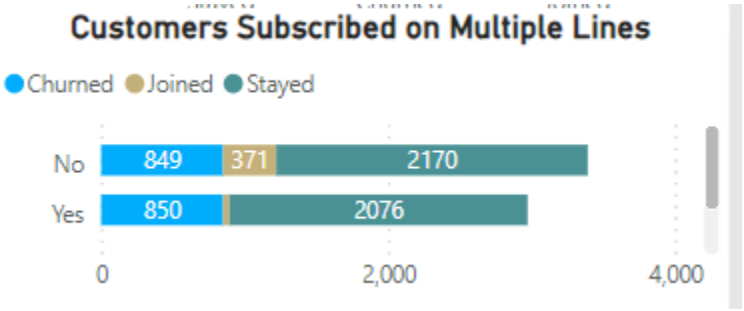


EDA revealed that month-to-month contracts were leading cause of churned customers and this chart proves that

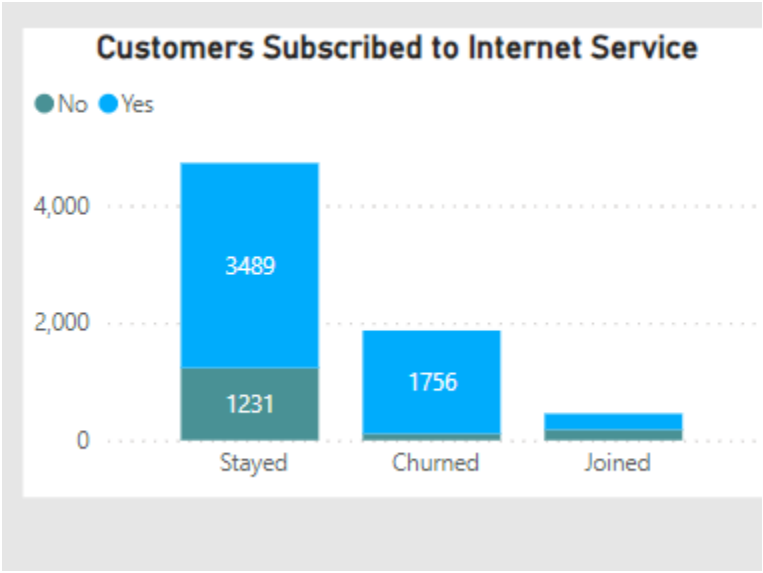
Services



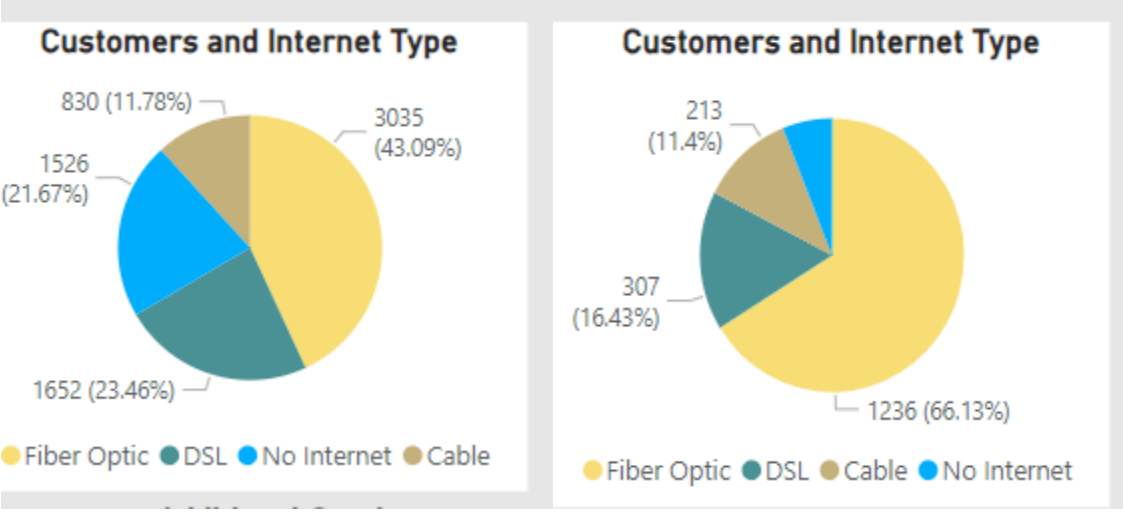
Most customers subscribed to phone service stayed so the phone service is up to standard



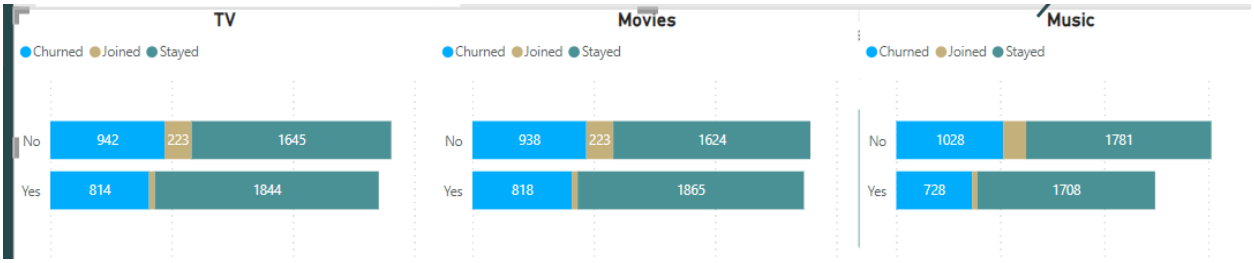
Stayed customers also opted for getting multiple lines (for phone service) which again proves how good phone service is.



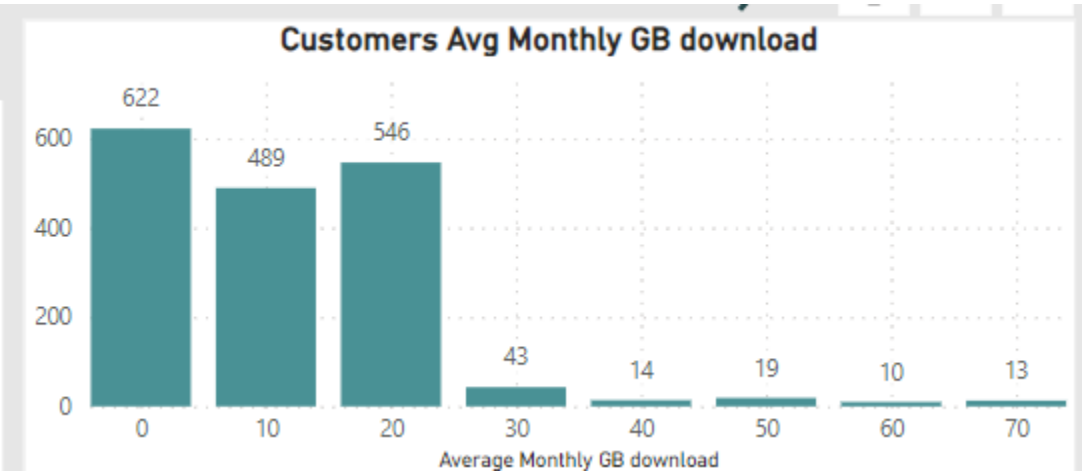
Customers subscribed to general internet service were likely to stay so generally it is also up to standard.



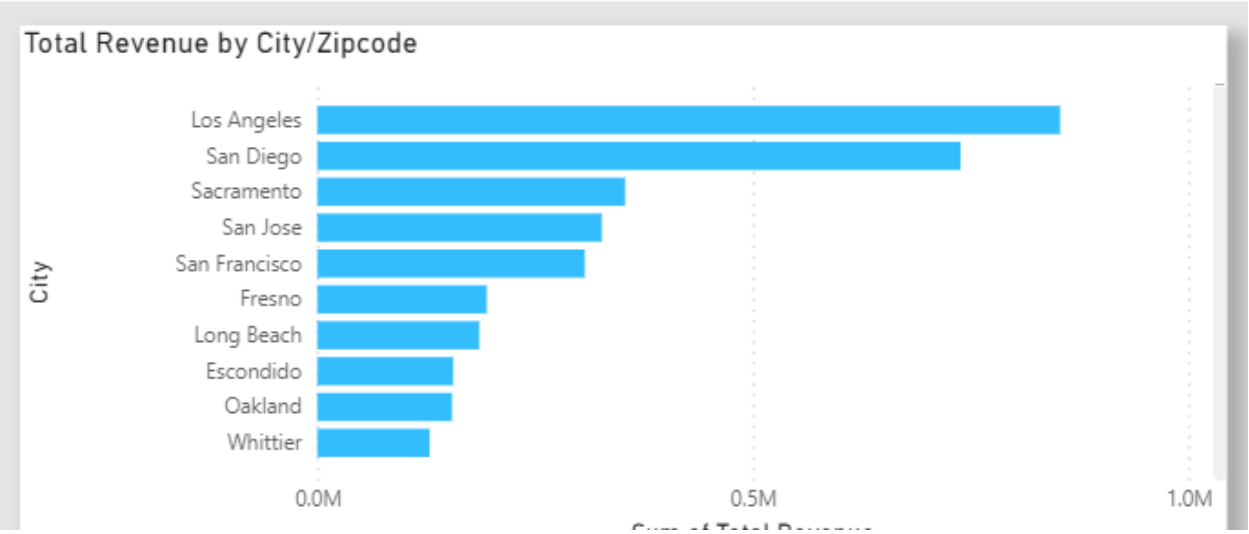
Despite fiber optic having an original share of 43% for internet type for all customers, it drastically increased to 66% for churned customers meaning fiber optic cables need to be made better.



Significant portion of customers churned using these streaming internet services, especially music, which means these services need to be improved upon



EDA showed that original spread of average monthly GB download was rightly skewed and by using the churned slicer here, again spread of churned customers is also right skewed which means monthly gb download does not have a direct impact on customer churning

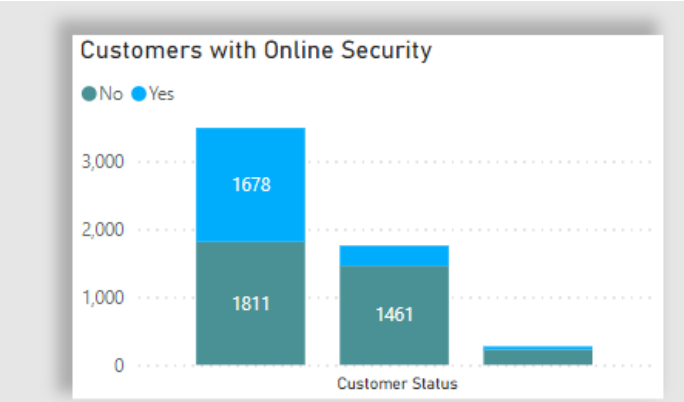


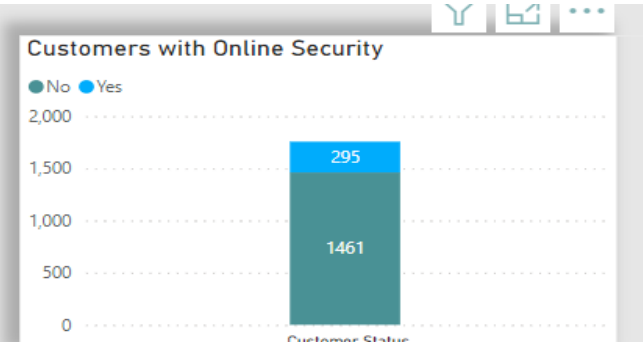
Utilizing slicers allows us to pinpoint high-value customers' locations and identify the cities where we are experiencing the most financial loss due to churned customers.

City	Churned	Joined	Stayed	Total
Los Angeles	78	18	197	293
San Diego	185	7	93	285
San Jose	29	2	81	112
Sacramento	26	6	76	108
San Francisco	31	7	66	104
Fresno	13		48	61
Long Beach	15	2	43	60

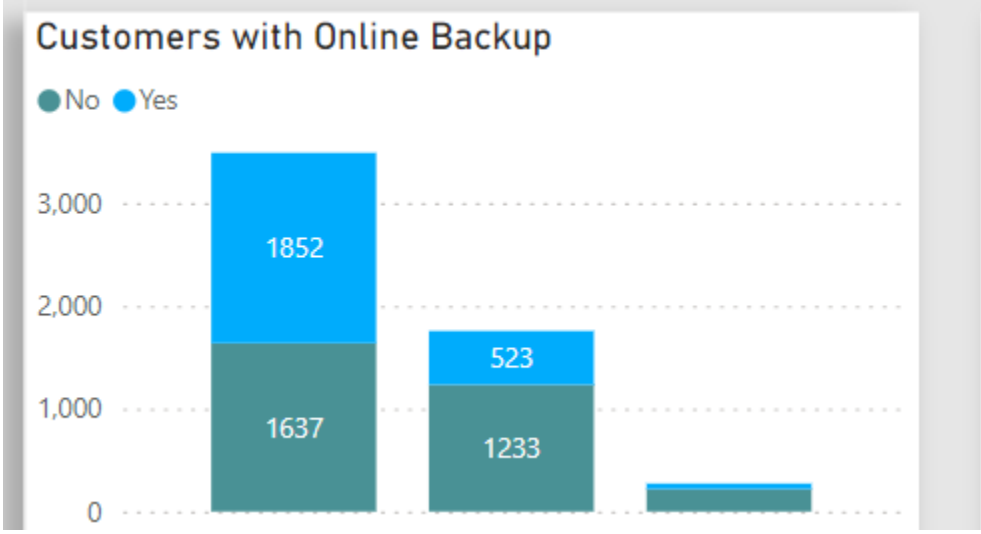
Customer count per city for a general overview.

Additional Services

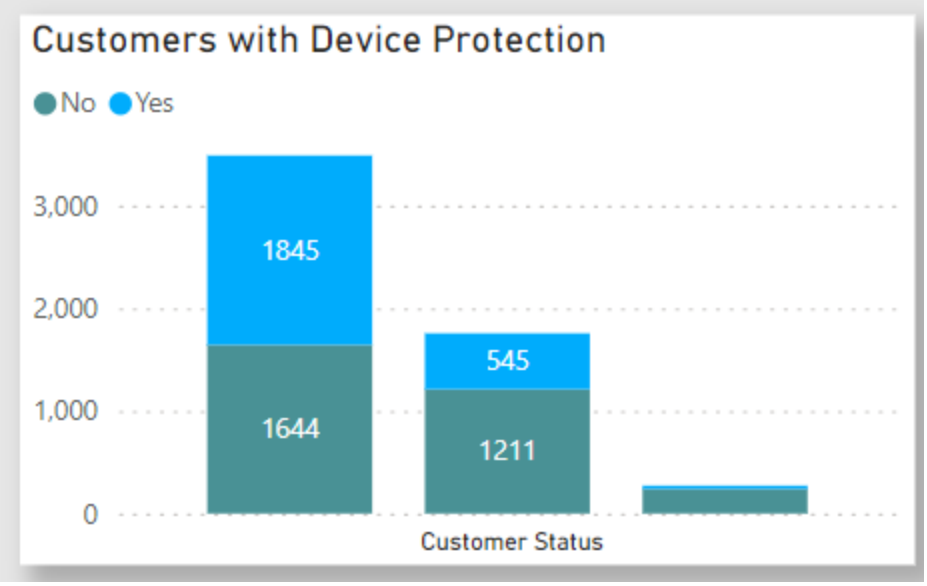




Using the slicer for churned, customers who aren't using online security plan are more likely to churn so we need to advertise this service to them for retention



Churned customers are more likely to not use online backup (1233 vs 523 customers for No: Yes) which means this service needs to be advertised to them.

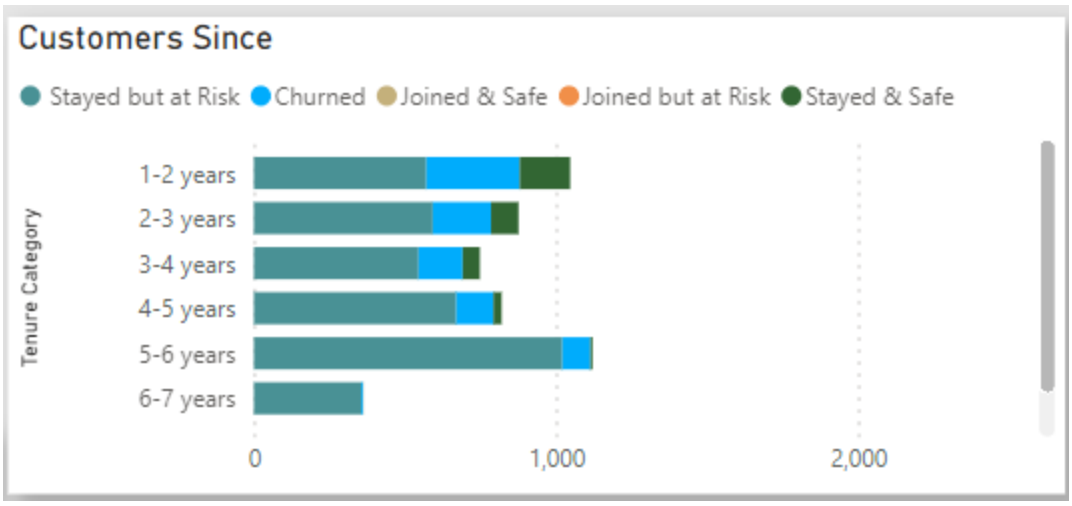


Customers not using device protection are more likely to churn so this service also needs to be advertised to them.

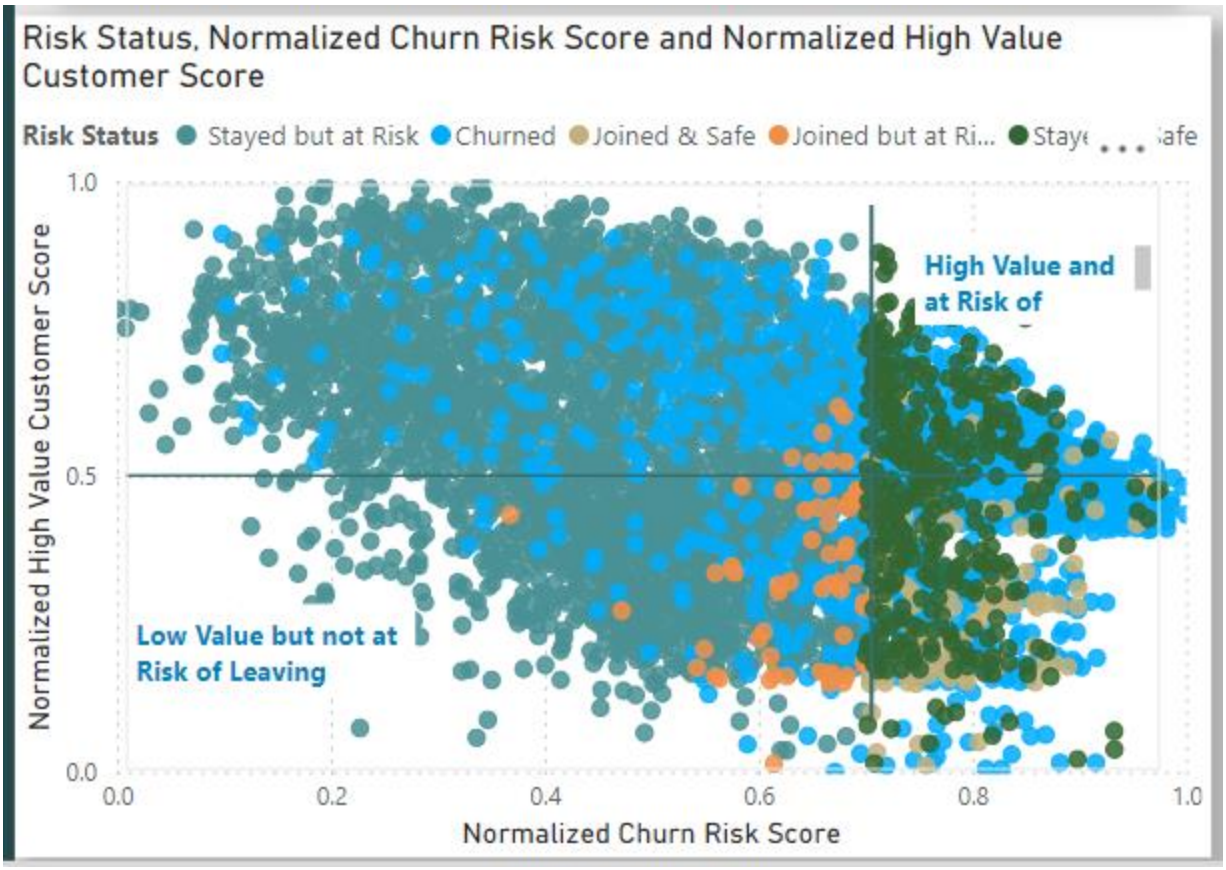
Risk of Churning

Risk Status	Sum of Total Revenue	Count of Customer ID
Stayed but at Risk	16,488,800.01	4058
Churned	3,684,459.82	1869
Joined & Safe	33,603.03	285
Joined but at Risk	20,676.72	169
Staved & Safe	1,143,592.11	662
Total	**21,371,131.69**	**7043**

Find out total revenue generated by each risk category, as evident 3.6$ million dollars has been lost due to customers churning.

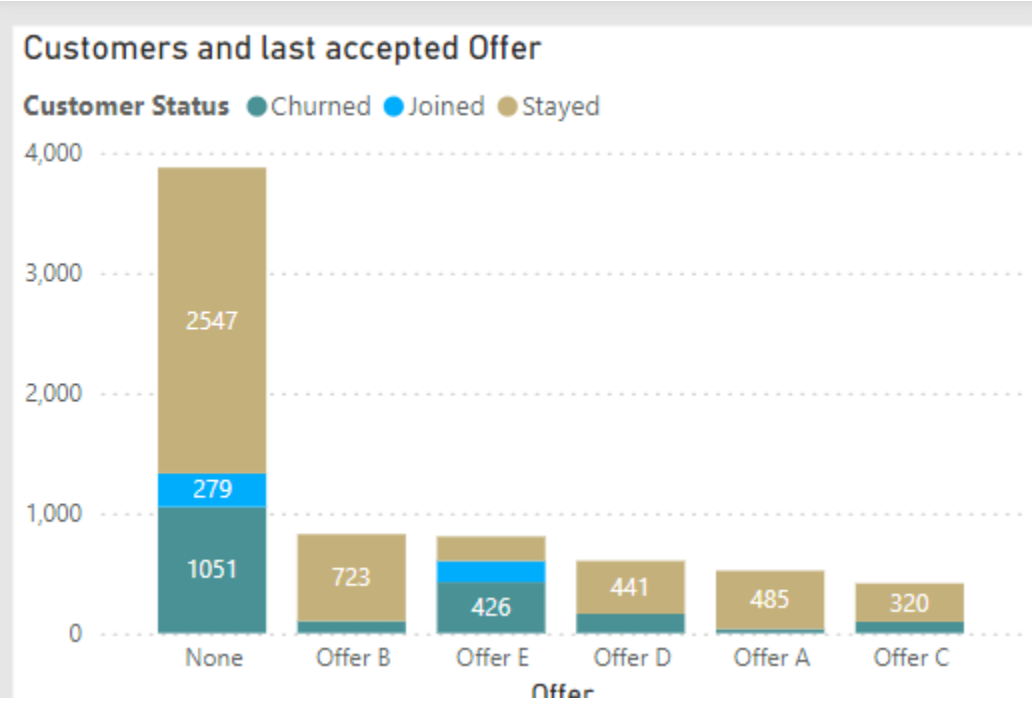


This chart is to show customer loyalty, especially useful for high value customers



The scatter plot assesses customers based on their churn risk and high value scores, facilitating easy identification. The top-right intersection indicates high-value customers

at risk of leaving, while the bottom-left intersection represents low-value customers who are not at risk of leaving.



As evident, offer E did not perform well as it has the highest churn rate compared to other offers.

Final Insights and Conclusions:

The churn analysis reveals several critical insights for strategic decision-making. Firstly, the primary reasons for churn include competitors offering better deals, superior devices, and the attitude of support personnel. These findings underscore the importance of improving our services and products to remain competitive in the market.

Additionally, specific factors like month-to-month contracts, Offer E, lack of online security, and non-utilization of premium tech support significantly contribute to the churn rate. Addressing these areas through targeted strategies can help mitigate churn risk and improve customer retention.

Location-based analysis highlights San Diego as having the highest churn rate, signaling a need for localized retention strategies tailored to address regional challenges.

Interestingly, customer demographics such as gender, marital status, age, and average monthly GB download do not show a significant impact on retention, indicating that retention strategies should focus more on service quality and customer experience.

Furthermore, the analysis identifies customers using fiber optic cable and streaming music services as more prone to churning. Leveraging insights from the churn risk score and high-value score metrics can help segment and prioritize retention efforts effectively, especially targeting high-value customers at risk of leaving.

In conclusion, the report recommends enhancing service quality, refining customer support experiences, and implementing targeted marketing strategies based on churn risk and high-value scores. These strategies aim to reduce churn, retain valuable customers, and ultimately improve business performance and customer satisfaction.

Team Member Contributions:

Ayesha Nayyer: Primarily focusing on EDA and Report Writing,

Humera Raheel: Primarily focusing on Information Gathering, Power BI Dashboard and Design, and Excel Analysis