|  |  |
| --- | --- |
| Capstone project  Telecom Customer Churn Analysis | Abstract  Reports on the EDA and modelling findings using a recent telecommunication dataset.  Jayden LI  GA Data Analytics Immersive |

[1. Executive Summary 2](#_Toc111770461)

[2. Problem, Goals and Audiences 3](#_Toc111770462)

[3. Patterns, trends and insights 3](#_Toc111770463)

[3.1 Methodology 3](#_Toc111770464)

[3.2 Customer Dataset Overview 4](#_Toc111770465)

[3.3 Geographical Analysis 5](#_Toc111770466)

[3.4 Customer Segmentation 6](#_Toc111770467)

[3.5 Data Modelling 8](#_Toc111770468)

[4. Recommendations and next steps 8](#_Toc111770469)

[5. Appendix – list of supplementary documents 9](#_Toc111770470)

[6. References: 9](#_Toc111770471)

1. Executive Summary

* This project set out to reveal insights behind customer churn at a U.S. telecommunication company in California.
* The dataset contains over 7,000 customer information and was made available to the public for free download.
* This analysis went through multiple phases including understanding data, preparing data, describing data, modelling data, as well as reporting for business decision-making.
* The analysis revealed interesting patterns behind the customer churns, including geographical differences that may require adjusting future performance metrics.
* The analysis revealed limited correlation or relationships between given variables, which suggests the need for further details with a more thorough scientific investigation.

1. Problem, Goals and Audiences

This project set out to analyse a full set of customer churn data from an anonymous telecommunication company located in California, the U.S. The data is reportedly released in Q2, 2022. The original data set and challenge brief can be found at: [link](https://www.mavenanalytics.io/blog/maven-churn-challenge).

The original problem was stated as how to ‘improve retention by identifying high-value customers and churn risks’ using the dataset. This research project adopted similar goals including:

* Identifying common characteristics of churned customers.
* Segment customers based on criteria such as revenue, loyalty, etc.
* Recommend future marketing initiatives and managerial practices.

The research is conducted in part of an intensive data analytics training program. The intended audience is both technical and non-technical. In full, this report should be read together will supplementary documents including SQL and Python codes in a JupyterNotebook, a Tableau presentation deck and a data dictionary excel file (Full list see Appendix).

1. Patterns, trends and insights
   1. Methodology

The project was conducted in several steps, including understanding data, checking, cleaning and preparing data, describing and comparing data, analysing correlations and relationships in data, data modelling and evaluation, and lastly, reporting and recommending insights. First, the dataset was exported using python Pandas to the researcher’s SQL server to construct a database. SQL allows the researcher to extract and view the dataset more timely.

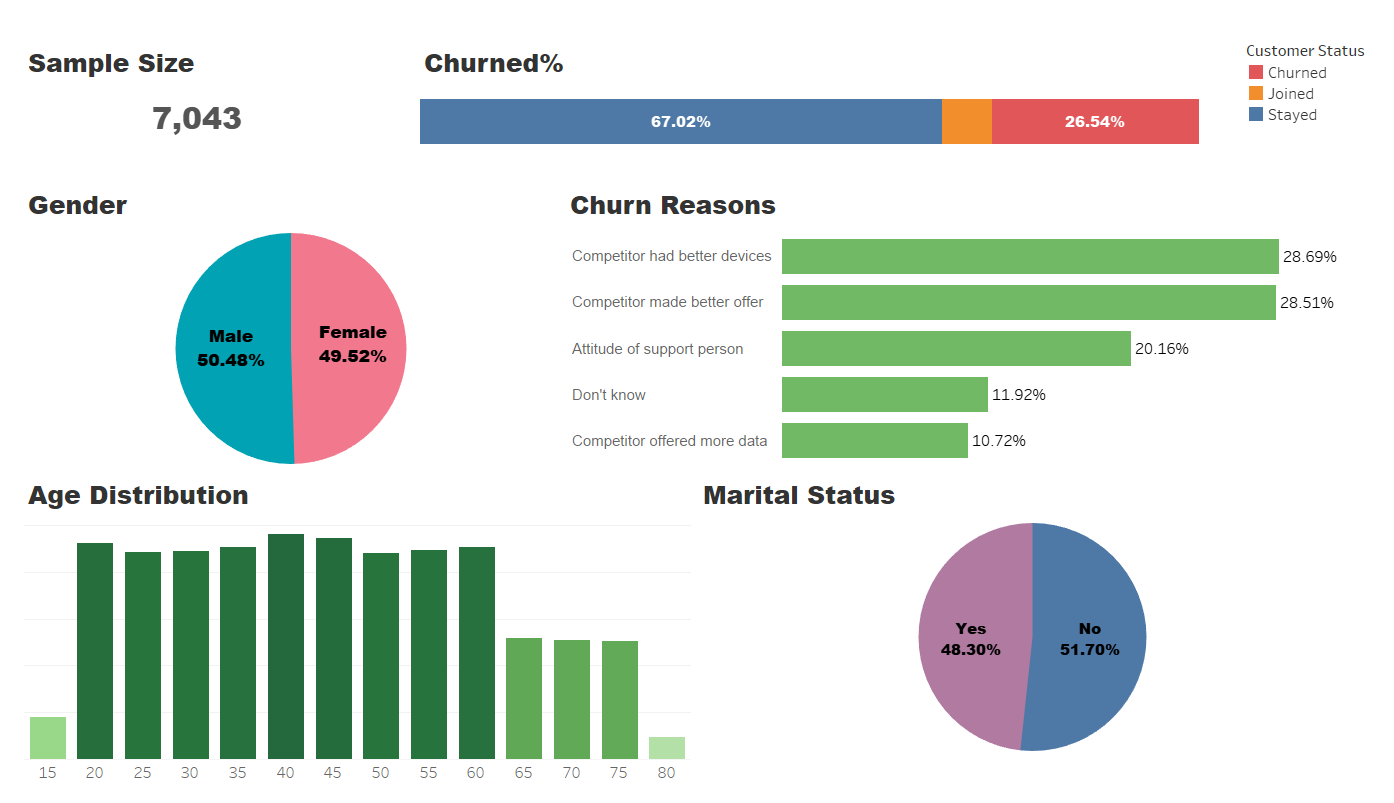
Next, the data was re-exported to Python for a more complicated, yet preliminary visualization to find correlations. Cleaning was also a part of the process, and a second version (V2) was exported for further analysis in Tableau. Tableau was used to generate visualizations and a presentation deck. The tool is not suitable for data modelling and prediction. So lastly, Python was used again for a KNN modelling practice.

From the above exploratory data analysis, this research did not find significant correlations or relationships between the distinctive variables contained within the dataset. This means that further information or scientific analysis may be required to construct any concrete predictive model. Nevertheless, the research does reveal some key insights from describing and importantly, successfully visualizing multi-dimensions using Tableau.

* 1. Customer Dataset Overview

The dataset contains 7,043 non-repeated customer IDs (‘customers’). Assumably a sampled data, the gender distribution appears to be even among all customers with each around 50% (see Figure 1, a presentation draft screen capture). We have found similar results in terms of marital status.

Figure 1 Customer Overview (draft, final version see presentation deck)

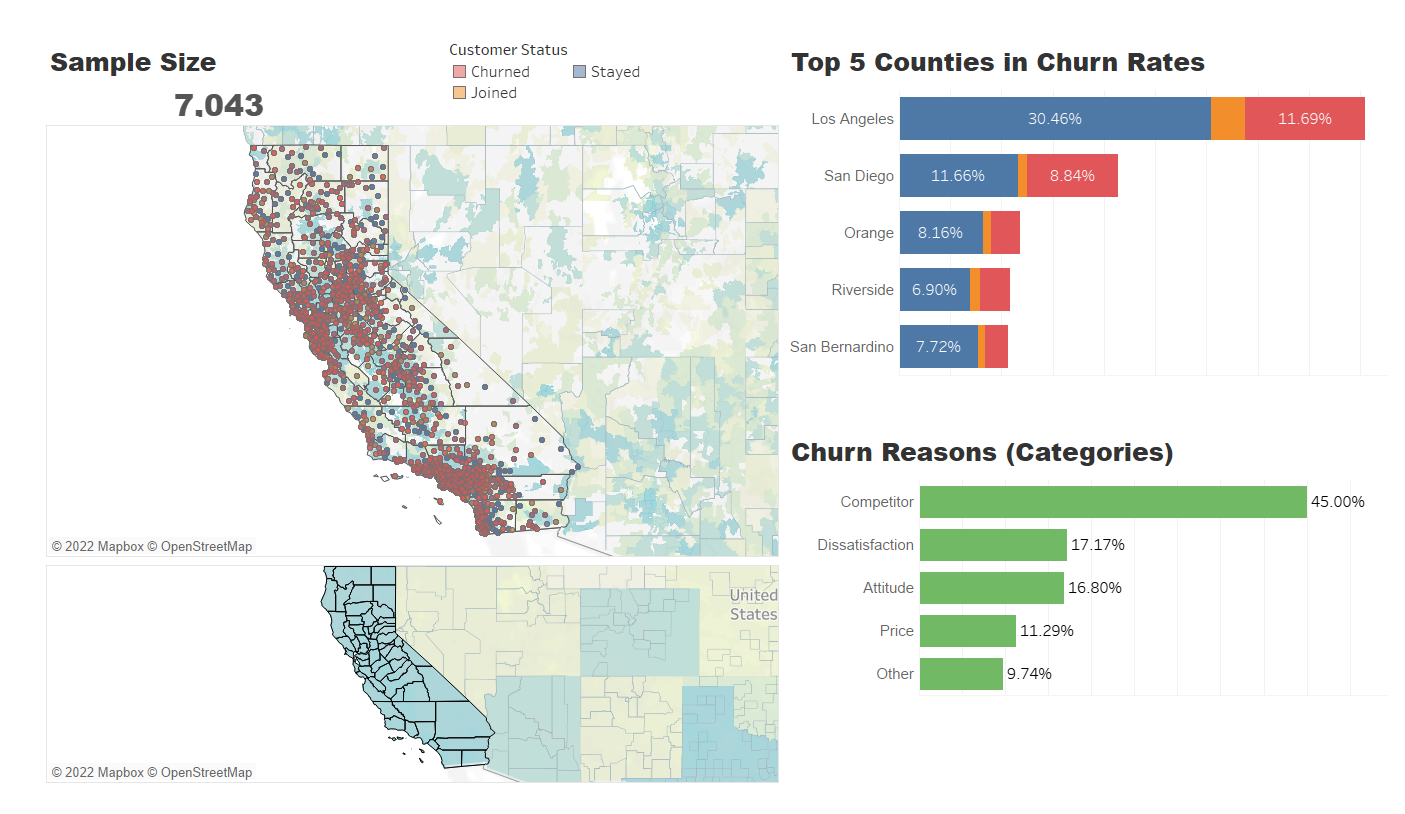


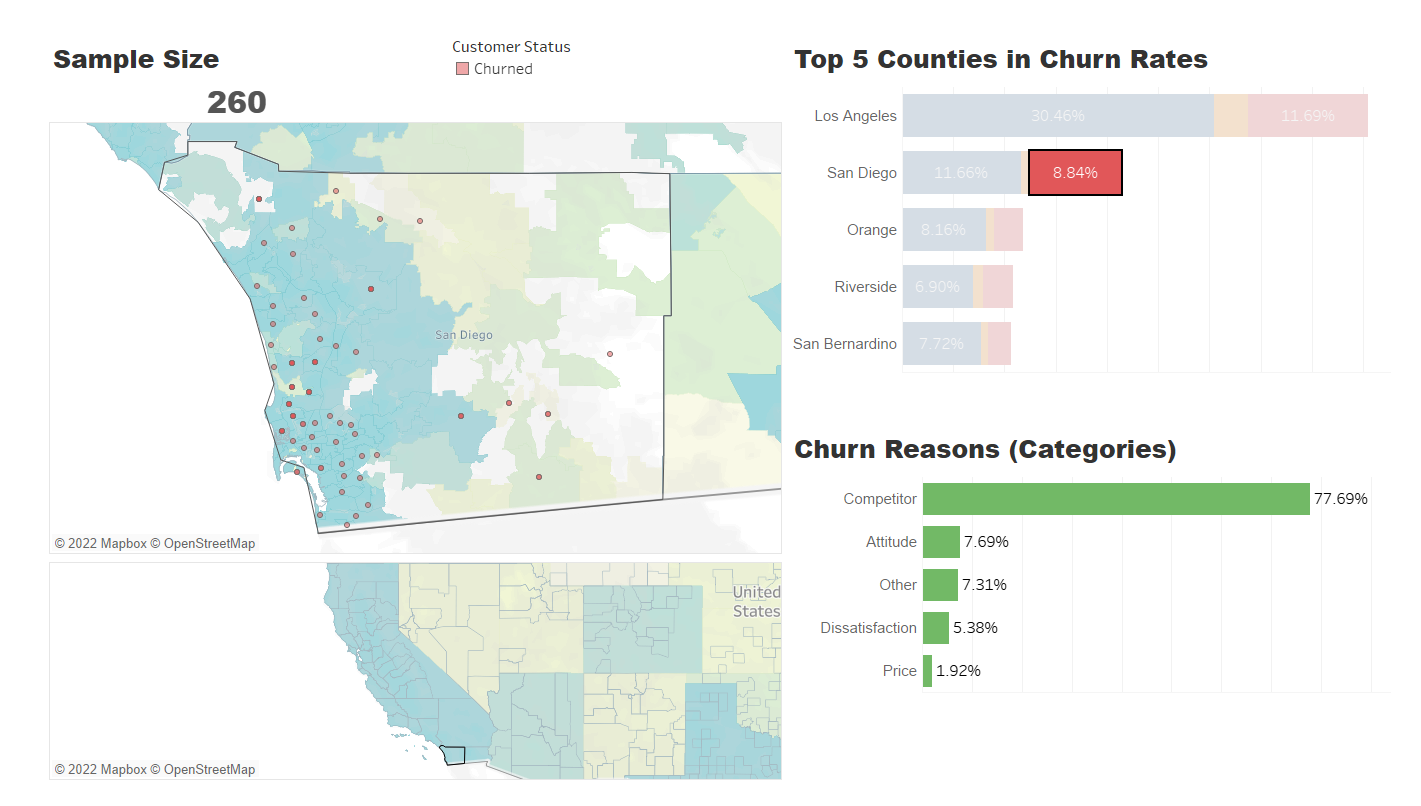
For those churned, both genders share similar reasons in terms of why turned away from the service provider. Better devices offers from competitors, and support person attitude are the top three churn reasons. However, the map visualization later would indicate that there are geographical differences between these reasons.

* 1. Geographical Analysis

The geographical mapping in Tableau provided interesting and intuitive insights regarding the customer churn landscape. In the visualization phase, a ‘county’ feature was added to the dataset based on an external source including county names and their zip codes. As shown in Figure 2, the analysis reveals that the top two most populated customer counties are Los Angele and San Diego, both locates in south California.

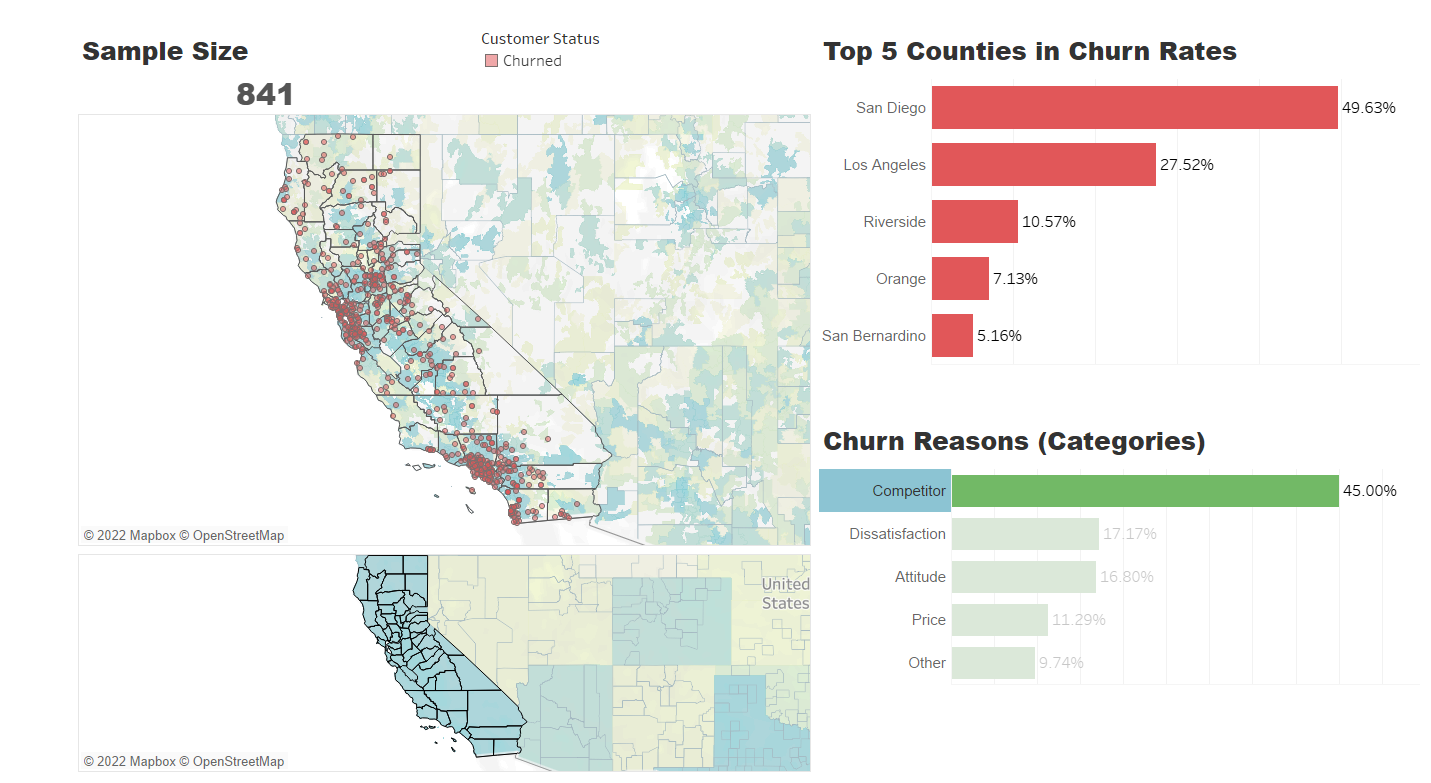
Figure 2 Top States in terms of Churn (Draft)





Interestingly, as both carry a significant number of stayed customers, they are also the top two counties with the highest rate of churned customers. In particular, San Diego county has a higher churn rate in comparison to Los Angeles. Whereas on the other hand, in terms of churn reasons (Figure 3), selecting the ‘Competitor’ category will reveal which counties the company faces fierce competition. Sacramento county, close to the northeast of San Francisco, is a sub-area with high competition. In terms of dissatisfaction and attitude, however, the churned customers are less concentrated beyond Los Angeles and San Francisco region. This would suggest differentiated performance evaluation across counties.

Figure 3 Locations with high competition (Draft)

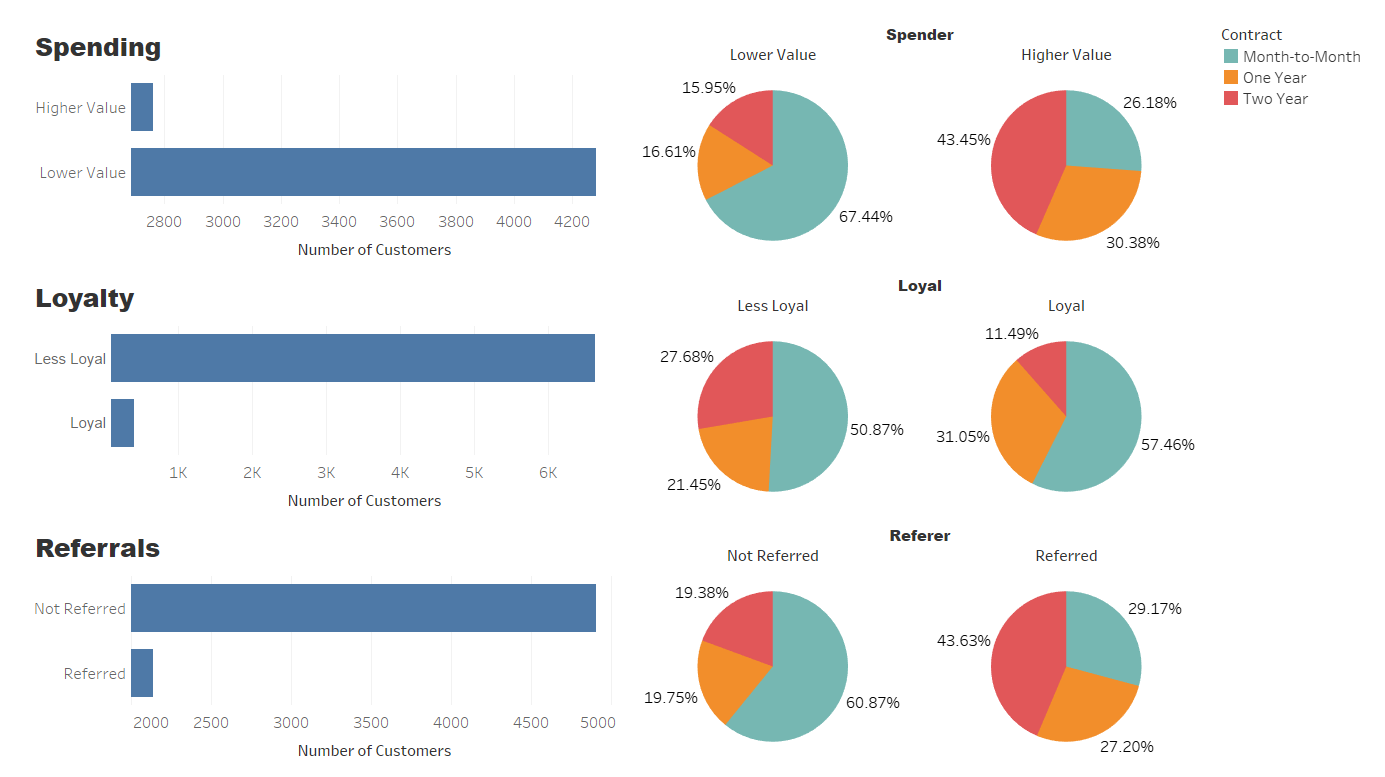


* 1. Customer Segmentation

This analysis separates customers based on three criteria, namely: total contributed revenue, customer loyalty and their networking ability (referrals). Based on whether a customer contributed above-average revenue, has above average tenure or has an above-average number of referrals, he or she would be tagged as high/low value, loyal/less loyal, and(or) referred/not referred.

Analysis (Figure 4) shows that the six groups of customers share very different traits in terms of contract terms and offers they have. For example, we can see that except for ‘Loyalty’, the spenders and network lovers subscribe more to yearly contracts (red and orange) than month-to-month ones (blue). Customers with less revenue and no referral, tend to sign month-to-month contracts. So the loyal customers in the middle seem to be a potential target for conversion.

Figure 4 Customer groups and contract patterns (Draft)



In an even more evident manner, we can see (Figure 5) that in terms of offer plans, the preferable customers (right side) tend to use Offer B and A much more than customers on the left side. While the latter appears to be more with Offer E, which is hardly observable among valued and loyal customers. This may indicate the plan has traits that are commonly avoided by valued and loyal customers. This is valuable insight for the marketing team, if they are considering converting less loyal or valued customers to preferable ones.

Figure 5 Customer groups and offer use pattern (Draft)



* 1. Data Modelling

This analysis used Python modules and conducted a K-nearest Neighbour modelling. Detailed codes and explanations can be found in the JupyterNotebook file submitted alongside this report.

1. Recommendations and next steps

Based on the above findings, this analysis puts forward the following recommendations:

1. The overall churn rate is pretty high as a result of market competition both in products and prices. The company should focus on maintaining and developing a core competitive edge and achieve brand differentiation.
2. Visualization reveals that there are geographical differences in terms of customer pain points. Recommend introducing a dynamic and localised performance evaluation framework.
3. Valued customer groups have a different contract and offer plan preferences. Despite the lack of detailed information, the analysis recommends considering what made them attractive to those customers. Findings can be used to tailor future contracts and offer plans.
4. Appendix – list of supplementary documents
5. Technical Analysis Report (this document)
6. Technical Analysis Presentation Deck
7. JupyterNotebook with SQL and Python codes
8. Data Dictionary in MS excel format
9. Datasets used for the analysis
10. References:

*Data source 1:*

https://www.mavenanalytics.io/blog/maven-churn-challenge

*Data source 2:*

https://www.kaggle.com/datasets/imakash3011/customer-personality-analysis