OVERVIEW

Full Analysis available here

Data obtained from: Kostas Diamantaras. (2020). Customer Churn Prediction 2020. Kaggle. https://kaggle.com/competitions/customer-churn-prediction-2020

Five questions will guide your case study

What type of company does your client represent, and what are they asking you to accomplish?	Customer Support department of Telecommunications company. They would like to increase retention rates.
What are the key factors involved in the business task you are investigating?	Obtaining, cleaning and analysing customer data. Segmenting the accounts and finding trends in churned customers, using the trends and correlations to implement a process to reduce churn rate.
What type of data will be appropriate for your analysis?	All customer data, quantitative and qualitative.
Where will you obtain that data?	Internally from the client, in this hypothetical case study, it is obtained from Kaggle.
Who is your audience, and what materials will help you present to them effectively?	The audience is the Director of GTM, Customer Support, Sales and Retentions teams. PPT with graphs and a retentions process and dashboards for ongoing use.

Report checklist

1. A clear statement of the business task you have selected to investigate Completed
2. A description of all data sources used Completed
3. Documentation of any cleaning or manipulation of data Completed
4. A summary of your analysis Completed
5. Supporting visualizations and key findings Completed
6. Based on what you discover, a list of additional deliverables you think would be helpful to include for further exploration Completed
7. Your top high-level insights based on your analysis Completed

ASK

Guiding questions

● What topic are you exploring?	Churn Rate in telecommunication customers
● What is the problem you are trying to solve?	Decrease churn rate
● What metrics will you use to measure your data to achieve your objective?	Churn rate, correlation rate between churn and account metrics
● Who are the stakeholders?	Director of GTM, Customer Support, Sales and Retentions teams
● Who is your audience?	Churn risk customers
● How can your insights help your client make decisions?	The insights will guide the retentions proactive response to limit voluntary churn

Key tasks

1. Identify the business task

Decrease churn rate by using customer data to identifying at risk customers and why they are at

risk. Using these insights to guide a proactive process to be used by the retentions team.

2. Determine key stakeholders

Director of GTM, Customer Support, Sales and Retentions teams

3. Choose a dataset

Kostas Diamantaras. (2020). Customer Churn Prediction 2020. Kaggle.

https://kaggle.com/competitions/customer-churn-prediction-2020

4. Establish metrics

Churn rate

Deliverable

A clear statement of the business task you have selected to investigate
Decrease churn rate by using customer data to identifying at risk customers through statistical analysis. Using these insights to guide a proactive process to be used by the retentions team.

PREPARE

Once you have decided on a dataset, gather and store the data appropriately, clean the data, and make sure it is reliable and error-free. Document your process, as clients often ask to see both raw and cleaned data. Now, prepare your data for analysis using the following Case Study Roadmap as a guide:

Guiding questions

● Where is your data located?	Big Query and locally
● How is the data organized?	As a dataframe. Each tuple is a customer and attributes are relating to customer account usage and type.
● Are there issues with bias or credibility in this data? Does your data ROCCC?	It is a hypothetical dataset sourced from a well known reputable location. It is within the last 3 years and has many attributes. The dataset has been used for a competition in kaggle with 629 entries. The data is csv format with a thorough description. It has been cited as coming from a scholar who is a co author on ‘A comparison of machine learning techniques for customer churn prediction, Simulation Modelling Practice and Theory’.
● How are you addressing licensing, privacy, security, and accessibility?	Citing the data source every where it is used. Saving copies locally and on cloud and ensuring permissions are for only myself.
● How did you verify the data’s integrity?	I cleaned and validated the data by checking for nulls, duplicates, type, code, range, format and length and checking for outliers. All of the tests returned few (or no) results.
● How does it help you answer your question?	The data is directly telling us which customers churned and which didn’t and what they have in common and the differences. Analysing the differences will help us identify churn risk customers
● Are there any problems with the data?	There are some attributes that are missing that could be useful, such as; month/day of cancellation, cs call reason, year on year data and most importantly, customers stated reason for churn.

Key tasks

- Completed Download data and store it appropriately.
- Completed Identify how it's organized.
- Completed Sort and filter the data.
- Completed Determine the credibility of the data.

Deliverable

A description of all data sources used
Kostas Diamantaras. (2020). Customer Churn Prediction 2020. Kaggle. https://kaggle.com/competitions/customer-churn-prediction-2020

PROCESS

Then, process your data for analysis using the following Case Study Roadmap as a guide:

Guiding questions

● What tools are you choosing and why?	SQL, big query saved to Jupiter for ease of sharing and because its nearly 5000 rows (excel would also be a valid option). - For cleaning and documentation of cleaning process - For analysis Tableau for visualisation - It allows for dynamic dashboards to share with teams - Graphs for presenting
● How have you ensured your data's integrity?	1. Checked the sources and reviewed the data 2. Cleaned the data 3. Reviewed the cleaned data
● What steps have you taken to ensure that your data is clean?	I ran the same steps I used to clean the data, after I cleaned it to ensure nothing was missing; Checking for nulls, duplicates, type, code, range, format and length and checking for outliers.

● How can you verify that your data is clean and ready to analyze?	By running the validation checks regularly, at each point of the process; Checking for nulls, duplicates, type, code, range, format and length and checking for outliers.
● Have you documented your cleaning process so you can review and share those results?	Yes, it is documented in the notebook alongside the sql used to clean and validate the data.

Key tasks

- Completed Check the data for errors.
- Completed Choose your tools.
- Completed Transform the data so you can work with it effectively.
- Completed Document the cleaning process.

Deliverable

Documentation of any cleaning or manipulation of data
Github Churn Rate Cleaning

ANALYSE

Now that your data is stored appropriately and has been prepared for analysis, start putting it to work. Use the following Case Study Roadmap as a guide:

Guiding questions

● How should you organize your data to perform analysis on it?	The data should be in a dataframe for analysis with SQL And Tableau.
● Has your data been properly formatted?	Yes, some of the variables have been grouped for analysis and boolean values transformed to binary for ease of analysis.
● What surprises did you discover in the data?	The significant impact that account cost has on Churn Rate. I didn't expect the higher paying accounts to have such a high churn.
● What trends or relationships did you find in the data?	There were trends with churn rate and increased CS calls, increased account length and increased yrr.
● How will these insights help answer your business questions?	These can help to identify future accounts at churn risk.

Key tasks

- Completed Aggregate your data so it's useful and accessible.
- Completed Organize and format your data.
- Completed Perform calculations.
- Completed Document your calculations to keep track of your analysis steps.
- Completed Identify trends and relationships.

Deliverable

A summary of your analysis
Github Churn Rate Analysis Github Churn Prevention Presentation Factors that can lead to an increased churn risk include: - Customer service calls > 4 - Account length > 16 years - Yrr > $80.00 Factors with above average churn rate: - Having international plan - Yrr $60-$80

SHARE

After analysis and gaining insights into your data, create visualizations to share your findings. You will be presenting to your client and other stakeholders, so visuals should be sophisticated and polished in order to eectively communicate your insights. Use the following Case Study Roadmap as a guide:

Guiding questions

● Were you able to answer the business question?	Yes
● What story does your data tell?	It tells us churned customers have certain account variables in common
● How do your findings relate to your original question?	This can help us to predict which customers are at churn risk, so we can use preventative care to prevent the churn
● Who is your audience? What is the best way to communicate with them?	Presenting these findings in dashboards and PPT will help the GTM directors/managers to understand the impact
● Can data visualization help you share your findings?	Yes, the graphs show a compelling story
● Is your presentation accessible to your stakeholders?	Yes, the dashboards are accessible for all stakeholders

Key tasks

- Completed Determine the best way to share your findings.
- Completed Create effective data visualizations.
- Completed Present your findings.
- Completed Ensure your work is accessible.

Deliverable

Supporting visualizations and key findings

Refer to:

- Github Churn Prevention Presentation
- Tableau Story

ACT

It's time to act on your findings. Organize the deliverables you created, including your top high-level insights based on your analysis. Use the following Case Study Roadmap as a guide:

Guiding questions

●What is your final conclusion based on your analysis?	Churn Rate can be decreased by contacting at risk customers and resolving outstanding issues, discussing their current plan and costs and tailoring the account for their individual needs
●How could your team and business apply your insights?	They can have the retentions team implement an outreach process that includes recommended churn reduction measures
●What next steps would you or your stakeholders take based on your findings?	Implement the process, Monitor the process and churn rate, review the process and gather more data for future analysis
●Is there additional data you could use to expand on your findings?	Yes, time series and more extensive data points

Key tasks

- Completed Create your portfolio.
- Completed Add your case study.
- Completed Practice presenting your case study to a friend or family member

Deliverable

Your top high-level insights based on your analysis Based on what you discover, a list of additional deliverables you think would be helpful to include for further exploration

Churn rate is highest among the accounts that have the greatest total charges, the longest account length and the most customer service calls. Speaking to customers with these attributes and addressing their issues and maximising their cost/value could help reduce churn rate.

Deliverables for future exploration include:

- Gather time series data
- Gather Qualitative data: reason for CS call, reason for churn, month subscription started, month churned
- Gather Quantitative data: NPS score, how many times X feature is used
- Gather Boolean variables: do they have an overdue balance, have they churned and resigned