

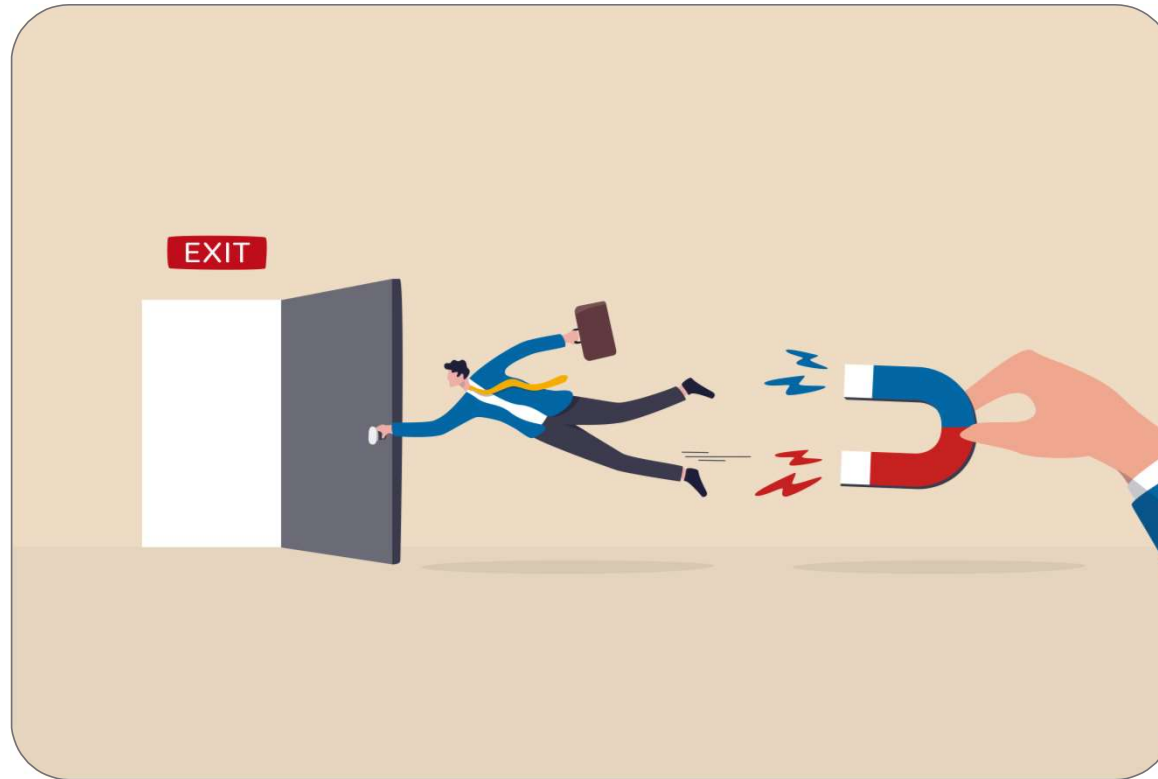
# WELCOME!

- Master degree in Physics
- Bachelor Degrees in Physics and Computer Science
- Currently working as Data Scientist for H2O.ai
- +10 years of experience as Data Scientist
- Worked in finance, telecommunications, banking and retail
- Founder of Closter (Webinars and Podcasts about Data Science)
- Teacher at Berkeley, Dartmouth, Wharton and Columbia
- Creator of more than 120 articles about data science and machine learning
- LinkedIn Top Voice in Data Analytics (2018)
- Top 100 Educator by GFEL (2020)



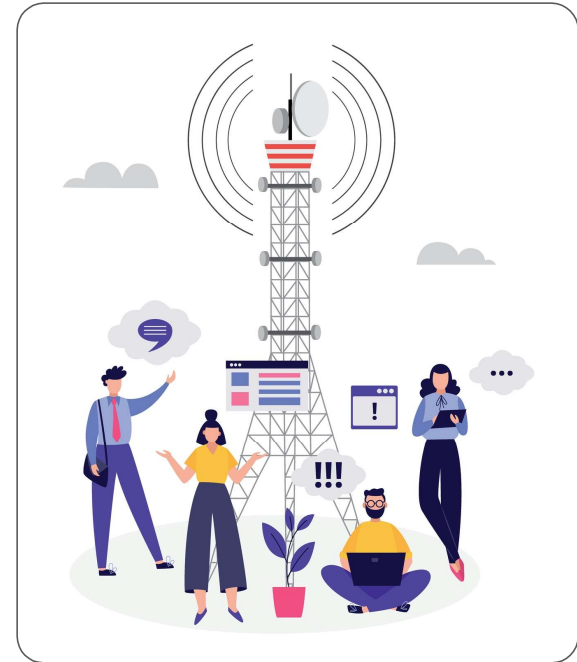
Favio Vázquez

## CASE STUDY: CHURN ANALYSIS FOR A TELECOM COMPANY



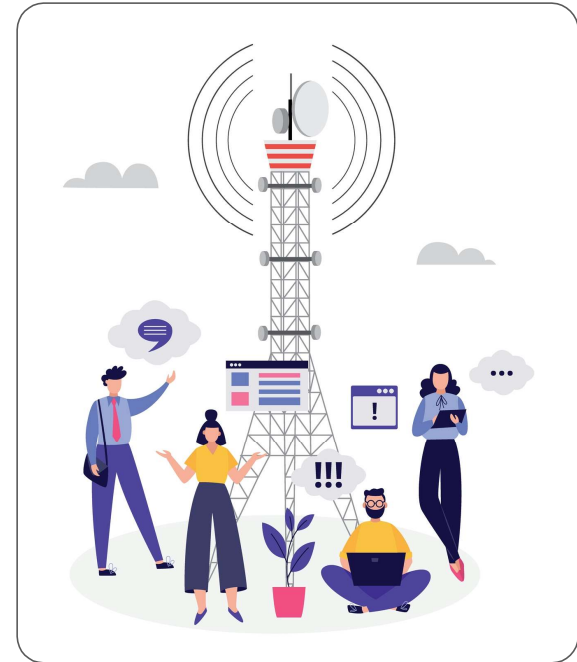
# PROBLEM STATEMENT

- You are working as an analyst in a telecom industry company that provides home phone and internet services to 7,043 customers across Southern California (in the San Diego area)
- Apart from phone and internet services, the company also provides other services such as online security, online backup, device protection and many other services that you will see going forward
- A new competitor has moved into the area and has started stealing your company's customers, leading to an issue known as customer churn
- To retain its existing customers, your company has decided to dig deeper into what is causing customers to leave the company



# BUSINESS OBJECTIVES

- Understand which of the services are not performing well so the product team can improve the quality of the service and ensure that the existing customers are happy with those services
- Identify which variables are affecting the customer churn regarding the existing data from the accounts
- Identify the services that are being received well by the customers so the company can use these services to attract new customers
- Identify the high-value customers so the company can give them a premium membership to retain them for as long as possible



## VISITING OUR DATA

### Files:

○ customer.csv



CustomerID	Gender	Senior Citizen	Partner	Dependents
3668-QPYBK	Male	No	No	No
9237-HQITU	Female	No	No	Yes

○ cust\_loc.csv



Cust_ID	State	Latitude	Longitude	Zip Code
3668-QPYBK	California	33.96413	-118.273	90003
9237-HQITU	California	34.05928	-118.307	90005

○ cust\_account.csv



Account_id	Tenure	Contract	Payment Method	Paperless Billing	Monthly Charges	Total Charges
3668-QPYBK	2	Month-to-month	Mailed check	Yes	53.85	108.15
9237-HQITU	2	Month-to-month	Electronic check	Yes	70.7	151.65

○ cust\_services.csv



Cust_ID	Phone Service	Multiple Lines	Internet Service	Online Security	Online Backup	Device Protection	Tech Support	Streaming TV	Streaming Movies
3668-QPYBK	Yes	No	DSL	Yes	Yes	No	No	No	No
9237-HQITU	Yes	No	Fiber optic	No	No	No	No	No	No

○ cust\_churn.csv



Id	Churn
3668-QPYBK	Yes
9237-HQITU	Yes

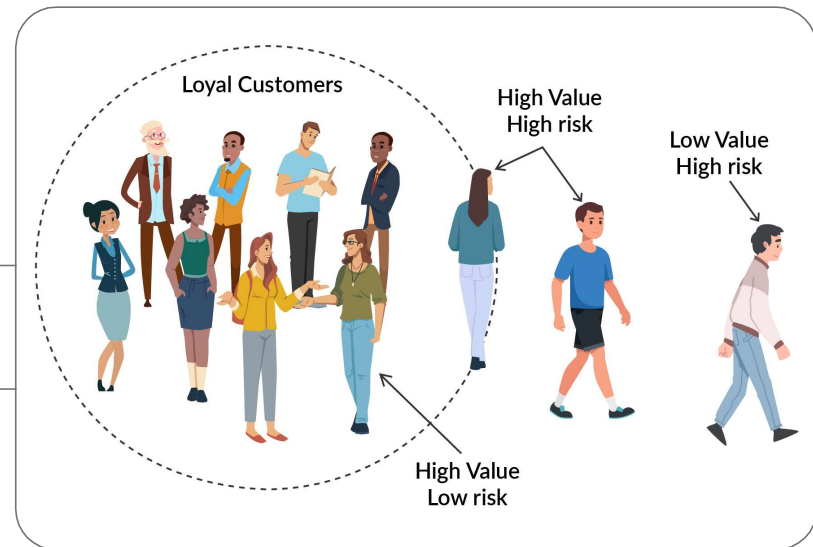
# UNDERSTANDING SOME KEY METRICS

1

Customer churn

2

Customer lifetime value (CLV)



# CUSTOMER CHURN

## 1. What is customer churn or churn rate?

- Customer churn is the percentage of customers who have discontinued their subscription to the service in a given period of time.

$$\frac{\text{(Number of Customers Lost)}}{\text{(Total Customers started with)}} = \text{Churn Rate}$$

# CUSTOMER CHURN

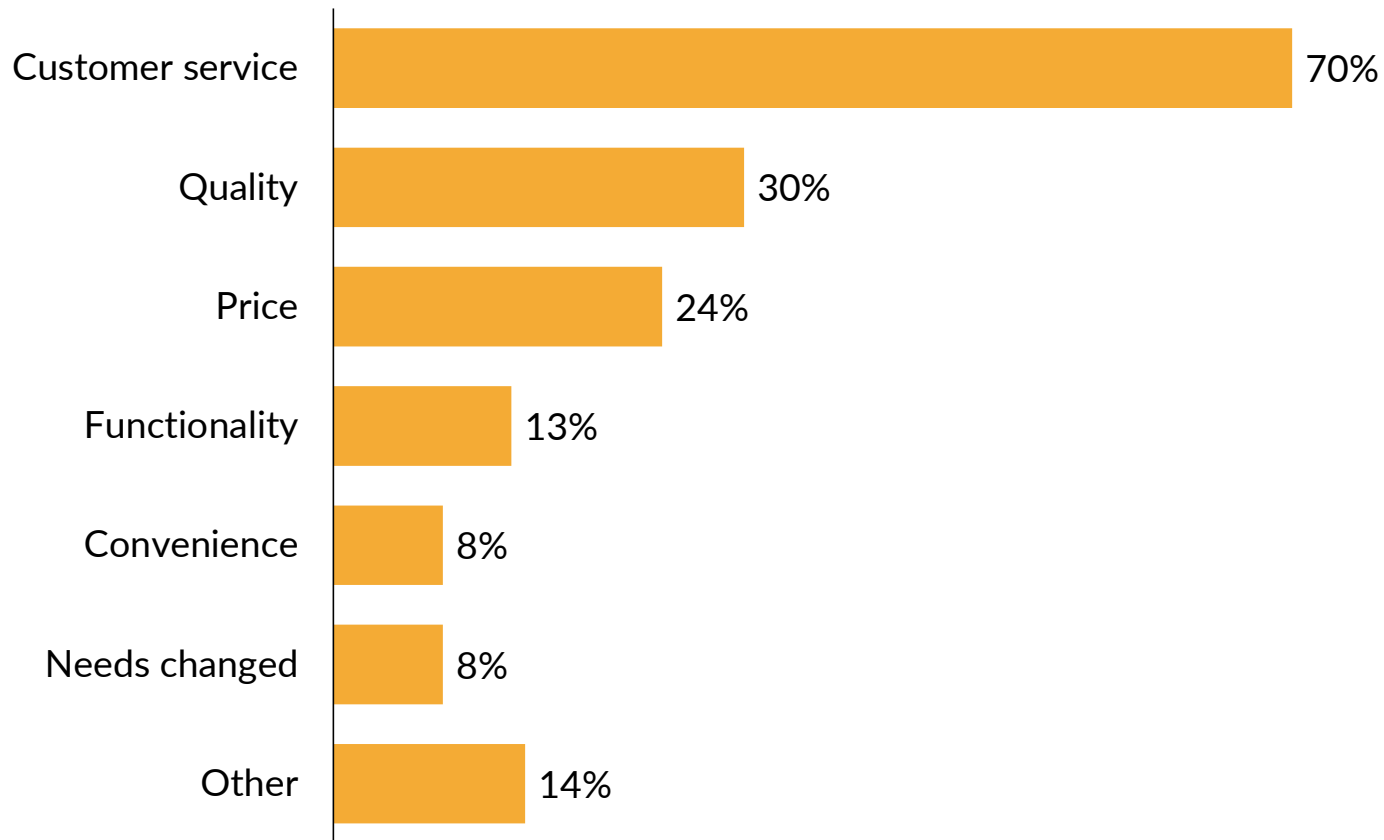
## 2. Why does customer churn matter?

- It is less expensive to retain customers than attract new customers
- Customer churn helps your competitors
- Existing customers, when satisfied, are an excellent source for brand promotions.
- High churn rate is an indicator of a bigger company problem



# CUSTOMER CHURN

## 3. Why do customers leave?



Source:- <https://www.kornferry.com/about-us/our-story/achieve-forum>

# CUSTOMER LIFETIME VALUE

1

**What is customer lifetime value?**

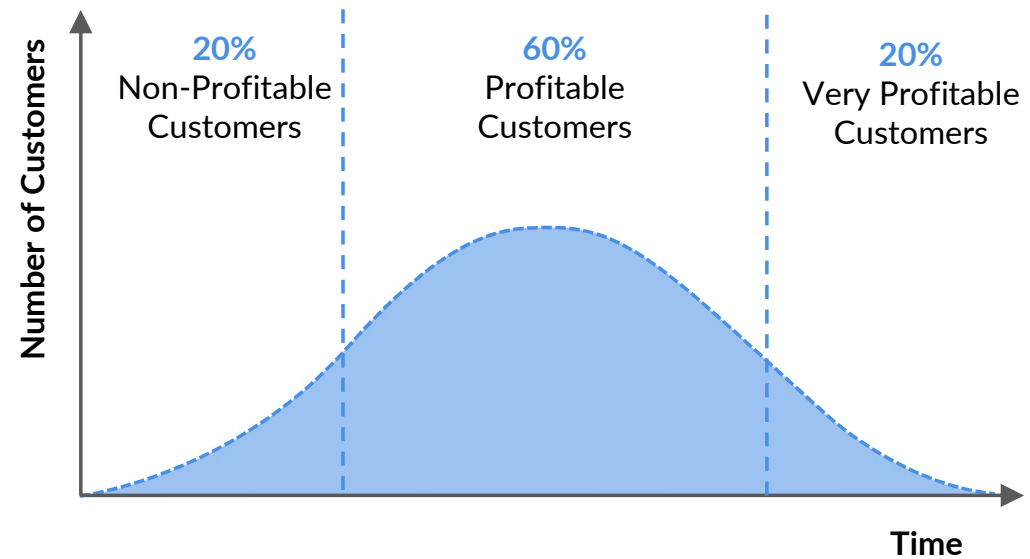
Customer lifetime value (CLV) is the net profit associated with the customer to the company over a fixed period of time

2

**Why should I care about CLV?**

3

**How is it related to customer churn?**



# CUSTOMER LIFETIME VALUE

1

## What is customer lifetime value?

Customer lifetime value (CLV) is the net profit associated with the customer to the company over a fixed period of time

2

## Why should I care about CLV?

3

## How is it related to customer churn?



Higher ROI



Less likely to churn



Custom-made customer programs

# CUSTOMER LIFETIME VALUE

1

## What is customer lifetime value?

Customer lifetime value (CLV) is the net profit associated with the customer to the company over a fixed period of time

2

## Why should I care about CLV?

3

## How is it related to customer churn?

I will probably purchase again



I'm likely to churn

# MY APPROACH

## Generate insights using SQL

- Analyze the dataset and identify potential factors that lead to improvement in CLTV and Churn

## Refine your database

- Make any required modifications to the schema

## Create the database

- Create the schema of the database
- Identify the primary key and foreign keys for each table
- Identify relationships between tables
- Apply normalization rules (1NF, 2NF, 3NF)
- Sketch the ERD diagram



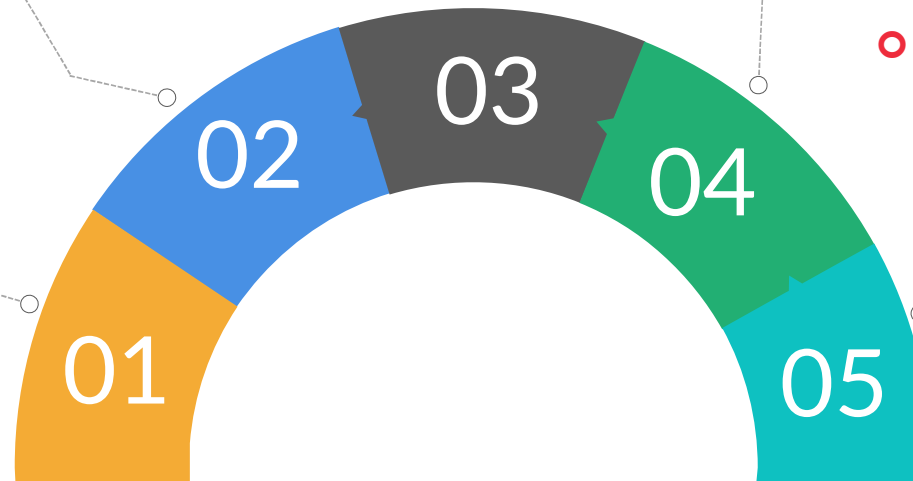
## Write business logics using python

- Connect python to the database and generate more complicated analysis on the database
- Analyze a lifetime value per group



## Visualize data using Tableau

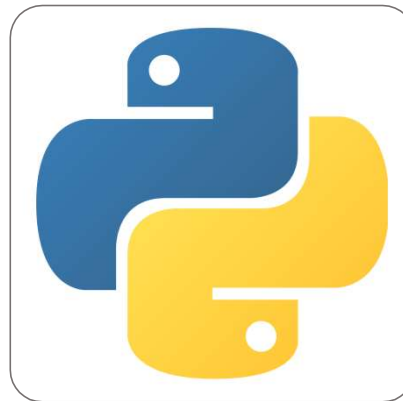
- Explore the data with Tableau
- Build Tableau dashboards
- Create a Tableau story with results



## TOOLS USED IN THIS CASE STUDY



SQLite



Python



Tableau

# INTRODUCTION TO SQLite

01

Already installed in Python; same code and installation for Mac, Linux, and Windows

02

Lightweight SQL server only a few kilobytes

03

Same SQL code from MySQL

○ Only small difference in data types

04

Easy integration with the Python-Tableau workflow

# WHEN TO USE SQLITE VS MYSQL

## Advantages of SQLite

Very easy to setup and use



SQLite is easily portable



Good for development and testing



## Advantages of MySQL

Good security features



Good speed and performance



Provides a lot of database related features





# RUNNING SQL ON PYTHON

- Python can work as a single interface to run simple queries and complicated calculations at the same time
- Single workflow to create databases, run queries and use Python functions and libraries that are common to data analytics
- Most big data libraries (e.g., Apache Spark) run SQL code through Python because it's the only way to parallelize and distribute the code in a cluster
- Required by most employers and companies because it's much easier to maintain and version SQL code when is embedded in Python

