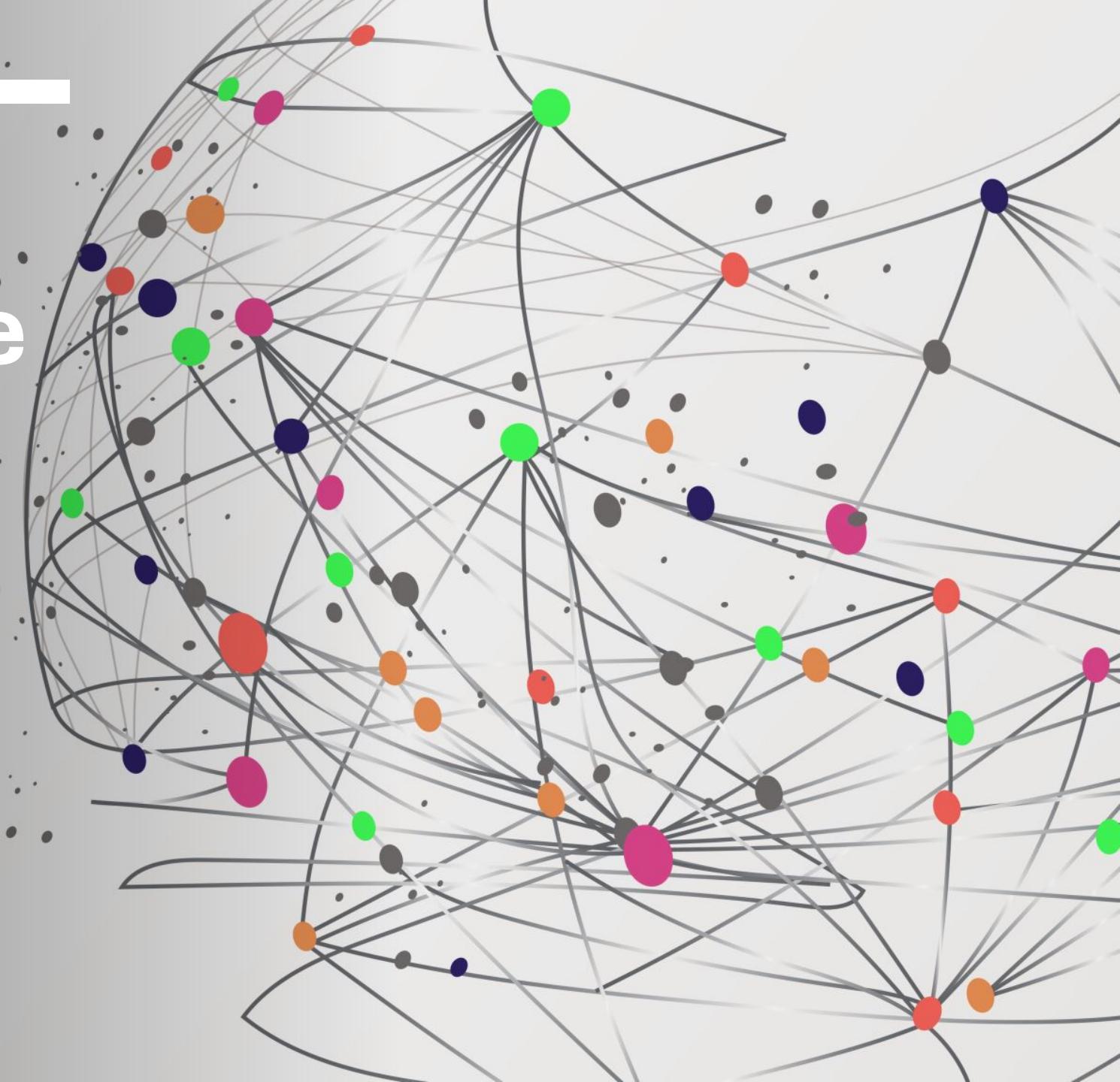


Practice Live Session - Case Study



Agenda



Part-I: Case study
specific business
problem (20 mins)



Part-II: Introduction
to SQLite (40 mins)



Part-III: Data
frames and EDA (40
mins)



Part IV: Tableau (40
mins)



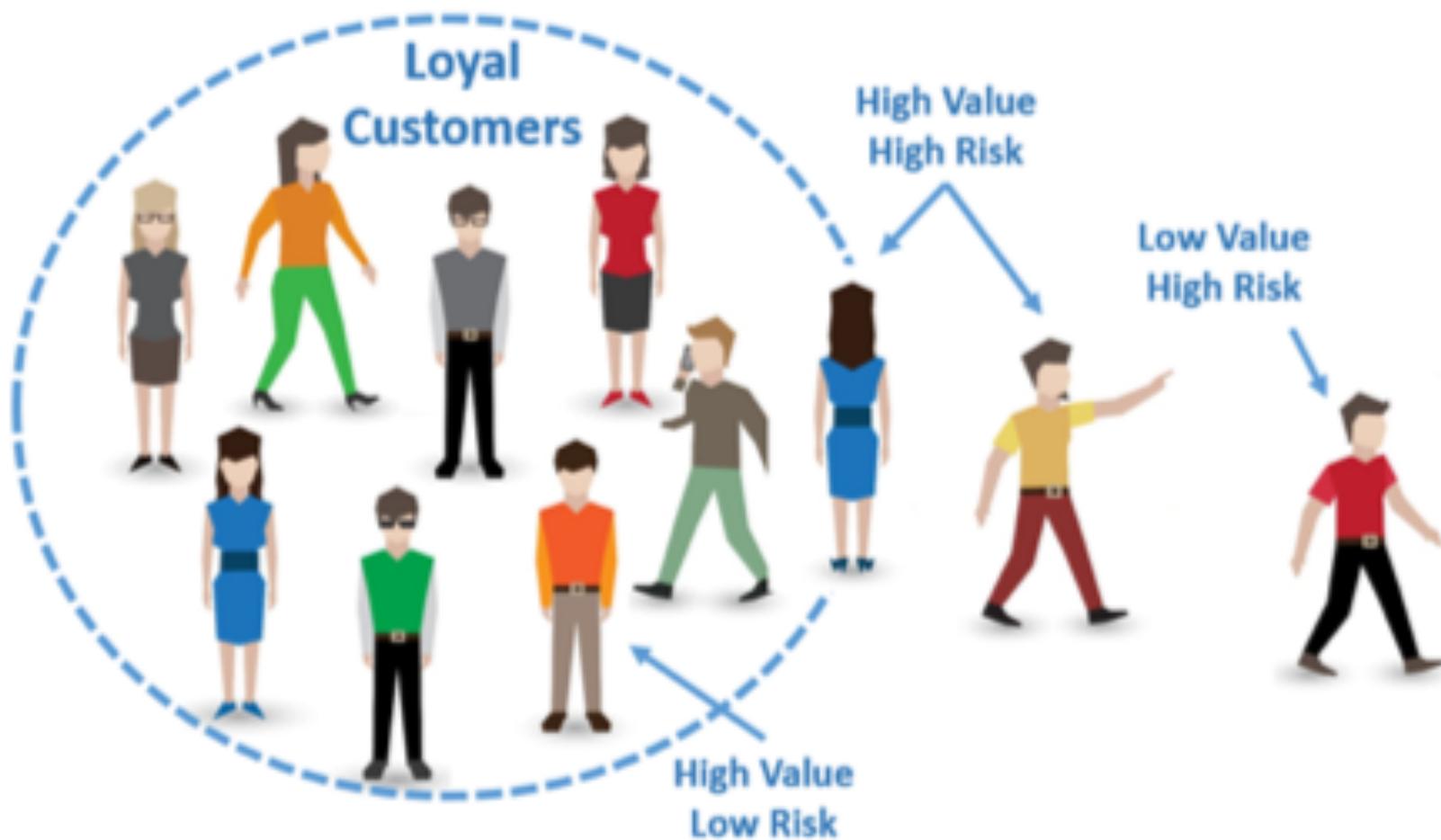
Part-V: Github (15
mins)



Part VI: Doubt
Resolving (15 mins)

Part I: Case study specific business problem

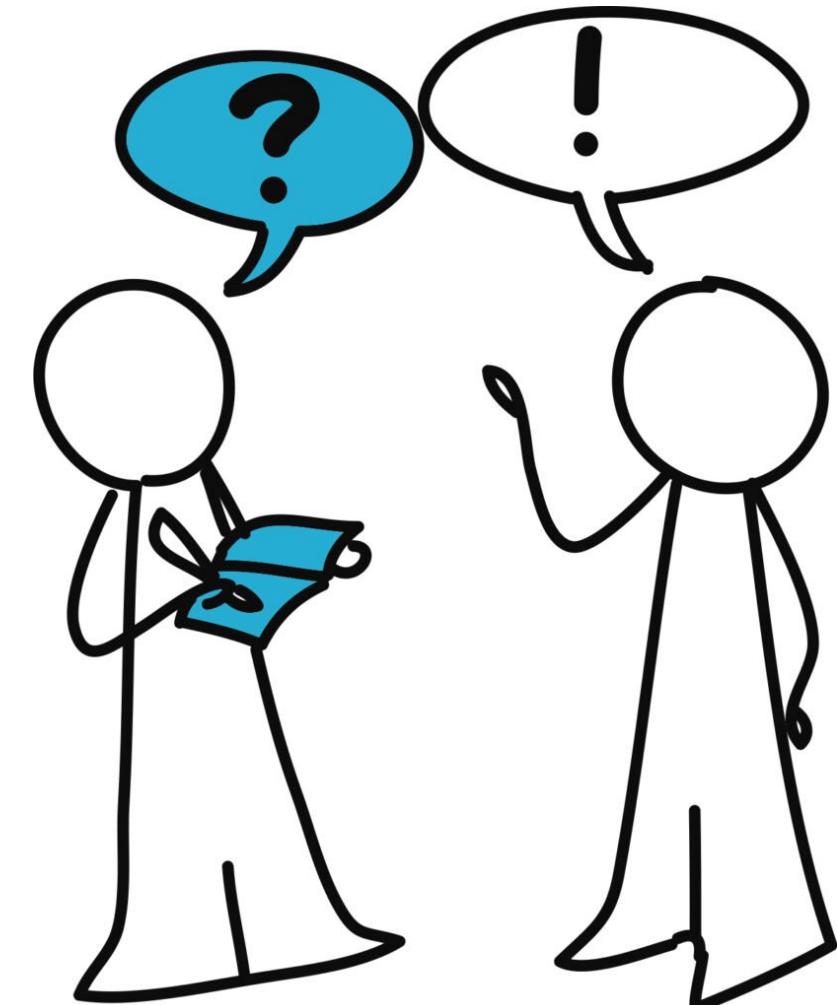
Church Analysis for a Telecoms Company



Part I: Case study specific business problem

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



Part I: Case study specific business problem

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

Customer.csv

Column Name	Description
CustomerID	A unique ID that identifies each customer.
Gender	The customer's gender: Male, Female
Senior Citizen	Indicates if the customer is 65 or older: Yes, No
Partner	Indicate if the customer has a partner: Yes, No
Dependents	Indicates if the customer lives with any dependents: Yes, No. Dependents could be children, parents, grandparents, etc.

Cust_net.csv

Column Name	Description
Cust_ID	A unique ID that identifies each customer.
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
Internet Service	Indicates if the customer subscribes to Internet service with the company: No, 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
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 programming 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.

Cust_account.csv

Column Name	Description
Account_id	A unique ID that identifies each customer.
Tenure Months	Indicates the total amount of months that the customer has been with the company by the end of the quarter specified above.
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.
Total Charges	Indicates the customer's total charges, calculated to the end of the quarter specified above.

Cust_loc.csv

Column Name	Description
Cust_ID	A unique ID that identifies each customer.
State	The state 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.

Cust_churn.csv

Column Name	Description
Id	A unique ID that identifies each customer.
Churn	Yes = the customer left the company this quarter. No = the customer remained with the company. Directly related to Churn Value.

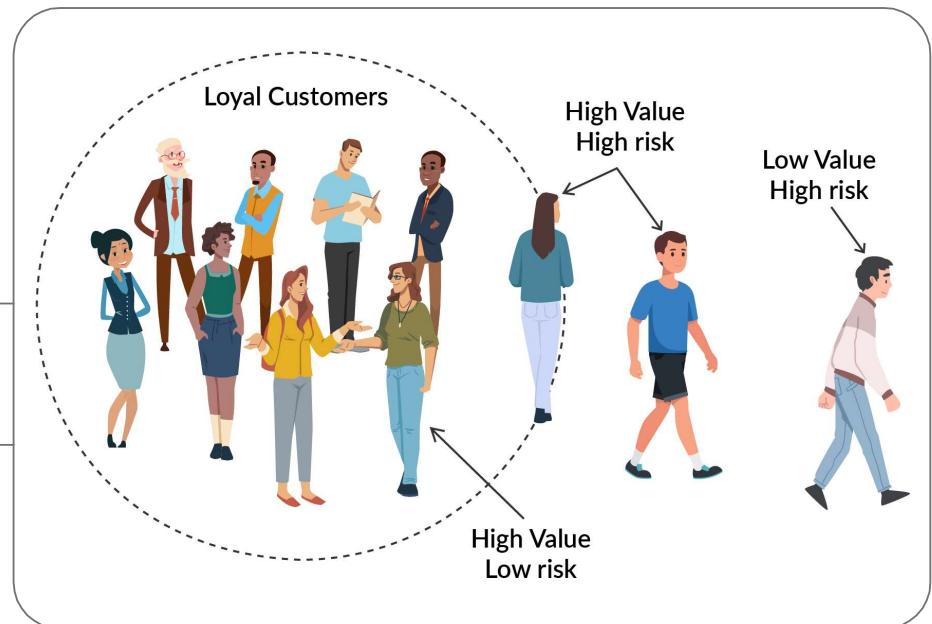
UNDERSTANDING SOME KEY METRICS

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 LIFETIME VALUE

1

What is customer lifetime value?

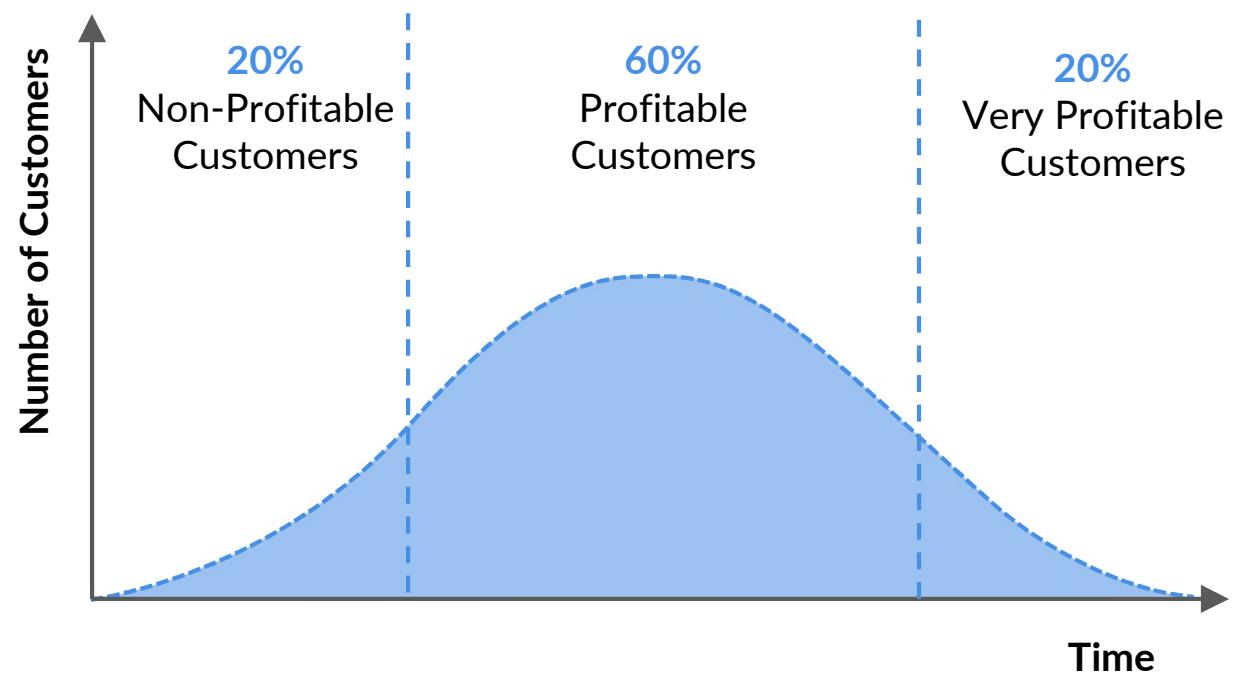
Customer lifetime value (CLV) is the net profit associated with the customer to the company over a fixed period of time

2

Why should I care about CLV?

3

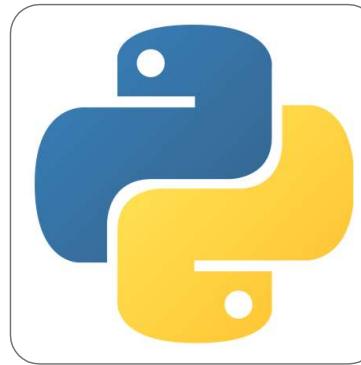
How is it related to customer churn?



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

SQLite vs MySQL vs PostgreSQL

Name	Advantages	Disadvantages	When to use	When not to use
SQLite	<ul style="list-style-type: none">• File based• Standards-aware• Great for developing and even testing	<ul style="list-style-type: none">• No user management• Lack of possibility to tinker with for additional performance	<ul style="list-style-type: none">• Embedded applications• Disk access replacement• Testing	<ul style="list-style-type: none">• Multi-user applications• Applications requiring high write volumes
MySQL	<ul style="list-style-type: none">• Easy to work with• Feature rich• Secure• Scalable and powerful• Speedy	<ul style="list-style-type: none">• Known limitations• Reliability issues• Stagnated development	<ul style="list-style-type: none">• Distributed operations• High Security• Web-sites and Web-applications• Custom solutions	<ul style="list-style-type: none">• SQL compliance• Concurrency• Lack of features
PostgreSQL	<ul style="list-style-type: none">• An open source SQL standard compliant RDBMS• Strong Community• Strong third-party support• Extensible• Objective	<ul style="list-style-type: none">• Performance• Popularity• Hosting	<ul style="list-style-type: none">• Data Integrity• Complex, custom procedures• Integration• Complex designs	<ul style="list-style-type: none">• Speed• Simple to set up• Replication

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

