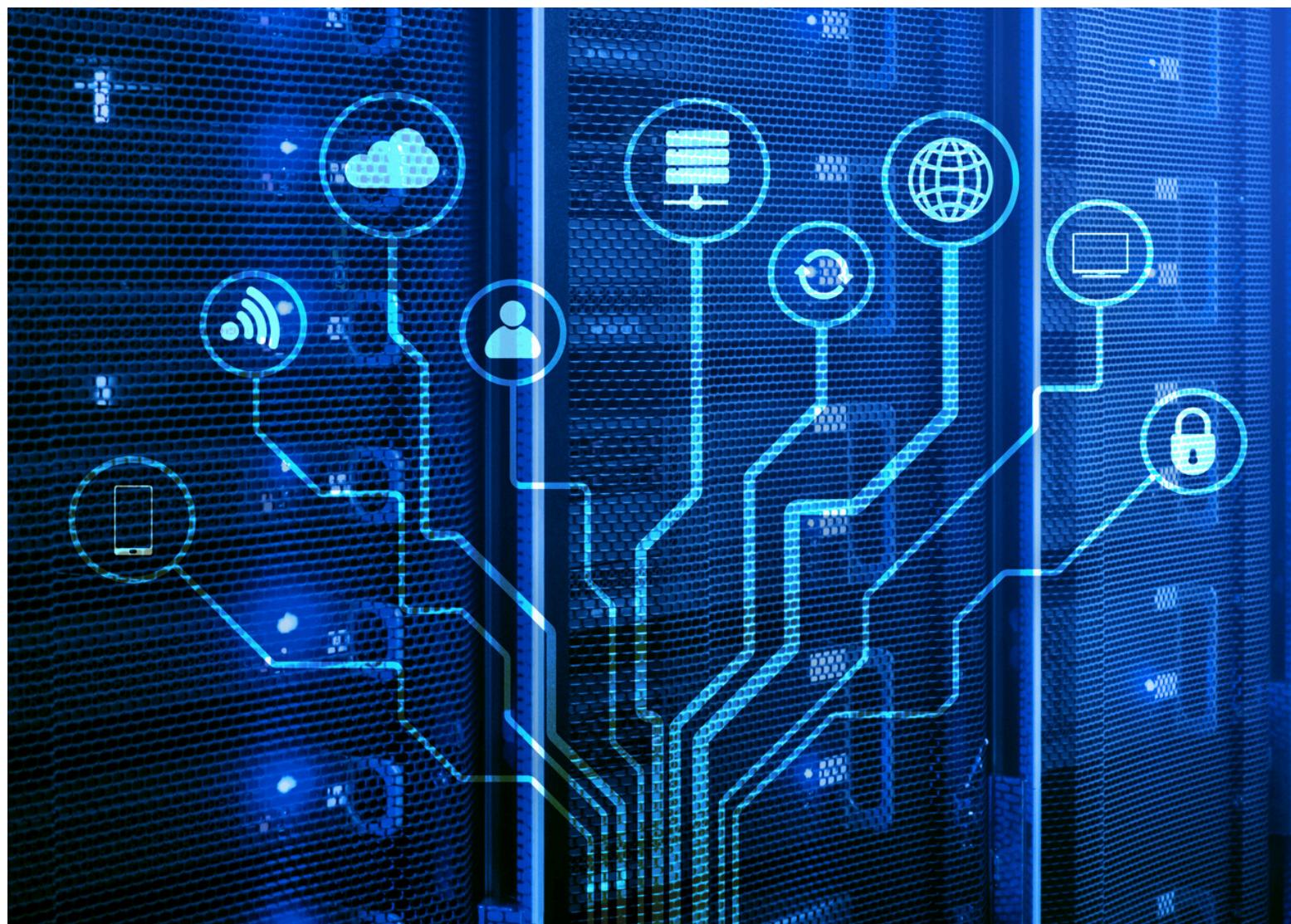
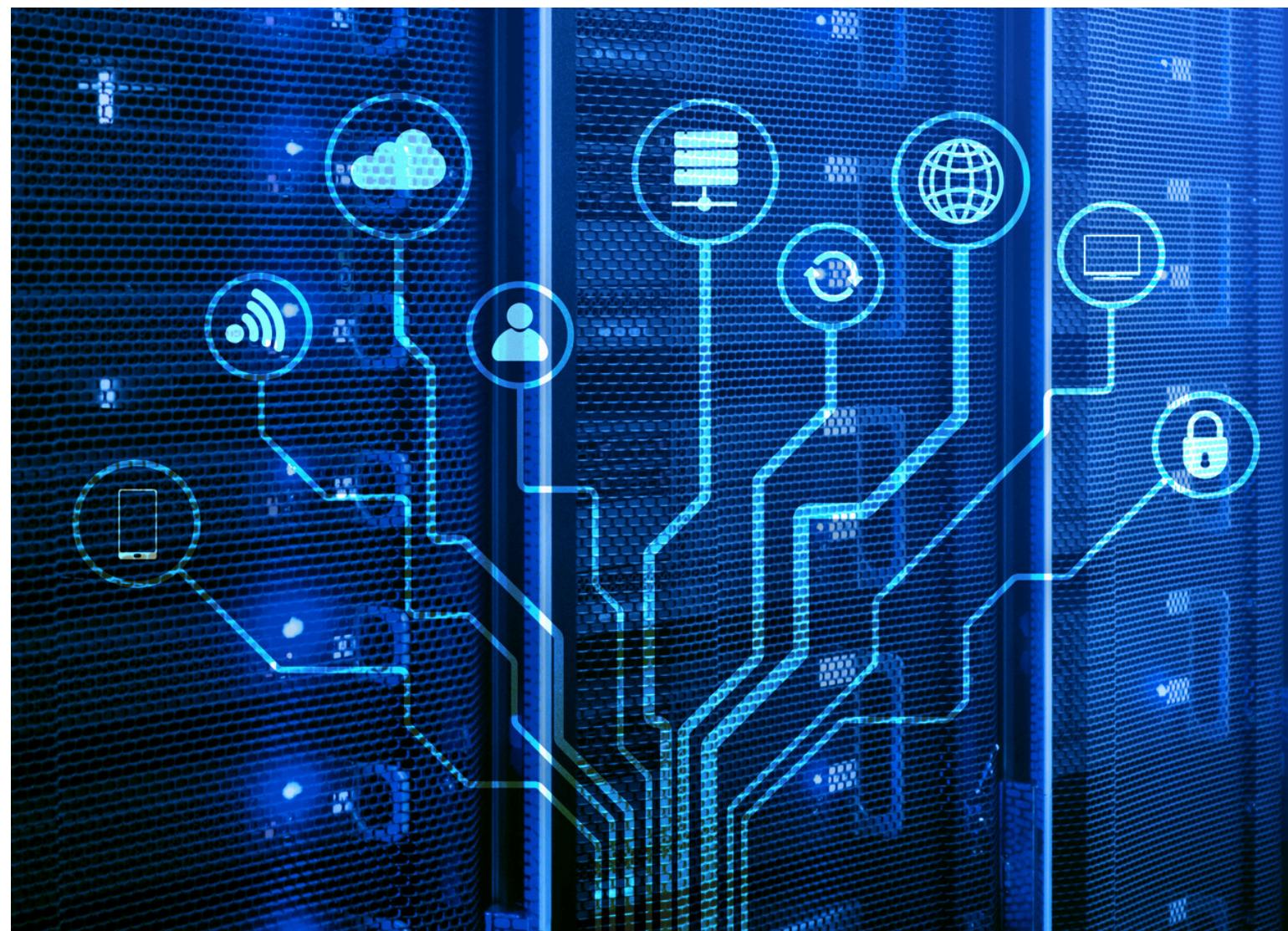


Unveiling Patterns in Telecom Churn

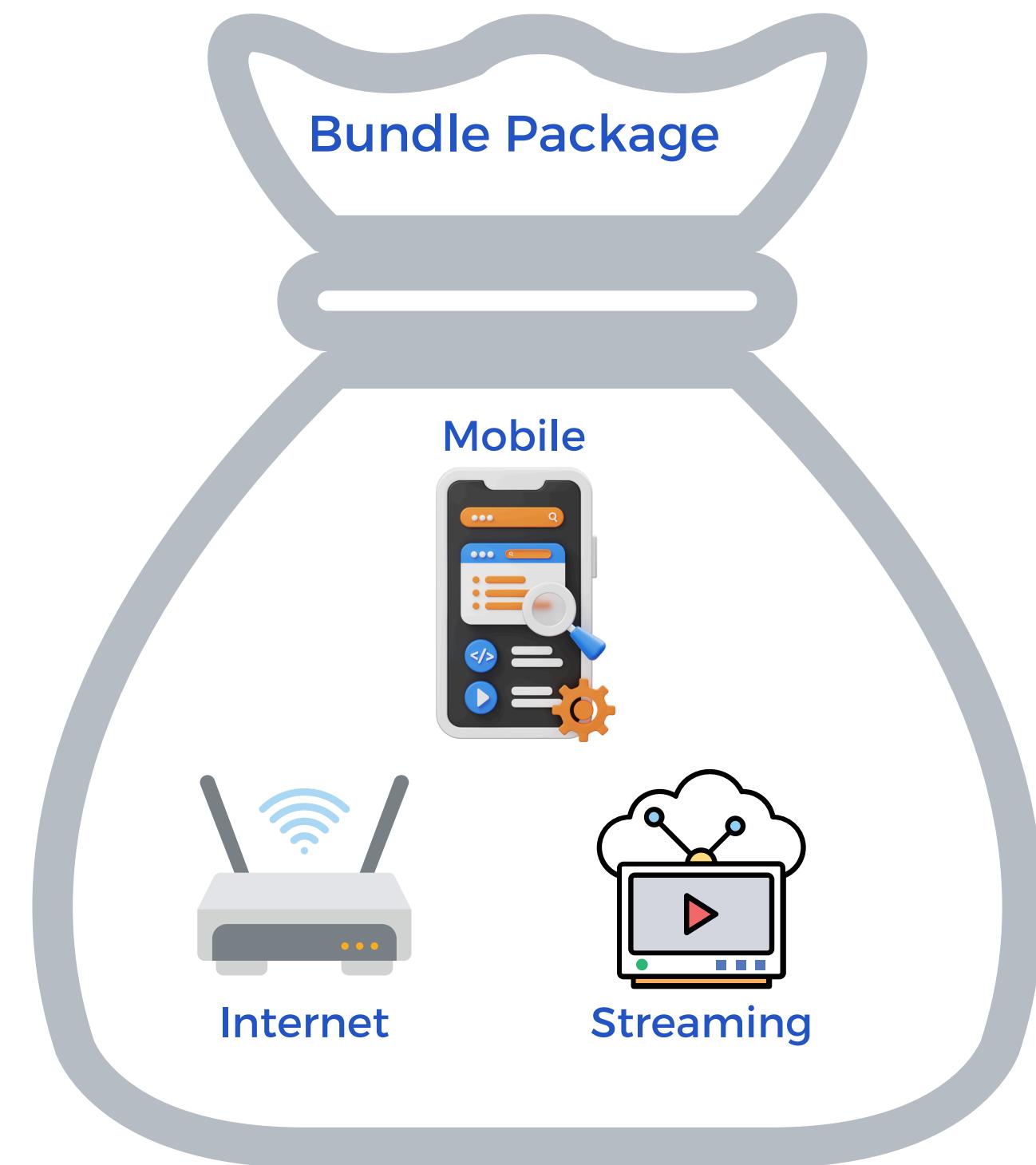
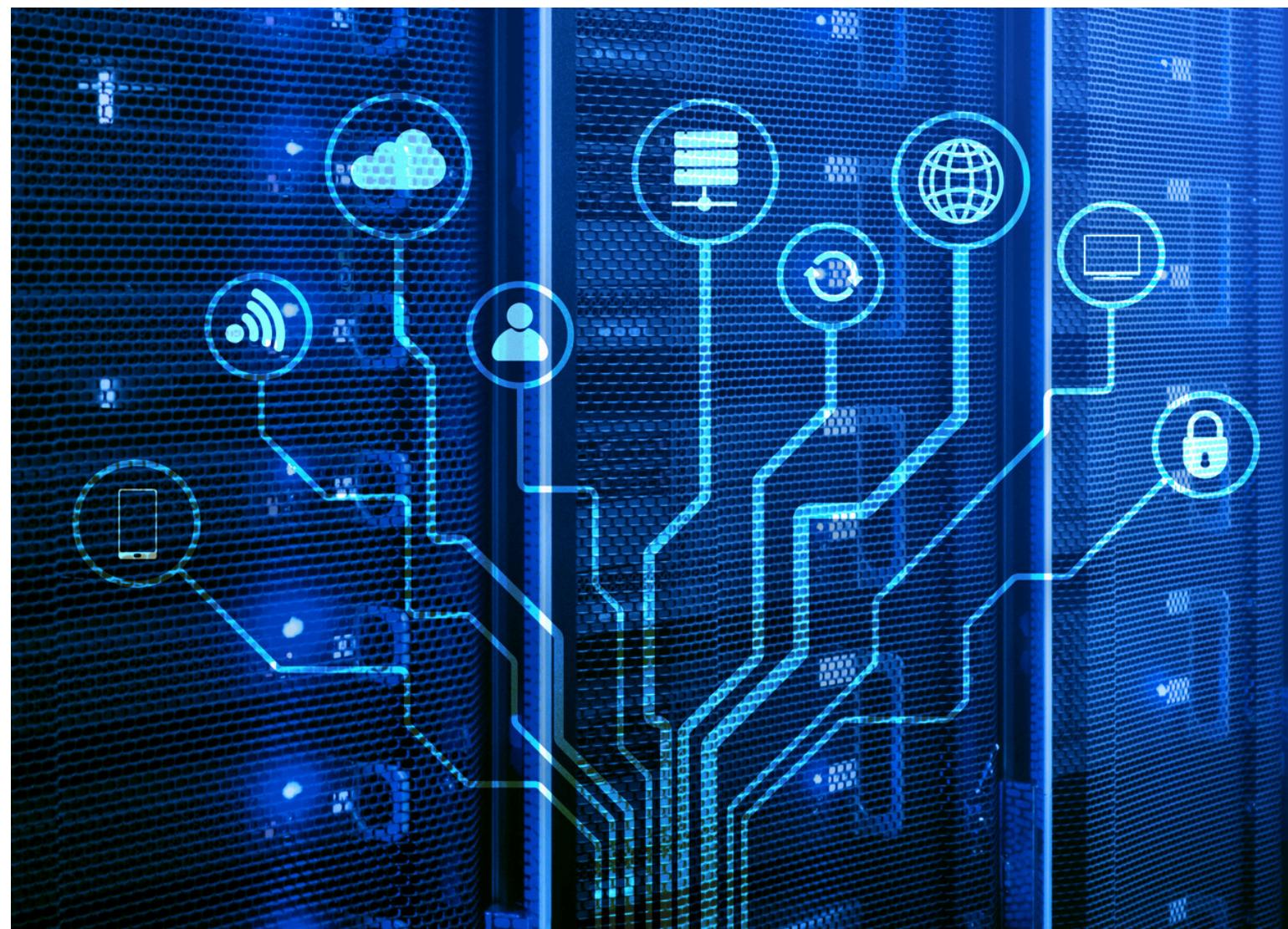
Telecommunication



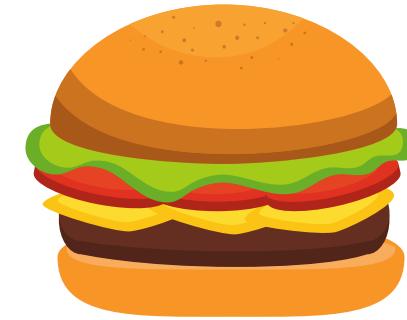
Telecommunication



Telecommunication



Why bundle?



\$ 3.49

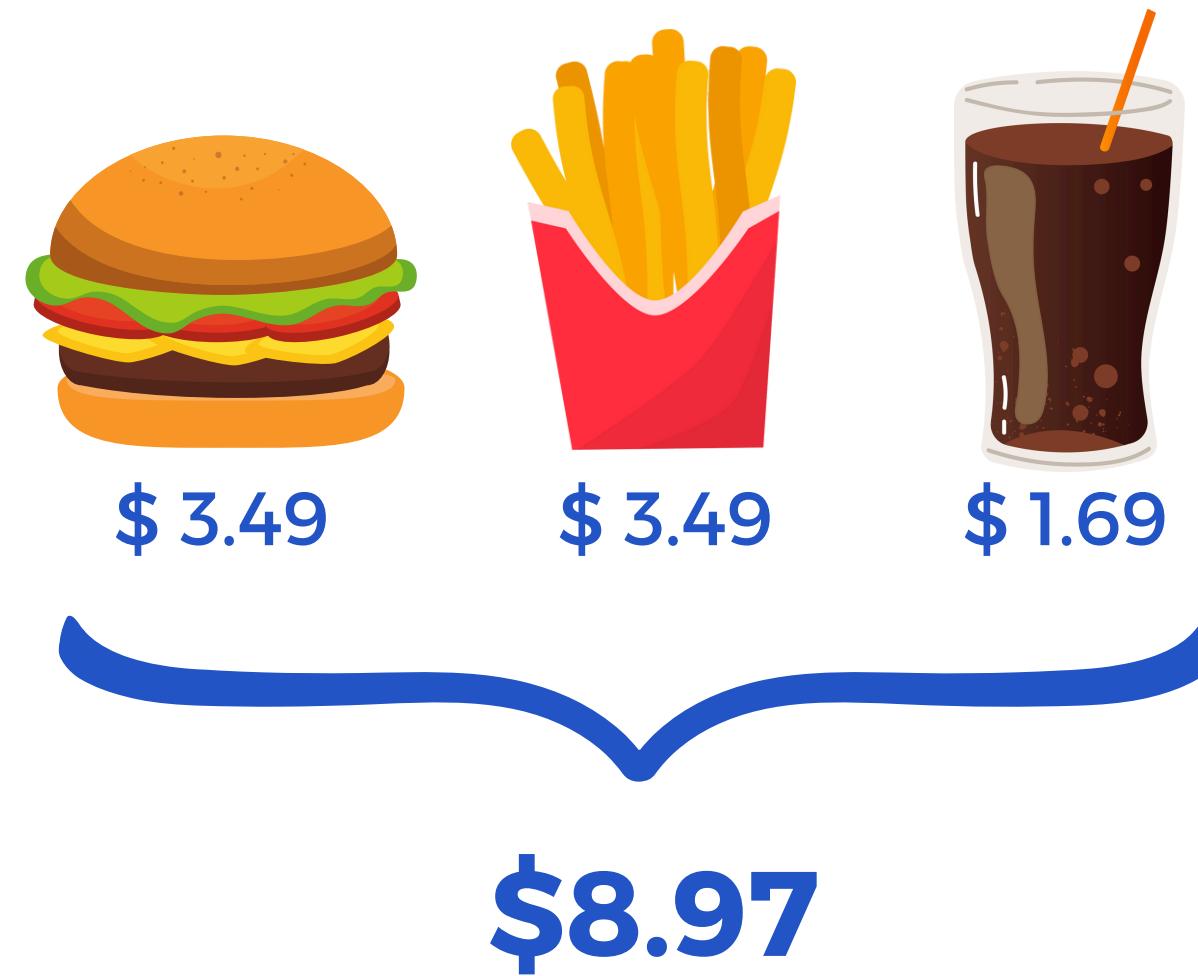


\$ 3.49

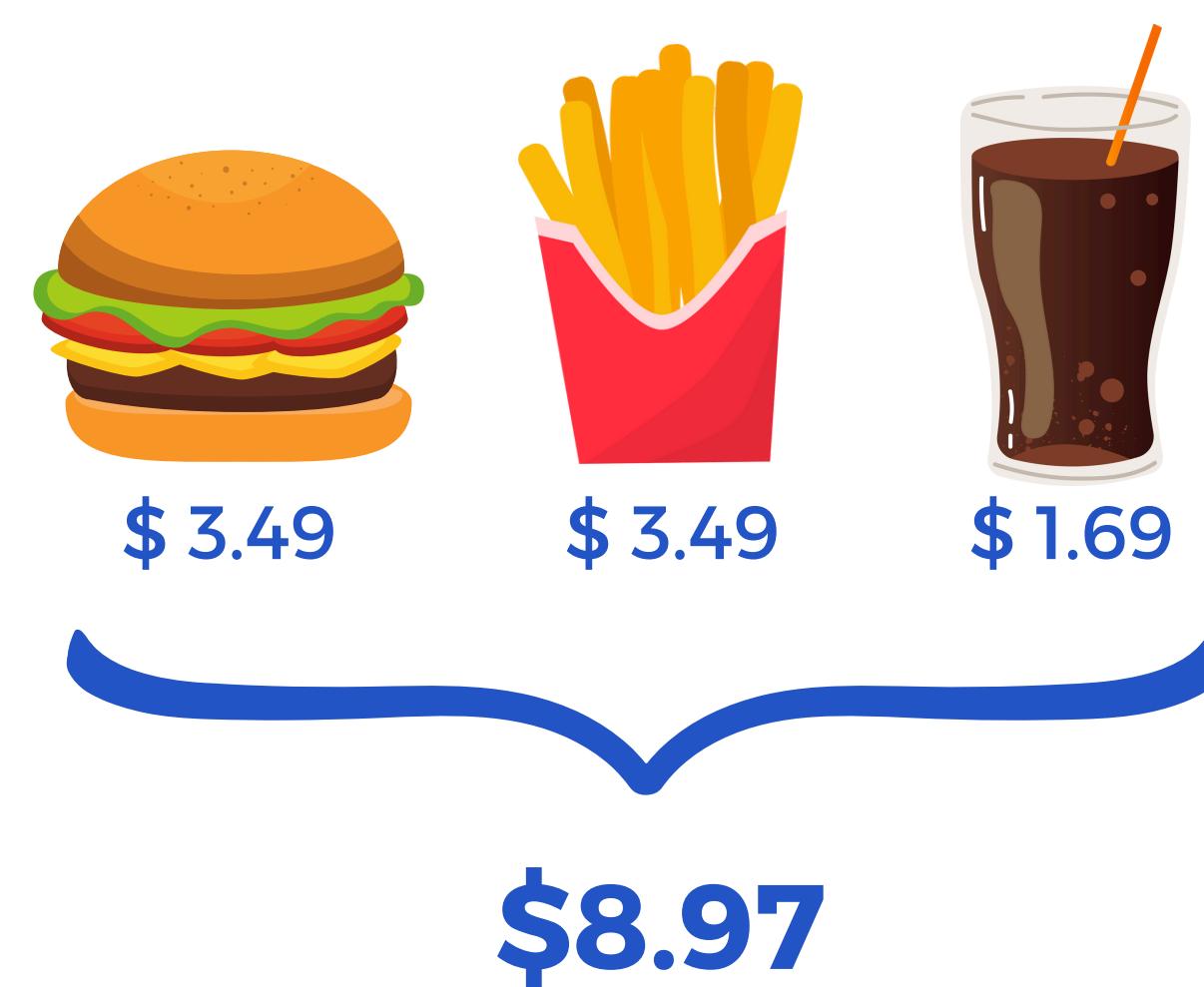


\$ 1.69

Why bundle?

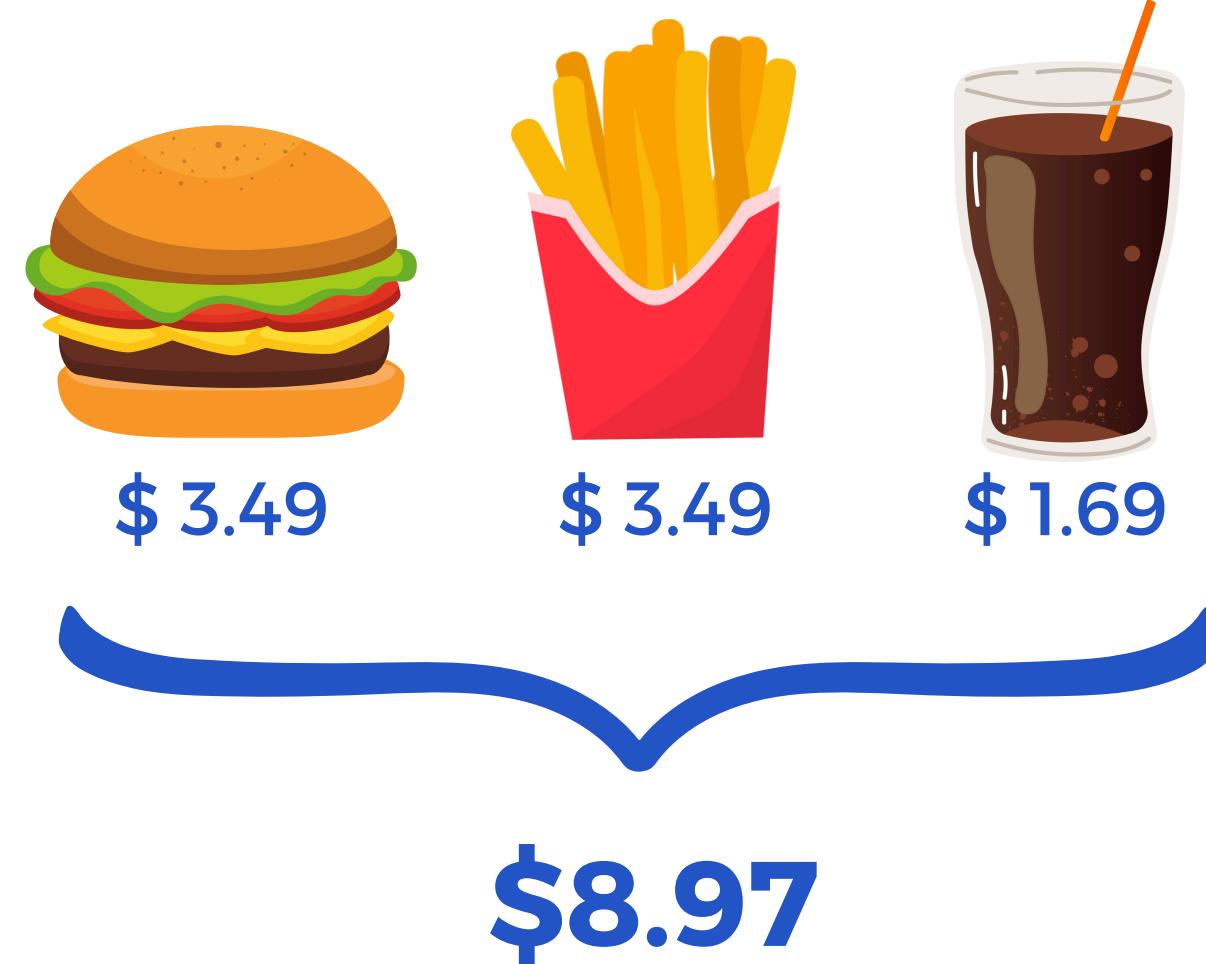


Why bundle?



Happy Meal

Why bundle?



Happy Meal



\$ 4.99

44.7% cheaper

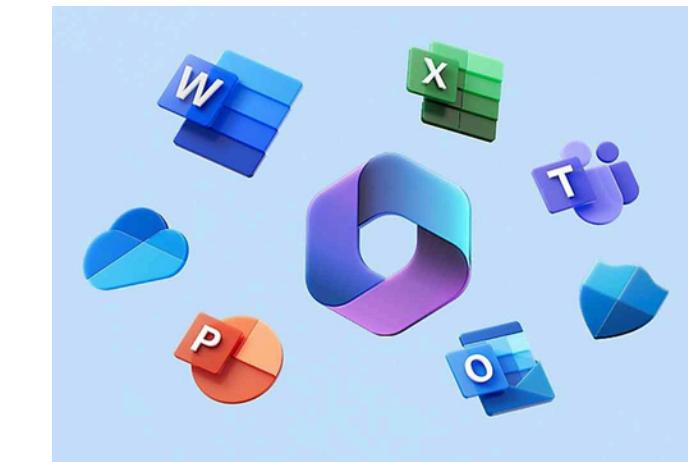
Benefits

For Customers

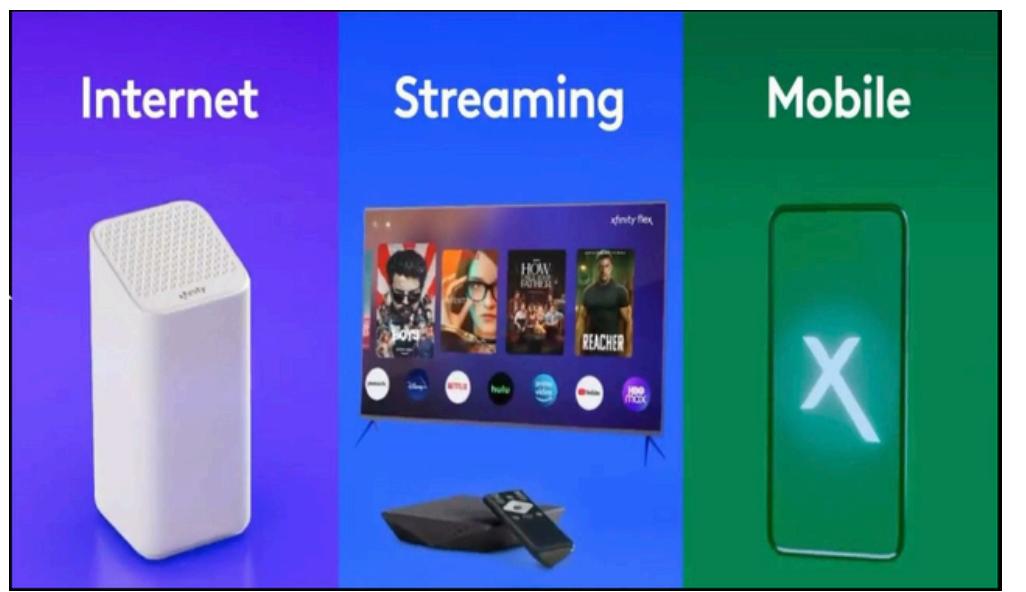
- “More bang for your buck”

For Businesses

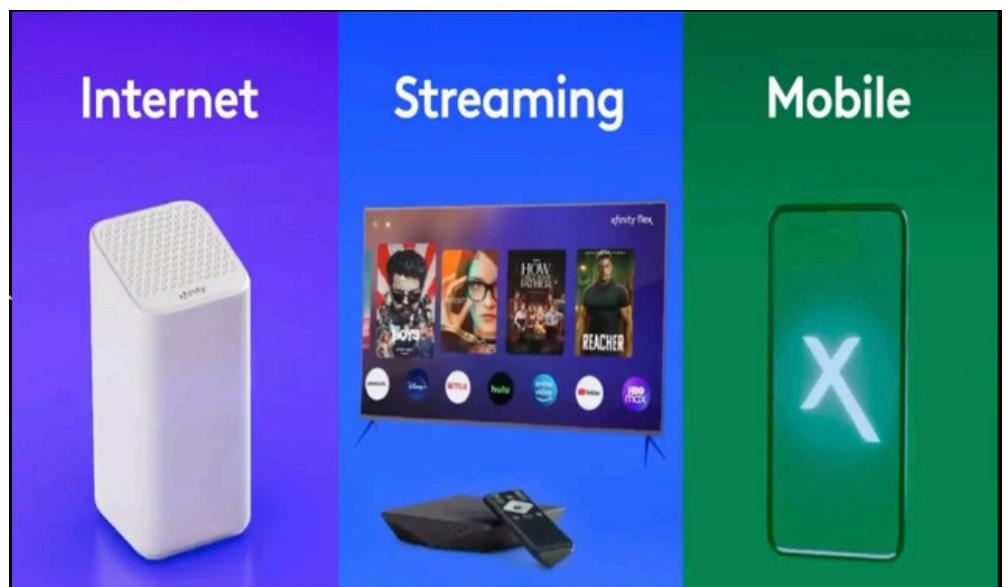
- Increase demand
- Increase retention
- Increase revenue



Situation



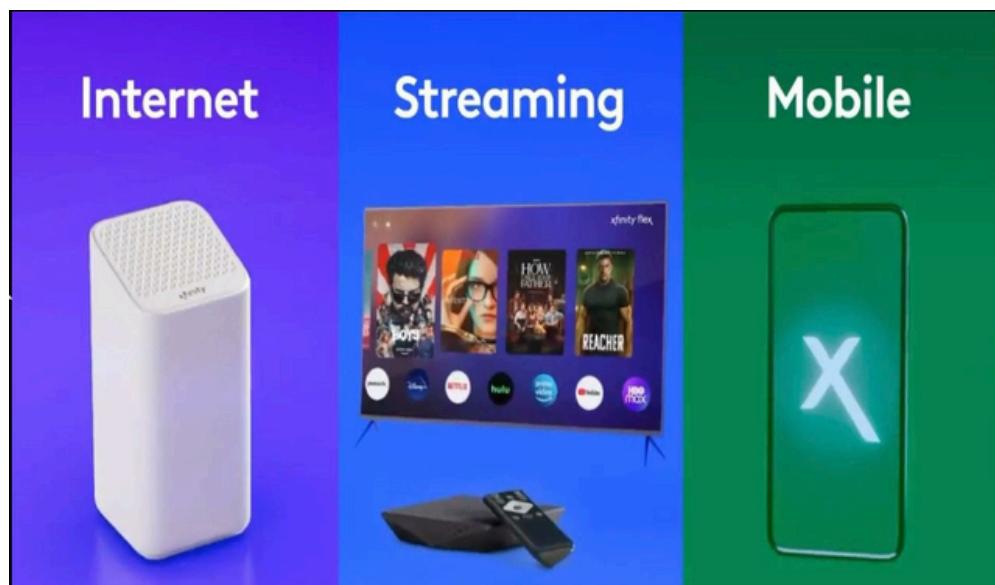
Problem: 27% of consumer churn each month



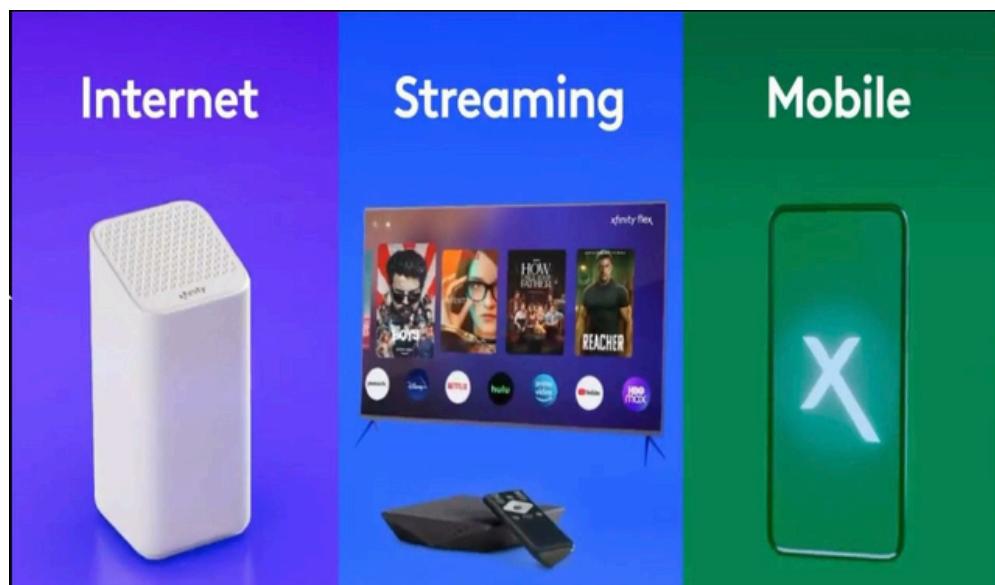
Problem: 27% of consumer churn each month



Monthly users:	7,043
Average monthly revenue per user:	\$64
Monthly lost revenue:	\$121,039



Problem: 27% of consumer churn each month



Monthly users: 7,043

Average monthly revenue per user: \$64

Monthly lost revenue: \$121,039

Monthly users: 1m

Average monthly revenue per user: \$64

Problem: 27% of consumer churn each month



Monthly users: 7,043

Average monthly revenue per user: \$64

Monthly lost revenue: \$121,039

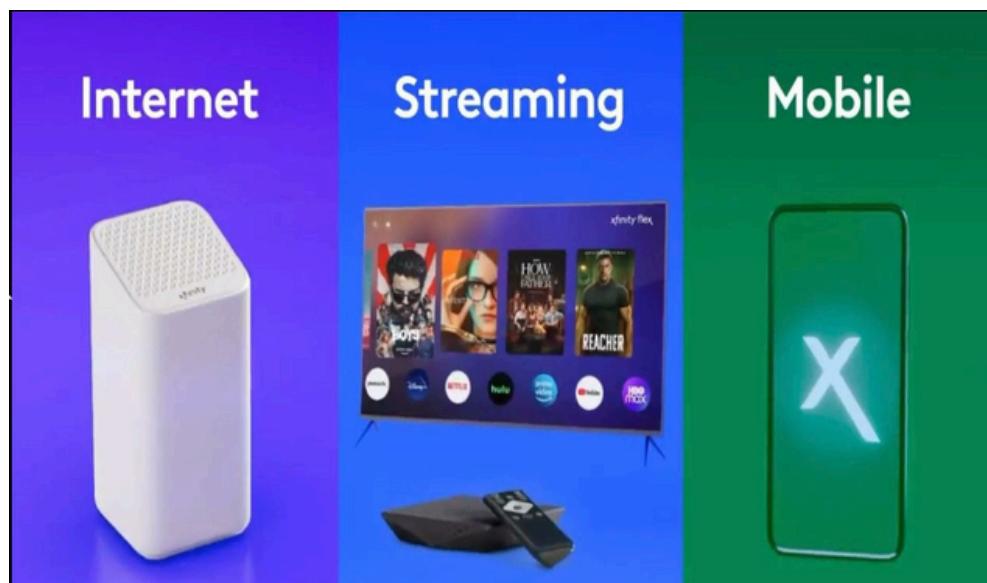
Monthly users: 1m

Average monthly revenue per user: \$64

Churn during presentation:

Lost revenue during presentation:

Problem: 27% of consumer churn each month



Monthly users: 7,043

Average monthly revenue per user: \$64

Monthly lost revenue: \$121,039

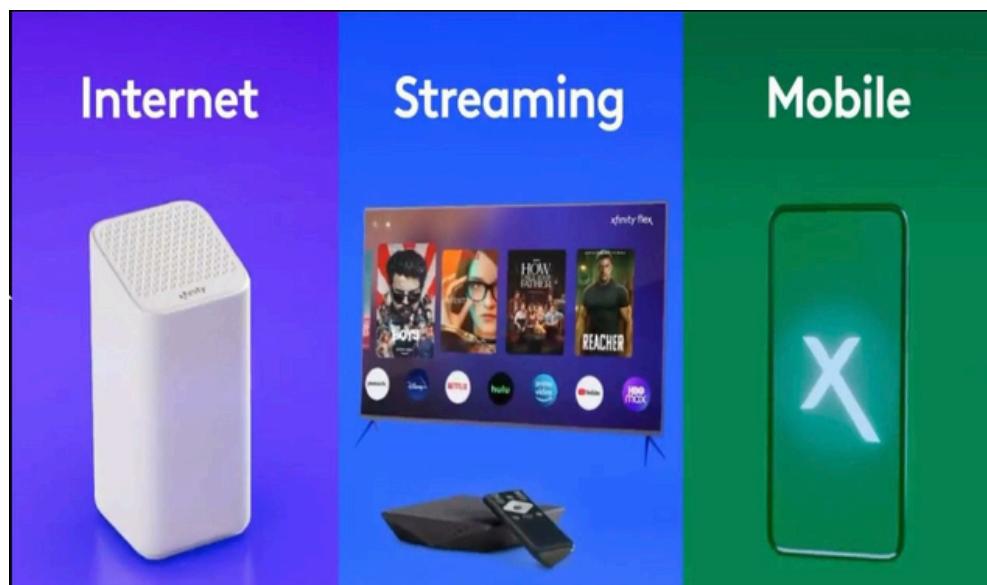
Monthly users: 1m

Average monthly revenue per user: \$64

Churn during presentation: 123 people

Lost revenue during presentation: \$7,872

Problem: 27% of consumer churn each month



Monthly users: 7,043

Average monthly revenue per user: \$64

Monthly lost revenue: \$121,039

Monthly users:

Average monthly revenue per user:

Churn during presentation:

Lost revenue during presentation:

Missed Annual Revenue

206.3M

We aim to reduce their monthly consumer churn from **27%** to **17%**

- Identifying the most influential features or reasons leading to consumer churn
- Predicting consumers at risk of churning

We aim to reduce their monthly consumer churn from **27%** to **17%**

- Identifying the most influential features or reasons leading to consumer churn
- Predicting consumers at risk of churning

Approach:

Implement descriptive, diagnostic, and predictive methods.

Project Budget:

\$ 37,150

Result:

Additional **\$522,000** annual revenue

ROI: After 01 year of successful implementation ~ **14X**

As a team, we generated a plan to approach this project by dividing the tasks into a series of 3 parts:

Part 01

- Determining the data mining goals as a team.

As a team, we generated a plan to approach this project by dividing the tasks into a series of 3 parts:

Part 01

- Determining the data mining goals as a team.

Part 02: Setup

- Generate an ERD @kevin
- Data dictionary @rachel
- Data governance policies @valentin
- ETL pipeline @mason
- Explore the data @minh
- Data warehouse @mason & @kevin

As a team, we generated a plan to approach this project by dividing the tasks into a series of 3 parts:

Part 01

- Determining the data mining goals as a team.

Part 02: Setup

- Generate an ERD @kevin
- Data dictionary @rachel
- Data governance policies @valentin
- ETL pipeline @mason
- Explore the data @minh
- Data warehouse @mason & @kevin

Part 03: Quantify the problem and solution

- Identify features - or behaviors - that have high predictive power on consumer churn
- Generate KPIs and objectives
- Visualize the KPIs / objectives / metrics in a dashboard

Action

Data Mining Goals

As a team,
we came together to
brainstorm from the
data periodic table.

Results

Action

Data Mining Goals

As a team,
we came together to
brainstorm from the
data periodic table.

Results

F01. Increased use of external data	F02. Increased use of internal data	F03. Increased use of all onboarding data	F04. Incentives for customers to provide more data
F05. More consistent data capture	F06. Improved quality of customer data	F07. Increased use of structured data	F08. Increased use of unstructured data
F09. Improved data methods and processes	F10. Improved data governance	F11. Increased understanding and insight of customers	F12. Increased no of contactable customers
F13. Single customer view	F14. Improved use of historical dialogue customer data	F15. Improved modelling accuracy	F16. Reduced modelling development and production time
F17. Improved use of data science models	F18. Data science models embedded in downstream systems and op processes	F19. Improved analytics tools	F20. Real-time streaming data architecture
F21. Improved use of artificial intelligence	F22. More accessible data	F23. Improved use of machine learning	F24. Reduced time taken to access data
F25. Improved transactional data analysis	F26. Improved event detection for rules-based decisioning	F27. Improved real-time decisioning	F28. Improved data team skills
F29. Stronger analytics community	F30. More knowledgeable data community	F31. More data-aware exec	F32. Improved data ethics
F33. Improved data sets for teaching models			

Action

Data Mining Goals

As a team,
we came together to
brainstorm from the
data periodic table.

Results

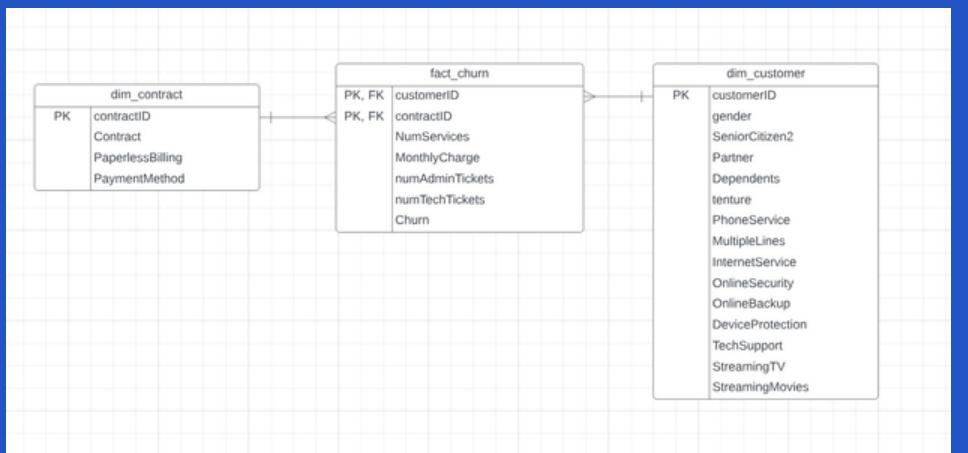
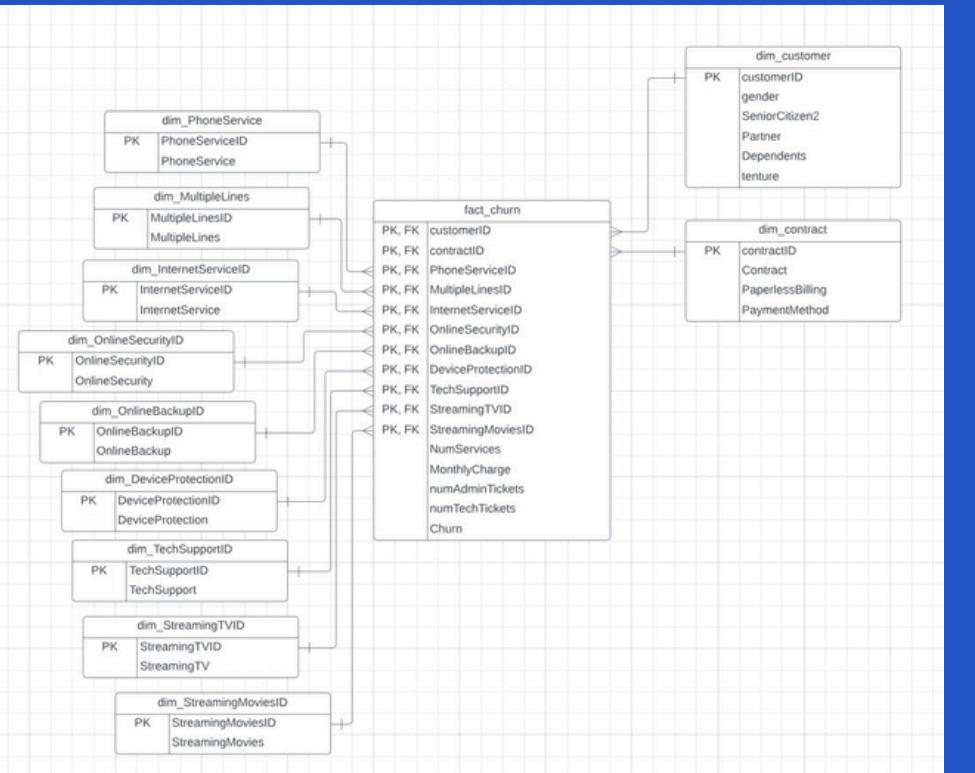
F01. Increased use of external data	F02. Increased use of internal data	F03. Increased use of all onboarding data	F04. Incentives for customers to provide more data
F05. More consistent data capture	F06. Improved quality of customer data	F07. Increased use of structured data	F08. Increased use of unstructured data
F09. Improved data methods and processes	F10. Improved data governance	F11. Increased understanding and insight of customers	F12. Increased no of contactable customers
F13. Single customer view	F14. Improved use of historical dialogue customer data	F15. Improved modelling accuracy	F16. Reduced modelling development and production time
F17. Improved use of data science models	F18. Data science models embedded in downstream systems and op processes	F19. Improved analytics tools	F20. Real-time streaming data architecture
F21. Improved use of artificial intelligence	F22. More accessible data	F23. Improved use of machine learning	F24. Reduced time taken to access data
F25. Improved transactional data analysis	F26. Improved event detection for rules-based decisioning	F27. Improved real-time decisioning	F28. Improved data team skills
F29. Stronger analytics community	F30. More knowledgeable data community	F31. More data-aware exec	F32. Improved data ethics
F33. Improved data sets for teaching models			

Data Mining Success Criteria

- Achieve a **specific** accuracy in churn prediction & identify significant churn predictors
- Achieve **80% recall score** & **75% F1-score** in the churn prediction model

Action

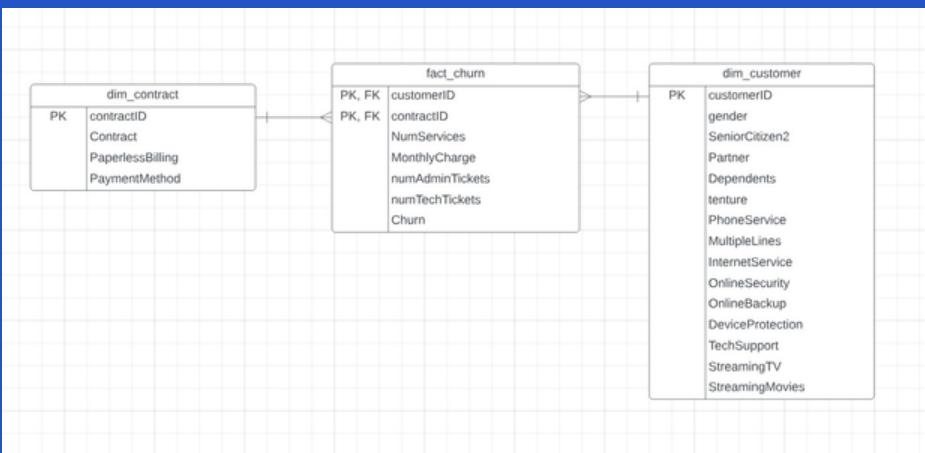
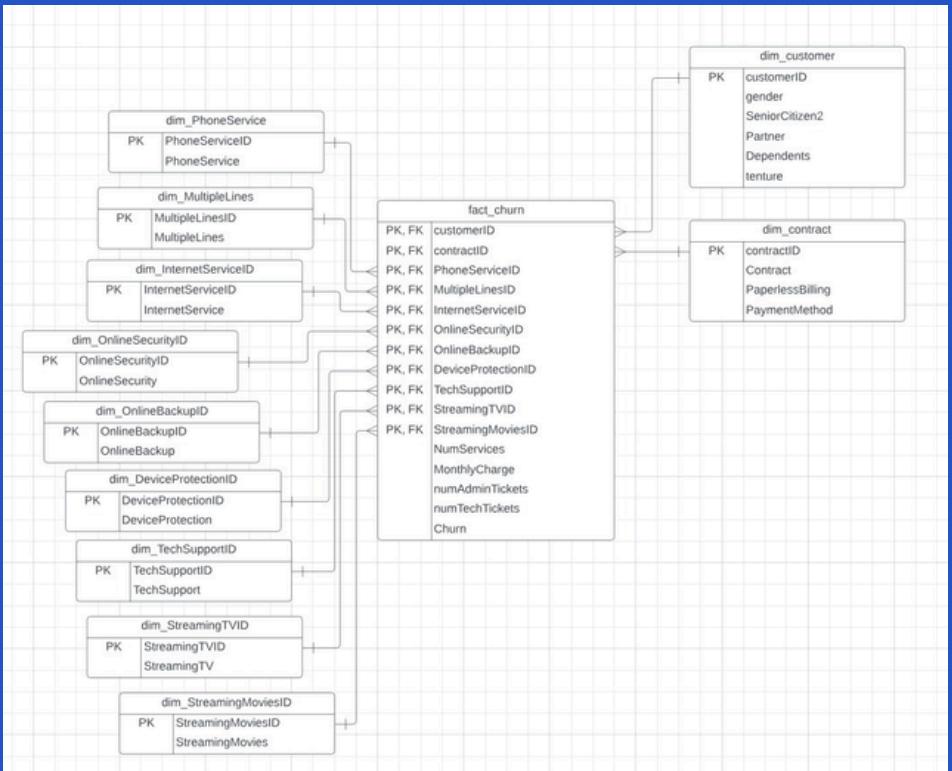
Generate an ERD



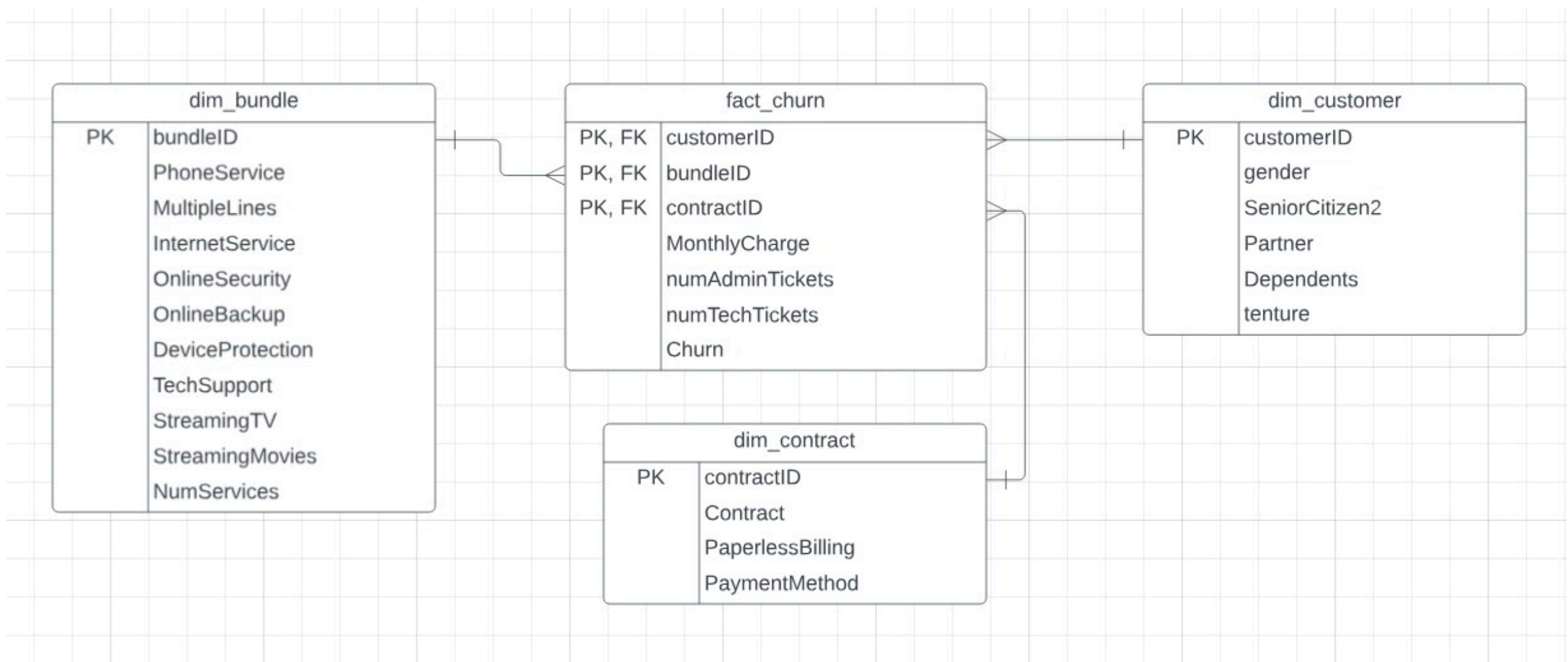
Results

Action

Generate an ERD



Results



Action

Data Dictionary

Metadata

- Source
- Owner
- Last Updated
- Notes

Dataset Name

Data Attributes

- Attribute Name
- Data Type
- Description

Results

Action

Data Dictionary

Metadata

- Source
- Owner
- Last Updated
- Notes

Dataset Name

Data Attributes

- Attribute Name
- Data Type
- Description

Results

Original Dictionary

Index	Column Name	Data Type	Description	NULL?
1	customerID	VARCHAR(25)	Unique identifier for the customer	No
2	gender	VARCHAR(6)	Customer's gender (Male/Female)	No
3	SeniorCitizen	BOOL	Indicates if the customer is a senior citizen (1 means senior citizen)	No
4	Partner	VARCHAR(3)	Whether the customer is married/ in a live-in relationship (Yes/No)	No
5	Dependents	VARCHAR(3)	Whether the customer has dependents - children/ retired parents (Yes/No)	No
6	tenure	INT	Number of months the customer has stayed with the company	No
7	PhoneService	VARCHAR(3)	Whether the customer has a phone service (Yes/No)	No
8	MultipleLines	VARCHAR(25)	Whether the customer has multiple lines (Yes/No/No phone service)	No
9	InternetService	VARCHAR(25)	Customer's internet service type (DSL/Fiber optic/No)	No
10	OnlineSecurity	VARCHAR(25)	Whether the customer has online security (Yes/No/No internet service)	No
11	OnlineBackup	VARCHAR(25)	Whether the customer has online backup (Yes/No/No internet service)	No
12	DeviceProtection	VARCHAR(25)	Whether the customer has device protection (Yes/No/No internet service)	No
13	TechSupport	VARCHAR(25)	Whether the customer has tech support (Yes/No/No internet service)	No
14	StreamingTV	VARCHAR(25)	Whether the customer has streaming TV (Yes/No/No internet service)	No
15	StreamingMovies	VARCHAR(25)	Whether the customer has streaming movies (Yes/No/No internet service)	No
16	NumServices	INT	The number of services the customer is using from the telecom	No
17	Contract	VARCHAR(25)	The term of the customer's contract (Month-to-month/One year/Two year)	No
18	PaperlessBilling	VARCHAR(3)	Whether the customer has paperless billing (Yes/No)	No
19	PaymentMethod	VARCHAR(25)	The customer's payment method	No
20	MonthlyCharges	DECIMAL(10, 2)	The amount charged to the customer every month	No
21	TotalCharges	DECIMAL(10, 2)	The total amount charged to the customer	No
22	numAdminTickets	INT	Number of administrative tickets raised by the customer	No
23	numTechTickets	INT	Number of technical tickets raised by the customer	No
24	Churn	VARCHAR(3)	Whether the customer has churned or not (Yes/No)	No

Action

Metadata

- Source
- Owner
- Last Updated
- Notes

Dataset Name

Data Attributes

- Attribute Name
- Data Type
- Description

Data Dictionary

Original Dictionary

Index	Column Name	Data Type	Description	NULL?
1	customerID	VARCHAR(25)	Unique identifier for the customer	No
2	gender	VARCHAR(6)	Customer's gender (Male/Female)	No
3	SeniorCitizen	BOOL	Indicates if the customer is a senior citizen (1 means senior citizen)	No
4	Partner	VARCHAR(3)	Whether the customer is married/ in a live-in relationship (Yes/No)	No
5	Dependents	VARCHAR(3)	Whether the customer has dependents - children/ retired parents (Yes/No)	No
6	tenure	INT	Number of months the customer has stayed with the company	No
7	PhoneService	VARCHAR(3)	Whether the customer has a phone service (Yes/No)	No
8	MultipleLines	VARCHAR(25)	Whether the customer has multiple lines (Yes/No/No phone service)	No
9	InternetService	VARCHAR(25)	Customer's internet service type (DSL/Fiber optic/No)	No
10	OnlineSecurity	VARCHAR(25)	Whether the customer has online security (Yes/No/No internet service)	No
11	OnlineBackup	VARCHAR(25)	Whether the customer has online backup (Yes/No/No internet service)	No
12	DeviceProtection	VARCHAR(25)	Whether the customer has device protection (Yes/No/No internet service)	No
13	TechSupport	VARCHAR(25)	Whether the customer has tech support (Yes/No/No internet service)	No
14	StreamingTV	VARCHAR(25)	Whether the customer has streaming TV (Yes/No/No internet service)	No
15	StreamingMovies	VARCHAR(25)	Whether the customer has streaming movies (Yes/No/No internet service)	No
16	NumServices	INT	The number of services the customer is using from the telecom	No
17	Contract	VARCHAR(25)	The term of the customer's contract (Month-to-month/One year/Two year)	No
18	PaperlessBilling	VARCHAR(3)	Whether the customer has paperless billing (Yes/No)	No
19	PaymentMethod	VARCHAR(25)	The customer's payment method	No
20	MonthlyCharges	DECIMAL(10, 2)	The amount charged to the customer every month	No
21	TotalCharges	DECIMAL(10, 2)	The total amount charged to the customer	No
22	numAdminTickets	INT	Number of administrative tickets raised by the customer	No
23	numTechTickets	INT	Number of technical tickets raised by the customer	No
24	Churn	VARCHAR(3)	Whether the customer has churned or not (Yes/No)	No

Results

Star Schema Dictionary

Column Name	Data Type	Description	NULL?	Key Type
FACT_CHURN				
1 customerID	VARCHAR(25)	Unique identifier for the customer	No	COMPOSITE PRIMARY KEY FOREIGN KEY to DIM_CUSTOMER
2 bundleID	INT	Unique identifier for the bundle	No	COMPOSITE PRIMARY KEY FOREIGN KEY to DIM_BUNDLE
3 contractID	INT	Unique identifier for the contract	No	COMPOSITE PRIMARY KEY FOREIGN KEY to DIM_CONTRACT
4 MonthlyCharges	DECIMAL(10, 2)	The amount charged to the customer every month	No	
5 TotalCharges	DECIMAL(10, 2)	The total amount charged to the customer	No	
6 numAdminTickets	INT	Number of administrative tickets raised by the customer	No	
7 numTechTickets	INT	Number of technical tickets raised by the customer	No	
8 Churn				
	VARCHAR(3)	Whether the customer has churned or not (Yes/No)	No	

Action

Results

Data Governance

Quality

Security

Compliance

Ethics

- Clear documentation of data source, transformations and lineage to maintain data integrity and accuracy of data.
- Prompt actions to correct any identified data quality issues.

- Strict RBAC to ensure that data is only accessible by authorized personnel.
- Apply LPA to achieve a more granular level of data security.

- Stay informed of and comply with all applicable data protection and privacy laws, such as GDPR, CCPA, etc.
- Establish procedures for data retention, archival, and destruction to meet compliance obligations.

- Respecting user privacy and handling data in a manner that is fair, transparent, and accountable.
- Ensuring that data use aligns with our Team values and the expectations of our stakeholders.

Action

Results

Data Governance

Quality

Security

Compliance

Ethics

- Clear documentation of data source, transformations and lineage to maintain data integrity and accuracy of data.
- Prompt actions to correct any identified data quality issues.

- Strict RBAC to ensure that data is only accessible by authorized personnel.
- Apply LPA to achieve a more granular level of data security.

- Stay informed of and comply with all applicable data protection and privacy laws, such as GDPR, CCPA, etc.
- Establish procedures for data retention, archival, and destruction to meet compliance obligations.

- Respecting user privacy and handling data in a manner that is fair, transparent, and accountable.
- Ensuring that data use aligns with our Team values and the expectations of our stakeholders.

Action

Results

Data Governance

Quality

Security

Compliance

Ethics

- Clear documentation of data source, transformations and lineage to maintain data integrity and accuracy of data.
- Prompt actions to correct any identified data quality issues.

- Strict RBAC to ensure that data is only accessible by authorized personnel.
- Apply LPA to achieve a more granular level of data security.

- Stay informed of and comply with all applicable data protection and privacy laws, such as GDPR, CCPA, etc.
- Establish procedures for data retention, archival, and destruction to meet compliance obligations.

- Respecting user privacy and handling data in a manner that is fair, transparent, and accountable.
- Ensuring that data use aligns with our Team values and the expectations of our stakeholders.

Action

Results

Data Governance

Quality

Security

Compliance

Ethics

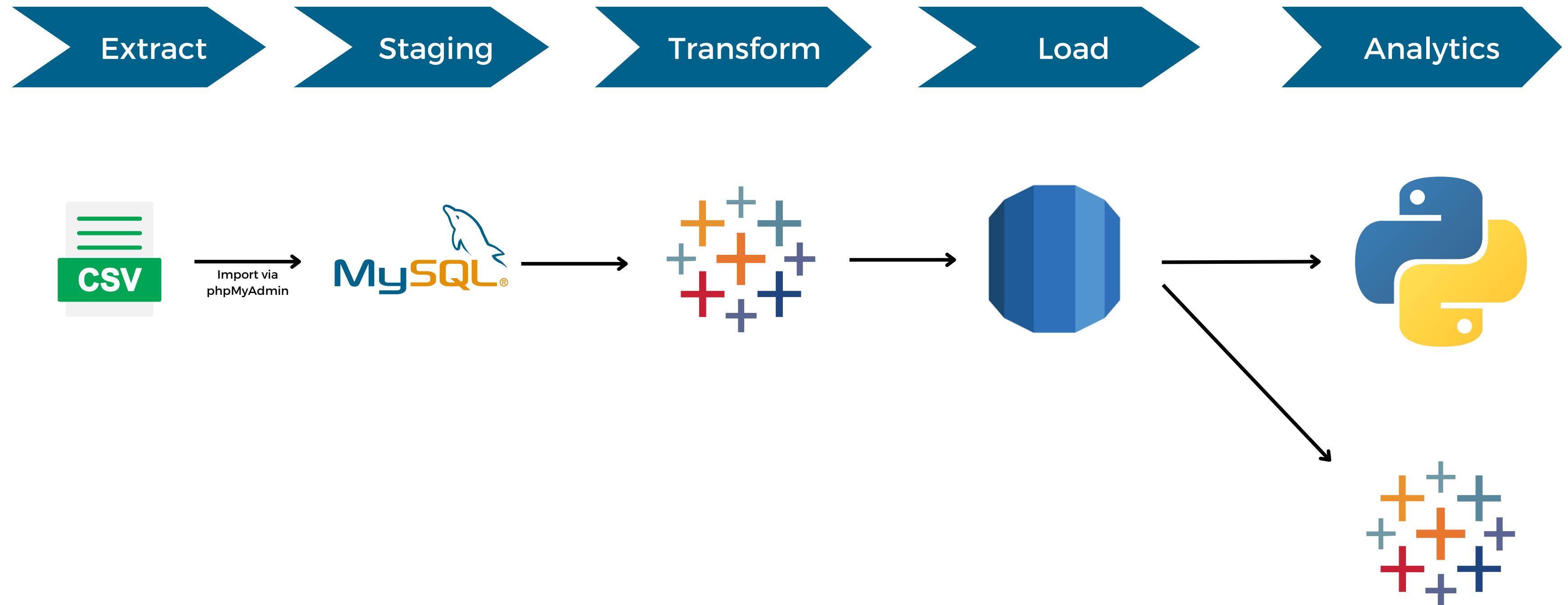
- Clear documentation of data source, transformations and lineage to maintain data integrity and accuracy of data.
- Prompt actions to correct any identified data quality issues.

- Strict RBAC to ensure that data is only accessible by authorized personnel.
- Apply LPA to achieve a more granular level of data security.

- Stay informed of and comply with all applicable data protection and privacy laws, such as GDPR, CCPA, etc.
- Establish procedures for data retention, archival, and destruction to meet compliance obligations.

- Respecting user privacy and handling data in a manner that is fair, transparent, and accountable.
- Ensuring that data use aligns with our Team values and the expectations of our stakeholders.

ETL Pipeline



Action

ETL Pipeline

Execution:

- Create new keys
- Set up the fact table
- Connect data source to AWS data warehouse

Results

Action

ETL Pipeline

Execution:

- Create new keys
- Set up the fact table
- Connect data source to AWS data warehouse

Results

The screenshot shows the AWS RDS console under the 'Databases' section. A single database named 'telecom-db' is listed. It is marked as 'Available' and is an 'Instance' of 'MySQL Community'. The engine is 'MySQL/Aurora' and the instance type is 'db.t3.micro'. The database has a size of 3.21% and 5 connections. Below the database list, there is a table titled 'Inbound rules (1/1)' which shows one rule allowing traffic from 0.0.0.0/0 on port 3306 via IPv4.

AWS

- Output Source is connected to AWS RDS
- Set-up remote access

The screenshot shows the MySQL Workbench interface. The 'Navigator' pane on the left lists two schemas: 'sys' and 'telecom_db'. The 'telecom_db' schema is currently selected. The main workspace is titled 'AWS Telecom'.

SQL

Can call on the data with SQL

FACT_CHURN

Result Grid								
Filter Rows: Export: Wrap Cell Content: Fetch rows:								
customerID	bundleID	contractID	MonthlyCharges	TotalCharges	numAdminTickets	numTechTickets	Churn	
7590-VHVEG	17	7	29.85	29.85	0	0	No	
5575-GNVDE	105	12	56.95	1889.50	0	0	No	
3668-QPYBK	113	8	53.85	108.15	0	0	Yes	
7795-CFOCW	45	9	42.30	1840.70	0	3	No	
9237-HQITU	129	7	70.70	151.65	0	0	Yes	
9305-CDSKC	269	7	99.65	820.50	0	0	Yes	

Record Count

- Fact_Churn: 7043 Rows
- Dim_Customer: 7043 Rows
- Dim_Bundle: 322 Rows
- Dim_Contract: 24 Rows

DIM_BUNDLE

Result Grid									
Filter Rows: Edit: Export/Import: Wrap Cell Content: Fetch rows:									
bundleID	PhoneService	MultipleLines	InternetService	OnlineSecurity	OnlineBackup	DeviceProtection	TechSupport	StreamingTV	Str
1	No	No phone service	DSL	No	No	No	No	No	No
2	No	No phone service	DSL	No	No	No	No	No	Yes
3	No	No phone service	DSL	No	No	No	No	Yes	No
4	No	No phone service	DSL	No	No	No	No	Yes	Yes
5	No	No phone service	DSL	No	No	No	Yes	No	No

DIM_CONTRACT

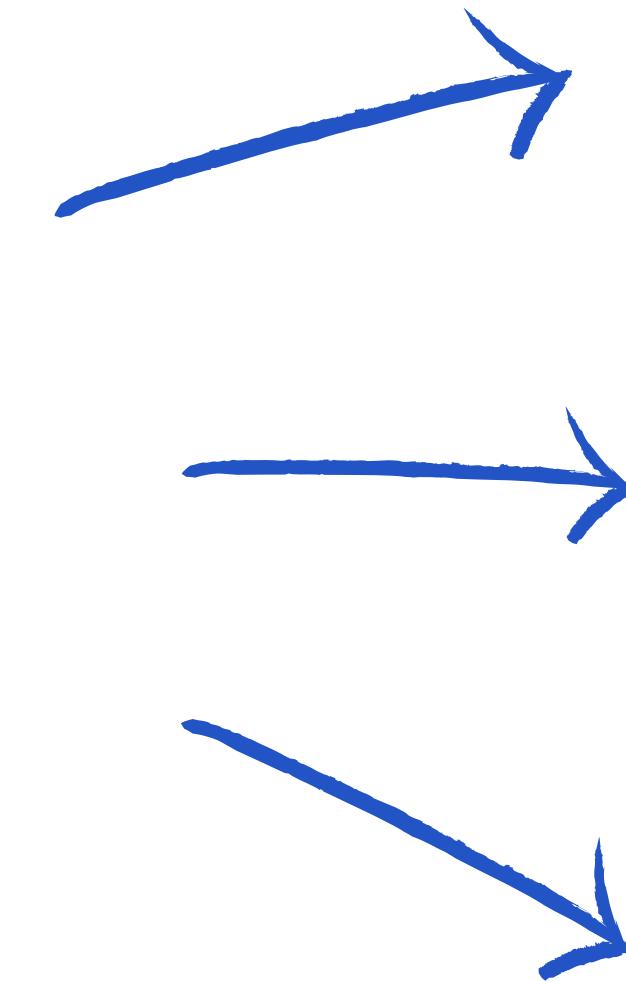
Result Grid				
Filter Rows: Edit: Export/Import: Wrap Cell Content: Fetch rows:				
	contractID	Contract	PaperlessBilling	PaymentMethod
1		Month-to-month	No	Bank transfer (automatic)
2		Month-to-month	No	Credit card (automatic)
3		Month-to-month	No	Electronic check
4		Month-to-month	No	Mailed check
5		Month-to-month	Yes	Bank transfer (automatic)

DIM_CUSTOMER

Result Grid						
Filter Rows: Edit: Export/Import: Wrap Cell Content: Fetch rows:						
	customerID	gender	SeniorCitizen	Partner	Dependents	tenure
0002-ORFBO	Female	0	Yes	Yes	9	
0003-MKNFE	Male	0	No	No	9	
0004-TLHLJ	Male	0	No	No	4	
0011-IGKFF	Male	1	Yes	No	13	
0013-EXCHZ	Female	1	Yes	No	3	

Descriptive Analytics

Technical
difficulties in
user base

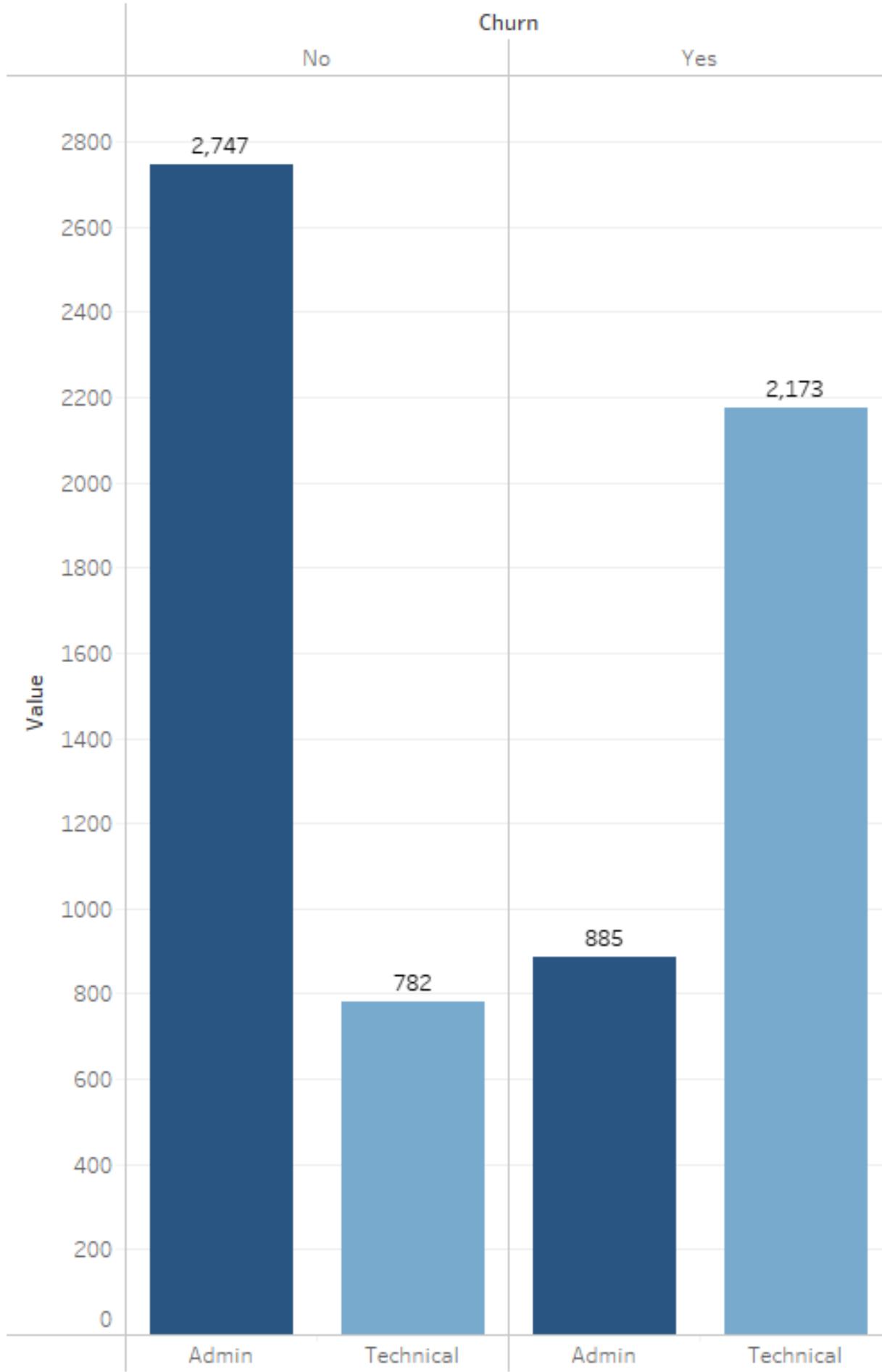


Support
Tickets

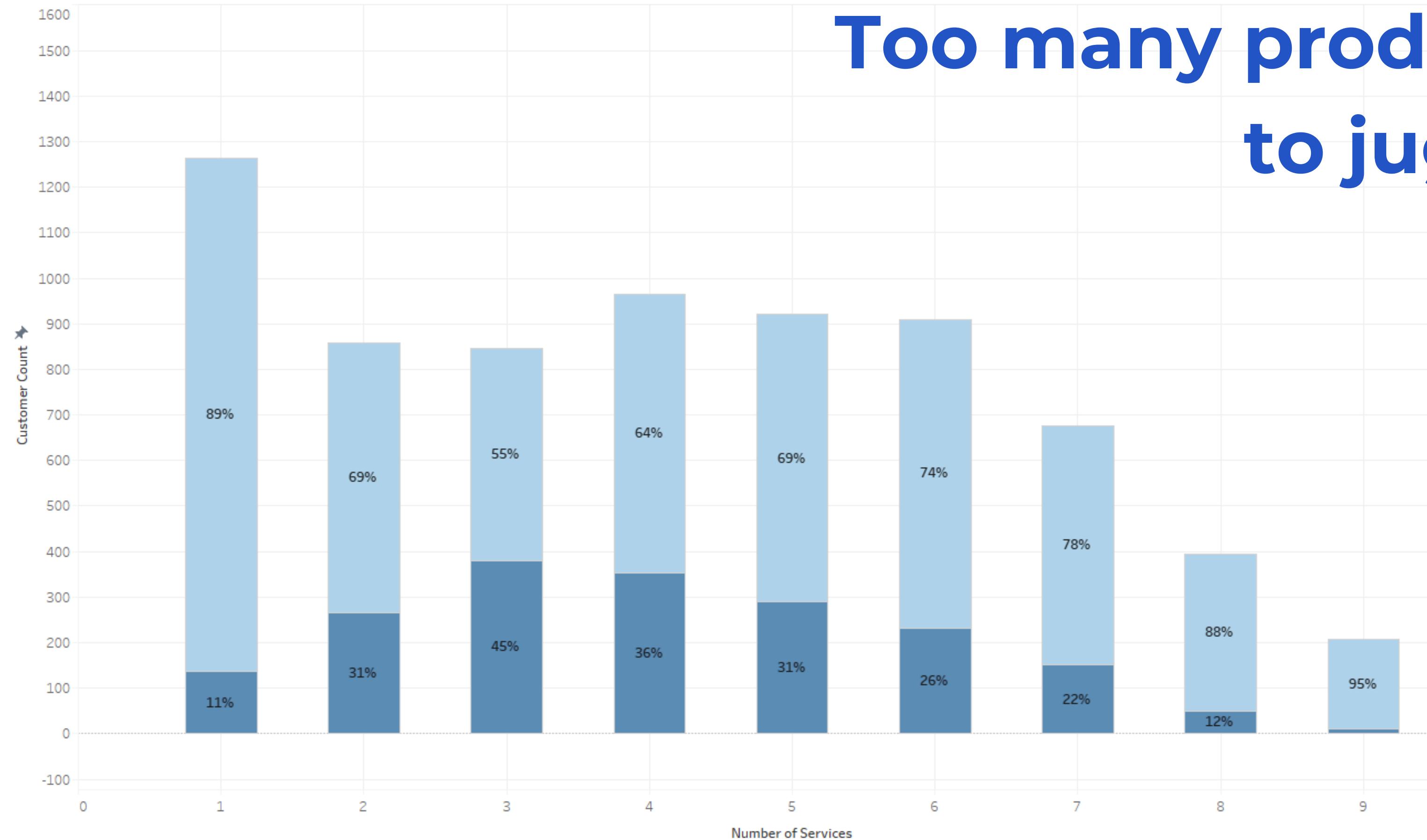
Tenure

Number of
Services

Technical tickets lead to higher churn

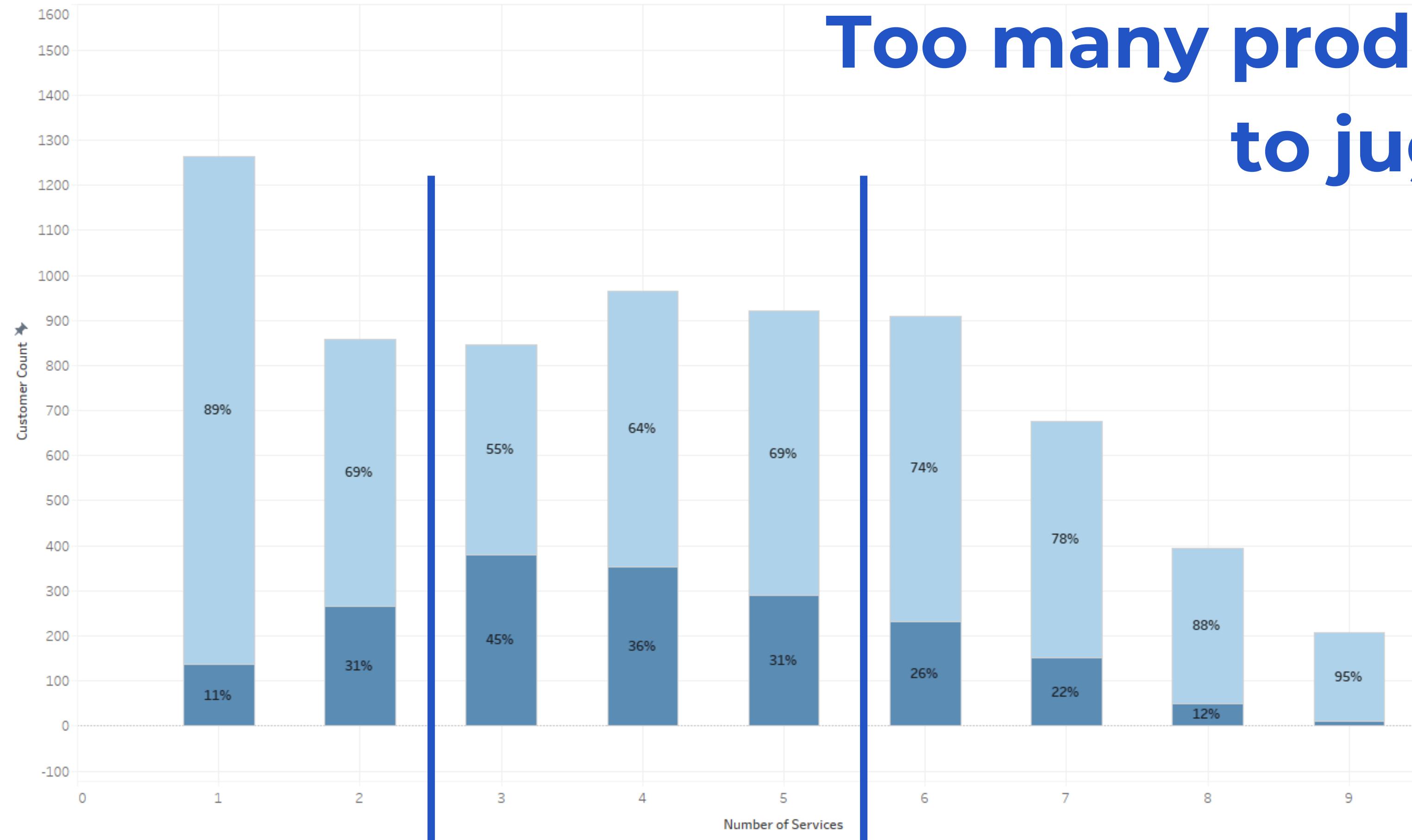


Distribution of Service Count



Too many products
to juggle

Distribution of Service Count



Too many products
to juggle

Too many products to juggle

Simple Users
1-2 services

- Only one or two services
- Small learning curve

Too many products to juggle

Simple Users
1-2 services

- Only one or two services
- Small learning curve

Power Users
6-9 services

- Technical comfort
- System familiarity

Too many products to juggle

Simple Users
1-2 services

Aspiring Users
3-5 services

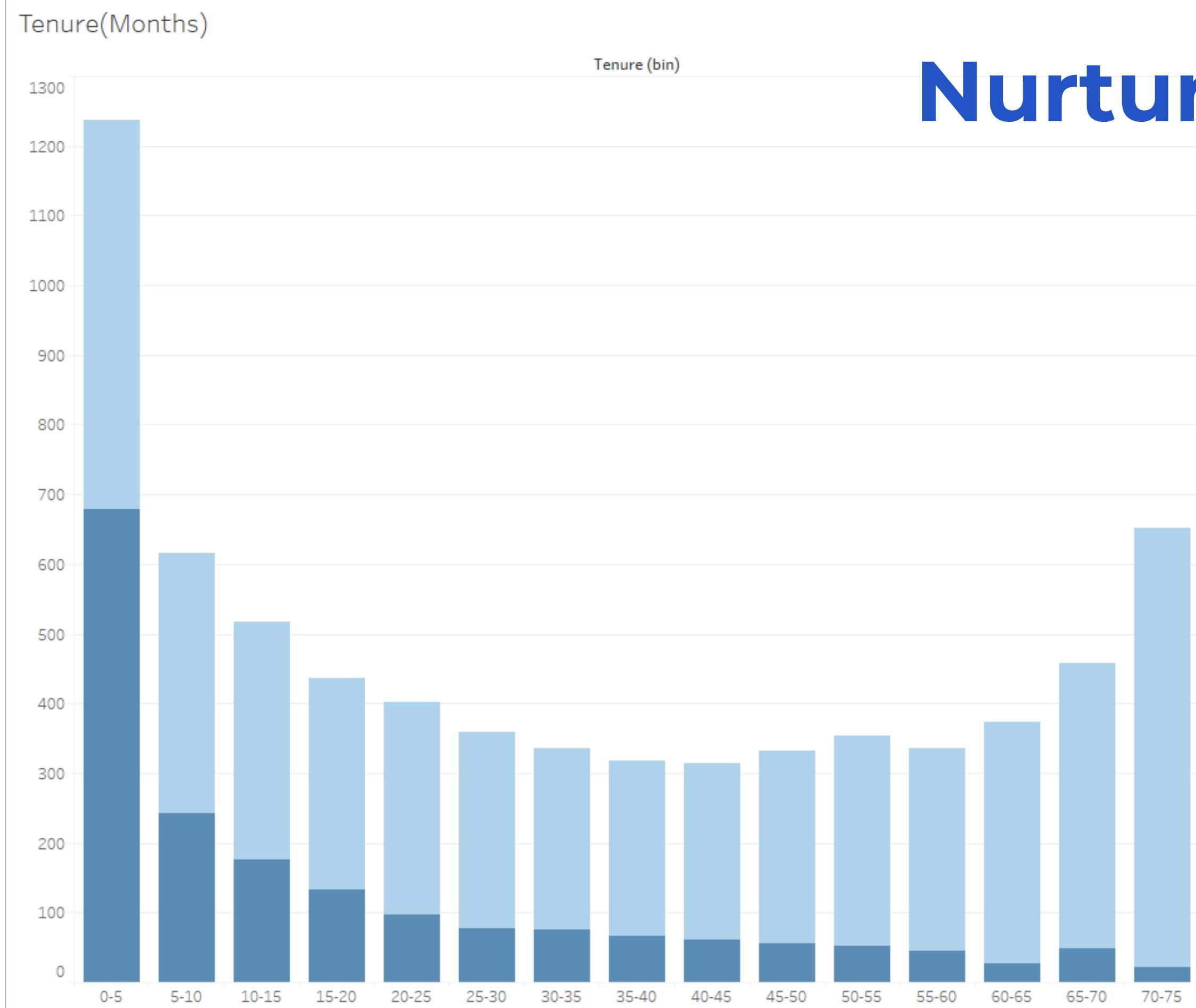
Power Users
6-9 services

- Only one or two services
- Small learning curve

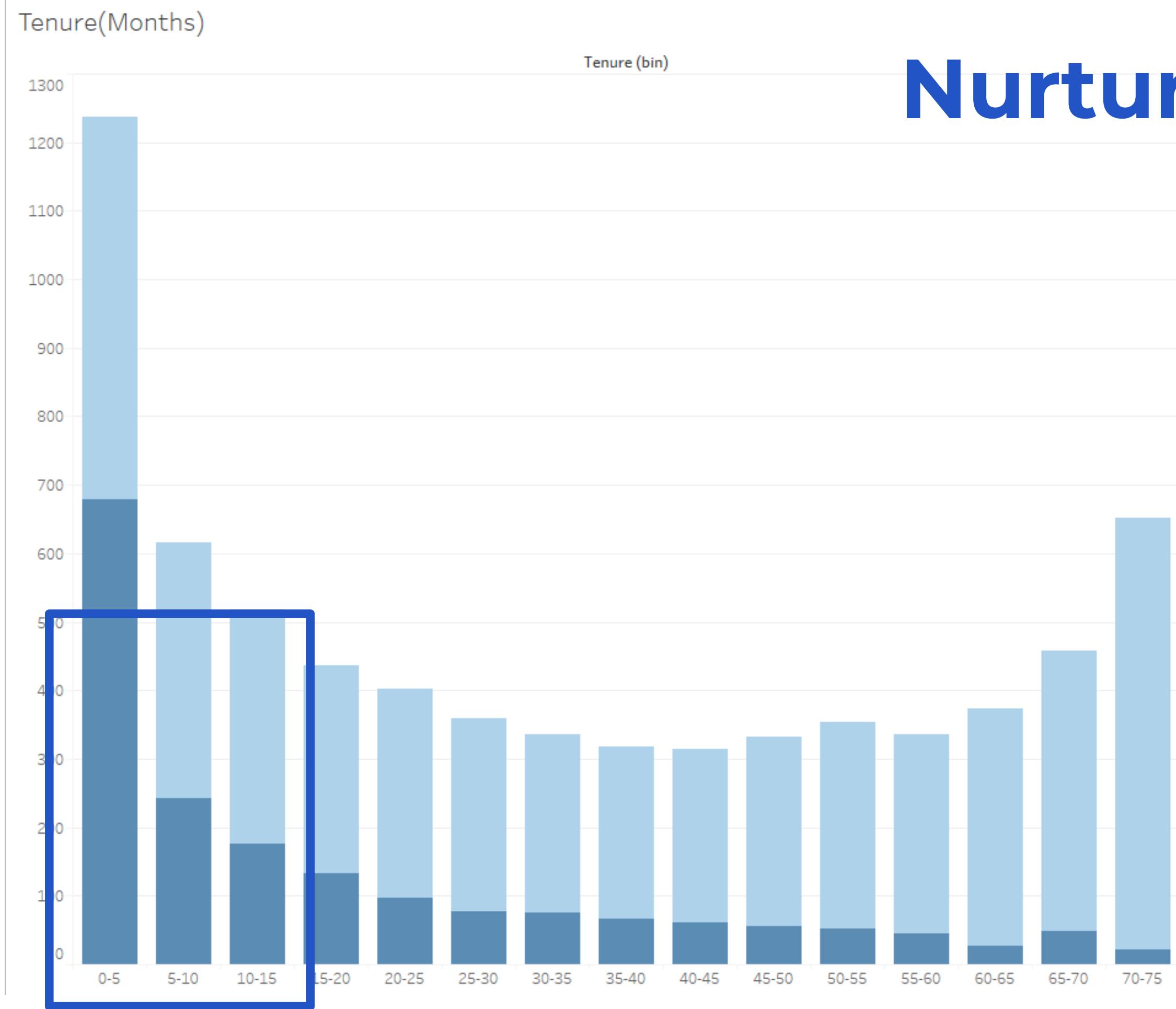
- Aspire to Bundle
- Technical discomfort

- Technical comfort
- System familiarity

Nurturing early user engagements



Nurturing early user engagements



Dashboard

AVG Charge

\$64.76

MoM Churn Rate

26.5%

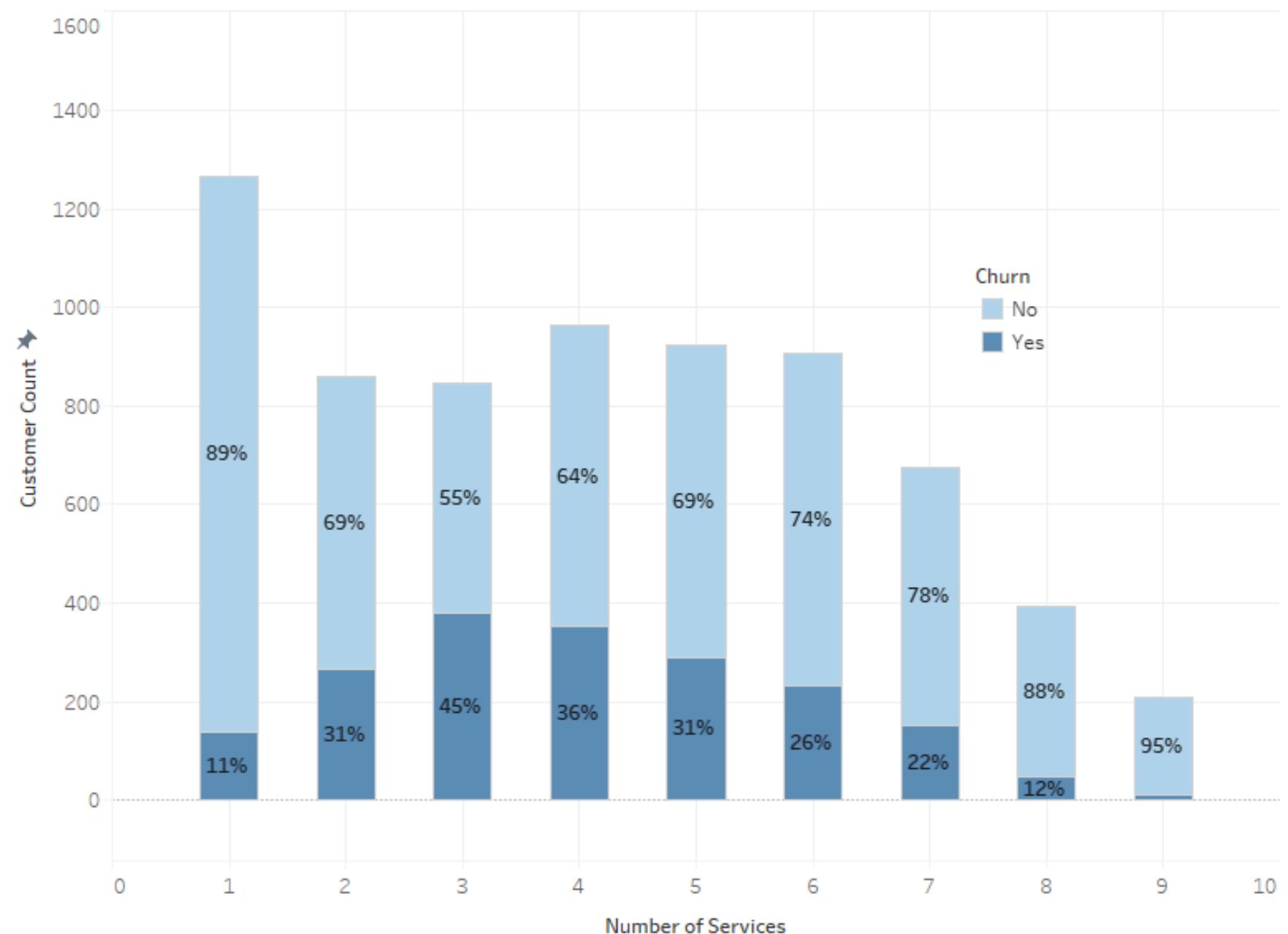
Total Annual Revenue Forgone

\$1,669,570

Support Tickets

3,632

Distribution of Service Count



Admin

Technical

2,955

Appendices

Appendix A

- The total budget allocated to this project is \$37,170 and the expected completion time is 3 weeks (i.e., 525 hours). See the breakdown of budget and time below.

Contractor Salary Breakdown			
Salary	Data Analyst	Data Engineer	Data Scientist
# of Employees	1	3	1
Hourly rate	\$57	\$72	\$81
Total hourly rate	\$57	\$216	\$81
Annual rate	\$120,000	\$150,000	\$170,000
Total annual rate	\$120,000	\$450,000	\$170,000
Total Budget and Hours Allocated To The Project			
Hours to complete the project	Total employee cost	Additional Comments	
525	\$37,170	Each employee spends 105 hours on this project	

Valentin: Product owner - business analyst
Minh: Scrum Master - data scientist
Kevin: Developer - data engineer
Mason: Developer - data engineer
Rachel: Developer - data engineer

