Vodafone Churn Analysis

1. Project Overview

This project analyzes Vodafone's customer churn data to understand why customers discontinue services. The objective is to identify churn patterns, high-risk customer segments, and key factors such as customer demographics, service usage, billing, contracts, and support interactions that influence customer attrition, in order to support data-driven customer retention strategies.

2. Dataset Summary

- **Rows:** 6,441
- **Columns:** 23

Key Features:

- **Customer Demographics:** Gender, senior citizen status, partner, dependents, and location
- **Customer Lifecycle:** Tenure with the company
- **Service Usage:** Phone service, internet service type, multiple lines, and value-added services
- **Billing & Payments:** Monthly charges, total charges, payment method, and paperless billing
- **Customer Support:** Administrative and technical support ticket counts
- **Target Variable:** Churn status (Yes/No)

3. Data Preparation & Cleaning (MySQL)

Before conducting any analysis, the raw Vodafone customer churn dataset was cleaned and prepared using MySQL to ensure data quality, consistency, and reliability. This step focused on identifying data inconsistencies, handling blank values, validating data integrity, and creating derived features to support meaningful analysis.

1. Initial Data Inspection

The dataset was first inspected to understand its structure and content.

Result Grid | Filter Rows: | Export: | Wrap Cell Content: | Fetch rows:

customerID	gender	SeniorCitizen	Dependents	tenure	PhoneService	MultipleLines	InternetService	OnlineSecurity	OnlineBackup	DeviceProtection	TechSupport	Stream
7590-VHVEG	Female	0	No	1	No	No phone service	DSL	No	Yes	No	No	No
5575-GNVDE	Unknown	0	No	34	Yes	No	DSL	Yes	No	Yes	No	No
3668-QPYBK	Male	0	No	2	Yes	No	DSL	Yes	Yes	No	No	No
7795-CFOCW	Male	0	No	45	No	No phone service	DSL	Yes	No	Yes	Yes	No
9237-HQITU	Female	0	No	2	Yes	No	Fiber optic	No	No	No	No	No
9305-CDSKC	Female	0	No	8	Yes	Yes	Fiber optic	No	No	Yes	No	Yes
1452-KIOVK	Male	0	Yes	22	Yes	Yes	Fiber optic	No	Yes	No	No	Yes
6713-OKOMC	Female	0	No	10	No	No phone service	DSL	Yes	No	No	No	No
7892-POOKP	Female	0	No	28	Yes	Yes	Fiber optic	No	No	Yes	Yes	Yes
6388-TABGU	Male	0	Yes	62	Yes	No	DSL	Yes	Yes	No	No	No
9763-GRSKD	Male	0	Yes	13	Yes	No	DSL	Yes	No	No	No	
7469-LKBCI	Male	0	No	16	Yes	No	No	No internet s...	No internet ...	No internet serv...	No internet ...	No inter
8091-TTVAX	Male	0	No	58	Yes	Yes	Fiber optic	No	No	Yes	No	Yes
0280-XJGEX	Male	0	No	49	Yes	Yes	Fiber optic	No	Yes	Yes	No	Yes
5129-JLPIS	Male	0	No	25	Yes	No	Fiber optic	Yes	No	Yes	Yes	Yes
3655-SNQYZ	Female	0	Yes	69	Yes	Yes	Fiber optic	Yes	Yes	Yes	Yes	Yes
8191-XWSZG	Female	0	No	52	Yes	No	No	No internet s...	No internet ...	No internet serv...	No internet ...	No inter
9959-WOFKT	Male	0	Yes	71	Yes	Yes	Fiber optic	Yes	No	Yes	No	Yes
4190-MFLUW	Female	0	Yes	10	Yes	No	DSL	No	No	Yes	Yes	No
4183-MYFRB	Unknown	0	No	21	Yes	No	Fiber optic	No	Yes	Yes	No	No

Result Grid | Filter Rows: | Export: | Wrap Cell Content: | Fetch rows:

StreamingTV	StreamingMovies	Contract	PaperlessBilling	PaymentMethod	MonthlyCharges	TotalCharges	numAdminTickets	numTechTickets	Churn	Location
No	No	Month-to-month	Yes	Electronic check	29.85	29.85	0	0	No	North - New York
No	No	One year	No	Mailed check	56.95	1889.5	0	0	No	West - California
No	No	Month-to-month	Yes	Mailed check	53.85	108.15	0	0	Yes	East - Illinois
No	No	One year	No	Bank transfer (automatic)	42.3	1840.75	0	3	No	West - Illinois
No	No	Month-to-month	Yes	Electronic check	70.7	151.65	0	0	Yes	East - Illinois
Yes	Yes	Month-to-month	Yes	Electronic check	99.65	820.5	0	0	Yes	South - California
Yes	No	Month-to-month	Yes	Credit card (automatic)	89.1	1949.4	0	0	No	North - New York
No	No	Month-to-month	No	Mailed check	29.75	301.9	0	0	No	East - Texas
Yes	Yes	Month-to-month	Yes	Electronic check	104.8	3046.05	0	2	Yes	North - Ohio
No	No	One year	No	Bank transfer (automatic)	56.15	3487.95	0	0	No	North - Florida
	No	Month-to-month	Yes	Mailed check	49.95	587.45	1	0	No	South - Florida
No internet ...	No internet serv...	Two year	No	Credit card (automatic)	18.95	326.8	0	0	No	South - Florida
Yes	Yes	One year	No	Credit card (automatic)	100.35	5681.1	0	0	No	North - California
Yes	Yes	Month-to-month	Yes	Bank transfer (automatic)	103.7	5036.3	5	4	Yes	West - Florida
Yes	Yes	Month-to-month	Yes	Electronic check	105.5	2686.05	0	0	No	East - Illinois
Yes	Yes	Two year	No	Credit card (automatic)	113.25	7895.15	0	0	No	East - Texas
No internet ...	No internet serv...	One year	No	Mailed check	20.65	1022.95	0	0	No	West - Florida
Yes	Yes	Two year	No	Bank transfer (automatic)	106.7	7382.25	0	4	No	North - California
No	No	Month-to-month	No	Credit card (automatic)	55.2	528.35	0	0	Yes	East - California
No	Yes	Month-to-month	Yes	Electronic check	90.05	1862.9	0	0	No	South - Florida

2. Standardizing Categorical Text Columns

During inspection, it was observed that some categorical columns contained leading or trailing spaces due to CSV import issues. These inconsistencies can cause incorrect grouping and misleading analysis.

Columns cleaned:

- gender
- StreamingMovies
- InternetService
- Contract
- PaymentMethod

Result Grid

gender
Female
Unknown
Male

Result Grid | Fi

StreamingMovies
No
Yes
No internet service

3. Handling Blank Values

During data preparation, a small number of categorical columns were found to contain blank values introduced during data import. These blanks were handled to ensure consistency in analysis. Gender-related blanks were categorized as unknown, and blank service-related entries were treated as non-usage of the service. After this step, no blank categorical values remained in the dataset.

4. Data Integrity Checks

Basic data integrity checks were performed to verify the total number of customer records and the distribution of churned and retained customers. Duplicate records and missing values in key columns were checked, and no data quality issues were found.

5. Feature Engineering

Two derived columns were created to support analysis. A binary churn indicator (churn_flag) was generated from the original **Churn** column, and tenure-based customer groups (tenure_group) were created from the **tenure** column to improve lifecycle-based churn analysis.

churn_flag	Total
0	4730
1	1711

tenure_group	Total
0-1 year	1891
2-4 years	1454
1-2 years	1053
4+ years	2043

4. Exploratory Data Analysis (EDA)

1. Overall Churn Overview

The first step of exploratory analysis focused on understanding the overall churn situation in the dataset. The distribution of churned and retained customers was reviewed to assess the scale of customer attrition and establish a baseline for further analysis.

Churn	customer_count
No	4730
Yes	1711

2. Churn by Key Customer Segments

This step explores how churn varies across **important customer segments** to identify early patterns before deeper business analysis. The focus is on high-impact categorical variables that commonly influence churn behaviour.

For EDA, we will look at churn distribution across:

- **Contract type**
- **Tenure group**
- **Payment method**

This helps highlight which segments show higher churn tendencies at a high level.

Contract type - Churn is highest among customers with month-to-month contracts, while customers on longer-term contracts show significantly lower churn.

Result Grid | Filter Rows:

Contract	Churn	customer_count
Month-to-month	No	2032
Month-to-month	Yes	1521
One year	No	1193
One year	Yes	144
Two year	No	1505
Two year	Yes	46

3. Numerical Overview

This step focuses on understanding the distribution of key numerical variables related to customer lifecycle and billing. Reviewing tenure and monthly charges helps provide context on how long customers stay with the company and the typical billing range before deeper churn analysis.

Result Grid | Filter Rows:

min_tenure	max_tenure	avg_tenure
1	72	32.3704

Result Grid | Filter Rows: | Export: | Wrap Cell

min_monthly_charges	max_monthly_charges	avg_monthly_charges
18.25	118.75	64.72530662940557

Observation

Customer tenure ranges from short-term to long-term subscriptions, indicating a mix of new and loyal customers. Monthly charges also show a wide spread, suggesting varied service usage and pricing plans across the customer base.

5.Business Analysis using SQL

1. Churn by Customer Demographics

This analysis examines how customer churn varies across key demographic attributes such as age group, relationship status, and household responsibilities. Understanding demographic churn patterns helps identify customer segments that may require targeted retention efforts.

Dependents vs Churn

Dependents	Churn	customer_count
No	No	3100
No	Yes	1410
Yes	No	1630
Yes	Yes	301

Senior Citizen vs Churn

SeniorCitizen	Churn	customer_count
0	No	4112
0	Yes	1271
1	No	618
1	Yes	440

Observation

- Senior citizens show a noticeably higher churn rate compared to non-senior customers.
- Customers without dependents churn significantly more than customers with dependents, indicating stronger retention among family-oriented households.

2. Churn by Contract Type

This analysis examines how customer churn varies across different contract types. Understanding contract-level churn helps assess the impact of long-term commitments on customer retention.

Contract	Churn	customer_count
Month-to-month	No	2032
Month-to-month	Yes	1521
One year	No	1193
One year	Yes	144
Two year	No	1505
Two year	Yes	46

Observation

Customers on month-to-month contracts show the highest churn, while churn decreases significantly for one-year and two-year contracts, indicating better retention with longer commitments.

3. Churn by Tenure Group

This analysis examines how customer churn varies across different stages of the customer lifecycle using tenure groups. It helps identify whether churn is more common among new or long-term customers.

Result Grid | Filter Rows:

tenure_group	Churn	customer_count
0-1 year	No	972
0-1 year	Yes	919
1-2 years	No	747
1-2 years	Yes	306
2-4 years	No	1159
2-4 years	Yes	295
4+ years	No	1852
4+ years	Yes	191

Observation

Customer churn is highest among **0-1 year** tenure customers and steadily decreases as tenure increases, indicating that early-stage customers are more likely to leave compared to long-term customers.

4. Churn vs Billing Behavior

This analysis examines how customer churn relates to billing behavior using monthly and total charges. It helps understand whether pricing and spending patterns influence customer attrition.

Result Grid | Filter Rows:

Churn	avg_monthly_charges
No	61.26
Yes	74.32

Result Grid | Filter Rows:

Churn	avg_total_charges
No	2549.34
Yes	1511.18

Observation

Churned customers have significantly higher average monthly charges but lower total charges, indicating that customers paying higher monthly fees tend to leave earlier in their lifecycle.

5. Churn by Payment Method

This analysis evaluates customer churn across different payment methods. It helps identify whether certain payment types are associated with higher customer attrition.

PaymentMethod	Churn	customer_count
Bank transfer (automatic)	No	1177
Bank transfer (automatic)	Yes	232
Credit card (automatic)	No	1180
Credit card (automatic)	Yes	213
Electronic check	No	1196
Electronic check	Yes	985
Mailed check	No	1177
Mailed check	Yes	281

Observation

Customers using **electronic check** show noticeably higher churn compared to other payment methods, while customers using **automatic bank transfer or credit card** exhibit better retention.

6. Impact of Customer Support on Churn

This analysis examines whether customer support interactions are associated with higher churn. It focuses on administrative and technical support ticket counts to assess their relationship with customer attrition.

Churn	avg_tech_tickets
No	0.15
Yes	1.15

Observation

Churned customers raise significantly more technical support tickets compared to retained customers, indicating a strong relationship between technical issues and customer churn.

7.Geographic Churn Analysis

This analysis examines churn patterns across different customer locations to identify regions with higher customer attrition. It helps detect any geographic concentration of churn risk.

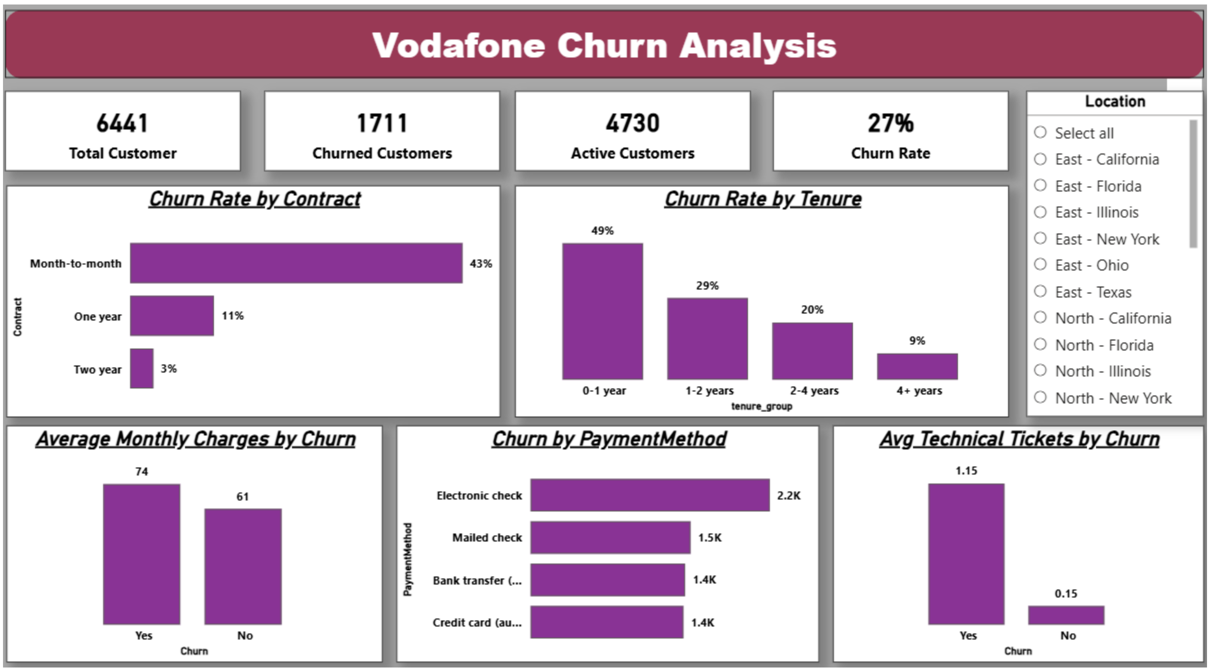
Location	churn_no	churn_yes	total_customers
North - Florida	186	95	281
South - California	192	84	276
South - Ohio	217	82	299
South - Florida	199	80	279
West - California	196	79	275
East - New York	211	76	287
North - New York	199	75	274
West - Illinois	186	75	261
South - Texas	187	75	262

Observation

Customer churn varies across locations, with certain regions such as **North–Florida and South–California** showing higher churn counts compared to others. This indicates potential regional differences in customer experience or service quality.

6. Dashboard in Power BI

Finally, we built an interactive dashboard in Power BI to present insights visually.



7. Key Insights

- Customer churn is highest among **month-to-month contract** customers, while long-term contracts show significantly lower churn rates.
- **Early-tenure customers (0–1 year)** have the highest churn rate (49%), indicating that churn risk is greatest during the initial customer lifecycle stage.
- Customers with **higher monthly charges** are more likely to churn, despite having lower overall lifetime value.
- **Electronic check** payment users exhibit higher churn compared to customers using automatic payment methods.
- Churned customers raise **substantially more technical support tickets**, highlighting a strong link between service issues and customer attrition.

8. Business Recommendations

- Introduce targeted **onboarding and engagement programs** for new customers, especially during the first year, to reduce early-stage churn.
- Encourage customers to move from **month-to-month contracts** to longer-term plans through discounts or bundled offers to improve retention.
- Review and optimize **pricing strategies** for customers with higher monthly charges to reduce price-sensitive churn.
- Promote **automatic payment methods** (bank transfer or credit card) to customers using electronic checks, as these users show higher churn.
- Improve **technical support responsiveness and service quality**, as higher technical ticket volumes are strongly associated with churn.