## Implementation

```
sandbox login: root
root@sandbox.hortonworks.com's password:
Last login: Tue Oct 22 02:09:12 2024 from 172.17.0.2
[root@sandbox ~]# hadoop fs -get /Telecom_customer_churn_analysis/input/telecom.csv
```

```
hive> show databases;
OK
aadhaardb
airline
default
finance_db
foodmart
inventory_db
lab8
lab9
market_db
product_db
sales_db
student_db
xademo
Time taken: 2.127 seconds, Fetched: 13 row(s)
hive> create database telecom;
OK
Time taken: 0.128 seconds
hive> use telecom;
OK
Time taken: 0.228 seconds
hive> show tables;
OK
```

### 3.1 Table Creation

The schema for the telecom_churn table is defined as follows:

```
CREATE TABLE telecom_churn (
    gender STRING,
    SeniorCitizen INT,
    Partner STRING,
    Dependents STRING,
    tenure INT,
    PhoneService STRING,
    MultipleLines STRING,
    InternetService STRING,
    OnlineSecurity STRING,
    OnlineBackup STRING,
    DeviceProtection STRING,
    TechSupport STRING,
    StreamingTV STRING,
    StreamingMovies STRING,
    Contract STRING,
    PaperlessBilling STRING,
    PaymentMethod STRING,
    MonthlyCharges FLOAT,
    TotalCharges FLOAT,
    Churn STRING
)
ROW FORMAT DELIMITED
FIELDS TERMINATED BY ','
LINES TERMINATED BY '\n';
```

```
hive> CREATE TABLE telecom_churn (
    >
    >     gender STRING,
    >
    >     SeniorCitizen STRING,
    >
    >     Partner STRING,
    >
    >     Dependents STRING,
    >
    >     tenure INT,
    >
    >     PhoneService STRING,
    >
    >     MultipleLines STRING,
    >
    >     InternetService STRING,
    >
    >     OnlineSecurity STRING,
    >
    >     OnlineBackup STRING,
    >
    >     DeviceProtection STRING,
    >
    >     TechSupport STRING,
    >
    >     StreamingTV STRING,
    >
    >     StreamingMovies STRING,
    >
    >     StreamingTV STRING,
    >
    >     StreamingMovies STRING,
    >
    >     Contract STRING,
    >
    >     PaperlessBilling STRING,
    >
    >     PaymentMethod STRING,
    >
    >     MonthlyCharges FLOAT,
    >
    >     TotalCharges FLOAT,
    >
    >     Churn STRING
    >
    > )
    >
    > ROW FORMAT DELIMITED
    >
    > FIELDS TERMINATED BY ','
    >
    > LINES TERMINATED BY '\n';
OK
Time taken: 0.446 seconds
```

### 3.2 Data Loading and Initial Query

**Load data into the table and display the first 5 records**

LOAD DATA INPATH '/path_to_data/telecom_churn.csv' INTO TABLE telecom_churn;

SELECT * FROM telecom_churn LIMIT 5;

```
hive> load data local inpath 'telecom.csv' into table telecom_churn;
Loading data to table telecom.telecom_churn
Table telecom.telecom_churn stats: [numFiles=1, numRows=0, totalSize=2777307, rawDataSize=0]
OK
Time taken: 1.011 seconds
hive> select * from telecom_churn limit 5;
OK
Female  No      Yes     No      1       No      No phone service        DSL     No      Yes     No      No      No      No      Month-to-mon
th      Yes     Electronic check        29.85   29.85   No
Male    No      No      No      34      Yes     No      DSL     Yes     No      Yes     No      No      No      One year        No      Mail
ed check        56.95   1889.5  No
Male    No      No      No      2       Yes     No      DSL     Yes     Yes     No      No      No      No      Month-to-month  Yes     Mail
ed check        53.85   108.15  Yes
Male    No      No      No      45      No      No phone service        DSL     Yes     No      Yes     Yes     No      No      One year   N
o       Bank transfer (automatic)       42.3    1840.75 No
Female  No      No      No      2       Yes     No      Fiber optic     No      No      No      No      No      No      Month-to-month  YesE
lectronic check 70.7    151.65  Yes
Time taken: 0.165 seconds, Fetched: 5 row(s)
```

### 3.3 Key Hive Queries and Outputs

### 1. Overall Churn Rate

SELECT

(COUNT(CASE WHEN Churn = 'Yes' THEN 1 END) * 100.0 / COUNT(*)) AS churn_rate_percentage

FROM telecom_churn;

```
hive> SELECT
    >
    >     (COUNT(CASE WHEN Churn = 'Yes' THEN 1 END) * 100.0 / COUNT(*)) AS churn_rate_percentage
    >
    > FROM telecom_churn;
Query ID = root_20241022035345_826796a4-4bcc-4f2c-aab2-5f70651aed80
Total jobs = 1
Launching Job 1 out of 1


Status: Running (Executing on YARN cluster with App id application_1729443979892_0090)

--------------------------------------------------------------------------------
        VERTICES      STATUS  TOTAL  COMPLETED  RUNNING  PENDING  FAILED  KILLED
--------------------------------------------------------------------------------
Map 1 ..........   SUCCEEDED      1          1        0        0       0       0
Reducer 2 ......   SUCCEEDED      1          1        0        0       0       0
--------------------------------------------------------------------------------
VERTICES: 02/02  [==========================>>] 100%  ELAPSED TIME: 11.88 s
--------------------------------------------------------------------------------
OK
49.464214314274294
Time taken: 12.842 seconds, Fetched: 1 row(s)
```

**2. Churn Rate by Gender**

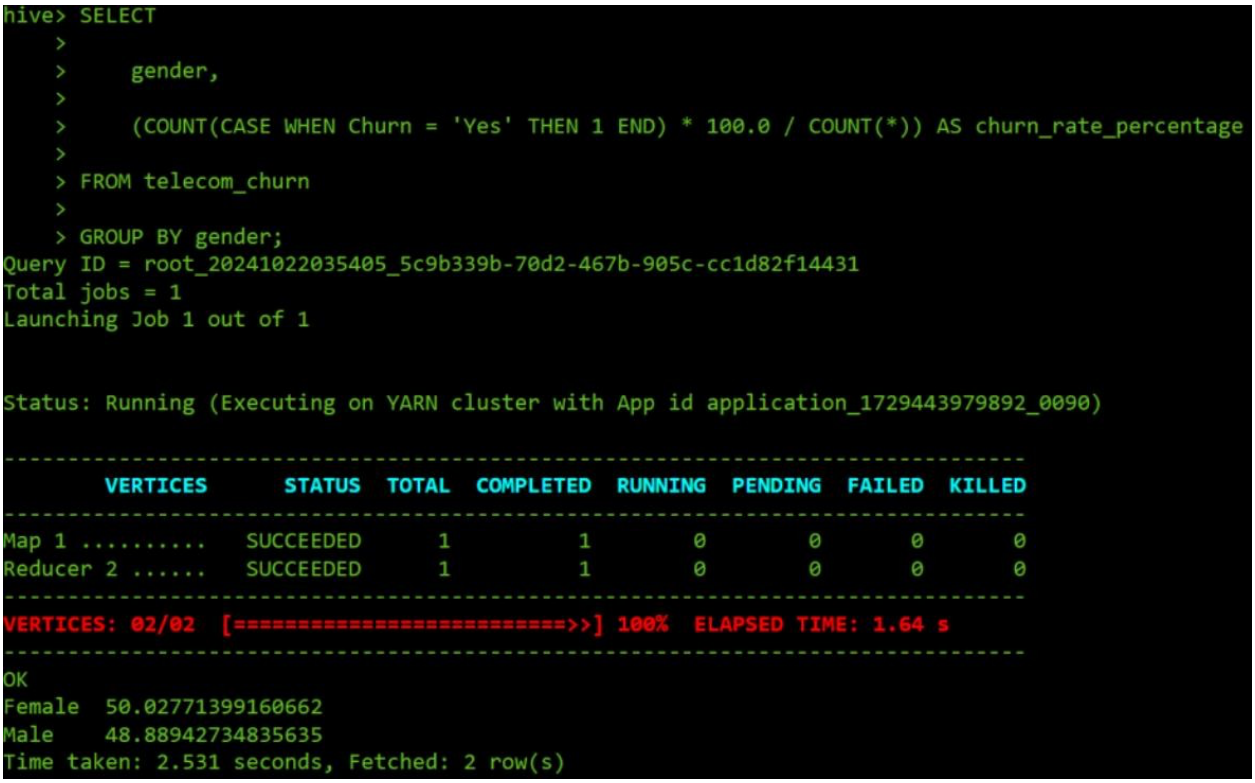
SELECT gender,

(COUNT(CASE WHEN Churn = 'Yes' THEN 1 END) * 100.0 / COUNT(*)) AS

churn_rate_percentage

FROM telecom_churn

GROUP BY gender;




**3. Churn Rate by Contract Type**

SELECT

Contract,

(COUNT(CASE WHEN Churn = 'Yes' THEN 1 END) * 100.0 / COUNT(*)) AS

churn_rate_percentage FROM telecom_churn GROUP BY Contract;

```
hive> SELECT
    >
    >     Contract,
    >
    >     (COUNT(CASE WHEN Churn = 'Yes' THEN 1 END) * 100.0 / COUNT(*)) AS churn_rate_percentage
    >
    > FROM telecom_churn
    >
    > GROUP BY Contract;
Query ID = root_20241022035417_8a77dcd2-85d9-4360-ab6c-3912e6311efe
Total jobs = 1
Launching Job 1 out of 1


Status: Running (Executing on YARN cluster with App id application_1729443979892_0090)

--------------------------------------------------------------------------------
        VERTICES      STATUS  TOTAL  COMPLETED  RUNNING  PENDING  FAILED  KILLED
--------------------------------------------------------------------------------
Map 1 ..........   SUCCEEDED      1          1        0        0       0       0
Reducer 2 ......   SUCCEEDED      1          1        0        0       0       0
--------------------------------------------------------------------------------
VERTICES: 02/02  [==========================>>] 100%  ELAPSED TIME: 8.76 s
--------------------------------------------------------------------------------
OK
Month-to-month  49.187823366622546
One year        50.14943215780036
Two year        49.05207583393329
Time taken: 9.597 seconds, Fetched: 3 row(s)
```

**4. Customer churn count by Internet Service**

```
SELECT
        InternetService,
        COUNT(*) AS customer_count
FROM telecom_churn
WHERE Churn = 'Yes'
GROUP BY InternetService
ORDER BY InternetService;
```

```
hive> SELECT
    >
    >     InternetService,
    >
    >     COUNT(*) AS customer_count
    >
    > FROM telecom_churn
    >
    > WHERE Churn = 'Yes'
    >
    > GROUP BY InternetService
    >
    > ORDER BY InternetService;
Query ID = root_20241022070642_eee60576-01d9-476d-a50d-2fccfc4f8706
Total jobs = 1
Launching Job 1 out of 1


Status: Running (Executing on YARN cluster with App id application_1729443979892_0094)

--------------------------------------------------------------------------------
        VERTICES      STATUS  TOTAL  COMPLETED  RUNNING  PENDING  FAILED  KILLED
--------------------------------------------------------------------------------
Map 1 ..........   SUCCEEDED      1          1        0        0       0       0
Reducer 2 ......   SUCCEEDED      1          1        0        0       0       0
Reducer 3 ......   SUCCEEDED      1          1        0        0       0       0
--------------------------------------------------------------------------------
VERTICES: 03/03  [==========================>>] 100%  ELAPSED TIME: 5.98 s
--------------------------------------------------------------------------------
OK
```

```
OK
DSL     4187
Fiber optic     4024
No      4160
Time taken: 6.853 seconds, Fetched: 3 row(s)
```

**5. Churn rate for each payment method along with average monthly charges**

SELECT

PaymentMethod,

AVG(MonthlyCharges) AS avg_monthly_charges,
COUNT(CASE WHEN Churn = 'Yes' THEN 1 END) * 100.0 / COUNT(*) AS churn_rate
FROM telecom_churn
GROUP BY PaymentMethod;

```
hive> SELECT
    >
    >     PaymentMethod,
    >
    >     AVG(MonthlyCharges) AS avg_monthly_charges,
    >
    >     COUNT(CASE WHEN Churn = 'Yes' THEN 1 END) * 100.0 / COUNT(*) AS churn_rate
    >
    > FROM telecom_churn
    >
    > GROUP BY PaymentMethod;
Query ID = root_20241022071843_a0e78ca7-4b56-41ec-9ef5-2aff0e3df8dc
Total jobs = 1
Launching Job 1 out of 1
Tez session was closed. Reopening...
Session re-established.


Status: Running (Executing on YARN cluster with App id application_1729443979892_0095)

--------------------------------------------------------------------------------
        VERTICES      STATUS  TOTAL  COMPLETED  RUNNING  PENDING  FAILED  KILLED
--------------------------------------------------------------------------------
Map 1 ..........   SUCCEEDED      1          1        0        0       0       0
Reducer 2 ......   SUCCEEDED      1          1        0        0       0       0
--------------------------------------------------------------------------------
VERTICES: 02/02  [==========================>>] 100%  ELAPSED TIME: 7.27 s
--------------------------------------------------------------------------------
OK
```

```
OK
Bank transfer (automatic)       70.31778999131679       49.77692797960484
Credit card (automatic) 69.32330386078783       49.582396402184386
Electronic check        69.38573621809996       50.08624745177983
Mailed check    70.62167670275294       48.37709998368945
Time taken: 14.62 seconds, Fetched: 4 row(s)
```

**6. Churn Rate by Senior Citizens**

SELECT

SeniorCitizen,

(COUNT(CASE WHEN Churn = 'Yes' THEN 1 END) * 100.0 / COUNT(*))

ASchurn_rate_percentage FROM telecom_churn GROUP BY SeniorCitizen;

```
hive> SELECT
    >
    >     SeniorCitizen,
    >
    >     (COUNT(CASE WHEN Churn = 'Yes' THEN 1 END) * 100.0 / COUNT(*)) AS churn_rate_percentage
    >
    > FROM telecom_churn
    >
    > GROUP BY SeniorCitizen;
Query ID = root_20241022035537_44fa55a1-c731-43ef-a881-a0778bf3cc82
Total jobs = 1
Launching Job 1 out of 1


Status: Running (Executing on YARN cluster with App id application_1729443979892_0090)

--------------------------------------------------------------------------------
        VERTICES      STATUS  TOTAL  COMPLETED  RUNNING  PENDING  FAILED  KILLED
--------------------------------------------------------------------------------
Map 1 ..........   SUCCEEDED      1          1        0        0       0       0
Reducer 2 ......   SUCCEEDED      1          1        0        0       0       0
--------------------------------------------------------------------------------
VERTICES: 02/02  [==========================>>] 100%  ELAPSED TIME: 10.66 s
--------------------------------------------------------------------------------
OK
No      49.39643456711168
Yes     49.53203743700504
Time taken: 11.461 seconds, Fetched: 2 row(s)
```

### 7. Total Revenue of churned customers and non-churned customers

SELECT SUM(CASE WHEN Churn = 'Yes' THEN TotalCharges ELSE 0 END) AS total_revenue_churn_yes, SUM(CASE WHEN Churn = 'No' THEN TotalCharges ELSE 0 END) AS total_revenue_churn_no FROM telecom_churn;

```
hive> SELECT
    >
    >     SUM(CASE WHEN Churn = 'Yes' THEN TotalCharges ELSE 0 END) AS total_revenue_churn_yes,
    >
    >     SUM(CASE WHEN Churn = 'No' THEN TotalCharges ELSE 0 END) AS total_revenue_churn_no
    >
    > FROM telecom_churn;
Query ID = root_20241022073016_45f450dd-7557-46c0-9bc1-0d307ba6ed64
Total jobs = 1
Launching Job 1 out of 1
Tez session was closed. Reopening...
Session re-established.


Status: Running (Executing on YARN cluster with App id application_1729443979892_0096)

--------------------------------------------------------------------------------
        VERTICES      STATUS  TOTAL  COMPLETED  RUNNING  PENDING  FAILED  KILLED
--------------------------------------------------------------------------------
Map 1 ..........   SUCCEEDED      1          1        0        0       0       0
Reducer 2 ......   SUCCEEDED      1          1        0        0       0       0
--------------------------------------------------------------------------------
VERTICES: 02/02  [==========================>>] 100%  ELAPSED TIME: 12.10 s
--------------------------------------------------------------------------------
OK
6.146969944462013E7     6.363080120060921E7
Time taken: 24.89 seconds, Fetched: 1 row(s)
```

### 8. Average Tenure of Customers

SELECT AVG(tenure) AS avg_tenure FROM telecom_churn;

```
hive> SELECT AVG(tenure) AS avg_tenure FROM telecom_churn;
Query ID = root_20241022035613_ec4a3ca2-d363-46a8-a461-a19fc9375449
Total jobs = 1
Launching Job 1 out of 1


Status: Running (Executing on YARN cluster with App id application_1729443979892_0090)

--------------------------------------------------------------------------------
        VERTICES      STATUS  TOTAL  COMPLETED  RUNNING  PENDING  FAILED  KILLED
--------------------------------------------------------------------------------
Map 1 ..........   SUCCEEDED      1          1        0        0       1       0
Reducer 2 ......   SUCCEEDED      1          1        0        0       0       0
--------------------------------------------------------------------------------
VERTICES: 02/02  [==========================>>] 100%  ELAPSED TIME: 13.67 s
--------------------------------------------------------------------------------
OK
36.65293882447021
Time taken: 14.555 seconds, Fetched: 1 row(s)
```

**9. Tenure Distribution by Churn Status**

SELEC

Churn,

ROUND(AVG(tenure), 2) AS avg_tenure FROM telecom_churn

GROUP BY Churn;

```
hive> SELECT
    >
    >     Churn,
    >
    >     ROUND(AVG(tenure), 2) AS avg_tenure
    >
    > FROM telecom_churn
    >
    > GROUP BY Churn;
Query ID = root_20241022035650_ba12dbef-82cb-4043-9845-401061e868da
Total jobs = 1
Launching Job 1 out of 1


Status: Running (Executing on YARN cluster with App id application_1729443979892_0090)

--------------------------------------------------------------------------------
        VERTICES      STATUS  TOTAL  COMPLETED  RUNNING  PENDING  FAILED  KILLED
--------------------------------------------------------------------------------
Map 1 ..........   SUCCEEDED      1          1        0        0       0       0
Reducer 2 ......   SUCCEEDED      1          1        0        0       0       0
--------------------------------------------------------------------------------
VERTICES: 02/02  [==========================>>] 100%  ELAPSED TIME: 8.31 s
--------------------------------------------------------------------------------
OK
No      36.86
Yes     36.44
Time taken: 9.03 seconds, Fetched: 2 row(s)
```

**10. Churn Rate by Streaming Services**

SELECT

StreamingTV,

StreamingMovies,

(COUNT(CASE WHEN Churn = 'Yes' THEN 1 END) * 100.0 / COUNT(*)) AS

churn_rate_percentage

FROM telecom_churn

GROUP BY StreamingTV, StreamingMovies;

```
hive> SELECT
    >
    >     StreamingTV,
    >
    >     StreamingMovies,
    >
    >     (COUNT(CASE WHEN Churn = 'Yes' THEN 1 END) * 100.0 / COUNT(*)) AS churn_rate_percentage
    >
    > FROM telecom_churn
    >
    > GROUP BY StreamingTV, StreamingMovies;
Query ID = root_20241022035723_8db98516-45da-4f2c-bd97-0e9dc5b8b0b6
Total jobs = 1
Launching Job 1 out of 1


Status: Running (Executing on YARN cluster with App id application_1729443979892_0090)

--------------------------------------------------------------------------------
        VERTICES      STATUS  TOTAL  COMPLETED  RUNNING  PENDING  FAILED  KILLED
--------------------------------------------------------------------------------
Map 1 ..........   SUCCEEDED      1          1        0        0       0       0
Reducer 2 ......   SUCCEEDED      1          1        0        0       0       0
--------------------------------------------------------------------------------
VERTICES: 02/02  [==========================>>] 100%  ELAPSED TIME: 11.98 s
--------------------------------------------------------------------------------
```

```
OK
No      No      47.595222724338285
No      Yes     50.330165888226766
Yes     No      49.62916206406817
Yes     Yes     50.28717294192725
Time taken: 12.856 seconds, Fetched: 4 row(s)
```

**11. Customers with No Phone Service**

SELECT

COUNT(*) AS no_phone_service_customers

FROM telecom_churn WHERE PhoneService = 'No';

```
hive> SELECT
    >
    >     COUNT(*) AS no_phone_service_customers
    >
    > FROM telecom_churn
    >
    > WHERE PhoneService = 'No';
Query ID = root_20241022035801_789ab3f8-bb0f-4e4a-9a1d-e91d773da6fb
Total jobs = 1
Launching Job 1 out of 1


Status: Running (Executing on YARN cluster with App id application_1729443979892_0090)

--------------------------------------------------------------------------------
        VERTICES      STATUS  TOTAL  COMPLETED  RUNNING  PENDING  FAILED  KILLED
--------------------------------------------------------------------------------
Map 1 ..........   SUCCEEDED      1          1        0        0       0       0
Reducer 2 ......   SUCCEEDED      1          1        0        0       0       0
--------------------------------------------------------------------------------
VERTICES: 02/02  [==========================>>] 100%  ELAPSED TIME: 4.96 s
--------------------------------------------------------------------------------
OK
12599
Time taken: 5.479 seconds, Fetched: 1 row(s)
```

**12. Churn Rate by Paperless Billing**

SELECT

PaperlessBilling,

(COUNT(CASE WHEN Churn = 'Yes' THEN 1 END) * 100.0 / COUNT(*)) AS

churn_rate_percentage

FROM telecom_churn

GROUP BY PaperlessBilling;

```
hive> SELECT
    >
    >     PaperlessBilling,
    >
    >     (COUNT(CASE WHEN Churn = 'Yes' THEN 1 END) * 100.0 / COUNT(*)) AS churn_rate_percentage
    >
    > FROM telecom_churn
    >
    > GROUP BY PaperlessBilling;
Query ID = root_20241022035825_809c3d8f-00a5-48aa-b2e4-2b480519dd94
Total jobs = 1
Launching Job 1 out of 1


Status: Running (Executing on YARN cluster with App id application_1729443979892_0090)

--------------------------------------------------------------------------------
        VERTICES      STATUS  TOTAL  COMPLETED  RUNNING  PENDING  FAILED  KILLED
--------------------------------------------------------------------------------
Map 1 ..........   SUCCEEDED      1          1        0        0       0       0
Reducer 2 ......   SUCCEEDED      1          1        0        0       0       0
--------------------------------------------------------------------------------
VERTICES: 02/02  [==========================>>] 100%  ELAPSED TIME: 5.61 s
--------------------------------------------------------------------------------
OK
No      49.35549250101338
Yes     49.570019723865876
Time taken: 6.188 seconds, Fetched: 2 row(s)
```

**13. Gender-Wise Monthly Charges**

SELECT

gender,

AVG(MonthlyCharges) AS avg_monthly_charges FROM telecom_churn GROUP BY gender;

```
hive> SELECT
    >
    >     gender,
    >
    >     AVG(MonthlyCharges) AS avg_monthly_charges
    >
    > FROM telecom_churn
    >
    > GROUP BY gender;
Query ID = root_20241022035840_44b26765-9fb6-4d64-88a2-196e59b3d7b6
Total jobs = 1
Launching Job 1 out of 1


Status: Running (Executing on YARN cluster with App id application_1729443979892_0090)

--------------------------------------------------------------------------------
        VERTICES      STATUS  TOTAL  COMPLETED  RUNNING  PENDING  FAILED  KILLED
--------------------------------------------------------------------------------
Map 1 ..........   SUCCEEDED      1          1        0        0       0       0
Reducer 2 ......   SUCCEEDED      1          1        0        0       0       0
--------------------------------------------------------------------------------
VERTICES: 02/02  [==========================>>] 100%  ELAPSED TIME: 1.24 s
--------------------------------------------------------------------------------
OK
Female  69.68928652989986
Male    70.12920442323444
Time taken: 2.053 seconds, Fetched: 2 row(s)
```

**14. Most Popular Contract Type**

SELECT

Contract,

COUNT(*) AS customer_count FROM telecom_churn

GROUP BY Contract

ORDER BY customer_count DESC;

```
hive> SELECT
    >
    >     Contract,
    >
    >     COUNT(*) AS customer_count
    >
    > FROM telecom_churn
    >
    > GROUP BY Contract
    >
    > ORDER BY customer_count DESC;
Query ID = root_20241022035919_4143ddf7-310e-4aa5-a8b8-a9e0d2dadb5c
Total jobs = 1
Launching Job 1 out of 1


Status: Running (Executing on YARN cluster with App id application_1729443979892_0090)

--------------------------------------------------------------------------------
        VERTICES      STATUS  TOTAL  COMPLETED  RUNNING  PENDING  FAILED  KILLED
--------------------------------------------------------------------------------
Map 1 ..........   SUCCEEDED      1          1        0        0       0       0
Reducer 2 ......   SUCCEEDED      1          1        0        0       0       0
Reducer 3 ......   SUCCEEDED      1          1        0        0       0       0
--------------------------------------------------------------------------------
VERTICES: 03/03  [==========================>>] 100%  ELAPSED TIME: 5.32 s
--------------------------------------------------------------------------------
```

```
OK
One year        8365
Two year        8334
Month-to-month  8311
Time taken: 5.9 seconds, Fetched: 3 row(s)
```

**15. High-Risk Customers (High Monthly Charges, Short Tenure)**

SELECT

tenure, MonthlyCharges, Churn

FROM telecom_churn
WHERE MonthlyCharges > 90 AND tenure < 6;

```
hive> SELECT
    >
    >     tenure,
    >
    >     MonthlyCharges,
    >
    >     Churn
    >
    > FROM telecom_churn
    >
    > WHERE MonthlyCharges > 90 AND tenure < 6;
OK
4       99.7    No
3       92.69   Yes
1       96.27   No
1       108.55  Yes
2       100.14  Yes
3       104.32  Yes
5       98.2    No
1       119.13  No
2       90.65   Yes
3       98.8    Yes
1       100.21  Yes
2       96.9    No
2       101.32  No
3       116.56  No
1       102.57  Yes
5       110.41  No
2       98.55   No
1       98.55   Yes
```

**16. Retention Opportunities (Long Tenure, Low Charges)**

SELECT

customerID, tenure, MonthlyCharges, Churn

FROM telecom_churn
WHERE MonthlyCharges < 30 AND tenure > 24 AND Churn = 'No';

```
hive> SELECT
    >
    >     tenure,
    >
    >     MonthlyCharges,
    >
    >     Churn
    >
    > FROM telecom_churn
    >
    > WHERE MonthlyCharges < 30 AND tenure > 24 AND Churn = 'No';
OK
68      29.39   No
32      28.21   No
55      20.09   No
35      27.5    No
52      24.44   No
67      27.05   No
51      26.24   No
59      29.87   No
49      22.01   No
71      26.47   No
63      23.06   No
35      28.85   No
25      29.06   No
39      21.52   No
62      22.41   No
50      29.02   No
63      22.68   No
30      29.92   No
```

**17. Payment Method and Churn Correlation**

SELECT

PaymentMethod,

(COUNT(CASE WHEN Churn = 'Yes' THEN 1 END) * 100.0 / COUNT(*)) AS

churn_rate_percentage FROM telecom_churn GROUP BY PaymentMethod;

```
hive> SELECT
    >
    >     PaymentMethod,
    >
    >     (COUNT(CASE WHEN Churn = 'Yes' THEN 1 END) * 100.0 / COUNT(*)) AS churn_rate_percentage
    >
    > FROM telecom_churn
    >
    > GROUP BY PaymentMethod;
Query ID = root_20241022045626_fe326a5a-6f11-42a6-87d5-2ca3be0efe68
Total jobs = 1
Launching Job 1 out of 1
Tez session was closed. Reopening...
Session re-established.


Status: Running (Executing on YARN cluster with App id application_1729443979892_0091)

--------------------------------------------------------------------------------
        VERTICES      STATUS  TOTAL  COMPLETED  RUNNING  PENDING  FAILED  KILLED
--------------------------------------------------------------------------------
Map 1 ..........   SUCCEEDED      1          1        0        0       0       0
Reducer 2 ......   SUCCEEDED      1          1        0        0       0       0
--------------------------------------------------------------------------------
VERTICES: 02/02  [==========================>>] 100%  ELAPSED TIME: 16.08 s
--------------------------------------------------------------------------------
OK
Bank transfer (automatic)       49.77692797960484
Credit card (automatic) 49.582396402184386
Electronic check        50.08624745177983
Mailed check    48.37709998368945
```

**18. Churn by Device Protection**

SELECT

DeviceProtection,

(COUNT(CASE WHEN Churn = 'Yes' THEN 1 END) * 100.0 / COUNT(*)) AS

churn_rate_percentage

FROM telecom_churn

GROUP BY DeviceProtection;

```
hive> SELECT
    >
    >     DeviceProtection,
    >
    >     (COUNT(CASE WHEN Churn = 'Yes' THEN 1 END) * 100.0 / COUNT(*)) AS churn_rate_percentage
    >
    > FROM telecom_churn
    >
    > GROUP BY DeviceProtection;
Query ID = root_20241022050718_b98afad5-995e-4a6a-bf30-471833e1c19f
Total jobs = 1
Launching Job 1 out of 1


Status: Running (Executing on YARN cluster with App id application_1729443979892_0091)

--------------------------------------------------------------------------------
        VERTICES      STATUS  TOTAL  COMPLETED  RUNNING  PENDING  FAILED  KILLED
--------------------------------------------------------------------------------
Map 1 ..........   SUCCEEDED      1          1        0        0       0       0
Reducer 2 ......   SUCCEEDED      1          1        0        0       0       0
--------------------------------------------------------------------------------
VERTICES: 02/02  [==========================>>] 100%  ELAPSED TIME: 9.33 s
--------------------------------------------------------------------------------
OK
No      48.90983148310838
Yes     50.020017615501644
Time taken: 10.165 seconds, Fetched: 2 row(s)
```

**19. Churn Rate by Tech Support**

SELECT

TechSupport,

(COUNT(CASE WHEN Churn = 'Yes' THEN 1 END) * 100.0 / COUNT(*)) AS

churn_rate_percentage FROM telecom_churn GROUP BY TechSupport;

```
hive> SELECT
    >
    >     TechSupport,
    >
    >     (COUNT(CASE WHEN Churn = 'Yes' THEN 1 END) * 100.0 / COUNT(*)) AS churn_rate_percentage
    >
    > FROM telecom_churn
    >
    > GROUP BY TechSupport;
Query ID = root_20241022051035_3f3cec78-1902-4efd-a106-c7096247e69d
Total jobs = 1
Launching Job 1 out of 1


Status: Running (Executing on YARN cluster with App id application_1729443979892_0091)

--------------------------------------------------------------------------------
        VERTICES      STATUS  TOTAL  COMPLETED  RUNNING  PENDING  FAILED  KILLED
--------------------------------------------------------------------------------
Map 1 ..........   SUCCEEDED      1          1        0        0       0       0
Reducer 2 ......   SUCCEEDED      1          1        0        0       0       0
--------------------------------------------------------------------------------
VERTICES: 02/02  [==========================>>] 100%  ELAPSED TIME: 8.42 s
--------------------------------------------------------------------------------
OK
No      49.28605425987625
Yes     49.64527571751048
```

**20. Revenue condition by internet services**

SELECT

InternetService,

SUM(MonthlyCharges) AS total_revenue FROM telecom_churn

GROUP BY InternetService;

```
hive> SELECT
    >
    >     InternetService,
    >
    >     SUM(MonthlyCharges) AS total_revenue
    >
    > FROM telecom_churn
    >
    > GROUP BY InternetService;
Query ID = root_20241022051224_3350f470-8de3-4d20-a0bc-83152ec0c56e
Total jobs = 1
Launching Job 1 out of 1


Status: Running (Executing on YARN cluster with App id application_1729443979892_0091)

--------------------------------------------------------------------------------
        VERTICES      STATUS  TOTAL  COMPLETED  RUNNING  PENDING  FAILED  KILLED
--------------------------------------------------------------------------------
Map 1 ..........   SUCCEEDED      1          1        0        0       0       0
Reducer 2 ......   SUCCEEDED      1          1        0        0       0       0
--------------------------------------------------------------------------------
VERTICES: 02/02  [==========================>>] 100%  ELAPSED TIME: 8.55 s
--------------------------------------------------------------------------------
OK
DSL     587006.9196338654
Fiber optic     574278.4000072479
No      587090.3599090576
Time taken: 9.267 seconds, Fetched: 3 row(s)
```

**21. Tenure Distribution among Customers**

```
hive> SELECT
    >
    >     tenure,
    >
    >     COUNT(*) AS customer_count
    >
    > FROM
    >
    >     telecom_churn
    >
    > GROUP BY
    >
    >     tenure
    >
    > ORDER BY
    >
    >     tenure;
Query ID = root_20241022093451_b95027aa-59cb-4e18-8f4f-bb60e3e7b76b
Total jobs = 1
Launching Job 1 out of 1
```

```
OK
1       363
2       339
3       354
4       332
5       344
6       343
7       359
8       329
9       355
10      330
11      367
12      335
13      338
```

**22. Compare Churned and Retained Customers with Similar Monthly charges**

SELECT

a.tenure AS churned_customer_tenure,

b.tenure AS retained_customer_tenure, a.MonthlyCharges AS common_monthly_charges

FROM telecom_churn a

JOIN telecom_churn b

ON a.MonthlyCharges = b.MonthlyCharges

AND a.Churn = 'Yes'

AND b.Churn = 'No' LIMIT 10;

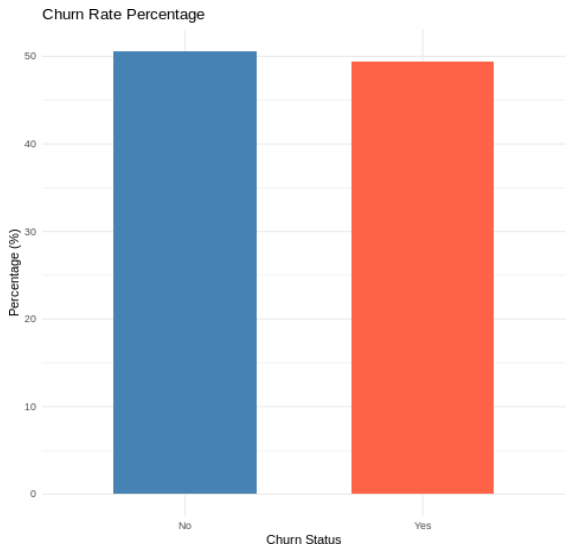
```
hive> SELECT
    >
    >     a.tenure AS churned_customer_tenure,
    >
    >     b.tenure AS retained_customer_tenure,
    >
    >     a.MonthlyCharges AS common_monthly_charges
    >
    > FROM telecom_churn a
    >
    > JOIN telecom_churn b
    >
    > ON a.MonthlyCharges = b.MonthlyCharges
    >
    >    AND a.Churn = 'Yes'
    >
    >    AND b.Churn = 'No'
    >
    > LIMIT 10;
Query ID = root_20241022052801_157514a8-0e63-4fcb-8521-cae4a09e9339
Total jobs = 1
Launching Job 1 out of 1
Tez session was closed. Reopening...
Session re-established.


Status: Running (Executing on YARN cluster with App id application_1729443979892_0092)
```

```
--------------------------------------------------------------------------------
        VERTICES      STATUS  TOTAL  COMPLETED  RUNNING  PENDING  FAILED  KILLED
--------------------------------------------------------------------------------
Map 1 ..........   SUCCEEDED      1          1        0        0       0       0
Map 2 ..........   SUCCEEDED      1          1        0        0       0       0
--------------------------------------------------------------------------------
VERTICES: 02/02  [==========================>>] 100%  ELAPSED TIME: 26.15 s
--------------------------------------------------------------------------------
OK
6       1       29.85
57      34      56.95
68      45      42.3
52      22      89.1
13      22      89.1
65      62      56.15
6       31      80.67
68      9       62.75
35      5       84.85
55      5       84.85
Time taken: 32.865 seconds, Fetched: 10 row(s)
```

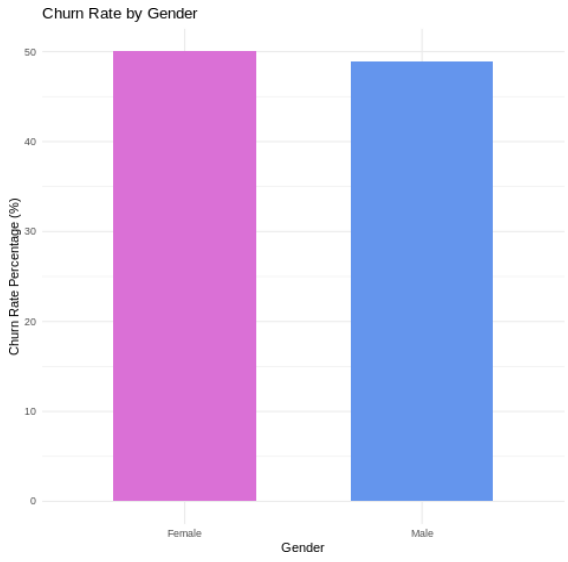
### 3.4. Data Visualization with R:

1. **Overall Churn Rate**



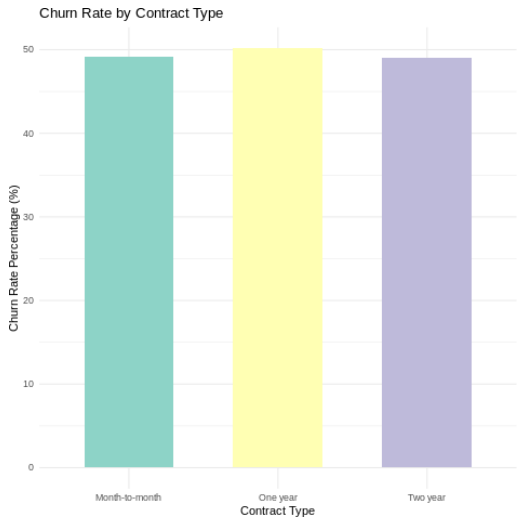

The bar chart illustrates the churn rate, showing that approximately 50% of customers have churned (left the company) and 50% have remained active.

2. **Churn Rate by Gender**



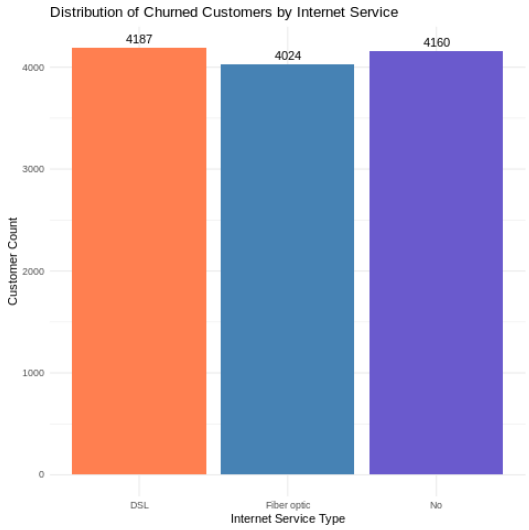

This bar chart shows that the churn rate is higher for females compared to males. Approximately 50% of female customers have churned, while the churn rate for male customers is around 48%.

### 3. Churn Rate by Contract type



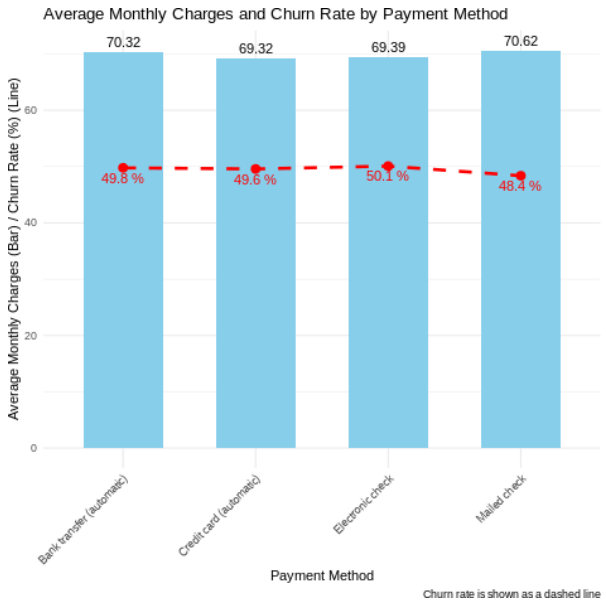

It shows that customers with month-to-month contracts have the highest churn rate, followed by those with one-year contracts. Customers with two-year contracts have the lowest churn rate.

### 4. Distribution of Churned Customers by Internet Service



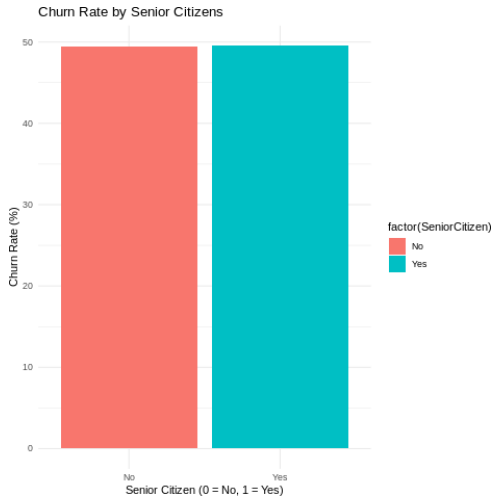

It shows that the highest number of churned customers have no internet service, followed closely by those with fiber optic service. The lowest number of churned customers have DSL internet service.

## 5. Average monthly charges and churn rate by payment method



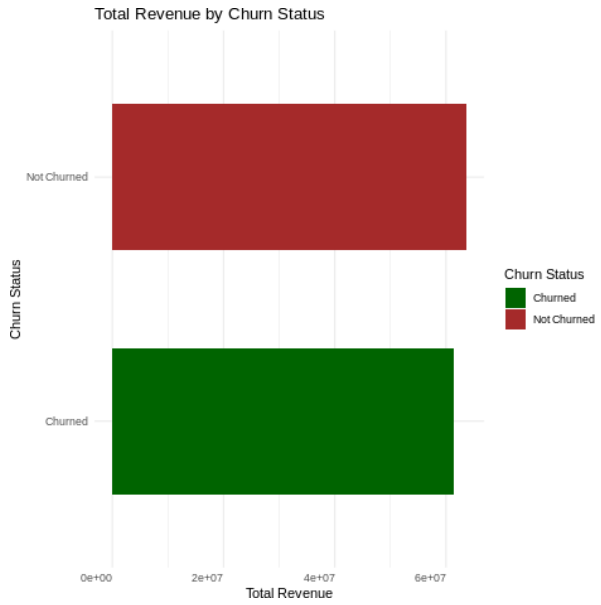

This graph shows that there is not a significant difference in average monthly charges between payment methods. However, customers who pay by mailed check have the lowest churn rate (48.4%), while those who pay by electronic check have the highest churn rate (50.1%).

## 6. Churn Rate by Senior Citizens



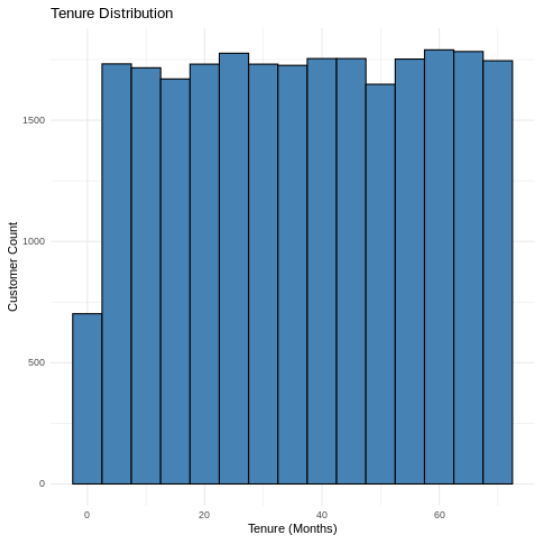

The bar chart shows that senior citizens have a higher churn rate compared to non-senior citizens. Approximately 50% of senior citizens have churned, while the churn rate for non-senior citizens is around 40%.

## 7. Total revenue for churned and non-churned customers



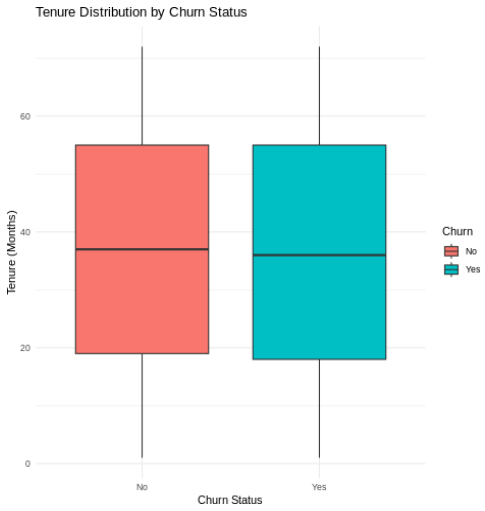

The bar chart shows that customers who have not churned generate significantly more total revenue compared to those who have churned. The revenue generated by customers who have not churned is approximately double that of those who have churned.

## 8. Tenure Distribution among Customers



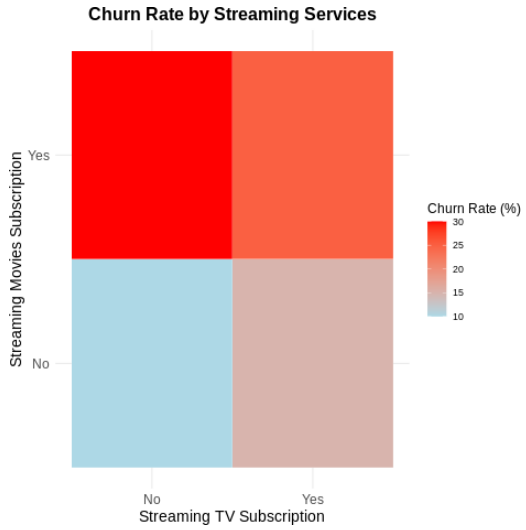

It shows the distribution of customer tenure in months. The distribution appears to be relatively uniform, with a slight increase in the middle and towards the end.This suggests that customers are generally distributed across different tenure lengths, with no significant peaks or troughs in the distribution.

## 9. Tenure Distribution by Churn Status



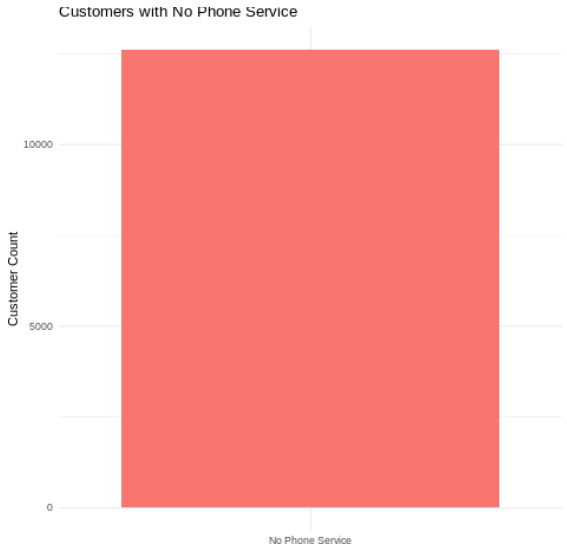

It shows that customers who churned had a lower median tenure compared to retained customers. This suggests that shorter tenure might be a factor contributing to churn, indicating a need for strategies to improve customer retention in the early stages of their relationship with the company.

## 10. Churn Rate by Streaming Services



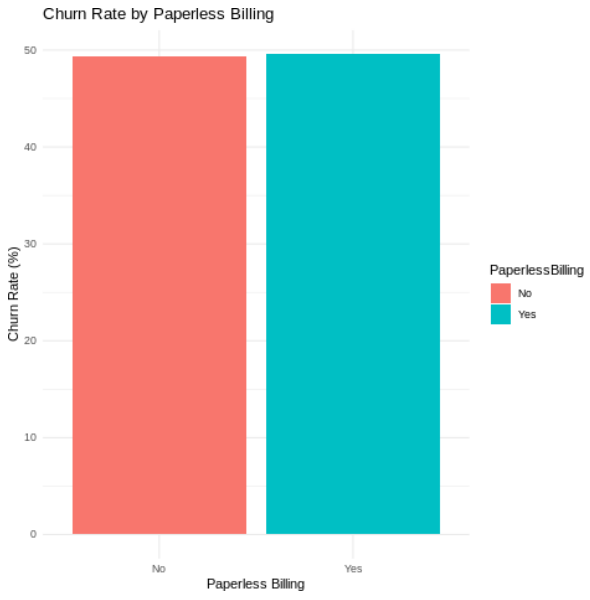

It shows the churn rate based on different combinations of streaming TV and movie subscriptions. Customers with both TV and movie subscriptions have the highest churn rate.This suggests that offering bundled packages or exclusive content might be necessary to retain customers who subscribe to both services.

## 11. Customer with no phone service



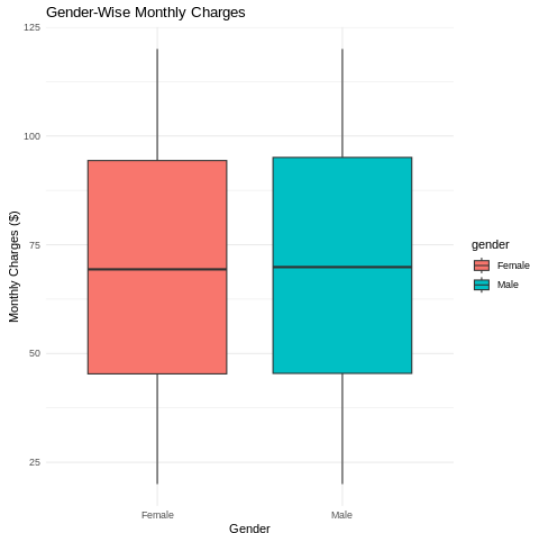

The bar chart titled shows a single bar representing a significant number of customers, likely exceeding 10,000, who do not have phone service. The absence of additional bars or data points suggests a focus on this specific customer segment and the potential implications for the company or service provider.

## 12. Churn Rate by Paperless Billing



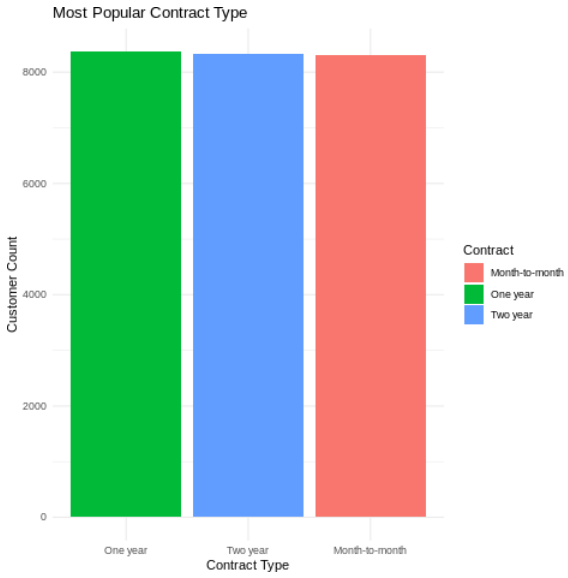

It shows a higher churn rate for customers without paperless billing compared to those with paperless billing. This suggests that offering paperless billing options might be an effective strategy to reduce customer churn.

## 13. Gender-Wise Monthly Charges



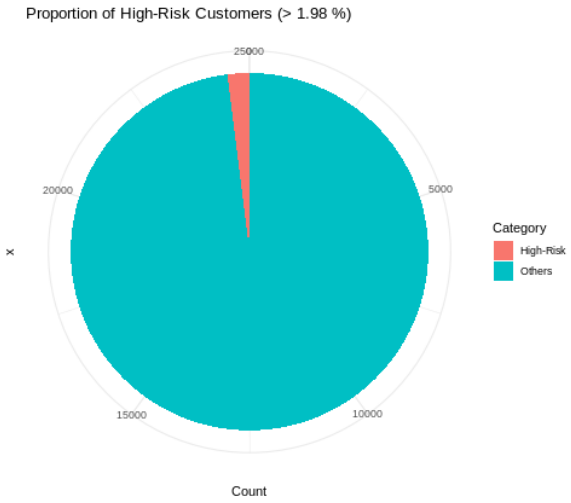

It shows that male customers tend to have slightly higher monthly charges compared to female customers. Both genders exhibit similar variability in monthly charges, with the median charge being around $75 for both males and females.

## 14. Most Popular Contract Type



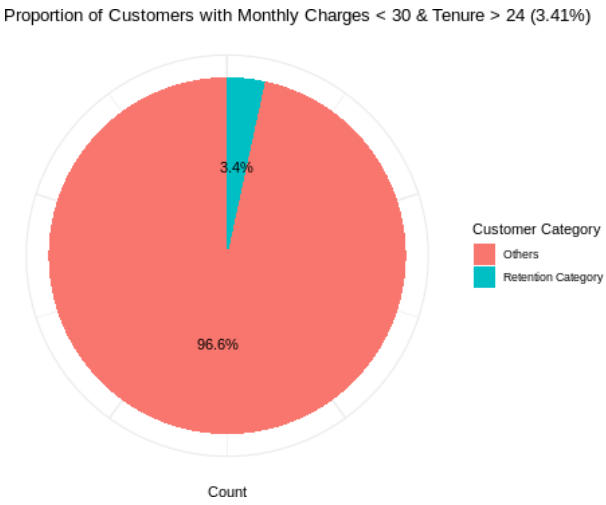

It shows the distribution of customers across different contract types. One-year contracts appear to be the most popular, followed closely by two-year contracts and month-to-month contracts. This suggests that offering longer-term contract options might be beneficial for customer retention and revenue stability.

### 15. Percentage of High-Risk Customers (High Monthly Charges, Short Tenure)



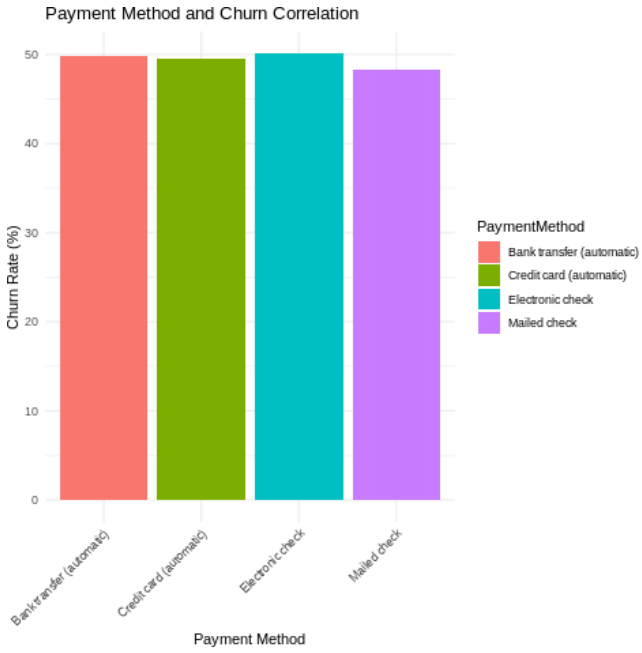

The pie chart shows that a very small percentage of customers are classified as high-risk.

### 16. Percentage of Retention Opportunities (Long Tenure, Low charges)



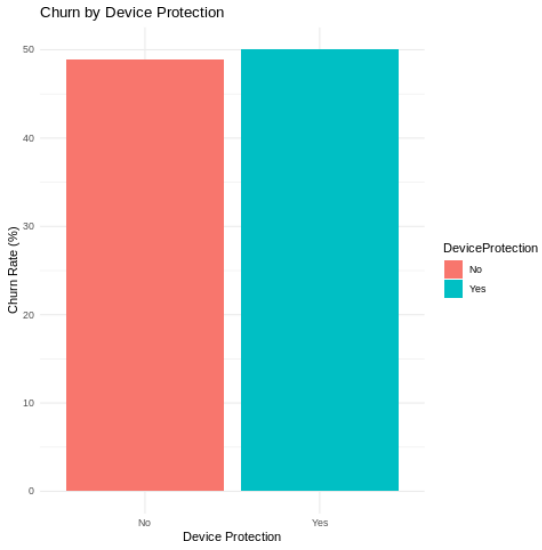

The pie chart titled "Proportion of Customers with Monthly Charges < 30 & Tenure > 24 (3.41%)" shows that a very small percentage of customers meet the criteria of having monthly charges less than $30 and a tenure greater than 24 months.This suggests that this specific customer segment represents a small portion of the overall customer base.

## 17. Payment Method and Churn Correlation



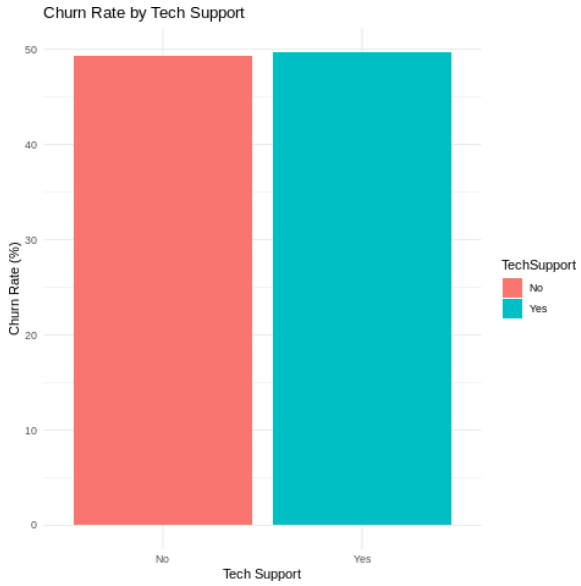

It shows the churn rate for different payment methods. Customers using mailed checks have the highest churn rate, while those using electronic checks have the lowest. This suggests that offering convenient and efficient payment methods like electronic checks might help reduce customer churn.

## 18. Churn by Device Protection



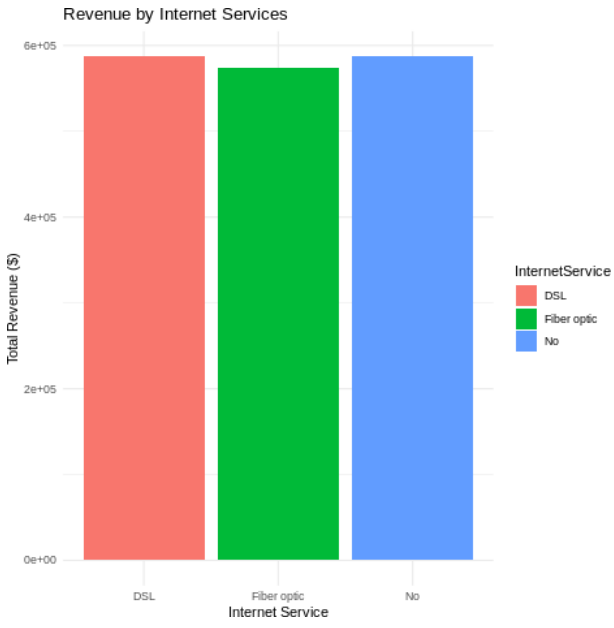

It shows a slightly higher churn rate for customers without device protection compared to those with device protection.This suggests that offering device protection plans might be a strategy to retain customers and potentially reduce churn.

### 19. Churn Rate by Tech Support



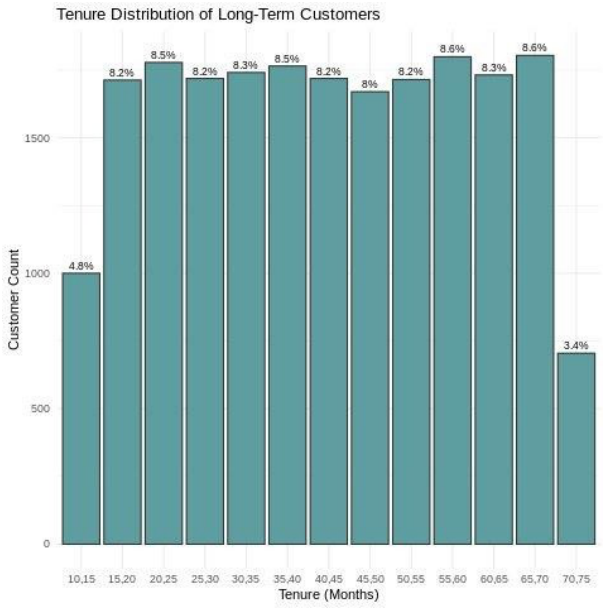

It shows a slightly higher churn rate for customers without tech support compared to those with tech support. This suggests that offering tech support might be a strategy to retain customers and potentially reduce churn.

### 20. Revenue by Internet Services



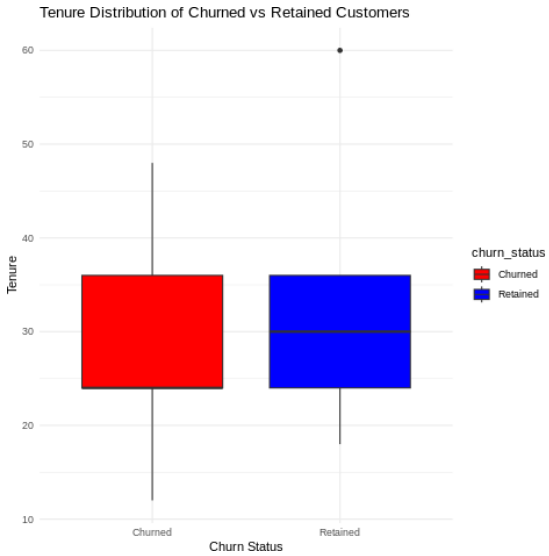

It shows that fiber optic internet service generates the highest revenue, followed by DSL and no internet service. This suggests that offering fiber optic internet service could be a strategy to increase revenue and attract more customers.

## 21. Long-Term Customers




It shows the distribution of customer tenure in months. The majority of long-term customers have a tenure between 65 and 70 months. This suggests that a significant portion of customers remain with the company for a considerable period, potentially indicating high customer satisfaction and loyalty.

## 22. Tenure Distribution of Churned and Retained Customers




It shows that customers who churned had a lower median tenure compared to retained customers. This suggests that shorter tenure might be a factor contributing to churn, indicating a need for strategies to improve customer retention in the early stages of their relationship with the company.

## Concluding Remarks

The analysis of telecom customer churn data using Hive and Hadoop provided valuable insights into customer retention challenges and opportunities. Key findings include:

Approximately 50% of customers have churned.

1. **Churn Rate by Gender**
   - Gender-based analysis reveals that female customers churn slightly more than male customers.

2. **Churn Rate by Contract Type**
   - Month-to-month contracts show the **highest churn rate** due to less customer commitment.
   - Two-year contracts exhibit the **lowest churn rate**, possibly due to better incentives.

3. **Internet Service Impact**
   - **Fiber Optic Internet:** High churn rate, but highest revenue contributor.
   - **DSL Service:** Low churn rate, indicating satisfied customers.

4. **Payment Methods**
   - **Electronic Checks:** High churn rate (50.1%) due to ease of cancellation.
   - **Mailed Checks:** Low churn rate, indicating potential customer loyalty.

5. **Senior Citizen Analysis**
   - Senior citizens churn more frequently, possibly due to cost sensitivity or technology challenges.

6. **Revenue from Churned vs. Non-Churned Customers**
   - Non-churned customers generate nearly double the revenue of churned customers, underscoring the financial impact of churn.

7. **Tenure Distribution**
   - Retained customers have significantly higher median tenures than churned ones.
   - Short tenures correlate strongly with churn.

8. **Streaming Services**
   - Customers with both TV and movie streaming services exhibit higher churn rates.
   - Bundling exclusive content could reduce churn.

9. **Retention Opportunities**
   - **High-Risk Customers:** Monthly charges > $90, tenure < 6 months.
   - **Retention Segment:** Monthly charges < $30, tenure > 24 months (only 3.41% of customers).

## 10. Data Visualization Insights

Visuals provide clarity on key trends:

- **Churn by Gender:** A slight difference in churn rates between males and females.
- **Churn by Payment Method:** Mailed checks have the least churn.
- **Revenue vs. Churn:** Non-churned customers generate significantly higher revenue.
- **Tenure and Churn Relationship:** Short-tenured customers have a higher propensity to churn.

This study demonstrated how distributed data processing and structured querying can effectively analyze large datasets, offering a roadmap for data-driven decision-making in reducing churn and improving telecom services.