Analyzing correlation between customer churn rate and charges in telecommunication setting

Overview:

As a business analyst at a telecommunications company, my primary objective is to identify customers likely to churn (leave the company). Using data sourced from Kaggle's Customer Churn Prediction 2020 dataset, I analyzed customer behavior patterns, service interactions, and billing data to uncover key churn indicators. This analysis aims to help the company proactively retain customers and focus on areas needing service improvements.

Step-by-Step Analysis Process:

1. **Data Preparation (Excel + Power Query):**

   o Calculated total service usage metrics:

      ▪ total_minutes = sum of total_day_minutes + total_eve_minutes + total_night_minutes + total_intl_minutes

      ▪ total_calls = sum of total_day_calls + total_eve_calls + total_night_calls + total_intl_calls

      ▪ total_charge = sum of total_day_charge + total_eve_charge + total_night_charge + total_intl_charge

   o Derived average customer service calls per state.

2. **PivotTables & Calculations:**

   o Summarized total charges by state and area code.

   o Identified patterns in customer service interactions.

   o Created a simple rule-based churn prediction model (customers with >3 service calls flagged as potential churners).


Row Labels	Average of number_customer_service_calls
AL	0.913043478
KS	1
CO	1.0625
OH	1.142857143
MN	1.235294118
WV	1.263157895
MT	1.263157895
ME	1.285714286
OK	1.333333333

|	
SD	1.4
NV	1.428571429
LA	1.461538462
NE	1.466666667
SC	1.473684211
DE	1.5
RI	1.5
OR	1.533333333
IN	1.533333333
MO	1.538461538
DC	1.5625
AR	1.571428571
ND	1.571428571
AZ	1.583333333
VA	1.611111111
TX	1.611111111
NM	1.615384615
MI	1.625
NH	1.647058824
IL	1.666666667
HI	1.666666667
WI	1.666666667
PA	1.7
MS	1.705882353
MA	1.714285714
CT	1.727272727
FL	1.785714286
WA	1.833333333
UT	1.866666667
NC	1.909090909
KY	1.928571429
MD	1.9375
AK	2
TN	2
GA	2
IA	2
WY	2.05
VT	2.066666667
NY	2.111111111
ID	2.230769231
CA	2.230769231
NJ	2.625
Grand Total	1.634666667

Average number of customer service calls
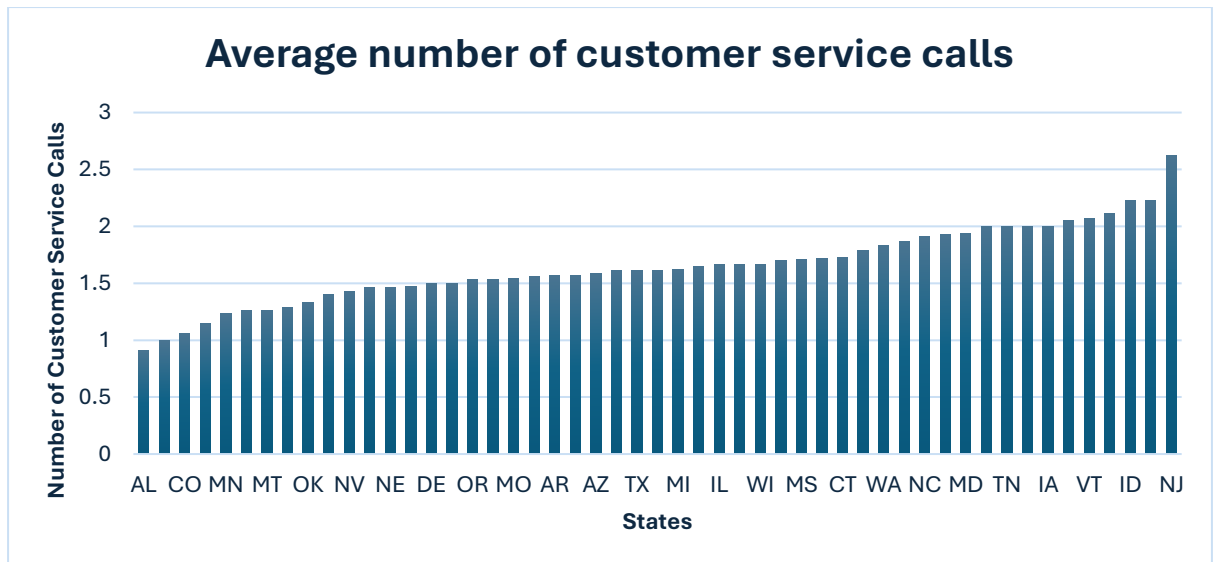
It is evident that Alabama (AL) has the lowest average number of customer service calls at 0.91, while New Jersey (NJ) records the highest average at 2.63. This suggests that customers in NJ face more issues, requiring frequent contact with customer service to resolve their problems. Among the 51 states, NJ appears to have the lowest customer satisfaction based on the higher volume of service interactions. Furthermore, I calculated the top 10 states with the highest total charges. From the results, it is clear that Alabama (AL) has the highest total charges, amounting to 1382.42, while West Virginia (WV) has the lowest total charges among the top 10, with 1105.26.

Row Labels	Sum of total_charges
WV	1105.26
WA	1118.97
WY	1121.24
VA	1122.41
SC	1135.05
MT	1180.64
OH	1236.32
ND	1244.27
AR	1339.75
AL	1382.42
Grand Total	11986.33

Sum of total charges
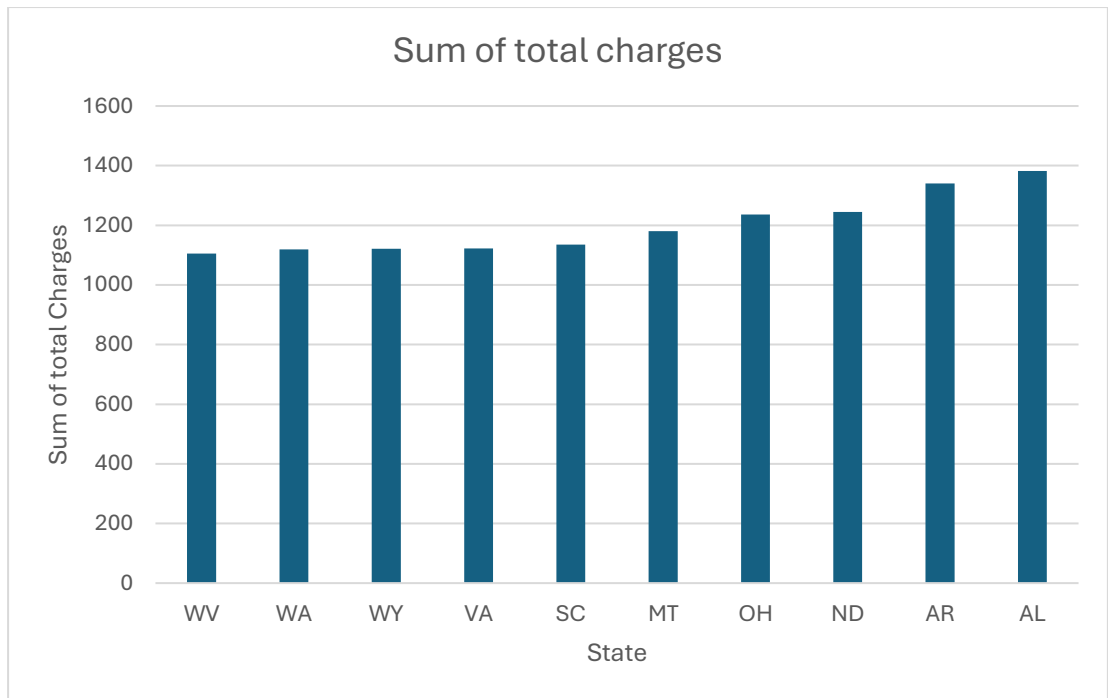
Total charge based on area code and the states:

Sum of total_charges	Column Labels			
Row Labels	area_code_408	area_code_415	area_code_510	Grand Total
AL	369.04	840.13	173.25	1382.42
AR	260.67	596.39	482.69	1339.75
DC	372.68	303	299.45	975.13
GA	277.92	561.11	261.25	1100.28
MD	199.77	693.62	111.85	1005.24
MI	220.76	599.74	214.22	1034.72
MN	116.44	627.13	285.12	1028.69
MS	133.6	440.04	470.82	1044.46
MT	237.03	764.45	179.16	1180.64
ND	290.27	727.5	226.5	1244.27
NH	357.45	480.08	187.29	1024.82
NJ	362.67	584.69	50.74	998.1
NY	250.64	699.78	125.6	1076.02
OH	106	911.41	218.91	1236.32
SC	121.46	669.61	343.98	1135.05
TX	181.4	573.54	328.42	1083.36
VA	382.5	492	247.91	1122.41
WA	122.4	672.58	323.99	1118.97
WV	123.07	639.08	343.11	1105.26
WY	264.43	524.36	332.45	1121.24
Grand Total	4750.2	12400.24	5206.71	22357.15
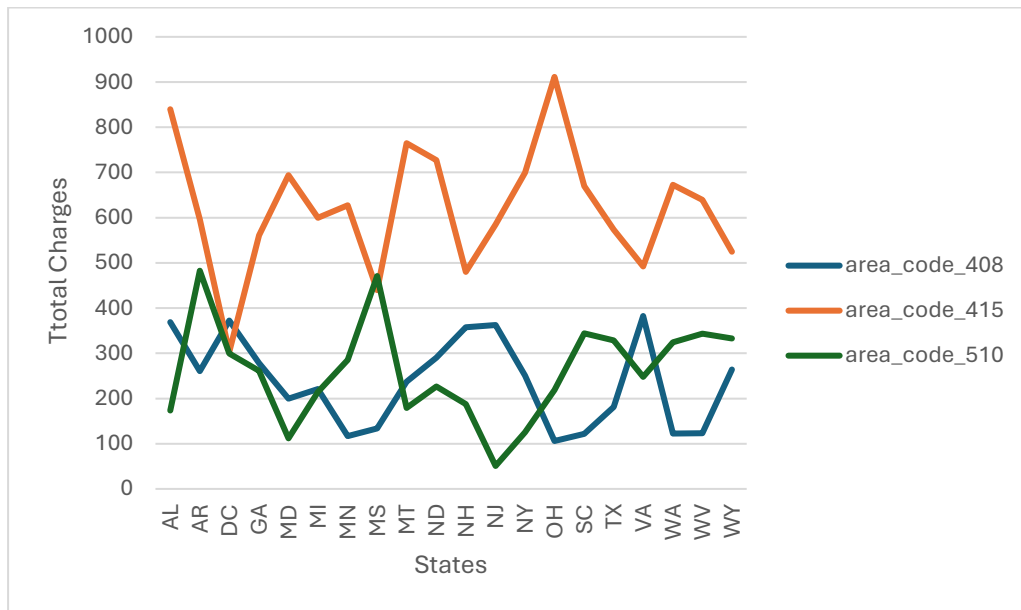
The analysis of total charges across different area codes and states reveals significant variation in customer billing patterns. By breaking down total charges into area codes (408, 415, 510) for each state, we can observe which regions contribute most to overall revenue. For example, Alabama (AL) records the highest total charges of 1382.42, with a large portion coming from area code 415. States like North Dakota (ND) and Montana (MT) also show high total charges, particularly concentrated in area codes 415 and 510, suggesting that customers in these regions are either using more telecom services or being charged at higher rates. This detailed segmentation helps understand geographic trends in service usage and can guide decisions on resource allocation, marketing efforts, and service improvements targeted to specific regions. To further assess customer satisfaction and predict potential churn, a new column called Predicted Churn was created based on the logic that customers making more than 3 customer service calls are more likely to churn.

$$\text{Accuracy} = \frac{\text{Number of correct predictions}}{\text{Number of total test samples}}$$

This simple rule-based model was compared against actual churn data to measure its effectiveness. The comparison showed that this method predicted churn correctly 45.6% of the time, which is a moderate accuracy and indicates that customer service interactions alone may not be sufficient to predict churn precisely. The analysis also highlighted that New Jersey (NJ)

has the highest average number of customer service calls, suggesting that customers in NJ face more issues and are thus more prone to dissatisfaction and churn. With a total of 287 churn cases, it's clear that churn is a significant concern, and identifying such patterns helps in developing better customer retention strategies.

Recommendations:

1. **Focus States for Retention Efforts:**

    o **High Churn Risk States:** New Jersey (NJ), California (CA), Idaho (ID)

    o **Action:** Improve service support, monitor customer complaints, offer retention deals.

2. **Leverage High-Value Low-Risk Regions:**

    o **Low Churn Risk States:** Alabama (AL), Ohio (OH), North Dakota (ND)

    o **Action:** Continue existing strategies, monitor for emerging issues, cross-sell additional services.

3. **Service Interaction Monitoring:**

    o Develop real-time dashboards to track spikes in customer service calls.

    o Conduct regular satisfaction surveys in states with high support activity.

4. **Model Improvement:**

    o Use advanced models (e.g., logistic regression, decision trees) combining more features for accurate churn prediction.

For further information:

Tableau link:
https://public.tableau.com/views/ChurnRateforTelecome/Dashboard1?:language=en-GB&:sid=&:redirect=auth&:display_count=n&:origin=viz_share_link

Dataset: https://www.kaggle.com/c/customer-churn-prediction-2020/overview