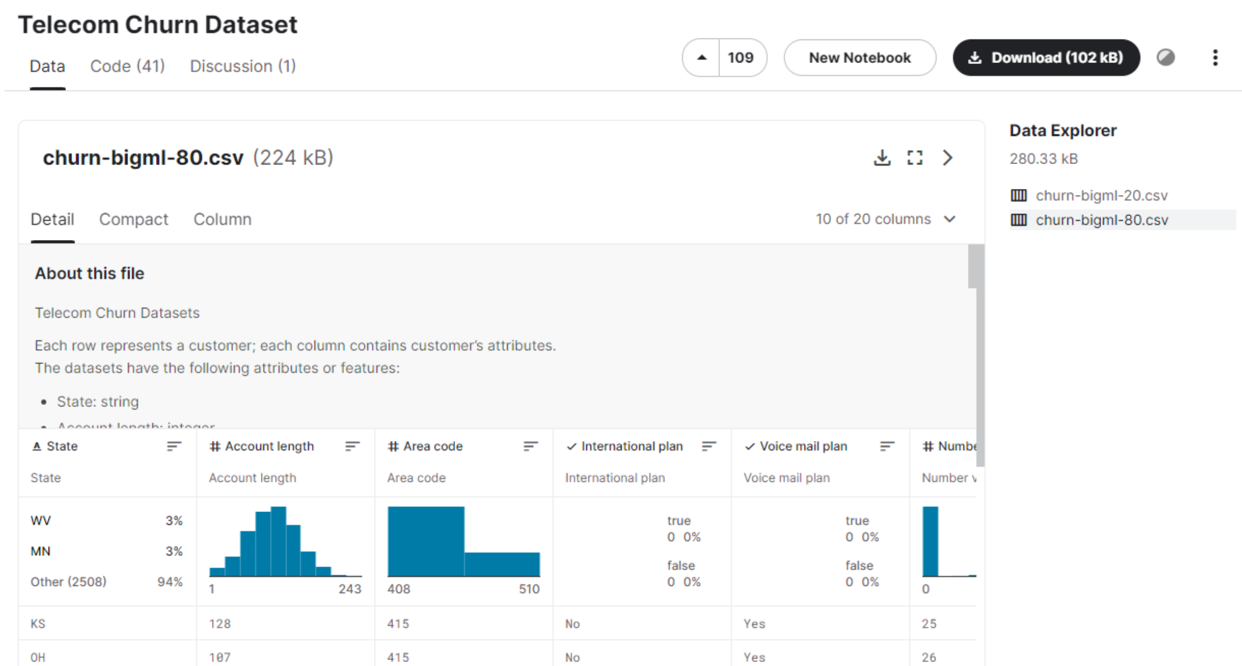
Problem Statement

How can The Orange Telecom decrease customer churn by X% before the end of the next fiscal year by focussing their efforts on the factors that are most closely related to custom churn.

Approach

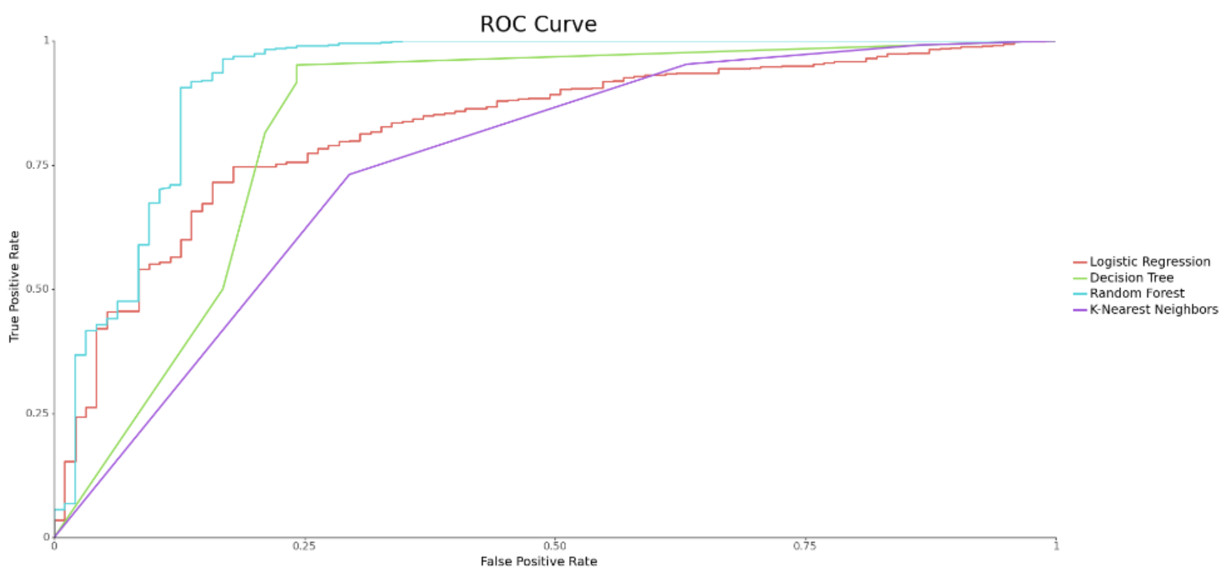
The Telecom Churn Dataset from Kaggle, provided data on the user base of The Orange Telecom Company. This dataset had information on number of calls, minutes, and changes split by day, evening, night, and international. It also had data on voicemail, location, and customer service calls as well. Then of course, there was the output variable churn.

The dataset was as clean as could be. It required very few adjustments to be usable. There were no missing values, outliers, incomplete rows, etc. The data came presplit 80/20 between training and test sets as well. For the most part, the only adjustments I made were in the pre-processing phase creating dummy variables and standardization.



Model Results

model	top_features	train_cv_mean	train_cv_std	accuracy	precision	recall	f1
Random Forest	[total_day_minutes, customer_service_calls, to...	0.952	0.0131	0.9535	0.7263	0.9324	0.8166
Decision Tree	[total_day_minutes, total_eve_minutes, interna...	0.9355	0.0084	0.9235	0.7579	0.72	0.7385
K-Nearest Neighbors	[NA1, NA2, NA3]	0.8811	0.0157	0.8696	0.3684	0.5645	0.4459
Logistic Regression	[voice_mail_plan, total_day_minutes, customer_...	0.8604	0.0243	0.8531	0.1789	0.4595	0.2576

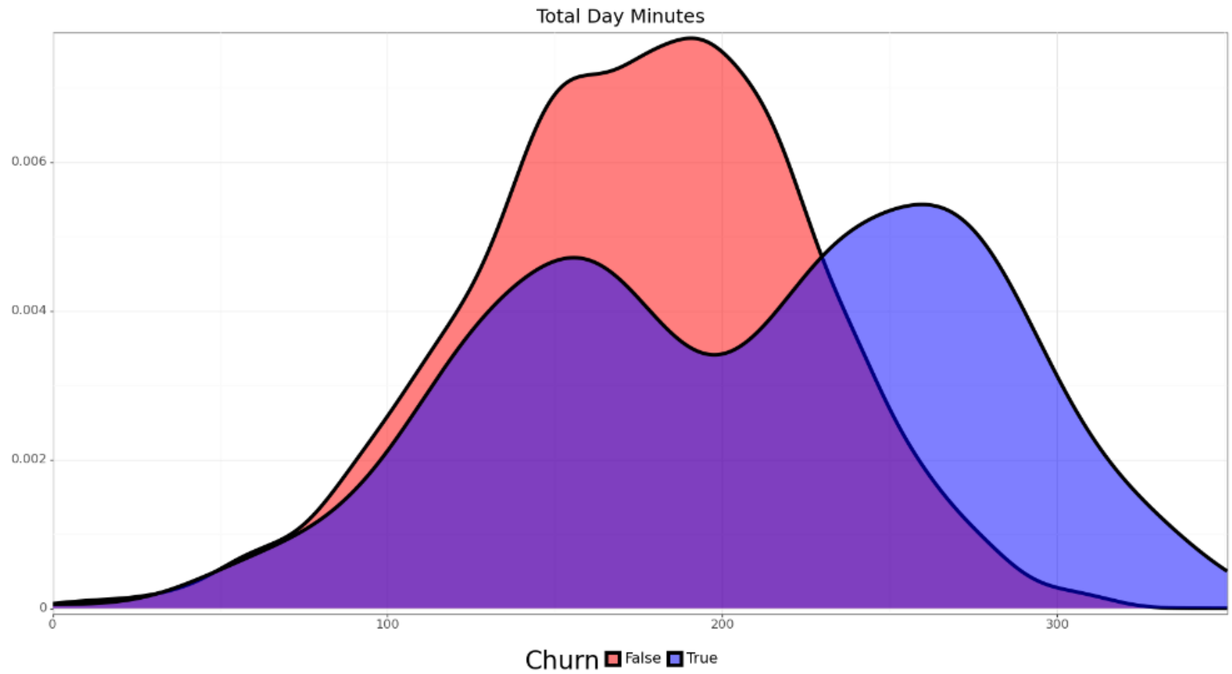


Recommendations

Recommendation 1: Lower Day Time Pricing

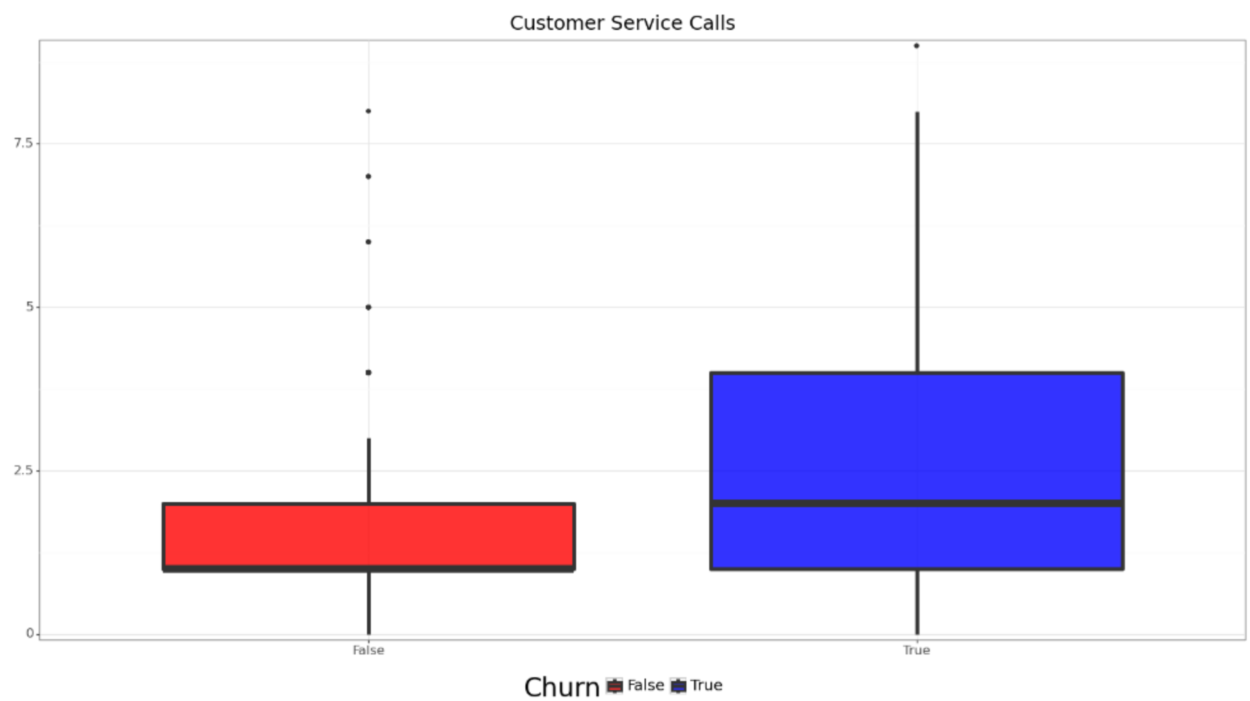
Daytime pricing was the biggest determining factor in predicting churn. The distribution of calls was even across daytime, evening, and nighttime, but daytime calls cost twice as much as evening calls and four times as much as nighttime calls. This pricing method makes sense for boosting revenue, but is hurting the churn rate. A new pricing strategy should be explored to help reduce the churn rate, while keeping revenue at a reasonable level. It’s better to get some money from a customer than have that customer leave entirely

- Day Price: $0.170/min
- Eve Price: $0.085/min
- Night Price: $0.045/min
- International Price: $0.270/min



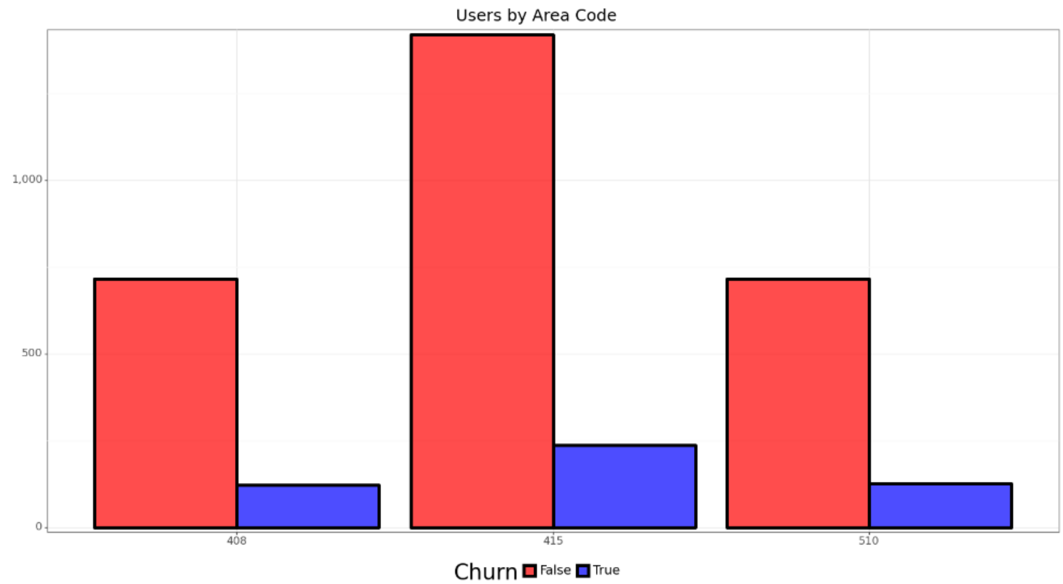
Recommendation 2: Flag repeat Customer Service callers

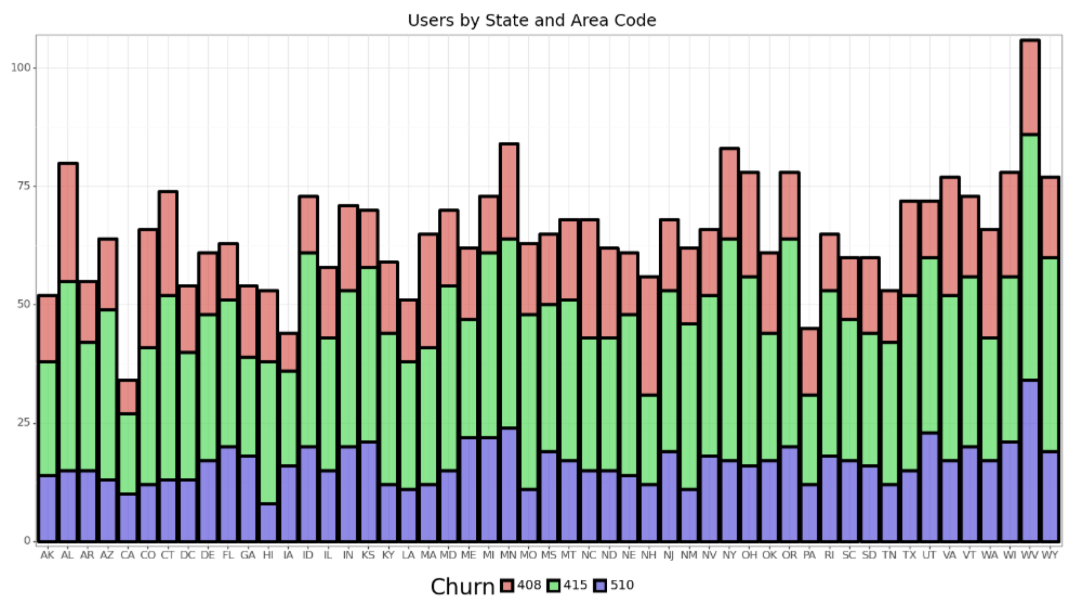
Repeat Customer Service callers were more likely to churn. Customers only use Customer Service when something is wrong so this is a great indicator for someone who may churn. These customers should be flagged and offered some sort of temporary discount to keep them satisfied while they're at their angriest with the service. I suggest reducing the cost of daytime calls as that is the biggest determining factor in predicting churn



<u>Recommendation 3: Location & Voicemail are Irrelevant</u>

Location and area code are irrelevant in determining churn. Customers of all three provided area codes had the same churn rate. Those area codes were also evenly distributed across the various states the user base is located in. This isn't an issue with lack of service in certain parts of the county.





Further Research

I may have the Data Science abilities to gather data, clean it, and build an accurate predictive model, but it doesn’t mean much without context. I lack knowledge of the Telecom industry so there’s only so much insight I can provide. There’s only so much I can understand about the data. I can explain what the results are but industry knowledge is required to explain what the results really mean.

I also lack company knowledge. I don’t know the processes, strategies, or reasoning behind their specific data. I don’t know why the daytime price is so much more than evening and nighttime. Maybe that high price is worth the churned customers. I don’t know.

Though this is a Data Science project, qualitative data is also important here as we are dealing with customers. Surveys would be very helpful. Interviews with various customers would be helpful. Service reviews would also be helpful. That data will only show so much. Speaking with the people who actually use the service would add a perspective data can’t provide.