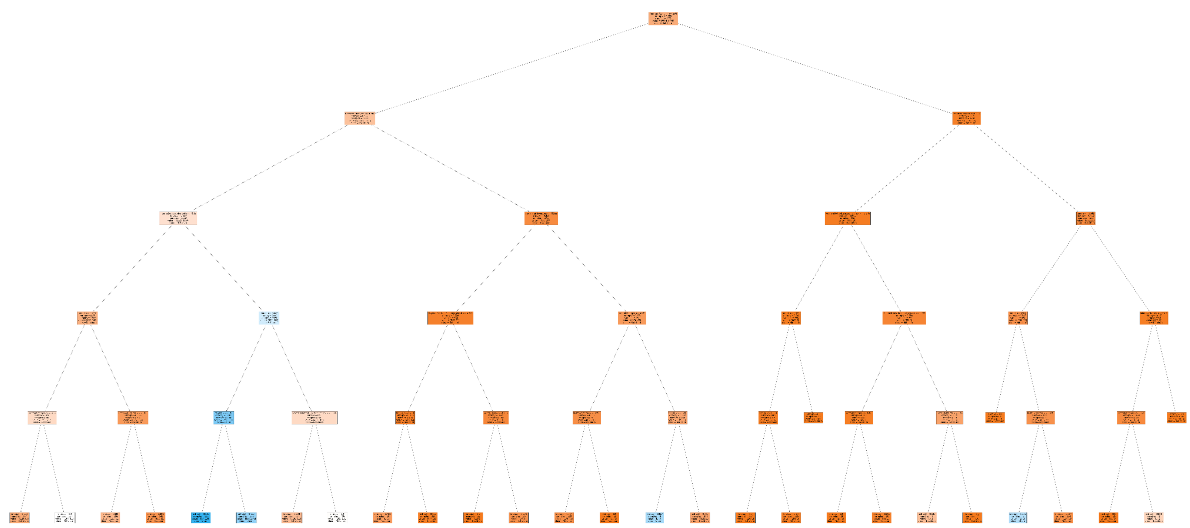
Question 1: what is the most effective decision tree you can build? To answer this question, you want to systematically experiment with different levels of model complexity. The easiest way is to change the “maximal tree depth” levels and leave other parameters unchanged. **Please report your findings and explain the rationale behind.** Once you settle with the most effective decision tree, **please characterize the high- and low-probability churning customers with the decision tree results. Is the output of data- driven analysis same or different from your intuitions?**

To find the most effective decision tree, we experiment using the try them all methodology with different levels of model complexity by varying the "maximal tree depth" parameter from 1 to 27 in the decision tree. We test it out by using the GridSearchCV to search over different values of max-depth hyperparameter. We applied the AUC (Area under the ROC curve) as the scoring metric to evaluate the performance of different max-depth.The result shows that best max_depth is 5. Then we instantiate a new decision tree model with the best performing depth (max_depth=5) and fit it to the training data.

From the decision tree, as the parent node is contract_two_year, customer with any contract terms longer than 2 years tends to not churn. And customer with a month to month contract term, with lower tenure, tends to churn.



Our institutions was that monthly charges and contract year tends to be the most significant indicator of customer churning behaviors. With higher the monthly charges, customer tends to churn as it incentivize customer to switch to a competitor with lower pricing. Customer tends to not churn as they signed a longer contract/ already stayed alone with the company for a longer period of time as customers who commit to longer contract terms may have a higher level of loyalty or satisfaction with the service. Our insight share similarity on the contract terms and tenure part, but monthly charges does not seem to be a significant factor here.

Question 2: Is decision tree or logistic regression better? Please first build a Logistic regression model with no regularization and compare with the best decision tree model you built from the previous question. **Please carefully examine various performance metrics and comment about the differences in performance between the two methods**. Next, use Lasso regularization. Please try to adjust the Strength parameter from Weak to Strong. As expected, the stronger the regularization strength is, the less variables are kept in the model. **Please compare the "good" features identified by decision tree and logistic regression and try to provide explanations about why they are different.**

Accuracy of Decision Tree: 0.7834123222748816
Confusion Matrix of Decision Tree
Precision = 1448/ (1448 + 356) = 0.802660754
Recall = 1448/ (1448 + 101) = 0.934796643

Accuracy of Logistic Regression Model without regularization: 0.7981042654028436
Confusion Matrix of Logistic Regression Model without regularization
Precision = 1382/ (1382 + 259) = 0.842169409
Recall = 1382/ (1382 + 167) = 0.892188509

The logistic regression model without regularization has an accuracy of 79.8%, higher than the accuracy score of the decision tree, which is 78.3%. The precision of decision tree is 80.2%, lower than the 84% of the logistic regression model without regularization. The recall of decision tree is 93.5%, higher than the 89.2% of logistic regression model. To sum up, the logistic regression model has higher overall measure of correctness and lower tendency to incorrectly classify negative instance as positive. And the decision tree model has lower tendency to incorrectly classify positive instances as negative.

To get the balance between precision and recall, we can also use f1 score to determine which model is better. The f1 score for the decision tree model is 0.86 and for the logistic regression model is 0.87. The higher the f1 score, the better the model performs. Therefore, based on f1 score, the logistic regression model is better.

We increase the strength by gradually decreasing C value from 1000000.0 to 0.001. Based on the AUC as the scoring, Best C for logsitic regression is when strength 0.1 (AUC = 0.8459220210368799)

Importance of Feature from Decision tree
SeniorCitizen 0.3500323808747644
tenure 0.2388885028785912
MonthlyCharges 0.15127495478743064

gender_Male 0.12088068028143875
Partner_Yes 0.06024989640561687
Dependents_Yes 0.03670618187228954
PhoneService_Yes 0.022655726111918094
MultipleLines_No phone service 0.014284229689695633
MultipleLines_Yes 0.005027447098254751

Coefficient of logistic with best AUC:
gender: 0.08144356270076884
SeniorCitizen: -0.8082656357418979
Partner: 0.0
Dependents: 0.0
PhoneService: 0.03311353736102287
MultipleLines: -0.09724779995348551
InternetService: -0.005122309757958441
OnlineSecurity: 0.09601137722902493
OnlineBackup: 0.09903860615131019
DeviceProtection: 0.47320974810330857
TechSupport: -7.326384475301997e-05
StreamingTV: 0.0
StreamingMovies: -0.15383371931956658
Contract: -0.1798305990551073
PaperlessBilling: -0.005751042858643435
PaymentMethod: -9.818199030374782e-07
MonthlyCharges: 0.0

Decision tree suggestion important features are contact year, monthly charges, payment method, tenure, dependents, multiplelines, phone services and gender. While logistic model suggests SeniorCitizen, DeviceProtection, TechSupport, StreamingMovies, Contract, and payment method tends to be good features.

Question 3: Reevaluate different models with the cost information and customer's contract value

One simple decision-making rule is if the model predicts a customer will churn, a retention offer will be sent to him/her. Now we take into account the differential economic benefits of sending retention offers to different customers. We assume that churning customers will actually renew a one-year contract with the same monthly payment as the previous contract, if they receive the retention offer. So, we define contract value (CV) of individual customer as: Monthly payment *12.

Also, let us make the following further assumptions. Offer itself costs $200 and sending offer costs $5. If a customer is predicted not churn but he/she actually churn, the company loose CV-205, which could have been earned if the prediction is correct. Also, if a customer is predicted churn but he/she actually doesn't, it costs only $5. The cost matrix is shown in the above Figure.

Please use the prediction output from decision tree and logistic regression model you built from Task 1 (with the default decision threshold) and compare which model gives you better overall performance with the cost information.

Next, if you are allowed to change the decision threshold, what would be best decision threshold and minimal total cost generated from either model?

In order to evaluate the true value of your data-driven solution in the previous question, we need to compare the returns with the two baselines strategies including doing nothing and sending offer to everyone, which is equivalent to predicting everyone as not churning, or predicting everyone as churning. Please make some conclusions about whether data-driven solutions are better?

With the given cost information, the logistic regression model without regularization is better than the decision tree model and logistic regression model with regularization in terms of average musclassification cost. By first calculating the mean customer value and creating the cost metrix, we have determined that the average misclassification cost of the decision tree model is 29.05, while the cost of logistic regression model with and without regularization is 21.6 and 21.2 respectively. Since the logistic regression model without regularization exhibits the lowest cost, it can be concluded that this model is preferable in terms of minimizing misclassification expenses.

For the decision tree model, the best decision threshold is 0.1 and the minimal total cost is $16081.6.
For the logistic regression model, the best decision threshold is also 0.1 and the minimal total cost is $21292.4.
The lower decision threshold means that more samples will be classified as churn. This may increase the true positive, in which the cost is 0, thus, the cost will be lower.

For cost calculator, we are using the logistic regression model without any regularization as it induce the lowest average cost. While both option would keep the customer who will not churn, so the profit (CV) here is skipped for calculation and we would analyze the cost only. Doing nothing would induce 0 marginal cost. Sending offer to everyone would instead induce 5 dollar per customers while keeping people who would churn and create marginal value of CV - 205. Thus the actual value would be ((777.5784982935158 - 200) * 1382 - 2110 * 5) / 2110 = 787663.485 in total of 373.30023 per customer.

Thus, giving offer to everyone would be drive high value as it transform churning customer into not churning while only induce additional 5 dollar promotion cost for each customer.

Question 4: Making retention offer decisions based on calculating the expected return.

Now, we further extend our analysis by using the expected value framework (Check Expected value - Wikipedia, if necessary). That is, we only send to customers whose expected return is positive, which means that their contract value (CV) multiplied by their probabilities of churning is worth more than retention cost ($200) and offer transaction cost ($5). That is, the retention offer decisions are conditional on:

CV * Prob(churning) – 205 > 0

For the probabilities of churning for each customer, you are going to use those generated by the best decision tree and logistic regression models from the first task.

Again, you can evaluate your retention offer decisions with the same cost matrix. Please report the evaluation results.

As the total cost for logistic regression model is 65183.6 which is lower than decision tree model's 72692.2. So the retention offer decisions made by according to logistic regression model will bring lower cost and is better in targeting customers who are more likely to churn and have a higher potential value to the business.

```
Decision Tree Model:
Total Cost: 72692.1552901024
Accuracy: 0.7270142180094786
Logistic Regression Model:
Total Cost: 65183.63481228669
Accuracy: 0.7393364928909952
```

Question 5: The assumption that the churning customers will not churn if they receive the retention offer might not be realistic. Please make a short proposal about how to further improve the solution.

In real life, customers may churn even if they receive a retention offer. To target a higher number of customers to stay, we suggest the following approaches:

1. Personalized Retention Offers:
   Using TelCo's data, we apply a clustering algorithm to segment customers based on their characteristics and behaviors. Let's consider two customer groups: Group A consists of price-sensitive customers, while Group B comprises customers who value additional features and services.

   For Group A, we could offer a discounted subscription price or a promotional package that addresses their price sensitivity. This personalized offer acknowledges their specific concerns and provides an incentive to stay with TelCo.

   For Group B, we could focus on offering additional features or services that align with their preferences. This could include exclusive access to premium content, priority customer support, or discounted upgrades to the latest phone models. By tailoring the offer to their specific needs, we increase the likelihood of retaining these customers.

2. Offer Optimization and Decision Threshold:
   After segmenting the customers, we analyze the historical data to determine the expected return for each customer group and set a decision threshold for offering retention incentives.

   For Group A, we calculate the expected return for each customer based on their churn probability and the cost of the retention offer. We set a decision threshold, such as targeting customers with a churn probability above 50% and an expected return higher than the cost of the offer. This ensures that the retention offers are directed towards customers who are more likely to churn and where the expected return justifies the cost.

   For Group B, we may find that the expected return for retention offers is generally higher due to their higher value perception and lower churn probability. In this case, we can set a lower decision threshold, such as targeting customers with a churn probability above 30% and an expected return that exceeds the cost of the offer. This allows us to capture a larger number of customers within the limited budget.