This is my README file for Milestone Project 3

This project is and was the final project for my certification program in Data Science and Analytics. In this project I was asked to propose and answer a real-world problem using machine learning.

Utilizing a data set from Kaggle ([here](https://colab.research.google.com/drive/1uVAyNFx6UQjV9AHvBOmlLM16b9x0sMNz#scrollTo=F6aD8mXffCBd&uniqifier=5:~:text=available%20on%20Kaggle-,here.,-I%20chose%20this)) based on telecommunication service sales and Churn rates I sought to answer the question: What demographic of customer Churns before 12 months? How Many after 12?

Following my review of the dataset my research question was refined to fit the data more precisely to: **Can I group consumers by tenure, and find out how many churn within 1 year as opposesed to post 1 year?**

The Topic of customer churn is at the top of mind when leading a successful telecom business. Knowing and understanding what products sell the most, our customer demographics, and most importantly how many of these customers stay the longest and which don't is imperative to company growth. The goal is to identify when a customer is more likely to cancel their contracts so we can better structure our packages and remain competitive. I test the data against current churn rates and contract types to identify which produces a higher % of churn and why.

My stakeholders are managers/decisions makers who strategize around the results of sales. I will communicate my results by comparing current customer data to the rest of the testing model that confirms a successful decrease in customer churn % based on contract type. The most important information to convey is if it makes sense to focus on a specific contract type to increase customer retention.

Following review of the data set it was found that:

* The number of my observations in my data is 7043
* The number of my features in my data is 23
* The number of missing values in my data is 0

Various categorical values were used so a classification model was used.

Logistic regression was utilized in conjunction with cross validation, but after completing this analysis I believe other models could have been used. Though I lack the specific knowledge on which others I could have tested with.

My goal was to use a classification model to predict whether or not contract types affect churn rates. I will then use exploratory data analysis or predictive modeling (maybe both) to answer my research question. I chose classification because the model fits my categorized dataset with enough data to split and test to analyze and test for potential churn rates. I will use logistic regressing to evaluate the model performance.

Based on the results (accuracy and Cross validation) from my logistic regression model there is lacking evidence necessary to suggest a project plan.

The dataset lacks detailed or specific features necessary for improving the accuracy of this model. I think with More time and further research I could have maybe created the features I need to create a better model.

74% isn't terrible, but to suggest a project for improvement to management based on this model's results would not be wise. I would love to continue research in the future.