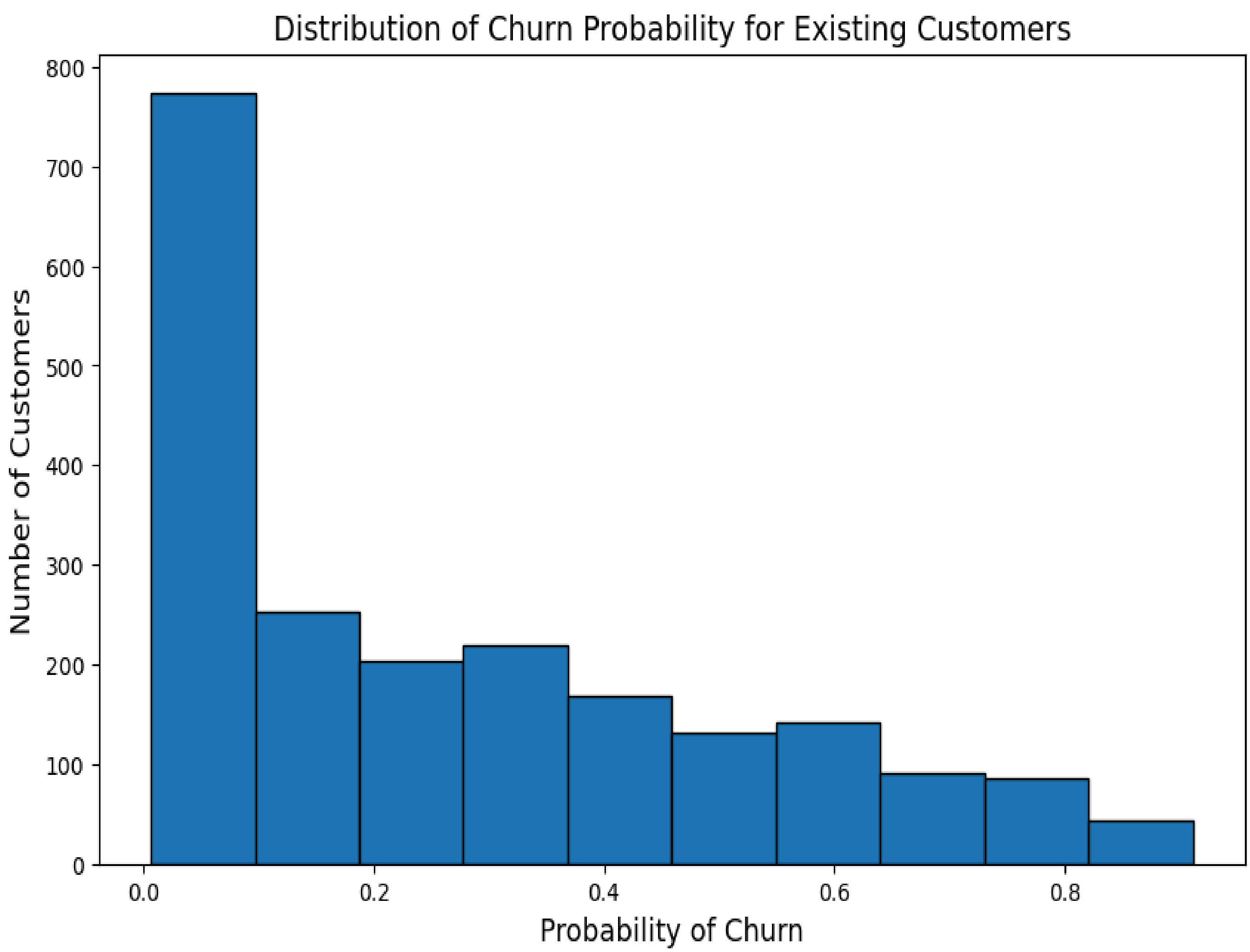
Project Report: Comprehensive Analysis of Customer Churn

1. Introduction

Objective and Scope

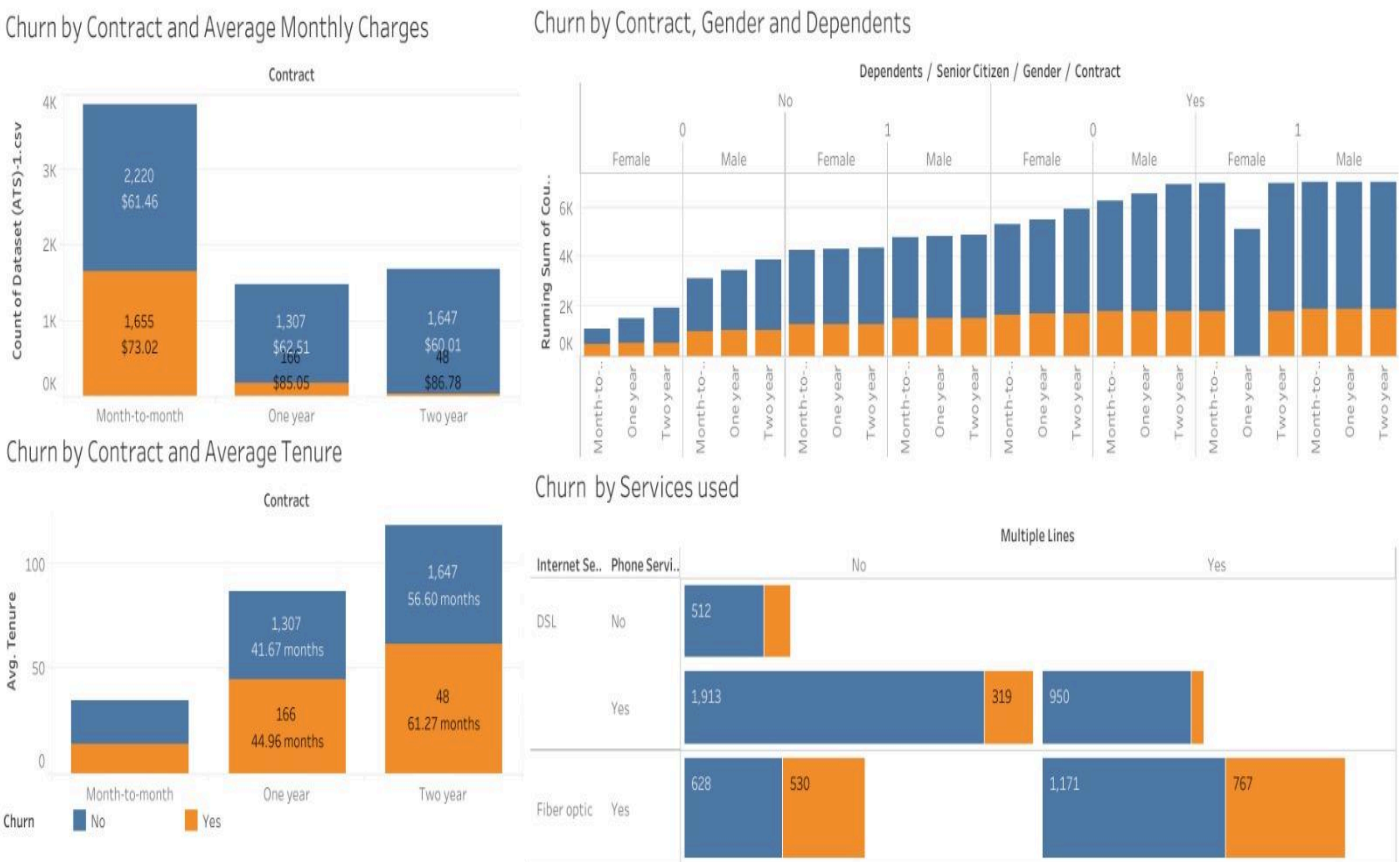
The major goal of this project is to analyze customer churn behavior using a dataset that includes a variety of factors such as contract kinds, internet services, and demographic data. This investigation seeks to identify the primary causes of churn, create predictive models to estimate customer attrition, and provide practical methods to enhance customer retention. The project includes exploratory data analysis (EDA), clustering analysis, and predictive modeling.



2. Exploratory Data Analysis (EDA)

2.1. Descriptive Statistics

Understanding the basic statistics of the dataset is crucial for uncovering trends and patterns. The descriptive statistics for customer churn are segmented by contract type, revealing significant variations in customer characteristics based on churn status and contract type.



https://public.tableau.com/app/profile/pratap.uchai.thakuri/viz/EDAChurnTelco/Dashboard1?publish=yes

Churn by Contract Type:

Month to Month Contracts:

Nonchurn Customers: The mean monthly charge is $61.5, with a standard deviation of $27.9, indicating substantial variability in charges.

Churn Customers: The mean monthly charge rises to $73.0, with a standard deviation of $24.1. This increase in average charges among churned customers suggests a relationship between higher monthly charges and increased churn.

OneYear Contracts:

Nonchurn Customers: The mean monthly charge is $62.5, with a standard deviation of $31.7.

Churn Customers: The mean monthly charge increases to $85.1, with a standard deviation of $25.6. This trend suggests that higher charges are also associated with higher churn rates for one year contracts.

TwoYear Contracts:

Nonchurn Customers: The mean monthly charge is $60.0, with a standard deviation of $34.5.

Churn Customers: The mean monthly charge is $86.8, with a standard deviation of $28.9. The higher mean charges for churned customers indicate that higher pricing may contribute to churn even in long term contracts.

Range of Monthly Charges ${ 18.25 - 118.75 }
This is not so okay, and new packages can be developed with least difference in the prices considering retention of customers.

	seniorcitizen	**tenure**	**monthly charges**	**churn_rate**
count	7043.000000	7043.000000	7043.000000	7043.000000
mean	0.162147	32.371149	64.761692	0.265370
std	0.368612	24.559481	30.090047	0.441561
min	0.000000	0.000000	18.250000	0.000000
25%	0.000000	9.000000	35.500000	0.000000
50%	0.000000	29.000000	70.350000	0.000000
75%	0.000000	55.000000	89.850000	1.000000
max	1.000000	72.000000	118.750000	1.000000

Churn Rates by Contract Type:

MonthtoMonth: 42.7%
OneYear: 11.3%
TwoYear: 2.8%

churn	contract	count	mean	std	min	25%	50%	75%	max
No	**Month-to-month**	**2220.0**	**61.5**	**27.9**	**18.8**	**38.5**	**65.0**	**84.9**	**116.5**
	One year	**1307.0**	**62.5**	**31.7**	**18.2**	**24.8**	**64.8**	**91.2**	**118.6**
	Two year	**1647.0**	**60.0**	**34.5**	**18.4**	**23.8**	**63.3**	**89.8**	**118.8**
Yes	**Month-to-month**	**1655.0**	**73.0**	**24.1**	**18.8**	**55.2**	**79.0**	**90.9**	**117.4**
	One year	**166.0**	**85.1**	**25.6**	**19.3**	**70.1**	**95.0**	**104.7**	**118.4**
	Two year	**48.0**	**86.8**	**28.9**	**19.4**	**73.6**	**97.3**	**108.4**	**116.2**

The analysis reveals that consumers with month-to-month contracts are substantially more likely to churn than those with longer-term contracts. This underlines the potential benefits of pushing consumers to sign longer-term contracts in order to enhance retention.

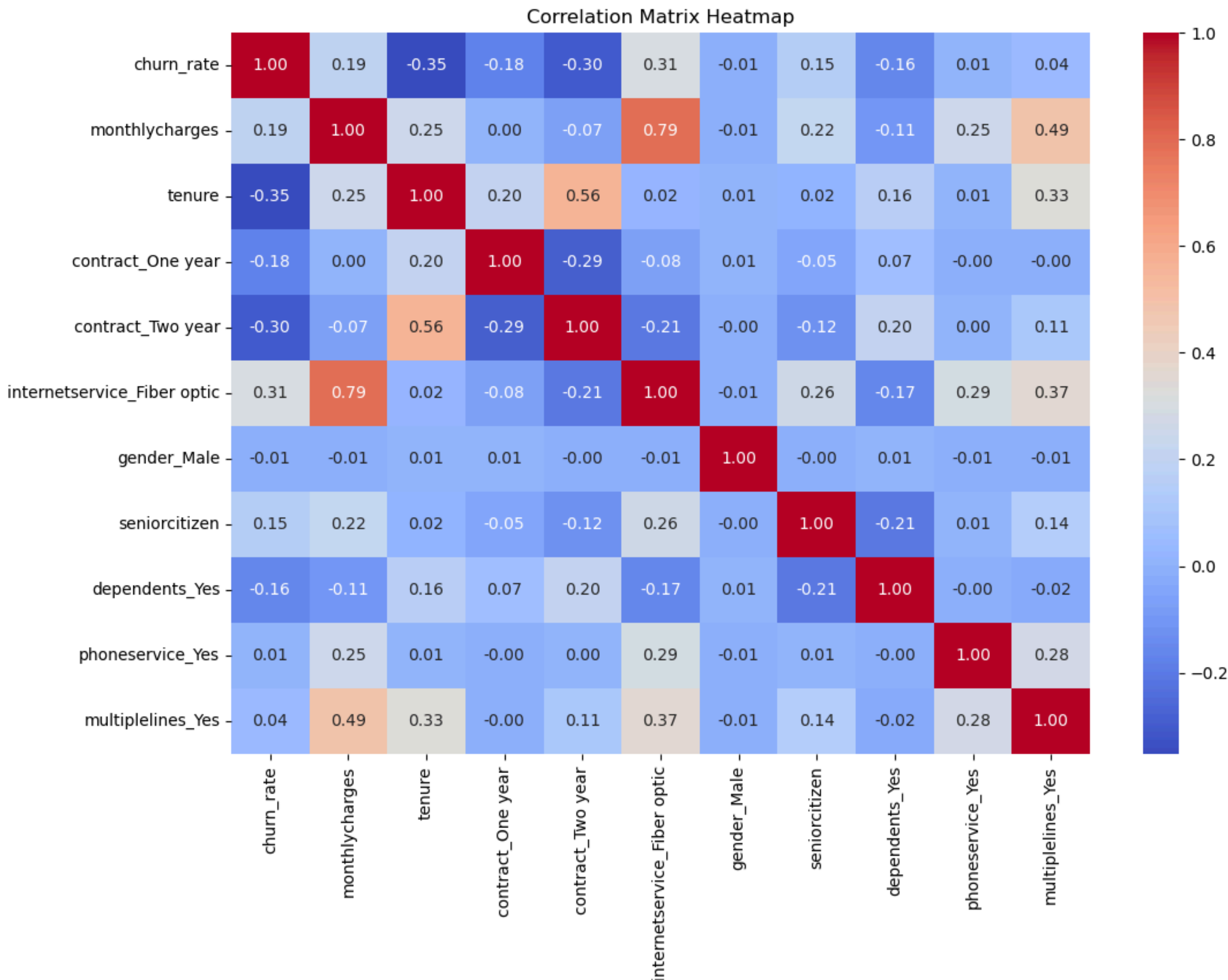
2.2. Feature Correlation

Correlation analysis was conducted to identify the relationships between various features and churn:

According to the exploratory data, clients who have fiber optic internet and pay higher monthly fees are more likely to churn. Furthermore, clients on month-to-month contracts are more likely to depart than those on longer terms.

Customers with one- or two-year contracts, as well as those with longer tenures, are less likely to churn. Having dependents also reduces the likelihood of turnover. Senior folks, on the other hand, show a modest rise in churn risk, with gender having no influence.

Based on these findings, retention efforts can be made at clients with fibre optic services, high prices, and month-to-month contracts. Offering incentives for longer contracts and promoting bundled services could be effective strategies to reduce churn and improve customer retention.



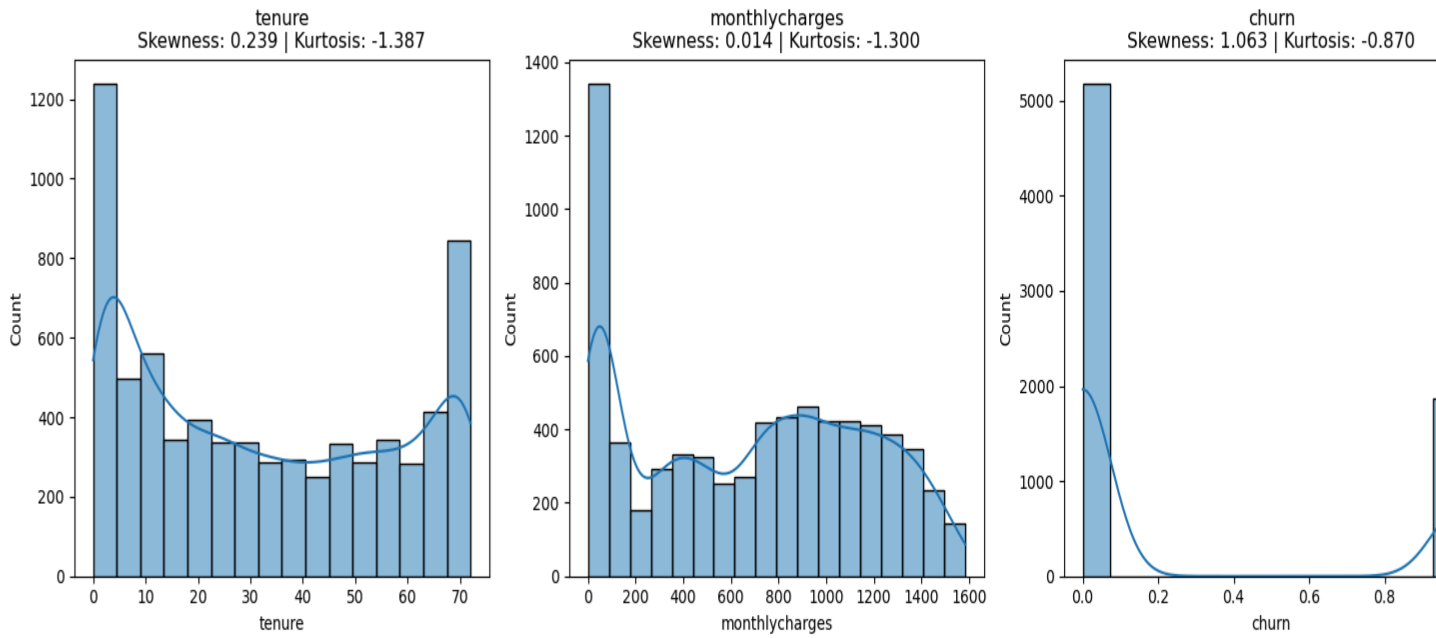
2.3. Data Distribution and Skewness

The distribution of key features was analyzed to understand their characteristics and how they may influence churn:

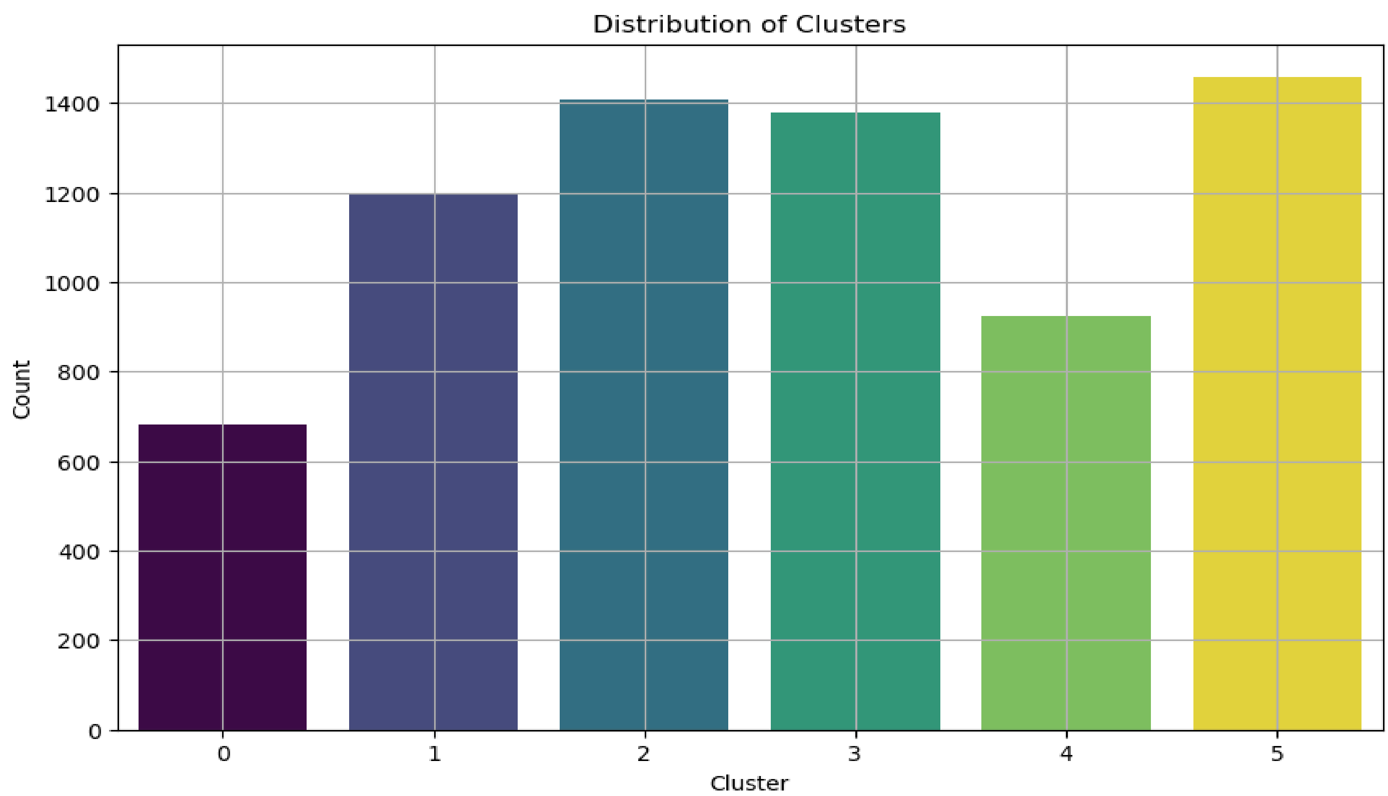
Tenure: Exhibits right skewed distribution (Skewness: 0.239, Kurtosis: 1.387). This suggests that a majority of customers have relatively short tenures, with a smaller number having longer tenures.

Monthly Charges: Approximates a normal distribution (Skewness: 0.014, Kurtosis: 1.300), indicating stable charge patterns across customers.

Churn: Shows significant right skewness (Skewness: 1.063) and flatness (Kurtosis: 0.870), reflecting that churn is less common but has a substantial impact when it occurs.



3. Clustering Analysis



3.1. Cluster Centers

Clustering analysis was performed to segment the customer base into distinct groups based on their attributes. The identified cluster centers provide insights into different customer segments:

1. Cluster 0: Characterized by high charges and high churn rates.
2. Cluster 1: Represents customers with moderate charges and moderate churn rates.
3. Cluster 2: Consists of customers with low charges and low churn rates.
4. Cluster 3: Features customers with high tenure and moderate charges.
5. Cluster 4: Includes customers with low tenure and high charges.
6. Cluster 5: Contains customers with moderate tenure and low charges.

gender	Senior Citizen	Dependents	tenure	Phone Service	Multiple Lines	Internet Service	Contract	
-0.329133	0.353338	-0.031946	0.022155	0.184638	0.327438	0.019903	-0.187209	1
-0.357524	-0.849719	-0.812817	0.027365	-0.038405	0.327438	-0.532067	-0.885660	-
-0.026192	-0.025801	-0.755556	0.019813	0.005380	-3.054010	-0.854176	-0.885660	0
-0.439916	-0.488332	0.728730	-0.018835	-0.204905	0.327438	0.233767	1.122395	-
2.273159	-0.007143	0.815520	-0.042136	-0.509131	0.327438	0.523846	0.995389	-
-0.326464	1.087654	-0.102234	-0.008792	0.424128	0.327438	0.344410	-0.372891	-

3.2. Cluster Distribution

The distribution of customers across the identified clusters is as follows:

Cluster 5: 1,458 customers
Cluster 2: 1,409 customers
Cluster 3: 1,378 customers
Cluster 1: 1,194 customers
Cluster 4: 922 customers
Cluster 0: 682 customers

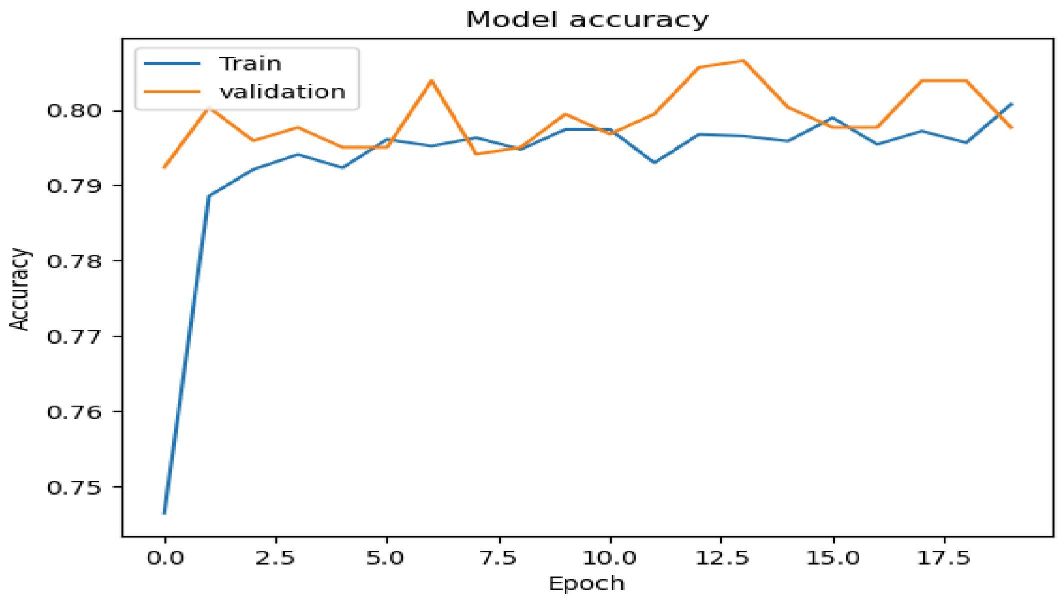
Clusters with high churn rates are characterized by month-to-month contracts and greater monthly fees. cheaper churn rates, on the other hand, are associated with longer tenures and cheaper costs. This clustering finding emphasizes the significance of contract type and pricing in predicting client attrition.

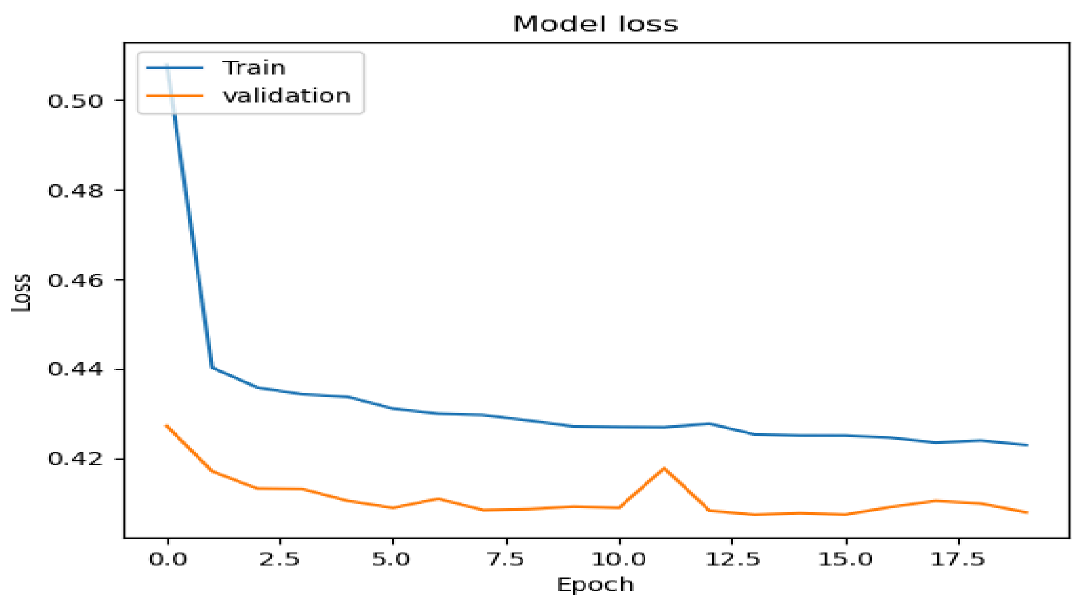
4. Predictive Modeling

4.1. Model Performance

An Artificial Neural Network (ANN) model was developed to predict customer churn. The performance metrics for the model are as follows:

Training Accuracy: 79.73%
Training Loss: 0.4192
Validation Accuracy: 81.19%
Validation Loss: 0.4057

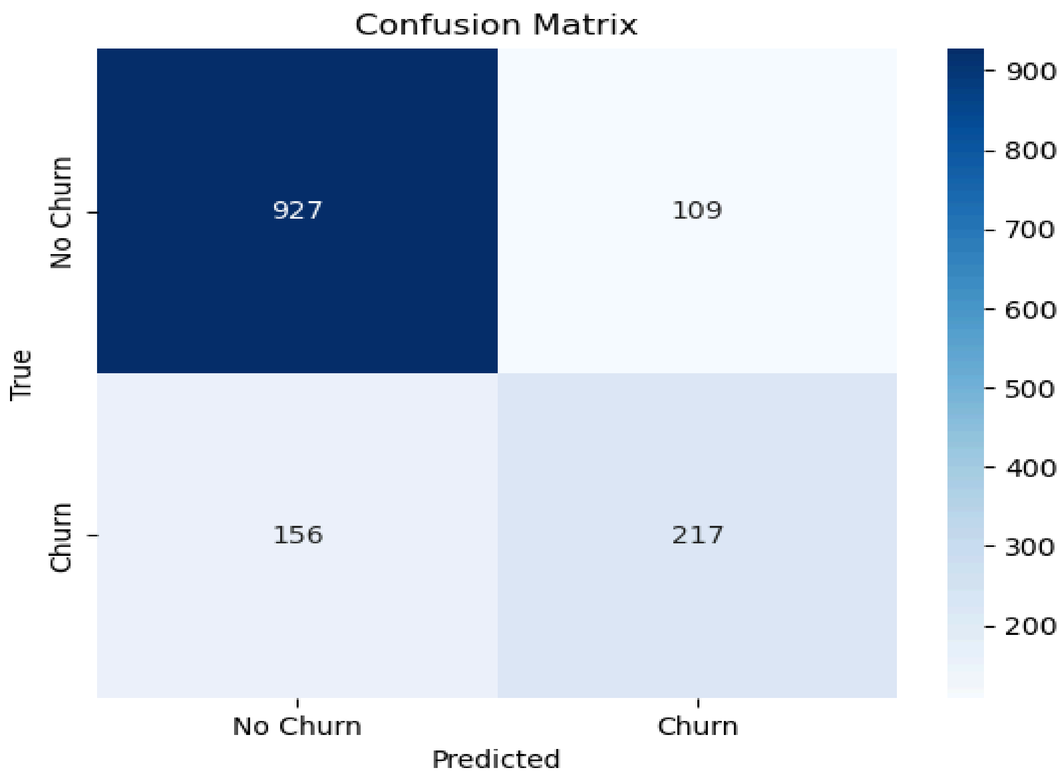




These metrics indicate that the model performs effectively in distinguishing between churn and non churn customers, with a slightly higher accuracy on the validation set compared to the training set.

4.2. Classification Report

The classification report for the validation set provides the following metrics:

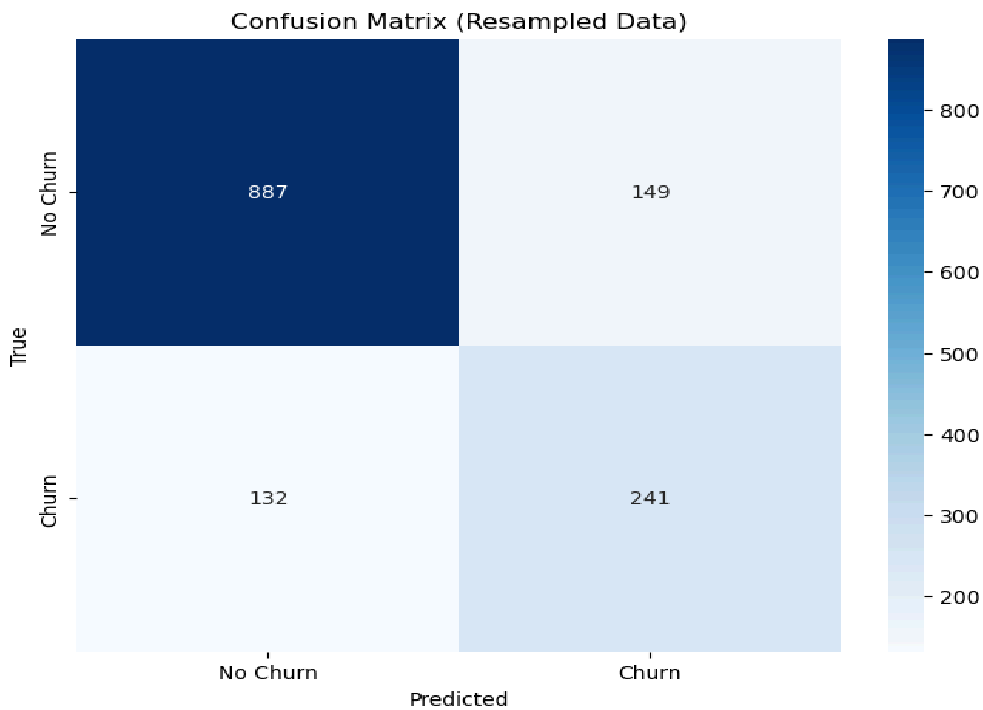


Classification Report:

	precision	recall	f1-score	support
No Churn	0.86	0.89	0.87	1036
Churn	0.67	0.58	0.62	373
accuracy			0.81	1409
macro avg	0.76	0.74	0.75	1409
weighted avg	0.81	0.81	0.81	1409

No Churn: Precision of 0.86 and Recall of 0.89, demonstrating strong performance in identifying customers who do not churn.

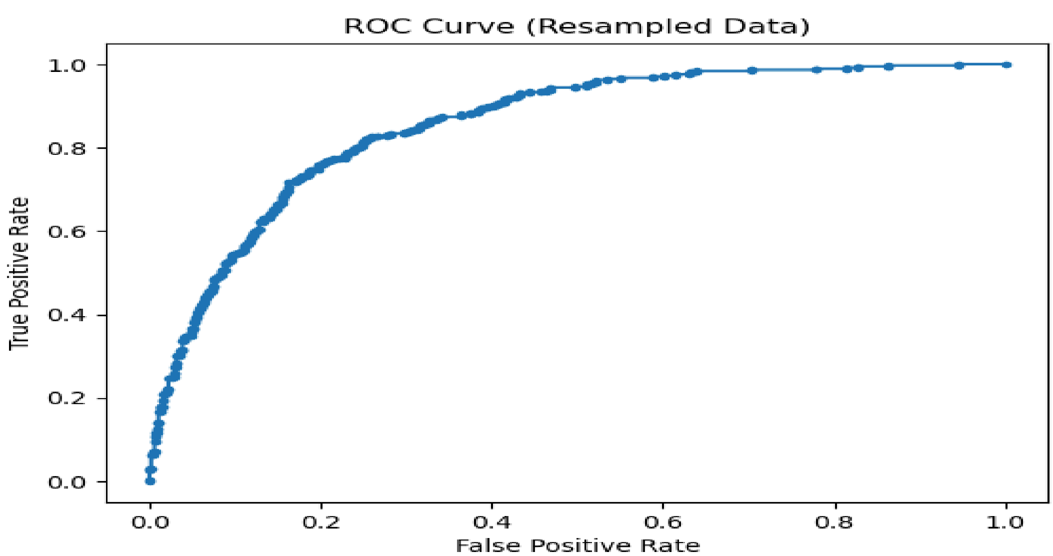
Churn: Precision of 0.67 and Recall of 0.58, showing that while the model can identify churned customers, there is room for improvement in this area.



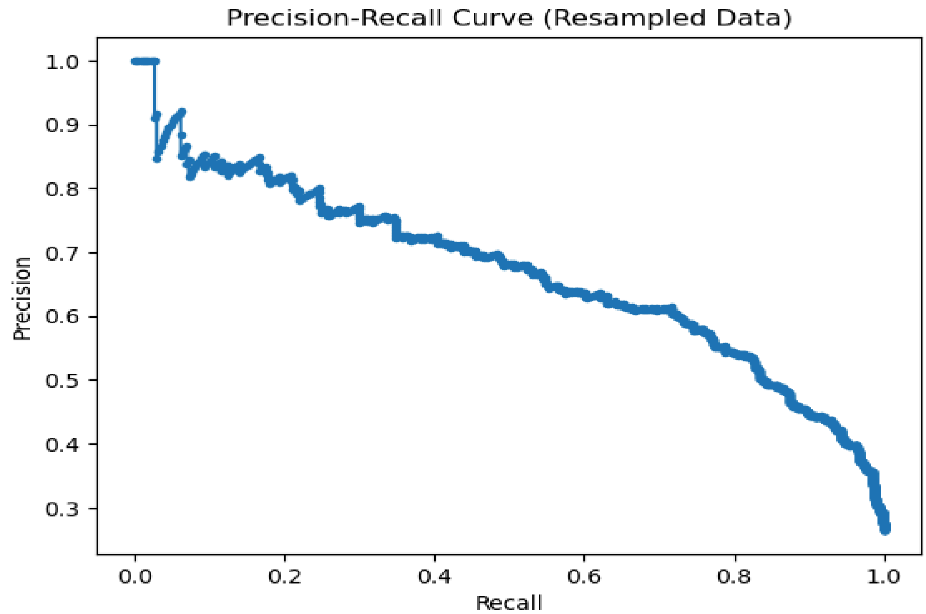
Classification Report (Resampled Data):

	precision	recall	f1-score	support
No Churn	0.87	0.86	0.86	1036
Churn	0.62	0.65	0.63	373
accuracy			0.80	1409
macro avg	0.74	0.75	0.75	1409
weighted avg	0.80	0.80	0.80	1409

ROC-AUC (Resampled Data): 0.8540



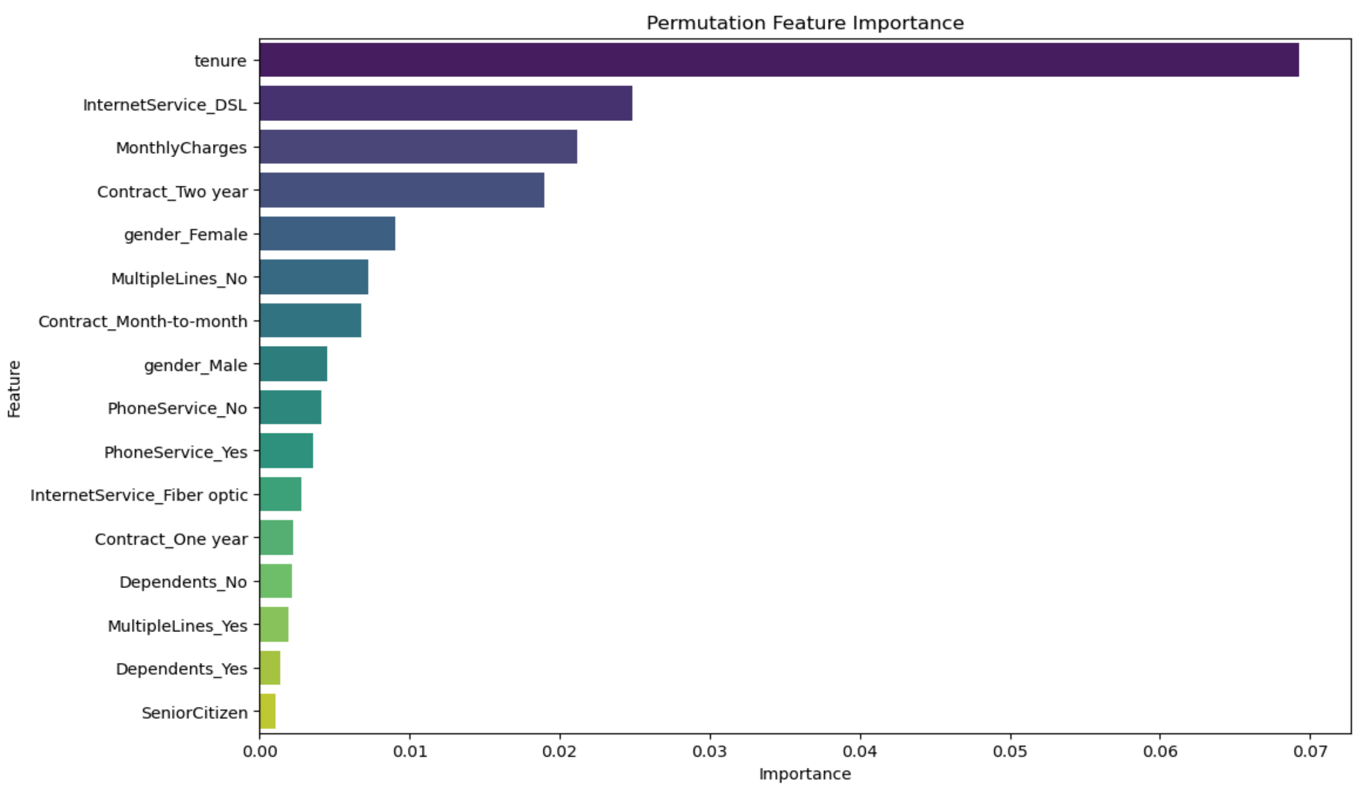
Average Precision (Resampled Data): 0.6703



These results highlight that while the model is effective at predicting non-churn customers, it could benefit from enhancements to improve its ability to detect churn cases.

4.3. Feature Importance

The importance of various features in predicting churn was assessed:



Tenure: Contributes 6.9% to the model's predictions.
Monthly Charges: Contributes 2.1%.
Internet Service (DSL): Contributes 2.5%.[See correlation factor, re encoded data as there was issue in ANN model preparation]

Tenure, monthly charges and Internet Service are identified as the most significant predictors of churn, emphasizing their critical role in the predictive model.

4.4. Model Insights

The ANN model reveals that tenure and monthly charges are the most influential factors in predicting churn. Customers with shorter tenures and higher monthly charges are more likely to churn. These insights suggest that focusing on these attributes could help in designing effective retention strategies.
https://github.com/Pratapuchai2056/Churn_Analysis_Project_WIL/blob/main/Predictive_Modeling/model_trained_on_resampled_data.keras

Note: Recall factor is great and very realistic in the resampled model so we have used a resampled model rather than the original one, however accuracy is slightly better on the original model. Please, refer to this
https://github.com/Pratapuchai2056/Churn_Analysis_Project_WIL/blob/main/Predictive_Modeling/model_trained_on_original_data.keras

5. Conclusions and Recommendations

5.1. Key Findings

The analysis provides several key insights into customer churn:
Customer Retention: Customers with month to month contracts and higher monthly charges are at a higher risk of churn.

Contract Types: longer term contracts are associated with significantly lower churn rates.

Feature Impact: Tenure and monthly charges are the most significant predictors of churn.

5.2. Recommendations

Based on the findings, the following recommendations are proposed:
Retention Strategies: Develop targeted retention strategies for customers with month to month contracts and high monthly charges. Consider offering incentives or discounts to encourage these customers to switch to longer term contracts.

Contract Policy: Revise contract options to make long term contracts more attractive. This could include offering competitive pricing or additional benefits for longer term commitments.

Customer Segmentation: Utilize clustering insights to tailor marketing and customer support strategies based on cluster characteristics. This approach can help address the specific needs and preferences of different customer segments.

5.3. Future Work
Future efforts should focus on the following areas:
Model Enhancement: Explore additional features and advanced modeling techniques to improve the predictive accuracy of churn forecasts.

Customer Feedback Integration: Implement mechanisms to gather and analyze customer feedback to gain deeper insights into the reasons behind churn and address these issues proactively.

This comprehensive analysis provides actionable insights into customer churn and offers strategic recommendations to enhance customer retention efforts and reduce churn rates effectively.

Appendices

Appendix A: Data Preparation
Details of data preprocessing, including encoding categorical variables and scaling features, are documented.
https://github.com/Pratapuchai2056/Churn_Analysis_Project_WIL/blob/main/Data_Preparation/DatPreProAndEncoding.md

Appendix B: Clustering Analysis Details
Full results of clustering analysis, including cluster characteristics and distribution, are provided in
https://github.com/Pratapuchai2056/Churn_Analysis_Project_WIL/blob/main/Clustering_Analysis/Trained_K-Means_Model.md

Appendix C: Predictive Modeling Code
The code used for developing and training the ANN model, including scripts and configurations, is included for reference.
https://github.com/Pratapuchai2056/Churn_Analysis_Project_WIL/blob/main/Predictive_Modeling/Ann_Modelling.md

Appendix D: Model Performance Metrics
Detailed performance metrics, including accuracy, loss, and classification reports, are available for review.
https://github.com/Pratapuchai2056/Churn_Analysis_Project_WIL/blob/main/Predictive_Modeling/Predictive_modelling.pdf

This final report summarizes a thorough investigation of customer churn, including a thorough assessment of the dataset, insights from EDA and clustering, and findings from predictive modeling based on level of our competency. The recommendations based on this study are intended to successfully improve client retention and reduce churn rates. For any corrections to me made it is highlighted in red.