Customer Churn
Descriptive and Predictive Analysis

Minh Nguyen Ngoc Tue

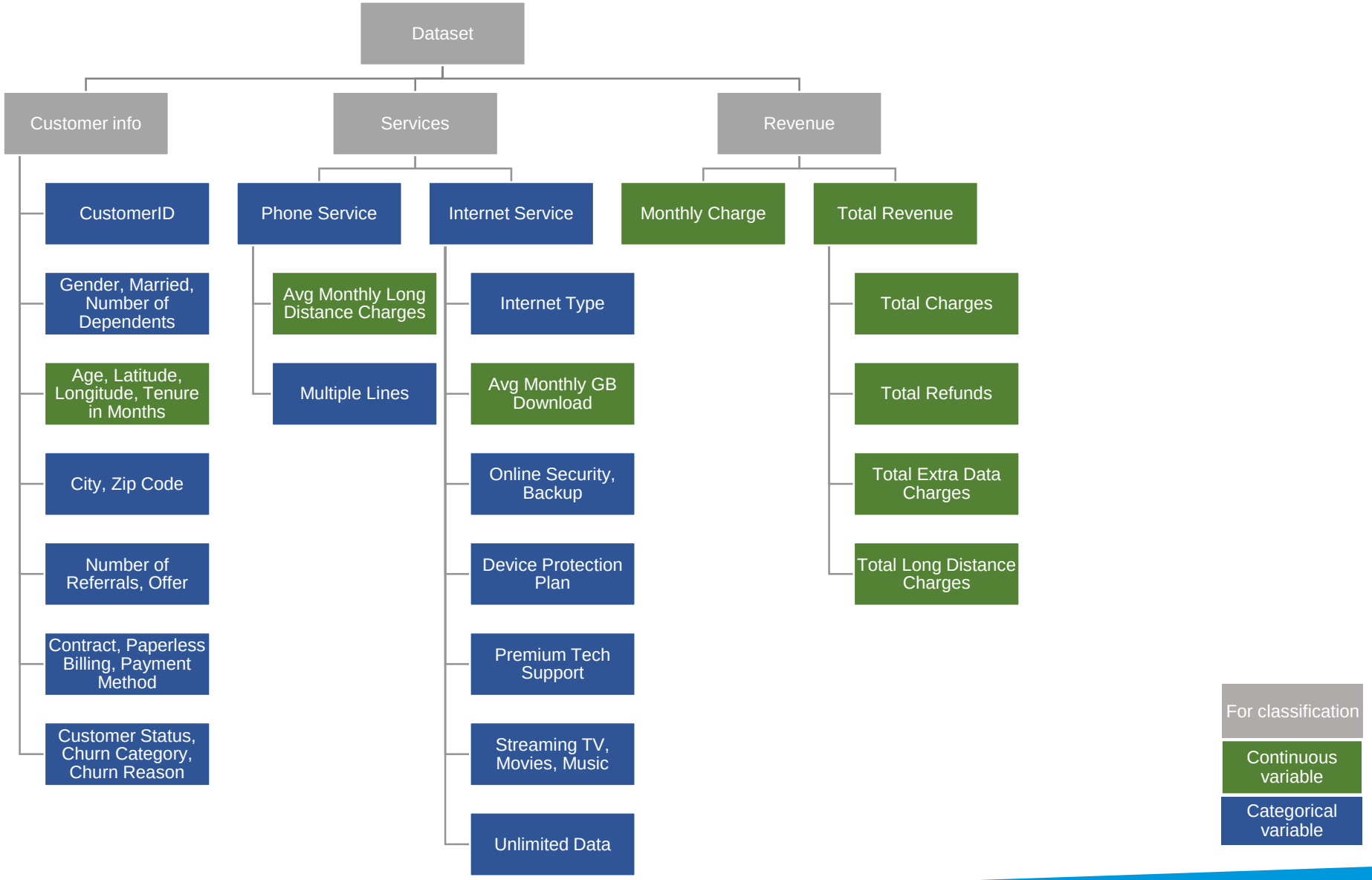
December 2023

Presentation Outline

- Data Overview
- Data Exploration
- Customer Churn Prediction

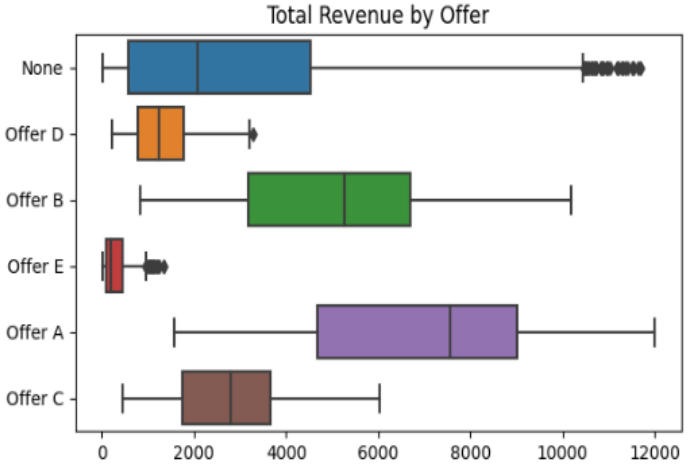
Data Overview

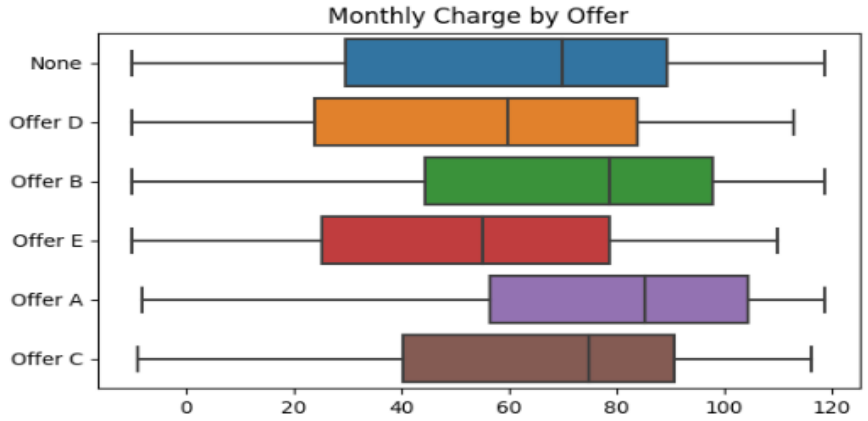
Data Overview

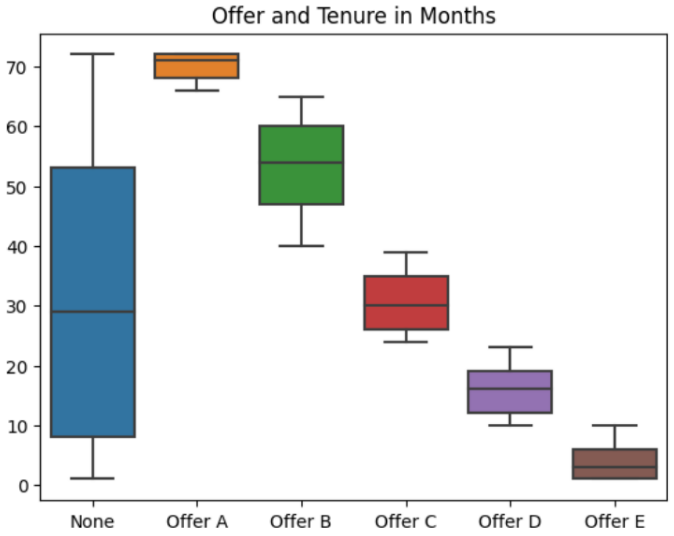


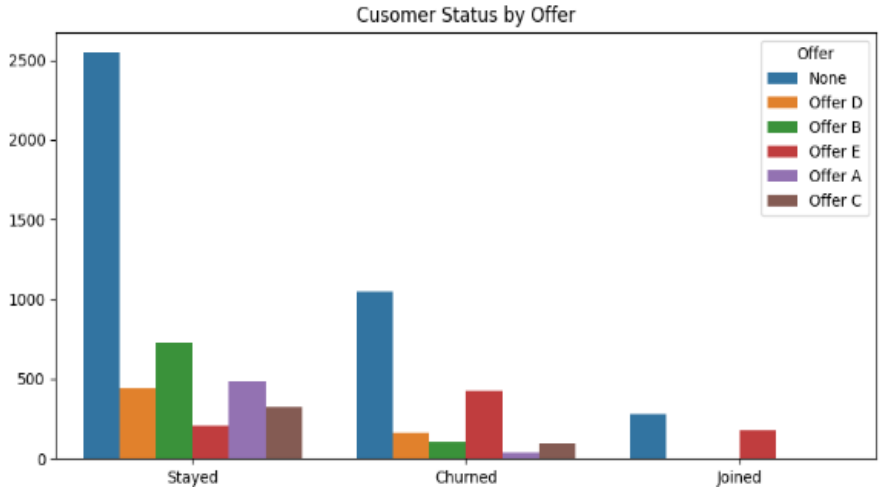
Data Exploration

Data Exploration - Offer types



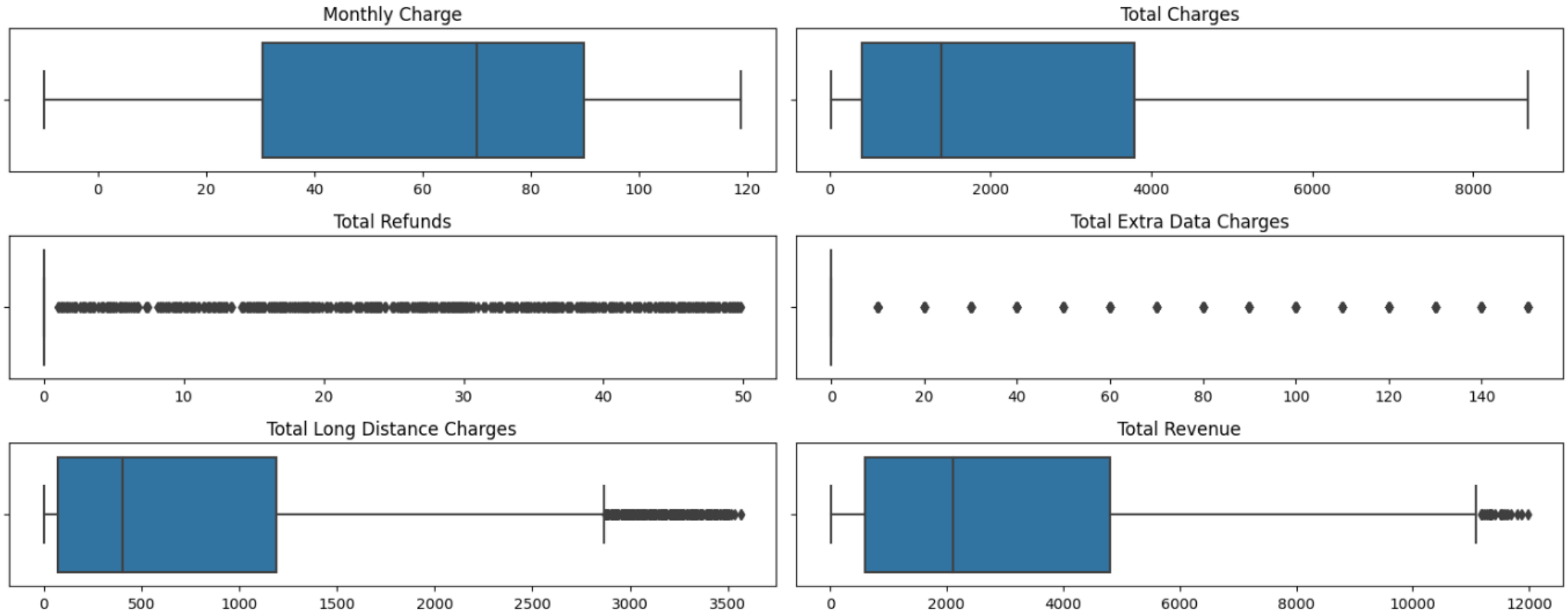






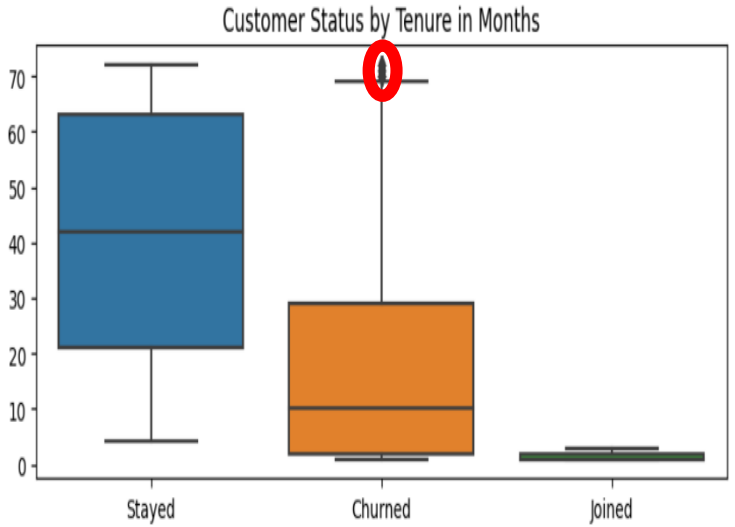
- Offer E is the least performing offer & also customers churned are also from Offer E.
- Company should either discontinue or make changes to make this offer more attractive.

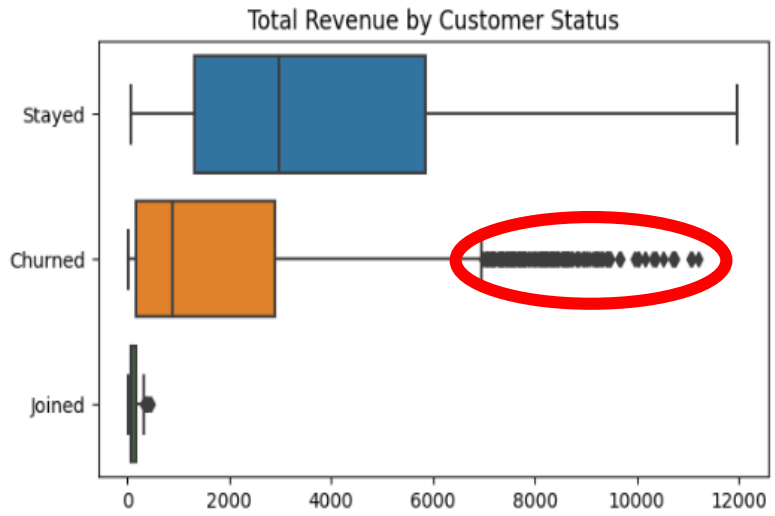
Data Exploration - Revenue Streams



- Total revenue is based on total charges & total charges are mainly contributed by the long-distance charges.
- Total refunds & extra data charges are very low.

Data Exploration – Churn reason



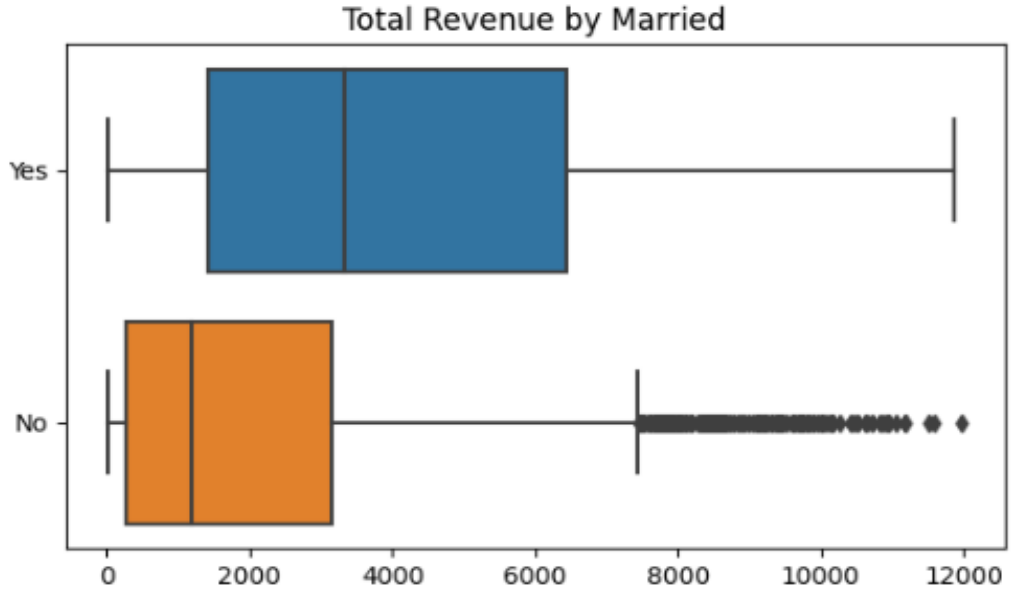


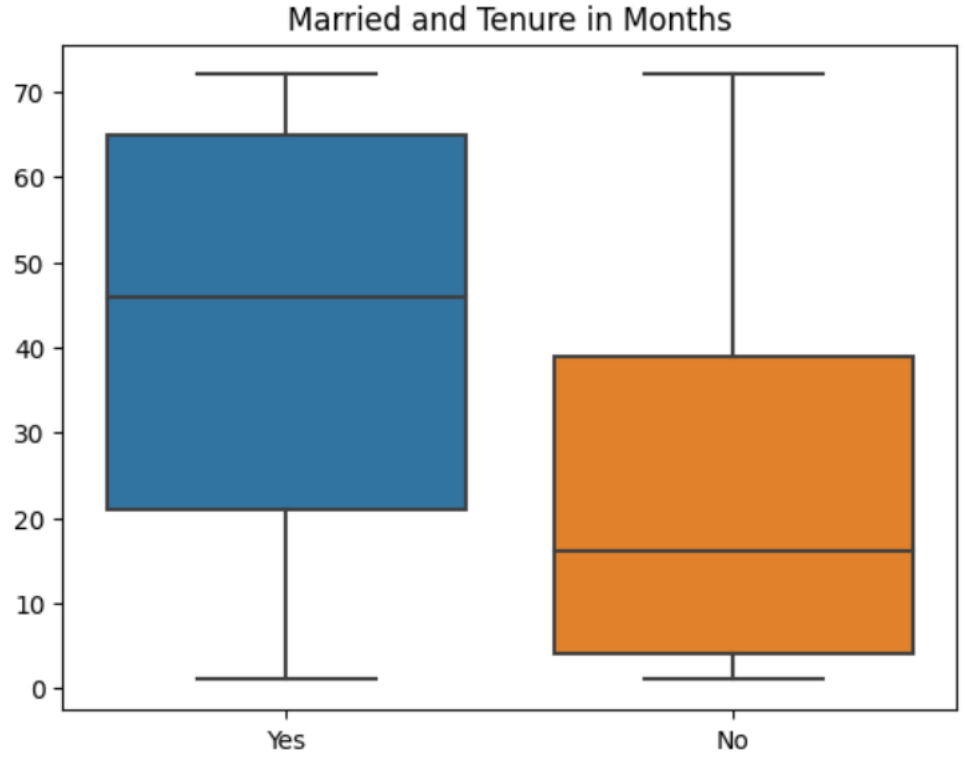
	All customers	Long-term customers	High-value customers
Total Revenue Mean	3,034	8,470	8403
Monthly Charge Mean	64	97	102

Churn Reason	Long-term customers	High-value customers
Competitor	39%	52%
Dissatisfaction	22%	15%
Price	22%	11%
Attitude	17%	16%
Other	0%	6%

- The company is losing its revenue mainly due to competitor's strategy, dissatisfaction & attitude of the company
- Company should conduct a thorough analysis of competitors' services, pricing models, and customer satisfaction levels.

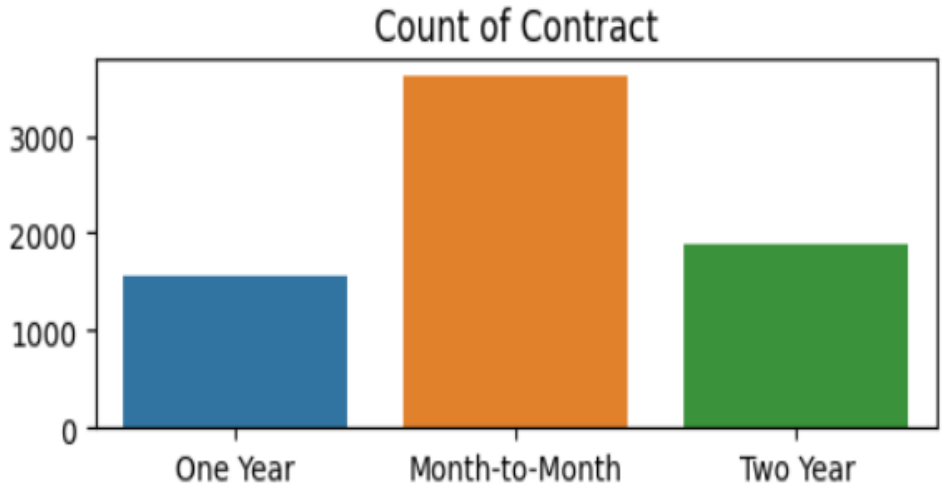
Data Exploration – Married/Single

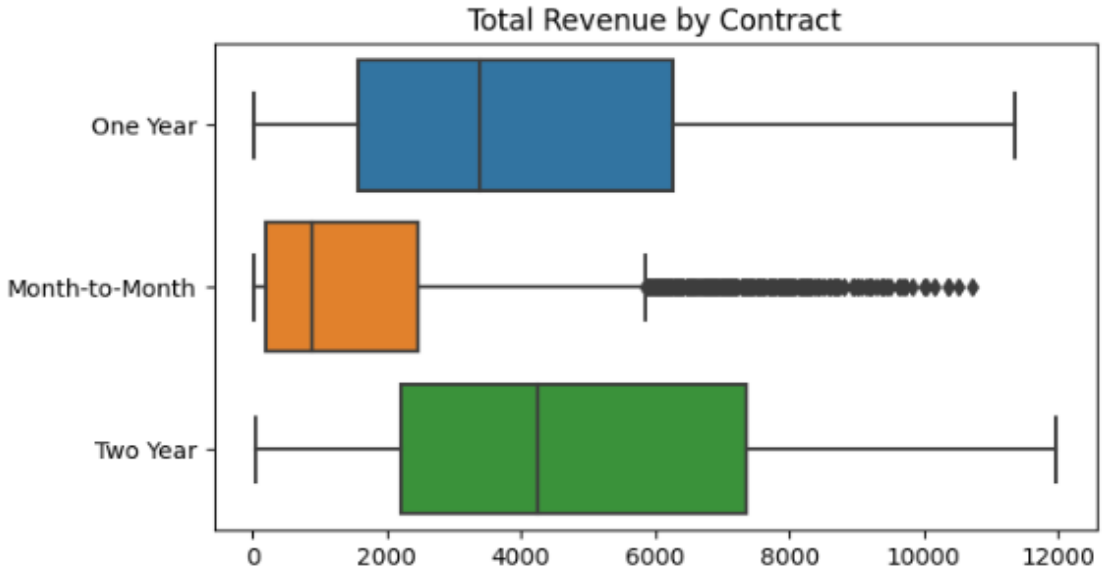




The company is generating more revenue from married people & the tenure of the contract is also more as compared to the single people.

Data Exploration – Contract types





- The maximum number of contracts are from customers who use services on a monthly basis,
- The total revenue from monthly contracts is lesser than all other types of contracts.
- The company should establish open communication channels to gather feedback from the customers who use monthly services & use this feedback to continuously refine and improve the services and promotions offered to meet their evolving needs.

Customer Churn Prediction

Logistics Regression

Version of data

- unscaled and unbalanced data
- scaled and unbalanced data
- unscaled and balanced
- scaled and balanced

In Logistic Regression, **scaled and balanced data** give the best performance

- accuracy_score: 81%
- balanced_accuracy_score: 83%
- confusion_matrix

```
[[322  53]
 [191 752]]
```

- classification_report

	precision	recall	f1-score	support
0	0.63	0.86	0.73	375
1	0.93	0.80	0.86	943
accuracy			0.81	1318
macro avg	0.78	0.83	0.79	1318
weighted avg	0.85	0.81	0.82	1318

Decision Tree

Version of data	Parameter combination
- unscaled and unbalanced data - scaled and unbalanced data - unscaled and balanced - scaled and balanced	- max_depths: 1-21 - criterions: “gini”, “entropy”, “log_loss”

In **unscaled and balanced data**, the combination of **max_depth=3, criterion=entropy,** give the best performance

- accuracy_score: 73%
- balanced_accuracy_score: 78%
- confusion_matrix

```
[[335  40]
 [315 628]]
```

- classification_report

	precision	recall	f1-score	support
0	0.52	0.89	0.65	375
1	0.94	0.67	0.78	943
accuracy			0.73	1318
macro avg	0.73	0.78	0.72	1318
weighted avg	0.82	0.73	0.74	1318

Gaussian Naïve Bayes

Version of data

- unbalanced data
- oversampling data
- undersampling data

In Gaussian Naive Bayes, **oversampling data** gives the best performance

- accuracy_score: 70%
- balanced_accuracy_score: 74%
- confusion_matrix

```
[[316  59]
 [342 601]]
```

- classification_report

	precision	recall	f1-score	support
0	0.48	0.84	0.61	375
1	0.91	0.64	0.75	943
accuracy			0.70	1318
macro avg	0.70	0.74	0.68	1318
weighted avg	0.79	0.70	0.71	1318

K Nearest Neighbours

Version of data	Parameter combination
- scaled and unbalanced data - scaled and balanced data	- n_neighbors: 1-11 - metrics: "minkowski","manhattan","euclidean","canberra" - weights: "uniform","distance"

In scaled and balanced data, the combination of **n=8, m=canberra, w=uniform** give the best performance

- accuracy_score: 76%
- balanced_accuracy_score: 79%
- confusion_matrix

```
[[330  45]
 [274 669]]
```

- classification_report

	precision	recall	f1-score	support
0	0.55	0.88	0.67	375
1	0.94	0.71	0.81	943
accuracy			0.76	1318
macro avg	0.74	0.79	0.74	1318
weighted avg	0.83	0.76	0.77	1318

Random Forest

Version of data	Parameter combination
- unscaled and unbalanced data - scaled and unbalanced data - unscaled and balanced data - scaled and balanced data	- max_depths: 1-21 - metrics: "gini", "entropy", "log_loss"

In unscaled and balanced data, the combination of **max_depth=8, citerion=entropy** give the best performance

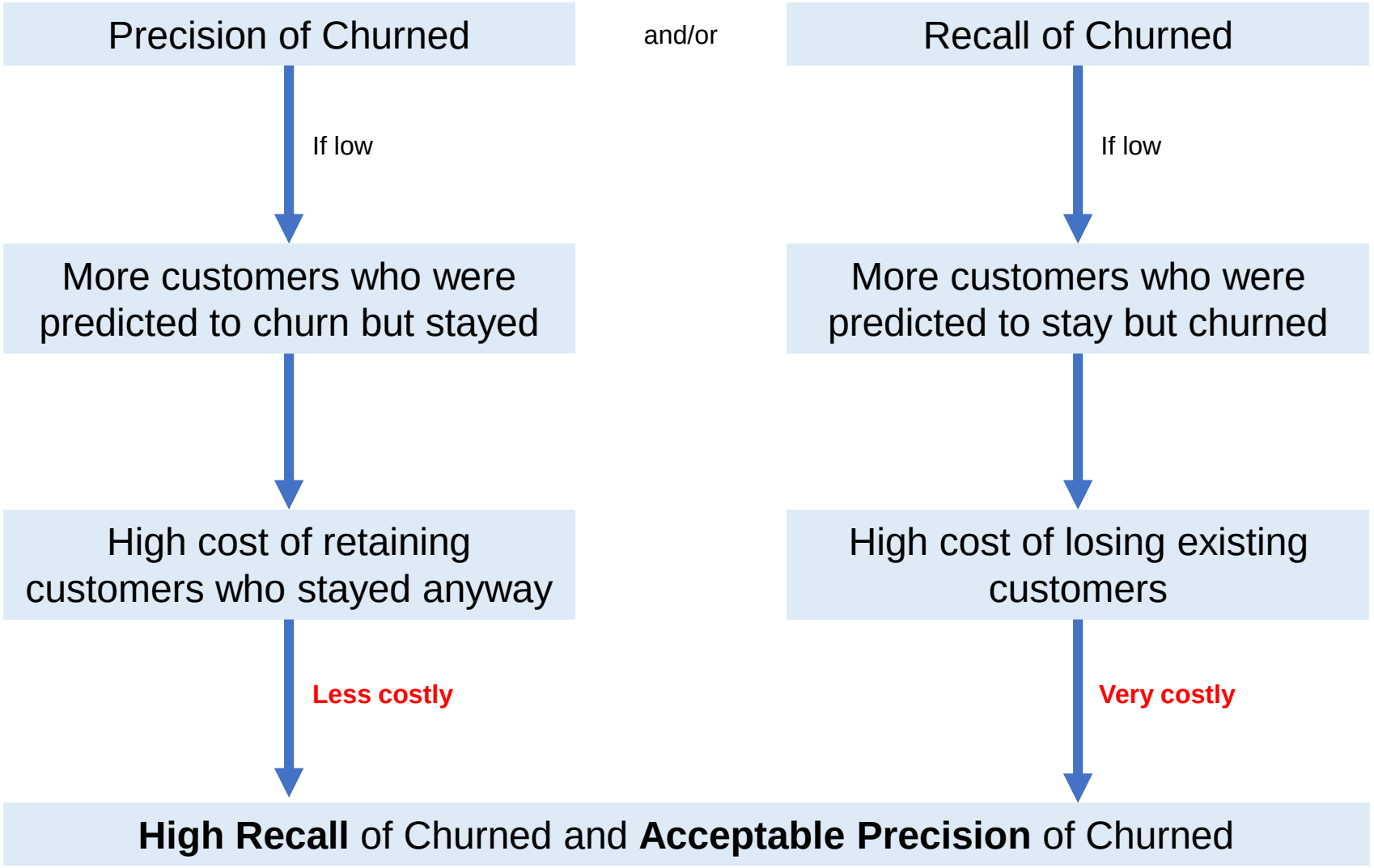
- accuracy_score: 84%
- balanced_accuracy_score: 85%
- confusion_matrix

```
[[320  55]
 [151 792]]
```

- classification_report

	precision	recall	f1-score	support
0	0.68	0.85	0.76	375
1	0.94	0.84	0.88	943
accuracy			0.84	1318
macro avg	0.81	0.85	0.82	1318
weighted avg	0.86	0.84	0.85	1318

Evaluation Principle



Conclusion

Model	Accuracy score	Balance accuracy score	Precision of Churned	Recall of Churned	F1	Ranking
Logistics	81.49%	82.81%	62.77%	85.87%	72.52%	2
KNN	75.80%	79.47%	54.64%	88.00%	67.42%	4
Decision Tree	73.07%	77.96%	51.54%	89.33%	65.37%	3
GaussianNB	69.58%	74.00%	48.00%	84.00%	61.00%	5
Random Forest	84.37%	84.66%	67.94%	85.33%	75.65%	1

Recommendation:

- **Stop offer E** which is a new offer (low Total Revenue) and only accept among new customers because it does not perform well (low Monthly Charge), and more people who churned when accepted this offer. **Promote Offer A and B**.
- **Promote Long Distance Service** because Total revenue mainly comes from Total Charges and Total Long Distance Charge (Total Refunds and Total Extra Data Charges are small).
- **Improve competitiveness** by researching about Competitors' services and improving their own services because there are many long-term customers and high-value customers who churned, they have higher average Total Revenue and Monthly Charge than all customers, and most of them leaving because of Competitor, Dissatisfaction and Price
- **Promote for Married customers** because they tend to stay longer and single customers and they generate more revenue.
- **Focus on converting monthly customers to yearly customers** instead of acquiring new customers, and change in business strategy to generate more revenue from monthly customers, because yearly customers create more revenue and the company already have so many monthly customers.