East-Coast Retention

14th November, 2023

Dhruv Patel, dp99977n@pace.edu
CS 667 Practical Data Science
MS in Data Science
Seidenberg School Of Computer Science and Information System

Agenda

- Executive summary
- Project plan
- Data
- Exploratory data analysis
- Modeling methods
- Findings
- Recommendations
- Next Steps
- Appendix

Executive Summary

Business problem: Pace Telecom is facing challenges in establishing a significant market presence in the East Coast region, despite successfully acquiring new customers.

Proposed solution: Accelerating customer acquisition without maintaining market share may lead to underperformance in the East Coast region. Given the current scenario, customer retention would possibly be the optimal solution.

Project plan Recap

Deliverable	**Details**	**Due Date**	**Status**
Data & EDA	Acquire relevant data, perform data processing & analysis, provide initial findings.	10/31/23	Completed
Modeling	Employe the potential of ML and make future predictions.	11/14/23	In Process
Recommendations	Propose best steps to successfully retain customers	12/05/23 (3 weeks from now)	Not Started

Data

Data

- Source: Marketing team
- Size: 7032 customers * 20 characteristics
- Time period: 1st january, 2023 – 1st october, 2023.

Data Exclusion

Considering the ethnicity concern, sensitive details such as name, address, banking details and joining date are excluded by the HR team.

Information this sensitive can lead to biases by any data operator.

Information like address and name must be preserved because incase of a data breach, customers can get on the verge of identity theft.

Data Assumptions

The dataset was provided by our marketing team which was extracted directly from the service dealers, so it is assumed that the dataset contains all the customers as its mandatory to register a customer before providing the service and the details in the dataset are correct.

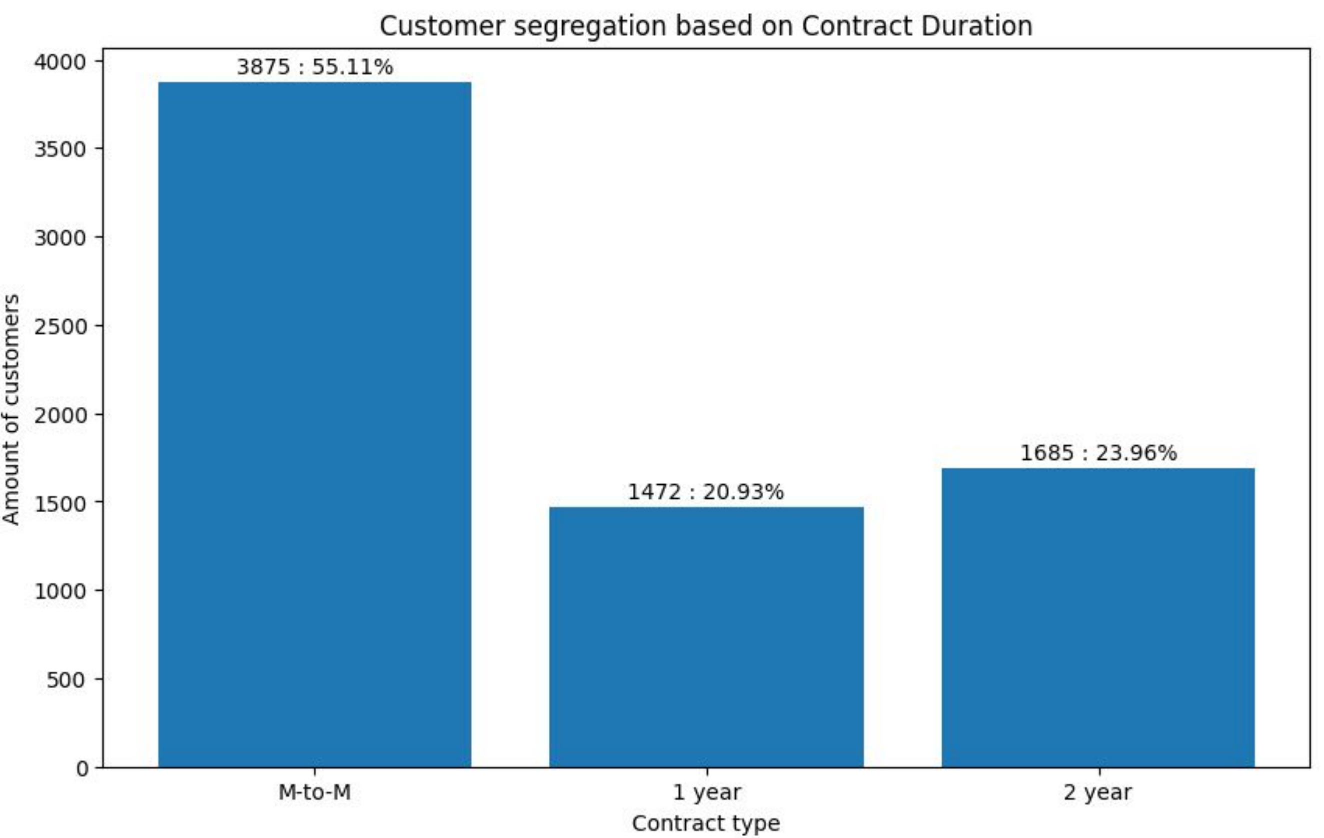
Refer this slide and for more detailed data description.

Key technical jargon

Churn: Referred when customer stops using the service or no longer does business with the company.

Exploratory Data Analysis

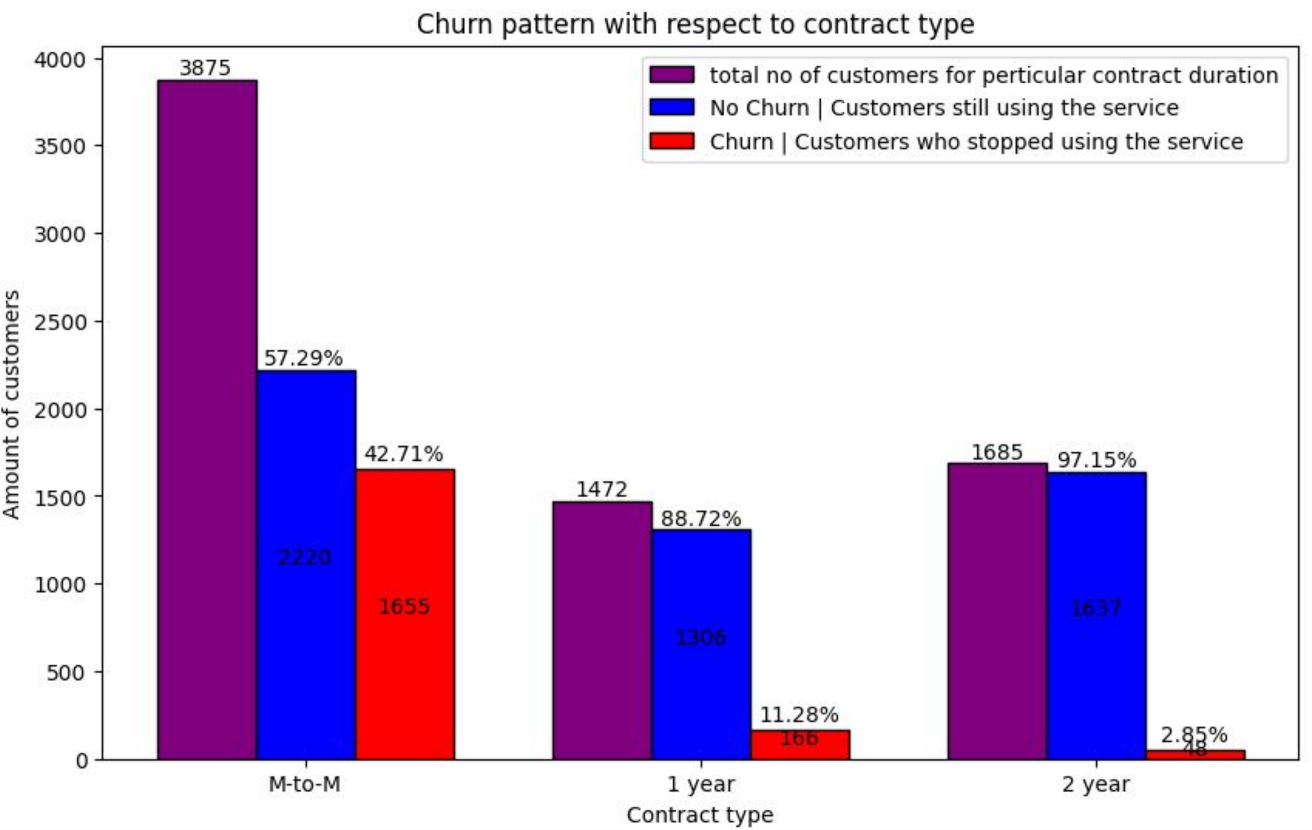
Exploring Churn pattern with types of Contract.



Key Takeaways

- Majority of customers choose Monthly contracts.
- Probably because they were new to the company and unsure about the service so chose monthly plan and kept using the same

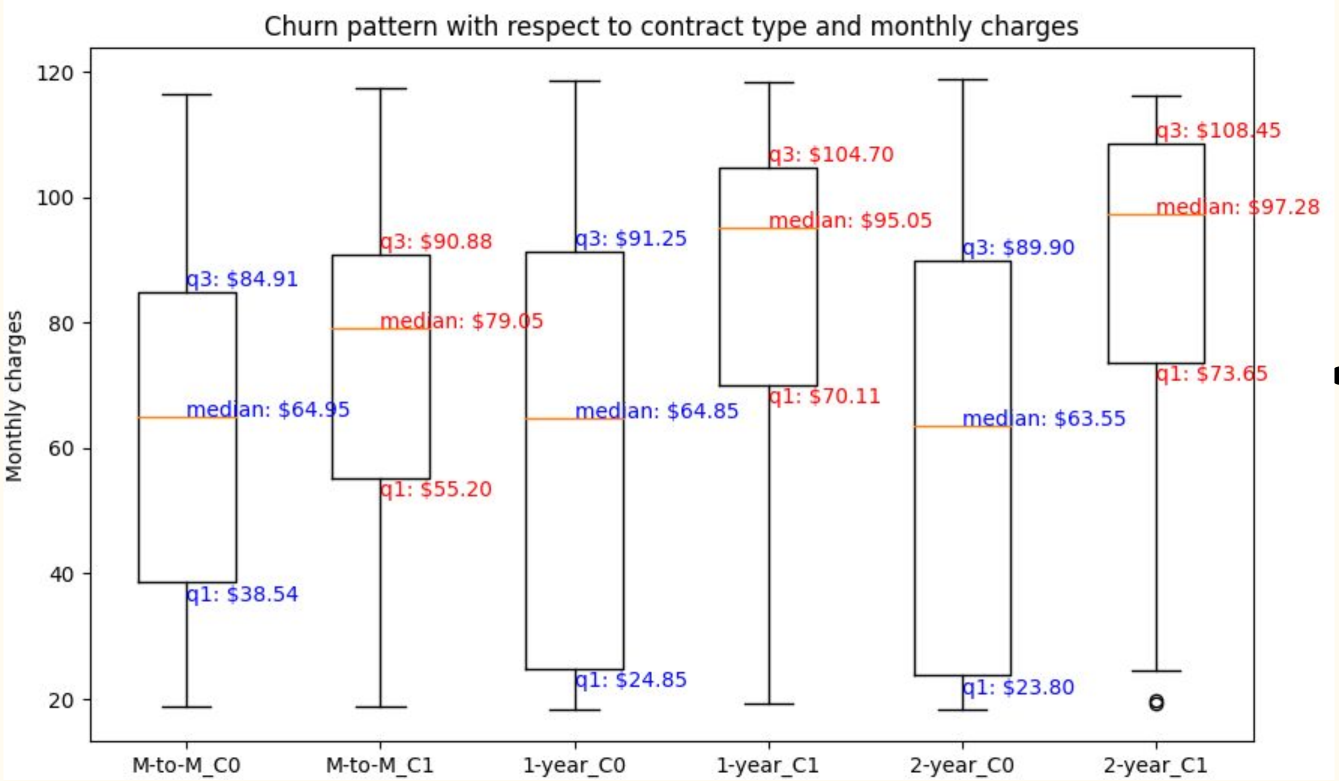
Exploring Churn pattern with types of Contract.



Key Takeaways

- Maximum churn happens in customers who choose monthly contract at 42.7%.
- Churn rate for customers with yearly contract is significantly lower at 11.28% and 2.85%.

Exploring Churn pattern with monthly charges



Graph Notes

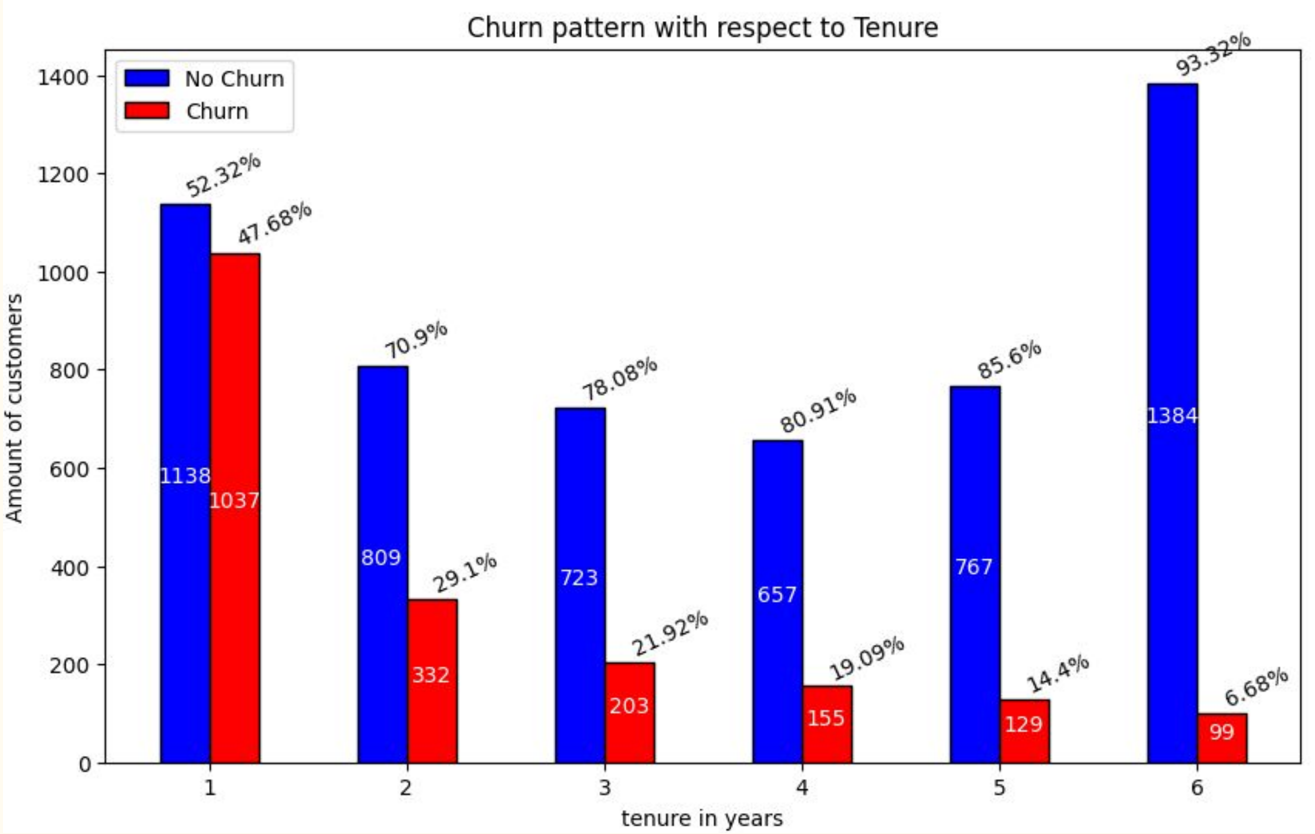
- Blue plots | No Churn
  Red plots | Churn

Key Takeaways

- Customers who churned had higher on-average monthly charges.

For detailed box-plot description

Exploring Churn pattern with Tenure.



Key notes

- Chances of customer turning down the service during the first year of operation are highest at 47.68%
- The longer the customer keeps using the service, the lesser are their chances of churning.

Modeling methods

Outcome variable

- Using this model, we are trying to predict what amount of our customers may churn by the end of this year.
- Predicting the future churn can be used in targeting retention to customers likely to churn.
- Retaining customers and acquiring new ones can help business achieve the solution proposed by marketing and data team.

Features

While predicting customer churn, we found following variables to be highly important,

- Contract : The duration of contract. (monthly, 1 year or 2 year)
- Tenure: Fow how long the customer has been using the service.
- Online Security: has customer opted for online security or not.
- Monthly Charges: the monthly billed amount.

Refer this slide for full list of features.

Random Forest Non Technical description

What's better than one expert giving advice for a problem? Having many experts.

Random forest is like having a huge team of decision-making experts, specialized in different aspects of data, and when asked, they all independently vote on best decision according to them.

This teamwork provides more accurate and reliable prediction compared to single expert. As our data has lots of different characteristics, this model would be a good fit for our problem case.

Refer this slides for technical details.

Random Forest Technical Description

Random Forest will be a perfect fit for our problem as

- The dataset is showcasing complex relationships among multiple features.
- It provides feature importance which can be used to understand the factors driving customer to churn.
- The ensemble effect will result in more stable and reliable model.

Training, testing and predicting data

- The data acquired from 1st january, 2023 – 1st october, 2023 will be used to train the model.
- A 20% randomly selected section of that data will be used to evaluate the model, to check whether the model is performing upto mark or not.
- Using the trained model, we will predict churn among customers who joined out service from october 1st to november 10th, 2023.

Findings

Model Evaluation / Recall

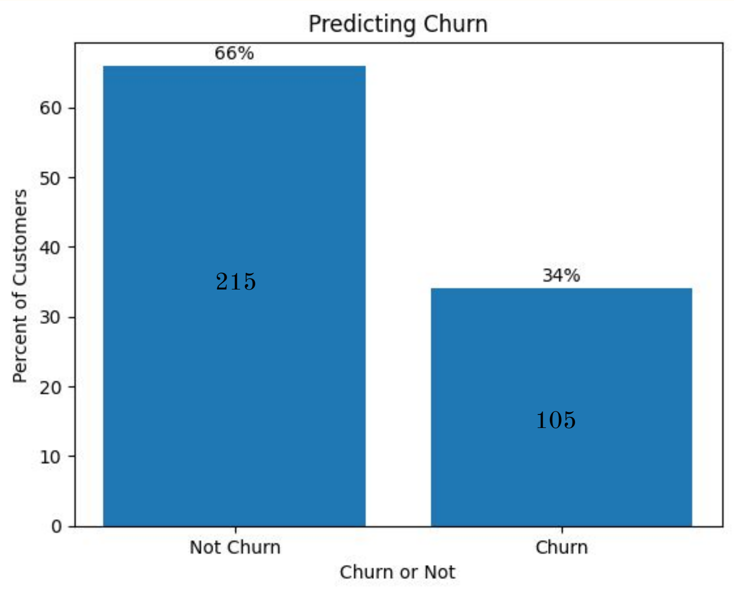
		Predicted	
		No	Yes
Actual	No	True Negatives Predicted Customers Stay Customers Actually Stay	False Positives Predicted Customers Churn Customers Actually Stay
	Yes	False Negative Predicted Customers Stay Customers Actually Churn	True Positives Predicted Customers Churn Customers Actually Churn

		Predicted	
		No Churn	Churn
Actual	No Churn	50%	15%
	Churn	5%	30%

False Negative can hide potential churns as not churns and can hurt our retention goals, So we need to make sure that our model is producing least FN.

Recall = ability of a model to correctly identify all True Positive = TP/ TP+FN = 85.71%

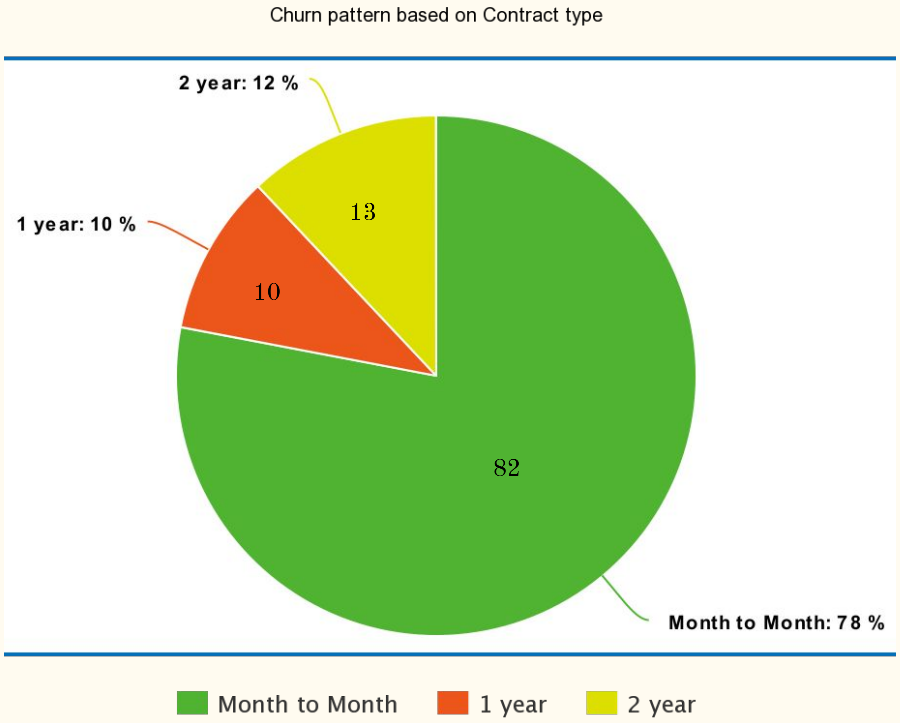
Model findings - Churn



Model Suggests that out of all the new customers we acquired from october 1st to november 10th, 34% of them might churn.

The churning of this customers can hurt our Market capture goals of 2023-Q4.

Model findings - Contract Type in churn



Among those 34% potential churning customers, 78% of them have opted for monthly contracts.

This proves the hypothesis we pointed during our analysis steps to be true. Redirect to analysis slide.

Recommendation

Recommendations

Based on the done analysis, the following steps can be recommended:

- Offer yearly contract to customers who opt monthly contract consecutively.
- Prioritize support and retain customers with tenure < 1 year without compromising support for other customers.

Recommendation - offering yearly contract.

Statistics show that customers with yearly contract has 11.28% (for 1 year) and 2.85% (for 2 year) chances of churning where 43% for monthly user.

This may be because during the short commitment, the customer might be looking for a better offer from other providers and if found, then leveraging their small commitment, they might switch to other provider.

Monthly contract also don't provide as greater value as yearly contract does, so offering them yearly contract can bring down their monthly bills, convincing them to not churn by the next contract.

By offering yearly contracts to customer, their churn rate may decrease, resulting in retaining more customers thus solving our business problem in hand.

Recommendation - Prioritize customers with tenure < 1 year

About 48% of customers churns by their first year of service, but only 29% within second year and keeps decreasing there on.

This can be because of the trust customers build by working with us, the more they stay, the more their trust, resulting in lower chances of churning.

Prioritizing support for customers with tenure < 1 year may retain group of customers who are most possible to churn.

Doing so may have good long time impact on company as more customers will build trust towards us and can be retained for longer tenures thus solving our business problem in hand.

Next Steps

Advanced model

The random forest model we build for this project was sufficient for our use case, however, it can be refined further to support our future projects.

Expanding data

Pace telecom is growing in the west coast region, we can leverage our growth and can acquire more data about our customers and the services they opt for, so that more complex problems can be answered in future.

Appendix

GitHub

Refer this link to redirect to github.

Random Forest Technical explanation

Random Forest is a machine learning model that combines multiple decision trees to make predictions. Each tree looks at a random subset of data and features, and they vote on the final prediction. This ensemble approach often leads to more accurate and robust results, handling complex relationships and reducing overfitting.

Data description 1

CustomerID: Customer ID unique for each customer

gender: Whether the customer is a male or a female

SeniorCitizen: Whether the customer is a senior citizen or not (1, 0)

Partner: Whether the customer has a partner or not (Yes, No)

Dependent: Whether the customer has dependents or not (Yes, No)

PhoneService: Whether the customer has a phone service or not (Yes, No)

MultipeLines: Whether the customer has multiple lines or not (Yes, No, No phone service)

InternetService: Customer’s internet service provider (DSL, Fiber optic, No)

OnlineSecurity: Whether the customer has online security or not (Yes, No, No internet service)

OnlineBackup: Whether the customer has an online backup or not (Yes, No, No internet service)

DeviceProtection: Whether the customer has device protection or not (Yes, No, No internet service)

Data description 2

TechSupport: Whether the customer has tech support or not (Yes, No, No internet service)

StreamingTV: Whether the customer has streaming TV or not (Yes, No, No internet service)

StreamingMovies: Whether the customer has streaming movies or not (Yes, No, No internet service)

Contract: The contract term of the customer (Month-to-month, One year, Two years)

PaperlessBilling: The contract term of the customer (Month-to-month, One year, Two years)

PaymentMethod: The customer’s payment method (Electronic check, Mailed check, Bank transfer (automatic), Credit card (automatic))

Data description 3

Next, there are 3 numerical features:

Tenure: Number of months the customer has stayed with the company

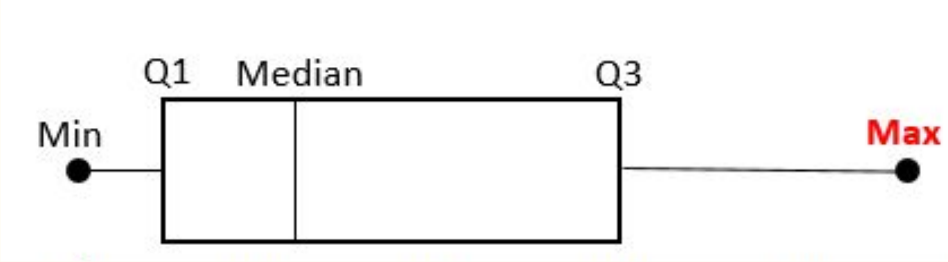
MonthlyCharges: The amount charged to the customer monthly

TotalCharges: The total amount charged to the customer

Finally, there's a prediction feature:

Churn: Whether the customer churned or not (Yes or No)

Box Plot description



A Box-plot is a quick snapshot of a set of data that helps you see where most of the data lies, how it spreads, and if there are any unusual values.

The Box: represents the middle 50% of the data. It shows where most of the data falls. The lower edge/left of the box is where the 25% (Q1) of data lies, and the upper edge/right is where the top 25%(Q3) lies. So, it gives you a sense of where the majority of data points are.

This **line inside the box** is the "middle" or median value of the data. It's like the middle point.

The **Whiskers** are the lines that extend out of the box like the arms of the plot. They show you how far the data spreads. The whiskers tell you about the range of the data—how low and high it goes.