1. **Now let's take a quick look at the relation between monthly and total charges**

We will observe that the total charges increases as the monthly bill for a customer increases.

Chart, scatter chart

Description automatically generated

1. **Finally, let's take a look at our predictor variable (Churn) and understand its interaction with other important variables as was found out in the correlation plot.**

Lets first look at the churn rate in our data

Chart, bar chart

Description automatically generated

In our data, 74% of the customers do not churn. Clearly the data is skewed as we would expect a large majority of the customers to not churn. This is important to keep in mind for our modelling as skeweness could lead to a lot of false negatives. We will see in the modelling section on how to avoid skewness in the data.

**Lets now explore the churn rate by tenure, seniority, contract type, monthly charges and total charges to see how it varies by these variables.**

1. **Churn vs Tenure**: As we can see form the below plot, the customers who do not churn, they tend to stay for a longer tenure with the telecom company.  
   **Chart, box and whisker chart

   Description automatically generated**
2. **Churn by Contract Type**: Similar to what we saw in the correlation plot, the customers who have a month to month contract have a very high churn rate.  
   **Chart, waterfall chart

   Description automatically generated**
3. **Churn by Seniority**: Senior Citizens have almost double the churn rate than younger population.  
   **Chart, waterfall chart

   Description automatically generated**
4. **Churn by Monthly Charges**: Higher % of customers churn when the monthly charges are high.  
   **Chart, histogram

   Description automatically generated**
5. **Churn by Total Charges**: It seems that there is higer churn when the total charges are lower.  
   **Chart, histogram

   Description automatically generated**

**SOLUTION :**

Random forests is a supervised learning algorithm. It can be used both for classification and regression. Random forests creates decision trees on randomly selected data samples, gets prediction from each tree and selects the best solution by means of voting. It also provides a pretty good indicator of the feature importance.

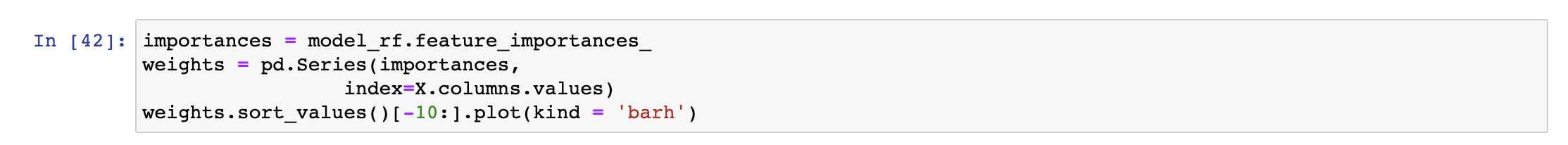
A big part of machine learning is classification — we want to know what class (a.k.a. group) an observation belongs to. The ability to precisely classify observations is extremely valuable for various business applications like predicting whether a particular user will buy a product or forecasting whether a given loan will default or not.

The problem at hand is a classification problem where we have multiple features and we need to classify if a particular customer will churn or not. So Random Forest is the go-to algorithm for classification problems like these and we will implement a Random Forest classifier as our first ML algorithm here.

**Code Snippet**

**Graphical user interface, text, application

Description automatically generated**

****

**Chart, bar chart

Description automatically generated**

**Observations:**

* From random forest algorithm, monthly contract, tenure and total charges are the most important predictor variables to predict churn.
* The results from random forest are in line to what we had expected from our EDA