**Analysis and Visualization of a Telecom Churn data set**

Most importantly, visual investigation of information is the main thing one will in general do while managing another undertaking. We do fundamental checks and examination utilizing designs and tables to sum up the information and leave out the less significant subtleties. It is considerably more advantageous for us, people, to get a handle on the primary concerns this way than by perusing many lines of crude information. It is astounding how much knowledge can be acquired from apparently basic outlines made with accessible perception devices.

Then, when we examine the exhibition of a model or report results, we additionally regularly use graphs and pictures. Some of the time, for deciphering a perplexing model, we really want to extend high-layered spaces onto all the more outwardly coherent 2D or 3D figures.

With everything taken into account, perception is a moderately quick method for discovering some new information about your information. Along these lines, it is imperative to get familiar with its most valuable strategies and make them part of your regular ML tool stash.

In this undertaking, we will get involved insight with visual investigation of information utilizing well known libraries like pandas, matplotlib and seaborn.

Reading and analyzing data set

In [3]:

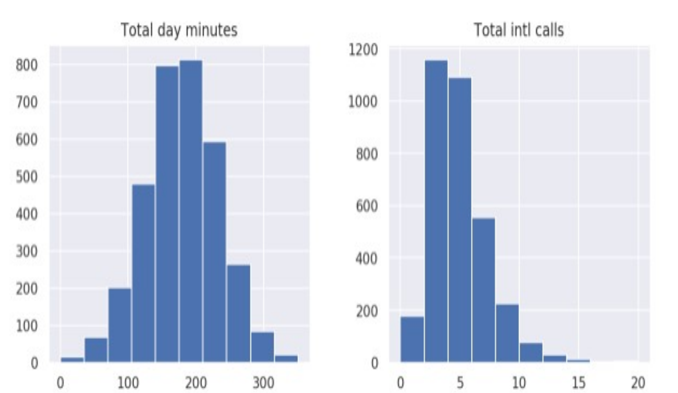
```
df.head()
```

Out[3]:

| | State | Account length | Area code | International plan | Voice mail plan | Number vmail messages | Total day minutes | Total day calls | Total day charge | Total eve minutes |
|---|---|---|---|---|---|---|---|---|---|---|
| 0 | KS | 128 | 415 | No | Yes | 25 | 265.1 | 110 | 45.07 | 197.4 |
| 1 | OH | 107 | 415 | No | Yes | 26 | 161.6 | 123 | 27.47 | 195.5 |
| 2 | NJ | 137 | 415 | No | No | 0 | 243.4 | 114 | 41.38 | 121.2 |
| 3 | OH | 84 | 408 | Yes | No | 0 | 299.4 | 71 | 50.90 | 61.9 |
| 4 | OK | 75 | 415 | Yes | No | 0 | 166.7 | 113 | 28.34 | 148.3 |

Above table shows that all the statistics of the data set with respect to summaries of each and every row

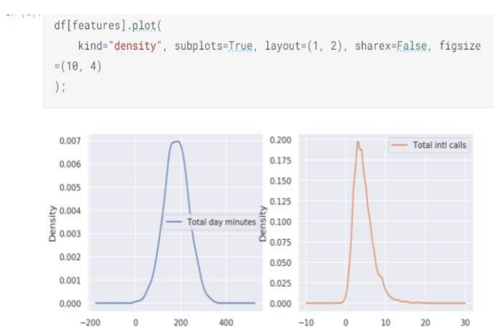
Analyzing features



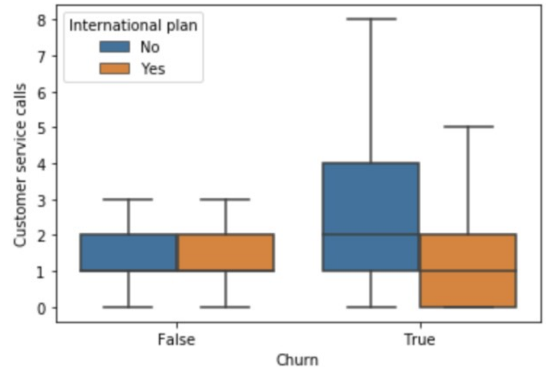
In the above plot, we see that the variable Total day minutes is typically disseminated, while Total intl calls is noticeably slanted right (its tail is longer on the right).

There is additionally another, regularly more clear, method for getting a handle on the circulation: thickness plots or, all the more officially, Kernel Density Plots. They can be viewed as a smoothed form of the histogram. Their principle advantage over the last option is that they don't rely upon the size of the receptacles. How about we make thickness plots for similar two factors:

**Features plot**




While California is the most populous state in the U.S, there are not as many customers from California in our dataset. Arizona (AZ), for example, has 64 customers, 4 of whom ended up churning. In comparison, California has a higher number (and percentage) of customers who churned. This is useful information for a company.




It looks like customers who do churn end up leaving more customer service calls unless these customers also have an international plan, in which case they leave fewer customer service calls. This type of information is really useful in better understanding the drivers of churn. It's now time to learn about how to preprocess your data prior to modelling.

I used multiple linear regression for these questions, so first, because most of them are categorical variables, so I needed to convert them into dummy variables,90%(low LTV) of spilled clients just remained under 10 months. What's more, the normal LTV of 90% of the individuals who withdrew is 750 dollars. Then again, the normal LTV of the top 20% of withdrawn is 4750 dollars. Furthermore, the proportion by the amount of absolute LTV by each gathering is 750*4: 4750 = 1: 1.6, which recommends we should zero in on serving those 20% clients with high LTV, which brought 60%(1.6/2.6) of our income from spilled clients.

Create marketing campaigns to upsell those currently subscribed to streaming movies and TV services on our other internet services.