**Multivariate Logistic Regression - Telecom Churn Example**

Let's now look at the process of building a logistic regression model in Python.

You will be looking at the **telecom churn** **prediction**example. You will use 21 variables related to customer behaviour (such as the monthly bill, internet usage etc.) to predict whether a particular customer will switch to another telecom provider or not (i.e. churn or not).

**Problem Statement**

You have a telecom firm which has collected data of all its customers. The main types of attributes are:

* Demographics (age, gender etc.)
* Services availed (internet packs purchased, special offers taken etc.)
* Expenses (amount of recharge done per month etc.)

Based on all this past information, you want to build a model which will predict whether a particular customer will churn or not, i.e. whether they will switch to a different service provider or not. So the variable of interest, i.e. the target variable here is ‘Churn’ which will tell us whether or not a particular customer has churned. It is a binary variable - 1 means that the customer has churned and 0 means the customer has not churned.

You can download the datasets here:

**[churn\_data.csv](https://cdn.upgrad.com/UpGrad/temp/9b593e11-856c-4ab9-982d-931c4548aea2/churn_data.csv" \o "churn_data.csv" \t "_blank)**

[file\_download](https://cdn.upgrad.com/UpGrad/temp/9b593e11-856c-4ab9-982d-931c4548aea2/churn_data.csv" \o "churn_data.csv" \t "_blank)**[Download](https://cdn.upgrad.com/UpGrad/temp/9b593e11-856c-4ab9-982d-931c4548aea2/churn_data.csv" \o "churn_data.csv" \t "_blank)**

**[internet\_data.csv](https://cdn.upgrad.com/UpGrad/temp/c62c0489-fc52-4b8a-a464-b1f5c94304af/internet_data.csv" \o "internet_data.csv" \t "_blank)**

[file\_download](https://cdn.upgrad.com/UpGrad/temp/c62c0489-fc52-4b8a-a464-b1f5c94304af/internet_data.csv" \o "internet_data.csv" \t "_blank)**[Download](https://cdn.upgrad.com/UpGrad/temp/c62c0489-fc52-4b8a-a464-b1f5c94304af/internet_data.csv" \o "internet_data.csv" \t "_blank)**

**[customer\_data.csv](https://cdn.upgrad.com/UpGrad/temp/f9ff4b35-a538-4910-9655-de03027f10b4/customer_data.csv" \o "customer_data.csv" \t "_blank)**

[file\_download](https://cdn.upgrad.com/UpGrad/temp/f9ff4b35-a538-4910-9655-de03027f10b4/customer_data.csv" \o "customer_data.csv" \t "_blank)**[Download](https://cdn.upgrad.com/UpGrad/temp/f9ff4b35-a538-4910-9655-de03027f10b4/customer_data.csv" \o "customer_data.csv" \t "_blank)**

Also, here’s the data dictionary:

**[Telecom Churn Data Dictionary.csv](https://cdn.upgrad.com/UpGrad/temp/b7c1276e-a890-42e5-a581-0074b72fbb20/Telecom%20Churn%20Data%20Dictionary.csv" \o "Telecom Churn Data Dictionary.csv" \t "_blank)**

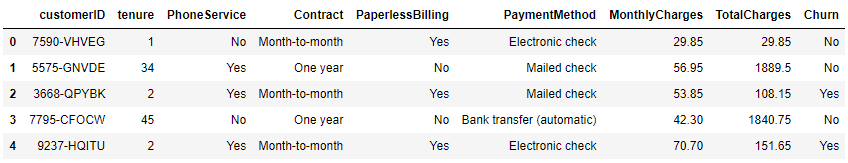
[file\_download](https://cdn.upgrad.com/UpGrad/temp/b7c1276e-a890-42e5-a581-0074b72fbb20/Telecom%20Churn%20Data%20Dictionary.csv" \o "Telecom Churn Data Dictionary.csv" \t "_blank)**[Download](https://cdn.upgrad.com/UpGrad/temp/b7c1276e-a890-42e5-a581-0074b72fbb20/Telecom%20Churn%20Data%20Dictionary.csv" \o "Telecom Churn Data Dictionary.csv" \t "_blank)**

You can also download the code file and may follow along.

**[Logistic Regression in Python - Telecom Churn Case Study](https://cdn.upgrad.com/UpGrad/temp/88e53ed7-0af8-4ee7-a543-501f6f8e50f0/Logistic+Regression+-+Telecom+Churn+Case+Study.ipynb" \o "Logistic Regression - Telecom Churn Case Study.ipynb" \t "_blank)**

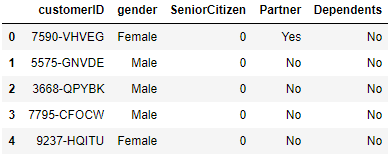
[file\_download](https://cdn.upgrad.com/UpGrad/temp/88e53ed7-0af8-4ee7-a543-501f6f8e50f0/Logistic+Regression+-+Telecom+Churn+Case+Study.ipynb" \o "Logistic Regression - Telecom Churn Case Study.ipynb" \t "_blank)**[Download](https://cdn.upgrad.com/UpGrad/temp/88e53ed7-0af8-4ee7-a543-501f6f8e50f0/Logistic+Regression+-+Telecom+Churn+Case+Study.ipynb" \o "Logistic Regression - Telecom Churn Case Study.ipynb" \t "_blank)**

So, here’s what the data frame churn\_data looks like:



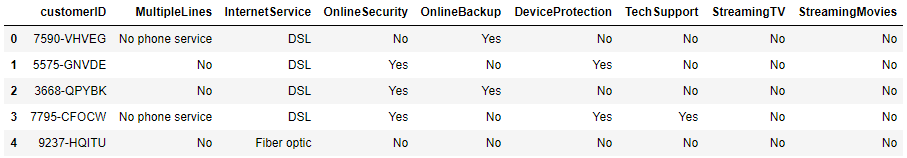
**Data Frame 1: churn\_data**

Also, here’s the data frame customer\_data:



**Data Frame 2: customer\_data**

Lastly, here’s the data frame internet\_data:



**Data Frame 3: internet\_data**

Now, as you can clearly see, the first 5 customer IDs are exactly the same for each of these data frames. Hence, using the column customer ID, you can collate or merge the data into a single data frame. We'll start with that in the next segment.

**Coming Up**

In the next segment, you will start with reading and inspecting the dataframes and then move on to preparing that data for model building.