# Conditional probability using pandas

*Special thanks to Roba, Data Analysis Session Lead for referring the dataset*

## Dataset

Telco Customer Churn | Kaggle
https://www.kaggle.com/datasets/blastchar/telco-customer-churn

Read through the dataset description and then answer the following questions. You can either
download the dataset and work locally or you can work in a Kaggle notebook

## Tasks

1- check for missing values in all columns at once
2- check for duplicates
3- Pick two categorical variables and check for typos in their entries
4- Choose two factors that may affect customer churn. Calculate the conditional probability P(Churn|factor)
5- Check if value_counts can give you proportions. If yes, can you use it for the previous calculation?
6- Can pandas crosstabs function do all previous steps in one line? Implement this **Hint:** check the function's arguments [use the shift + tab option]
7- check `Minimally sufficient pandas` for a comparison between crosstabs vs pivot table vs groupby

Minimally Sufficient Pandas
https://www.dunderdata.com/blog/minimally-sufficient-pandas

Minimally-Sufficient-Pandas Similarity between groupby, pivot_table, crosstab
https://github.com/tdpetrou/Minimally-Sufficient-Pandas/blob/master/9.%20Similarity%20between%20groupby%2C%20pivot_table%2C%20crosstab.ipynb

8- present your work