Objective: I want to find out if the customer will churn or not.

Given the data, we should first find out if the churn result is correlated to any other data points. What factors seem to influence customer churn?

Hypothesis: If the “Contract” period is longer, or “Tenure” is longer or “TotalCharges” is bigger, the churn result is likely to be negative. (People will stay with the company)

For the linear correlation, it is important to use the past behavior (train.csv) as the basis to predict the future behavior (homework 1.csv) of clients

Approach:

1. Get the distribution of each independent variables (How many people are males, senior citizens, have dependents, … in which Quartiles of TotalCharges and etc.)
2. Find the correlation between each independent variables and churn result. (Of those people, who chose to churn and who chose to stay?)
3. Get the probability of each customer who would churn or not based on how strong the correlation is and apply it to actual data (homework 1.csv)

Problems:

* Multiple factors seem to suggest correlation with Churn result. (Multivariate regression)
* How to find the probability of an event (Churn result) dependent on multiple variables that are also represented as probabilities?

\*If I multiply the probability of two independent variables that is over 50%, it should reinforce, not diminish, the likelihood of the dependent variable. (But 0.6\*0.6 = 0.36)

Conclusion:

Therefore, TelCo company should work on 1. Providing longer contract period, 2. Giving benefits to new customers and 3. Increasing the transaction value per customer.