1. What is the problem you want to solve?

Through this project, I want to identify the root causes for customer churn and predict it by machine learning.

2. Who is your client and why do they care about this problem? In other words, what will your client DO or DECIDE based on your analysis that they wouldn't have otherwise?

My client, Telco Company, is a telephone and internet service provider with over 5000 customers. In order to grow and maintain profitability, it's essential that they learn how to maintain a dedicated customer base and reduce churn. Based on my analysis, Telco can:

1. Identify customers that are likely to churn and reach out to them to try to stop them from churning via special offers targeted to their needs
2. Focus marketing on customers that are more likely to be long term customers
3. Modify their services to improve the likelihood customers will stay longer term

3. What data are you going to use for this? How will you acquire this data?

I am going to use Telco Customer Churn dataset which is provided by Kaggle. I will use the panda library to import csv file.

Attribute	**Description**
CustomerID	Customer ID
Gender	Customer gender (female, male)
SeniorCitizen	Whether the customer is a senior citizen or not (1, 0)
Partner	Whether the customer has a partner or not (Yes, No)
Dependents	Whether the customer has dependents or not (Yes, No)
Tenure	Number of months the customer has stayed with the company
PhoneService	Whether the customer has a phone service or not (Yes, No)

MultipleLines	Whether the customer has multiple lines or not (Yes, No, No phone service)
InternetService	Customer's internet service provider (DSL, Fiber optic, No)
OnlineSecurity	Whether the customer has online security or not (Yes, No, No internet service)
OnlineBackup	Whether the customer has online backup or not (Yes, No, No internet service)
DeviceProtection	Whether the customer has device protection or not (Yes, No, No internet service)
TechSupport	Whether the customer has tech support or not (Yes, No, No internet service)
StreamingTV	Whether the customer has streaming TV or not (Yes, No, No internet service)
StreamingMovies	Whether the customer has streaming movies or not (Yes, No, No internet service)
Contract	The contract term of the customer (Month-to-month, One year, Two year)
PaperlessBilling	Whether the customer has paperless billing or not (Yes, No)
PaymentMethod	The customer's payment method (Electronic check, Mailed check, Bank transfer (automatic), Credit card (automatic))
MonthlyCharges	The amount charged to the customer monthly
TotalCharges	The total amount charged to the customer
Churn	Whether the customer churned or not (Yes or No)

4. In brief, outline your approach to solving this problem (knowing that this might change later).

Data Wrangling·

1. Acquire database tables in CSV format from kaggle.
    - Connect tables
    - Convert columns with yes/ no to 0/1
    - Use get_dummy turning a column to multiple column
    - Clean data and check for missing values.
2. Construct and develop data story
    - Explore data, trends based on services and senior discount.
    - Produce histograms, plot to visualize patterns in data
    - Identify insights, correlations between columns.

Inferential Statistics:

- Identify important variables in predicting customer churn
- Run statistical tests to find significance of the relationship between these variables and customer churn
- Examine correlation between features to remove or combine variables that are too closely correlated with one another
- Implement inferential statistical analyses and report ·

Machine learning:

- Build classification model to predict customer churn Machine learning with linear regression.

Write Final Report:

- Summarize findings
- Build presentation
- Present project to springboard

5. What are your deliverables? Typically, this would include code, along with a paper and/or a slide deck.

- Github with code, and documentation
- Final paper to describe the project
- Project presentation
- Client report