Capstone Project Proposal

1. What is the problem you want to solve?

I want to predict the attrition of the company's valuable employees, Uncover the factors that lead to employee attrition and explore important questions such as 'show me a breakdown of distance from home by job role and attrition' or 'compare average monthly income by education and attrition'.

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 is IBM human resources director. He is trying to figure out the roots of employee attrition and improve the performance of company. For that, he focuses on defining the parameters which cause the employee attrition via proactive approach and tries to overcome that/those with the project's outcome.

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

I will use IBM HR Analytics Employee Attrition & Performance data from Kaggle, which is created by IBM data scientists. It has 1470 rows x 35 columns and contains numeric and categorical data types in columns.

Data Source:
https://www.kaggle.com/pavansubhasht/ibm-hr-analytics-attrition-dataset

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

I will approach this machine learning project by following the steps below:

a. Create a repository in Github (in addition, create google driver to use shareable documents with my mentor)
b. Gather the data from Kaggle and load it into Python.
c. Analyze the data to determine the data quality

d. Prepare the data:

(1) Clean that which may require it (remove duplicates, deal with missing values, correct errors, normalization, data type conversions, etc.)

(2) Randomize data, which erases the effects of the particular order

(3) Visualize data to help detect relevant relationships between variables and perform exploratory analysis.

e. Feature engineering and selection
f. Split the data as training and test data
g. Choose a machine learning algorithm
h. Train the model
i. Evaluate the model
j. Parameter Tuning
k. Make predictions.
l. Prepare a report

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

My deliverables will be a paper, a PowerPoint presentation summarizing my paper, and the code associated with my project.