Create a Google Doc and use the steps you learned in the Problem Identification unit to develop answers to each of the bullets listed below:

- Problem statement formation

Worldwide financial losses caused by credit card fraudulent activities are worth tens of billions of dollars. One American over ten has been a victim of credit card fraud (median amount of $399), according to the Statistic Brain Research Institute

Building ML-based credit card fraud detection systems generally provides fast response times and verification of large volumes of information that, if done manually, would not currently be feasible.

- Context

It is important that credit card companies are able to recognize fraudulent credit card transactions so that customers are not charged for items that they did not purchase.

- Criteria for success

Average Precision (AP)
Card Daily Accuracy (CP @ k)

The area under the ROC curve (known as AUC ROC) is the most widely used metric to evaluate the performance of fraud detection systems. However, it is poorly adapted to class imbalance problems, such as fraud detection. Two other metrics, Average Accuracy (AP)
We will use the Average Precision (AP), which summarizes such a plot as the weighted mean of precisions achieved at each threshold, with the increase in recall from the previous threshold used as the weight [BEP13,FZ11].

$$AP=\sum_{n}(R_n-R_{n-1})*P_n$$

where Pn and Rn are the precision and recall at the n-th threshold
Pn Precision=TP/(TP+FP)
Rn Recall=TP/(TP+FN)

and Card Daily Accuracy (CP @ k) motivated me to better characterize the performances of a CCFD.

$$P @ k = (1 / | D |) * \sum_{d \in D} Pk\ (d)$$

k the number of alerts that investigators can process during a day
d for a set of days d∈D

The AUC ROC, AP, and CP@k are complementary. The AUC ROC reflects the accuracy of the detection system for all possible thresholds. The AP also reflects the accuracy of the detection system for all possible thresholds but gives more importance to regions of the decision thresholds where the precision remains high. Finally, the CP@k provides a more concrete metric for CCFD, by assessing the average daily precision of the system, assuming that a maximum of k cards can be checked daily by investigators.

● Scope of solution space

I’ll focus in a Supervising CNN ML classification system

● Constraints

My computer hardware I don't have a GPU for CNN and the short time to finish the project.

● Stakeholders

My mentor Richard Ball.
My source data is from Kaggle https://www.kaggle.com/mlg-ulb/creditcardfraud
I’ll present this recommendation to Modernizing Medicine

● Data sources

https://www.kaggle.com/mlg-ulb/creditcardfraud

Here are some questions to consider to help you get started:

- 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?

- What data are you using? How will you acquire the data?

- Briefly outline how you’ll solve this problem. Your approach may change later, but this is a good first step to get you thinking about a method and solution.

- What are your deliverables? Typically, this includes code, a paper, or a slide deck.