



# Capstone 3 Project Ideas

🕒 Created	@October 16, 2023 9:36 AM
📂 Class	Capstone Three
📁 Type	Project
☑ Reviewed	<input type="checkbox"/>

## Idea 1:

Use data on credit card transactions made by European cardholders in the year 2023 to build a classification model. I'm learning towards using tree based methods again, likely XGBoost and Gradient Boost.

A few things to note about the particular dataset:

- It contains a tad over 550,000 transactions, all made in 2023.
- It contains a unique ID and 28 distinct features (like time, location, etc) that have all been **anonymized. (V1 - V28)**
  - Will this create problems, or is this ok?
- Each feature has been scaled using PCA. The only feature that has not been scaled is the amount of the purchase.

This is the idea I would most like to pursue.

My goal for capstone three is to use the skills I've built to solve a more practical and real world business problem. Credit card fraud commonly ranks among the most important ML projects any bank can undertake, and I'm certain every major bank in the world has models already in place.

Given big banks are a target of mine post graduation, I'm most drawn to this idea.

## Idea 2:

Use data on a fictional multi state bank to predict customer churn. The methods I'd plan on using and my reasoning behind this idea are similar to idea 1, to hone in one something useful and valuable to predict. Churn as opposed to fraud in this case.

## Idea 3:

Use data on bank costumers to segment them into distinct clusters, primarily using K means and hierarchical clustering. The reasoning and objectives for this idea are the same from idea 1 and 2, although it's the idea I'd least like to pursue.