**Project Objective:** (taken directly from our project proposal)

Using this dataset and focusing on the **customer complaints narrative**, explore how we can use advanced machine learning techniques to:

1. identify the key problem in the complaint;
2. and once the main problems have been identified, how we can assign each complaint to those key problems.

Steps:

1. Filtered the Dataset by “Customer Consent Obtained” = yes, because these are the records that contain the customer complaint narrative.
2. Export Dataset – Remove all the columns except for customer complaint narrative and Issue.
3. Since there are 151 different issues, and that is too many for meaningful, high level categorization, used business knowledge to map each issue to a higher level **Main Problem**. For example:
   1. “Charged fees or interest I didn't expect” and “Excessive Fees” both mapped to the **Main Problem** of “Billing or Debt Disputes”
   2. “Managing, Opening, or Closing an Account” and “Trouble During Payment Process” both mapped to the **Main Problem** of “Administrative Help”
   3. “Can’t repay my Loan” and “Loan modification, collection, foreclosure” both mapped to the **Main Problem** of “Loan or Lease Help”
4. After this process, nine **Main Problem** categories were identified:

|  |
| --- |
| Incorrect Information |
| Credit Card or Credit Reporting Help |
| Loan or Lease Help |
| Mortgage Help |
| Administrative Help |
| Identity Theft or Fraud |
| Debit Card Help |
| Billing or Debt Disputes |
| Service Disputes |

1. I then did a vlookup to match the **Main Problem** to the issue field in the original dataset.
2. Saved a copy of the csv file and in the copy, I removed the issue field, so all that is left is the **Main Problem** in column A and the complain narrative in column B.
3. From here, refer to the google collab notebook to see what was done next to clean and analyze the complaint narrative.

Ultimately, this is the idea I have about how we can use the **Main Problem** field created to teach the model how to categorize complaints:

With a portion of the data used as a training set - We’re teaching the model this:

When you see a complaint like this: \*Complaint narrative\*,

it belongs in this category: \* **Main Problem** \*

With a portion of the data used as a testing set – we will just be giving the model the complaint narratives, NOT the **Main Problem**. The model itself will assign the complaint narrative to one of the nine **Main Problem** defined above.