**ETL Financial Complaint Project Report**

By: Roberto Pinel and Christy Balaji

**Extract**

We pulled two CSV formatted data source files from data world. (https://data.world/ )

1. Bank Complaints.csv
2. Credit Card complaints.csv

**Transform**

For each of the data source file we transformed the data into five unique data frames.

1. products and sub-products
2. Complaint issues types
3. Financial companies
4. Banks complaints
5. Credit Card complaints
6. We droped duplicates and added a unique identifier column to each of the data frames.
7. We concatenated the bank and credit card data frames for products, complaint types and companies.
8. We renamed columns with spaces to include underscores and removed special characters.
9. We converted date columns from character format to datetime format.

**Load**

We loaded the five data frames into MySql database. We created individual tables using 2f database normalization to eliminate redundancy and unnecessary data.

1. Financial\_products (products and sub-products)
2. Complaint\_type (Complaint issues types)
3. Financial\_company (Financial companies)
4. Bank\_compliants (Banks complaints)
5. Credit\_card\_complaint (Credit Card complaints)