

In this project we are demonstrating creating a data pipeline and creating code to extract raw data and transforming it to something useful then loading it into a database where we could explore the data through queries.

First, we explored the data to get a better understanding of what we were working with.

Next came the transformation which involved breaking up an excel file into multiple csv tables to load them into SQL. We broke out the category and subcategory information, split them to create two separate tables. Once the four tables were created and broken into category, subcategory, contact info and campaign info. We checked each table to make sure all the data types were correct and cleaned and ready to load.

While the data was being cleaned, we also created an ERD with Quickdbd.com, which allowed us to create a code to organize our schema in SQL.

Once our database was set up, we were ready to load it in our tables utilizing SQL alchemy to connect our jupyter notebook to the SQL database. When loading our tables, we had to make sure to do it in the proper order which meant loading the Campaign table last because it has foreign keys which is what creates the relationships between the tables.

After that we did some exploratory queries to look at the data to see if there were any actionable items or things, we could use the data to help our business. We created a visual using the data which showed our breakdown of outcomes.

