For this Data Project 1, I first performed ETL through SQL and then performed ETL through Python, and finally used Python to integrate MongoDB into an ETL process – all the related folders for each performance are in the DS-2002 Data Project 1 folder. I searched for and found a sample SQL database online to use for this project and named it "humanresources".

Within SQL, I first created the original humanresources database in the "Create Data" sql file, and then loaded the data in the "Load Data" sql file. I then created the dimensional tables in the "Create Dim Data" sql and then extracted, transformed, and loaded the data for the new humanresources_dw database (specific details, thought processes, and processes are in comments in the actual files). I also included a date dimension in the "Create Populate Dim Date" sql.

I also performed ETL through Python and used Python to integrate MongoDB into ETL as well. These are all in the "Python and Python MongoDB ETL" folder in my DS-2002 Data Project 1 folder. I also have a "data" folder in this folder that contains all source data in json format that I used to populate MongoDB. For both the Python ETL and Python MongoDB ETL files, all details, thought processes, and process explanations are in the headers and comments of each process which can be referenced to for further annotation. I also perform aggregations at the end of my Python ETL file where I average the salaries of each department and look at them in descending order to see which departments have the highest salary and which departments have the lowest salary in relation to the others.