

Melanie Eppright ETL Report | 12.13.18

Initially, I imported the following:

- import pandas as pd
- from sqlalchemy import create_engine
- import pymysql
- pymysql.install_as_MySQLdb()
- from config import mypass
 - This masked my MySql password

Next, I read in my data files into pandas data frames. For the .csv I used the following:

- pd.read_csv(museum_directory_csv_file, encoding = "UTF-8")
 - Upon trying to import final pandas dataframe to MySql, I received an error stating that the csv was encoded in UTF-8. Therefore, I had to go back and add this in to the cell in which the .csv files is read.

In the next cell, I transformed my museum_directory dataframe to keep the following columns:

- Museum Name
- Museum ID
- State (Administrative Location)

Then, I renamed them as such:

- Museum Name : Musuem_Name
- Museum ID : Museum_ID
- State (Administrative Location) : State

The next step was to set the index = Museum_Name

Then, I dropped duplicates

Upon importing the json file which contained our museum reviews, I noticed that each column's title was the name of a museum and the review data (series of 5 ratings) was stored in each of the following rows.

In order to create a dataframe that was a similar shape to the csv Museum Directory data, we had to do the following:

- Find the averages of all reviews using the following code:
Museum_Reviews_df.mean(axis=0)
- Rename the columns as follows:
 - Index : Museum_Name,
 - 0 : Rating
- Then I set the index to Museum_Name

At this point, both of my data frames were relatively similar and were exported to MySql.