Group 5 ETL Project Final Report

Updated: 14-Oct

Group Name:

Group 5

Group Members:

Adrian Strecker / Jay Jackson / Mindy Ketchum / Kristan Pruett / Timothy O’Malley

Here is the final report of our ETL Project:

1. **Extract**

Our team used data from the Kaggle website for our project and extracted a series of csv files:

* 1. Michelin Restaurants:
     1. <https://www.kaggle.com/jackywang529/michelin-restaurants>
        1. one-star-michelin-restaurants.csv
        2. two-star-michelin-restaurants.csv
        3. three-star-michelin-restaurants.csv
  2. Datafiniti ‘s Business Database Reviews:
     1. <https://www.kaggle.com/datafiniti/hotel-reviews>
        1. Datafiniti\_Hotel\_Reviews.csv

1. **Transform**

Our team cleaned the csv files to filter and shape the data down to the information we needed:

1. The Hotel\_Reviews.csv file was read and converted into a database (hotel\_df), dropping out the columns deemed unnecessary.
2. A second database was created (hotel\_means) that complied the average reviewer ratings.
3. These two databases were merged together by similar states (CA) giving us our final database (hotel\_df).
4. Three additional databases were created (one\_df, two\_df, three\_df), and were used to display the data for restaurants and their respective star Michelin rating.
5. **Load**:
6. The databases were connected to a local SQL database and pgAdmin PostgreSQL was used to complete our data queries.
7. Used Jupyter Notebook to perform the queries to check for accuracy.