**GOAL: compare and contrast Hotel and Airbnb guest reviews in the US, segmented out by State.**

*#1 The sources of data that you will extract from.*

Hotels

https://www.kaggle.com/datafiniti/hotel-reviews#Datafiniti\_Hotel\_Reviews.csv

This is a list of 1,000 hotels and their reviews provided by Datafiniti's Business Database. The dataset includes hotel location, name, rating, review data, title, username, and more.

Airbnb

https://public.opendatasoft.com/explore/dataset/airbnb-listings/information/?disjunctive.host\_verifications&disjunctive.amenities&disjunctive.features

Inside Airbnb is an independent, non-commercial set of tools and data that allows you to explore how Airbnb is really being used in cities around the world.

*#2 The type of transformation needed for this data (cleaning, joining, filtering, aggregating, etc).*

- Reading database using Python Pandas.

- Ensuring delimiters and character encoding are aligned.

- Testing available data columns

- Selecting target data columns for subsequent merging: filtering, renaming, dropping.

- Normalizing reviews to match scale 1-5

GOAL: compare and contrast Hotel and Airbnb guest reviews in the US, segmented out by State.

*#3 Load: the final database, tables/collections, and why this was chosen.*

- Selected overlapping data points

Name Address, City, State, Zip, Type, Average Rating

These will let us have some hotel/Airbnb identifiers and corresponding rating.

- Loading cleaned python data into MongoDB

-Creating database connection.

-Creating database vacation

-Creating collection Accommodations

*Possible analysis drawn from curated database.*

States where Airbnb review is higher Hotel

States where Hotel review is higher than Airbnb

Top 5 states in terms of review count

Top 5 states in terms of ratings

Zip code

*#5 Creating an API app for users to easily query curated data by entering two query fields.*

* Used Flask to call on MongoDB database.
* Issues: JSON output was giving us some issues due to syntax discrepancy; we were able to adjust dictionaries for correct API output.