|  |  |
| --- | --- |
| Project Name: | US Housing Market Database |
| Team Name: | 5 Masketeers |
| Team Members: | Alanna Javier, Ali Deniz, Donna Tigrett, Elizabeth Koehler, Anand Sharan |
| Project Description: | ETL of housing data, and |
| Data Sources: | 1. Zillow – multiple Zillow .csv files, one for each type of unit  * Single Family * Condo/Co-op * 1 bed * 2 bed * 3 bed * 4 bed * 5+ bed  1. json or api to gather businesses in each zip code |
| Database Goal: | To be able to provide information on the San Francisco zip codes, including rental prices by unit type and surrounding amenities to prospective home buyers in San Francisco. |
| Extract: | 1. Pull the ZHVI home prices from each Zillow csv in SF files 2. Pull the different businesses in each SF zip code using the json or API |
| Transform: | 1. Aggregate the different home types into one table, and add a column for home type 🡪 This will allow us to apply filters based on home type 2. Add a column to show whether the home prices are increasing or decreasing for that zip code in relation to the previous month. 3. Json or api business data |
| Database: | Mongo DB to further explore the capabilities of MongoDB. |

Brief Schemata outline:

A few tables of housing and amenities data:

* Some tables will hold housing data
* Some tables will hold business data
* Primary key will be the zip code