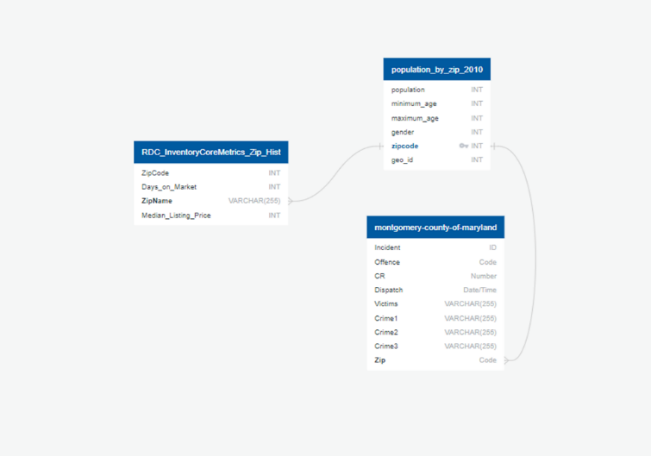
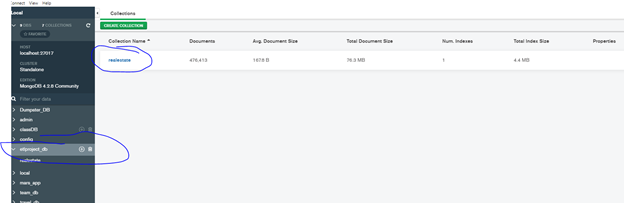
ETL Project – Overall Process

1. **Project Proposal**
   1. Gathering 3 complex datasets that involved real estate, crime and populations.
2. **Data Sources** 
   1. We utilized 3 datasets from the following resources that gave use the opportunity to discover our findings.
      1. Kaggle.com – Population by Zip
         1. <https://www.kaggle.com/census/us-population-by-zip-code>
      2. Kaggle.com - US Real Estate Listings by Zip Code
         1. [https://www.kaggle.com/waragones/us-real-estate- listings-by-zip-code](https://www.kaggle.com/waragones/us-real-estate-%20listings-by-zip-code)
      3. Data world.com - Crime Dataset
         1. <https://data.world/montgomery-county-of-maryland/56c7bd9b-34f4-40be-990b-efd13eb1a6d2/workspace/file?filename=comma-separated-values-file-1.csv>
3. **Data Cleanup & Analysis**
   1. Extract:
      1. The original data was imported from the Kaggle.com website.
      2. We used Pandas to import a file called: population\_by\_zip\_2010 and made it a DataFrame.
      3. We used Pandas to import a file called: RDC\_InventoryCoreMetrics\_Zip\_Hist and made it a DataFrame.
      4. We used Pandas to import a file called: comma-separated-values-file-1.csv
      5. 3 datasets were in zipped form as a csv.
   2. Transform:
      1. We used Pandas to Transform the data
      2. For the population file
         1. We dropped columns, geo\_id and renamed zip code columns
         2. We created a DataFrame
         3. Converted the DataFrame into a list of dictionaries
      3. For the Inventory file,
         1. We dropped several columns and only kept the columns we needed
         2. We did a group by to keep consolidate the zip codes
      4. Consolidated files
         1. We merged the data and grouped by zip code
         2. We got a count of the data
   3. Load:
      1. We created a connection to our mongo DB (relational)
      2. Defined the collection
      3. We pushed the records into Mogo DB.
      4. The data is uploaded to GitHub.
4. **Visualizations**

****

1. **MondoDB**
   1. **Screen shot of Database**



* 1. **Screen Shot of Tables**

