Data Management - Dataset Selection
Sravya Bhaskara - Section 2

Source of the dataset:

- I have chosen my dataset from kaggle, Zillow Prize: Zillow's Home Value Prediction (Zestimate), which contains real estate property information from 2017
- This dataset contains information on various property characteristics like tax details, building features, and regional information. My goal is to analyze this data to understand property trends, assess regional property values, and explore patterns in property characteristics across different areas.

Data File:

- There is a single data file (properties_2017.csv) with around 2M rows and 58 columns. (I am thinking to trim number of rows and will drop few columns)

Number of Tables:

I am considering 6 tables right now:

1. Property Table - This table includes basic property information, such as geographic location, lot size, and the year built (fields include parcelid, latitude, longitude, buildingqualitytypeid, squarefeet, yearbuilt, etc..)
2. Building Table - This table provides detailed building characteristics for each property, (fields include parcelid, bathroomcnt, bedroomcnt, garagecarcnt, airconditioningtypeid, heatingorsystemtypeid, and others).
3. Pool Table - This table lists details about pools on properties, including pool count and pool size (fields include parcelid, poolcnt, poolsizesum, and pooltypeid).
4. Region Table - This table includes regional information such as the city, county, and ZIP code associated with each property (fields include parcelid, regionidcity, regionidcounty, and regionidzip).
5. Price Table - This table provides price and tax-related information for each property, including assessed value, land and structure tax value, and total tax amount (fields include parcelid, taxvaluedollarcnt, structuretaxvaluedollarcnt, landtaxvaluedollarcnt, taxamount, and assessmentyear).
6. Catalog Table - This table contains information on description of each of the types or ids in the above mentioned tables. Some of them are heatingorsystemtypeid, buildingqualitytypeid, airconditioningtypeid.
7. One more table - Yard Table - Has information about backyard (Optional)