**Extract, Transform and Load the Final:**

**Extract**

All three CSV files were read and loaded into a Dataframe.

* Source for home\_price\_final.csv: Kaggle: Indiana - Property Sales Disclosure.
* Source for average\_indiana\_weather.csv: US Climate Data <https://www.usclimatedata.com/climate/indiana/united-states/3184>
* Source for IN\_census\_file.csv: US Census Data [www.census.gov](http://www.census.gov)

**Transform**

* **For ERD**
* CSV files had some random empty columns and gave error “Column violates ‘Not Null’ constraint” while trying to create ERD diagram and create Tables.
* CSV files after filling empty columns to replace with new address where address was null, etc.
  + The home\_price\_final.csv was renamed to home\_price\_final\_ERD.csv
  + The average\_indiana\_weather.csv was renamed to average\_indiana\_weather\_ERD.csv
  + The IN\_census\_file.csv was renamed to IN\_census\_file\_ERD.csv
* **For Ipynb**
* Column names were made pretty by removing the space and brackets by using .str.replace(' ', '\_') ('(', '') (')', '') function
* Columns with NaN were replaced with key value pair

**Load**

* Created connection to PostgreSql
* Final DB named HomeSales\_DB is loaded to PostgreSQL
* All the three DFs were converted to SQL
* Finally Queried to check if the DB has all the tables.
* Final **E**xtract**T**ransform**L**oad work is shown in: Final\_ETL.ipynb

**ERD Notes**

* ERD diagram is saved as a PNG file QuickDBD-export.png
* DB Table creation script is stored in HomeSales\_DB.sql