

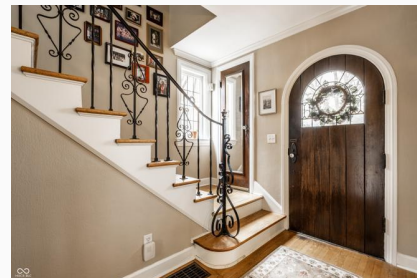
Real Estate Market Research App Demo



by: Ronald Daley

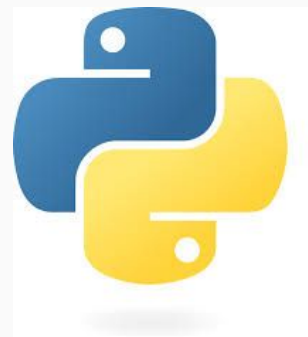
Introduction

- My Background
 - Passionate about real estate investing
 - Worked in real estate industry for 4 years
- Project Motivation
 - Develop a natural language (NL) tool that allows investors to interact with real estate-related databases to make data driven decisions when investing in a new market.
- Project Goal
 - The tool will convert questions from investors into postgresSQL queries for efficient data retrieval, analysis, and modification.



Data

- Zillow Housing Data tables
 - House value
 - For Sale Inventory
 - New Listing Inventory
 - Median List Price
- Assumptions
 - Data Type = Time series (ALL TABLES)
 - Frequency = Monthly
 - Geography = Metro level (City, State)
 - Duration = 5 years of data:
 - start = '2020-02-28'
 - end = '2025-02-28'
- Data prep and EDA in Python



Sample data: Home Value table

| | RegionID | 2020-02-29 | 2020-03-31 | 2020-04-30 | 2020-05-31 | 2020-06-30 | 2020-07-31 | 2020-08-31 | 2020-09-30 | 2020-10-31 | ... | 2024-07-31 | 2024-08-31 | 2024-09-30 | 2024-10-31 | 2024-11-30 | 2024-12-31 |
|---|----------|--------------|--------------|--------------|--------------|--------------|--------------|--------------|--------------|--------------|-----|--------------|--------------|--------------|--------------|--------------|--------------|
| 0 | 102001 | 4.710990e+05 | 4.734481e+05 | 4.753970e+05 | 4.762181e+05 | 4.763861e+05 | 4.770996e+05 | 4.797424e+05 | 4.847230e+05 | 4.914446e+05 | ... | 6.930996e+05 | 6.941272e+05 | 6.957628e+05 | 6.978330e+05 | 6.998435e+05 | 7.013544e+05 |
| 1 | 394913 | 7.417605e+05 | 7.429928e+05 | 7.441946e+05 | 7.442005e+05 | 7.434160e+05 | 7.432395e+05 | 7.457132e+05 | 7.515983e+05 | 7.604155e+05 | ... | 1.102751e+06 | 1.108185e+06 | 1.113485e+06 | 1.118237e+06 | 1.121632e+06 | 1.123018e+06 |
| 2 | 753899 | 1.129119e+06 | 1.131086e+06 | 1.132641e+06 | 1.129972e+06 | 1.123791e+06 | 1.121600e+06 | 1.128527e+06 | 1.145479e+06 | 1.166966e+06 | ... | 1.729674e+06 | 1.738847e+06 | 1.750788e+06 | 1.761684e+06 | 1.770720e+06 | 1.777742e+06 |
| 3 | 394463 | 4.267689e+05 | 4.283210e+05 | 4.296867e+05 | 4.299479e+05 | 4.291088e+05 | 4.284059e+05 | 4.287944e+05 | 4.313542e+05 | 4.358986e+05 | ... | 5.897441e+05 | 5.916298e+05 | 5.940620e+05 | 5.967046e+05 | 5.994185e+05 | 6.015685e+05 |
| 4 | 394514 | 4.487976e+05 | 4.509599e+05 | 4.520351e+05 | 4.519391e+05 | 4.514519e+05 | 4.516843e+05 | 4.536836e+05 | 4.577002e+05 | 4.631510e+05 | ... | 7.124083e+05 | 7.112924e+05 | 7.113514e+05 | 7.125698e+05 | 7.139178e+05 | 7.141215e+05 |

5 rows x 64 columns

| 4-08-31 | 2024-09-30 | 2024-10-31 | 2024-11-30 | 2024-12-31 | 2025-01-31 | 2025-02-28 | City | State |
|---------|--------------|--------------|--------------|--------------|--------------|--------------|---------------|-------|
| 272e+05 | 6.957628e+05 | 6.978330e+05 | 6.998435e+05 | 7.013544e+05 | 7.023234e+05 | 7.027477e+05 | United States | NaN |
| 185e+06 | 1.113485e+06 | 1.118237e+06 | 1.121632e+06 | 1.123018e+06 | 1.124979e+06 | 1.128410e+06 | New York | NY |
| 847e+06 | 1.750788e+06 | 1.761684e+06 | 1.770720e+06 | 1.777742e+06 | 1.778364e+06 | 1.775890e+06 | Los Angeles | CA |
| 298e+05 | 5.940620e+05 | 5.967046e+05 | 5.994185e+05 | 6.015685e+05 | 6.039141e+05 | 6.060862e+05 | Chicago | IL |
| 924e+05 | 7.113514e+05 | 7.125698e+05 | 7.139178e+05 | 7.141215e+05 | 7.132128e+05 | 7.112661e+05 | Dallas | TX |

Data Insert Process

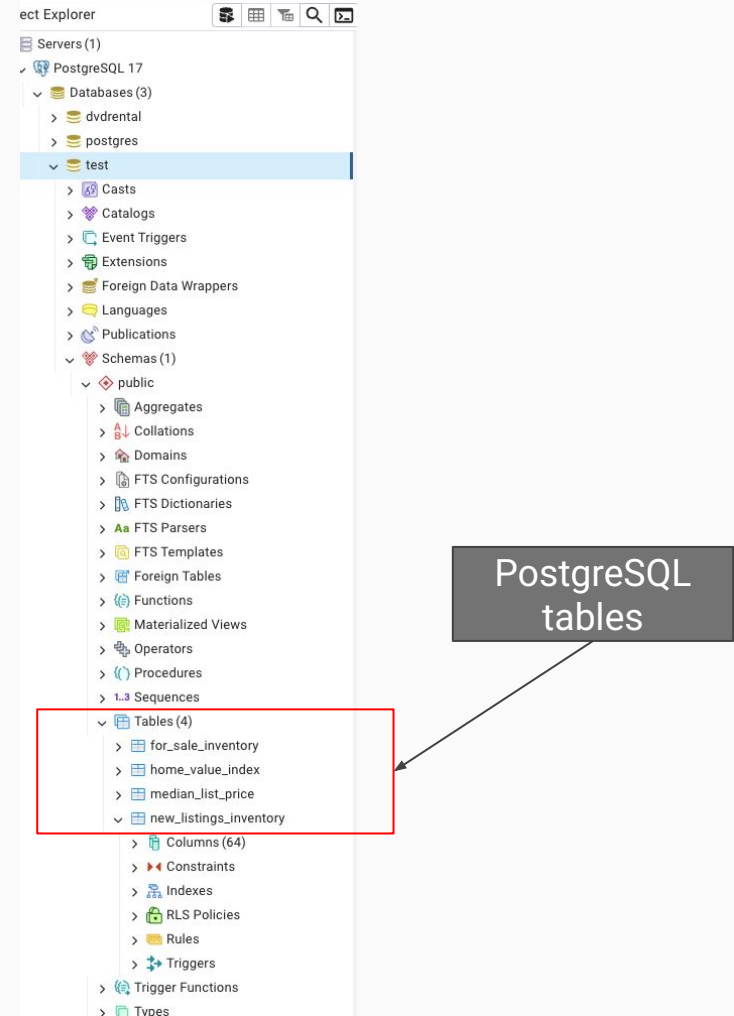
- Download data from Zillow
- Clean data and create Dataframe tables via Python
- Insert data from DataFrames into Postgres tables via Python

df1_to_table: home value index table

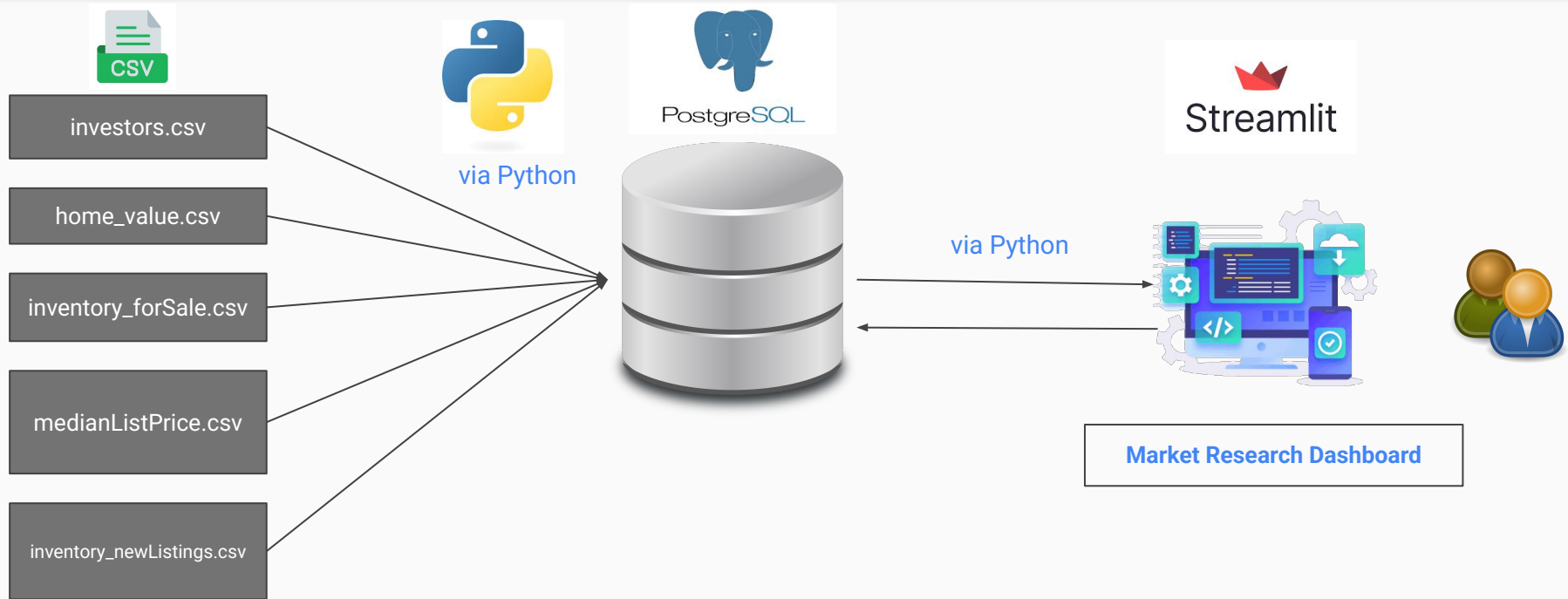
```
1 # map pandas dtype to Postgres types
2
3 type_map = {
4     "int64": "BIGINT",
5     "float64": "DOUBLE PRECISION",
6     "object": "TEXT",
7     "datetime64[ns]": "TIMESTAMP"
8 }
9
10 columns = []
11
12 for col, dtype in zip(df1_to_table.columns, df1_to_table.dtypes):
13     pg_type = type_map.get(str(dtype), "TEXT")
14     columns.append(
15         sql.SQL("{}-{}").format(
16             sql.Identifier(col),
17             sql.SQL(pg_type)
18         )
19     )
20
21 create_table = sql.SQL("CREATE TABLE {}-{}").format(
22     sql.Identifier("home_value_index"),
23     sql.SQL(",").join(columns)
24 )
25
26 cur.execute(create_table)
27
28 # Bulk insert into table
29
30 insert_stmt = sql.SQL("INSERT INTO {}-{} VALUES %s").format(
31     sql.Identifier("home_value_index"),
32     sql.SQL(",").join(map(sql.Identifier, df1_to_table.columns))
33 )
34
35 # convert Nan to Null
36
37 data = [
38     tuple(None if pd.isna(val) else val for val in row)
39     for row in df1_to_table.itertuples(index = False, name = None)
40 ]
41
42 "
```

RDBMS Database

- PostgreSQL
- PgAdmin



Architecture Design



Home Affordability Dashboard

- Streamlit app
- Dashboard Functionality:
 - Explore Tables
 - Pre-built Sample Queries
 - Query Execution (NL→SQL)
 - Insert, Create, Delete into database via NL

Market Research Dashboard

Real Estate Market Insights

Explore Tables

Choose table

home_value_index

Sample rows

5

1

1000

Sample Data

| | RegionID | 2020-02-29 | 2020-03-31 | 2020-04-30 | 2020-05-31 | 2020-06-30 | 2020-07-31 | 2020-08-31 | 2020-09-30 | 2020-10-31 | 2020-11-30 | 2020-12-31 | 2021-01-31 | 2021-02-28 | 2021-03-31 | 2021-04-30 | 2021-05-31 | 2021-06-30 |
|---|----------|--------------|--------------|-------------|--------------|--------------|--------------|--------------|--------------|--------------|-------------|--------------|--------------|--------------|--------------|--------------|--------------|-------------|
| 0 | 102001 | 471099.0476 | 473448.1 | 475397.0273 | 476218.1375 | 476386.1414 | 477099.6136 | 479742.4251 | 484723.0063 | 491444.6073 | 499370.8814 | 507581.9296 | 515710.5959 | 524154.1069 | 532898.6097 | 542596.0724 | 553336.0539 | 564283.4739 |
| 1 | 394913 | 741760.5052 | 742992.788 | 744194.64 | 744200.4914 | 743416.0214 | 743239.452 | 745713.2349 | 751598.2576 | 760415.4896 | 771998.5285 | 785608.0578 | 798783.4547 | 812209.2624 | 824364.1418 | 836809.4588 | 848905.6694 | 860134.6617 |
| 2 | 753899 | 1129118.8766 | 1131085.8841 | 1132641.048 | 1129972.0607 | 1123790.8945 | 1121599.7408 | 1128526.5267 | 1145478.5108 | 1166966.3962 | 1191562.328 | 1215427.6426 | 1232332.4634 | 1242659.8568 | 1248974.1792 | 1261671.6508 | 1281473.3356 | 1307256.88 |
| 3 | 394514 | 448797.5688 | 450959.8523 | 452035.0822 | 451939.0624 | 451451.8842 | 451684.261 | 453683.6057 | 457700.2225 | 463150.9923 | 469526.9203 | 476131.7598 | 483327.7063 | 491169.8139 | 500204.4814 | 511086.0988 | 524518.6527 | 538856.7629 |
| 4 | 394692 | 392183.6528 | 393124.4053 | 393314.977 | 392957.6394 | 392526.101 | 392978.4198 | 394695.1751 | 397685.5424 | 401576.0018 | 406246.6351 | 411227.6233 | 416612.0883 | 422303.3845 | 428175.6643 | 434547.6273 | 441975.5062 | 449975.6499 |

DEMO

(v2app.py)

Conclusion

- Developed a **relational database using PostgreSQL** to efficiently store and query time-series real estate market data, facilitating structured and scalable analysis.
- Created an interactive **market research dashboard that utilizes large language models** to answer questions and provide prospective investors with data-driven insights and customizable queries.



Next Steps

- Add more data variety (i.e. income, population, occupation, demographic statistics, crime rate, % of renters)
- Add more visualizations (interactive maps, charts, etc..)
- Improve Overall User Experience of dashboard
 - Migrate to Tableau

Thank you!

Appendix

Demo Notes

- Update query execution
- Include a JOIN clause sql
- IMPORTANT: Make sure ReadMe includes descriptive steps on implementation
- For the database, submit the .tar file?
- Remove API keys