DEPI Real Estate

رواد مصر الرقمية

Team Member

Mohamed Mohsen

Mohamed Fo’ad

Abdelrahman Mohamed

Zyad Mohamed

Ahmed ElSayed

رواد مصر الرقمية

Table of contents

رواد مصر الرقمية

01

Data Collection

رواد مصر الرقمية

Zip_Code	No_Beds	No_Baths	City	State	House_Size_s_feet	Sold_Date	Price	No_Bedrooms	Size_Status	Year
6066	3	2	Vernon	Connecticut	1850	Monday, November 9, 2015	$299,900	2	Small House	2015
6040	3	2	Manchester	Connecticut	1828	Monday, July 31, 2000	$349,900	2	Small House	2000
6029	3	2	Ellington	Connecticut	1174	Wednesday, July 9, 2008	$248,900	2	Small House	2008
6098	3	2	Winchester	Connecticut	1248	Wednesday, March 17, 2004	$59,900	2	Small House	2004
6018	3	2	North Canaan	Connecticut	1620	Tuesday, August 23, 2011	$299,000	2	Small House	2011
6063	3	2	Barkhamsted	Connecticut	1584	Wednesday, July 24, 2002	$389,000	2	Small House	2002
6098	3	2	Winchester	Connecticut	1346	Monday, July 28, 2008	$225,000	2	Small House	2008
6096	3	2	Windsor Locks	Connecticut	1440	Wednesday, February 25, 1998	$179,900	2	Small House	1998
6026	3	2	East Granby	Connecticut	1470	Tuesday, September 1, 1998	$254,900	2	Small House	1998
6081	3	2	Simsbury	Connecticut	1416	Monday, February 25, 2019	$219,900	2	Small House	2019
6070	3	2	Simsbury	Connecticut	1260	Wednesday, April 1, 2015	$275,000	2	Small House	2015
6082	3	2	Enfield	Connecticut	1504	Thursday, January 16, 2003	$224,900	2	Small House	2003
6035	3	2	Granby	Connecticut	1574	Monday, November 22, 1999	$324,900	2	Small House	1999
6082	3	2	Enfield	Connecticut	1587	Wednesday, July 1, 2015	$274,900	2	Small House	2015
6082	3	2	Enfield	Connecticut	1479	Tuesday, July 12, 2016	$290,000	2	Small House	2016
6082	3	2	Enfield	Connecticut	1428	Monday, July 29, 2019	$249,900	2	Small House	2019
6082	3	2	Enfield	Connecticut	1008	Tuesday, November 24, 2009	$290,000	2	Small House	2009
6082	3	2	Enfield	Connecticut	1480	Friday, April 1, 2016	$339,999	2	Small House	2016
6082	3	2	Enfield	Connecticut	1456	Thursday, August 27, 2020	$295,000	2	Small House	2020
6026	3	2	East Granby	Connecticut	1470	Tuesday, April 30, 2019	$279,900	2	Small House	2019
6082	3	2	Enfield	Connecticut	1200	Monday, December 4, 1989	$324,900	2	Small House	1989
6098	3	2	Winchester	Connecticut	1606	Friday, October 27, 2017	$349,900	2	Small House	2017
6026	3	2	East Granby	Connecticut	1512	Thursday, July 28, 2016	$295,000	2	Small House	2016
6098	3	2	Winchester	Connecticut	1654	Friday, December 18, 2009	$169,000	2	Small House	2009
6078	3	2	Suffield	Connecticut	1864	Wednesday, December 1, 2021	$359,800	2	Small House	2021

02

Explore Data with Python

1.1 Import And Read Data

[1]:
```
import pandas as pd
import numpy as np
import matplotlib.pyplot as plt
import seaborn as sns
import sweetviz as sv
import warnings
warnings.filterwarnings('ignore')
plt.style.use('ggplot')
```

[2]:
```
df=pd.read_csv("realtor-data.csv")
```

[6]:
```
df.shape
```

[6]: (407890, 11)

[7]:
```
df.describe()
```

[7]:

	bed	bath	acre_lot	zip_code	house_size	price	bedr
count	320108.000000	321618.000000	331873.000000	407693.000000	3.243650e+05	4.078900e+05	407890.00
mean	3.500200	2.566545	17.418487	3299.396838	2.222783e+03	6.758307e+05	1.57
std	2.320135	2.391618	931.723094	2222.641467	3.333098e+03	1.178266e+06	1.32
min	1.000000	1.000000	0.000000	601.000000	1.000000e+02	1.000000e+00	0.00
25%	2.000000	2.000000	0.200000	1890.000000	1.206000e+03	1.999000e+05	1.00
50%	3.000000	2.000000	0.560000	2822.000000	1.767000e+03	3.979000e+05	2.00
75%	4.000000	3.000000	2.200000	4630.000000	2.640000e+03	7.090000e+05	2.00
max	99.000000	198.000000	100000.000000	99999.000000	1.450112e+06	6.000000e+07	50.00

[8]:
```
df.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 407890 entries, 0 to 407889
Data columns (total 11 columns):
 #   Column          Non-Null Count   Dtype
---  ------          --------------   -----
 0   status          407890 non-null  object
 1   bed             320108 non-null  float64
 2   bath            321618 non-null  float64
 3   acre_lot        331873 non-null  float64
 4   city            407838 non-null  object
 5   state           407890 non-null  object
 6   zip_code        407693 non-null  float64
 7   house_size      324365 non-null  float64
 8   prev_sold_date  140950 non-null  object
 9   price           407890 non-null  int64
 10  bedroom         407890 non-null  int64
dtypes: float64(5), int64(2), object(4)
memory usage: 34.2+ MB
```

[10]:
```
df.isnull().sum()
```

[10]:
```
status              0
bed             87782
bath            86272
acre_lot        76017
city               52
state               0
zip_code          197
house_size      83525
prev_sold_date  266940
price               0
bedroom             0
dtype: int64
```

[11]:
```
df.duplicated().sum()
```

[11]: 371626

03

Data Cleaning

1.3 Preparing Data

[12]:
```
df.columns
```

[12]:
```
Index(['status', 'bed', 'bath', 'acre_lot', 'city', 'state', 'zip_code',
       'house_size', 'prev_sold_date', 'price', 'bedroom'],
      dtype='object')
```

[13]:
```
df=df[[#'status',
    'bed', 'bath',
    #'acre_lot',
    'city', 'state', 'zip_code',
    'house_size', 'prev_sold_date', 'price','bedroom']].copy()
```

[14]:
```
new_order=['zip_code','bed','bath','bedroom','city','state','house_size','prev_sold_date','price']
```

[15]:
```
df=df[new_order]
```

[16]:
```
df.rename(columns={'prev_sold_date':'Sold_Date','zip_code':'Zip_Code','bed':'No.Beds','bath':'No.Baths','city':'City','state':'State','house_size':'House
```

[17]:
```
df.head(1)
```

[17]:

	Zip_Code	No.Beds	No.Baths	No.Bedrooms	City	State	House_Size(s feet)	Sold_Date	Price
0	601.0	3.0	2.0	2	Adjuntas	Puerto Rico	920.0	NaN	105000

1.3.1 Dealing With Nulls

[18]:
```
df.query('Zip_Code.isnull()')
```

[18]:

	Zip_Code	No.Beds	No.Baths	No.Bedrooms	City	State	House_Size(s feet)	Sold_Date	Price
679	NaN	3.0	1.0	2	Ponce	Puerto Rico	790.0	NaN	58000
917	NaN	3.0	1.0	2	Ponce	Puerto Rico	790.0	NaN	58000
1022	NaN	3.0	1.0	2	Ponce	Puerto Rico	790.0	NaN	58000
1195	NaN	NaN	NaN	0	Rincon	Puerto Rico	NaN	NaN	650000
1451	NaN	3.0	1.0	2	Ponce	Puerto Rico	790.0	NaN	58000
...	...	...	...	...	...	...	...	...	...
72823	NaN	NaN	NaN	0	Phillipston	Massachusetts	NaN	NaN	27000
76868	NaN	NaN	NaN	0	Phillipston	Massachusetts	NaN	NaN	27000
87708	NaN	NaN	NaN	0	Phillipston	Massachusetts	NaN	NaN	27000
251103	NaN	NaN	NaN	0	Phillipston	Massachusetts	NaN	NaN	27000
267827	NaN	NaN	NaN	0	Phillipston	Massachusetts	NaN	NaN	27000

197 rows × 9 columns

[19]:
```
df.dropna(subset=['Zip_Code','No.Beds','No.Baths','City','House_Size(s feet)','Sold_Date'],inplace=True)
```

[20]:
```
df.isnull().sum()
```

[20]:
```
Zip_Code              0
No.Beds               0
No.Baths              0
No.Bedrooms           0
City                  0
State                 0
House_Size(s feet)    0
Sold_Date             0
Price                 0
dtype: int64
```

[22]:
```
df.shape
```

[22]:
```
(133333, 9)
```

1.3.2 Dealing With Duplicates

[22]:
```
df.loc[df.duplicated()]
```

[22]:

	Zip_Code	No.Beds	No.Baths	No.Bedrooms	City	State	House_Size(s feet)	Sold_Date	Price
745	729.0	3.0	2.0	2	Canovanas	Puerto Rico	1200.0	2/28/2020	76900
802	729.0	3.0	2.0	2	Canovanas	Puerto Rico	1200.0	2/28/2020	76900
932	729.0	3.0	2.0	2	Canovanas	Puerto Rico	1200.0	2/28/2020	76900
1000	729.0	3.0	2.0	2	Canovanas	Puerto Rico	1200.0	2/28/2020	76900
1031	729.0	3.0	2.0	2	Canovanas	Puerto Rico	1200.0	2/28/2020	76900
...	...	...	...	...	...	...	...	...	...
407885	6226.0	4.0	1.0	2	Windham	Connecticut	1254.0	12/21/2018	197000
407886	6118.0	2.0	1.0	1	East Hartford	Connecticut	624.0	12/13/2017	150000
407887	6109.0	4.0	1.0	2	Wethersfield	Connecticut	1238.0	9/29/2000	199900

[23]:
```
df.query('Zip_Code==6109.0')
```

[23]:

	Zip_Code	No.Beds	No.Baths	No.Bedrooms	City	State	House_Size(s feet)	Sold_Date	Price
54128	6109.0	4.0	1.0	2	Wethersfield	Connecticut	1238.0	9/29/2000	199900
54135	6109.0	2.0	2.0	1	Wethersfield	Connecticut	1277.0	4/7/1987	222000
59251	6109.0	4.0	4.0	2	Wethersfield	Connecticut	5117.0	5/26/1978	439000
59541	6109.0	3.0	2.0	2	Wethersfield	Connecticut	1652.0	9/12/1983	179900
59641	6109.0	1.0	1.0	1	Wethersfield	Connecticut	643.0	11/8/2011	95000
...	...	...	...	...	...	...	...	...	...
407789	6109.0	2.0	1.0	1	Wethersfield	Connecticut	792.0	8/30/2019	199000
407799	6109.0	3.0	2.0	2	Wethersfield	Connecticut	1416.0	8/16/2016	250000
407835	6109.0	5.0	5.0	3	Wethersfield	Connecticut	3020.0	8/17/1994	539900

[24]:
```
df=df.loc[~df.duplicated(subset=['Zip_Code','House_Size(s feet)','Sold_Date','Price','No.Beds','No.Baths','City','State'])]
```

[25]:
```
df.duplicated().sum()
```

[25]:
```
0
```

[26]:
```
df.shape
```

[26]:
```
(11676, 9)
```

Explore Data after Cleaning

[33]:
```
def calculate_upper_limit(column):
    upper_limit = column.mean() + 3 * column.std()
    return upper_limit
```

[34]:
```
def calculate_lower_limit(column):
    lower_limit = column.mean() - 3 * column.std()
    return lower_limit
```

[35]:
```
upper_limit = calculate_upper_limit(df['House_Size(s feet)'])
lower_limit = calculate_lower_limit(df['House_Size(s feet)'])
print(f'Upper Limit: {upper_limit}\n Lower Limit: {lower_limit}')
```

```
Upper Limit: 7117.794714075286
 Lower Limit: -2969.5903632702157
```

04

Explore Data with SQL

```sql
--AVERGAE PRICE BY ZIP_CODE--

SELECT [Zip_Code] , AVG([Price]) AS Avg_Price  FROM [dbo].[R_Estatee]
GROUP BY [Zip_Code] ORDER BY [Zip_Code];

--TOTAL NUMBER OF SALES  PER YEAR

SELECT [Year] , COUNT(*) AS TOTAL_SALES FROM [dbo].[R_Estatee]
GROUP BY [Year] ORDER BY [Year];
```

100 %

Results | Messages

	Zip_Code	Avg_Price
1	725	159000.00
2	729	76900.00
3	802	3924500.00
4	926	1100000.00
5	949	91700.00
6	1001	269944.3333
7	1002	555700.00
8	1005	409675.00
9	1007	296824.875
10	1008	770814.00
11	1010	438425.00
12	1011	115900.00
13	1013	226716.6666
14	1020	265177.6428
15	1022	175325.00
16	1026	152000.00
17	1027	351622.2222
18	1028	390918.7692
19	1030	427425.00

```sql
--TOTAL NUMBER OF SALES  PER YEAR

SELECT [Year] , COUNT(*) AS TOTAL_SALES FROM [dbo].[R_Estatee]
GROUP BY [Year] ORDER BY [Year] DESC;
```

100 %

Results | Messages

	Year	TOTAL_SALES
1	2022	151
2	2021	738
3	2020	580
4	2019	653
5	2018	595
6	2017	474
7	2016	449
8	2015	379
9	2014	294
10	2013	286
11	2012	285
12	2011	205
13	2010	246
14	2009	390
15	2008	394
16	2007	477
17	2006	543
18	2005	566
19	2004	495

Query executed successfully.

```sql
--AVERGAE PRICE BY HOUSE SIZE--

SELECT [Size_Status],AVG([Price]) AS AVG_PRICE_BY_SIZE
FROM [dbo].[R_Estatee] GROUP BY[Size_Status]
ORDER BY AVG_PRICE_BY_SIZE DESC ;

--TOP 10 MOST EXPENSIVE HOUSES--

SELECT [Zip_Code],[No_Beds],[No_Baths],[City],[State],[House_Size_s_feet],[Sold_Date],
[Price],[No_Bedrooms],[Size_Status] FROM [dbo].[R_Estatee]
ORDER BY[Price] DESC ;
```

100 %

Results | Messages

	Size_Status	AVG_PRICE_BY_SIZE
1	Large House	4373477.9702
2	Medium House	915957.5488
3	Small House	382028.2094

```sql
--TOP 10 MOST EXPENSIVE HOUSES--

SELECT Top(10) [Zip_Code],[No_Beds],[No_Baths],[City],[State],[House_Size_s_feet],[Sold_Date],
[Price],[No_Bedrooms],[Size_Status] FROM [dbo].[R_Estatee]
ORDER BY[Price] DESC ;
```

100 %

Results | Messages

	Zip_Code	No_Beds	No_Baths	City	State	House_Size_s_feet	Sold_Date	Price	No_Bedrooms	Size_Status
1	3862	6	9	North Hampton	New Hampshire	9443	2021-08-31	17500000.00	3	Large House
2	1259	3	5	New Marlborough	Massachusetts	4273	2011-01-28	15500000.00	2	Medium House
3	2138	7	7	Cambridge	Massachusetts	5709	1982-10-22	15300000.00	4	Medium House
4	2535	6	6	Chilmark	Massachusetts	4706	2005-10-04	15000000.00	3	Medium House
5	2540	4	6	Falmouth	Massachusetts	4244	2008-01-15	14500000.00	2	Medium House
6	2138	7	11	Cambridge	Massachusetts	9080	1992-09-25	14000000.00	4	Large House
7	2633	5	7	Chatham	Massachusetts	5079	1995-09-21	13600000.00	3	Medium House
8	1944	4	9	Manchester	Massachusetts	8122	2006-10-02	12900000.00	2	Large House
9	2554	8	8	Nantucket	Massachusetts	5085	2008-01-09	12750000.00	4	Medium House
10	2554	5	9	Nantucket	Massachusetts	8333	2005-12-21	12450000.00	3	Large House

```sql
--PRICE TREND OVER THE YEARS--

SELECT [Year],AVG([Price]) AS AVG_PRICES_OVER_YEARS FROM [dbo].[R_Estatee]
GROUP BY [Year] ORDER BY [Year] DESC ;
```

100 %

Results | Messages

	Year	AVG_PRICES_OVER_YEARS
1	2022	663799.7615
2	2021	505034.7791
3	2020	493269.4241
4	2019	516918.6477
5	2018	518449.9714
6	2017	434834.597
7	2016	437752.9354
8	2015	475534.9815
9	2014	591681.3231
10	2013	515660.2027
11	2012	542496.1192
12	2011	652629.6195
13	2010	608491.7886
14	2009	549085.9025
15	2008	721845.7081
16	2007	606711.7379
17	2006	607579.523
18	2005	697499.4876
19	2004	685577.7757

Query executed successfully.

```sql
--SALES DISTRIBUTION BY STATE--

SELECT [State] ,
COUNT(*) AS TOTAL_SALES ,
AVG([Price]) AS AVG_PRICE_BY_STATE
FROM [dbo].[R_Estatee]
GROUP BY [State]
ORDER BY TOTAL_SALES DESC ;

--
```

100 %

Results | Messages

	State	TOTAL_SALES	AVG_PRICE_BY_STATE
1	Connecticut	4045	369886.80
2	Massachusetts	3917	943697.6484
3	Rhode Island	1454	512981.6162
4	New Hampshire	823	548288.5419
5	Vermont	499	438523.3026
6	New York	496	648785.9052
7	Maine	383	476098.9112
8	Puerto Rico	6	436550.00
9	Georgia	3	499046.6666
10	Virgin Islands	2	3924500.00

```sql
--SALES DISTRIBUTION BY CITY--

SELECT [City] ,
COUNT(*) AS TOTAL_SALES ,
AVG([Price]) AS AVG_PRICE_BY_CITY
FROM [dbo].[R_Estatee]
GROUP BY [City]
ORDER BY TOTAL_SALES DESC ;
```

100 %

Results | Messages

	City	TOTAL_SALES	AVG_PRICE_BY_CITY
1	Boston	466	1389262.5901
2	Providence	256	411467.5585
3	Waterbury	229	222744.4803
4	Bristol	185	333942.8594
5	Manchester	176	426147.1534
6	Hartford	165	231410.8484
7	West Hartford	131	445483.0152
8	Meriden	127	251571.2598
9	Hamden	125	330554.36
10	Cranston	125	377262.656
11	New Britain	111	234676.5585
12	New Haven	103	357507.2427
13	Pawtucket	100	350010.00
14	Springfield	99	238486.5656
15	East Hartford	94	229480.8297
16	Torrington	87	220606.885
17	Middletown	85	501259.9411
18	Warwick	82	369909.7439
19	Pittsfield	80	289874.75

Query executed successfully.

```sql
--Houses Price Range--
WITH PriceRanges AS (
    SELECT
        CASE
            WHEN [Price] < 1000000 THEN 'Under 1M'
            WHEN [Price] BETWEEN 1000000 AND 5000000 THEN '1M-5M'
            WHEN [Price] BETWEEN 5000000 AND 10000000 THEN '5M-10M'
            WHEN [Price] BETWEEN 10000000 AND 15000000 THEN '10-15M'
            ELSE '15M and Above'
        END AS price_range
    FROM
        [dbo].[R_Estatee]
)

SELECT
    price_range,
    COUNT(*) AS NUMBER_OF_HOUSES
FROM
    PriceRanges
GROUP BY
    price_range
ORDER BY
    NUMBER_OF_HOUSES DESC;
```

105 %

Results | Messages

	price_range	NUMBER_OF_HOUSES
1	Under 1M	10434
2	1M-5M	1100
3	5M-10M	76
4	10-15M	15
5	15M and Above	3

05

Data visualization

رواد مصر الرقمية

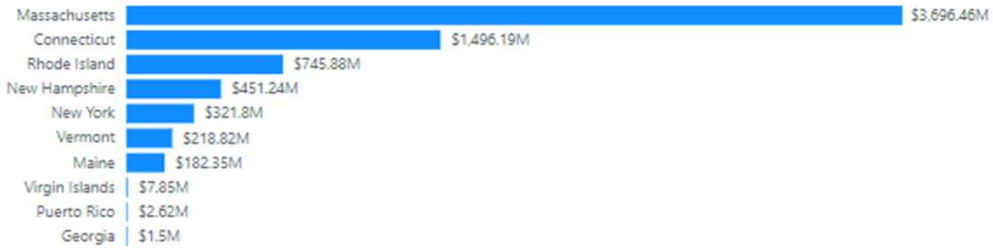
Real Estate Sales Dashboard

D E P I

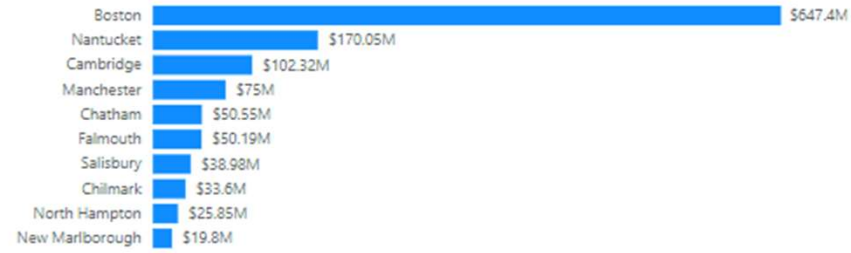
Total Sales	No.of States	No.of Cities	Total Sold Houses
$7.12bn	10	894	11.63K

Sales | Details | Map

Sum of Price by State



Sum of Price by City



Top 10 Prices

Zip_Code	City	House_Size_s_feet	No_Baths	No_Bedrooms	No_Beds	Price	Size_Status
3862	North Hampton	9443	9	3	6	$17,500,000	Large House
1259	New Marlborough	4273	5	2	3	$15,500,000	Medium House
2138	Cambridge	5709	7	4	7	$15,300,000	Medium House
2535	Chilmark	4706	6	3	6	$15,000,000	Medium House
2540	Falmouth	4244	6	2	4	$14,500,000	Medium House
2138	Cambridge	9080	11	4	7	$14,000,000	Large House
2633	Chatham	5079	7	3	5	$13,600,000	Medium House
1944	Manchester	8122	9	2	4	$12,900,000	Large House
2554	Nantucket	5085	8	4	8	$12,750,000	Medium House
2554	Nantucket	8333	9	3	5	$12,450,000	Large House

Sum of Price by Year

Real Estate Sales Dashboard

D E P I

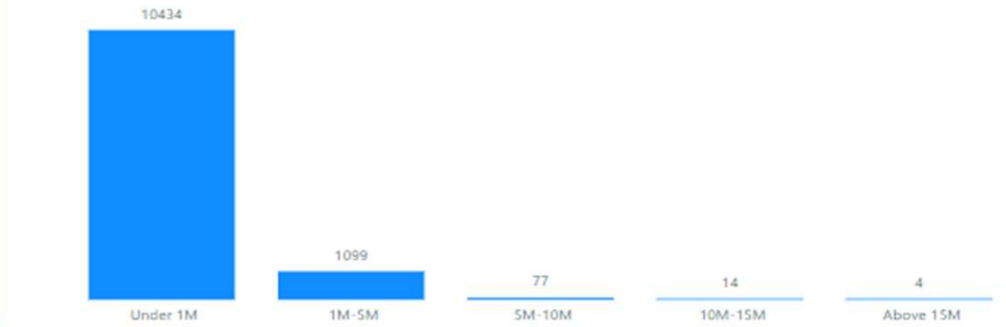
City: All

states: All

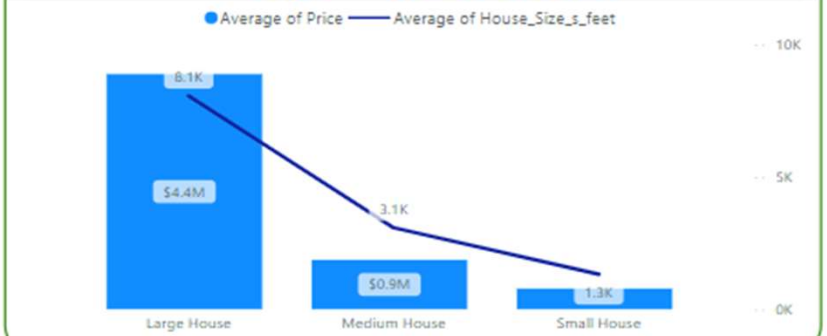
Years: 1901 – 2022

Sales | Details | Map

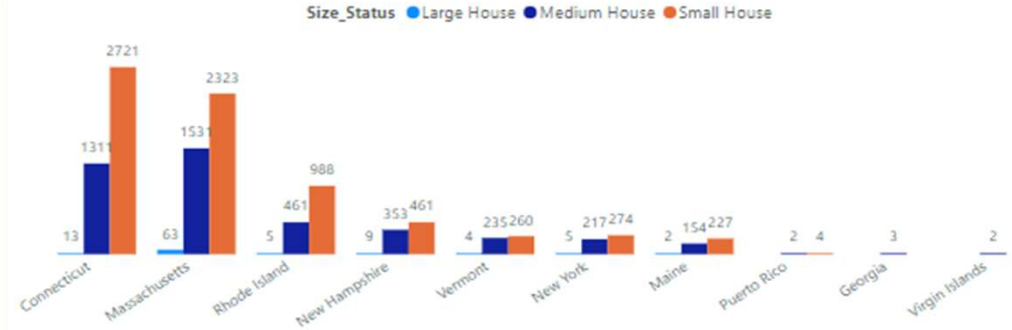
Price Ranges



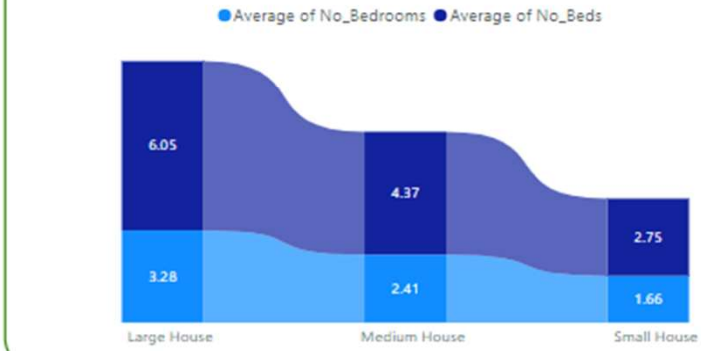
Average of Price and Size by Size

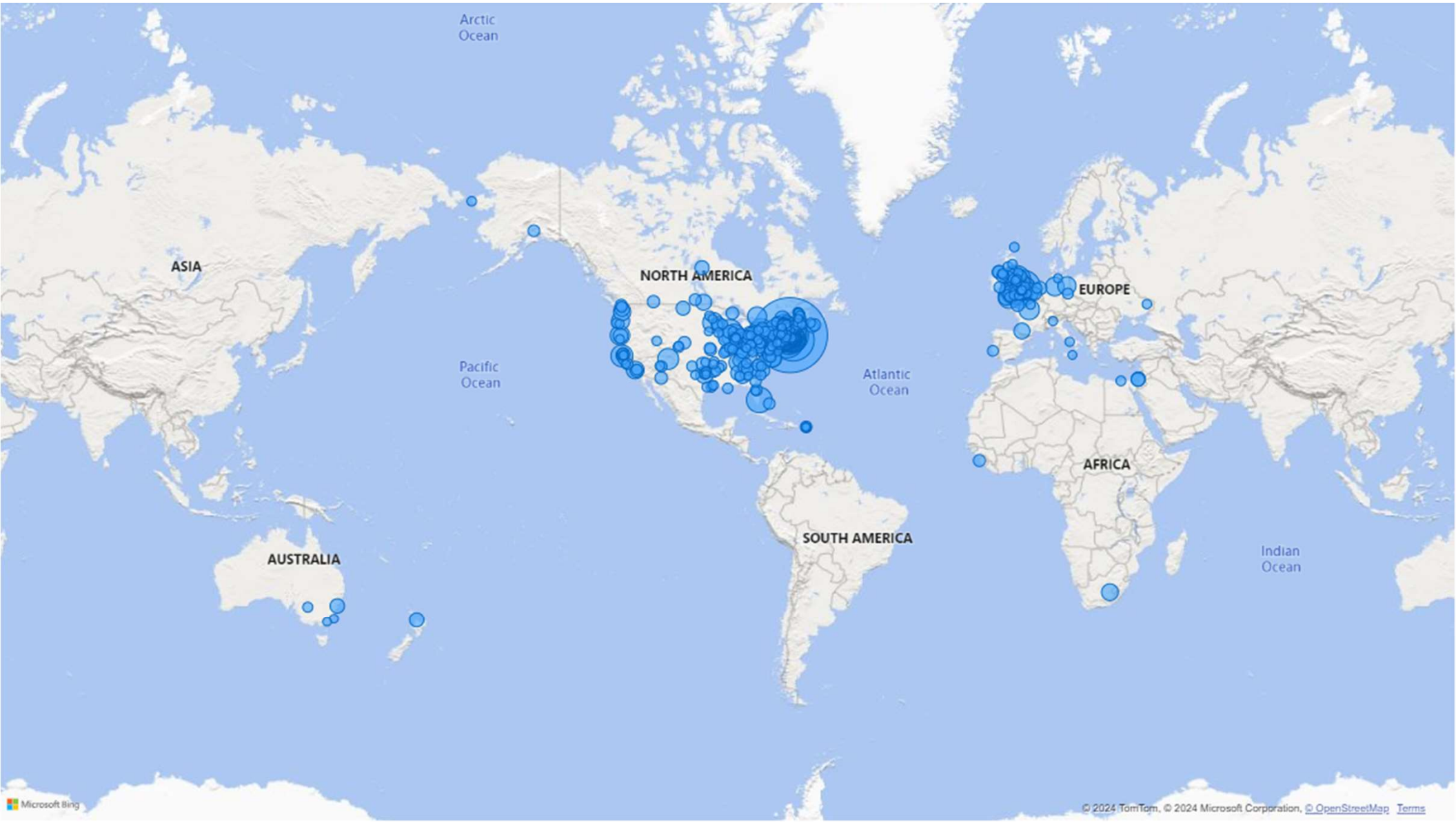


Count of Size_Status by State and Size_Status



Average of No_Bedrooms and No_Beds by Size_Status




Arctic Ocean
ASIA
NORTH AMERICA
EUROPE
Pacific Ocean
Atlantic Ocean
AFRICA
SOUTH AMERICA
AUSTRALIA
Indian Ocean
Microsoft Bing
© 2024 TomTom, © 2024 Microsoft Corporation, © OpenStreetMap Terms

Tools and Technologies

Collect Data

Data Cleaning

Data Exploring

Data Exploring & Data Cleaning

Design Dashboard

Team Connections

Team meetings

Project GP Presentation

رواد مصر الرقمية

Teamwork

	Collecting Data	Data Cleaning	Data Exploring	Data visualization	Presentation
Mohamed Mohsen	✓		✓	✓	
Mohamed fo’ad	✓	✓	✓		
Abdelrahman Mohamed		✓		✓	✓
Zyad Mohamed	✓	✓			
Ahmed ElSayed			✓	✓	✓

Any Questions?


Thank You