



Home Purchase Decision Analysis

Florida, New York, North Carolina

BaruchCOLLEGE | CU
NY



Group 2

**Team
Members:**

Syed Shah

Ena Sula

Braily
Noesi





The problem

Most people want a home, but don't even know where to begin

What the homebuyer should focus on

- Location
- Price and budget



Business Pitch

Help clients find the home they desire

Objective

Help people make an informed decision about home buying





Personas

- Homebuyers
- Dashboard Developers
- IT Project Manager
- Data Warehouse Specialist



Requirements

- We will find the average price in each county and zip code
- In New York, North Carolina, and Florida

What makes a good home purchase?

- Data accuracy
- Location
- Price transparency



What is the process behind?

- Data Sourcing

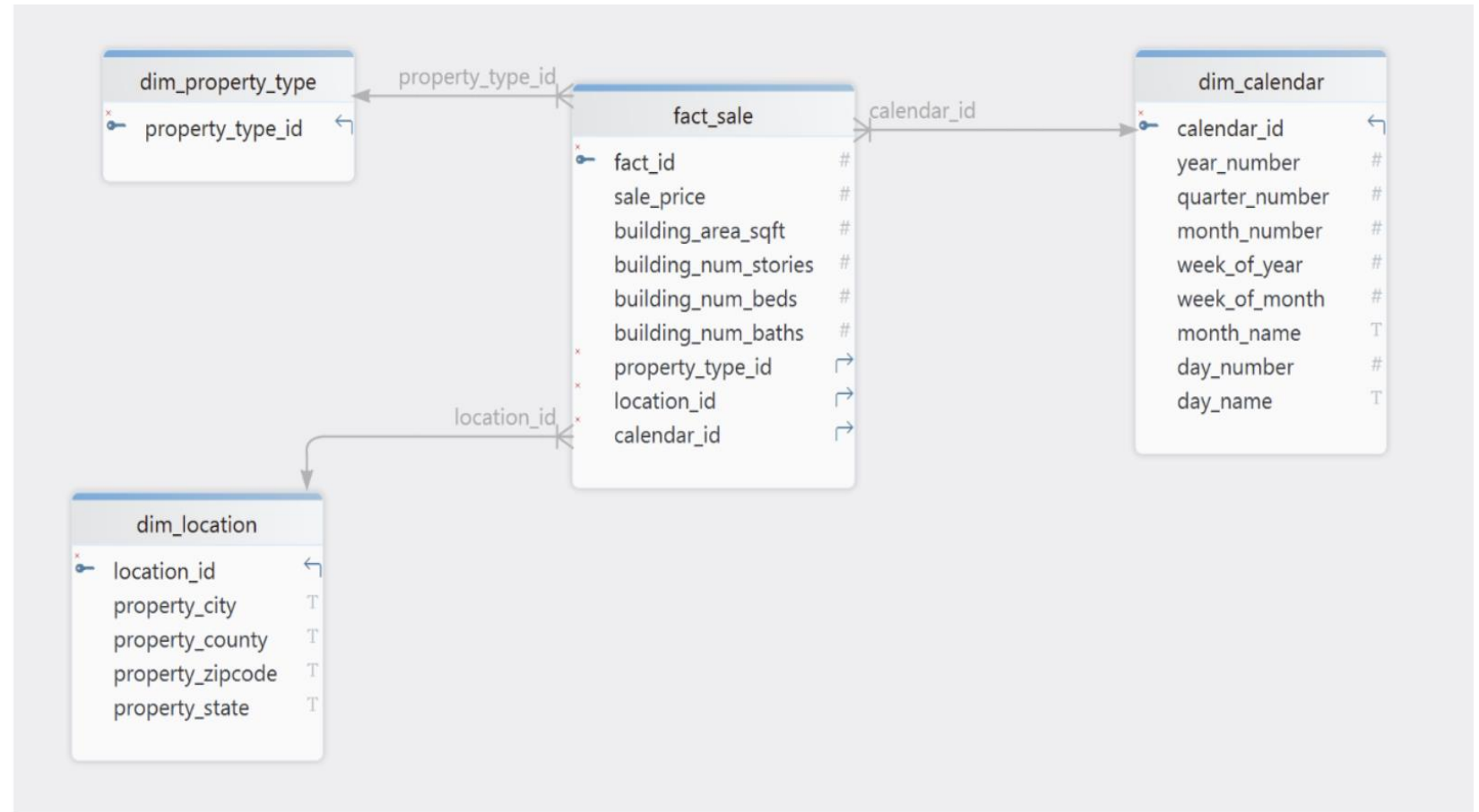
The screenshot shows the DoltHub interface for a database named 'us-housing-prices-v2'. The interface includes a sidebar with navigation links (Databases, Pricing, Documentation, Blog) and a main content area. The main content area displays the 'sales' table schema and a list of data rows. The schema includes columns: state, property_zip5, property_street_address, property_city, property_county, property_id, and sale_date. The data rows show various property listings with their respective details.

state	property_zip5	property_street_address	property_city	property_county	property_id	sale_date
AZ	85143	1 E CORAL BEAN DR	SAN TAN VALLEY	PINAL	210572230	2021-04-
AZ	85143	1 E PEPPERGRASS PL	SAN TAN VALLEY	PINAL	210571490	2020-09-
AZ	85173	1 N MESQUITE DR	SUPERIOR	PINAL	10526004A	2021-02-
AZ	85143	1 W CANYON ROCK RD	SAN TAN VALLEY	PINAL	210702840	2011-09-
AZ	85143	1 W MILL REEF DR	SAN TAN VALLEY	PINAL	210591110	2009-08-
AZ	85143	10 E MILL REEF DR	SAN TAN VALLEY	PINAL	B021FF7A-227A-4260-BFE5-543E3B64D432	2021-10-
AZ	85143	10 E MILL REEF DR	SAN TAN VALLEY	PINAL	210590970	2022-01-
AZ	85194	10 N AGUA FRIA LN	CASA GRANDE	PINAL	505900230	2020-11-
AZ	85623	10 N BACHMAN WASH RD	ORACLE	PINAL	30823026B	2009-07-
AZ	85132	10 N CENTRAL AVE	FLORENCE	PINAL	200440810	2003-11-
AZ	85138	10 N CHERRY LN	MARICOPA	PINAL	501290850	2000-07-
AZ	85172	10 N FUQUA RD	STANFIELD	PINAL	501310010	2004-11-
AZ	85138	10 N PEPPER PL	MARICOPA	PINAL	501290170	2001-07-
AZ	85138	10 N PRIMROSE LN	MARICOPA	PINAL	501290510	2004-02-
AZ	85131	10 N SANTA CRUZ AVE	ELOY	PINAL	411050040	2012-06-
AZ	85194	10 S MESILLA LN	CASA GRANDE	PINAL	505903150	2021-01-
AZ	85120	10 W 14TH AVE	APACHE JUNCTION	PINAL	102300010	2021-12-
AZ	85143	10 W CANYON ROCK RD	SAN TAN VALLEY	PINAL	210702320	2014-08-



What is the process behind?

- Data Modeling
 - Azure
 - Snowflake
 - dbt





What is the process behind?

- Data cleaning

← New relationship

BUILDING_AR...	BUILDING_N...	BUILDING_N...	BUILDING_N...	DATE_ID	FACT_SALES_ID	LOCATION
null	null	null	null	20210806	5f761380cc39...	6c5363e
null	null	null	null	20191008	cf0f90a57fc70...	c19f878!
null	null	null	null	20191008	cf0f90a57fc70...	c19f878!

To table

DIM_CALENDAR

DATE_ID	DATE_ISO	DAY_NAME	DAY_NUMBER	DAY_OF_WEEK	MONTH_NAME	MONTH_N
19830801	8/1/1983 12:0...	Mon	1	1	Augth	8
20180212	2/12/2018 12:...	Mon	12	1	Febth	2
20100513	5/13/2010 12:...	Thu	13	4	Mayth	5

Cardinality

Many to one (*:1)

Cross-filter direction

Both

☒ Make this relationship active

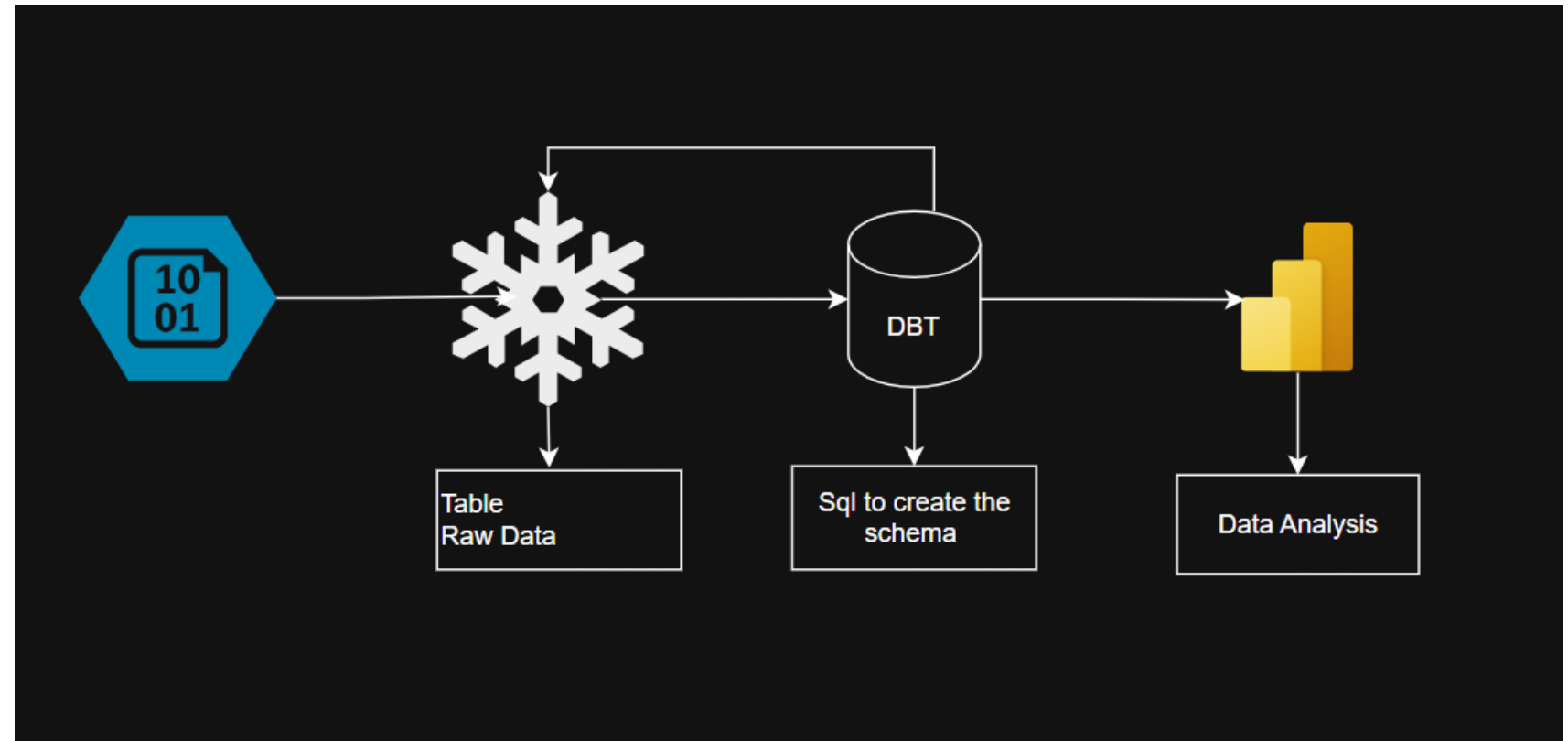
☐ Apply security filter in both directions

☐ Assume referential integrity [Learn more](#)

✖ Column DATE_ID in Table DIM_CALENDAR contains blank values and this is not allowed for columns on the one side of a many-to-one relationship or for columns that are used as the primary key of a table.

What is the process behind?

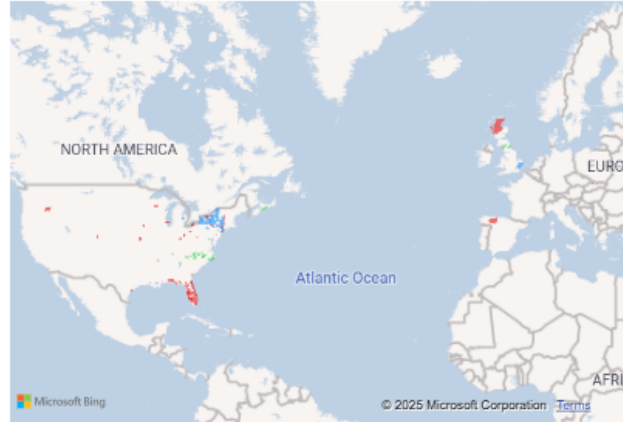
- Technical
- Visualization (Power BI)



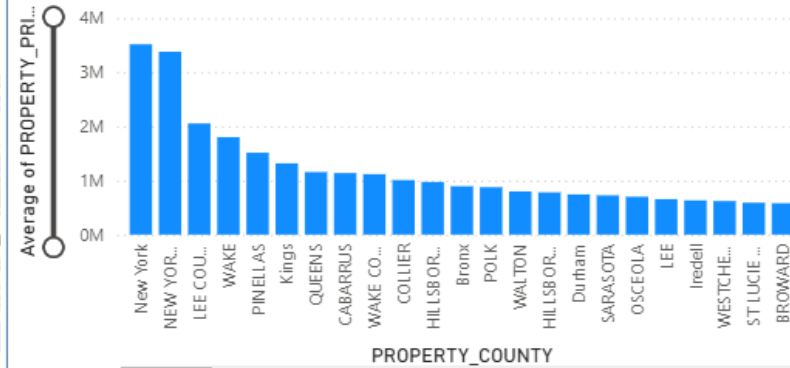
The Visualization

Average of PROPERTY_PRICE by PROPERTY_COUNTY and PROPERTY_STATE

PROPERTY_STATE ● FL ● NC ● NY



Average of PROPERTY_PRICE by PROPERTY_COUNTY



Average of PROPERTY_PRICE

555.16K

Sum of PROPERTY_PRICE

12.99T

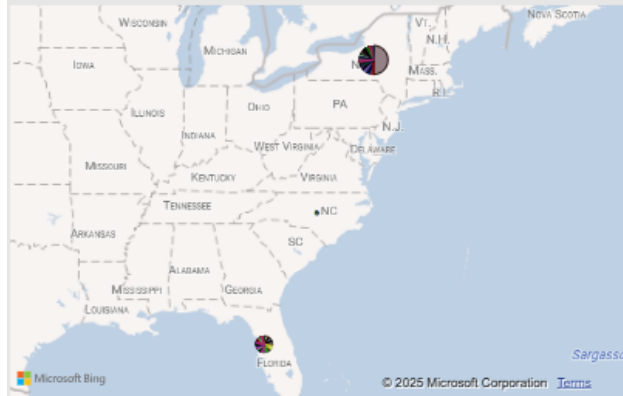
Sum of PROPERTY_PRICE by PROPERTY_STATE

FL

NC

NY

Average of PROPERTY_PRICE by PROPERTY_STATE and PROPERTY_ZIP



YEAR_NUMBER	PROPERTY_ZIP	PROPERTY_COUNTY	Average of PROPERTY_PRICE	Sum of PROPERTY_PRICE
2022			1,017,630.29	3,434,1
2022		ALBANY	286,051.19	965,7
2022		ALLEGANY	50,006.57	24,8
2022		BROOME	164,031.39	314,4
2022		BUNCOMBE	384,673.08	50,0
2022		CATTARAUGUS	86,446.51	107,0
2022		CAYUGA	123,056.59	141,3
2022		CHAUTAUQUA	118,346.66	21,3
2022		CHEMUNG	1,000.00	
Total			555,162.05	12,985,736.2

New York

PROPERTY_COUNTY

Search

ALBANY

ALLEGANY

Bronx

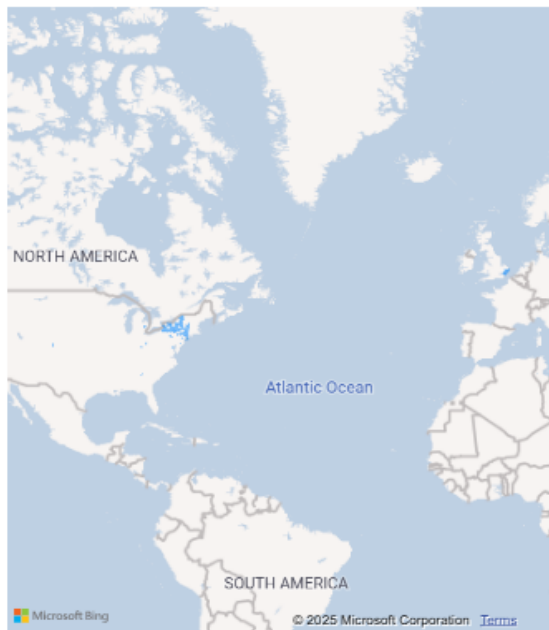
Average of PROPERTY_PRICE

750.97K

Sum of PROPERTY_PRICE

4.54T

PROPERTY_COUNTY and PROPERTY_ZIP

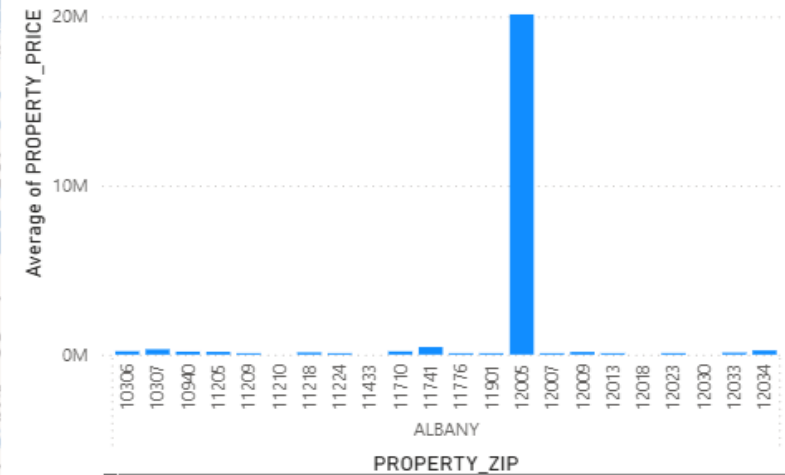


Average of PROPERTY_PRICE by
PROPERTY_COUNTY



PROPERTY_ZIP	PROPERTY_COUNTY	Average of PROPERTY_PRICE	YEAR_NUMBER
99664	STEUBEN	76,300.00	2020
97501	SUFFOLK	240,000.00	2019
971	SUFFOLK	1,400,000.00	2020
96816	ERIE	42,350.00	2013
94019	WESTCHESTER	33,000.00	2019
94019	WESTCHESTER	26,000.00	2021
9390	SUFFOLK	1.00	2016
93021	SULLIVAN	100,126.00	2015
92868	ERIE	240,000.00	2017
Total		861,789.45	

Average of PROPERTY_PRICE by PROPERTY_COUNTY and
PROPERTY_ZIP





North Carolina

PROPERTY_COUNTY

Search

ALEXANDER

BUNCOMBE

Burke

Average of PROPERTY_PRICE

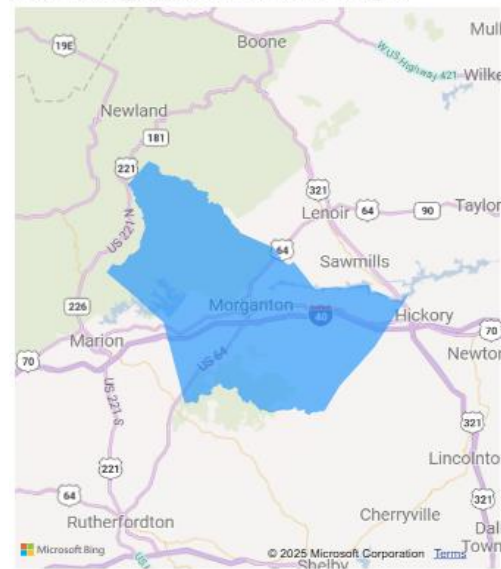
26.54K

Sum of PROPERTY_PRICE

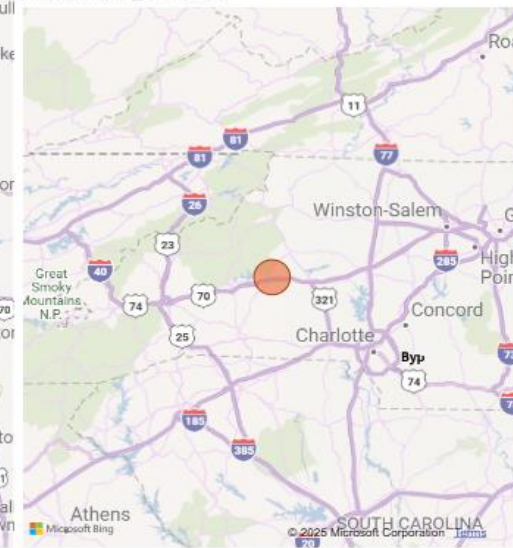
1.47bn

PROPERTY_ZIP	PROPERTY_COUNTY	Average of PROPERTY_PRICE	YEAR_NUMBER
28761	Burke	166,400.00	2002
28761	Burke	57,083.33	2003
28761	Burke	168,854.17	2004
28761	Burke	798,220.00	2005
28761	Burke	189,779.87	2006
28761	Burke	386,269.23	2007
28761	Burke	238,026.32	2008
28761	Burke	88,539.77	2009
28761	Burke	106,074.35	2010
Total		28,276.19	

PROPERTY_COUNTY and PROPERTY_ZIP



Average of PROPERTY_PRICE by PROPERTY_COUNTY



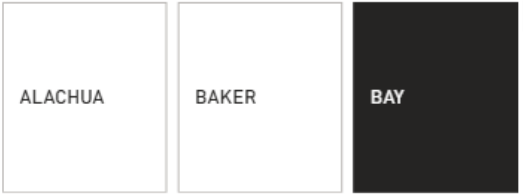
Average of PROPERTY_PRICE by PROPERTY_COUNTY and PROPERTY_ZIP





Florida

PROPERTY_COUNTY



Average of PROPERTY_PRICE

340.62K

Sum of PROPERTY_PRICE

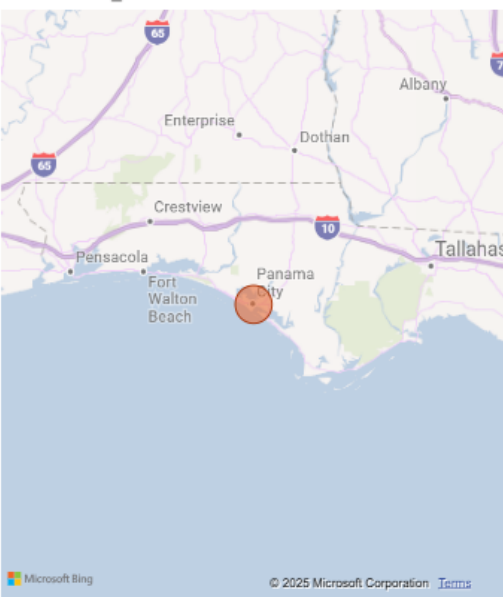
8.87bn

PROPERTY_ZIP	PROPERTY_COUNTY	Average of PROPERTY_PRICE	YEAR_NUMBER
32466	BAY	42,945.07	2020
32466	BAY	76,020.16	2021
32466	BAY	171,115.22	2022
32444	BAY	534,472.19	2020
32444	BAY	308,284.77	2021
32444	BAY	1,197,273.18	2022
32438	BAY	32,748.55	2020
32438	BAY	29,610.28	2021
32438	BAY	50,076.09	2022
Total		340,617.76	

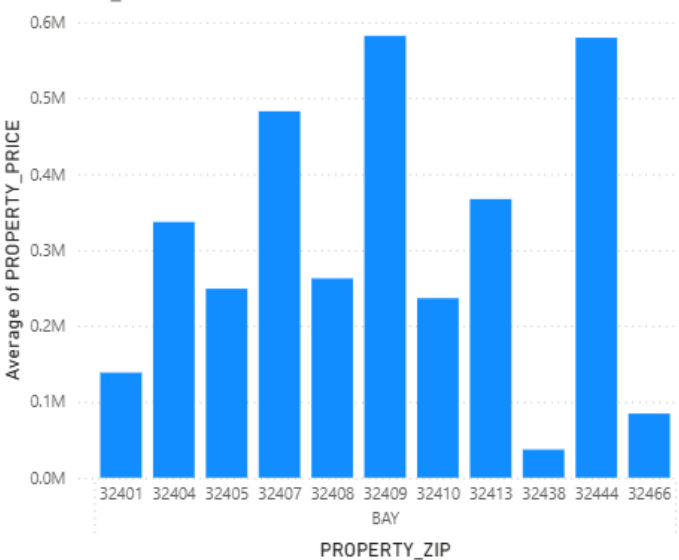
PROPERTY_COUNTY and PROPERTY_ZIP



Average of PROPERTY_PRICE by PROPERTY_COUNTY



Average of PROPERTY_PRICE by PROPERTY_COUNTY and PROPERTY_ZIP





Results

- **New York**
 - Highest Average: New York **\$3.52M**
 - Lowest Average: St Lawrence **\$58.16**
- **North Carolina**
 - Highest Average: Wake **\$1.80M**
 - Lowest Average: Lee **\$2.16**
- **Florida**
 - Highest Average: Lee County **\$2.05M**
 - Lowest Average: Bradford **\$50.96K**



Conclusion

❖ Make an informed, data-driven decision.



Future Work

- ❖ Other considerations that could be examined while making a decision regarding buying a property are the crime rate, school ranking, public transportation, and groceries suppliers.



Members Contribution



Syed Shah: The leader and has been in charge of keeping us up to date with the analysis.



Ena Sula: In charge of dimensional modeling and visualization.



Braily Noesi: In charge of storing the data in temporary storage, and code process.



Thank You!

{Have any question? }

