East Brunswick, NJ
Credit: Douglas Bauman
https://mycoast.org/reports/85367

Precipitation: Impact on Real Estate Values

1

Contents

Group 5

Archana Jadhav

Carolina Ramos

Mae Gaudio

Selected Topic

Group 5 selected this topic because we wanted to explore the relationship, if any, between storm events and central NJ residential real estate values.

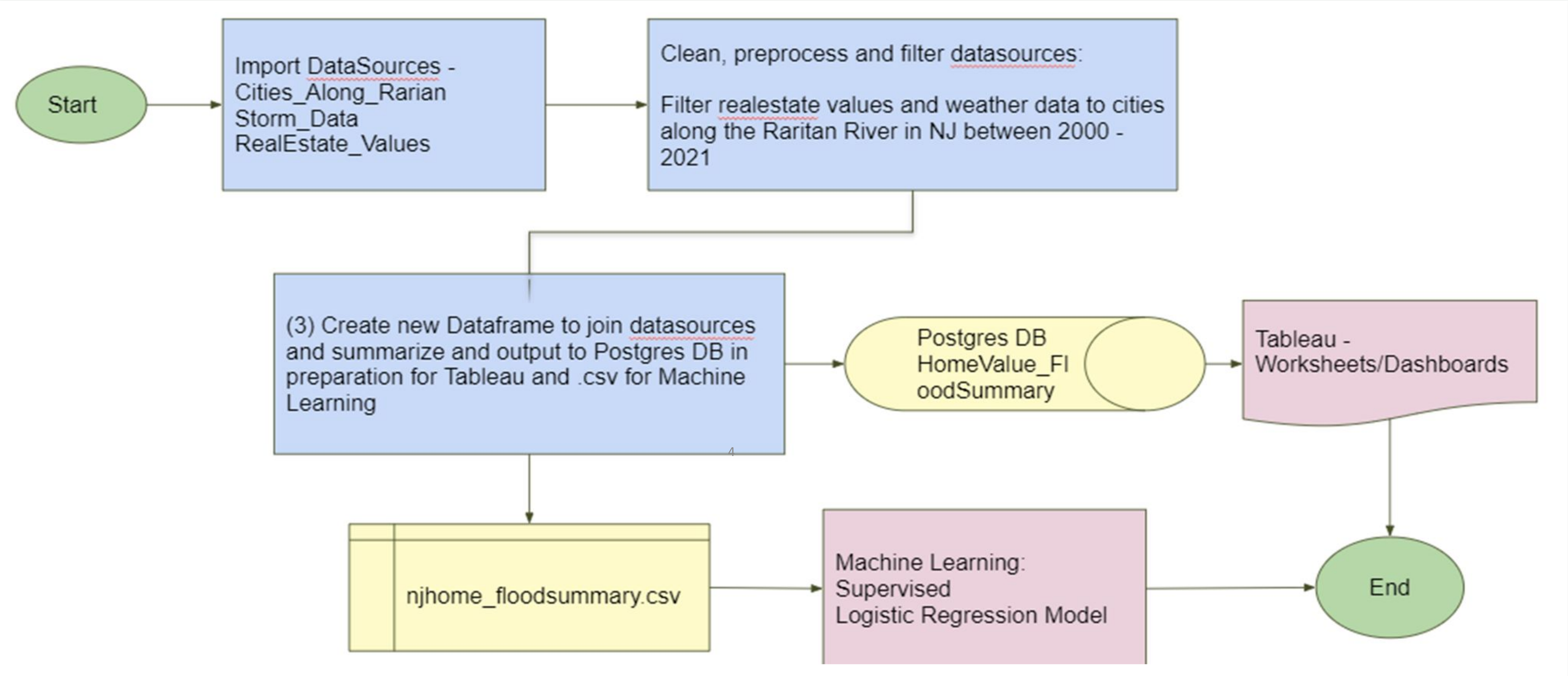
Our project focuses on municipalities that border the Raritan River, single-family homes in the area, and precipitation data.

Geographic location: Central Jersey, namely New Brunswick & Bridgewater, surrounding the Raritan River

Real estate data: Single-family homes & list price

Rain overflow from major Storm Events like Hurricanes Sandy and Ida

Source Data Flow to Analysis Presentation



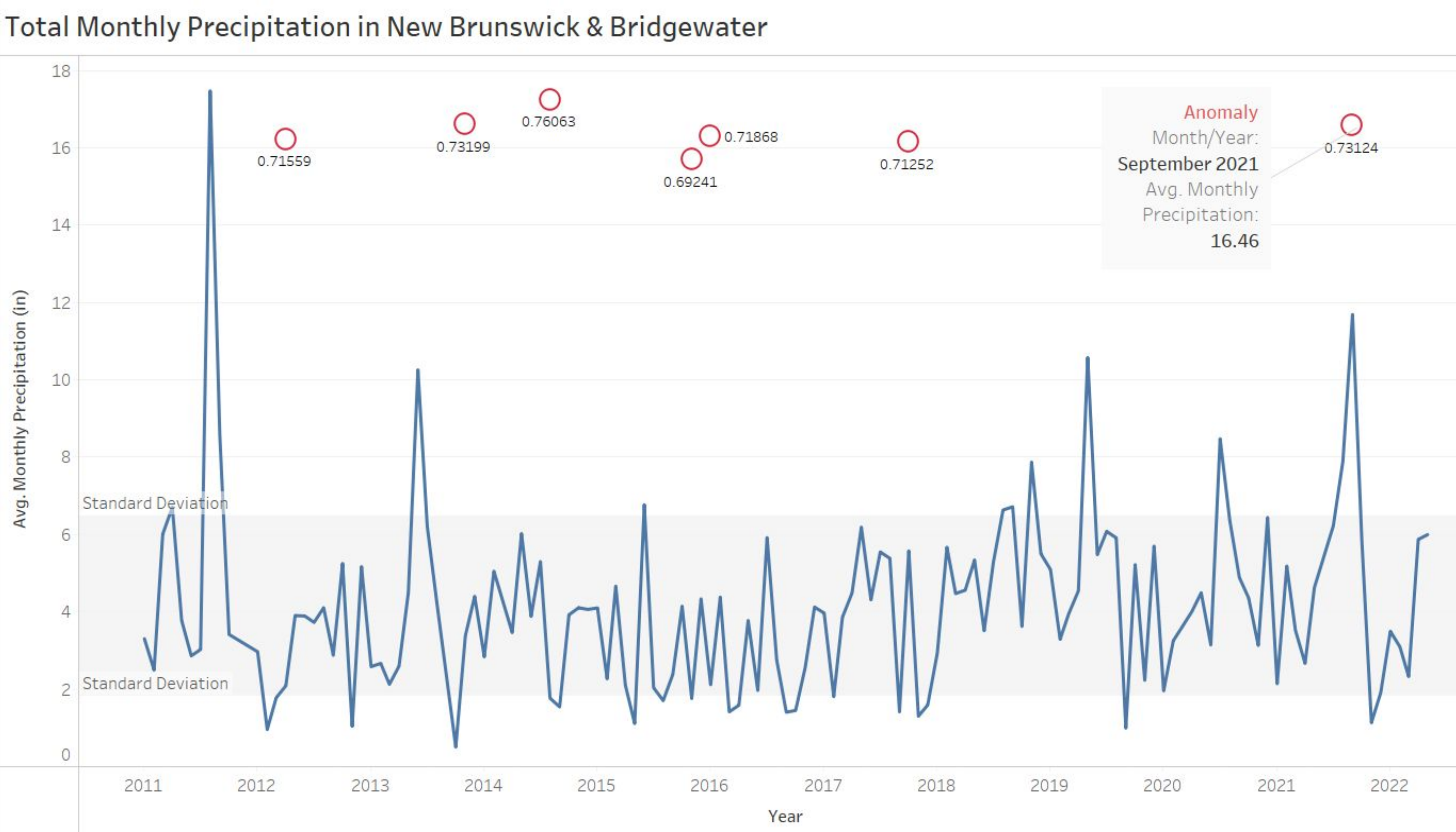
Github Repo: https://github.com/c-ramos/NJ_Flood_Risk_Capstone

Jupyter Notebooks, SQL scripts and Output Files

(Preparation for ML and Tableau data)

Jupyter Notebooks & SQL Scripts	Description	Input Files	Output to CSV	Postgres DB Table Name
ConvertRain.ipynb	Cleans daily_rainfall, calculates mean, max, total rainfall for the month and identifies if there was an anomaly event (storm) that took place within the month.	daily_rainfall.csv	per_city_rainfall_final.csv	per_city_rainfall
CleanCities_and_HomePrice.ipynb	Cleans cities, transposes dataframe to calculate historical realestate list prices by city, month and year.	ZipcodePricealltypeshouse.csv	njhomeprice_final.csv	njhomeprice
schema.sql	Creates tables in Postgres DB	n/a	n/a	per_city_rainfall cities njhomeprice njhome_floodsummary
njhome_floodsummary.sql	Joins the rainfall table to the real estate home listing table to export to .csv for machine learning and Tableau worksheets/dashboard	n/a	n/a	njhome_floodsummary
NJ_FloodSummary.ipynb	Generates final .csv for use by Tableau and Machine Learning.	AWS	njhome_floodsummary.csv	njhome_floodsummary
ML_Pricedrop.ipynb	Data split into test and train. Models analyzed and most accurate model determined.	njhome_floodsummary.csv	Tableau pictures provided	ML_PriceDrop

Analysis - Average Monthly Precipitation(in)



The trends of Avg. Monthly Precipitation and anomalies above 70% for New Brunswick & Bridgewater. Blue line shows details about average monthly precipitation. The data is filtered by Year and Month. The view is filtered on red open circles, which indicates an anomaly.

Analysis - Monthly Precipitation Anomaly (in)

Precipitation Anomaly by City & Month/Year

City	Month of Eo..	2011	2012	2013	2014	2015	2016	2017	2018	2019	2020	2021	2022
Bridgewater	January	0.3931	0.3607	0.4427	0.1940	0.3967	0.7513	0.3821	0.5743	0.2539	0.6030	0.3750	0.4348
	February	0.2599	0.2870	0.3411	0.2632	0.4541	0.3596	0.2591	0.1632	0.2607	0.2374	0.2829	0.2370
	March	0.3750	0.3408	0.3128	0.5621	0.1786	0.2340	0.2487	0.3364	0.3065	0.3018	0.3592	0.2267
	April	0.3251	0.7577	0.2815	0.3671	0.6262	0.2061	0.2995	0.3114	0.2215	0.3270	0.4676	0.3877
	May	0.2151	0.1956	0.2611	0.5812	0.5214	0.2284	0.2090	0.1906	0.1432	0.5669	0.3875	0.2218
	June	0.3471	0.3973	0.2716	0.2634	0.2890	0.3283	0.4420	0.2101	0.4011	0.5485	0.4760	
	July	0.4295	0.2263	0.4045	0.3438	0.1803	0.2200	0.2020	0.2374	0.3053	0.4013	0.3058	
	August	0.3997	0.3756	0.3214	0.8462	0.3883	0.2378	0.2661	0.2976	0.2595	0.2574	0.5362	
	September	0.3375	0.3048	0.5072	0.3106	0.5207	0.3305	0.3913	0.2275	0.7033	0.5010	0.7295	
	October	0.3412	0.3630	0.6774	0.4322	0.5446	0.3828	0.7023	0.6133	0.3216	0.4060	0.4638	
	November	0.5000	0.4086	0.7260	0.2456	0.7446	0.6951	0.4019	0.2546	0.4247	0.3067	0.4508	
	December	0.5192	0.2927	0.2708	0.2319	0.2412	0.3108	0.2968	0.2922	0.1609	0.3640	0.2692	
New Brunswick	January	0.4484	0.5751	0.3387	0.1967	0.5234	0.6208	0.3680	0.4274	0.3673	0.6383	0.4238	0.3456
	February	0.4524	0.3500	0.3435	0.2356	0.4510	0.3478	0.4690	0.1687	0.2773	0.2268	0.2870	0.2070
	March	0.3603	0.4318	0.3562	0.5972	0.1616	0.2361	0.2479	0.3507	0.3271	0.2679	0.3356	0.2218
	April	0.6839	0.5893	0.3843	0.2637	0.5893	0.2286	0.2870	0.3624	0.2197	0.2478	0.3077	0.2718
	May	0.3100	0.2507	0.3855	0.6624	0.5684	0.3410	0.2519	0.2262	0.1531	0.2407	0.3908	0.2186
	June	0.2637	0.3521	0.2531	0.4259	0.3070	0.4308	0.3698	0.2866	0.2407	0.4944	0.4830	
	July	0.5098	0.3420	0.2998	0.3373	0.3184	0.4364	0.2364	0.3007	0.2877	0.2637	0.2229	
	August	0.5009	0.3419	0.3303	0.5041	0.5042	0.2500	0.2743	0.3055	0.7815	0.1765	0.2932	
	September	0.3449	0.2609	0.3760	0.4815	0.5215	0.4880	0.5621	0.2172	0.5238	0.3828	0.7365	
	October	0.3696	0.3063	0.3529	0.3351	0.4227	0.2792	0.7432	0.4571	0.2215	0.3931	0.5837	
	November	0.6046	0.4855	0.7500	0.2729	0.5359	0.4826	0.3682	0.3152	0.4103	0.2921	0.2809	
	December	0.6613	0.3582	0.2559	0.3362	0.2445	0.2889	0.2791	0.2026	0.1835	0.2328	0.1277	

Avg. Anomalypct

0.1277 0.8462

Anomaly broken down by Month/ Year vs. City. Darker color shows anomaly. The marks are labeled by average anomaly.

Analysis - Average Sale Price by City/Storm Event

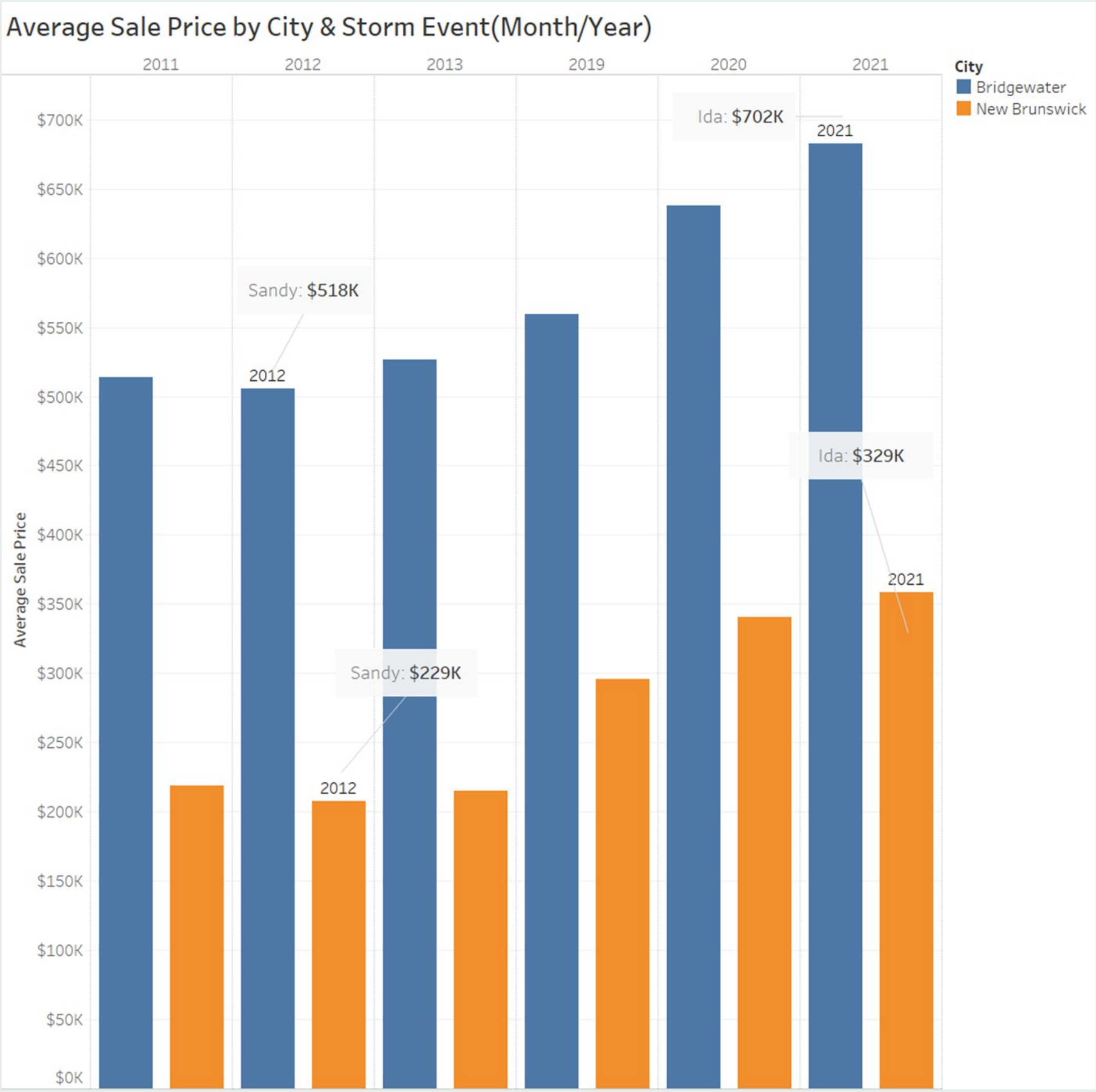
Average sale price per city broken down by storm event (Hurricanes Sandy & Ida and Month/Year). Blue color shows details about Bridgewater, NJ. Orange color is New Brunswick, NJ. The marks are labeled by storm event month and year.

Sandy Oct. 2012

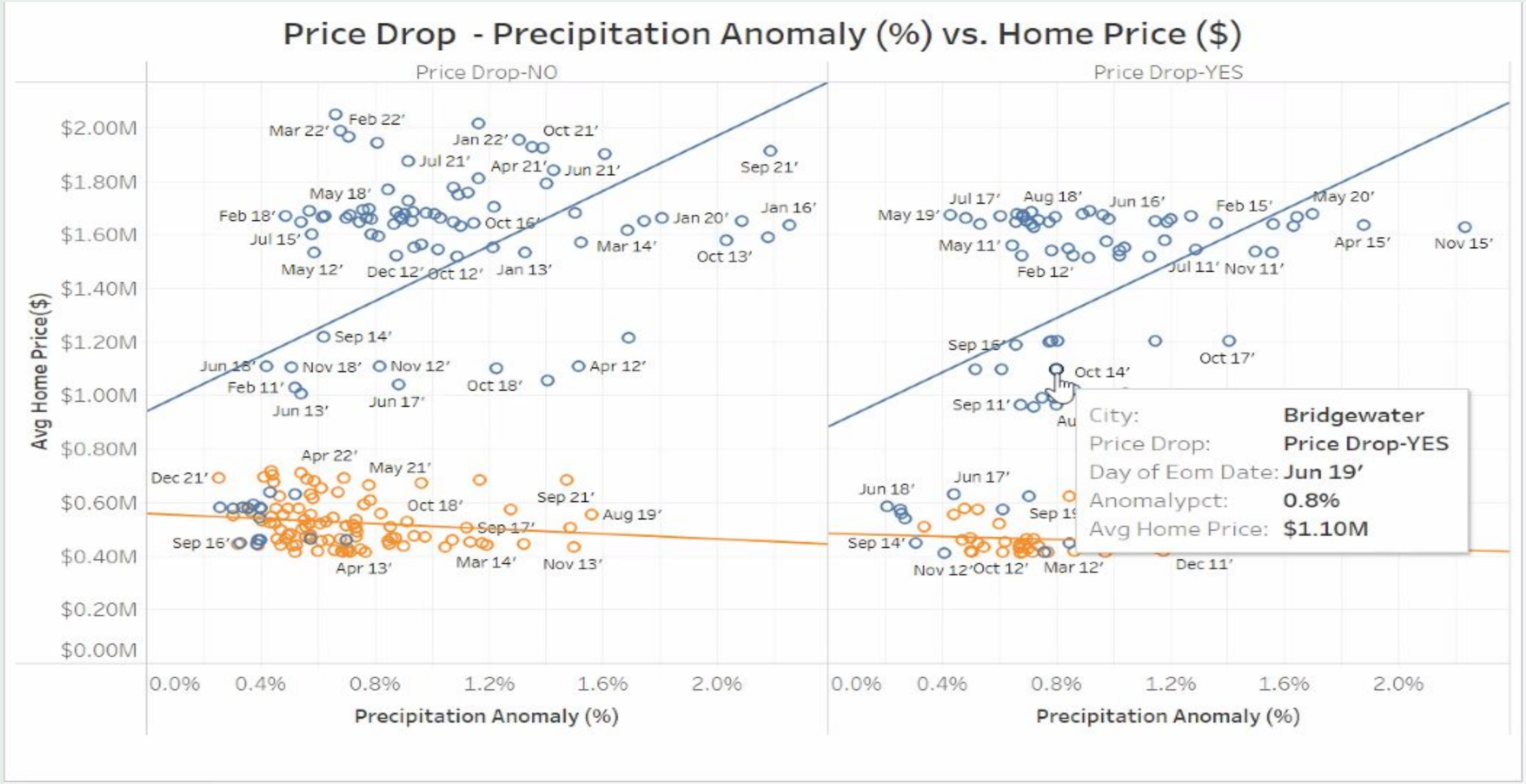
Pre- and post-event data from 2011 and 2013

Ida Sept. 2021

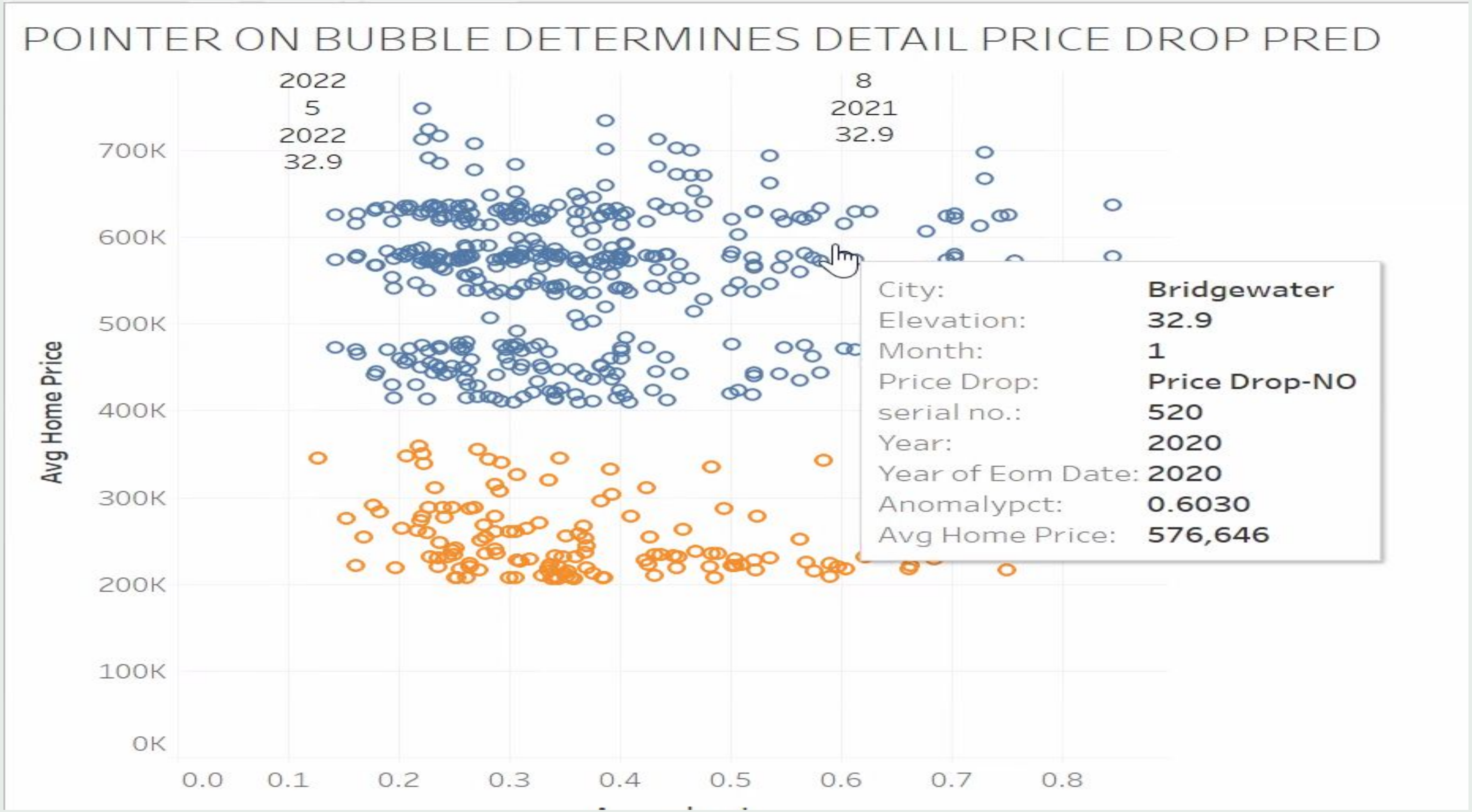
Pre- and post-event data from 2019-2021



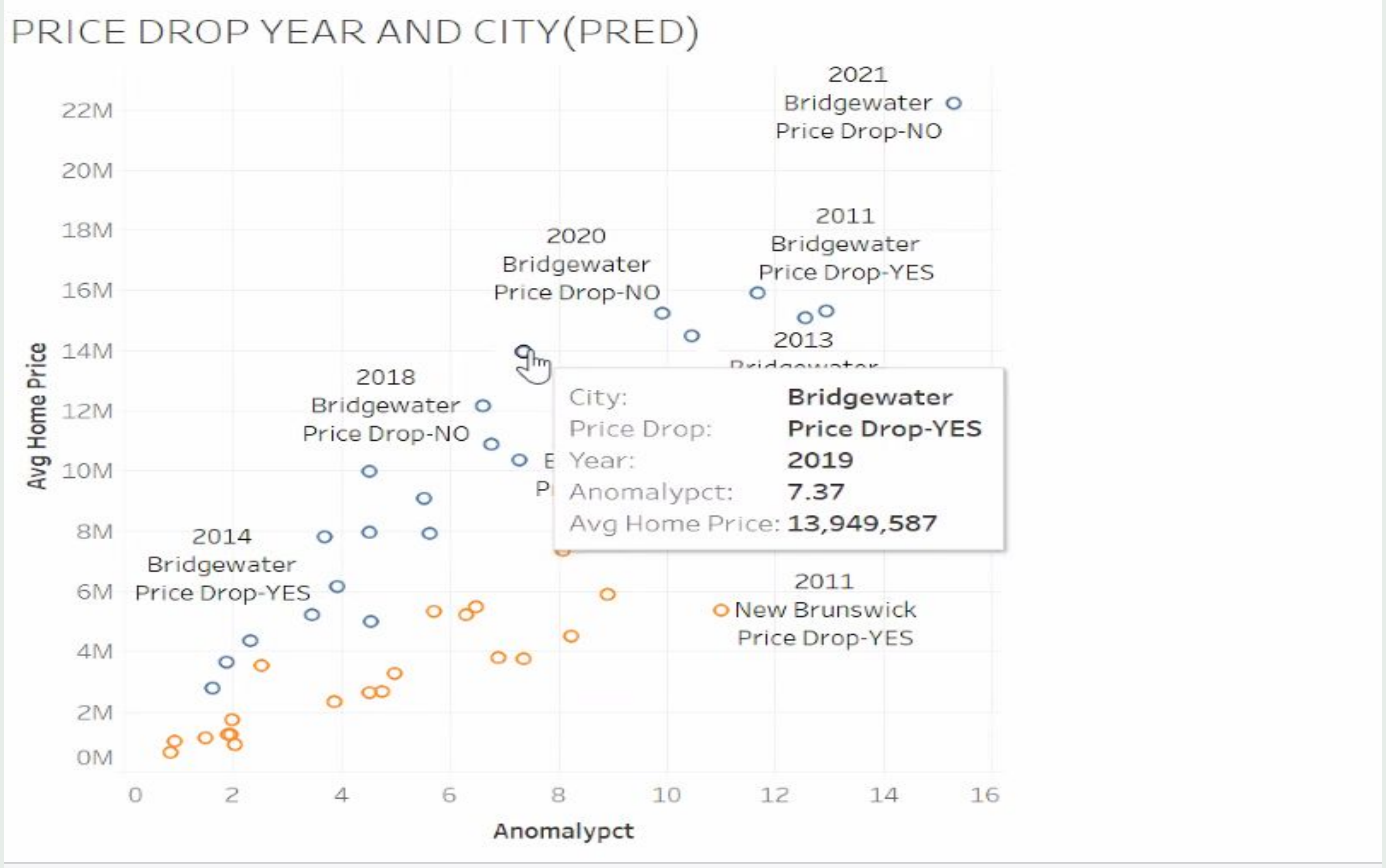
Machine Learning - Price Drop & Precipitation Anomaly vs Home Price



Machine Learning - Precipitation Anomaly vs Home Price



Machine Learning - Price Drop by City



Our Findings

MACHINE LEARNING MODELS(rubric 2)

ModelsML_Pricedrop.ipynb (This has detail test and train model for ML to predict
NJ_flood risk and house Pricedrop variaton Prediction).

SUPERVISED LEARNING

1. Preliminary code preprocessing (at this moment only New Brunswick and Bridgewater).
2. using ML codes ran
(Logistic Regression,Decision Tree,
RandomForest)

RESULTS

1.logistic Regression gave 0.5988372093023255 accuracy
2.Decision Tree gave 0.7848837209302325
3.Random forest gave Accuracy Score : 1.0 -(which is to the perfection but in reality no data can be so perfect ,so ignoring this Model.)

VISUALIZATION OF MODEL

Output is in Tableau visualization

Our Findings (Rubric3)

The Data was was processed using Pandas dataframe it has (2 cities and 2 counties) through year 2011-2022 the data of Precipitation and home price
is in the CSV. Input is independent variable is Precipitation(anamoly pct)and house Price drop ("Y" or "N") is projected over years.
The njhome_floodsummary csv was updated

CURRENTACCURACY SCORE for VARIOUS MODELS

LOGISTIC REGRESSION-0.5766423357664233
DECISION TREE-0.7883211678832117
,RANDOM FOREST -1.0 (Using the equation (TP + TN) / Total, we can determine our accuracy)-but thisis unrealistic so droping this model .

so keeping DECISION TREE as the regression .

Tableau shows the output with Price drops VS Anomaly pct

Future Recommendations

Based on our findings, we recommend looking at other factors that drive the real estate market such as:

- Location
  - Elevation
  - High- and low-density housing
- Demographics such as income and age buyers
- Housing type - condominiums, multi-family, and home's age and condition
- Legislation and policies - FIRMs (Flood Insurance Rate Map), building code requirements, tax incentives or subsidies
- State of the economy

Looking Back

- New Brunswick elevation is 12.5 and Bridgewater is 32.9
- There are a number of factors that impact real estate values, such as economic health, market cycles, interest rates, etc.
- FEMA flood zones and maps could provide more information if we looked into census tracts/census blocks versus zip code or municipality only

Appendix A

Key Data Sources and URLs

Dataset	Data Source	Details
Cities_Along_Raritan.csv	FEMA Flood Map Service Center | Search By Address	Picked select cities along the Raritan River (New Brunswick and Bridgewater, NJ)
daily_rainfall.csv (combined from all .csv's listed in /Resources/Cities CSVs)	https://www.ncdc.noaa.gov/cdo-web/datasets	Precipitation: daily rainfall in inches by city (historical from 2011 to 2021)
ZipcodePricealltypeshouse.csv	Housing Data - Zillow Research	Average list price all homes (USD) by city (historical from 2000 to 2021

Appendix B

Key Data Sources and URLs

Infrastructure	Link
Jupyter Notebook / SQL scripts	Github Repo: https://github.com/c-ramos/NJ_Flood_Risk_Capstone
AWS RDS - Postgres	Endpoint: finalprojectgroup5.c1jelrjhbrlm.us-east-1.rds.amazonaws.com
Tableau	https://public.tableau.com/views/Group5_Dashboard_16569394107950/Group5_Dashboard?:language=en-US&:display_count=n&:origin=viz_share_link
Machine Learning Models	Supervised => Logistic Regression, Decision Tree, Random Forest Regression. Determine best model.

Thank You

Group 5

Archana Jadhav

Carolina Ramos

Mae Gaudio

MACHINE LEARNING

PART PRESENTATION

CHOICE OF MODEL

SUPERVISED LEARNING-

Supervised learning deals with labeled data.

The reason to choose this is we have labeled data (njhome_floodsummary.csv)

We are Predicting 2011-2022 Precipitation(Anomaly_pct) vs House _Price drop

The Prediction is “Y” and “N”.

So its Binary Classification.

FIRST SEGMENT-RUBRIC

(Decision for binary classification Algorithms)

REGRESSION-**regression** is used to predict continuous variables

LOGISTIC REGRESSION-**Logistic regression** predicts binary outcomes, meaning that there are only two possible outcomes.

CLASSIFICATION-**Classification**, on the other hand, is used to predict discrete outcomes

The outcome, in this case, is whether the person will vote "Yes" or "No." The classification model's algorithms would attempt to learn patterns from the data, and if the model is successful, gain the ability to make accurate predictions for new voters.

1.DECISSION TREE

2.RANDOM FOREST(ENSEMBLE)

SECOND SEGMENT-RUBRIC

(DECISION TO CREATE A MACHINE LEARNING WITH AVAILABLE DATA)

Converted all object variable to float/Integer

Result for all algorithm accuracy as below

1.logistic Regression gave 0.5988372093023255 accuracy

2.Decision Tree gave 0.7848837209302325

3.Random forest gave Accuracy Score : 1.0 -(which is to the perfection but in reality no data can be so perfect ,so ignoring this Model.

1.logistic Regression gave 0.5988372093023255 accuracy
2.D1.logistic Regression gave 0.5988372093023255 accuracy
2.Decision Tree gave 0.7848837209302325
3.Random forest gave Accuracy Score : 1.0 -(which is to the perfection but in reality no data can be so perfect ,so ignoring this Model.
ecision Tree gave 0.7848837209302325
3.R1.logistic Regression gave 0.5988372093023255 accuracy
2.Decision Tree gave 0.7848837209302325
3.Random forest gave Accuracy Score : 1.0 -(which is to the perfection but in reality no data can be so perfect ,so ignoring this Model.andom forest gave Accuracy Score : 1.0 -(which is to the perfection but in reality no data can be so perfect ,so ignoring this Model.
1.logistic Regression gave 0.5988372093023255 accuracy
2.Decision Tree gave 0.7848837209302325
3.Random forest gave Accuracy Score : 1.0 -(which is to the perfection but in reality no data can be so perfect ,so ignoring this Model.)

THIRD SEGMENT RUBRIC

THE .CSV was updated fro 2 CITIES AND 2 COUNTIES

The whole df(with X_train all 16 columns) was used

And it gave

Results for 3 ALGORITHMS as below-

1.LOGISTIC REGRESSION-0.5766423357664233

2.DECISION TREE-0.7883211678832117

3.RANDOM FOREST -1.0 (Using the equation (TP + TN) / Total, we can determine our accuracy)-but thisis unrealistic so droping this model

NOTHING CHANGED MUCH -and decided to go with DECISION TREE

X_train shape selection- when ran all columns

```
In [71]:  # We can sort the features by their importance.
          sorted(zip(rf_model.feature_importances_, X.columns), reverse=True)

Out[71]: [(0.7985322390260998, 'price_drop_amt'),
          (0.05593839549998928, 'year'),
          (0.03440773750384153, 'avghomeprice_month'),
          (0.02272429981367929, 'city_max_day_rain'),
          (0.021444610569882075, 'city_avg_daily_rain'),
          (0.02037542199981193, 'Anomalypct'),
          (0.019108115818803586, 'city_month_total_rain'),
          (0.012722273931325145, 'month'),
          (0.006012076323998551, 'SizeRank'),
          (0.0031114961841169977, 'zipcode'),
          (0.001027733731375137, 'CountyName'),
          (0.001006643330261235, 'LONGITUDE'),
          (0.0009991019917467102, 'ELEVATION'),
          (0.0009764576405017474, 'CITY'),
          (0.0008949748195014253, 'LATITUDE'),
          (0.0007184218148963946, 'Anomaly'),
          (0.0, 'State')]
```

To get better Algorithm results- X_train reframed

Rank the Importance of Features#

In [51]:

```
# Calculate feature importance in the Random Forest model.
importances = rf_model.feature_importances_
importances
```

Out[51]:

```
array([0.01233577, 0.01042029, 0.11632184, 0.24015627, 0.18402527,
       0.19950475, 0.23723581])
```

In [52]:

```
# We can sort the features by their importance.
sorted(zip(rf_model.feature_importances_, X.columns), reverse=True)
```

Out[52]:

```
[(0.24015627453179084, 'year'),
 (0.23723581138264804, 'avghomeprice_month'),
 (0.19950474894836076, 'Anomalypct'),
 (0.18402526930943405, 'city_month_total_rain'),
 (0.11632184064232612, 'month'),
 (0.01233576996187017, 'CITY'),
 (0.01042028522356992, 'ELEVATION')]
```

FINAL RUBRIC

The njhome_floodsummary csv was updated

and we got CURRENTACCURACY SCORE for VARIOUS alogorithms

by minimising/reframing on X_train.shape

X = X = df [['CITY', 'ELEVATION', 'month', 'year',

'city_month_total_rain',

'Anomalypct',

'avghomeprice_month']]-7 columns -projected in Feature importance

1.LOGISTIC REGRESSION-0.5912408759124088

2.DECISION TREE-0.8029197080291971

3.RANDOM FOREST -0.8321167883211679 (Using the equation (TP + TN) / Total, we can determine our accuracy)- this gave more accuracy m

The final algorithm for the project is RANDOM FOREST with highest accuracy.

MACHINE LEARNING FINAL RESULT

It showed that Random forest gave best accuracy result

If compared Final JUPYTER NOTEBOOK -ML_Pricedrop.ipynb

RANDOM FOREST and DECISION TREE showed same Features with different ratio of importance

The Random forest has almost the same hyperparameters as a decision tree. Its ensemble method of decision trees is generated on randomly split data.

-continued next page

RANDOM FOREST OVER DECISION TREE

DECISION TREE-

-A decision tree is a tree-like model of decisions along with possible outcomes.

-There is always a scope for overfitting, caused due to the presence of variance.

-The results are not accurate.

RANDOM FOREST-

-A classification algorithm consisting of many decision trees combined to get a more accurate result as compared to a single tree.

-Random forest algorithm avoids and prevents overfitting by using multiple trees.

-This gives accurate and precise results.

SO RANDOM FOREST BEST FOR THIS MODEL TO PREDICT PRICEDROP

VISUALIZATION
-TABLEAU-https://public.tableau.com/views/FinalDashboard_16571266071880/Dashboard1?:language=en-US&:display_count=n&:origin=viz_share_link

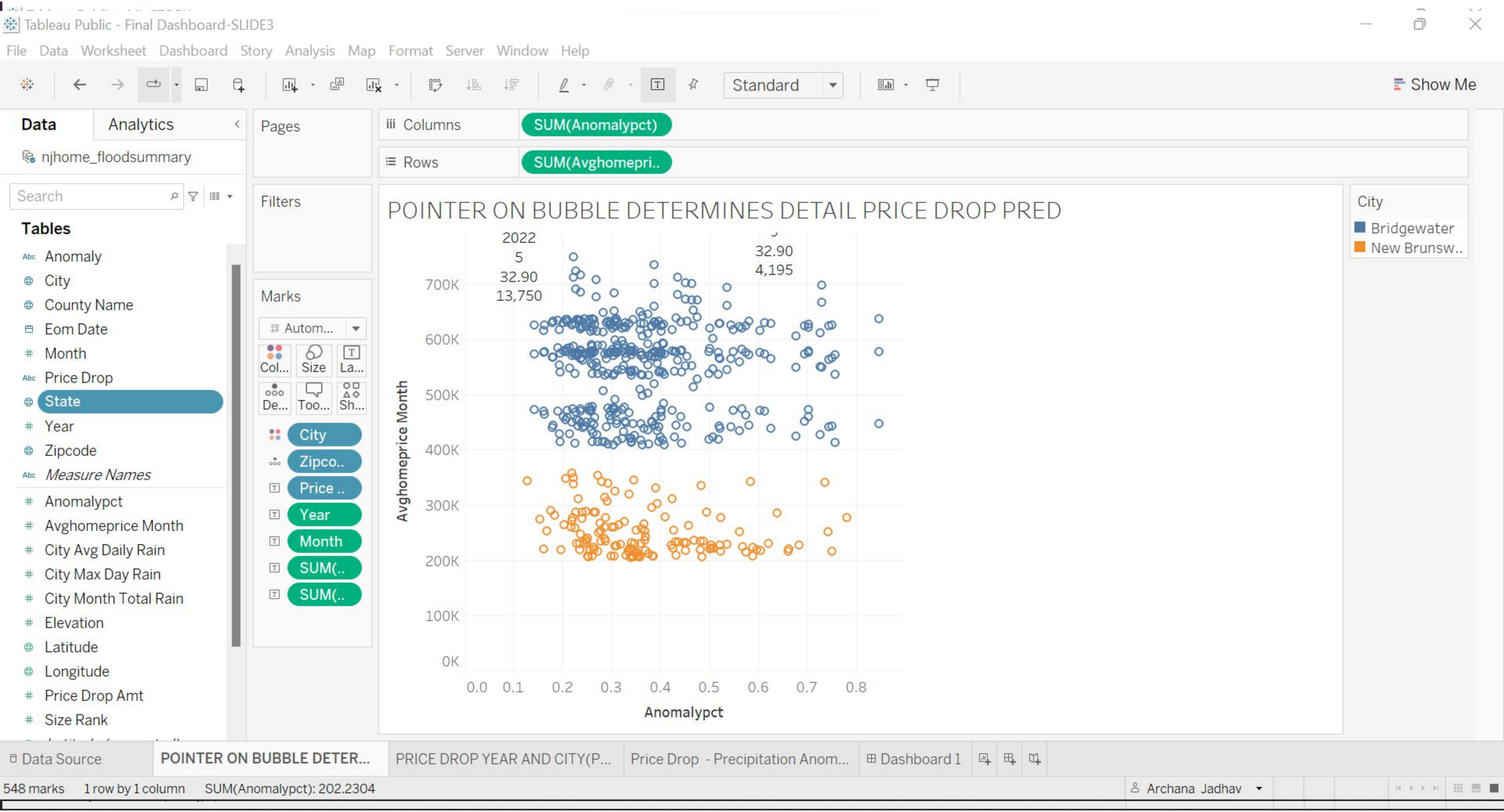


Tableau Public - Final Dashboard-SLIDE3
File Data Worksheet Dashboard Story Analysis Map Format Server Window Help
Standard
Show Me
Data
Analytics
njhome_floodsummary
Search
Tables
Anomaly
City
County Name
Eom Date
Month
Price Drop
State
Year
Zipcode
Measure Names
Anomalypct
Avghomeprice Month
City Avg Daily Rain
City Max Day Rain
City Month Total Rain
Elevation
Latitude
Longitude
Price Drop Amt
Size Rank
Pages
Filters
Marks
Autom...
Col...
Size
La...
De...
Too...
Sh...
City
Zipco..
Price ..
Year
Month
SUM(..
SUM(..
Columns
SUM(Anomalypct)
Rows
SUM(Avghomepri..
POINTER ON BUBBLE DETERMINES DETAIL PRICE DROP PRED
2022
5
32.90
13,750
32.90
4,195
700K
600K
500K
400K
300K
200K
100K
0K
Avghomeprice Month
0.0 0.1 0.2 0.3 0.4 0.5 0.6 0.7 0.8
Anomalypct
City
Bridgewater
New Brunsw..
Data Source
POINTER ON BUBBLE DETER...
PRICE DROP YEAR AND CITY(P...
Price Drop - Precipitation Anom...
Dashboard 1
548 marks
1 row by 1 column
SUM(Anomalypct): 202.2304
Archana Jadhav

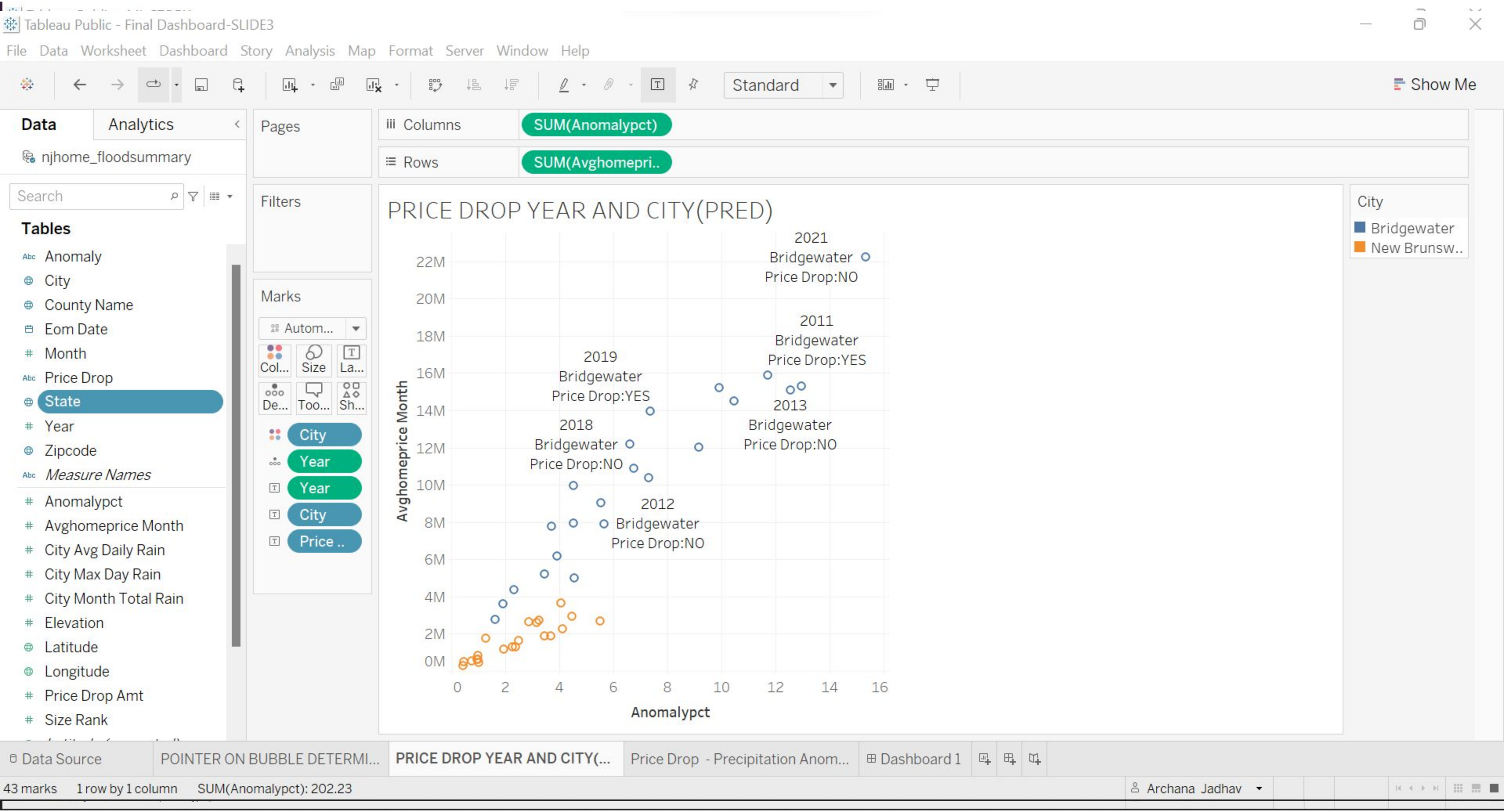


Tableau Public - Final Dashboard-SLIDE3
File Data Worksheet Dashboard Story Analysis Map Format Server Window Help
Standard
Show Me
Data
Analytics
njhome_floodsummary
Pages
Columns
SUM(Anomalypct)
Rows
SUM(Avghomepri..
Filters
Marks
Autom...
Col...
Size
La...
De...
Too...
Sh...
City
Year
Year
City
Price ..
Search
Tables
Anomaly
City
County Name
Eom Date
Month
Price Drop
State
Year
Zipcode
Measure Names
Anomalypct
Avghomeprice Month
City Avg Daily Rain
City Max Day Rain
City Month Total Rain
Elevation
Latitude
Longitude
Price Drop Amt
Size Rank
PRICE DROP YEAR AND CITY(PRED)
2021
Bridgewater
Price Drop:NO
2011
Bridgewater
Price Drop:YES
2019
Bridgewater
Price Drop:YES
2013
Bridgewater
Price Drop:NO
2018
Bridgewater
Price Drop:NO
2012
Bridgewater
Price Drop:NO
Avghomeprice Month
22M
20M
18M
16M
14M
12M
10M
8M
6M
4M
2M
0M
0
2
4
6
8
10
12
14
16
Anomalypct
City
Bridgewater
New Brunsw..
Data Source
POINTER ON BUBBLE DETERMI...
PRICE DROP YEAR AND CITY(...
Price Drop - Precipitation Anom...
Dashboard 1
43 marks
1 row by 1 column
SUM(Anomalypct): 202.23
Archana Jadhav

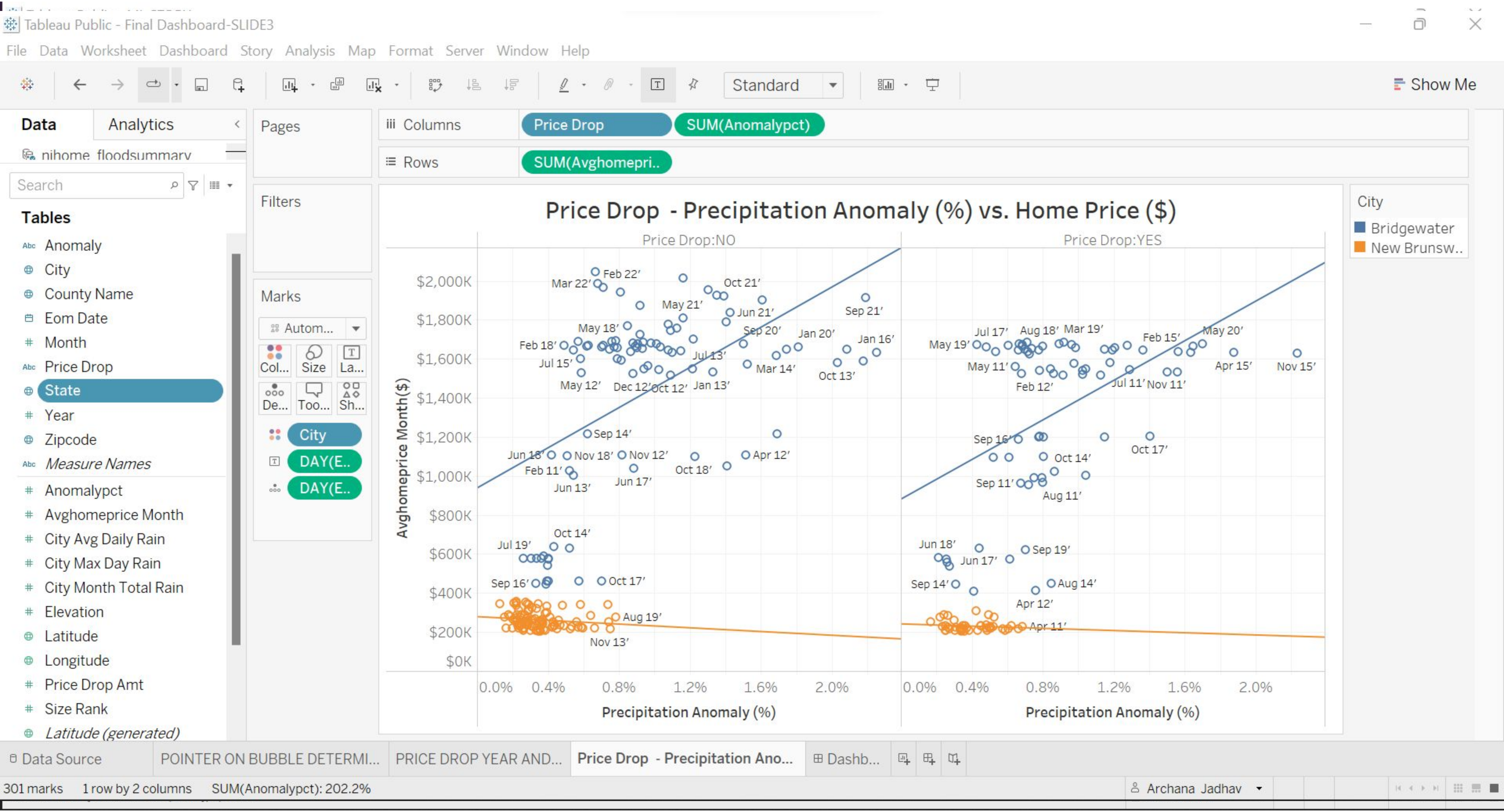


Tableau Public - Final Dashboard-SLIDE3
File Data Worksheet Dashboard Story Analysis Map Format Server Window Help
Standard
Show Me
Data
Analytics
nihome_floodsummary
Search
Tables
Anomaly
City
County Name
Eom Date
Month
Price Drop
State
Year
Zipcode
Measure Names
Anomalypct
Avghomeprice Month
City Avg Daily Rain
City Max Day Rain
City Month Total Rain
Elevation
Latitude
Longitude
Price Drop Amt
Size Rank
Latitude (generated)
Pages
Columns
Price Drop
SUM(Anomalypct)
Rows
SUM(Avghomepri..
Filters
Marks
Autom...
Col...
Size
La...
De...
Too...
Sh...
City
DAY(E..
DAY(E..
Price Drop - Precipitation Anomaly (%) vs. Home Price ($)
Price Drop:NO
Price Drop:YES
Avghomeprice Month($)
$2,000K
$1,800K
$1,600K
$1,400K
$1,200K
$1,000K
$800K
$600K
$400K
$200K
$0K
0.0% 0.4% 0.8% 1.2% 1.6% 2.0%
Precipitation Anomaly (%)
0.0% 0.4% 0.8% 1.2% 1.6% 2.0%
Precipitation Anomaly (%)
City
Bridgewater
New Brunsw..
Data Source
POINTER ON BUBBLE DETERMI...
PRICE DROP YEAR AND...
Price Drop - Precipitation Ano...
Dashb...
301 marks
1 row by 2 columns
SUM(Anomalypct): 202.2%
Archana Jadhav

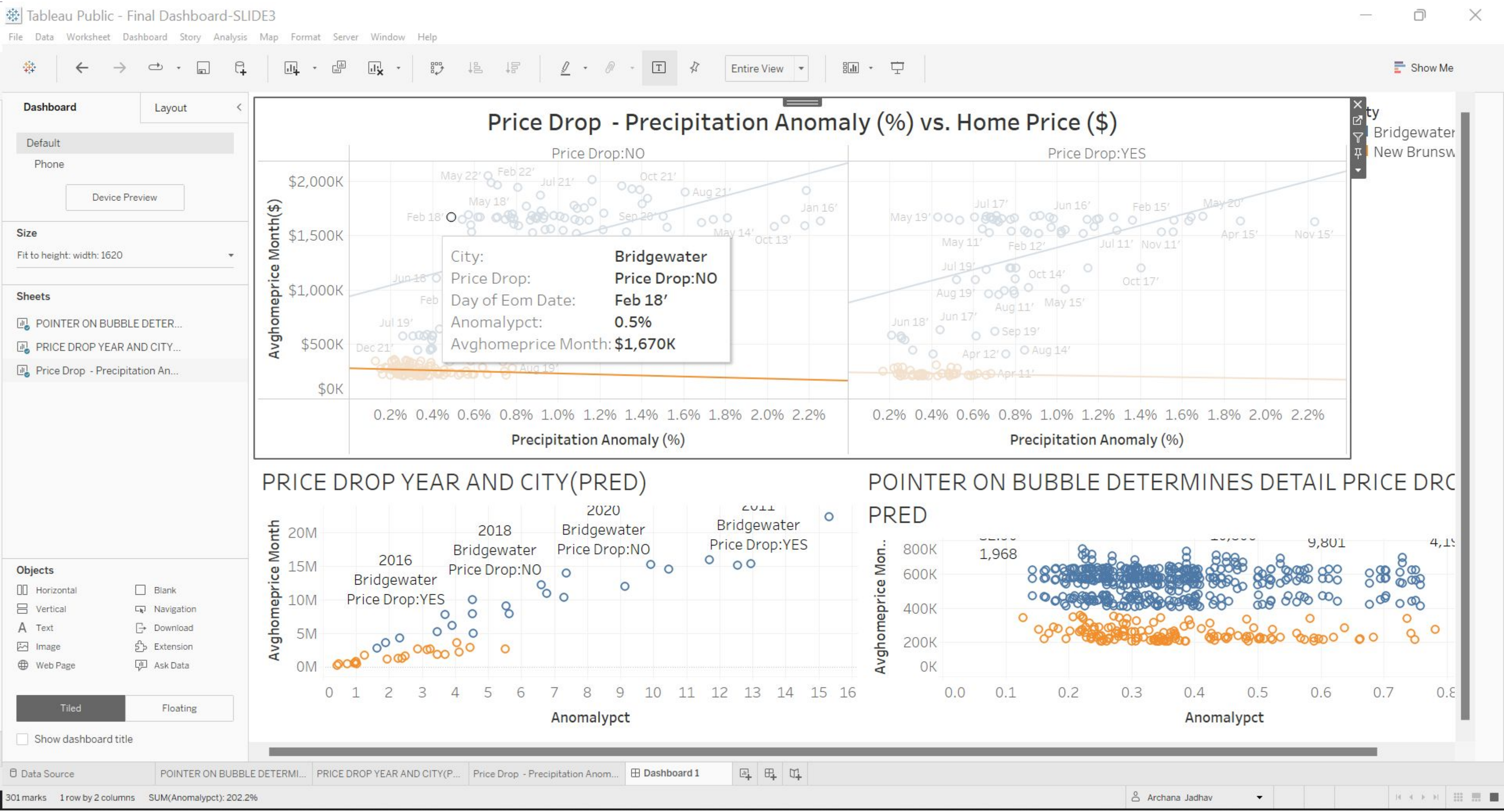


Tableau Public - Final Dashboard-SLIDE3
File Data Worksheet Dashboard Story Analysis Map Format Server Window Help
Entire View
Show Me
Dashboard
Layout
Default
Phone
Device Preview
Size
Fit to height: width: 1620
Sheets
POINTER ON BUBBLE DETER...
PRICE DROP YEAR AND CITY...
Price Drop - Precipitation An...
Objects
Horizontal
Blank
Vertical
Navigation
Text
Download
Image
Extension
Web Page
Ask Data
Tiled
Floating
Show dashboard title
Price Drop - Precipitation Anomaly (%) vs. Home Price ($)
Price Drop:NO
Price Drop:YES
Avghomeprice Month($)
$2,000K
$1,500K
$1,000K
$500K
$0K
City: Bridgewater
Price Drop: Price Drop:NO
Day of Eom Date: Feb 18'
Anomalypct: 0.5%
Avghomeprice Month: $1,670K
0.2% 0.4% 0.6% 0.8% 1.0% 1.2% 1.4% 1.6% 1.8% 2.0% 2.2%
Precipitation Anomaly (%)
Bridgewater
New Brunsw
PRICE DROP YEAR AND CITY(PRED)
Avghomeprice Month
20M
15M
10M
5M
0M
2016 Bridgewater Price Drop:YES
2018 Bridgewater Price Drop:NO
2020 Bridgewater Price Drop:NO
Bridgewater Price Drop:YES
0 1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16
Anomalypct
POINTER ON BUBBLE DETERMINES DETAIL PRICE DRO
PRED
Avghomeprice Mon..
800K
600K
400K
200K
0K
1,968
9,801
0.0 0.1 0.2 0.3 0.4 0.5 0.6 0.7 0.8
Anomalypct
Data Source
POINTER ON BUBBLE DETERMI...
PRICE DROP YEAR AND CITY(P...
Price Drop - Precipitation Anom...
Dashboard 1
301 marks
1 row by 2 columns
SUM(Anomalypct): 202.2%
Archana Jadhav