



Mid-term Project Question: Data-driven Decision NYC Property Sales

Markus Afonso



Project Requirements and Objectives

- Understand housing trends
- Identify key areas
- Optimizing future decisions

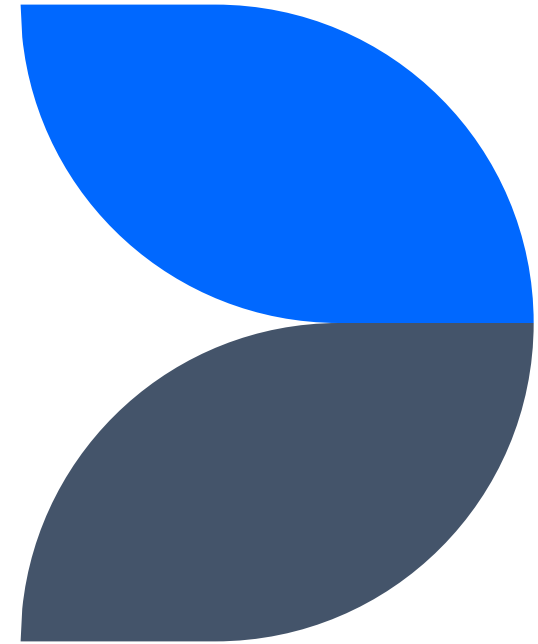
Translate objectives

- Analyze market trends
- Extract insights to improve decision making

Business/Research Understanding Phase

Preliminary Strategy

- Utilize NumPy and
- Visualization techniques




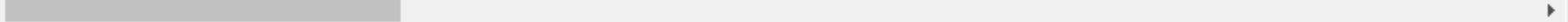
Collect data


- From <https://data.world/>

 **NYC_property_sales.csv**
[Request more info](#)

[View](#) [Download](#) [Filter](#) [Columns](#)

	 borough	 neighborhood	 building_class_category	 tax_class_at_present	#	block 	#	lot
1	Bronx	Bathgate	01 One Family Dwellings	1		3030		
2	Bronx	Bathgate	01 One Family Dwellings	1		3030		
3	Bronx	Bathgate	01 One Family Dwellings	1		3039		
4	Bronx	Bathgate	01 One Family Dwellings	1		3043		
5	Bronx	Bathgate	01 One Family Dwellings	1		3046		



 Showing 1-5 of 38,177 rows, 20 columns [See all](#)

[Switch to column overview](#)

Perform exploratory data analysis (EDA)

- Data for each property sold in NYC
- Average Sale Price by Building Class Category
- Total Residential Units Sold
- Total Commercial Units Sold
- Each subject has a unique GUID
- KPIs: # of Units, Land and Gross sq ft etc.

Assess data quality

- Consistent entries
- Includes all sales

Select interesting subsets

- Buildings
- Townhomes
- Duplex's

Data Visualization Phase

Preparing for modeling in subsequent phases

- Address any missing values or outliers
- Ensure the dataset is formatted correctly for analysis

```
data[(data['residential_units'].isin([1, 2, 3])) | (data['apartment_number'].notnull())]
```

Executed at 2024.02.22 01:02:51 in 13ms

	borough	neighborhood	building_class_category	tax_class_at_present	block	lot	building_class_at_present	
0	Bronx	Bathgate	01 One Family Dwellings	1	3030	62	A1	44
1	Bronx	Bathgate	01 One Family Dwellings	1	3030	70	A1	44
2	Bronx	Bathgate	01 One Family Dwellings	1	3039	63	A1	46
3	Bronx	Bathgate	01 One Family Dwellings	1	3043	55	A1	19
4	Bronx	Bathgate	01 One Family Dwellings	1	3046	35	A1	26
5	Bronx	Bathgate	01 One Family Dwellings	1	3046	39	A1	26
6	Bronx	Bathgate	01 One Family Dwellings	1	3053	62	S1	23
7	Bronx	Bathgate	02 Two Family Dwellings	1	2912	152	B1	51
8	Bronx	Bathgate	02 Two Family Dwellings	1	2927	119	B1	14
9	Bronx	Bathgate	02 Two Family Dwellings	1	2929	119	B1	38

Select cases and variables appropriate for analysis

- Sale Price
- Building Class Category
- Year Built
- Geographic regions.

Perform transformation of certain variables, if needed

- Clean building class category

```
1 data['building_class_category'] = data['building_class_category'].str.split(' ').str[0]
2 data
```

Executed at 2024.02.22 01:13:01 in 19ms

	borough	neighborhood	building_class_category	tax_class_at_pres
0	Bronx	Bathgate	01	1
1	Bronx	Bathgate	01	1
2	Bronx	Bathgate	01	1
3	Bronx	Bathgate	01	1
4	Bronx	Bathgate	01	1
5	Bronx	Bathgate	01	1
6	Bronx	Bathgate	01	1
7	Bronx	Bathgate	02	1
8	Bronx	Bathgate	02	1

- Calculate \$ per sq ft

```
data['price_per_sqft'] = data['sale_price'] / data['gross_square_feet']
data
```

Executed at 2024.02.22 01:17:09 in 40ms

38177 rows x 21 columns						
id	year_built	tax_class_at_time_of_sale	building_class_at_time_of_sale	sale_price	sale_date	price_per_sqft
1470	1899	1 A1	455000	2018-11-28T00:00:00Z	309.523810	
1497	1899	1 A1	388500	2019-07-23T00:00:00Z	259.519038	
1296	1910	1 A1	419000	2018-12-20T00:00:00Z	323.302469	
2047	1901	1 A1	470000	2019-03-08T00:00:00Z	229.604299	
1560	1899	1 A1	445000	2018-12-17T00:00:00Z	285.256410	
1744	1899	1 A1	430000	2018-05-24T00:00:00Z	312.500000	



Thank you