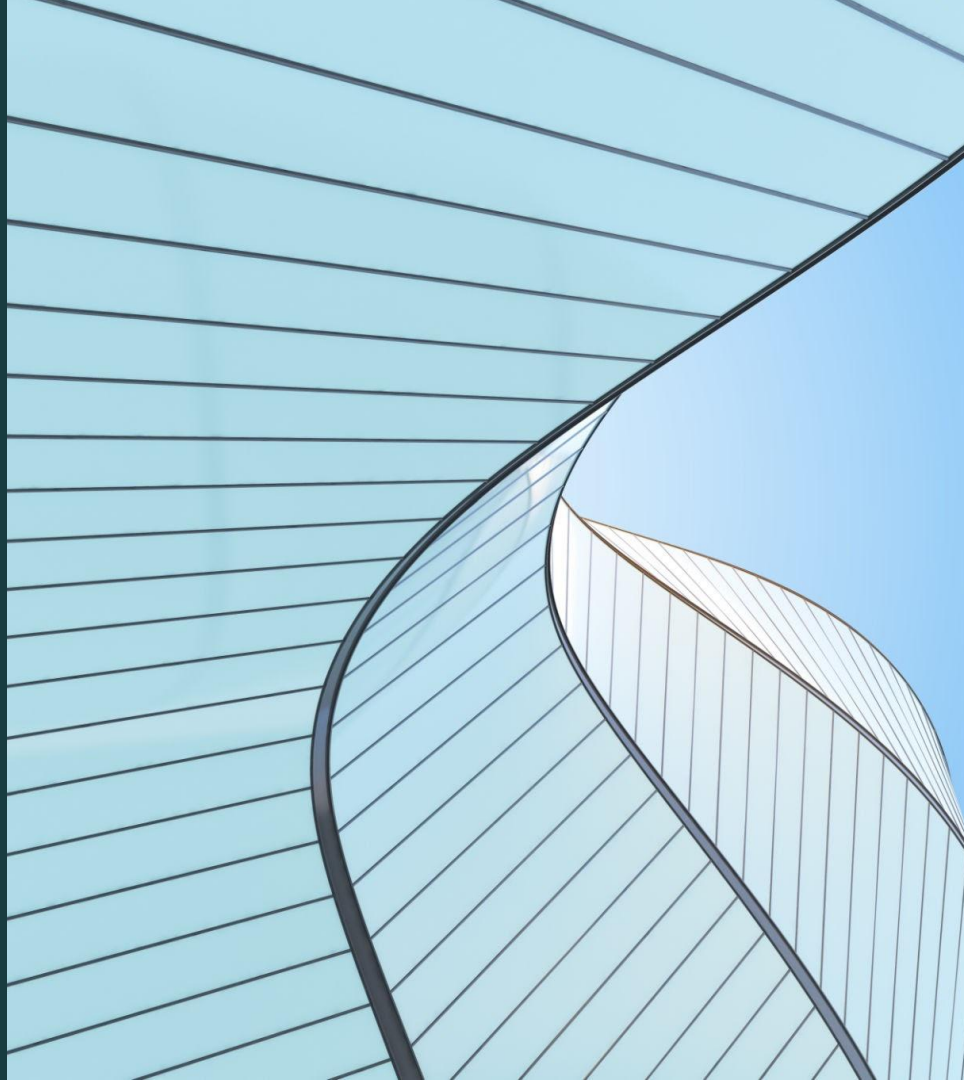
Tufts University
Fall 2025

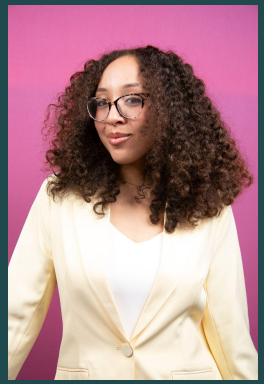
Foundations of Data Analytics

DATA 200 Final Project

Ezey Duru and Lily Davis



Our *team*



Ezey Duru

Data Scientist and Analyst

Ezenwanyi.Duru@tufts.edu

Lily Davis

Data Scientist and Analyst

Lily.Davis@tufts.edu

Agenda

1
Motivation

2
Our Investigation

3
Methodology

4
Main Linear Regression

5
Borough Linear Regression

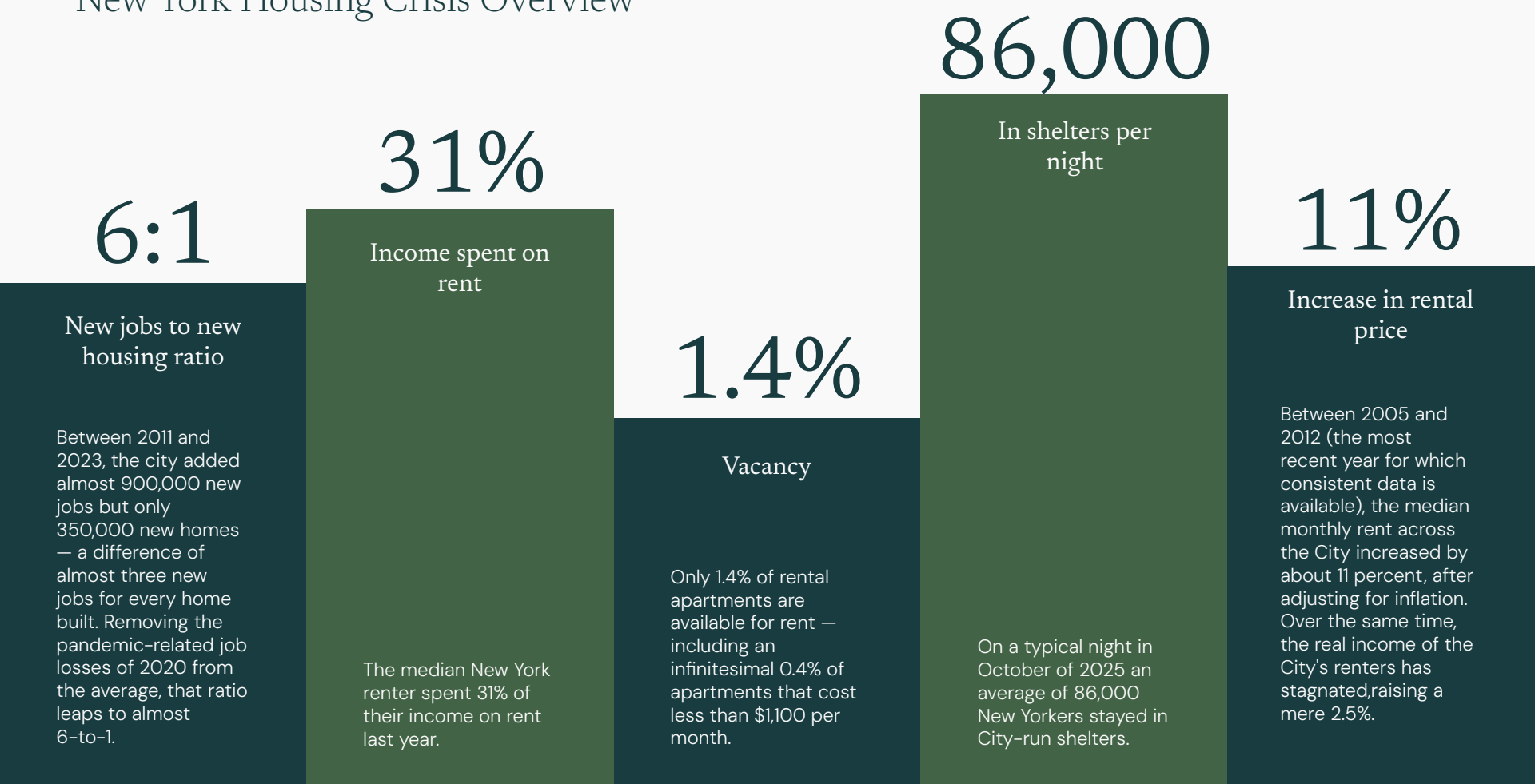
6
Conclusion

Motivation

New York City Housing Crisis

New York City is currently facing its worst housing affordability crisis in a century. Rent is increasing faster than income is increasing, and people are having to spend a major portion of their income on rent leaving very little money for basic necessities. More jobs are being created than new housing, and generally there is a lack of available-to-rent housing.

New York Housing Crisis Overview



Our Investigation

Research Question

How do number of new businesses, schools (public, private, and independent), subway stations, grocery stores, population, crime rate, unemployment rate, education rate, and median household income affect property value by zip code in New York City's 5 boroughs?

Our Model

$$\text{PropertyValue}_i = \beta_0 + \beta_1(\text{NewBusinesses}_i) + \beta_2(\text{PublicSchools}_i) + \beta_3(\text{NonPublicSchools}_i) + \beta_4(\text{Population}_i) + \beta_5(\text{TransitAccess}_i) + \beta_6(\text{CrimeRate}_i) + \beta_7(\text{GroceryStores}_i) + \beta_8(\text{Education}_i) + \beta_9(\text{Unemployment}_i) + \beta_{10}(\text{Income}_i) + \epsilon_i$$

*i represents each individual zip code. The data is at zip-code-level.

Methodology

Data Description

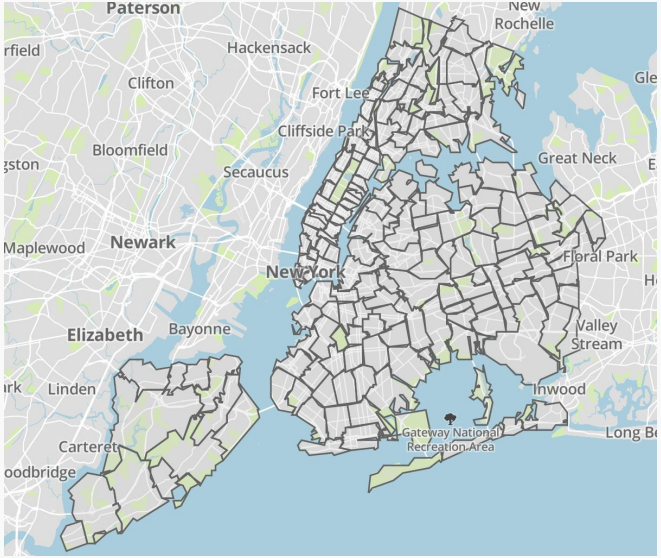
Variable	Description
Property Value*	Average property value in the year 2024
New Businesses	Count of new businesses in the year 2024
Public Schools	Count of public schools
Non-public Schools	Private and independent schools
Population	Number of people
Transit Access	Subway station count
Crime Rate	Crime count in the year 2024
Grocery Stores	Count of licensed retail food stores
Education	% Population 25 Years and Over: Bachelor's Degree or More
Unemployment	% Civilian Population in Labor Force 16 Years and Over: Unemployed
Income	Median household income

*Dependent variable

Cleaning and Data Preparation

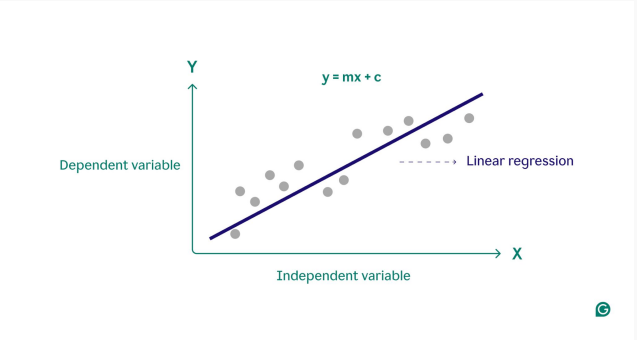
Identify all the **zip codes in the 5 boroughs**

Joined predictors by ZIP codes to create **master data set**



Why Multiple Linear Regression

- All **numeric and continuous** variables
- Goal to **quantify how much property values change** when neighborhood characteristics change
- **Explanation rather than prediction**
- Aggregated socioeconomic variables typically show **stable, monotonic relationships** with housing prices

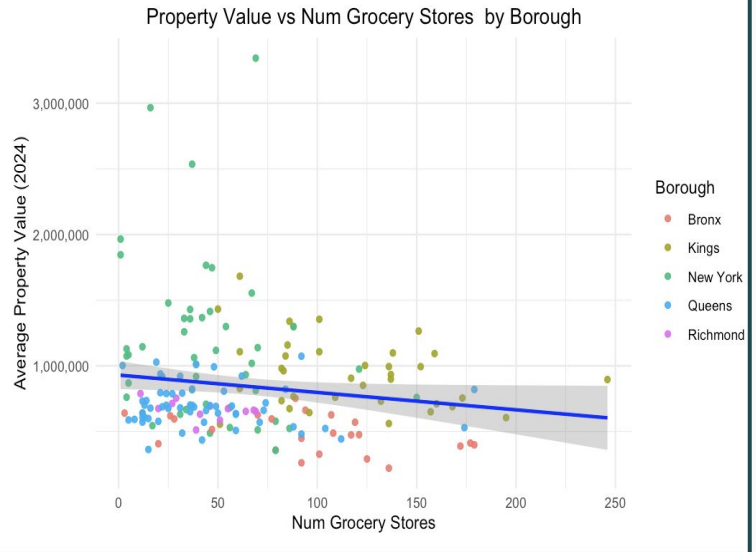


Linear Regression Plan

1. Run linear regression including **all predictors** on **all zip codes**

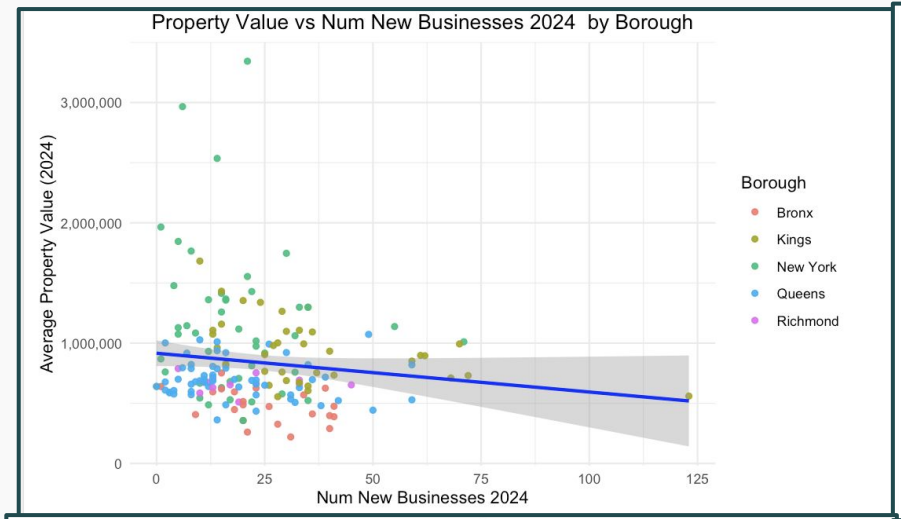
2. Run the same model **by borough**
    a. Group data by borough
    b. Run model 5 times (once per borough)

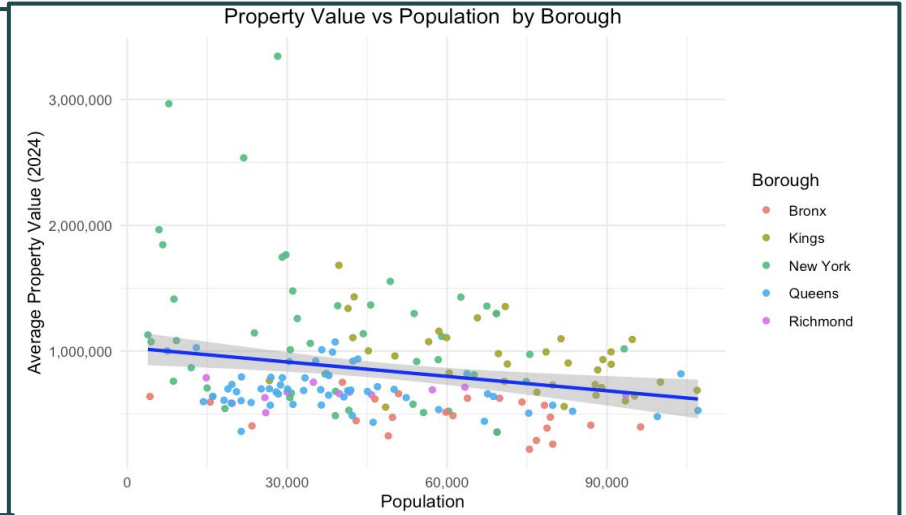
Main Linear Regression

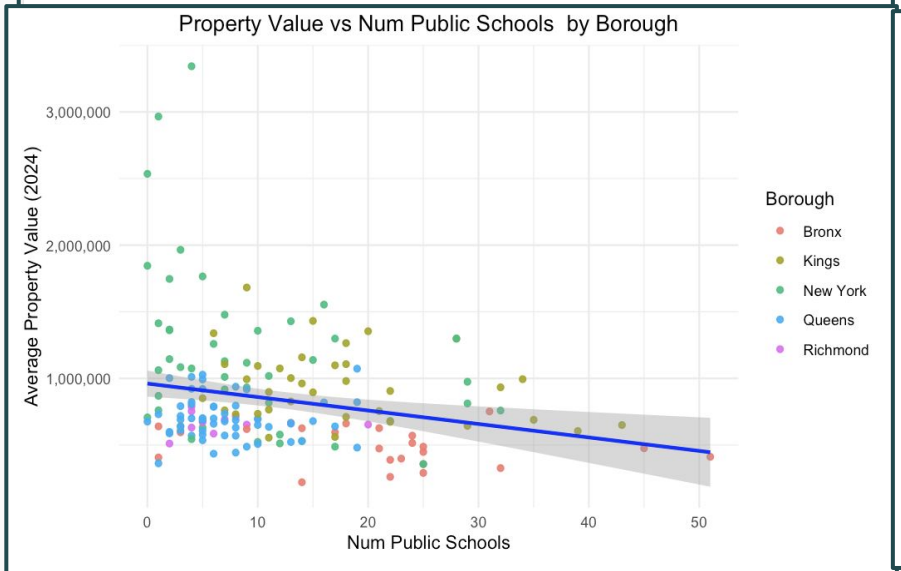


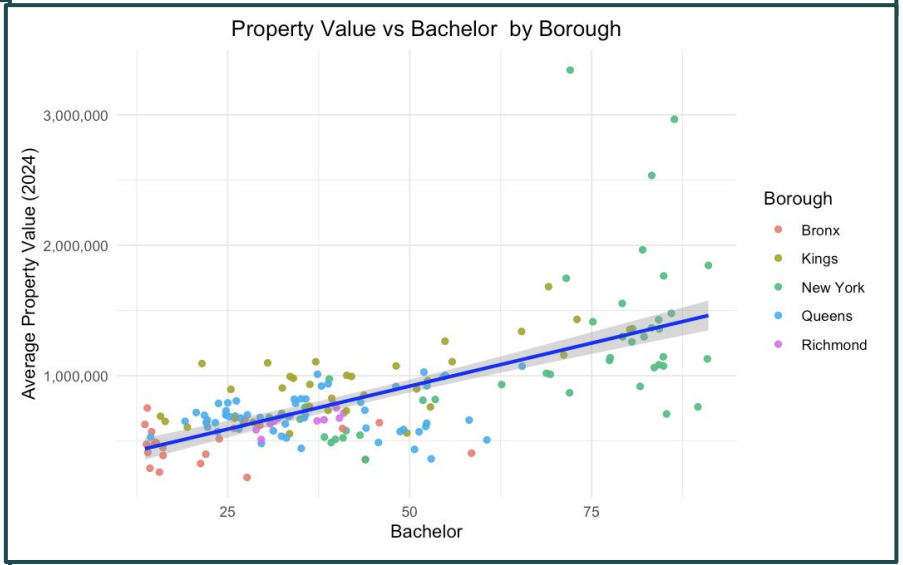
Key Predictors of Property Values in NYC

Variable	Coefficient	P-Value
(Intercept)	42443.075	0.76487
num_subway_stations	22274.685	0.06245
income	***5.191***	***0.00000276***
num_new_businesses_2024	-4350.187	0.03388
population	-3.822	0.04772
num_public_schools	-8008.954	0.0494
num_non_public_schools	6291.650	0.0869
num_crimes_2024	2524.763	0.05585
num_grocery_stores	***3175.642***	***0.00227***
bachelor	4174.100	0.04525
unemployed	9703.365	0.30214


Property Value vs Num New Businesses 2024 by Borough
Average Property Value (2024)
3,000,000
2,000,000
1,000,000
0
0
25
50
75
100
125
Num New Businesses 2024
Borough
Bronx
Kings
New York
Queens
Richmond


Property Value vs Population by Borough
Average Property Value (2024)
3,000,000
2,000,000
1,000,000
0
30,000
60,000
90,000
Population
Borough
Bronx
Kings
New York
Queens
Richmond


Property Value vs Num Public Schools by Borough
Average Property Value (2024)
3,000,000
2,000,000
1,000,000
0
10
20
30
40
50
Num Public Schools
Borough
Bronx
Kings
New York
Queens
Richmond


Property Value vs Bachelor by Borough
Average Property Value (2024)
3,000,000
2,000,000
1,000,000
25
50
75
Bachelor
Borough
Bronx
Kings
New York
Queens
Richmond

Non-Significant Indicators of Property Values in NYC

Across all 10 indicators, **subway** access, **non-public schools**, **crime** levels, and **unemployment** rates showed no statistically significant relationship with property values

Borough Linear Regression

Multi Linear Regression by Borough

- Drivers of property value may vary by borough.
- Variable significance changes once each borough is analyzed on its own.

Statistical Significance by Borough

subway stations (positive), **new businesses** (negative), **public schools** (negative), **income** (positive).

no statistically significant relationship

no statistically significant relationship

population levels (negative).

grocery store (positive) and **new businesses** (negative).

Conclusion

Limitations & Improvements

Limitations

- Data sourced from multiple platforms
- Potential sampling and coverage differences
- Single-year snapshot

Improvements

- Incorporate time-series data
- Build predictive models
  - Forecast future property value shifts
- Add more neighborhood-level variables
  - Better capture what drives property value
- Spatial scale

Conclusion

Significant drivers of property values include:

income,
grocery store access,
population,
new businesses,
public schools, and
bachelor's-level education

The direction and level of impact varies across boroughs

Thank you

Appendix

References

https://www.vitalcitynyc.org/articles/new-yorks-housing-crisis-self-inflicted-and-solvable

https://www.nyc.gov/site/housing/about/problem.page

https://www.nytimes.com/2023/10/25/nyregion/nyc-housing-crisis-plan.html

https://www.jstor.org/stable/25067439?seq=1