



 BrainStation®

HOUSING MARKET AND LIVABILITY

Developing a model that evaluates the adequate price of housing units based on the livability score

HAESUN JUNG





Scarborough neighbourhood becomes food desert after closure of grocery stores



GTA

Toronto the grey: the city's green space is disappearing, and it's likely to get worse

From physical health and climate-saving effects, to being a 'lifeline' for renters, advocates say politicians need to focus on green space before it's too late.

RESEARCH • SPATIAL PLANNING

By May Warren Housing Reporter

Sunday, June 25, 2023 | 5 min to read

By Tina Yazdani

Posted Aug 4, 2023, 5:07PM EDT. Last Updated Aug 4, 2023, 6:27PM EDT.

"Yes, but not like this!" – Why densification often lacks public acceptance

Densification is a fundamental principle of urban planning and development today. Nevertheless, it often encounters local resistance. A group of ETH researchers has now systematically investigated factors influencing public acceptance of densification, focusing on the canton of Zurich and six cities of global importance. One key factor: affordable housing.

14.09.2022 by Florian Meyer

Share

Problem

34%

Of canadians said the lack of available public transit was a pain point in their every-day activities

23% Said lack of greenspace

23% Said lack of safe bike lanes and walkways



DATA PROCESSING:



BASE DATASET: REAL ESTATE SALES

**RAW DATASET CONTAINS: 997K DATAPOINTS
TIMESPAN: 2001-2020
PLACE: CONNECTICUT**

Columns with negligible missing values :

- Serial Number
- Year Listed
- Date Recorded
- Town
- Address
- Assessed Value
- Sale Amount
- Sales Ratio

Columns with considerable missing values :

- Property Type
- Residential Type
- Non Use Code
- Assessor Remarks
- OPM remarks
- Location (coordinates)

PROGRESS ON DATA PROCESSING

1. Non Use Code

a. One-hot-encoding (32 values)

2. Assessor Remarks & OPM remarks

a. Vectorization as there are 70000 string values in total

3. Location

a. Clean the address column and use Open Street Map API to obtain coordinates



MODELS PERFORMED:

Baseline Regression Model:

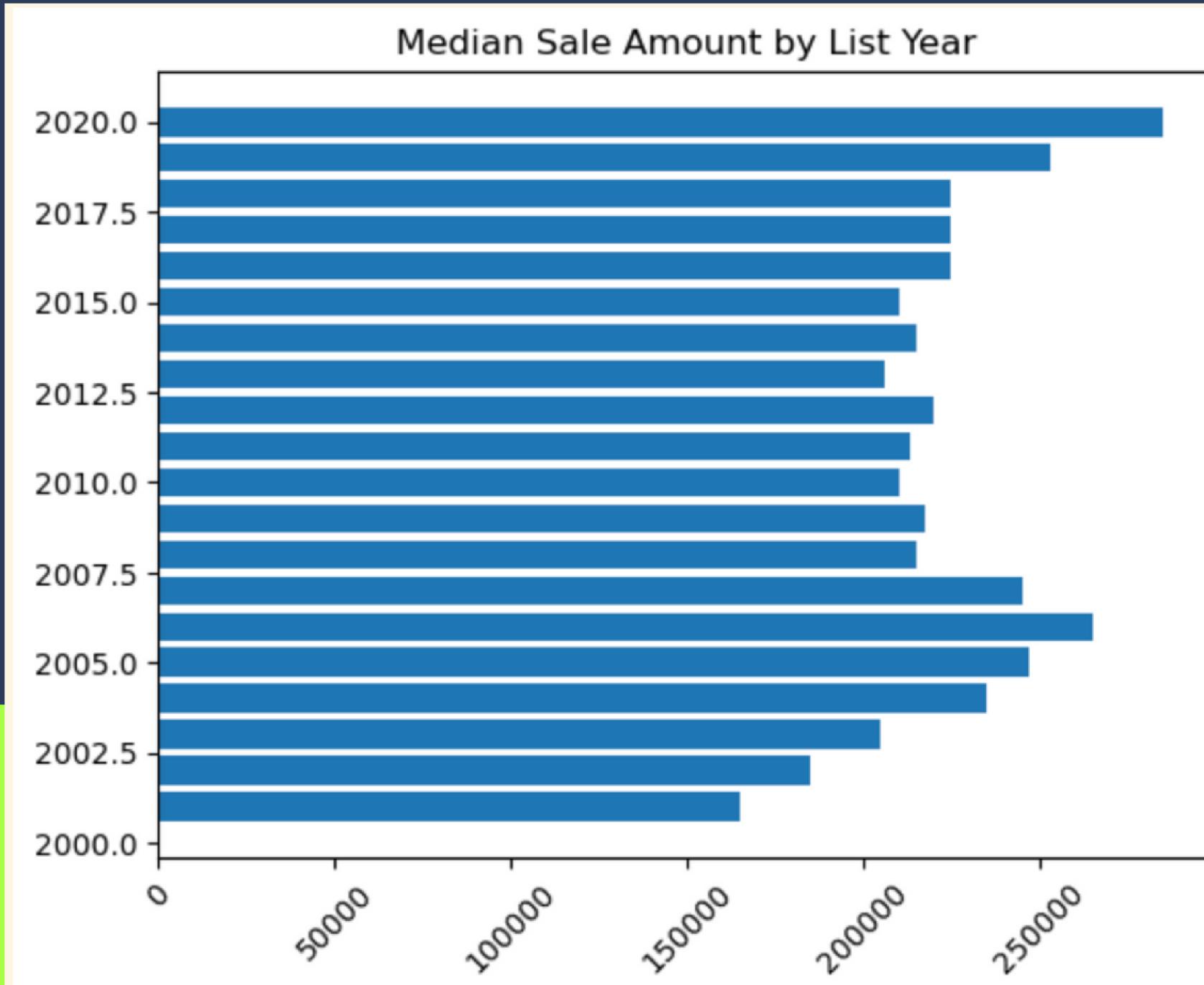
Using data from the original datasets, R² of 0.10. Used Year Listed, Town, Residential Type, Location (coordinates), and Non Use Code, all one-hot-encoded.

Baseline K-Best Analysis:

The top ten factors of highest coefficients are: 'latitude', 'longitude', 'List Year_2020', 'Town_Darien', 'Town_Greenwich', 'Town_New Canaan', 'Town_Stamford', 'Town_Westport', 'Property Type_Apartments', 'Property Type_Commercial'.

Shortcomings:

Lack of unit size information
Significance currently low due to a large number of missing coordinates.



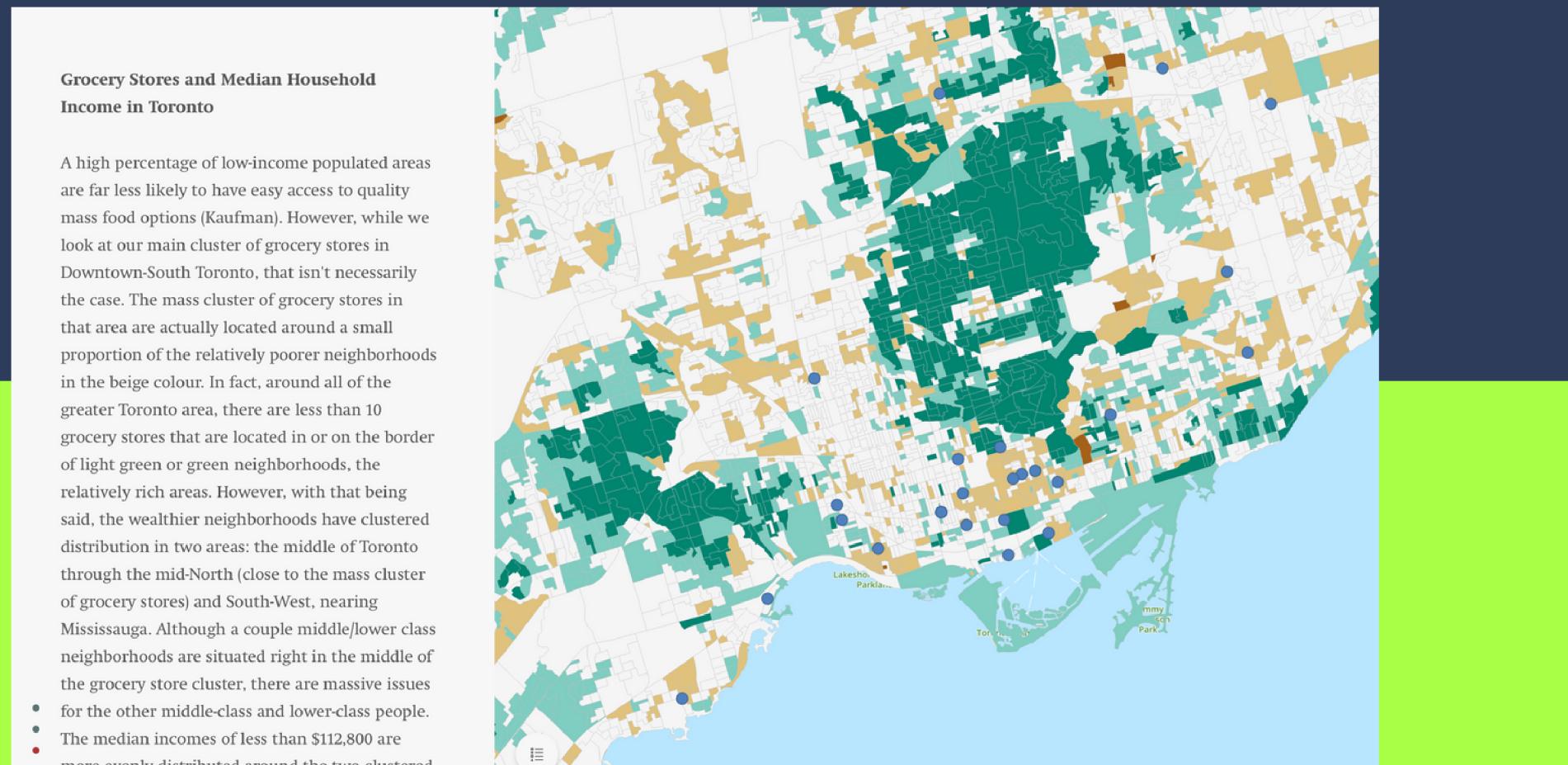
WHAT TO DO NEXT:

Feature engineering using [Year] on the mortgage rate, housing price inflation

Possibly use SQL to import mortgage and inflation rates from other datasets



WHAT TO DO NEXT:



Count-vectorize the string remark columns

Use API to get more coordinates of the housing units, green space, healthy food source, public transportation, etc. Visualize the result using map (see example)

Use other methods of supervised/ unsupervised learning

https://storymaps.arc

gis.com/stories/97351

b99eda44d7b976efdd

7b4f1a0f7

