



# Battle of the Neighborhoods

---

Analysis of the effect of Venues on Real estate prices

# Introduction

---

Purpose is to investigate effect of surrounding venues on price of real estate.

Target audience:

- Buyers
- Sellers
- Real estate developers.



# Description of data

Average real estate price (propertyshark.com)

	Area	Neighborhood	AvgPrice
0	Brooklyn	Bedford-Stuyvesant	750000
1	Brooklyn	Boerum Hill	1.69e+06
2	Brooklyn	Brooklyn Heights	2.15e+06
3	Brooklyn	Bushwick	967000
4	Brooklyn	Carroll Gardens	1.51e+06

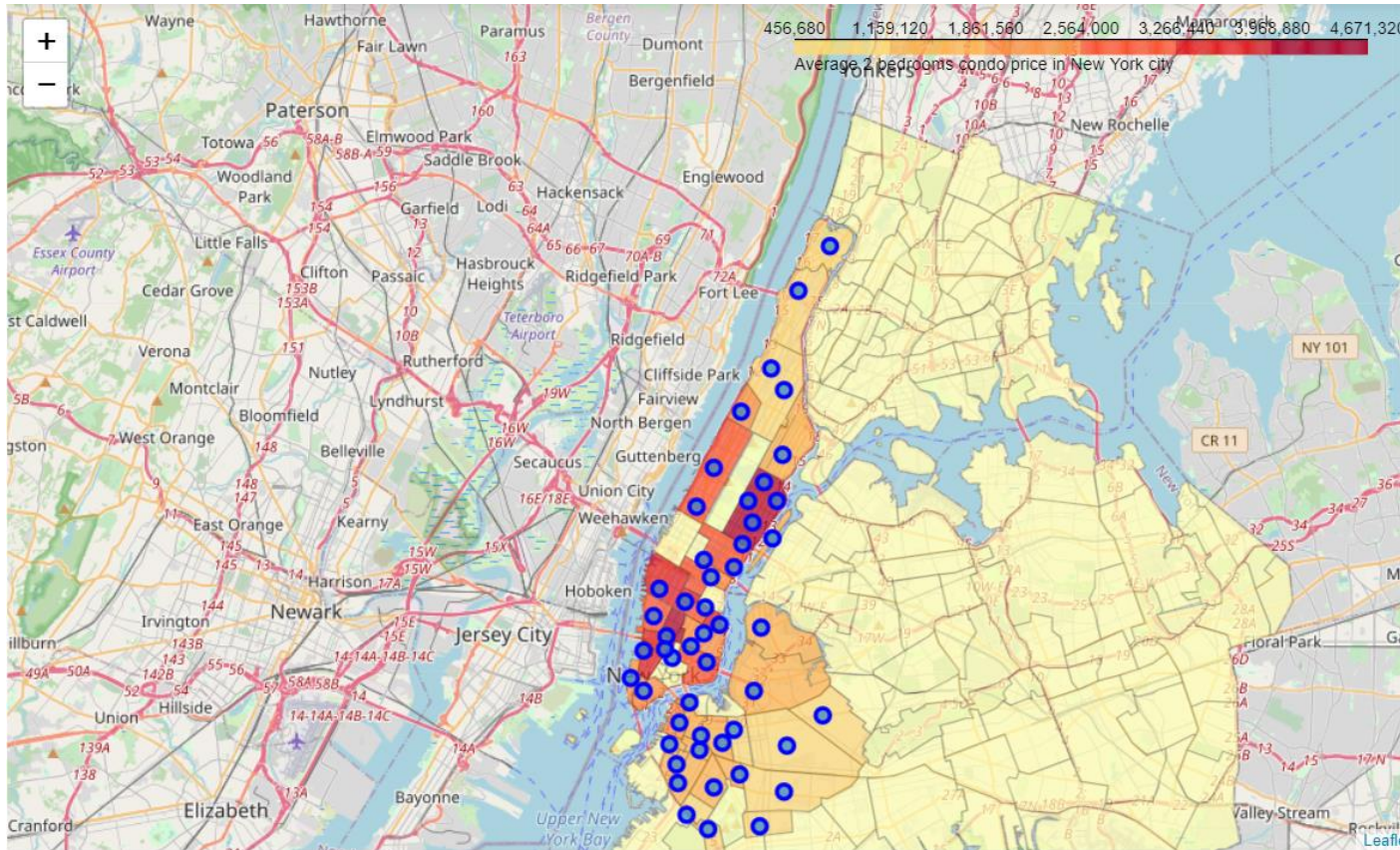
Latitude and longitude (Foursquare)

	Borough	Neighborhood	Latitude	Longitude
0	Manhattan	Marble Hill	40.876551	-73.910660
1	Brooklyn	Bay Ridge	40.625801	-74.030621
2	Brooklyn	Bensonhurst	40.611009	-73.995180
3	Brooklyn	Sunset Park	40.645103	-74.010316
4	Brooklyn	Greenpoint	40.730201	-73.954241

Joined dataframes

	Neighborhood	AvgPrice	Latitude	Longitude
0	Bedford-Stuyvesant	750000	40.687232	-73.941785
1	Boerum Hill	1.69e+06	40.685683	-73.983748
2	Brooklyn Heights	2.15e+06	40.695864	-73.993782
3	Bushwick	967000	40.698116	-73.925258
4	Carroll Gardens	1.51e+06	40.680540	-73.994654

# Methodology



Choropleth map showing venues in different neighborhoods in New York City.

# Regression analysis

- Simple linear regression is applied to the data.
- MSE and R2 score give insight in correlation

```
# Let's see how well Linear Regression fit the problem
y_pred = lreg.predict(X_test)

print('R2-score:', r2_score(y_test, y_pred)) # r2_score
print('Mean Squared Error:', mean_squared_error(y_test, y_pred)) # mse

print('Max positive coefs:', lreg.coef_[np.argsort(-lreg.coef_)[:10]])
print('Venue types with most postive effect:', X.columns[np.argsort(-lreg.coef_)[:10]].values)
print('Max negative coefs:', lreg.coef_[np.argsort(lreg.coef_)[:10]])
print('Venue types with most negative effect:', X.columns[np.argsort(lreg.coef_)[:10]].values)
coef_abs = abs(lreg.coef_)
print('Min coefs:', lreg.coef_[np.argsort(coef_abs)[:10]])
print('Venue types with least effect:', X.columns[np.argsort(coef_abs)[:10]].values)
```

R2-score: -0.0035500373879067126  
Mean Squared Error: 0.35125220645735017  
Max positive coefs: [0. 0. 0. 0. 0. 0. 0. 0. 0. 0.]  
Venue types with most postive effect: ['Athletics & Sports' 'Insurance Office' 'Kitchen Supply Store' 'Lawyer'  
'Liquor Store' 'Miscellaneous Shop' 'Moving Target' 'Music Venue'  
'Other Repair Shop' 'Outdoors & Recreation']  
Max negative coefs: [0. 0. 0. 0. 0. 0. 0. 0. 0. 0.]  
Venue types with most negative effect: ['Athletics & Sports' 'Insurance Office' 'Kitchen Supply Store' 'Lawyer'  
'Liquor Store' 'Miscellaneous Shop' 'Moving Target' 'Music Venue'  
'Other Repair Shop' 'Outdoors & Recreation']  
Min coefs: [0. 0. 0. 0. 0. 0. 0. 0. 0. 0.]  
Venue types with least effect: ['Athletics & Sports' 'Insurance Office' 'Kitchen Supply Store' 'Lawyer'  
'Liquor Store' 'Miscellaneous Shop' 'Moving Target' 'Music Venue'  
'Other Repair Shop' 'Outdoors & Recreation']

# Result, Discussion and Conclusion

## Result

- R2 score -0.0035
- MSE 0.351

## Discussion

- Dataset might be too small for conclusion
- Larger dataset should be used
- Multiple linear regression might be more effective

## Conclusion

- Surrounding venues have weak effect on prices