Analytics Project Symposium - Fall 2016

Analytics Project: Real Estate Evaluation Based on Crime & Complaints

Team: Pratik Sunil Rane(psr280)
Glancy Cajitan Rodrigues(gcr253)
Shivraj Patil(srp468)

Abstract: Study on how real estate prices are affected by crimes (Emergency) & complaints (Non-Emergency) in New York City.

Slide 1

Real Estate Evaluation Based on Crime & Complaints

Motivation

Who are the users of this analytic? Apartment seekers

Who will benefit from this analytic? Apartment seekers who are unaware of the property rates, safety and conditions of neighborhoods

Why is this analytic important?

The analytic suggests an apartment seeker the list good neighborhoods in each of the boroughs of NYC from which he/she can select to live in one of them

Real Estate Evaluation Based on Crime & Complaints

Goodness

What steps were taken to assess the 'goodness' of the analytic?

The analytics was ran on multiple years of data and most of the areas that were predicted to be the best areas in an year remained in the list of the results in the next years.

Which implies that a user can rely on our analytic result and can decide the area to rent or buy an apartment

Real Estate Evaluation Based on Crime & Complaints

Data Sources

Name: Real Estate DatasetReal Estate Dataset

Description: It consists of the location in terms of neighborhood as a common attribute for comparison, and the unique attribute is the market rate per sq ft.

Size of data: 3 MB

Name: NYPD 7 Major Felony Incidents Dataset

Description: It consists of the location in terms of precincts as a common attribute for comparison, and the unique attribute is the type of the crime.
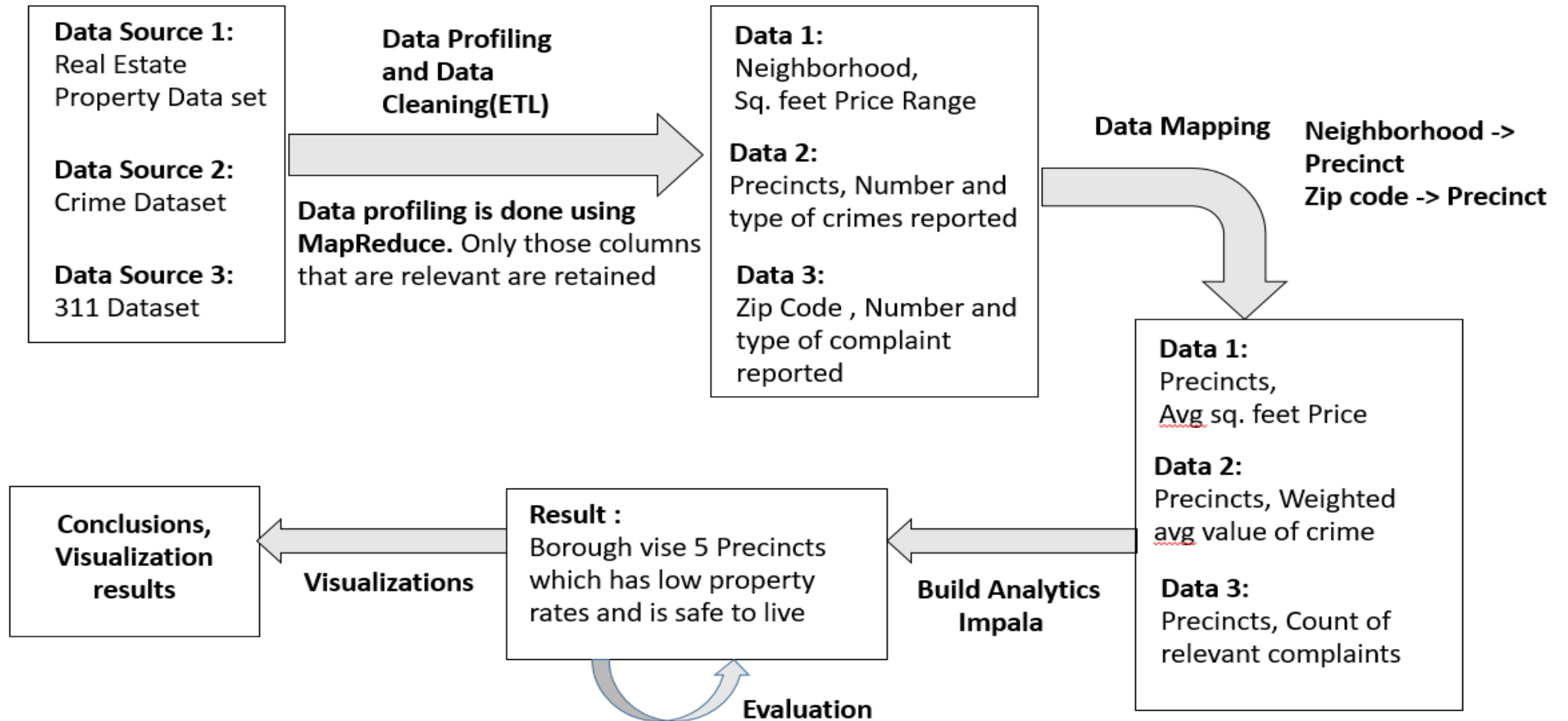
Size of data: 200 MB

Name: 311 Service Requests Dataset

Description: It consists of the location in terms of zip code as a common attribute for comparison, and the unique attribute is the type of the complaint request.

Size of data: 1.5 GB

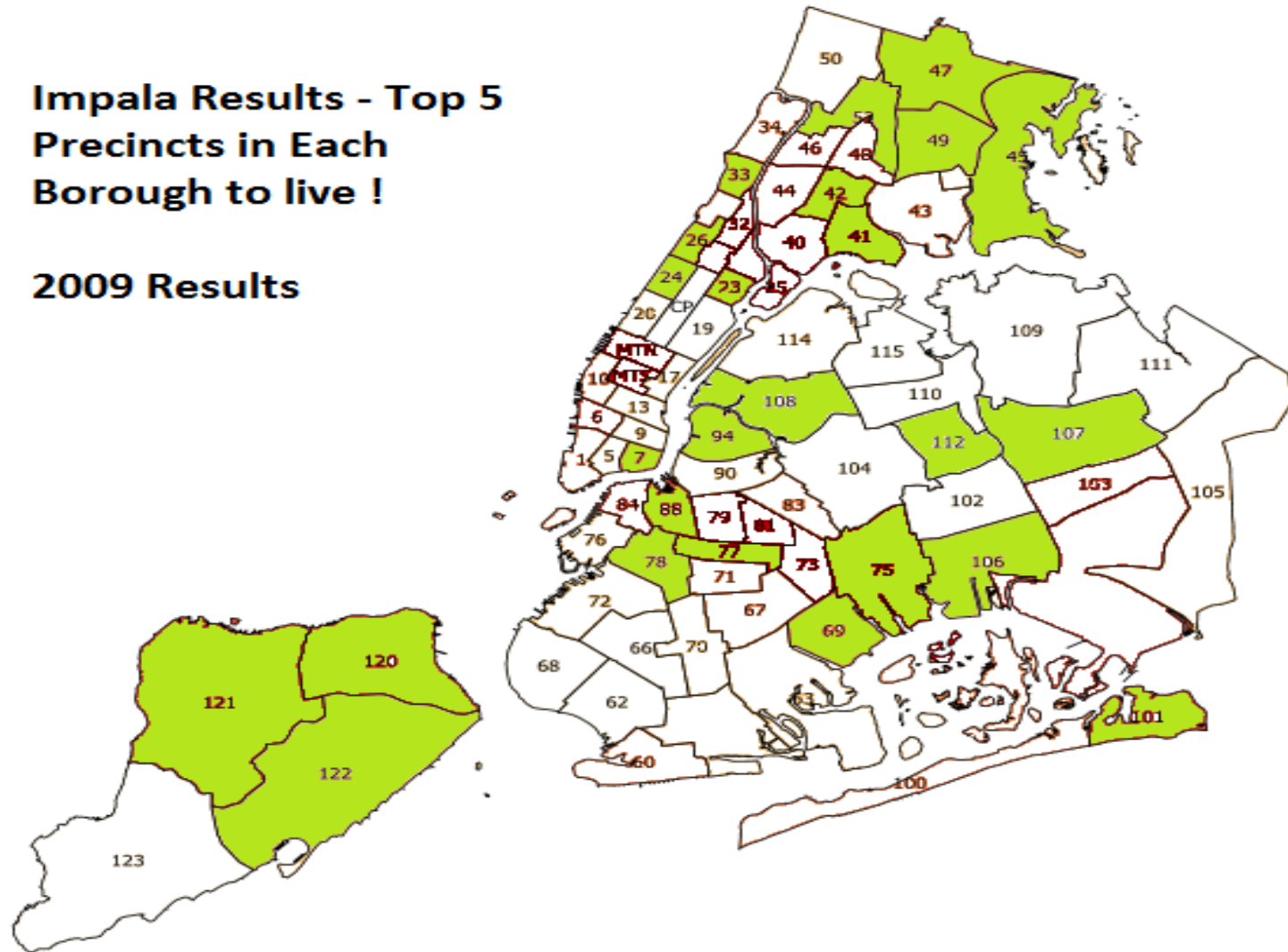# Real Estate Evaluation Based on Crime & Complaints

Design Diagram

Data Source 1:
Real Estate Property Data set

Data Source 2:
Crime Dataset

Data Source 3:
311 Dataset

Data Profiling and Data Cleaning(ETL)

Data profiling is done using MapReduce. Only those columns that are relevant are retained

Data 1:
Neighborhood, Sq. feet Price Range

Data 2:
Precincts, Number and type of crimes reported

Data 3:
Zip Code , Number and type of complaint reported

Data Mapping   Neighborhood -> Precinct Zip code -> Precinct

Data 1:
Precincts, Avg sq. feet Price

Data 2:
Precincts, Weighted avg value of crime

Data 3:
Precincts, Count of relevant complaints

Build Analytics Impala

Result :
Borough vise 5 Precincts which has low property rates and is safe to live

Evaluation

Visualizations

Conclusions, Visualization results

Platform(s) on which the analytic ran: Cloudera VM

Slide 5

Real Estate Evaluation Based on Crime & Complaints

Results

Impala Results - Top 5 Precincts in Each Borough to live !
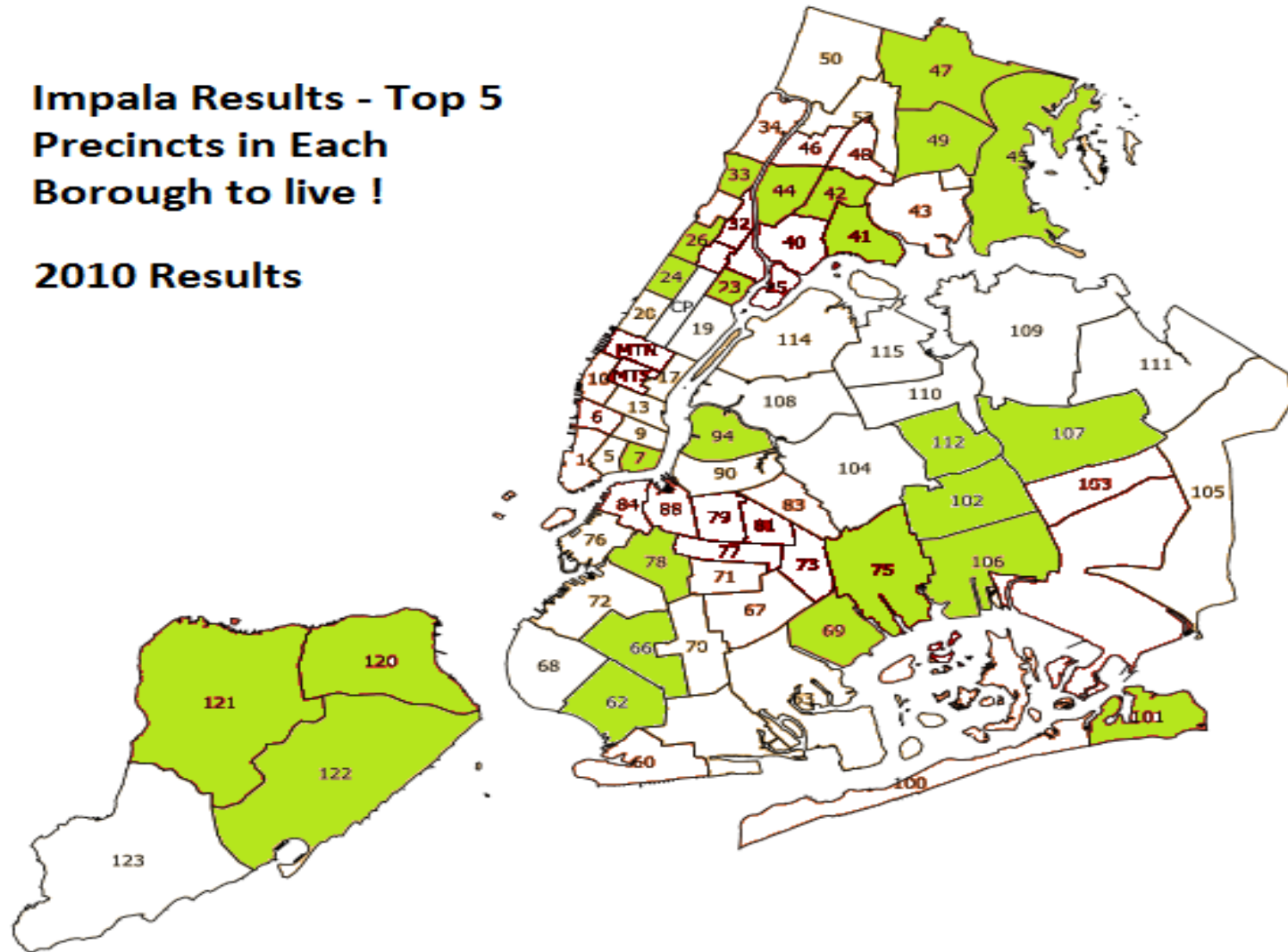
2009 Results

Real Estate Evaluation Based on Crime & Complaints

Results

Impala Results - Top 5 Precincts in Each Borough to live !
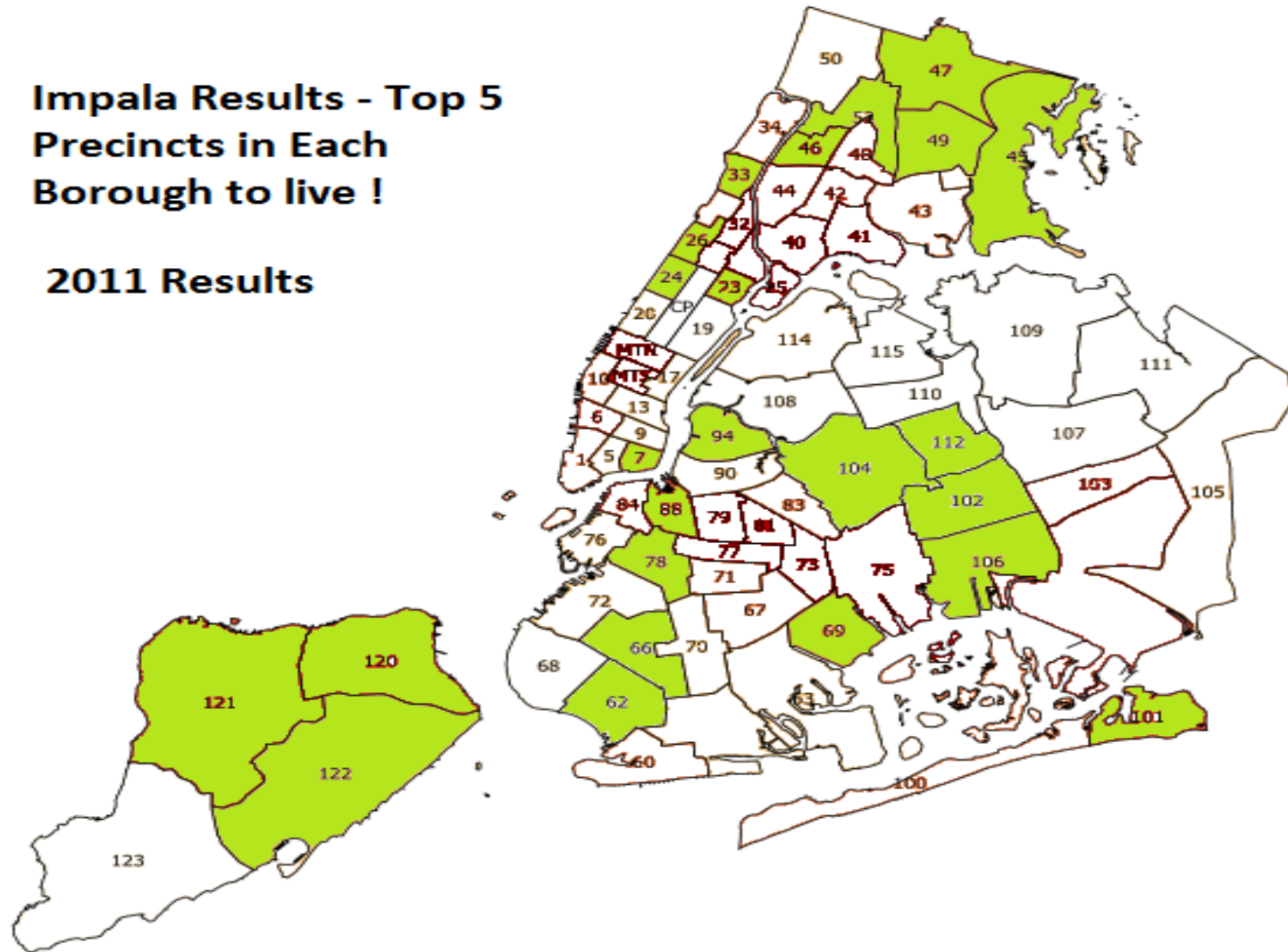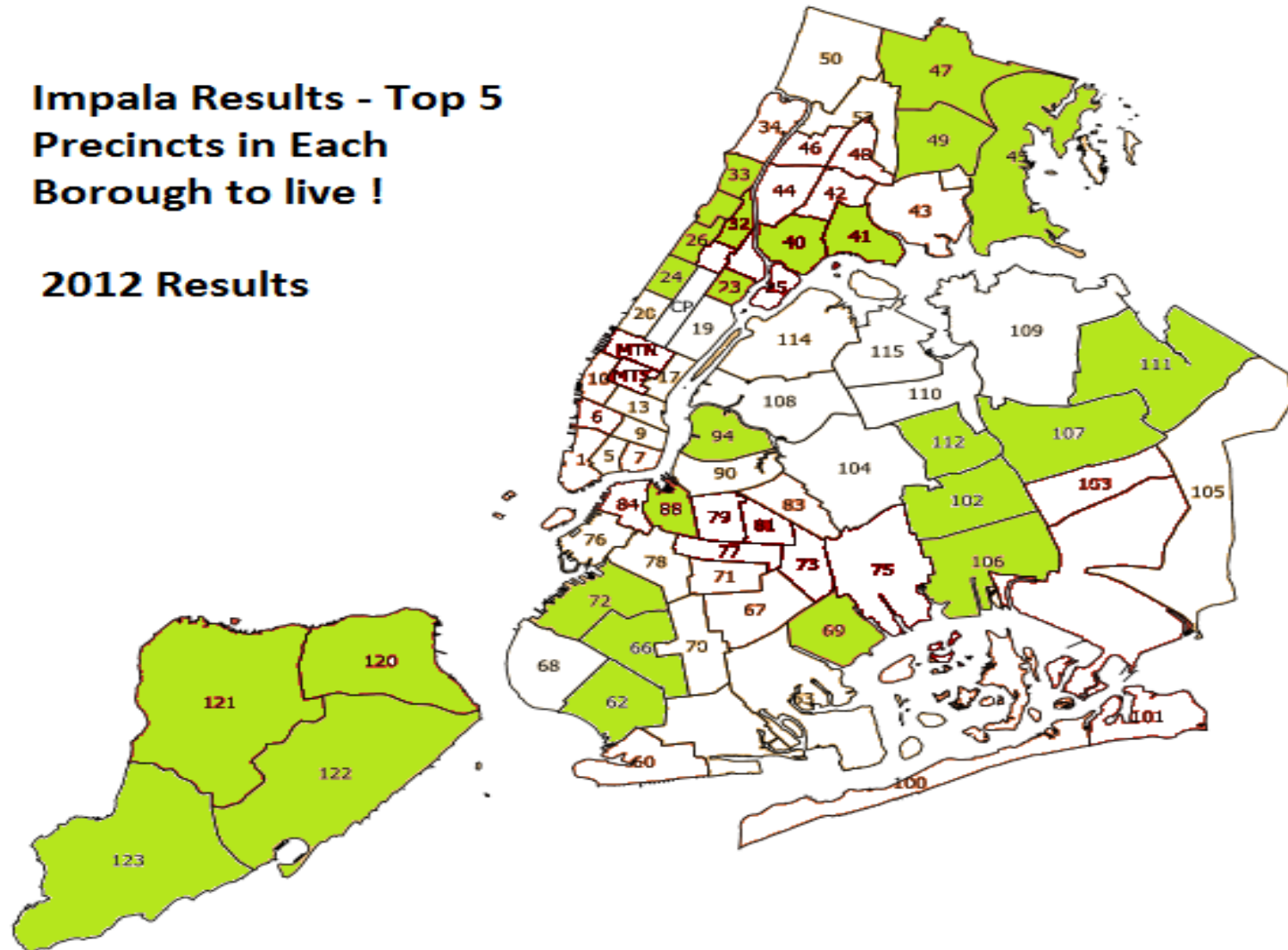
2010 Results

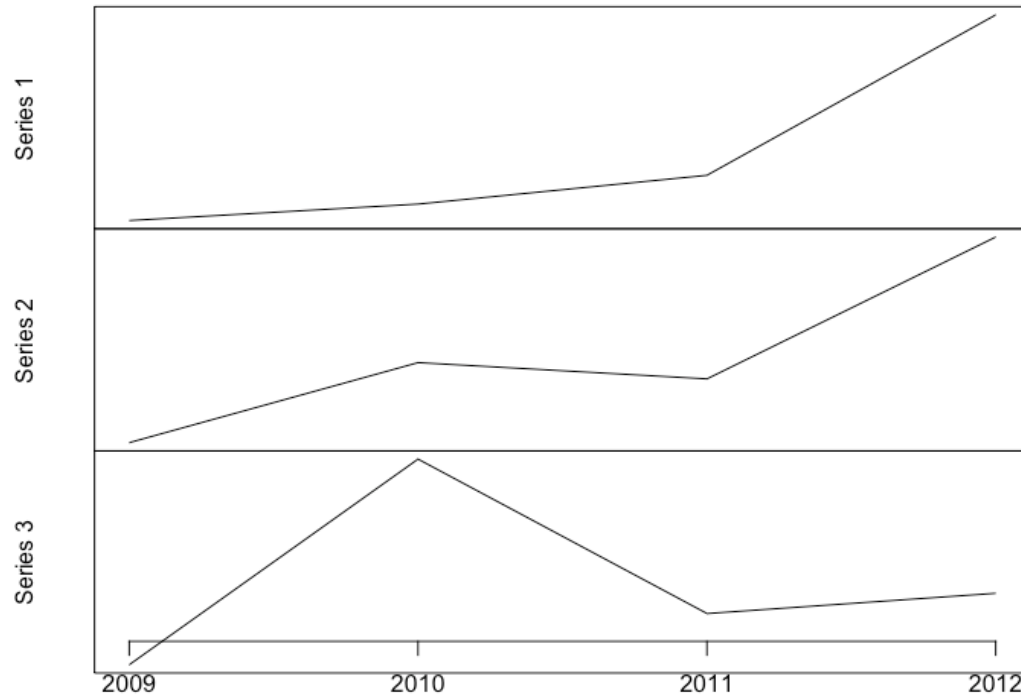# Real Estate Evaluation Based on Crime & Complaints

Results

Impala Results - Top 5
Precincts in Each
Borough to live !

2011 Results

Real Estate Evaluation Based on Crime & Complaints

Results

Impala Results - Top 5
Precincts in Each
Borough to live !

2012 Results

Real Estate Evaluation Based on Crime & Complaints

Results for Brooklyn borough

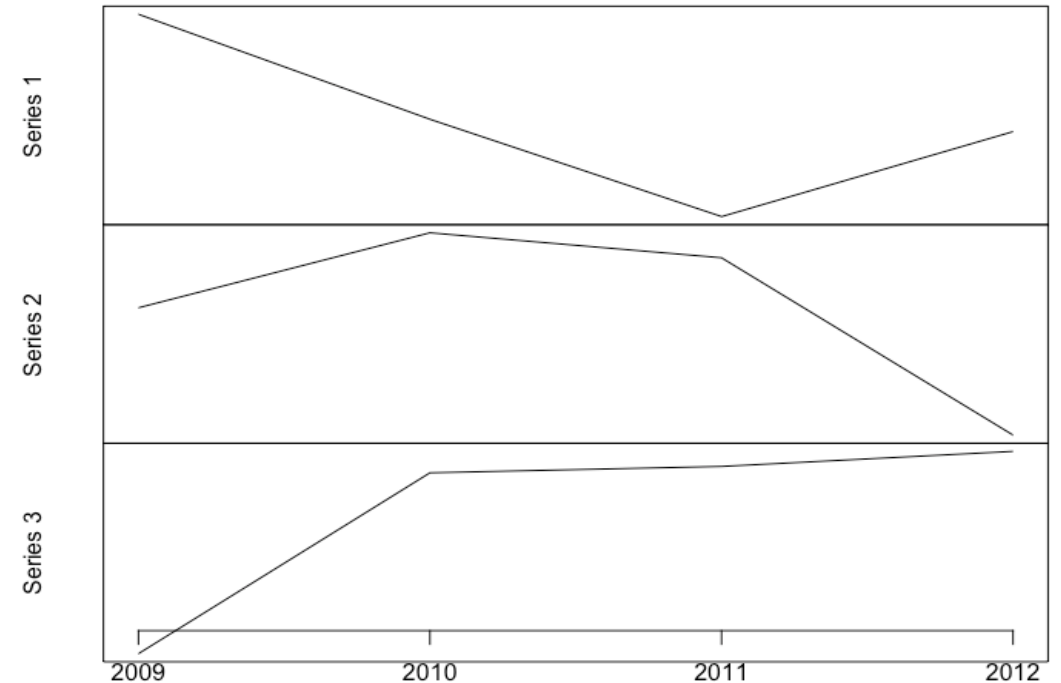| Year	Approach 1				Approach 2			
	2009	2010	2011	2012	2009	2010	2011	2012
Top 5 Precincts	94	94	78	69	77	69	69	69
	78	69	69	72	69	62	62	66
	88	62	88	66	78	94	66	62
	69	66	62	62	94	66	88	72
	77	60	66	60	88	78	78	88

Real Estate Evaluation Based on Crime & Complaints

Results

For Precinct 7: Crime(Series 1), Property(Series 2) & 311(Series 3)

For Precinct 26: Crime(Series 1), Property(Series 2) & 311(Series 3)

Real Estate Evaluation Based on Crime & Complaints

Obstacles

Mapping each of the data set : finding a common attribute with respect to location
We performed the mapping between precincts, neighborhood and zip code using NYPD precinct finder application

Real Estate Evaluation Based on Crime & Complaints

Summary

We have successfully evaluated the best neighborhoods / precincts based on our datasets from 2009 to 2012 using big data technologies and proved the goodness of our analytics.

Acknowledgements

NYC Open data

Real Estate Evaluation Based on Crime & Complaints

References

[1] George E. Tita, Tricia L. Petras, Robert T. Greenbaum, "Crime and Residential Choice: A Neighborhood Level".

[2] Shyam Varan Nath, "Crime Pattern Detection Using Data Mining".

[3] Varian, Hal R, "Big Data: New Tricks for Econometrics".

[4] Avinash Lakshman, Prashant Malik, "Cassandra - A Decentralized Structured Storage System".

[5] George Owusu, Martin Oteng-Ababio, Adobea Y Owusu, Charlotte Wrigley-Asant, "Can Poor Neighborhoods be Correlated with Crime? Evidence from Urban Ghana".

[6] Fay Chang, Jeffrey Dean, Sanjay Ghemawat, Wilson C. Hsieh, Deborah A. Wallach Mike Burrows, Tushar Chandra, Andrew Fikes, Robert E. Gruber, "Bigtable: A Distributed Storage System for Structured Data".

Real Estate Evaluation Based on Crime & Complaints

Thank you!