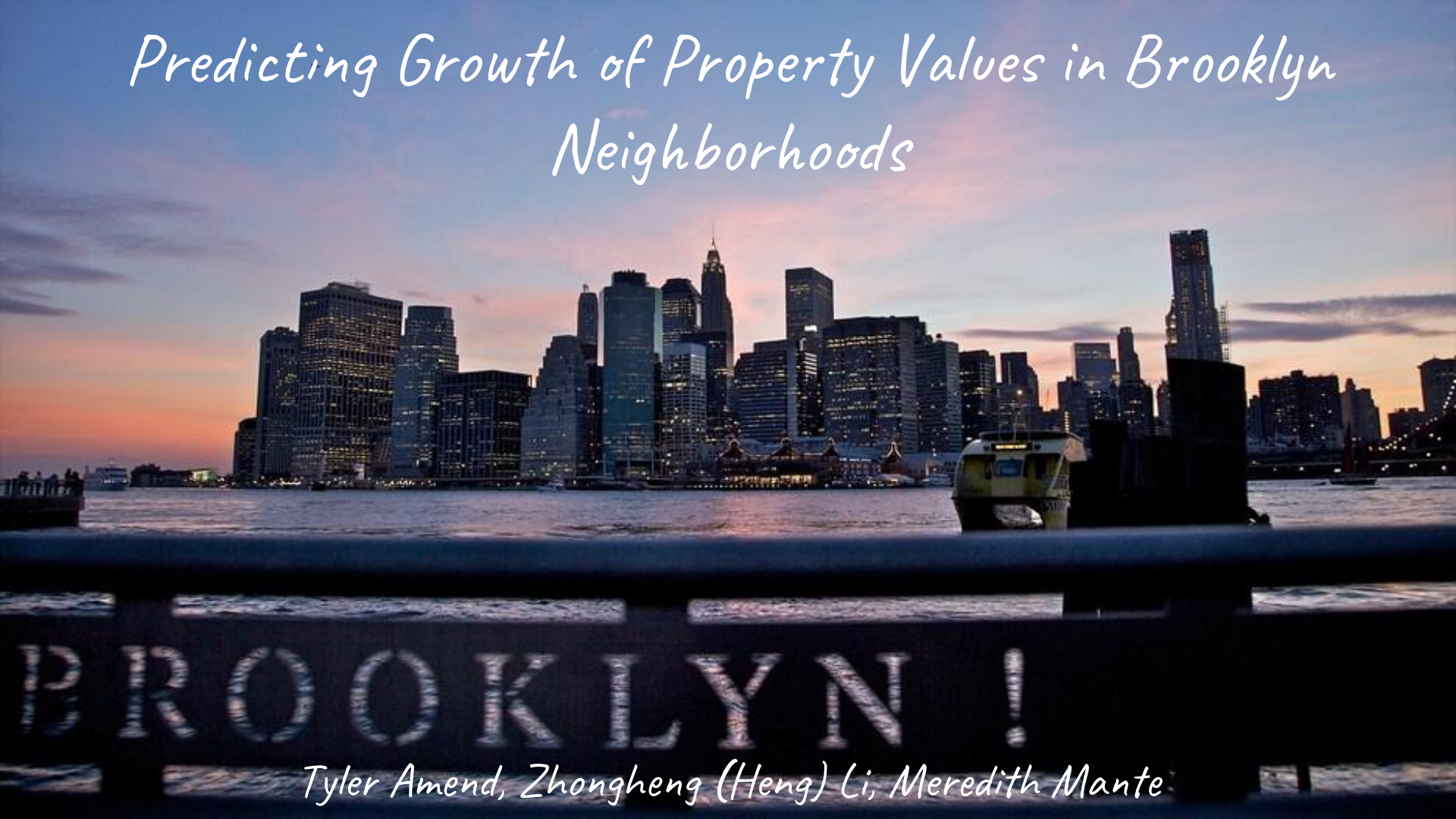
Predicting Growth of Property Values in Brooklyn Neighborhoods



Tyler Amend, Zhongheng (Heng) Li, Meredith Mante

Goal: Provide Insight on Future Property Value for First-time Homebuyers in Brooklyn

- Project the growth rate for individual properties in Brooklyn for the year 2020
- Identify what features most contribute to growth

CATEGORIES OF GROWTH

1: $x > 20\%$

2: $20\% > x > 5\%$

3: $5\% > x > -5\%$

4: $-5\% > x > -20\%$

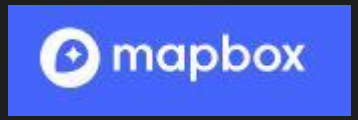
5: $x < -20\%$

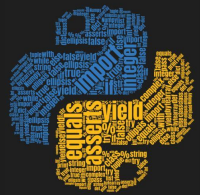
Data Sources

- NYC Property Assessment Values from 2009 to 2018
- NYC Community Data with economic and social indicators from 2009 to 2016
- Tweets scraped using Tweepy streaming. Script listened for Real Estate and Brooklyn related hashtags

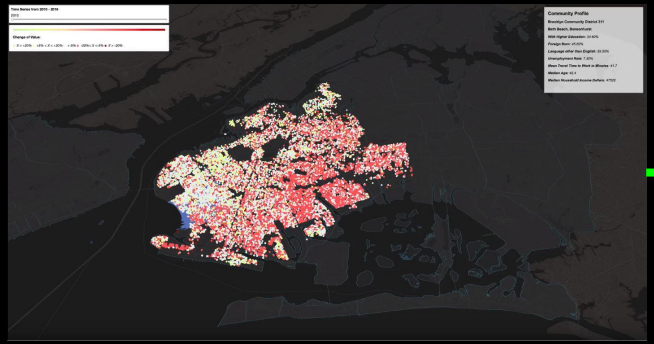
Technologies Used

- PySpark for Data Preprocessing
- MLlib using Scala for Machine Learning
- MongoDB to store data for Map
- MapBox for visualization
- Python for Data Collection





Project Workflow



Data Preprocessing

- Join data from disparate sources and format specific fields so that related data fields may be stored in the same relation
- Create attributes (owner change, % change in mean age, etc.)
- Create labels (based on our 5 categories)
- Formatting and storing Twitter data
- Cleaning data with broken values or inconsistent formats
- Understanding and linking government data with the keys provided in the actual data sets

Features for Prediction 1 (Neighborhood Indicators)

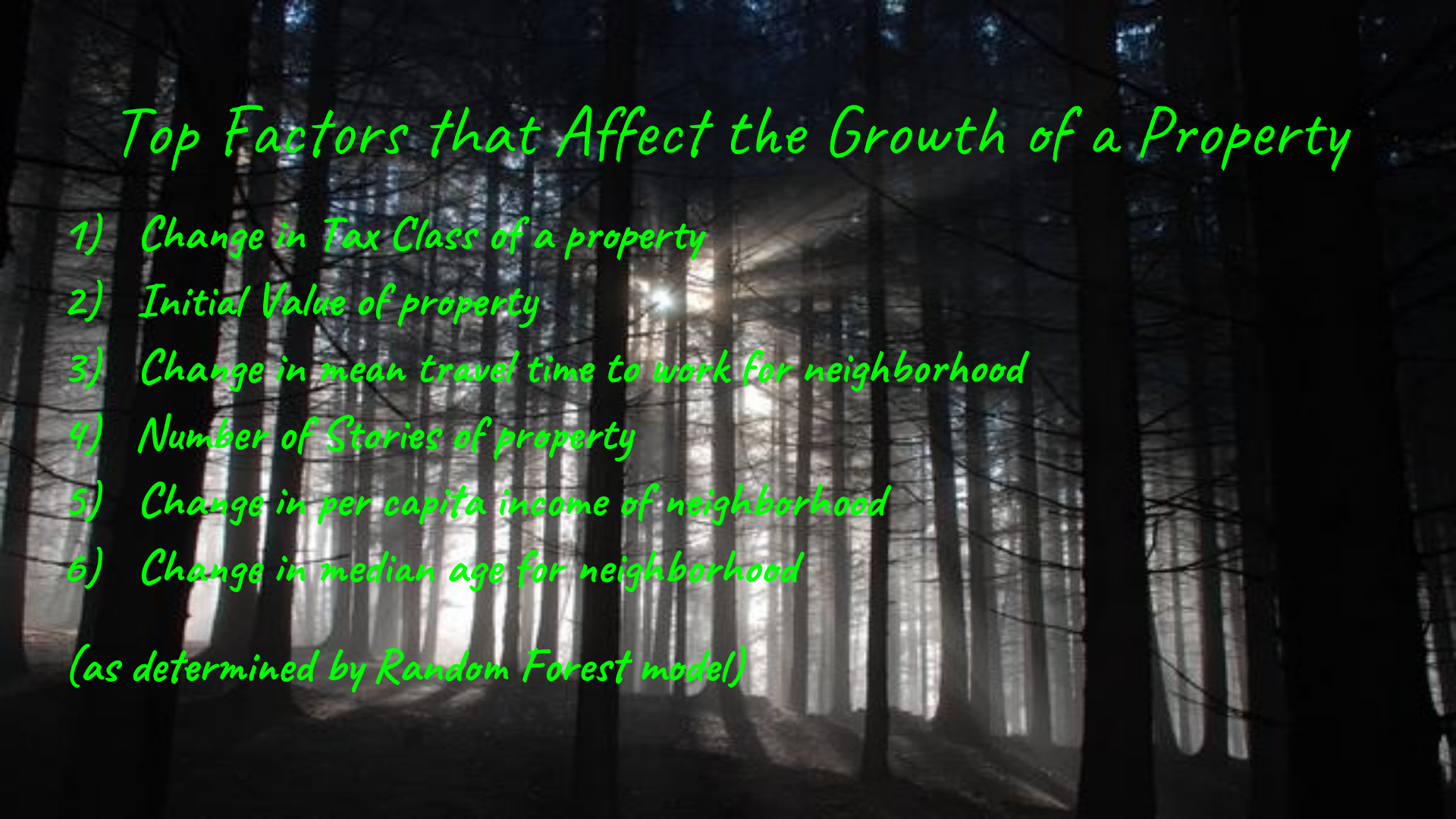
- Median age (years)
- Unemployment Rate
- Mean travel time to work (minutes)
- Median household income (dollars)
- Per capita income (dollars)
- No health insurance coverage
- Average household size
- Bachelor's degree or higher
- Lived in a different house 1 yr ago
- Foreign-born
- Naturalized U.S. citizen
- Language other than English

Features for Prediction 2 (Other Indicators)

- Average Mortgage Rate for Year
- Property sold within last year
- Building Class Changed
- Tax Class Changed
- Assessed Property Value for
- Current Year

Prediction Model - Random Forest

- In order to attract and retain investor capital, Hedge Funds and other investment vehicles often must be able to give some detail to why an investment decision was made. Unless our “firm” had an impeccable track record (we don’t), we must be able to provide our investors with coefficients from our model, which would not be possible using a more blackbox ML model
- Good fit for our categorical and numeric attributes
- Good fit for our multi-class prediction model
- We were able to extract most important features as determined by the model



Top Factors that Affect the Growth of a Property

1) Change in Tax Class of a property
2) Initial Value of property
3) Change in mean travel time to work for neighborhood
4) Number of Stories of property
5) Change in per capita income of neighborhood
6) Change in median age for neighborhood

(as determined by Random Forest model)

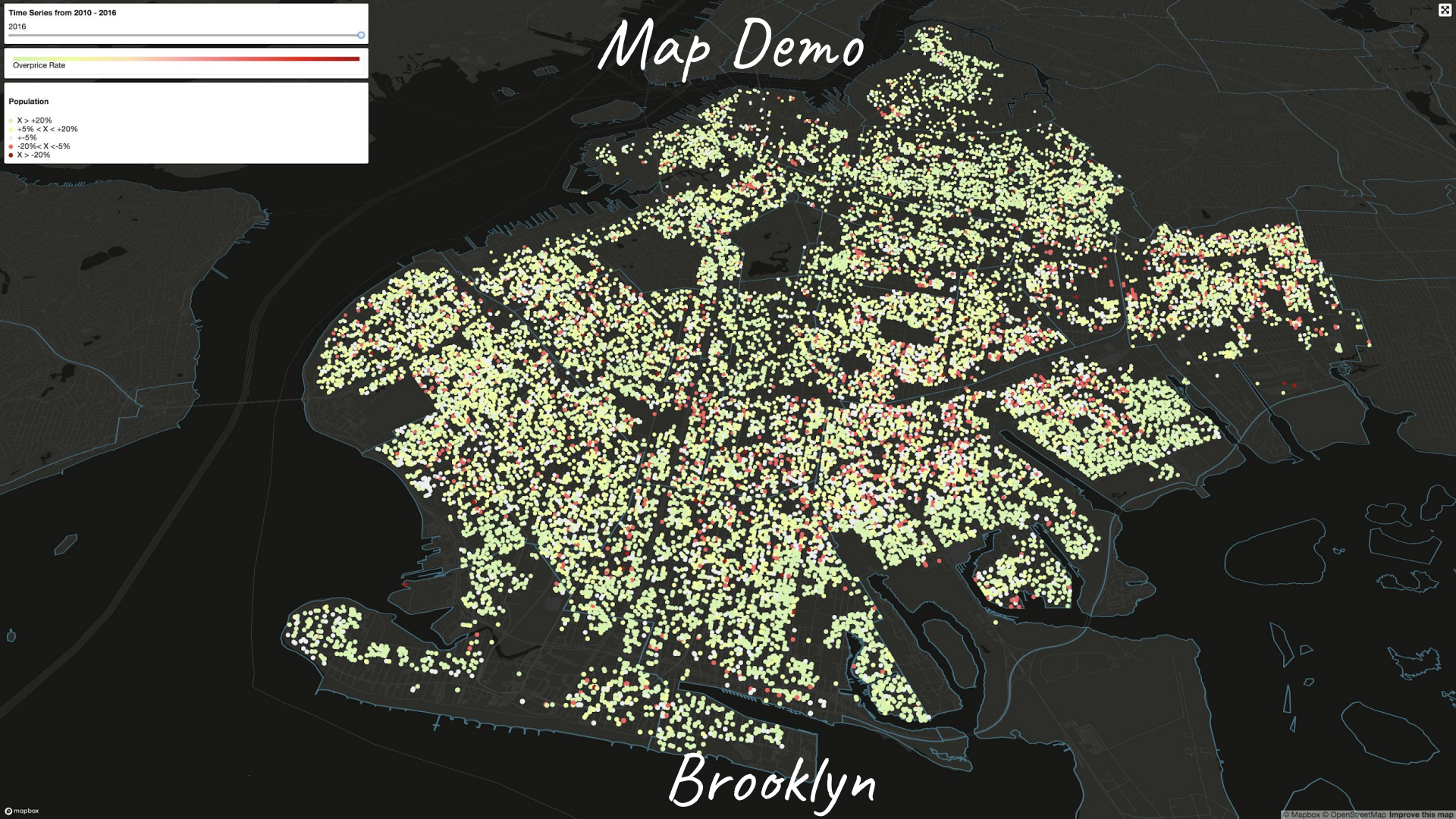
Difficulties

- Geolocating properties by address - limitations of Google Map API, 2500 per day -> limited to BK.
- Missing data - Only have social and economic data through 2016 -> aggregating by 2 years
- Twitter API limits
- Joining disparate data sources; creating attributes and labels
- Connecting MongoDB to our MapBox frontend
- Incorrect format storage of Twitter data. Required custom parsing and update of the format

(Premium API services available for professional projects to allow for scalability of Google Map & Twitter APIs)

Future Directions for Project

- Apply model to other boros
- Apply model to other urban areas across the country
- Incorporate more features like Crime, School quality, Business Type, etc
- Try different models for our classification problem
- Source more granular data per property
- Develop automated method to collect proprietary property data
- Develop sentiment analysis model to better analyze and source neighborhood specific Twitter data
- Deploy capital to invest in properties the model suggests are currently valued at a discount


Time Series from 2010 - 2016
2016
Overprice Rate
Population
X > +20%
+5% < X < +20%
+-5%
-20%< X <-5%
X > -20%
Map Demo
Brooklyn
mapbox
© Mapbox © OpenStreetMap Improve this map