



New Office Location Assessment:

A COMPARISON OF BOSTON, PHILLY, AND NYC

ANJALI KAYAL

Circumstance:

Expansion of operations to Northeast

- Expansion to northeast requires convincing employees to move and start up the new office
 - Compare Boston, Philadelphia, and NYC in order to determine best city to live in
- Assess across the following topics:
 - Safety of the city
 - Educational quality
 - Access to healthcare
 - Other Economic indicators

Datasets for comparison were leveraged from Open Data program

Boston data

- Historical crime data (2012 – 2015)
- Location of health centers
- Nationwide survey data on health centers by hospital
- Nationwide dataset on math and reading performance by school
- Government employee salaries
- Restaurant inspection data

Philadelphia data

- Historical crime data (2016 – 2015)
- Location of health centers
- Nationwide survey data on health centers by hospital
- Nationwide dataset on math and reading performance by school
- Government employee salaries
- Air pollution data

NYC data

- Historical crime data (2000 – 2011)
- Location of health centers
- Nationwide survey data on health centers by hospital
- Nationwide dataset on math and reading performance by school
- Government employee salaries
- Restaurant inspection data
- Air pollution data

Architecture

- Large static datasets; only updated annually
- Load data into data lake and transform to fit model
- Leverage Hadoop / hive / spark sql

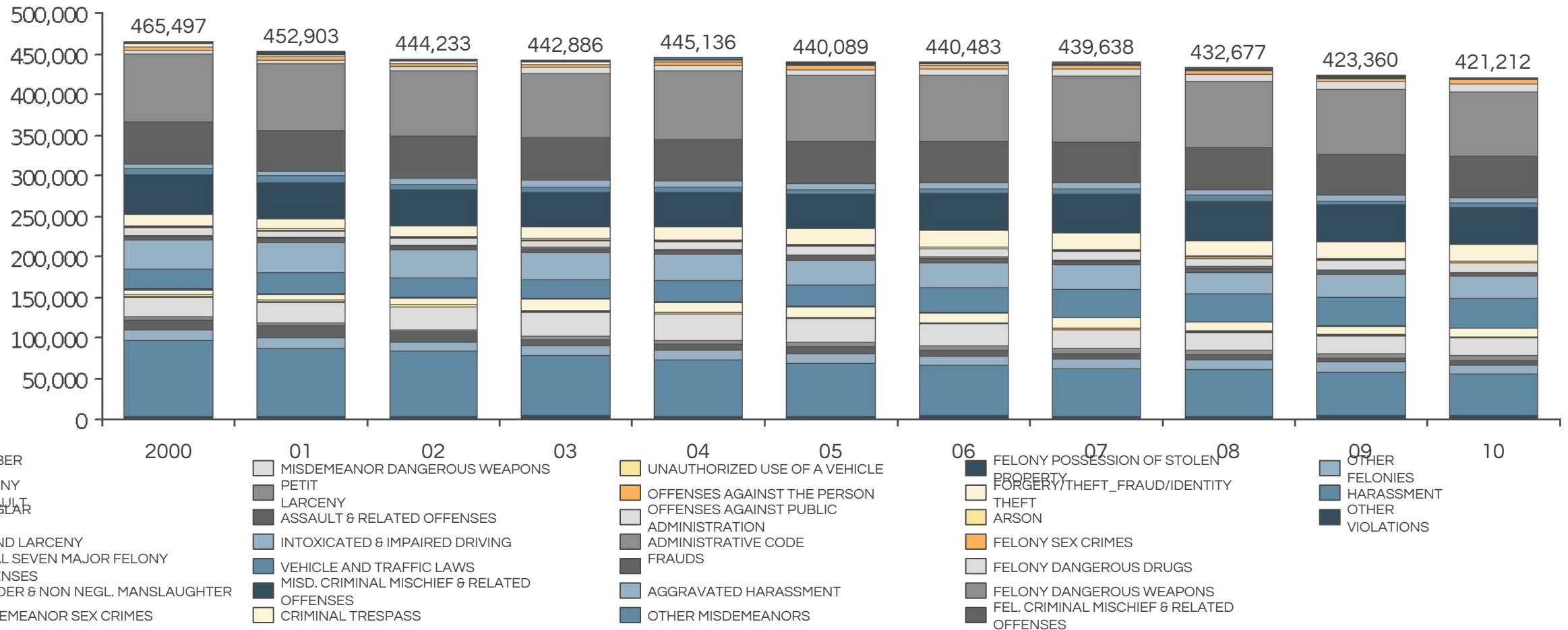


Problems

- Data sets available in multitude of formats (.xls, .xlsx)
- Data tracked differently a cross states
- Similar data not available across states
- Variety of spellings / words used for same concepts

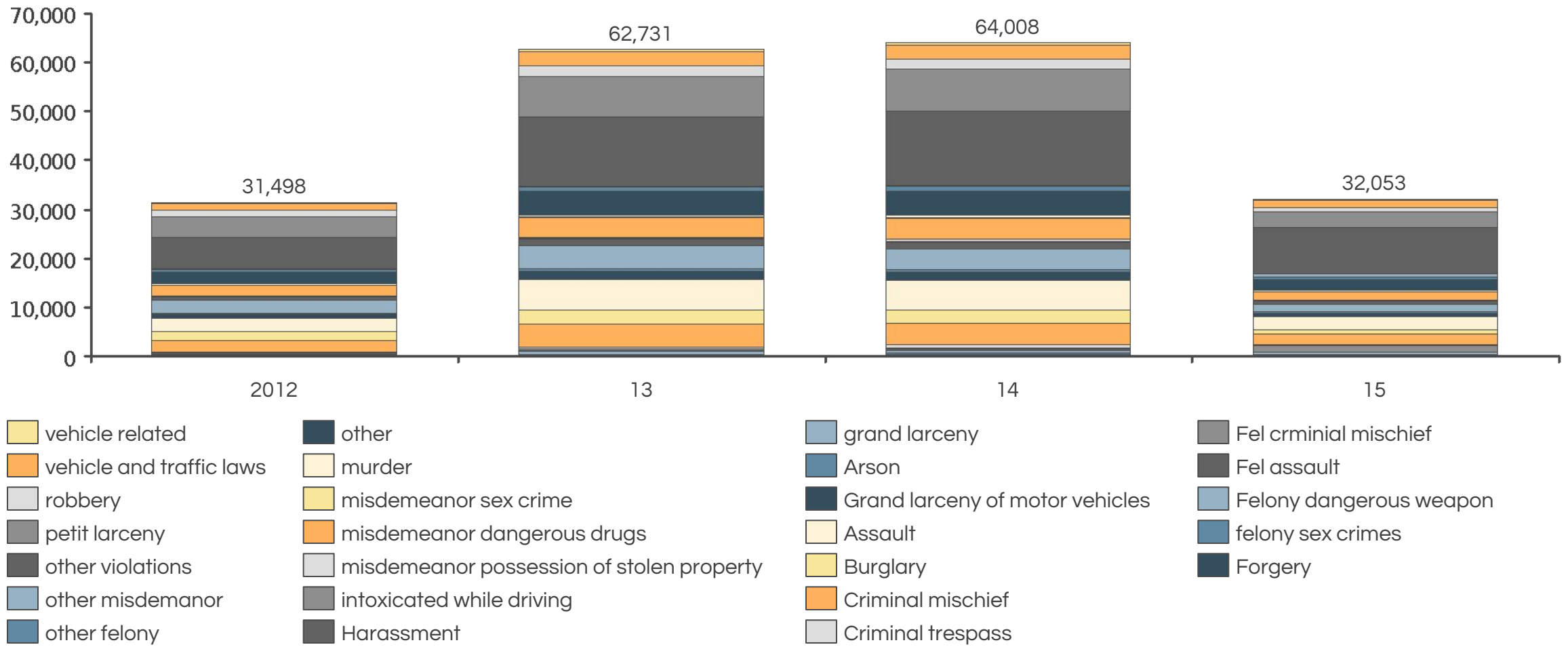
Preliminary results (1 of 3)

New York City Crimes



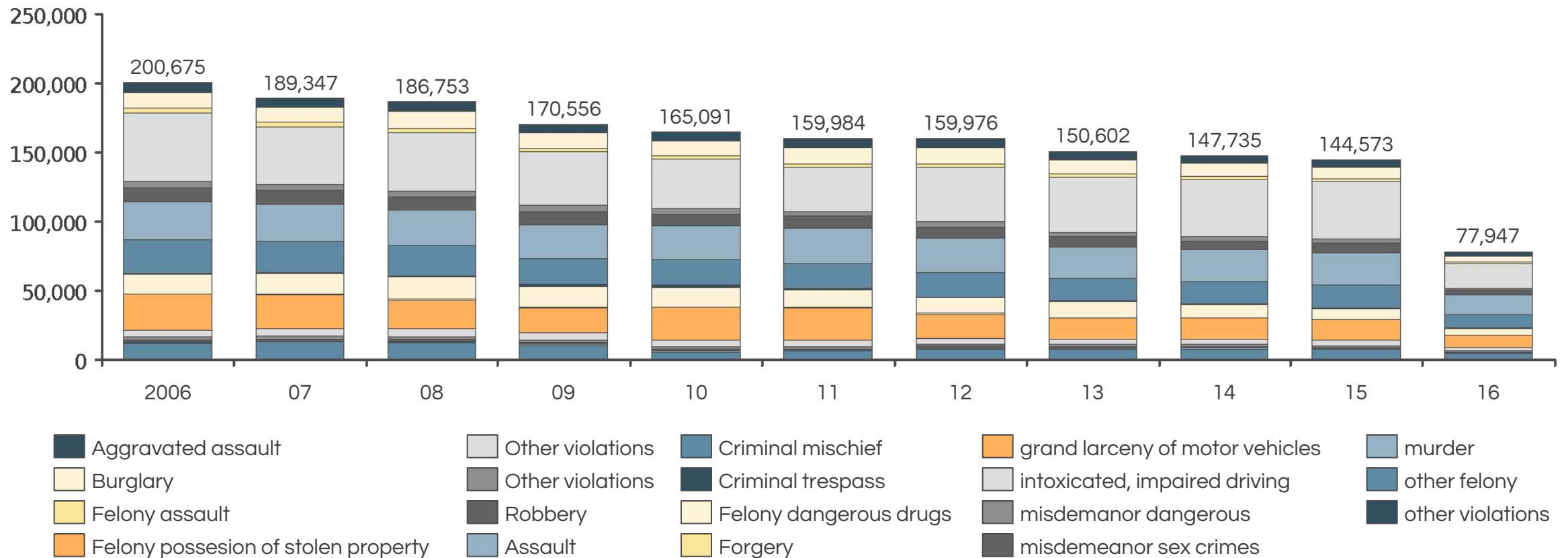
Preliminary results (2 of 3)

Boston City Crimes



Preliminary results (3 of 3)

Philadelphia City Crimes



Future considerations

- Architecture depends on static dataset
 - Significant portions are manually mapped
- Automate process of aligning data sets
- Collect data in standardized way across different all metrics
- Leverage corporate retention data by current state