

Selecting the best neighborhood for life in New York

March 28, 2020

1. Introduction

1.1. Background

People desire to live in a convenient and safe place. However, it is not so easy to understand, which neighborhood has best match with a person's criteria: each person has a different requirements and there are a lot of data to analyze. Especially, choice is difficult, when people change city or even country of living.

1.2. Problem

The report target is to provide recommendation for best place for life in New York for a person, who doesn't live currently in New York. The following criteria are important for a customer:

- It should be a safe place
- There must be markets and parks
- Bakeries, pharmacies and gyms are preferable

1.3. Audience

There is a single person, who desires to change his place of living to New York and doesn't know which neighborhood will be most convenient for him.

2. Data

2.1. Data sources

New York crimes data will be obtained from official source:

<https://data.cityofnewyork.us/api/views/qgea-i56i/rows.csv>.

New York neighborhoods geo information will be downloaded from

https://cocl.us/new_york_dataset.

Forsquare will be used to obtain information about venues in each neighborhood.

2.2. Data cleaning

Crime data has the following look:

CMPLNT_TO_TM	ADDR_PCT_CD	RPT_DT	KY_CD	OFNS_DESC	PD_CD	PD_DESC	CRM_ATPT_CPTD_CD	LAW_CAT_CD	BORO_NM
NaN	73.0	04/10/2008	341	PETIT LARCENY	321.0	LARCENY,PETIT FROM AUTO	COMPLETED	MISDEMEANOR	BROOKLYN
NaN	28.0	06/03/2007	236	DANGEROUS WEAPONS	782.0	WEAPONS, POSSESSION, ETC	COMPLETED	MISDEMEANOR	MANHATTAN
20:50:00	102.0	02/16/2010	105	ROBBERY	375.0	ROBBERY,PHARMACY	COMPLETED	FELONY	QUEENS

The following data is available:

```

CMPLNT_NUM          int64
CMPLNT_FR_DT        object
CMPLNT_FR_TM        object
CMPLNT_TO_DT        object
CMPLNT_TO_TM        object
ADDR_PCT_CD         float64
RPT_DT              object
KY_CD               int64
OFNS_DESC           object
PD_CD               float64
PD_DESC             object
CRM_ATPT_CPTD_CD    object
LAW_CAT_CD          object
BORO_NM             object
LOC_OF_OCCUR_DESC   object
PREM_TYP_DESC       object
JURIS_DESC          object
JURISDICTION_CODE   float64
PARKS_NM            object
HADEVELOPT          object
HOUSING_PSA         object
X_COORD_CD          float64
Y_COORD_CD          float64
SUSP_AGE_GROUP      object
SUSP_RACE            object
SUSP_SEX            object
TRANSIT_DISTRICT    float64
Latitude             float64
Longitude            float64
Lat_Lon             object
PATROL_BORO          object
STATION_NAME         object
VIC_AGE_GROUP        object
VIC_RACE             object
VIC_SEX             object

```

Customer didn't mention any special requirements for safety, so, the total number of crimes per borough (BORO_NM) will be as safety criteria. Other information will be not used.

New York geo data consists of following features:

```

{'type': 'Feature',
 'id': 'nyu_2451_34572.1',
 'geometry': {'type': 'Point',
 'coordinates': [-73.84720052054902, 40.89470517661]}},

```

```
'geometry_name': 'geom',  
'properties': {'name': 'Wakefield',  
  'stacked': 1,  
  'annoline1': 'Wakefield',  
  'annoline2': None,  
  'annoline3': None,  
  'annoangle': 0.0,  
  'borough': 'Bronx',  
  'bbox': [-73.84720052054902,  
    40.89470517661,  
    -73.84720052054902,  
    40.89470517661]}}
```

The following data will be used:

- Borough name
- Neighborhood name (properties.name)
- Coordinates

Forsquare data will be used to collect information of venues in each neighborhood. Following data will be used:

- Venue name
- Venue coordinates
- Venue category name