

STEVENS INSTITUTE OF TECHNOLOGY

BIA660-NYC311

Quality of Life complaints in New York City

5/3/2013

John Procyk Lianze Ma Mengyi Gong Pallavi Rastogi
--

Introduction

New York City is an amalgam of neighborhoods. The city is divided geographically into 5 boroughs – Manhattan, Bronx, Brooklyn, Queens, and Staten Island. Each of the boroughs is divided into neighborhoods such as the Upper East Side and Chelsea in Manhattan, and Morrisania and Riverdale in the Bronx. There are several zip codes in each neighborhood. Some of the neighborhoods are as different as night and day; others are nearly identical in their socio-economic makeup. Quality of Life can be looked at in different ways, depending on the observer. In a wealthy neighborhood, quality of life issues might be considered trivial by somebody living in a low income neighborhood. Issues can change with the seasons. They can change with the weather.

In this project, we chose to analyze the following aspects through our web application:

- We thought it would be interesting to see what the top 10 quality of life complaints were in different parts of New York City at different times of the year.
- We also analyzed 311 complaints in two months of the year- October 2012 (during the time of the Hurricane) and January 2013 (Winter).
- Comparisons were done among high-poverty areas, among high-income areas, between these neighborhoods, and between urban and suburban neighborhoods. We looked at seasonal differences, the heat of the summer vs the chill of winter and the effects of Hurricane Sandy.

Data Collection

- The main data was collected from ***nycopendata.org***. The 311 data on nycopendata.org has several years of information. Manhattan, Bronx, Staten Island, and Queens are represented in the logs. Brooklyn is not. For our purposes, we removed Queens; Staten Island stood in for the suburban areas in Queens. Zip code information are provided in nycopendata.org, making it possible to group the data into neighborhoods. GPS information is included, giving the ability to map the data.

NYC OpenData

Home

About NYC Open Data

FAQ

Contact NYC Open Data

Unsaved View

Save As...

Revert

Based on 311 Service Requests from 2010 to Present
All 311 Service Requests from 2010 to present. This information is automatically updated daily.

Manage

409 More Views

Filter

Visualize

Export

Find in this

	Created Date	Closed Date	Agency Name	Agency	Complaint Type	Descriptor	Location Type	Incident Zip
17	04/30/2013 12:00:00 AM		Department of Housing Preservation and D	HPD	ELECTRIC	BELL-BUZZER/INTERCOM	RESIDENTIAL BUILDING	10452
18	04/30/2013 12:00:00 AM		New York City Police Department	NYPD	Noise - Commercial	Loud Music/Party	Club/Bar/Restaurant	11225
19	04/30/2013 12:00:00 AM	04/30/2013 12:00:	New York City Police Department	NYPD	Blocked Driveway	No Access	Street/Sidewalk	10467
20	04/30/2013 12:00:00 AM	04/30/2013 12:00:	New York City Police Department	NYPD	Illegal Parking	Blocked Hydrant	Street/Sidewalk	10473
21	04/30/2013 12:00:00 AM	04/30/2013 12:00:	New York City Police Department	NYPD	Noise - Park	Loud Talking	Park/Playground	11434
22	04/30/2013 12:00:00 AM		New York City Police Department	NYPD	Illegal Parking	Commercial Overnight Parking	Street/Sidewalk	11234
23	04/30/2013 12:00:00 AM	04/30/2013 12:00:	New York City Police Department	NYPD	Homeless Encampment	N/A	Street/Sidewalk	10038
24	04/30/2013 12:00:00 AM	04/30/2013 12:00:	New York City Police Department	NYPD	Noise - Commercial	Banging/Pounding	Club/Bar/Restaurant	10002
25	04/30/2013 12:00:00 AM		Taxi and Limousine Commission	TLC	Taxi Complaint	Driver Complaint		11220

- Before gathering data, it was necessary to segment the boroughs by neighborhood. Another website was found which gave good segmentation by zip codes. The breakdowns we used were those used by the **NYS Cancer Registry and Cancer Statistics** website. We set up a table in a MySQL database and entered those values. Figure 2 shows a sample of the table.

New York State State Agencies		
Department of Health Information for a Healthy New York		
You are Here: Home Page > Appendices Menu > ZIP Code Definitions of New York City Neighborhoods		
ZIP Code Definitions of New York City Neighborhoods		
Borough	Neighborhood	ZIP Codes
Bronx	Central Bronx	10453, 10457, 10460
	Bronx Park and Fordham	10458, 10467, 10468
	High Bridge and Morrisania	10451, 10452, 10456
	Hunts Point and Mott Haven	10454, 10455, 10459, 10474
	Kingsbridge and Riverdale	10463, 10471
	Northeast Bronx	10466, 10469, 10470, 10475
	Southeast Bronx	10461, 10462, 10464, 10465, 10472, 10473
Brooklyn	Central Brooklyn	11212, 11213, 11216, 11233, 11238

Figure 2 - Groupings of Zip Codes into Neighborhoods

Demographic data was pulled from another site, **zipskinny.com**, that has demographic data by zip code. The website is difficult to work with – a mass of HTML tables – but we were able to retrieve the required data using Beautiful Soup and Python and move it directly to the database. Figure 3 shows a sample of the website with one zip code's information.

[What's My House Worth?](#)
www.HouseValues.com
 Find your home's current market value online with HouseValues.com.

➔

AdChoices

the skinny on:
10012
 (NEW YORK NY)

General Information:
 Latitude: 40.72596
 Longitude: -73.998345
Population: 26000
 Density: 81345.07
 (people per square land mile)
 Housing Units: 14625
 Land Area: 0.32 sq. mi.
 Water Area: 0 sq. mi.

Social Indicators		Economic Indicators			
Educational Achievement: <small>(among people 25 years or older)</small>		Household Income			
Marital Status: <small>(among people 15 years or older)</small>		Occupation <small>(among employed persons over 16)</small>			
Less than 9th grade: 8.3%	Never married: 53.4%	<\$10,000 10.5%	Mgt./Professional 65.8%		
9th-12th grade (nongrad): 5.1%	Married: 34.4%	\$10,000-\$14,999 5.3%	Service 7.1%		
High school graduate: 9.3%	Separated: 1.3%	\$15,000-\$24,999 9.1%	Sales/Office 19.3%		
Some college: 10.6%	Widowed: 3.8%	\$25,000-\$34,999 7.5%	Farm/Fishing/Forestry 0%		
Associate degree: 2%	Divorced: 7.1%	\$35,000-\$49,999 11.5%	Construction/Extraction/Maint. 2.3%		
Bachelors degree: 34.7%	Stability/Newcomer Appeal: <small>Same home 5+ years: 51.2%</small>	\$50,000-\$74,999 17%	Production/Transportation 5.5%		
Graduate/Professional: 30%		\$75,000-\$99,999 11.4%	Unemployment/Poverty		
High school or higher: 86.6%	Social and economic indicators <small>based on 2000 Census sample data.</small>	\$100,000-\$149,999 13.7%	Unemployed 4.1%		
Bachelors or higher: 64.6%		\$150,000-\$199,999 5.3%	Below Poverty Line 13.8%		
Green = Above U.S. Avg Red = Below U.S. Avg		\$200,000+ 8.8%	Median Household Income: \$58,313		
Red bkg. = Above U.S. Avg Green bkg. = Below U.S. Avg					

Figure 3 - zipskinny.org, with Demographics Data highlighted

The information we have gathered from nycopendata.org was kept to a fairly high level – date, zip code, complaint type, and the number of each complaint by month. New York’s 311 phone system logs quality of life complaints with address, zip code, gps code, complaint type and sub-type, and date. Other fields (School Zone, nearest intersection, Taxi Pickup Location) are available for logging as well, but most are rarely logged. This data can be gathered directly from nycopendata.org in tabular format, but the ability to compare 2 neighborhoods and their corresponding economic status simultaneously is currently not available.

Once the neighborhoods were prepared from zip codes, a Python script was set up to pull from nycopendata.org. Socrata has an API where SQL-like WHERE clauses could be used, but they didn’t work well when the clauses grew in size and were interrupted with multiple space characters in the WHERE clause. There was an addition to Socrata to bypass this using a direct SQL query in the GET. That was able to be programmed to substitute date ranges and zip code ranges to quickly fetch the data into MySQL. Both aggregated data and single-event data were gathered from this website. The query string submitted to retrieve counts of complaints is shown in the Appendix.

Programming Languages/Scripting Languages/Visualization software

- Python was used to pull data from nycopendata.org.
- MySQL relational database is used as the local data store. A mysqldump file was taken to upload the data to Google Cloud SQL database.
- Tableau Public: It is product of Tableau Software. With Tableau Public, we can create interactive graphs, dashboards, maps and tables from virtually any data and embed them on our website. Since it is free software, not all features were available for use. Therefore, we used a small amount of static data from our database to show the visualization. We used only October 2012 and January 2013 data to show the trends of NYC 311 complaints. But there is definitely a possibility of uploading dynamic data as well.
- Python, Javascript, CSS and HTML5 for coding the Google App Engine.
- C# and HTML for coding the Active Server Pages.

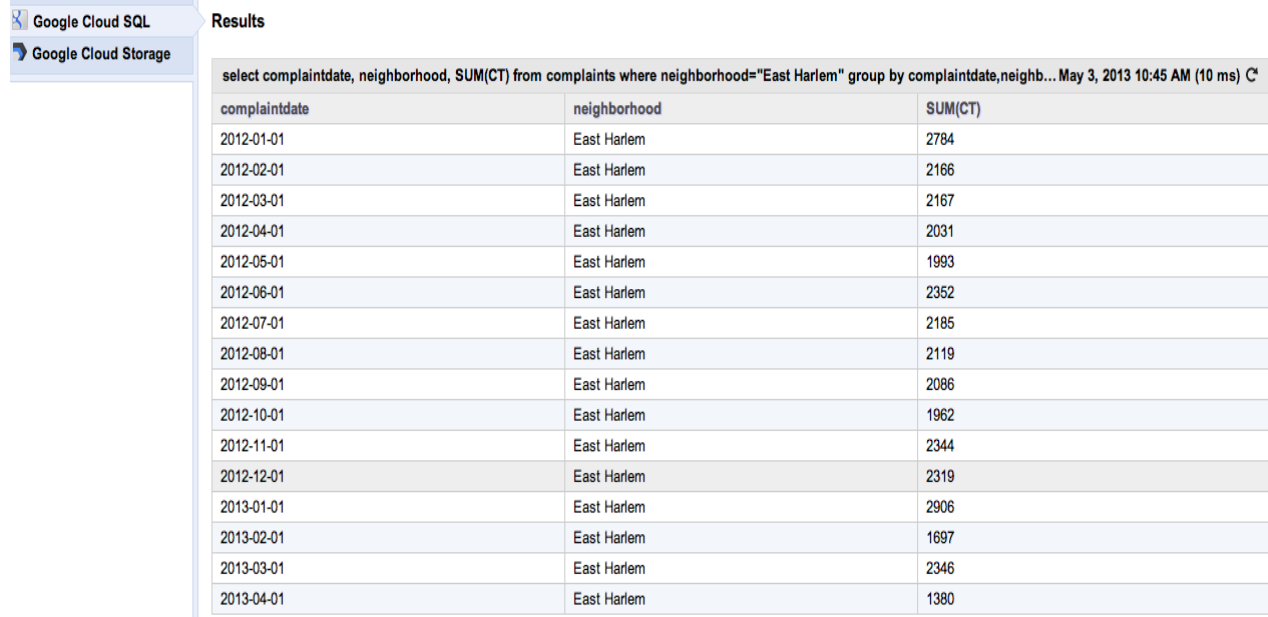
Preparing the Website

The website has been created in Google App Engine. We decided that it would be advantageous to quickly deliver a working prototype in order to begin viewing and analyzing data while the final application was being built. We built a prototype with Microsoft Active Server Pages and Tableau with MySQL as the data store on the back end. The ASP application was used for data investigation while the final application was being built.

The data for the Google App Engine is stored in Google Cloud SQL as well. The data was transferred from MySQL to Google Cloud SQL using the mysqldump utility to transfer database schema and data in a single file. The database will be updated with new data by running the Python script after changing the dates in the script. Nycopendata.org updates their 311 database daily; our script will be run on a monthly basis. We used Webapp2 and Jinja2 template to build our final version on Google App Engine. The server-end code is written in

Python. Google visualization libraries were used to generate charts on our website. And JQuery was used as well for the interaction between the server and the front-end web page.

We concentrated on socio-economic demographics. In this phase, there isn't a need to look beyond average income and poverty rate. A great deal of demographic data was available (gender, race, educational level, occupational) but for this application, we felt the best indicators were those we have chosen.



The screenshot shows the Google Cloud SQL console interface. On the left, there are tabs for 'Google Cloud SQL' and 'Google Cloud Storage'. The main area displays a 'Results' section with a SQL query: 'select complaintdate, neighborhood, SUM(CT) from complaints where neighborhood="East Harlem" group by complaintdate,neighb...'. The query was executed on May 3, 2013, at 10:45 AM, taking 10 ms. The results are shown in a table with three columns: 'complaintdate', 'neighborhood', and 'SUM(CT)'. The table contains 24 rows of data, showing the number of complaints (SUM(CT)) for each month from January 2012 to April 2013, all for the 'East Harlem' neighborhood.

complaintdate	neighborhood	SUM(CT)
2012-01-01	East Harlem	2784
2012-02-01	East Harlem	2166
2012-03-01	East Harlem	2167
2012-04-01	East Harlem	2031
2012-05-01	East Harlem	1993
2012-06-01	East Harlem	2352
2012-07-01	East Harlem	2185
2012-08-01	East Harlem	2119
2012-09-01	East Harlem	2086
2012-10-01	East Harlem	1962
2012-11-01	East Harlem	2344
2012-12-01	East Harlem	2319
2013-01-01	East Harlem	2906
2013-02-01	East Harlem	1697
2013-03-01	East Harlem	2346
2013-04-01	East Harlem	1380

Figure 4 - data from Google Cloud SQL for generating line chart

Observations

One of the most important findings we had in comparing neighborhoods using this application was that more questions were generated than answered. Some of the interesting observations are detailed below:

- Comparisons were done between wealthy and lower income neighborhoods in New York.
 - Taking Gramercy Park and Murray Hill in Manhattan against Morrisania and High Bridge in the Bronx for January 2012 show some distinct differences in Quality of Life calls that we will see repeated throughout this project. Heating was the top issue in both locations, but the difference in number of complaints is extraordinary.
 - For a 50% higher population in this area of the Bronx, there were over 8 times as many complaints about heat (over 2000 vs just under 250.)
 - Building issues were very high on the list also, numbering around 500 complaints for General Construction, Paint-Plaster, Plumbing.
- These Quality of Life issues, along with Heating, Nonconst, and Electric came up so frequently together that we thought it would be useful to group them into a segment called "Landlord/Housing Issues". We added a checkbox to remove the whole segment

of 6 issues from the top 10 list. That way, we could uncover an additional 6 issues not related to Public Housing.

- Once Landlord/Housing issues were removed from the view, we were surprised to find there were quite a bit fewer counts of top 10 complaints from the South Bronx than for Murray Hill.
 - Complaints such as Rodents, Street Conditions and Dirty Conditions were found to be common in low-income neighborhoods.
- We wondered why the number of complaints dropped off so drastically once the Landlord/Housing Issues were removed. Was it because of relative satisfaction, few problems in the area, or resignation that complaining will not have an effect anyway? Is it a tendency to not get involved? Or is it because of the population density, thinking that somebody else will call in a complaint? There are more police patrolling the area because of high crime. Once authorities are present and improvements aren't happening, what's the point of complaining? It is a known issue that many people call 911 instead of 311 for many calls. It would be interesting to see if the neighborhood calls the emergency number for quality of life issues instead of 311 in the belief that action is more likely to be taken with a 911 call. Unfortunately, logs of emergency calls are not in the public domain at this time, so this comparison cannot be easily accomplished.

Month-wise Analysis

- We analyzed the month of October 2012 with Tableau Public. In Tableau, we showed the NYC 311 map with 311 trends in October. We focused on the overall number of complaints during this month.
- We observed that during Hurricane Sandy period (Oct 22 - Oct 31), the number of complaints of damaged trees was enormously high (as expected). The maximum number of damaged tree complaints were from Staten Island where major destruction took place. Such kind of visualization could help the concerned authorities to get more prepared for such complaints in the future.
- Heating issues are a major concern in Manhattan and Bronx. In Staten Island, heating is not a major issue. As mentioned earlier, Staten Island is a residential community of primarily owner-occupied homes. Owners are responsible for their own heat. A call to 311 won't solve the issue. Noise is also a major problem in Manhattan in comparison with other neighborhoods.
- When we compared October to January complaints, we observed that heating was the major concern in all boroughs during these winter months.

Limitations

- Tableau Public has limited features because of which only excel, txt and access files can be uploaded for interactive visualization. It does not support connection to big databases like MySQL ,Oracle etc. Therefore, in our current web application, we tried to show the

visualization on a subset of data. But in future, if we use the enterprise version then we will be able to project the complete picture of our data.

- In our current web application, tableau public does not accept input from our web page. But we believe that this is also doable with the enterprise Tableau version.

Future Development

- For future development, an improvement would be to fetch the 311 data dynamically from nycopendata.org rather than from a database
- Another improvement would, of course, be to include Queens and Brooklyn. With the inclusion of this many neighborhoods, though, additional work on the User Interface would be beneficial.
- The Top 10 chart could benefit from a control with a scroll-bar or swipe to see further down the list from the initial 10.
- It might be useful to see other types of demographics for the neighborhoods being searched. The conditions in a neighborhood can change rapidly from one block to the next and varies with neighborhood activism, history, and pride of place.

Conclusions

With this application, we have made it easier to compare disparate neighborhoods in New York. A social scientist can look at data across neighborhoods and periods of time quickly and efficiently. The questions that immediately arise are interesting and the breadth is intriguing.

- What are the quality-of-life issues in different parts of the city, segregated by income level?
- Why are there relatively fewer complaints in low-income areas than in privileged areas, once corrected for housing issues?
- Residents of low-income areas don't have parking complaints and have no taxi complaints, presumably because of low car ownership and use of mass transportation. Is mass transportation adequate for these residents? Is the current level of availability a hindrance to upward mobility?
- Is 311 being used properly? How can it be used more effectively? What are the response rates to 311 complaints? What are the solution rates?

Continued studying of various combinations of neighborhoods and expanding the scope of this application, as suggested in Further Developments, can bring up more interesting situations that a social scientist would investigate further.

References:

NYC-311 Application

Main application:

<http://nyc-311.appspot.com/>

Prototype application:

<http://www.bigpalooka.com/bia660nyc2/default.aspx>

Tableau:

<http://www.tableausoftware.com/public/community>

Dashboard1(for the map):

http://public.tableausoftware.com/views/NYC311inJan2013/Dashboard2?:embed=y&:display_count=no

Dashboard2(for analysis statics):

http://public.tableausoftware.com/views/NYC_311_OCT/Dashboard3?:embed=y&:display_count=no

Data Sources

nycopendata.com 311 Service Requests from 2010 to Present

nycopendata.socrata.com/Social-Services/311-Service-Requests-from-2010-to-Present/erm2-nwe9

zipskinny.com

zipskinny.com/index.php?zip=10012

ZIP Code Definitions of New York City Neighborhoods

www.health.ny.gov/statistics/cancer/registry/appendix/neighborhoods.htm

Appendix

- Code for the website can be found in GitHub at [github.com/winteram/BIA660-2013S/BIA660-2013S/assigns/Final/Team NYC311/](https://github.com/winteram/BIA660-2013S/BIA660-2013S/assigns/Final/Team%20NYC311/)
- The Main application code for Google App Engine is in the subdirectory:
[/nyc-311](#)
- The prototype application code for Microsoft ASP is in the subdirectory:
[/BIA660NYC2](#)
- Python code used for extracting the data to MySQL and to CSV files is in the root directory as:
[bia660finalproject_v2.py](#)
- Sample code for extracting data from nycopendata

```
https://data.cityofnewyork.us/resource/erm2-nwe9.csv?
$query=select+complaint_type,count(complaint_type)+AS+CT+
where+(created_date>'2013-04-01'+AND+created_date<'2013-05-01')+
AND+(incident_zip='10029'+OR+incident_zip='10030'+OR+
incident_zip='10035'+OR+incident_zip='10039')+GROUP+BY+complaint_type
```

An example of the resulting SQL generated from this query string by Python :

```
INSERT INTO complaints (complaintdate,borough,neighborhood,complaint,CT)
VALUES ('2013-04-01','Manhattan','East Harlem','Derelict Bicycle',1);
```