NY CITY GOVERNEMNET FROM NYC OPEN DATA

The domain that I am dealing with is “NY CITY GOVERNMENT” domain which is a subdomain of NYC OPEN DATA domain.

Brief Introduction to NYC OPEN DATA:

Data is created every day by New York City government is a daily business. The people of NY can use this data to make informed decisions, become more engaged in their communities, solve tough problems, or turn their dreams into a reality.

The NYC OPEN DATA contains lot of sub domains making people life easier to go through exactly what sub domain they want to work on. Some of them are “CITY GOVERNMENT”, “TRANSPORTATION”,”ENVIRONMENT”, etcetera, which holds data related to their respective domain.

Every day, New Yorkers look for NYC Open Data. NYC Open Data is an invitation for anyone, anytime, anywhere to engage with New York City.

HOW THE NYC OPEN DATA EVOLVED:

In the early days of online government, cities rightly boasted about taking on "e-gov" as a project.

That process is well on its way in New York City, an early leader along with Chicago in using big (and little) data to drive innovation. New York is looking to use data science more broadly and deeply into existing institutions with an intention to remake governance.

The then-mayor Michael Bloomberg, a leader in technological innovation in both business and government, knew the importance of automating procedures and to control and make use of data analytics. Their major task is to of advance the city's innovation plan in the Office of Operations, which coordinates the agencies that provide the services most visible to the public. Data was a crucial part of their strategy.

Another officer called Mike Flowers, NY’s first chief analytics officer formed a team of young data scientists to produce a deep understanding of data which is not yet used. They worked fiercely on a few specific which they thought it could benefit the most from data analytics. In 2013, the mayor formalized team and created MODA (the Mayor’s Office of Data Analytics) as the “CIVIC INTELLIGENCE CENTER” where data from across agencies is aggregated, analyzed and turned into solutions which are helpful in performing actions.

MODA's Operations helped in a better change in city’s governance which now includes the Center for Economic Opportunity, which uses metrics to strengthen the city's antipoverty initiatives; NYC Open Data, which makes data from various city agencies available to the public.

Introduction to NY CITY GOVERNMENT:

New York City government gathers and reports data collected through a wide variety of government operations. It contains different types of data contributed by different agencies. The kind of data that the domain holds include data on inspections, tickets and parking and environmental violations, data on health of environment, data on water, air and urban landscape. Some of the main contributors of data to the CITY GOVERNMENT domain are:

Department of Information Technology and Tele Communications (DoITT)

Office of Administrative Trials and Hearings (OATH)

Department of Citywide Administrative Services (DCAS)

Department of Youth and Community Development (DYCD)

Department of Finance (DOF)

New York City Police Department (NYPD)

Department of Buildings (DOB)

Fire Department of NYC (FDNY)

Human Resources Administration (HRA) and many more.

New York City Fights Fire with Data (HOW FDNY PERFORMS ANALYSIS ON DATA)

The data driven analytics programs have created very useful and smart technologies to help in modernizing the manual tasks and make jobs more easy and efficient. Today, in most industries smart tools are developed which try to help in predicting the problems that occur in future so the decision makers can make better use of resources.

Predictive Analysis, for example, helps law enforcement pinpoint where a crime is likelier to occur on a given day based on the number of factors about a certain location. This helps in analyzing how the factors differed from daily, which hopefully helps officers to go to right places at right time to prevent crimes.

The New York City Fire Department (FDNY) has been using the Risk-Based Inspection System (RBIS), an Oracle-based program that has data mining capabilities to better check for fire accidents before they occur. The important piece of that tool is Fire Cast, an algorithm that organizes data from five city agencies into approximately 60 risk factors, which are then used to create lists of buildings that are most vulnerable to fire.

RBIS provides an inspection data warehouse the entire FDNY can access. Instead of being broken up among the city's 49 fire companies, any one of them can now look up the latest

information about when buildings and structures were last inspected. “This system digitally coordinates all of that and understands our various business rules.”

FUTURE ENHANCEMENTS

Despite its success in modernizing how the risk of fire accidents are calculated and the building inspections are done, the Fire Department is not resting on its success but is on it’s way to develop Fire Cast 3.0 and is expecting it to be a game changer in the city.

This third-generation algorithm will examine 7,500 factors across 17 city agency data streams. The new Fire Cast will also contain an element of artificial intelligence that helps in tracking the trends in all over the city.

They envision a machine that would notice trash violations in the South Bronx, and then if there is a fire accident in the same building within a period of 90 days , the program should learn and give trash violations in that area a higher risk rate while computing the area’s fire threat level. But the artificial intelligence element which is present in the algorithm must also recognize differing timeframes and should predict issues based on number of measuring tools.

In other words, the AI can’t “cry wolf” every time an incident occurs. Fire Cast 2.0 looks at constant factors about a building. But if successful, its successor will examine behavioral characteristics that could definitely bring a huge success and increase the level of capability on FDNY’s ability to compute a more accurate risk assessment.

WORKS AND ANALYSIS DONE BY MODA

Being a data-driven city is about more efficiently and effectively delivering the core services of the city. Being data-driven is not primarily a challenge of technology; it is a challenge of direction and organizational leadership.

http://beyondtransparency.org/chapters/part-4/beyond-open-data-the-data-driven-city/

For New York, a series of 2011 apartment fires helped galvanize our focus on the ability of data, in this case, the data that we already had is to save lives.

http://beyondtransparency.org/chapters/part-4/beyond-open-data-the-data-driven-city/

The challenge MODA faced was:

APARTMENT FIRES IN THE BRONX AND BROOKLYN:

In the spring 2011, a pair of house fires in apartment buildings in the Bronx and Brooklyn killed five people as a result of unsafe living conditions. When many people stay under unsafe

apartment conditions, with portable cooking devices, questionable electrical wiring, and inadequate fire escape access, catastrophic fires will occur for sure. The occurrence is all too common in a densely populated city like New York. The City receives over 20,000 citizen complaints a year from buildings suspected of being unsafe boarding houses.

New York collects an immense amount of information about every single one of their buildings. The data contains when and how buildings were built; it also contains if the building is receiving water service and if buildings are in good order based upon the location's history of ECB (environmental complaint board) violations on quality of life issues. Every day, they receive over 30,000 service requests (complaints) through 311 from New Yorkers, which gives them more location-specific intelligence.

With the help of data they know even more about the neighborhood where the building is located: they know how often 911 runs are made to that block, whether a road construction is being done, whether there are accidents in the intersections.

In the case of the fire in the two buildings, by the time they occurred, the City had information on tax liens, building complaints, sanitation violations, and building wall violations.

Now the questions they have are: Did we know enough about these buildings before the fire that should have intimated before? Could we determine which pieces of information are the most valuable predictors of catastrophic outcomes?

http://beyondtransparency.org/chapters/part-4/beyond-open-data-the-data-driven-city/

The team of the Mayor's Office of Data Analytics, are set to work to answer those questions.

WHEN THEY TRIED TO ANALYZE:

MODA team has started to use suitable tools and data, they started this project with two desktops and MS Excel which crashed after 36,000 rows of data**.** Using the undeveloped tools that are found on virtually every business machine, a talented young analyst was able to conduct the correlative analysis that helped them what to look for in the 311 complaints.

SOLUTION THEY CAME UP WITH:

By prioritizing the complaints that are most likely to be dangerous, they were remediating dangerous conditions faster, without any additional inspector manpower. That is a safety-based resource allocation plan.

ANALYSIS OF GOVERNMENT DATA USING PREDICTIVE ANALYSIS:

Predictive data analytics, like any data analysis, can only be as effective as the data

that they're based upon.

In that context, the approach that Mike Flowers and his data analytics team in New York City government have taken to detecting financial fraud and other crimes or problems is very interesting.

According to Flowers, applying predictive data analytics toward "preemptive government" in New York City has resulted in:

- A five-fold return on the time of building inspectors looking for illegal apartments.
- An increase in the rate of detection for dangerous buildings that are highly likely to result in firefighter injury or death.
- More than doubling the hit rate for discovering stores selling bootlegged cigarettes.
- A five-fold increase in the detection of business licenses being flipped.
- Fighting the prescription drug epidemic through detection of the 21 pharmacies (out of an estimated total of 2,150 in NYC) that accounted for more than 60% of total Medicaid reimbursements for Oxycodone in the city.

http://radar.oreilly.com/2012/06/predictive-data-analytics-big-data-nyc.html

EXAMPLE-1: ANALYSIS ON ILLEGAL CONVERSION:

The city receives roughly 20,000 to 25,000 complaints for something called an 'illegal conversion' every year. An illegal conversion is a situation where you have an apartment or a house that's zoned for six people to live in safely but the landlord's to make their own benefit puts 60 members together. They represent significant public safety hazards and not just from fire, but from crime and from epidemiological issues. To throw at those 20,000 to 25,000 complaints, we have roughly 200 inspectors to the Department of Buildings.

SOLUTION:

They first tried to prioritize the complaints which represent a greatest catastrophic risk as a structural fire. With his, they have built a basic flat file of about 900,000 structures in NY city and filled them with data from about 20 agencies. It contains data which contains whether an owner was in arrears on property taxes or if the property was in foreclosure, etc., Then they cross tabulated the data they were able to gather with the historical data which is about five years based on the structural fires in the city, with high severity.

After all this process is done, they found some things which are highly correlative to a fire which they cross checked with the inspectors of DOB and FDNY and asked them few questions to check whether the analysis they are going is in a right way and they got a positive response.

Next, they run every new complaint that comes in against that flat file. They find those complaints which represent the top five percent for historic fire risk and then send that top five percent back out to the inspectors to follow up on with urgency.

Conventionally, when the DOB went out to inspect a property because of a complaint, they were finding seriously high-risk conditions 13 percent of the time, they usually give a "vacate order" at such situations which is an order to empty the building in whole or in part is an extreme outcome but that's the only solution for such risks.

Using this newly developed and modernized system, they're finding these risky conditions at a sustained level of about 70 to 80 percent of the time in the complaints that these people send them out to [investigate]. From a DOB point of view that's a fivefold return on inspection man hours. From the FDNY standpoint, it also turns out that these buildings sent to them are 15 to 17 times more likely to result in a fireman being injured or killed. It's been going on for a year and this process is done on a weekly basis.

EXMAPLE-2: ANALYSIS ON CIGARETTE TAX INSPECTIONS:

Another area that MODA do predictive data analysis is on cigarette tax inspections.

We have a big problem in New York City with cigarette tax evasion. It costs about $12 a pack here. And that's because of taxes. And there's a reason for that. It's a social policy decision that government don't want people smoking because it has a lot of impact on public services.

But if we go to Virginia, however, we can pay $5 a pack. It's very easy to load up a van with 50 to 70 cartons that you buy in Virginia and then sell in the city.

SOLUTION:

In this instance, conventionally, the Department of Finance, in the Office of the Sheriff, goes out and does cigarette tax inspections around the city. Historically, their hit rate was just about 30 percent. We applied the same approach and the same methodology. Using our system, they're now somewhere close to 82 percent successful in finding the stores that are selling unstamped cigarettes.

Using this technique it is very useful for them because the conventional inspections they used make would cost them and are very expensive. For Sheriff, this has become a great tool because there is nearly threefold increase in efficiency.

EXAMPLE-3 ANALYSIS OF CIGARETTE TAX INSPECTION DATA TO FIND WHETHER THERE IS A FLIP:

The data that MODA use for the cigarette tax inspection also includes business licensing information from another agency called the Department of Consumer Affairs (DCA). They

license roughly 57 categories of businesses in New York City where there are about 150,000 different businesses in the city.

SOLUTION:

They first took the violations issued by the inspectors and used them to find out whether somebody would be violating the cigarette tax regulations or not. During this analysis they found out that they could use the data from other agencies to assist in DCA inspections.

With this kind of analysis they were able to find people who are unlawfully flipping their license. They could find this based on number of violations a particular person gets from DCA or any other agencies.

If a person gets a certain number of violations from DCA, then his license would be at risk at this time he tries to flip (paying someone for using their name to get a license on a clean slate but he is the actual one to run the business). If it's changed hands, but the violation activities remain consistent it is highly likely that there was a flip. This happens at some frequency because there are only about 65 inspectors to look after such things.

These people did the same analysis and sent their findings to DCA and the success rate here has been a fivefold again.

OPEN GOVERNMENT DATA: COMPANIES CASH IN

A big question is who is creating value to government data and how?

But who is actually creating that value and how?

A new website that has emerged www.OpenData500.com is created by the Governance Lab (GovLab) at New York University, is believed to be the most complete catalog that contains list of US based firms which totally or partially depend on government data.

It published a list of 500 companies. In addition to it, it also published profiles of 50 companies creating value from open government data. The companies range from startups such as Bill Guard, which analyzes publicly available billing dispute data to detect financial fraud, to more established firms such as iTriage, which helps consumers answer common medical questions and is now owned by Aetna.

Using this we can find which companies are using which agency datasets.
http://www.opendata500.com/us/

WITH GOVERNMENT DOMAIN IT IS EASY TO ANALYZE DATA:

The government domain provides public with very ease of access to data as well as sophisticated tools online to analyze the data then and there.

Recent Manhattan transplant Peter Gilks creates a dashboard that allows him to map and explore data on the city's 311 calls, so he can safely navigate his new home while avoiding rats, roosters and loud noises. He does it in 3 minutes flat**.**

This video can be found at http://3minutewin.tumblr.com/post/70416962762/recent-manhattan-transplant-peter-gilks-creates-a

One best example is NYC Jobs dataset which is updated on a daily basis providing the new Yorkers best way to search for jobs based on location. http://www1.nyc.gov/jobs/index.page really helps people in searching for jobs related to their specific area of interest.

CONCLUSION:

This paper shows how the NY City Government has transformed its way of governance using big data analysis. This paper covered a considerable number of problem statements and kind of analysis made to find a solution for that problem statements. It also included a list of companies that are really dependent on the government data. Big Data analysis on Government domain has become a new trend which really helped the governments in improving their conventional way of handling and analyzing the data and the way of governance. The concept of making government datasets public really helped people in many ways. There are really lot of startups and well developed companies that are fully or partially dependent on Government data and are benefited to a large extent.

REFERENCES:

http://www.nyc.gov/html/analytics/html/initiatives/open_data.shtml

http://beyondtransparency.org/chapters/part-4/beyond-open-data-the-data-driven-city/

https://github.com/tlevine/open-data-500/blob/master/download/500_Companies.csv

http://www.opendata500.com/us/

http://www.informationweek.com/government/open-government/open-government-data-companies-cash-in/d/d-id/1113143

http://radar.oreilly.com/2012/06/predictive-data-analytics-big-data-nyc.html

http://www.governing.com/blogs/bfc/gov-new-york-city-data-driven-governance.html

http://www.govtech.com/public-safety/New-York-City-Fights-Fire-with-Data.html

http://3minutewin.tumblr.com/post/70416962762/recent-manhattan-transplant-peter-gilks-creates-a

http://www1.nyc.gov/jobs/index.page