Urban Event Detection using Time Series Analysis

Employing taxi data for identifying happening places in NYC.

Akshay Trivedi

Shivendra Panicker

Users:

City planners, NYC public

Beneficiaries:

City planners : building new transportation systems, eateries, etc.

People : places to hangout, happening/interesting spots

Importance:

Cities are complex systems and exploring the data they generate is challenging due to its sheer volume and the inherent spatio-temporal complexity of this data. We enable both real-time monitoring and long-term analysis of social dynamics of a city.

The Dataset

1. NYC TLC Dataset

- 1.1 billion yellow and green taxi trip records from 2009 capturing pick-up and drop-off dates/times, pick-up and drop-off locations, trip distances, fares and passenger counts
- 267 GB

2. NYC Uber Dataset

- 4.5 million Uber pickups in NYC from April to September 2014, and 14.3 million from January to June 2015.
- 1 GB

Our Approach

Construct time series dataset

Decompose Time Series

Extract Important Anomalies

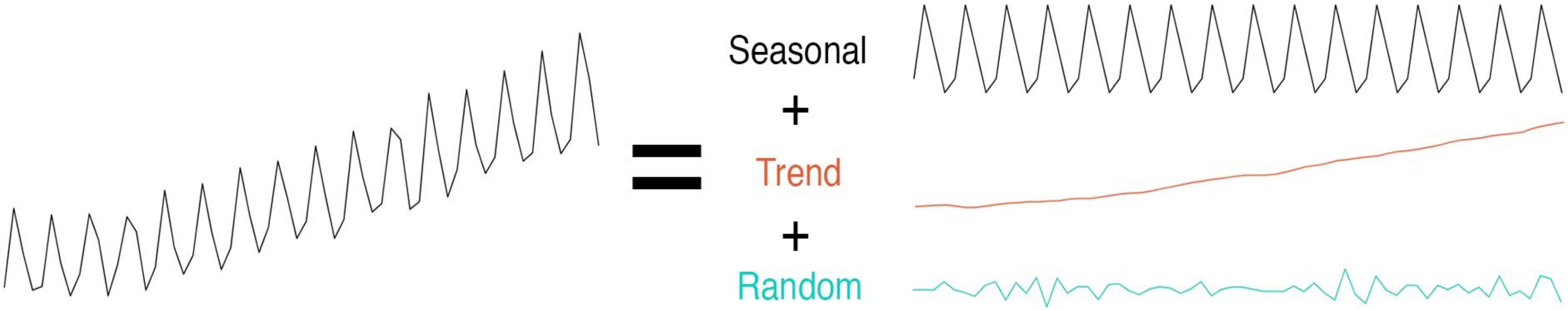
What is Time Series Data?

"A sequence of measurements of the same variable collected over time. Most often, the measurements are made at regular time intervals."

But trip data is not at regular intervals and doesn't measure anything?

1. Divided NYC into 20*20 grids.
2. Measured frequency of pickups and dropoffs for each grid.
3. Generated 3 time series with intervals - **Daily, Weekly, Annual**

Time Series Decomposition



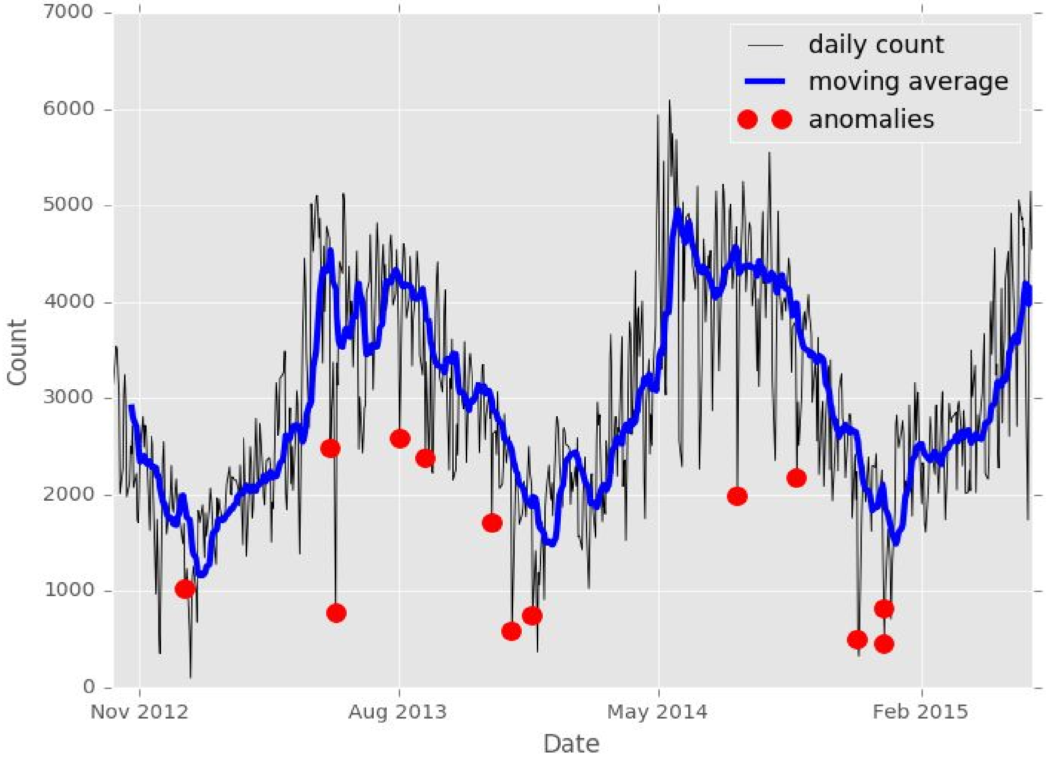
Each location has seasonal (continuous) + trend (expected) and residual (anomalous) traffic.

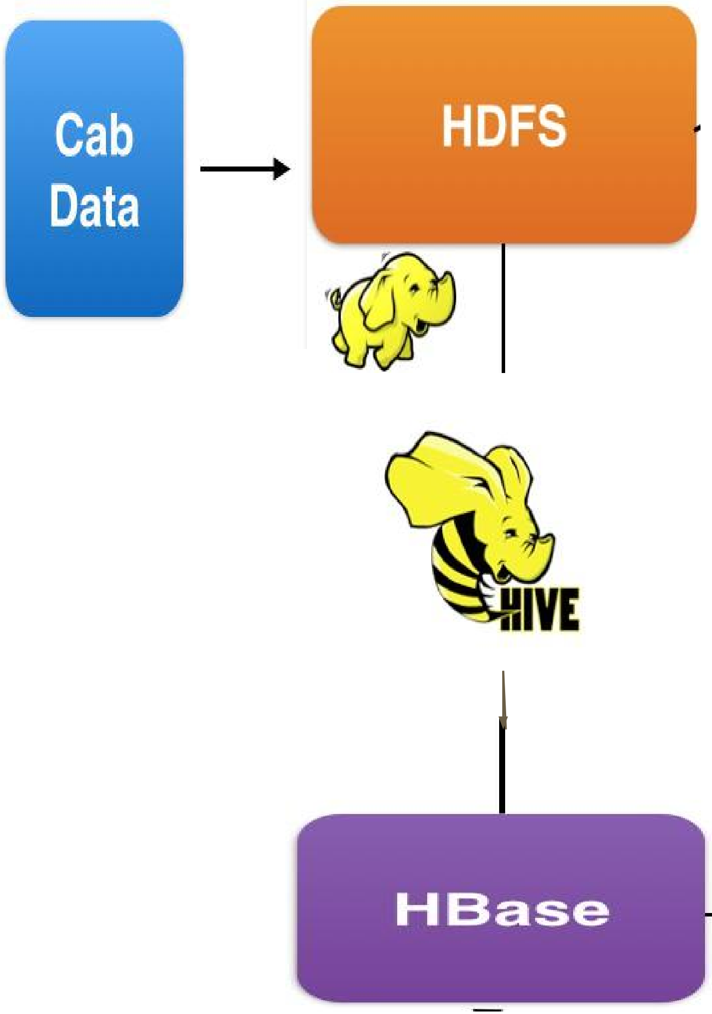
Importance of Anomalies

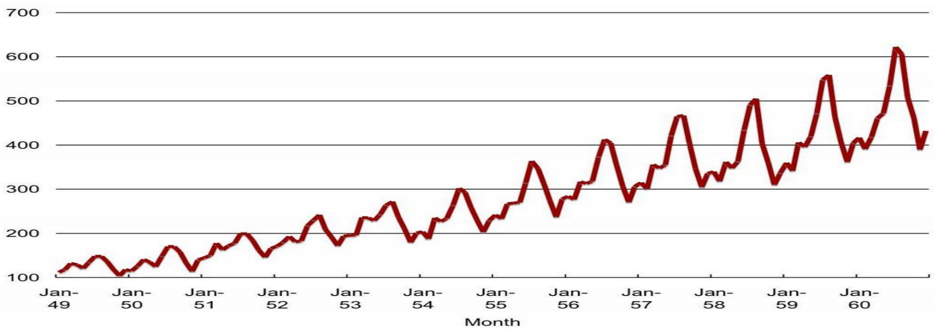
Not all anomalies are created equal!!

We wanted the top -100 most anomalous locations.

Utilized Z-score for finding statistically important anomalies.







Hourly Aggregates by Day of Week

y_m_dow	c:0	c:1	c:2	c:3	c:4	…	c23	c:Totals
Day of Week	Hour 0 Attributes	hr 1	hr 2	hr 3	hr 4	…	hr 23	..
2008_01_Mon	pickups, dropoffs, avg_occ, avg_dist	..	..	..	..	..	..	sum(pickups), sum(drop offs), avg(occ), avg(dist)

Goodness of our Hypothesis

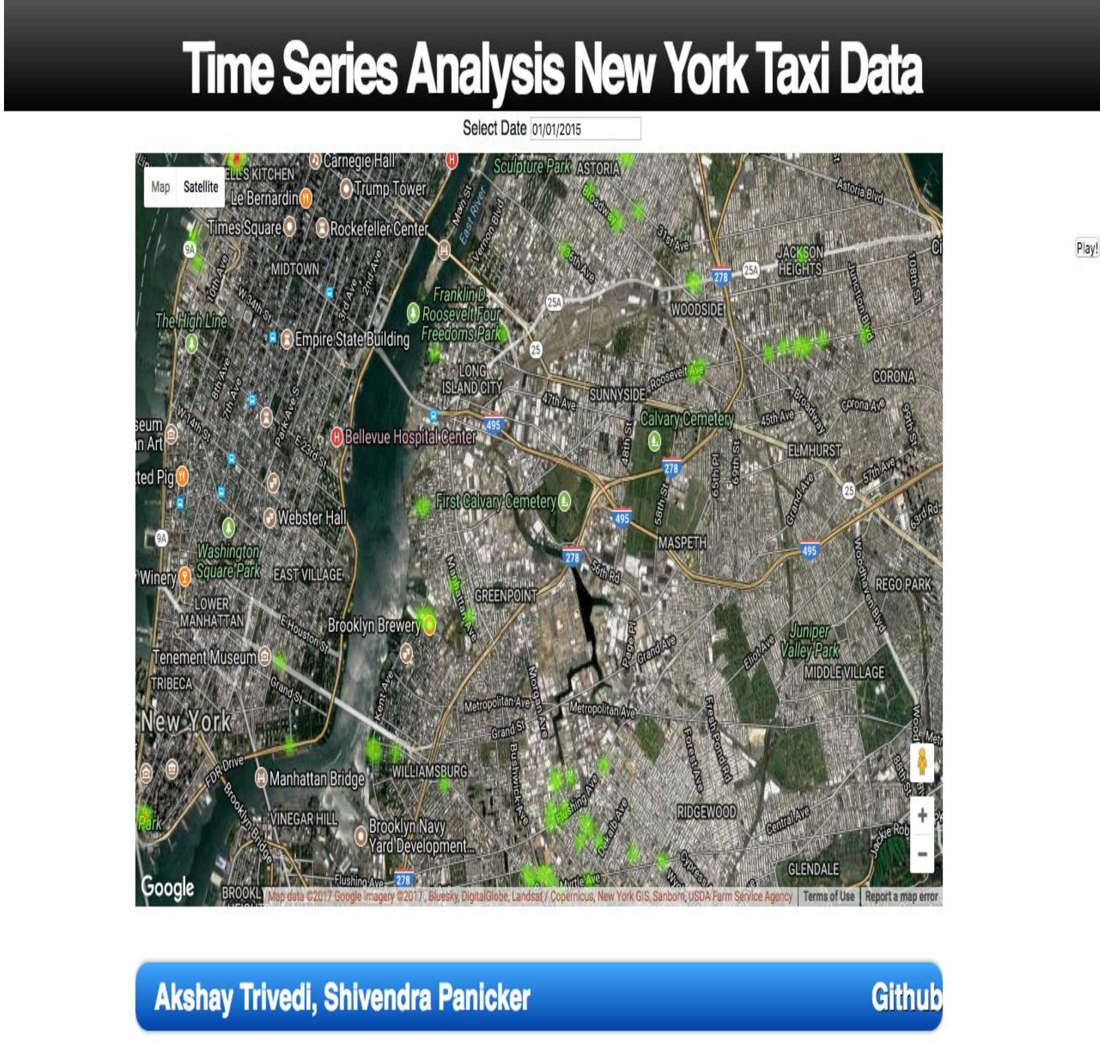
Our approach accurately detected a collection of interesting events at different spatial and temporal scales ranging from regional events such as festivals, mass gatherings to local events such as exhibitions, football games, parades.

Some representative examples:

1. Visit of Indian Prime Minister at Madison Square Garden
2. Macy's Thanksgiving parade
3. 9/11 memorial gathering

Demo

http://linserv1.cims.nyu.edu:45673



Results

1. Visualise and pinpoint top 100 most significant anomalous locations in New York per day.

2. Long term analysis of social dynamics of NYC.

Challenges

1. Working with geospatial data:

   Some trips have pickup/dropoff coordinates that are in Hudson!! and few trips are outside NY. Identifying and cleaning up invalid coordinates and then grouping trips into 20*20 grid was challenging.

2. Learning about Time Series and implementing TS Decomposition using Map-Reduce paradigm.

Conclusion & Acknowledgements

- Decompose time series per location into trend, seasonal, and remainder.
- Detect and filter statistically significant events from the extracted remainder component.
- Visualize the resultant heat map.
- Easily extendible to real-time data feed, showing current events via historical records.

We acknowledge

- HPC team for helping us along the way.
- Xi Zhu from University of South Carolina

References

[1] Zhu, Xi, and Diansheng Guo. "Urban event detection with big data of taxi OD trips: A time series decomposition approach." *Transactions in GIS* 21.3 (2017): 560-574.

[2] Ferreira, Nivan, et al. "Visual exploration of big spatio-temporal urban data: A study of new york city taxi trips." *IEEE Transactions on Visualization and Computer Graphics* 19.12 (2013): 2149-2158.

[3] Doraiswamy, Harish, et al. "Using topological analysis to support event-guided exploration in urban data." *IEEE transactions on visualization and computer graphics* 20.12 (2014): 2634-2643.

Thank you!