Square cities: time dimension

Philipp Kats

12/16/15

1 Problem Description

This project is a continuation of two previous research projects:

- Senseable Moscow "Moya Moskva" project 1 ()
- NYU CUSP Applied Data science research project 2 ()

All three projects are based on the same idea of explaining significant differences between cities stats, using foursquare venues data. This particular project, however, bounds to the the temporal dimension of data, analyzing venue creation through last 5 years for 8 cities, including New York, San Francisco, Shanghai, Mumbai, Moscow, Singapur, Kiev and Minsk.

This research aims to explore three questions stated below:

- Do all cities have similar "temporal behavior" on venue registration?
- Do they perform similar behaviour in terms of vanue-category granulated timelines?

While we are driven by scientific curiosity, This is not the research for the sake of research, as the answers to those questions, potentially, may lead us to the between ? understanding of the spatial-economical behaviour of cities.

2 Data

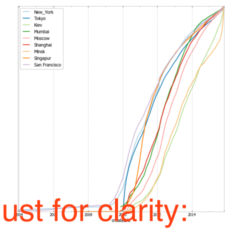
Research is based on foursquare service data on venues (locations), collected through official API with the custom `data collector`. Scraper was collecting all venues created at any time and still existing[1] for the given location. Location at this moment is defined by the coordinate rectangular.

3 Methodology

Research will be based on different time-series analysis techniques, time-series correlation, KS2 tests and K-mean clustering. more detail here!

3.1 TimeLines Overview

…. i know you know better than this: i csannot see this plot!! its waaaay too small!



you keep saying they opened at this date, but just for clarity:
this is not the open date but the date they appear on foursquare, whih includes a bias on the foursquare usage

Figure 1: Venue creation normalised timeline

On the general time line we can see that the general trend is similar, as most of the growth for each city happened on the range from 2010 till now. However, even without clustering we can see 4 groups of cities with similar behaviour:

- American cities, **San Francisco** and **New York** started much earlier, in 2004. Both cities grew fast from 2009 till 2013, where they were slowing down. We may assume that in 2013 forsquare covered most of *existing* places in thse cities.

- **Tokyo** and **Singapur** started in 2010, and both skyrocketed to the level of american cities in 2011, bahaving similarly after that.

- **Minsk** and **Kiev** are the most “late” cities in the set, as they started growing rapidly in 2012, and both have similar dynamics till now - we can assume that both of them did not finish the “extensive” phase, in other words, not all existing significant places were described in the service.

- All other cities (**Shanghai**,**Moscow**, and **Singapur**) have average behavior, grewing relatively slowly from 2010, and slowing down to “american” behaviour from 2014.

why average? intermediate perhaps

4 Categories

At this point we decided to go into details, looking at time line patterns for particular general categories, defined by the Foursquare itself.

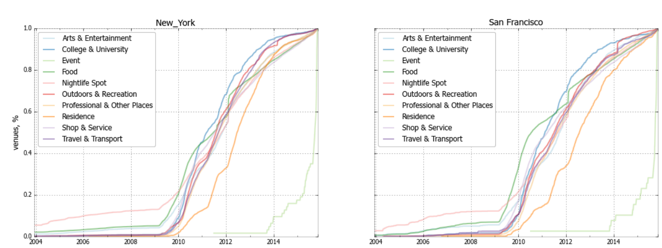


Figure 2: Normalized Timeline of venue creation, for New York and San Francisco

Two time lines reveal interesting (though expected) behaviour: while at the start most active venues were Nightlife, Food and Entertinements (those most interested in the free advertisement of this sort), on the peak of dynamics most active categories were “Hight Education”, “Travel” and “Outdoor” - we can (wildly) assume that while Education reveals the “true face of the user”- a student, other two categories are the most important and useful from the general user’s point of view.

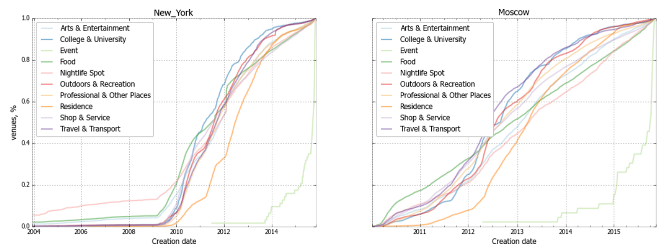


Figure 3: Normalized Timeline of venue creation, for New York and Moscow

To a surprise, those two plots are quite similar as well, with the same top-3 categories for the growth period. The only interesting feature here is low starting action of “Entertinement” and “Nightlife”, bringing up the question, ~~whatever~~ they are not that developed or not that interested in this kind of advertisement. whether

you should address the behavior of “events”: a new category?

5 Conclusions

In this short research we looked at the time patterns for 8 cities, analysing their differensies and similarities. We were able to establish 4 patterns of general activities for those cities and look at category-based time lines, Which allowed us to refine “pro-active” (self-advertising) categories and those covered by users themselves - Transportation, Outdoor/Recreation and Education, which might be interpreted as the sign of their value and importance, even bearing in mind the propensity of the users over general population.

6 Future work

At the next stage we are planning to drill deeper into the data, looking for the time patterns for particular areas. S Same time, It looks to be a good idea to cluster the cities by their category-based time behaviour.

7 Links

so in the end did you do any statistical analysis (correlation tests etc?) or just by-eye clustering? if you did not you should removed them from the methodology, if you did you should mention their result

- `Project repository`
- `Main project repo`

References

Philipp Kats. Senseable Moscow “Moya Moskva” project, jul 2012.
Philipp Kats, Xia Wang, and Vipassana Vijayarangan. Square cities: Measuring cities with Foursquare API, nov 2015.