Economics Bulletin

Volume 35, Issue 2

Using location-based social networks as a novel source of marketing data: the case of shopping malls

Elena B. Pokryshevskaya
National Research University Higher School of Economics

Evgeny A. Antipov
National Research University Higher School of Economics

Abstract

In this paper we demonstrate how publicly available data from location-based social networks can be used to model the popularity of different places. Our empirical analysis is based on a sample of 112 shopping malls located in Saint-Petersburg, the second largest city in Russia. Regression models that explain various measures of shopping malls' popularity using the characteristics of a place are built. It appeared that the number of visitors, check-ins, tips and even the frequency of visit can be largely explained by the basic features of a shopping mall, while successful modeling of a place's user rating requires more detailed data. Combining the data on the features of places, text reviews and popularity indicators from social networks is a promising approach for building sales, traffic, satisfaction and loyalty projection models, which are especially useful in business planning.

We thank Darya Morozova, the student of Higher School of Economics, for her invaluable help with data collection.
Citation: Elena B. Pokryshevskaya and Evgeny A. Antipov, (2015) "Using location-based social networks as a novel source of marketing data: the case of shopping malls", *Economics Bulletin*, Volume 35, Issue 2, pages 1292-1298
Contact: Elena B. Pokryshevskaya - epokryshevskaya@hse.ru, Evgeny A. Antipov - eantipov@hse.ru.
Submitted: August 22, 2014. **Published:** June 01, 2015.

1. Introduction

When customers check in using a location-based social network, they automatically tell their friends about a place they are visiting. Not accidentally companies are so interested in stimulating customers to check in. Nowadays the number of check-ins can easily become one of the target indicators for marketing departments of many firms from service industries. Marketers are designing ways to get customers to check-in as often as possible (Gallo, 2011).

Besides being useful as a social media marketing tool, location-based social networks provide a wealth of data for marketing analytics. In this paper we suggest using Foursquare.com data to conduct the analysis of competition on various markets. We apply our approach to the analysis of shopping malls in Saint-Petersburg, one of the largest European cities. Besides using its unique ratings of various places of interest, we suggest using the number of visitors, check-ins and tips as indicators of a place's popularity. Taking into account that the characteristics of shopping malls (as well as most other places) are publicly available, such kind of data on place's popularity is great for forecasting purposes: namely, for projecting the number of visitors, check-ins, etc. It is especially valuable taking into account that while you may know the exact number of visitors of your own shopping mall, it is unlikely that you have that information about your competitors. Such information is either very expensive or is not collected by anybody. Anyway, updating such kind of data regularly through primary research would cost an astronomical amount of money.

Social networking websites and applications offer new opportunities for conducting market research based on publicly available secondary data. Most research studies used location-based social networks for attaining a deeper understanding of human mobility (Cheng, Caverlee, Lee, & Sui, 2011; Colombo, Chorley, Williams, Allen, & Whitaker, 2012; Cramer, Rost, & Holmquist, 2011; Nguyen & Szymanski, 2012; Noulas, Scellato, Mascolo, & Pontil, 2011) including next place prediction (Noulas, Scellato, Lathia, & Mascolo, 2012). A few papers are related to fraud in social networks (Carbunar & Potharaju, 2012; Glas, 2013). Most of these papers are related to the technical issues of location-based social networks, whereas there seem to be many opportunities for using such networks in market research, which are worth studying. In our opinion location-based social networks can offer even more useful data for marketing scholars than traditional social networks such as Facebook (Chu, 2011; Erdogmus & Cicek, 2012) as they contain explicit information on the popularity of different places, such as cafes, stores, museums, etc.

We have found no applications of location-based social networks such as Foursquare.com to studying the determinants of a place's attractiveness to customers or to assessing a new place's commercial potential. To the best knowledge of the authors, we are the first to model the popularity of places as a function of their characteristics, i.e. combining data from a location-based social network with the data on a place's characteristics from other sources.

2. Data

We have collected data on 112 shopping malls located in Saint-Petersburg, the second largest city in Russia. According to Colliers International the city had 489m^2 of retail space per 1000 residents by the end of 2013, which ranked it first in Russia and sixth in Europe (after Warsaw, Budapest, Prague, Zagreb и Athens). The growth rate of retail real estate market in this city is also impressive and the market is far from saturation: nine shopping centers with a total area of 346000m^2 opened in Saint-Petersburg in 2013.

The data was collected in May 2014 from several websites:

1. ***Foursquare.com*** is a web and mobile application that allows registered users to post their location at a venue ("check-in") and connect with friends. This community totals over 50 million people worldwide. Information on such variables as **a place's rating**,

number of votes, **visitors**, **check-ins** and **tips** was gathered from this source. While votes, visitors and check-ins are self-explanatory, we should define a place's rating and tips. Tips are public notes that you leave at a place for other Foursquare users about experiences you recommend (or recommend be avoided). A place's rating is based on a number of signals that Foursquare's team gathered from their social data mines; likes and dislikes, and positive versus negative tips. When users like a business or leave a tip, the score won't change automatically, as it's based on a number of different factors. So it is important to highlight that Foursquare uses an intellectual rating system that is not prone to the manipulations that are typical of some websites with less intricate rating systems.

2. ***Shopandmall.ru*** is a website devoted to commercial real estate, and particularly to shopping centers in Russia. Most of the technical characteristics of shopping malls can be obtained from this website. These characteristics include ***age of the shopping mall***, ***total area***, ***leasing area***, ***total number of stories***, ***number of trade stories***, ***number of parking places***, ***number of working hours a day*** and entertainment facilities (i.e. ***cinema*** and ***food court***).
3. Data on price offer for rent per square meter was collected from ***Account.spb.ru***. This particular source provides relevant material on commercial real estate market in Saint-Petersburg.
4. In addition, distance from each shopping mall to the nearest subway station was calculated using ***Google Maps***.

After excluding a few cases with missing data we got a dataset of 112 observations which represent most of the shopping malls in Saint-Petersburg.

3. Empirical analysis

In our regression analysis we used the following inputs:

- Leasing area (in 1000s square meters)
- Other area (Building area-Leasing area, in 1000s square meters)
- Number of hours the mall is open every day
- Age of the shopping mall (in months)
- Number of parking places
- Distance from the nearest subway station (in 100s meters)
- Movie theater (1 – yes, 0 – no)
- Food court (1 – yes, 0 – no)
- Monthly rental (rubles per square meter a month)

Using monthly rental as input may seem questionable, but we think its inclusion as a control variable is justified by several reasons:

- Tenants have to take market rental rates into account, but they still can vary to a large extent, so it is a controlled variable for the management, rather than a fully endogenous one.
- Rental serves as a proxy for the district in which the mall is situated. The R^2 of the regression of rental on district-specific dummies is 0.45, so almost half of the variation in monthly rental is explained by the geographic differences.

The set of explanatory variables is characterized by low multicollinearity (none of the variance inflation factors exceed 2.4), which is good for the further regression analysis.

The outputs are:

- Place rating (out of 10)
- Number of visitors
- Number of check-ins
- Number of tips

- Frequency of visit=Number of check-ins/Number of visitors

Descriptive analysis (Table 1) has shown that the number of check-ins, the number of visitors and the number of tips are very heterogeneous variables (standard deviation exceeds mean), while the frequency of visit and the rating are much more homogeneous (standard deviations are just 36% and 11% of the mean, correspondingly).

Table 1. Descriptive statistics

Variable	**Obs**	**Mean**	**Std. Dev.**	**Min**	**Max**
Rating	112	8.00	0.90	5.20	9.20
Number of check-ins	112	29015.79	59728.19	325.00	567866.00
Number of visitors	112	9328.39	12774.82	184.00	102697.00
Number of tips	112	86.04	114.05	1	925
Frequency of visit	112	2.61	0.93	1.02	8.31
Leasing area (in 1000s square meters)	112	28.85	26.78	0.25	139.00
Other area (in 1000s square meters)	112	13.82	18.78	0.00	100.00
Number of hours the mall is open every day	112	11.36	0.98	7.00	13.00
Age of the shopping mall (in months)	112	87.46	60.77	7.00	557.00
Number of parking places	112	789.73	1201.56	0.00	10000.00
Distance from the nearest subway station (in 100s meters)	112	17.59	31.42	0.59	227.00
Monthly rental (rubles per square meter a month)	112	1097.95	459.79	590.00	2500.00
Movie theater (1 – yes, 0 – no)	112	0.34	0.48	0.00	1.00
Food court (1 – yes, 0 – no)	112	0.73	0.44	0.00	1.00

The number of visitors, the number of check-ins and the number of tips are strongly correlated with one another, which can be explained by the fact that they are all correlated with the actual mall's traffic (Table 2). Rating and frequency are associated with satisfaction and loyalty, correspondingly, and, therefore, are less correlated with traffic measures.

Table 2. Pearson correlation coefficients

	Rating	Visitors	Check-ins	Tips	Frequency
Rating	1.00				
Visitors	0.45*	1.00			
Check-ins	0.36*	0.96*	1.00		
Tips	0.46*	0.95*	0.93*	1.00	
Frequency	0.31*	0.40*	0.44*	0.40*	1.00

Factor analysis using principal components extraction with varimax rotation confirmed the idea that the number of visitors, check-ins and tips represent a different dimension than frequency and rating (Table 3). The 2-factor solution explains 84% of the total variance.

Table 3. Rotated component matrix

	Component	
	1	2
Check-ins	0.950	0.240
Visitors	0.948	0.274
Tips	0.933	0.292
Frequency	0.208	0.789
Rating	0.234	0.769

The factor scores from this dimension-reduction procedure can be used to segment shopping malls based on traffic and satisfaction/loyalty dimension. Our simplified segmentation (Table 4) has shown that less than 20% of shopping malls in Saint-Petersburg have both high traffic and high satisfaction/loyalty (i.e. both factor scores are positive).

Table 4. Traffic vs. satisfaction segmentation of shopping malls (% of all shopping malls)

		satisfaction	
		low	high
traffic	low	34.8%	35.7%
	high	9.8%	19.6%

Robust regression analysis[1] has shown that visitors give higher ratings to large (in terms of leasing area) malls with a food court (Table 5). A food court on average adds almost 0.4 points out of 10 to a shopping mall's rating. Other things equal, a mall with the leasing area of 100000 square meters is awarded with 0.64 points higher rating than a mall with the area of just 20000 square meters. Cinema increases the number of check-ins and visitors but not the place's rating, though. The effect of the proximity to a subway station is either statistically insignificant, or statistically significant, but practically negligible. For example, when distance increases by 1 km (i.e. 10 hundred meters) the number of tips decreases by about 2.6, which is just 3% of the average number of tips and 2.3% of its standard deviation.

Table 5. Robust regression analysis parameter estimates

	(1) Rating	(2) Visitors	(3) Check-ins	(4) Tips	(5) Frequency
Leasing area (1000s sq. meters)	0.008*	69.300**	138.681**	0.713***	0.002
	(0.004)	(27.247)	(64.482)	(0.235)	(0.003)
Other area (1000s sq. meters)	0.007	98.794***	184.687**	1.333***	0.028***
	(0.005)	(31.468)	(74.470)	(0.272)	(0.004)
Number of hours a day	-0.093	984.563*	2011.315	5.580	-0.001
	(0.087)	(561.754)	(1329.429)	(4.850)	(0.066)
Age (months)	-0.000	14.502*	41.505**	-0.022	-0.001
	(0.001)	(8.474)	(20.053)	(0.073)	(0.001)

[1] Stata 13's *rreg* command first performs an initial screening based on Cook's distance> 1 to eliminate gross outliers before calculating starting values and then performs Huber iterations followed by biweight iterations.

Parking places	0.000	2.445***	8.705***	0.015***	-0.000
	(0.000)	(0.516)	(1.222)	(0.004)	(0.000)
Distance from subway station (100s meters)	0.000	-24.716	-33.373	-0.259*	0.006***
	(0.002)	(16.038)	(37.955)	(0.138)	(0.002)
Monthly rental (in RUR)	0.000	2.969**	4.376	0.018*	0.000
	(0.000)	(1.219)	(2.885)	(0.011)	(0.000)
Movie theater	0.298	2051.559*	7810.860***	14.914	0.202
	(0.180)	(1213.802)	(2872.543)	(10.479)	(0.144)
Food court	0.391*	1185.961	1276.239	8.991	0.013
	(0.207)	(1199.564)	(2838.848)	(10.356)	(0.141)
Constant	8.193***	-14323.655**	-29592.572**	-65.207	2.110***
	(0.992)	(6261.148)	(14817.428)	(54.055)	(0.739)
R^2	0.264	0.637	0.693	0.629	0.549
adj. R^2	0.199	0.604	0.666	0.596	0.509

Standard errors in parentheses
* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Every additional thousand square meters of leasing area increases the number of visitors by 69 people. The number of visitors is also positively influenced by the number of working hours, the age of the shopping mall, the number of parking places and the presence of a movie theater. Other area and monthly rental are also positively correlated with the number of visitors. Increasing the number of parking places by 400 on average brings about 1000 extra visitors to the shopping mall, so it may be a reasonable investment. Shopping malls with an insufficient number of parking places obviously loses visitors that drive a car, since they will tend to substitute such a mall with another where the parking lot is larger. A movie theater attracts on average over 2000 additional visitors to a shopping mall. We do not have a clear explanation of why “other area” matters, but it may be the case that people often check-in while seating in a shopping mall’s lobby or some other recreation area. So it is desirable to organize such zones to stimulate check-ins. The number of check-ins is driven by almost the same factors as the number of visitors, except for the number of operational hours.

The frequency of visit (check-ins to visitors ratio) is determined by “other area” and distance from a subway station. “Other area” may stimulate people to visit the mall not only when they need to buy or eat something, but use it as a walking area or a place to meet with friends. Surprisingly, despite the popularity of subway in Saint-Petersburg, frequency of visit is the only one dependent variable that is significantly (at 1% significance level) influenced by the proximity of subway. A possible explanation: it is easier for people to *regularly* reach those malls that are situated near a subway station, because most people use subway to travel around the city.

Older shopping malls attract more visitors and therefore check-ins, which is not surprising because they have more time to collect check-ins on Foursquare.com. We discuss this limitation in Section 4 of this paper. However, rating, the number of tips and the frequency of visit are not influenced by the age of the shopping mall.

4. Conclusion

Social networking indicators of a place’s popularity are not direct substitutes of traditional financial indicators, but still complement them well as they measure how interesting and demanded a place is for users of social networks.

Besides the indicators that are readily available from Foursquare.com (number of visitors, check-ins, number of tips and a rating score), we suggest using the frequency of visit, which is the ratio of check-ins to the number of visitors. This measure reflects customer loyalty. While location-based social networks give a wealth of indicators that can be used for market segmentations, rankings, etc., we have concentrated on using them for building explanatory and predictive models.

According to robust regression analysis big shopping centers with a food court get higher ratings than others. The number of check-ins is driven by the leasing area, other area, mall's age, the number of parking places and the presence of a movie theater.

The explanatory power of models with such dependent variables as the number of visitors, check-ins, tips and even frequency of visit is rather high (R^2 exceeds 0.54). This means that investors can assess the potential of a shopping mall based even on a few characteristics. However, the precision of our model of a place's rating leaves much to be desired. Explaining, what determines customer satisfaction with shopping malls using publicly available data is a challenging direction for future research, which may require using text mining techniques as well as collecting more detailed information about the shopping mall's characteristics.

Finding a link between the number of visitors or check-ins (based on Foursquare.com data) and the actual traffic counting data could give an opportunity to do traffic projections for new shopping malls. This task requires proprietary data and may be hard to do, but even without actual traffic data one can predict the popularity of the place within a location-based social network's users and make a conclusion which shopping mall concept or place is better.

Even though shopping mall management may really be interested in all-time popularity measures (as users of Foursquare.com see them and may use them in their decision making), for future research we suggest limiting sample to a specific period of time (say, last 3 months) to reflect only recent trends. This is especially important in the case of the number of visits and the number of check-ins, since marketing departments are interested primarily in current figures rather than cumulative ones. Regular tracking allows limiting the time period to a few recent months easily.

5. References

1. Carbunar, B., & Potharaju, R. (2012). You unlocked the Mt. Everest badge on foursquare! Countering location fraud in geosocial networks. In *Mobile Adhoc and Sensor Systems (MASS), 2012 IEEE 9th International Conference on* (pp. 182–190).
2. Cheng, Z., Caverlee, J., Lee, K., & Sui, D. Z. (2011). Exploring Millions of Footprints in Location Sharing Services. *ICWSM*, *2011*, 81–88.
3. Chu, S.-C. (2011). Viral advertising in social media: Participation in Facebook groups and responses among college-aged users. *Journal of Interactive Advertising*, *12*(1), 30–43.
4. Colombo, G. B., Chorley, M. J., Williams, M. J., Allen, S. M., & Whitaker, R. M. (2012). You are where you eat: Foursquare checkins as indicators of human mobility and behaviour. In *Pervasive Computing and Communications Workshops (PERCOM Workshops), 2012 IEEE International Conference on* (pp. 217–222).
5. Cramer, H., Rost, M., & Holmquist, L. E. (2011). Performing a check-in: emerging practices, norms and'conflicts' in location-sharing using foursquare. In *Proceedings of the 13th International Conference on Human Computer Interaction with Mobile Devices and Services* (pp. 57–66).
6. Erdogmus, I. E., & Cicek, M. (2012). The impact of social media marketing on brand loyalty. *Procedia-Social and Behavioral Sciences*, *58*, 1353–1360.
7. Gallo, C. (2011). *The Power of Foursquare: 7 Innovative Ways to Get Your Customers to Check In Wherever They Are* (1st ed.). Mcgraw-Hill.

8. Glas, R. (2013). Breaking reality: Exploring pervasive cheating in Foursquare. *Transactions of the Digital Games Research Association*, *1*(1).
9. Nguyen, T., & Szymanski, B. K. (2012). Using location-based social networks to validate human mobility and relationships models. In *Advances in Social Networks Analysis and Mining (ASONAM), 2012 IEEE/ACM International Conference on* (pp. 1215–1221).
10. Noulas, A., Scellato, S., Lathia, N., & Mascolo, C. (2012). Mining User Mobility Features for Next Place Prediction in Location-Based Services. In *ICDM* (Vol. 12, pp. 1038–1043).
11. Noulas, A., Scellato, S., Mascolo, C., & Pontil, M. (2011). An Empirical Study of Geographic User Activity Patterns in Foursquare. *ICWSM*, *11*, 70–573.