Proposal for a Forecasting Methodology to Predict Commercial Real Estate Values in Istanbul Using Social Big Data

Maral TAŞCILAR
Istanbul Technical University
Graduate School, Real Estate Development

Dr. Kerem Yavuz ARSLANLI
Istanbul Technical University
Department of Urban and Regional Planning

ERES 28th Annual Conference 2022
SDA Bocconi School of Management, Milan, Italy
24.06.2022

Contents

Problem Statement & Motivation

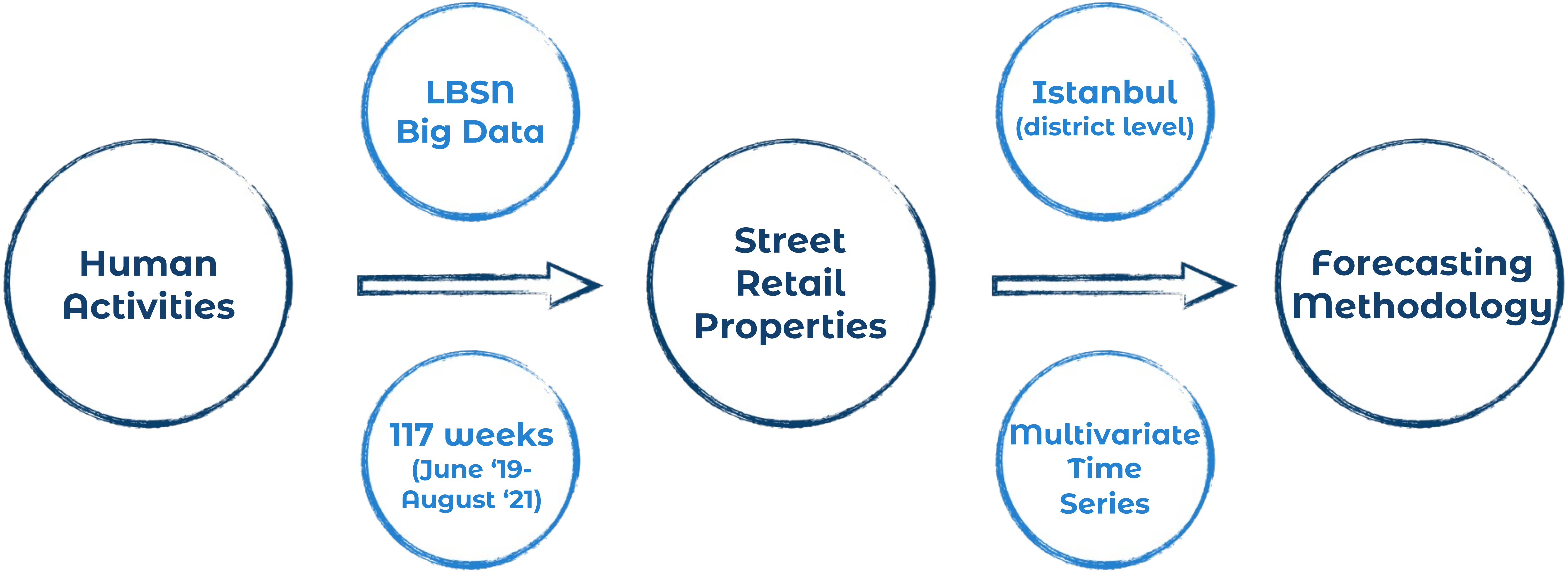
Problem:

The **fluctuations in the retail property prices** in a city with 16 million inhabitants combined with the **insufficient public indicators** about commercial real estates in Turkey result in significant **risks for the decision-makers**.

Motivations:

- The unveiled dependence of real estate sector to **human activity**
- The potential of adapting **new technologies** real estate studies
- Covering the need for a **reliable forecasting methodology** to minimize the investor risks

Introduction



Literature Review

Author(s)	Year	LBSN Platform	Outcomes
Cheng et al.	2011	Twitter	A correlation between **user mobility** with the **city population density** and **city average household income** is found.
Salas-Olmedo et al.	2018	Twitter, Foursquare & Panoramio	**Tourist activities** are mapped as an opportunity to **enhance retail investments.**
Zhou & Zhang	2016	Twitter & Foursquare	The **venues are classified** for Boston and Chicago are used for, such as shop & service, nightlife, food & restaurant areas.
Gupta et al.	2018	Twitter	The **higher activity zones** are visualized and **different economic activities** within the city of London are described.
Cranshaw et al.	2012	Foursquare	**Distinctly characterized clusters** called "Livehoods" within the city of Pittsburgh are created using 18 million **check-ins.**
Marti et al.	2017	Foursquare	The **public space usage of people** is illustrated in Alicante.
Agryzkov et al.	2016	Foursquare	A correlation between traditional methods and check-in frequencies is found in **analyzing public spaces.**
Song et al.	2020	Instagram & Flickr	LBSN data are found to reflect the **urban park popularity** in Singapore better than traditional surveys.
Zamani & Schwartz	2017	Twitter	**Twitter language** is found useful while **forecasting residential indicators.**
Hannum et al.	2019	Twitter	**Twitter sentiment** is found to be correlated with **housing prices** and **price appreciation** in Istanbul.
Tan & Guan	2021	Twitter	**Twitter sentiment** is applied to **housing prices** in the U.S. and a positive correlation is found between them.
Lifang et al.	2020	Sina Weibo (Chinese Twitter)	Check-ins are used to **map people's activity** within Wuhan and the **changes in rental housing prices** are investigated.

Framework of the Study

Data Collection & Preparation

Data Sources
- LBSN Data
  - Instagram
  - Twitter
- Real Estate Data

Language & Platform
- Python 3.8
- Jupyter Notebook

Model Specification

- Preliminary Analyses
- Constructing the Methodology

Language & Platform
- R
- R-Studio

Forecasting

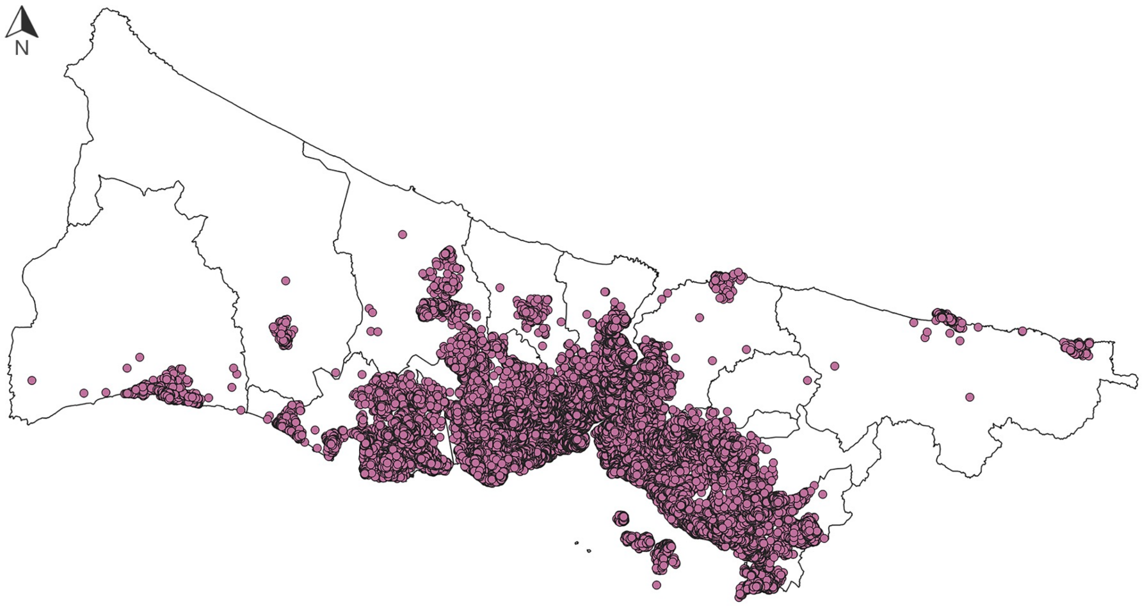
Scope
27 Districts of Istanbul

Language & Platform
- R
- R-Studio

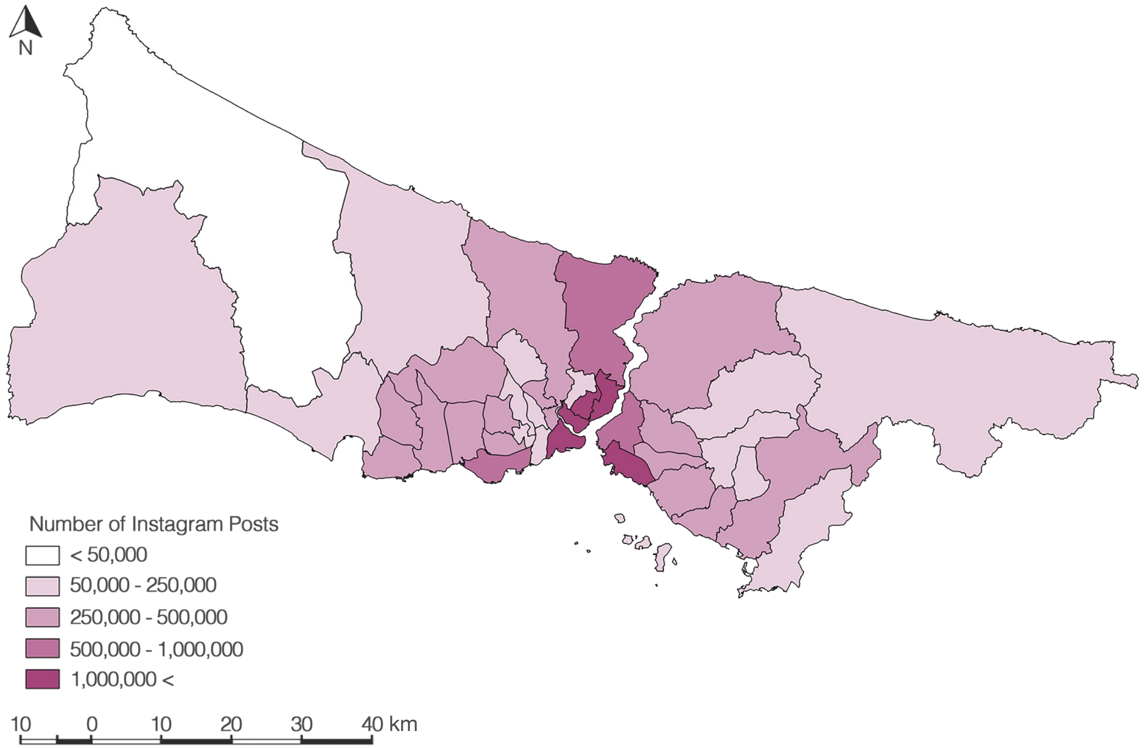
Data Collection & Preparation

INSTAGRAM DATA

- Venue Scraper – **152,145** venues



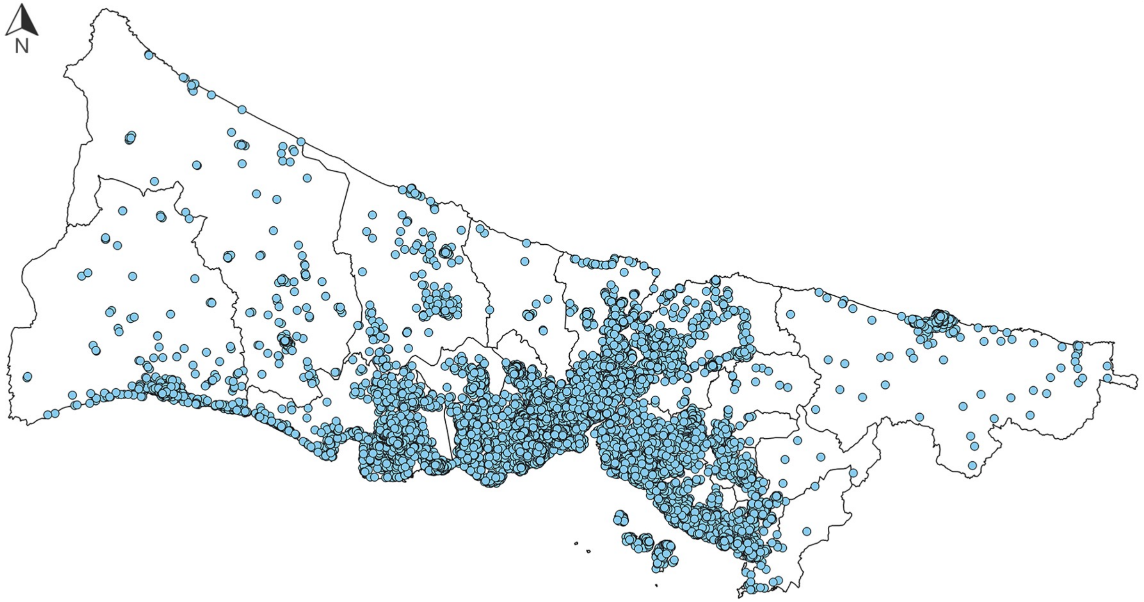
- Post Scraper - **17,161,015** posts



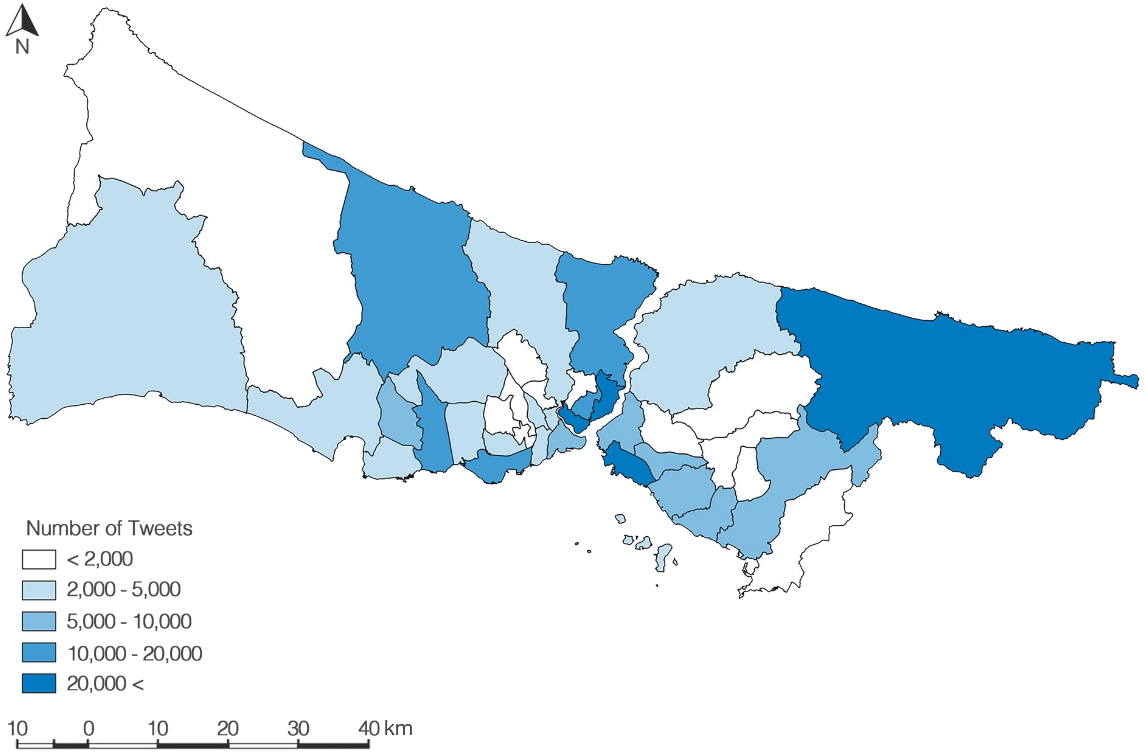
Data Collection & Preparation

TWITTER DATA

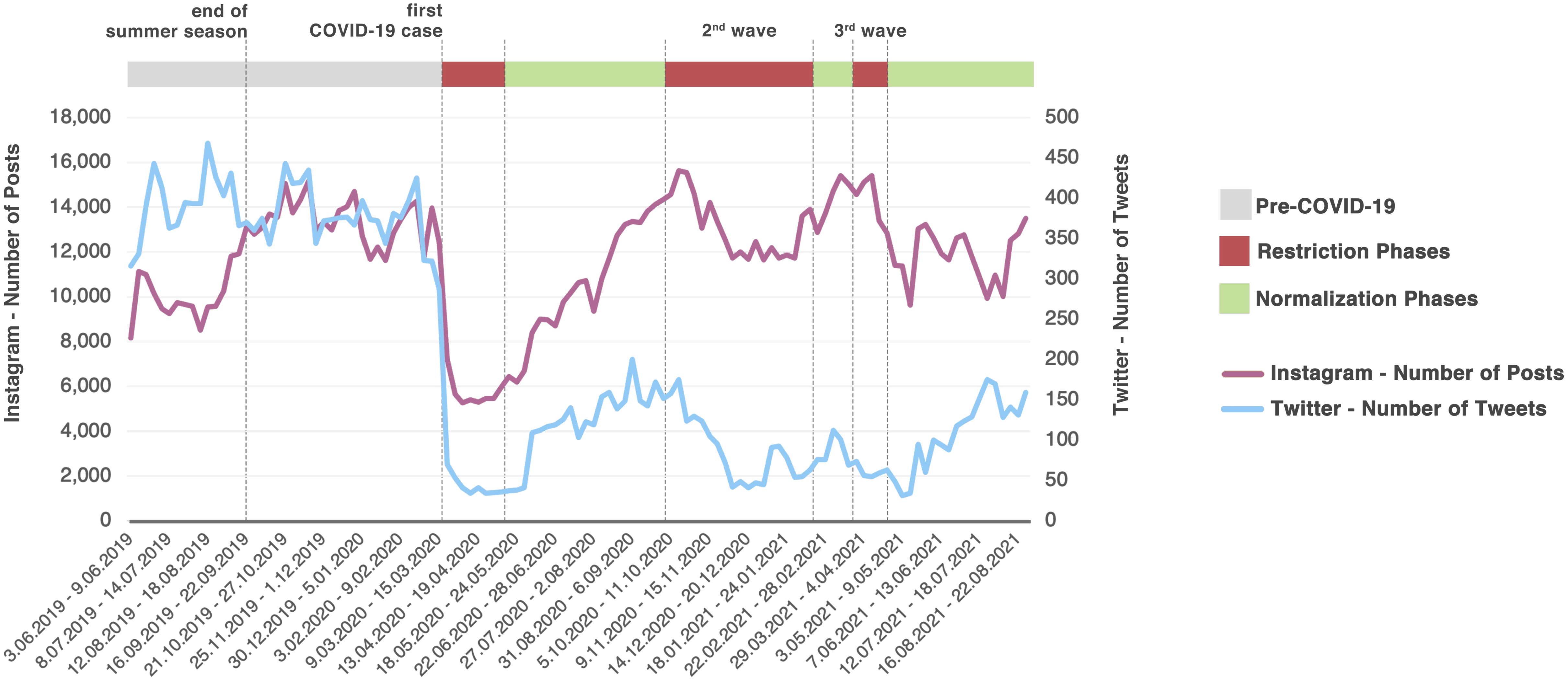
- Venue Scraper – **21,875** venues



- Tweet Scraper – **250,458** tweet



Data Collection & Preparation



Data Collection & Preparation

REAL ESTATE DATA

emlakjet

shop & store listings in Istanbul
(June 2019-August 2021)

38,185 listings for sale

52,235 listings for rent

FINAL DATA SET

week : week range of the time series

c_rental : count of listings for rent per week
av_rent : average rental level per week
c_sales : count of listings for sale per week
av_price : average price level per week
av_days : average days on market

c_post : count of Instagram posts per week
c_tweet : count of Tweets per week

Model Specification

- Time Series Forecasting → ARIMA(p,d,q)
- **Multivariate** Time Series Forecasting → R package ***marima***

*(based on an estimation method for **VARMA models** with **exogenous variables)***

$$\boldsymbol{y_t} = \begin{pmatrix} y_{t,1} \\ y_{t,2} \\ \dots \\ y_{k,t} \end{pmatrix}, \quad \textbf{\textit{and}} \quad \boldsymbol{u_t} = \begin{pmatrix} u_{t,1} \\ u_{t,2} \\ \dots \\ u_{k,t} \end{pmatrix}, \quad \boldsymbol{t = \{1, 2, \dots, N\}} \quad (1)$$

$$\boldsymbol{y_t + \varphi_1 y_{t-1} + \cdots + \varphi_p y_{t-p} = u_t + \theta_1 u_{t-1} + \cdots + \theta_q u_{t-q}} \quad (2)$$

$$\boldsymbol{y_t + \varphi_1 y_{t-1} + \cdots + \varphi_p y_{t-p} + x_t + \beta_1 x_{t-1} + \cdots + \beta_p x_{t-l} = u_t + \theta_1 u_{t-1} + \cdots + \theta_q u_{t-q}} \quad (3)$$

Model Specification

Collect and prepare the data

Apply auto.arima() to determine *d* order differencing for the minimum AIC

Ensure stationary with unit root tests

Determine the potential variables with pairwise Granger causality test

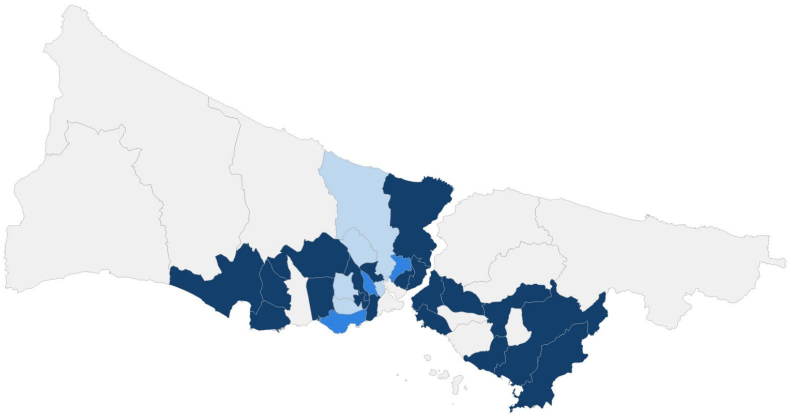
	Model	Model Dimension	Exogenous Variables	
			n_rental, n_sales, av_days	n_insta AND/OR n_tweet
Baseline	A	univariate	-	-
	B	multivariate	-	-
	C1	multivariate	X	-
LBSN-Supported	C2	multivariate	-	X
	C3	multivariate	X	X

- ***p* and *q* pairs from {0, 1, 2, 1-2} → 15 combinations (except for p=0 ∧ q=0)**
- **135 models in total with different *p* & *q* orders and variables**

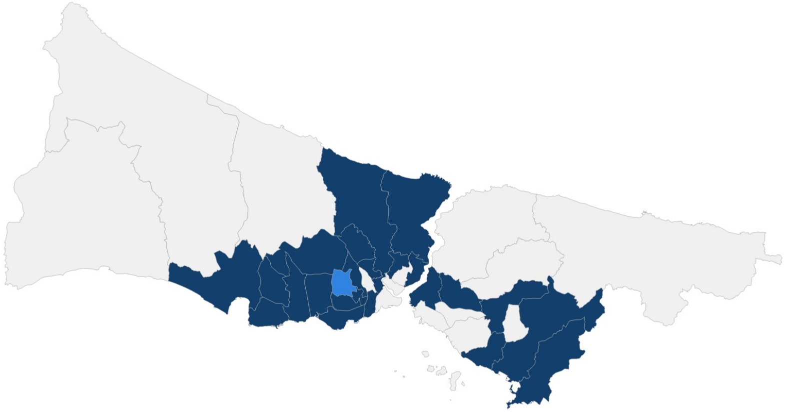
Forecasting Results

District	Sales Price Forecasts: Continuity – Occasional	Sales Price Forecasts: Continuity – Consistent	Sales Price Forecasts: Degree of Improvement – Sligh	Sales Price Forecasts: Degree of Improvement – Moderate	Sales Price Forecasts: Degree of Improvement – Strong	Rental Price Forecasts: Continuity – Occasional	Rental Price Forecasts: Continuity – Consistent	Rental Price Forecasts: Degree of Improvement – Slight	Rental Price Forecasts: Degree of Improvement – Moderate	Rental Price Forecasts: Degree of Improvement – Strong
Avcılar		X			X		X			X
Bağcılar		X	X				X		X	
Bahçelievler	X		X				X			X
Bakırköy		X		X			X			X
Başakşehir		X			X		X			X
Beşiktaş		X			X		X			X
Beylikdüzü		X			X		X			X
Büyükçekmece		X			X		X			X
Esenler		X			X		X			X
Esenyurt		X			X		X			X
Eyüpsultan		X	X				X			X
Gaziosmanpaşa	X				X	X				X
Güngören		X			X		X			X
Kağıthane	X			X			X			X
Kartal		X			X		X			X
Küçükçekmece		X			X		X			X
Pendik		X			X		X			X
Sancaktepe		X			X		X			X
Sarıyer		X			X		X			X
Sultangazi		X	X				X			X
Tuzla	X				X		X			X
Ümraniye		X			X		X			X
Üsküdar		X			X		X			X
Zeytinburnu		X			X		X			X
Bayrampaşa		X		X						
Kadıköy		X			X					
Şişli		X			X					

Sales Price Forecasts



Rental Price Forecasts



Forecasting Results

Beşiktaş District (N=62)

Week(s)	Sales Value Forecasts					
	MAPE			RMSE		
	Baseline	LBSN-Supported	Improvement	Baseline	LBSN-Supported	Improvement
1	0.0091	0.0024	**73.4%**	284.25	75.52	**73.4%**
2	0.0415	0.0183	**56.0%**	1,421.02	661.09	**53.5%**
3	0.1409	0.0691	**50.9%**	4,523.15	2,670.06	**41.0%**
4	0.1171	0.0569	**51.4%**	4,363.77	2,151.81	**50.7%**

Week(s)	Rental Value Forecasts					
	MAPE			RMSE		
	Baseline	LBSN-Supported	Improvement	Baseline	LBSN-Supported	Improvement
1	0.0411	0.0257	**37.4%**	3.91	2.44	**37.4%**
2	0.0412	0.0238	**42.2%**	4.43	3.04	**31.5%**
3	0.0715	0.0707	**1.1%**	10.64	10.01	**5.9%**
4	0.0877	0.0798	**9.0%**	11.51	9.67	**16.0%**

Forecasting Results

Kadıköy District (N=75)

Week(s)	Sales Value Forecasts					
	MAPE			RMSE		
	Baseline	LBSN-Supported	**Improvement**	Baseline	LBSN-Supported	**Improvement**
1	0.0678	0.0354	**47.7%**	991.25	518.06	**47.7%**
2	0.1869	0.0667	**64.3%**	2,689.32	934.22	**65.3%**
3	0.1490	0.1018	**31.7%**	2,296.58	1,702.31	**25.9%**
4	0.1178	0.0875	**25.7%**	1,867.27	1,400.95	**25.0%**

Şişli District (N=65)

Week(s)	Sales Value Forecasts					
	MAPE			RMSE		
	Baseline	LBSN-Supported	**Improvement**	Baseline	LBSN-Supported	**Improvement**
1	0.1520	0.0522	**65.7%**	3,946.48	1,353.91	**65.7%**
2	0.1208	0.0467	**61.4%**	3,613.78	1,674.73	**53.7%**
3	0.1205	0.1053	**12.6%**	3,275.44	3,050.69	**6.9%**
4	0.1344	0.1173	**12.7%**	4,174.00	4,053.79	**2.9%**

Conclusion

- This study discovers the **power of social media data** (Instagram & Twitter) to forecast the future **rent/price levels** of retail properties in Istanbul.

- The methodology is constructed by creating baseline models with the variations of real estate related variables, and **introducing the social media** related variables upon them.

- The multivariate time series analyses are conducted in R package ***marima***, and the **improvements in MAPE and RMSE metrics** are observed.

Conclusion

- The proposed methodology is tested for **27 districts of Istanbul**, and the results demonstrate that the big data arising from LBSN platforms **strongly improve** the retail property forecasts.

- The methodology allows the forecasters to **better predict the trends** in the sector and **adapt to potential changes** in the market in the volatile nature of an emerging market.

- Further research could extend the application of this methodology to **micro levels**, including street level forecasts.

Many thanks for your attention!

Maral TAŞCILAR
tascilar15@itu.edu.tr

Dr. Kerem Yavuz ARSLANLI
arslanli@itu.edu.tr

ERES Annual Conference 2022
SDA Bocconi School of Management, Milan, Italy
24.06.2022

References

Agryzkov, T., Martí, P., Nolasco-Cirugeda, A., Serrano-Estrada, L., Tortosa, L., & Vicent, J. F. (2016). Analysing successful public spaces in an urban street network using data from the social networks Foursquare and Twitter. Applied Network Science, 1(1), 12. https://doi.org/10.1007/s41109-016-0014-z

Carpio-Pinedo, J., & Gutiérrez, J. (2020). Consumption and symbolic capital in the metropolitan space: Integrating 'old' retail data sources with social big data. Cities, 106, 102859. https://doi.org/10.1016/j.cities.2020.102859

Cranshaw, J., Schwartz, R., Hong, J., & Sadeh, N. (2012). The livehoods project: Utilizing social media to understand the dynamics of a city. Proceedings of the International AAAI Conference on Web and Social Media, 6(1).

Gupta, N., Crosby, H., Purser, D., Javis, S., & Guo, W. (2018). Twitter Usage Across Industry: A Spatiotemporal Analysis. 2018 IEEE Fourth International Conference on Big Data Computing Service and Applications (BigDataService), 64-71. https://doi.org/10.1109/BigDataService.2018.00018

Hannum, C., Arslanli, K. Y., & Kalay, A. F. (2019). Spatial analysis of Twitter sentiment and district-level housing prices. Journal of European Real Estate Research, 12(2), 173–189. https://doi.org/10.1108/JERER-08-2018-0036

Lifang, Z., Ting, Y., Yang, L., & Li, Z. (2020). Analyses on the Spatial Distribution Characteristics of Urban Rental Housing Supply and Demand Hotspots Based on Social Media Data. 126–130.

Martí, P., Serrano-Estrada, L., & Nolasco-Cirugeda, A. (2017). Using locative social media and urban cartographies to identify and locate successful urban plazas. Cities, 64, 66–78. https://doi.org/10.1016/j.cities.2017.02.007

Salas-Olmedo, M. H., Moya-Gómez, B., García-Palomares, J. C., & Gutiérrez, J. (2018). Tourists' digital footprint in cities: Comparing Big Data sources. Tourism Management, 66, 13–25. https://doi.org/10.1016/j.tourman.2017.11.001

Song, X. P., Richards, D. R., He, P., & Tan, P. Y. (2020). Does geo-located social media reflect the visit frequency of urban parks? A city-wide analysis using the count and content of photographs. Landscape and Urban Planning, 203, 103908. https://doi.org/10.1016/j.landurbplan.2020.103908

Tan, M. J., & Guan, C. (2021). Are people happier in locations of high property value? Spatial temporal analytics of activity frequency, public sentiment and housing price using twitter data. Applied Geography, 132, 102474. https://doi.org/10.1016/j.apgeog.2021.102474

Zamani, M., & Schwartz, H. A. (2017). Using Twitter Language to Predict the Real Estate Market. Proceedings of the 15th Conference of the European Chapter of the Association for Computational Linguistics: Volume 2, Short Papers, 28–33. https://doi.org/10.18653/v1/E17-2005

Zhou, X., & Zhang, L. (2016). Crowdsourcing functions of the living city from Twitter and Foursquare data. Cartography and Geographic Information Science, 43(5), 393–404. https://doi.org/10.1080/15230406.2015.1128852