Understanding the City Behaviour through Data Analysis: A Case Study of Barcelona

Engin Zeydan
CTTC, Castelldefels, Barcelona, Spain, 08860.
Email: engin.zeydan@cttc.cat

Abstract—Incentivizing the usage of public transportation services among city residents to improve the air quality of the city has been the major target of city officials all around the world. In this paper, we perform an exploratory data analysis on the daily usage and patterns of bicying service usage (Barcelona’s bicycling system and a part of public transport system) and measured air quality values for the city of Barcelona in Spain. Our data analysis results yields three main observations (i) Regarding bicying service, the highest usage of bikes is observed to be during morning and evening rush hours as well as weekdays which shows that bicying service is mostly used by work commuters. (ii) Regarding air quality measurements, the analysis results indicate that Observatori Fabra has the highest air quality whereas the district of Eixample has the poorest air quality among the considered districts of Barcelona and high air pollution are observed during weekdays due to heavy morning rush traffic. (iii) Regarding the relation between bikes service usage and air quality, the correlation between bicying usage and air quality levels is observed to be low.

Index Terms—data analysis, bicying service, air quality, Barcelona, transportation.

I. INTRODUCTION

Intelligent systems powered by data analytics can help to address the contemporary challenges experienced by cities around the world including climate change, transportation problems. To accomplish this, real time data generated by advanced measurement devices such as sensors, actuators and controllers that are available on stations dispersed throughout the city can be used to analyze and describe the ongoing events and situations in real time. This large amount of data can help city officials to make proactive and reactive decisions based on the level of significance of the events. Moreover, intelligent analytics can help to unravel the city dynamics to improve both the service quality and life quality of residences.

Usage of public transportation and measured air quality are interrelated in most of the urban areas around the world [1]. All efforts by public government officers have been on improving the utilization of public transport services so that reduction in private transport services can improve the air pollution levels. In this paper, we analyze the publicly available dataset provided by city of Barcelona to measure the city dynamics in terms of shared bicycle usage as part of public transport system and air quality metrics.

A. Related Work

Data analytics on dataset that are collected city-wide large scale regions can yield important and useful insights into the behaviour of the residents of the city. For example, the paper in [2] has run empirical results on mobile cellular data to find interesting conclusions on tunnels and transport paths selected by residents of city of Istanbul in Turkey. In the mean time, the amount of available dataset for researchers provided by city officials have also recently gained momentum, e.g. New York City [3], London [4], Barcelona [5], etc. City of Barcelona has initiated an Open Data Barcelona project where real time data collected from diverse measurement points throughout the city are collected and published through their website [5]. The data portal contains dataset on different themes, including administration, city and services, economy and business, population and territory. Researchers around the world have used the dataset provided Open Data Barcelona initiative and obtained diverse insights on various domains.

With regard to publicly available shared bicycling system of Barcelona, there are various analysis works instantiated by researchers around the world in different domains. Some these works include improving the operational and usage efficiency, better planning to maintain fair load distribution at each station or inferring daily human behaviour, pattern or cultural behaviour of the city residents [6], [7]. For example, the authors in [6] have provided a spatio-temporal analysis of bicying during a period of six weeks where the analysis results have extracted different daily routines and patterns of Barcelona’s citizens. The paper in [8] analyzes the bike sharing data of Barcelona and Sevilla cities to improve the bicycle infrastructure operation.

In terms of air quality measurements, many works have started to emerge to investigate the effects of air quality measurements on daily lives of urban settlers due to effect of global climate change [1], [9]–[11]. Air quality can also impact several industries such as cruise tourism industry as demonstrated in [9] for the Port of Barcelona, subway systems in [10] or in education [11]. Interrelation between public transportation usage and air quality of the city has been under study in [1] where the effect of absence
of public transport systems on air quality is measured in Barcelona. The authors in [12] investigate associations between cycling paths and exposure to air pollution in Glasgow UK using crowd-sourced dataset. However, finding the interrelation between bicycle usage of city residents and air quality levels observed around the city is still an open research area.

B. Main Observations

Different than above works, in this paper we perform Exploratory Data Analysis (EDA) on bicycle service usage and air quality metrics of the city of Barcelona and seek to find the relationship between them. Through empirical analysis, we explore some of the key findings in terms of bicycle usage and air quality information and deduct conclusions that can be helpful for city planners and decision makers. The major observations of the paper are as follows: (i) Regarding bicycling service usage, the highest usage of bikes is observed to be around 6pm and 8 am (relatively) which shows that bicycling is mostly used by commuters for work. High usage of bikes are observed during weekdays compared to weekends and more number of bikes are observed in morning and evening rush hours which is again due to commuters for work. (ii) Regarding air quality measurements, the analysis results indicate that Observatori Fabra has the highest air quality with regards to $O_3$ (tropospheric Ozone), $NO_2$ (Nitrogen dioxide), $PM_{10}$ (measures solid and suspended particles and liquid droplets (such as dust, dirt, soot, and smoke) found in the air) values whereas the district of Eixample has the poorest air quality among all the considered districts of Barcelona where observations are done. During weekdays in morning between 7am and 11am, the $O_3$ values are observed to be the lowest and corresponding $NO_2$ are observed to be the highest in the city which can be due to morning rush traffic. (iii) Finally, low correlations are evaluated between the bicycle usage and observed air quality metrics during the observation duration.

The rest of the paper is organized as follows. In Section II, the utilized dataset collected from city of Barcelona’s open data initiative is described. In Section III, the analysis results based on EDA are presented using the previously collected dataset. In Section IV, discussions on the main findings the results are revealed. Finally in Section V, conclusions and future work of the paper are presented.

II. Analysis Steps and Utilized Dataset

For the analysis in the rest of the paper, Open Data Barcelona’s publicly available bicying service and air quality value dataset are used [5]. The analysis involved four main steps. These steps include collection, processing, analysis and visualization. During data collection, the dataset for bicying service usage of the city of Barcelona is obtained from [13] and data for air quality of the city of Barcelona is acquired from [14]. During processing step, the missing data in the analyzed air quality dataset is pre-processed by using both “forward and backward fill” methods to propagate last valid observation forward and use next valid observation to fill the backward gap. For analysis, we used Python’s Pandas data analytics library and for visualization of the results on map, we used Folium’s leaflet library. The air quality data includes measured data of $O_3$, $NO_2$, $PM_{10}$. The bicying service use dataset observation duration is from 01 – 08 – 2018 to 24 – 03 – 2019 whereas air quality data observation duration is from 01–08–2018 to 31–12–2018. Air quality dataset contains 35,983 rows of measurements with 18 different features. Bicying dataset contains 47,270 rows of measurements with 5 different features.

III. Analysis Results

In this section, EDA on the above described dataset is run to explore Barcelona city dynamics in terms of bicying usage and air quality levels based on space and time.

A. Bicing Usage Analysis

In all the figures below, bikes in usage represents the total number of bikes in usage including both mechanical and electrical bikes. Fig. 1 shows the mean number of bikes usage per day over the observation duration. It can be observed from this figure that the highest usage has occurred on 26 September 2018. Note that the amount of bikes in usage has decreased after January 2019 and diminishes to only one bicycle on 23 March 2019. This is due to gradual switch to new bicying service on different stations starting on 1 January 2019 [15].

Fig. 2 shows the median values of bikes in usage distributed per hour. The highest usage of bikes is observed to be around 6pm. We can also observe relatively high bikes usage in the morning hours around 8am. The reason behind these observations is due to working hours period where bicying service users are using bikes either to go to work or come back to their home after official work hours. Fig. 3 shows the number of bikes in usage distributed per day of the week. Fig. 3 clearly indicates high usage of bikes during weekdays compared to weekends. Usage of bikes on Sunday can be observed to be the lowest. Fig.
Fig. 2. Median values of bikes in usage distributed per hour.

Fig. 3. Number of bikes in usage based on days of the week.

Fig. 4. Heatmap of median bikes usage per hour and days of the week.

Fig. 5. Median values of $O_3$, $NO_2$ and $PM_{10}$ by hours of the day.

Fig. 6. Median values $O_3$, $NO_2$ and $PM_{10}$ distribution based on the days of the week.

Fig. 7 shows the heatmap of median bikes usage per hour and days of the week. Similar to observations in Fig. 2, the highest number of bikes usage has occurred during morning and evening rush hours of weekdays. This again validates our analysis on Fig. 2 that bicying service is mostly used by commuters for work.

**B. Air Quality Analysis**

There are 8 main observed districts of Barcelona in the air quality dataset which are namely Gracia, Vall Hebron, Ciutadella, Eixample, Sants, Observ Fabra, Poblenou and Palau Reial. Table I shows the minimum and maximum measured levels corresponding to each quality levels and metrics.

|            | $O_3$ | $NO_2$ | $PM_{10}$ |
|------------|-------|--------|-----------|
| Good       | 1-111 | 1-91   | 1-56      |
| Regular    | 112-180 | 92-197 | 57-50     |
| Poor       | 181-1000 | NA     | 51-113    |

Fig. 5 shows the median values of $O_3$, $NO_2$ and $PM_{10}$ distributed over hours of the day during the observation duration. It can be observed that the highest levels of $O_3$ has been observed around 7 pm in the evening. On the other hand, $PM_{10}$ levels are almost constant during all hours of the day. $NO_2$ values are observed to be the highest at 10 am in the morning. Fig. 6 shows the median values $O_3$, $NO_2$ and $PM_{10}$ distribution based on the days of the week. $O_3$ values are observed to be the highest on Wednesday whereas $NO_2$ levels are the highest on Friday. On the other hand, Monday and Sunday produce the lowest $PM_{10}$ values in the city. Fig. 7 shows the median values $O_3$, $NO_2$ and $PM_{10}$ distributed over the months of the year. The highest $O_3$ values are observed on August. Regarding $NO_2$ values, December is observed to be the month that has given the highest value. On the other hand, $PM_{10}$ values are observed to be the lowest on months of November and December.

Fig. 8 shows the percentages of median values $O_3$, $NO_2$ and $PM_{10}$ distribution over the districts of Barcelona. The highest $NO_2$ levels are observed in Eixample district,
whereas the highest $O_3$ values are observed in districts of Sants and Observ Fabra. Regarding $PM_{10}$ values, the district of Observ Fabra has given the lowest values.

Figs. 9, 10 and 11 show the heatmap of median values per hour and days of the week for $O_3$, $NO_2$ and $PM_{10}$ respectively over the observation duration. Figs. 9 shows that interestingly the highest values of $O_3$ and the lowest values of $NO_2$ have been observed on Saturday at 7pm. On the other hand during weekdays in morning from 7am to 11am, the observed $O_3$ values in the city are the lowest. Fig. 11 shows that the highest values of $PM_{10}$ is from midnight to early morning hours on Saturday.

Fig. 12 shows the distribution of total number of reported poor air quality locations whereas the areas around Ciutadella and Sants districts have the lowest number of observed poor air quality report locations. Note that this does not necessarily signify the quantified values of air quality levels (as shown in Fig. 8) but only the number of measured poor quality locations.

Final analysis result indicates that low correlations are observed between bicying usage and air quality values over the observation duration from August 2018 to December 2018. The Pearson correlations between bicying usage and $O_3$, $NO_2$ and $PM_{10}$ are calculated to be $-0.16$, $0.25$ and $0.26$ respectively.

**IV. Discussions on Analysis Results**

Regarding bicying service usage as expected the highest usage of bikes is observed to be around 6pm and 8am (relatively) which is mainly due to usage of bicying service
by work commuters (which is mainly used either to go to work or come back to home after official work hours). High usage of bikes are also observed during weekdays compared to weekends where Sunday is observed to be day that has the lowest number of bikes usage. This signifies that weekday usage of bicying service is more common than weekend usage. Additionally, higher number of bikes were observed in morning and evening rush hours. This again validates our assumption that bicying service is mostly used by commuters for work and not necessarily for leisure activities (shopping, touristic visits, etc) as mostly done during weekends.

Regarding air quality measurements, $O_3$ and $NO_2$ values are observed to be the lowest values during weekdays in the morning from 7am to 11am which can be due to morning rush traffic. High values of $PM_{10}$ on Saturday from midnight to early morning can indicate possibility of ongoing constructions works (e.g. roads, buildings, etc) throughout the city planned until early morning hours. Unexpectedly observed high value of $O_3$ value and low value of $NO_2$ on Saturday at 7pm can be due to several reasons including higher utilization of pedestrian paths by the city residents during Saturday evenings. Additionally, low numbers of poor air quality observation points do not necessarily imply high values of air qualities as observed for the district of Ciutadella in Fig. 8 and Fig. 12. The highest $O_3$ values are observed to be on month August where most of the local residents in Barcelona are generally on vacation. The district of Observ Fabra is observed to have the highest air quality with regards to $O_1$, $NO_2$ and $PM_{10}$ values whereas the district of Eixample has the poorest air quality among the considered districts of Barcelona.

The low correlations between air quality and bicying usage indicates that bicying usage does not necessarily have high impact on reduction of air quality metrics in the city of Barcelona. This can be due to various reasons. One of the reasons can be high usage of bicying system does not necessarily yield air quality reductions. Air quality can depend on many factors (e.g. other private and public transportation services) and even high usage of bicying service can yield small reductions air quality levels that can not be substantially observable.

V. Conclusions and Future Work

In this paper, air quality measurements and bicying usage in the city of Barcelona, Spain are analyzed using EDA technique. The analysis resulted in three main observations. First one is regarding bicying system usage where high bicying service is utilized during weekdays and morning and evening rush hours by work commuters. Second one is regarding air quality measurement where the district of Observatori Fabra is observed to have the highest and the district of Eixample is observed to have the poorest air qualities. Moreover, during weekdays the $O_3$ values are observed to be the lowest and corresponding $NO_2$ are observed to be the highest in the city of Barcelona due to heavy morning rush traffic. Finally, the correlations between bicycle usage and air quality measurements are observed to be low. As a future work, analysis results can be extended to accommodate larger dataset spanning more than one year of measurements. Moreover, new recommendations to incentivize the public transport usage can be created by exploiting additional dataset provided by Open Data Barcelona Initiative project.

Acknowledgment

This work was partially funded by Spanish MINECO grant TEC2017-88373-R (5G-REFINE) and by Generalitat de Catalunya grant 2017 SGR 1195.

References

[1] X. Basagaña, M. Triguero-Mas, D. Agis, N. Pérez, C. Reche, A. Alastuey, and X. Querol, “Effect of public transport strikes on air pollution levels in barcelona (spain),” Science of the total environment, vol. 610, pp. 1076–1082, 2018.

[2] F. Kalyoncu, E. Zeydan, A. Yildirim, and I. O. Yigit, “An experimental study of factor analysis over cellular network data,” in 2018 IEEE 87th Vehicular Technology Conference (VTC Spring), pp. 1–5, IEEE, 2018.

[3] New York State, “New York State Open Data Initiative.” https://data.ny.gov/, 2019. [Online; accessed 28-Sept.-2019].

[4] Greater London Authority (GLA), “London Datastore.” https://data.london.gov.uk/dataset, 2019. [Online; accessed 30-Sept.-2019].

[5] Ajuntament de Barcelona, “Open Data or Public Sector Information Openness Project.” https://opendata-ajuntament.barcelona.cat/en/open-data-bcn, 2019. [Online; accessed 28-Sept.-2019].

[6] J. Froehlich, J. Neumann, and N. Oliver, “Measuring the pulse of the city through shared bicycle programs,” Proc. of UrbanSense08, pp. 16–20, 2008.

[7] C. Li, Y. Liu, and C. Li, “Analysis and research on the use situation of public bicycles based on spark machine learning,” in Proceedings of the 2019 International Conference on Data Mining and Machine Learning, pp. 160–165, ACM, 2019.

[8] A. Faghils-Imani, R. Hampshire, L. María, and N. Eluru, “An empirical analysis of bike sharing usage and rebalancing: Evidence from barcelona and seville,” Transportation Research Part A: Policy and Practice, vol. 97, pp. 177–191, 2017.

[9] I. Ruiz-Guerra, V. Molina-Moreno, F. J. Cortés-García, and P. Núñez-Cacho, “Prediction of the impact on air quality of the cities receiving cruise tourism: the case of the port of barcelona,” Helgolyon, vol. 5, no. 3, p. e01280, 2019.

[10] T. Moreno, V. Martíns, C. Reche, M. C. Minguillón, E. de Miguel, and X. Querol, “Air quality in subway systems,” in Non-Exhaust Emissions, pp. 293–321, Elsevier, 2018.

[11] N. Castell, P. Schneider, S. Grossberndt, M. F. Fredriksen, G. Sousa-Santos, M. Vogt, and A. Bartonova, “Localized real-time information on outdoor air quality at kindergartens in oslo, norway using low-cost sensor nodes,” Environmental research, vol. 165, pp. 410–419, 2018.

[12] Y. Sun and A. Mobasher, “Utilizing crowdsourced data for studies of cycling and air pollution exposure: A case study using strava data,” International journal of environmental research and public health, vol. 14, no. 3, p. 274, 2017.

[13] Open Data BCN, “Bicing service use of the city of Barcelona.” https://opendata-ajuntament.barcelona.cat/data/en/dataset/us-del-servi-bicing, 2019. [Online; accessed 28-Sept.-2019].

[14] Open Data BCN, “Air quality of the city of Barcelona.” https://opendata-ajuntament.barcelona.cat/data/en/dataset/qualitat-aire-detall-bcn, 2019. [Online; accessed 28-Sept.-2019].

[15] Ajuntament de Barcelona, “The new Bicing is here.” http://ajuntament.barcelona.cat/bicicleta/en/noticia/the-new-bicing-is-here, 2019. [Online; accessed 30-Sept.-2019].