Aggregate Human Mobility Using Mobile Network Big Data

Okkie Putriani[1,2], Sigit Priyanto[1](✉), Imam Muthohar[1], and Mukhammad Rizka Fahmi Amrozi[1]

[1] Universitas Gadjah Mada, Yogyakarta, Indonesia
spriyanto2007@ugm.ac.id
[2] Universitas Atma Jaya Yogyakarta, Yogyakarta, Indonesia

Abstract. Processing big data into human mobility analysis requires more in-depth stages and calculations with each consideration or assumption at the beginning of the step. Data from the Ministry of Transportation Research and Research Section with City Data for March 2022 on 15–16, Sunday and Monday as a representative of weekends and weekdays. The data received is translated into a UTM 48 S coordinate map because it is located in DKI Jakarta. The method used is processing the Python and Kepler programming languages as data visualization. The data cleansing required at the beginning removes data that is not moving and eliminates data that moves too far at speeds above 60 km/h. Data aggregation is indicated by spatial join with the zoning area. It eliminates the need to do an internal movement. Furthermore, set a dwelling time of 30 min and a travel time of 50 km/h. The movement taken is the movement of more than one activity. In one zone, the area can be represented from one sub-district area or a particular zoning area, then translated into one midpoint simultaneously. It becomes interesting when the analysis of human mobility is seen from the internal movement so that the internal movement is eliminated. By looking at the data aggregation, it can be calculated for preparing the Origin-Destination Matrix table. The condition of DKI Jakarta with calculations using the MNBD is undoubtedly a new finding because this is the initial condition of the Covid-19 pandemic coming to Indonesia. So that human mobility can be seen more clearly due to less movement due to travel restrictions. After experiencing the data aggregation process, approximately only 2.17% of the data can be used. It can be seen that the five highest zones are Gambir, Karawaci, South Bekasi, Sunter Agung, and Pondok Aren.

Keywords: Aggregate Data · Human Mobility · Big Data

1 Introduction

The essence of human movement can be seen from the origin and purpose of their activities [1]. Billions of people move daily, and these activity trips can be aggregated into movement data that can be analyzed further [2]. The existence of technological sophistication makes cell phones capable of recording human mobility data [3–5], and the behaviour pattern of human mobility can be used as a transportation policy [6, 7].

W. A. Wirawan et al. (Eds.): ICORT 2022, AER 220, pp. 58–66, 2023.
https://doi.org/10.2991/978-94-6463-126-5_7

1.1 Human Mobility

Human mobility is a commonly used but loosely defined term that represents the notion of a person's spatio-temporal occupancy, involving the interactions between people, society, and the surrounding physical environment [4, 8]. A better understanding of human mobility is key to understanding human interactions with their environment and the use of geospatial, which aids in transportation and urban planning, political decision-making, epidemiology, economic development, emergency management, etc. [9]. Human activities generate large amounts of geospatial data. Processing big data into human mobility analysis requires more in-depth stages and calculations with each consideration or assumption at the beginning of the step [10].

1.2 Mobile Network Big Data

Given the volumes of data that are now generated by mobile networks due to the almost ubiquitous use of mobile phones [11] by the majority of the population, this data can be considered as big data. Spurred by the exponential growth of mobile connectivity [12], the attendant large volumes of Mobile Network Big Data (MNBD), offer the possibility to obtain rich behavioral insights [13] at a scale that was never possible before. The movement data model can be obtained from MNBD data collected from aggregations. One of the MNBD data provider companies comes from the United States, namely citydata.ai, which utilizes a geospatial platform related to daily movement data from the areas we need.

1.3 Previous Condition

For Indonesian conditions, they began utilizing cellular data through the country's leading telecoms carrier in 2017–2018. Processing this prefix data presents difficulties because it does not reflect actual situations. One Indonesian citizen may use multiple service providers, and a single territory is split up under the dominance of various providers, as well as mapping tasks that continue to employ traditional techniques, particularly by carrying out manual surveys [14].

2 Material and Methods

The information was gathered by the Republic of Indonesia's Ministry of Transportation from citydata.ai. With the initial entry strategy for the Covid-19 pandemic, sample data for 2020 were collected in Indonesia in March. The sample consists of the weekends and weekdays of March 15 and 16, 2020, respectively. The boundaries of the given area are Jakarta and its surroundings. The Python programming language combines this daily data into one massive data set, which is then processed to enable analysis and visualization.

3 Results and Discussion

3.1 Collect the Data

Information from the Ministry of Transportation Research and Research Section with City Data for March 2022 on 15–16, Sunday and Monday, as an example of weekends and weekdays. Due to its location at DKI Jakarta, the data is converted into a map in UTM 48 S coordinates. Figure 1 depicts the UTM 48 S coordinate.

3.2 Data Cleansing

Data cleansing is essential in preparing data during operations or downstream analysis. When many data sources are integrated, there are several opportunities for data to be duplicated or incorrectly categorized. If the data is incorrect, the results and algorithms are unreliable, even if they appear to be correct. [6]. Given that these procedures differ from data set to data set, there is no definite way to specify the distinct phases in the data cleaning process. The initial data cleansing process removes any data that is not moving and eliminates any data that moves too quickly at speeds more than 60 km/h. The data cleansing can be seen in Table 1.

3.3 Data Aggregation

Data aggregation [16] is the process of gathering data and summarizing it for uses like statistical analysis. Aggregation is used to get more information about a particular team based on a particular variable. Spatial connection with the zoning area is a visual indicator of data aggregation. It does away with the requirement for an internal movement. Set a dwell period of 30 min and a travel speed of 50 km/h. The movement made is the result of several different activities. One zone can be represented by one subdistrict area or one specific zoning area, simultaneously translated into one middle. When internal movement is eliminated, and the analysis of human motion is done from the outside, it

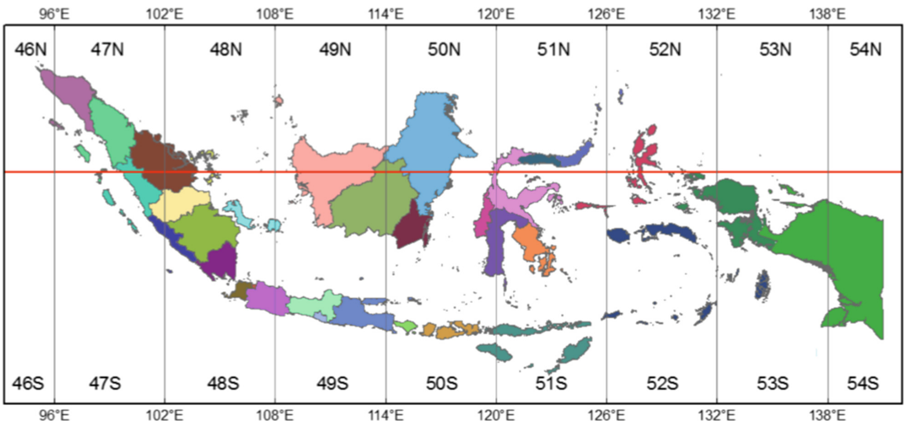


Fig. 1. UTM 48 S Coordinate Map: Indonesia 102°E–108°E [15]

Table 1. Example Data < 60 km/h

DID	Timestamp	App Name	Kel/Kec/Ka	No	X	Y	Dist_M	Delta_T	Time	Velocity_km/h
200OHV1….9qcG5j	2020–03-15 03:56:10	None	SUKMAJAYA	0	704320.533456	9.290947E + 06	NaN	NaT	NaN	NaN
200OHV1….9qcG5j	2020–03-15 03:56:22	None	SUKMAJAYA	1	704314.350359	9.290951E + 06	7.135486	0 days 00:00:12	12.0	2.140646
200OHV1….9qcG5j	2020–03-15 03:30:50	None	SUKMAJAYA	2	704311.485666	9.290954E + 06	4.391413	0 days 01:34:28	5668.0	0.002789
200OHV1….9qcG5j	2020–03-15 03:31:10	None	SUKMAJAYA	3	704317.028426	9.290957E + 06	6.286575	0 days 00:00:020	20.0	1.131584
200OHV1….9qcG5j	2020–03-15 03:31:18	None	SUKMAJAYA	4	704315.82143	9.290960E + 06	3.021022	0 days 00:00:08	8.0	1.359460

Table 2. The Origin - Destination Table

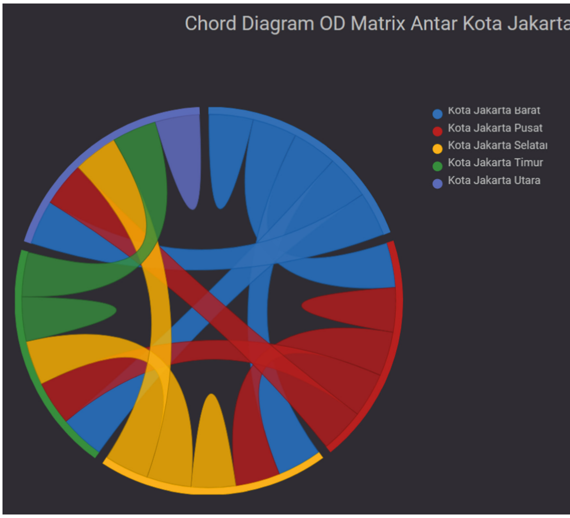
DID	T_start	Zone_start	Ion_start	Lat_start	T_end	Zone_end	Ion_end	Lat_end
M201c2.....yZ3Vr	2020-03-15 03:44:19	KEBON KACANG	106.817037	−6.190325	2020-03-15 04:25:47	BENDA	106.665930	−6.128539
M201c2.....yZ3Vr	2020-03-15 04:25:47	BENDA	106.665930	−6.128539	2020-03-15 16:32:55	SERPONG	106.680793	−6.311255
M201c2.....yZ3Vr	2020-03-15 16:32:55	SERPONG	106.680793	−6.311255	2020-03-15 17:29:37	GUNUNG SINDUR	106.686688	−6.387298
M201c2.....yZ3Vr	2020-03-15 17:29:37	GUNUNG SINDUR	106.686688	−6.387298	2020-03-15 23:44:05	SERPONG	106.680793	−6.311255
M201c2.....yZ3Vr	2020-03-15 00:03:48	PONDOKGEDE	106.930083	-6.268901	2020-03-15 00:47:24	BEKASI TIMUR	107.019131	−6.242045

15 Mar 2020 - 16 Mar 2020

end_kecamatan / freq_od_kecamatan

start_kecama...	Gambir	Cakung	Cengkar...	Duren S...	Tanjung...	Jagakar...	Grogol ...	Tebet	Penjarin...	Pasar M...	Kebayor...	Kebon J...	Kramat ...	Tanah A...	Jatineg...
Gambir	19,359,859	7,639	9,777	9,263	6,436	7,234	8,798	18,923	7,992	8,198	6,416	7,680	5,677	6,471	5,888
Cakung	7,817	677,395	171	2,018	206	125	93	162	46	130	43	110	90	46	372
Cengkareng	9,913	171	666,413	139	62	205	840	151	1,998	172	273	1,128	27	144	137
Duren Sawit	9,404	2,040	135	587,663	298	208	585	215	178	457	76	175	230	230	1,231
Tanjung Priok	6,475	202	68	301	584,405	191	120	86	168	197	37	119	261	239	166
Jagakarsa	7,446	118	205	203	195	568,190	267	985	89	2,565	610	180	373	218	197
Grogol Petam...	9,020	85	821	582	136	267	522,828	163	936	63	273	1,830	65	377	129
Penjaringan	8,113	42	1,959	173	157	86	899	82	525,777	115	93	533	51	106	38
Tebet	12,967	166	147	201	91	978	154	511,313	84	509	357	94	190	305	2,412
Pasar Minggu	8,355	134	188	458	193	2,405	63	540	119	516,056	256	113	894	62	175
Kebayoran La...	6,595	43	286	84	28	590	287	336	95	273	496,446	1,468	28	1,066	145
Kebon Jeruk	7,850	106	1,145	156	128	181	1,888	117	547	106	1,498	471,581	24	262	164
Kramat Jati	5,863	89	26	236	263	383	73	200	56	921	29	15	439,310	32	1,351
Jatinegara	6,099	340	134	1,206	179	188	119	2,436	36	203	135	161	1,313	56	432,756

Fig. 2. Jakarta Origin Destionation Matrix 15–16 March 2020 Using MNBD



	origin	destination	frequention
1.	Kota Jakarta Barat	Kota Jakarta Barat	3,416,595
2.	Kota Jakarta Barat	Kota Jakarta Pusat	61,427
3.	Kota Jakarta Barat	Kota Jakarta Selatan	11,596
4.	Kota Jakarta Barat	Kota Jakarta Timur	4,675
5.	Kota Jakarta Barat	Kota Jakarta Utara	8,764
6.	Kota Jakarta Pusat	Kota Jakarta Barat	60,398
7.	Kota Jakarta Pusat	Kota Jakarta Pusat	21,257,248
8.	Kota Jakarta Pusat	Kota Jakarta Selatan	83,439
9.	Kota Jakarta Pusat	Kota Jakarta Timur	60,231
10.	Kota Jakarta Pusat	Kota Jakarta Utara	38,296
11.	Kota Jakarta Selatan	Kota Jakarta Barat	11,641
12.	Kota Jakarta Selatan	Kota Jakarta Pusat	78,071
13.	Kota Jakarta Selatan	Kota Jakarta Selatan	4,338,098
14.	Kota Jakarta Selatan	Kota Jakarta Timur	14,592
15.	Kota Jakarta Selatan	Kota Jakarta Utara	3,553
16.	Kota Jakarta Timur	Kota Jakarta Barat	4,725
17.	Kota Jakarta Timur	Kota Jakarta Pusat	61,826
18.	Kota Jakarta Timur	Kota Jakarta Selatan	15,103
19.	Kota Jakarta Timur	Kota Jakarta Timur	4,140,310
20.	Kota Jakarta Timur	Kota Jakarta Utara	8,706

Fig. 3. The Chord Diagram of Origin-Destination Matrix Intercity in Jakarta Administration: 15–16 March 2020

becomes interesting. It can be determined to create the Origin-Destination Matrix table by looking at the data aggregation. The table can be seen in Table 2.

A table of origin and destination is then used to map the aggregate of human mobility, as seen in Fig. 2. The computation shows that internal motions continue to remain dominant. Internal movement is not considered when calculating trip displacement in the transportation sector.

The data is then aggregated back into Jakarta regional units in Fig. 3 to make it simpler to see the results of the OD Matrix calculations. Afterwards, they were grouped to create a Chord Diagram to determine which movements represent excellent human mobility.

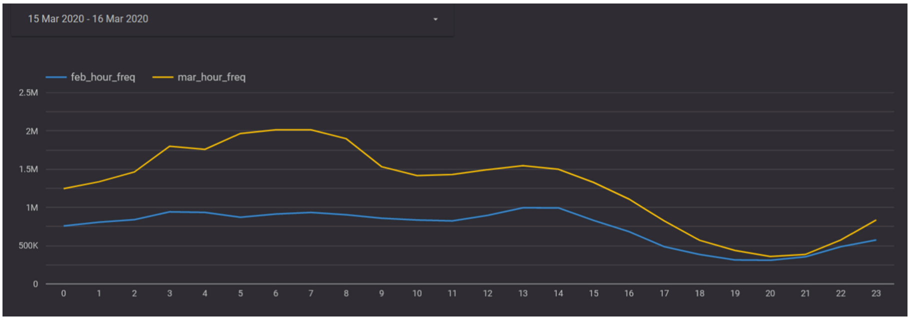


Fig. 4. The Frequency of Jakarta Human Mobility on March 15–16, 2020

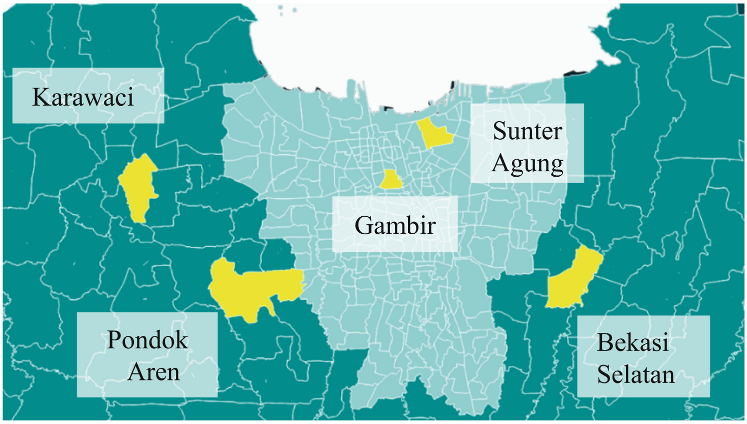


Fig. 5. Five Highest Zones of Human Mobility in Jakarta on March 15–16, 2020

The movement has decreased compared to the general conditions in Jakarta, as seen in Fig. 4, which displays the frequency of data throughout 24 h on two days in March 2022. This condition is a result of work-from-home policies and activity limitations. In order to effectively represent actions with colour. The pandemic is officially moving toward Indonesia, as indicated by the orange tint. A decreased activity is apparent since the blue denotes a restricted travel activity limitation scheme. This is the initial state of the Covid-19 epidemic reaching Indonesia, the situation of DKI Jakarta, as determined by calculations utilizing the MNBD, is unquestionably a novel discovery. In order to better understand human mobility by reducing movement owing to travel constraints.

It can be seen that the five highest human mobility zones during March 15–16, 2020 are Gambir, Karawaci, South Bekasi, Sunter Agung, and Pondok Aren. The five highest mobility zones can be seen in Fig. 5.

4 Conclusion

Using Mobile Network Big Data data samples collected on March 15 and 16, 2020, the movement of people in Jakarta, Indonesia, demonstrated the effects of the Covid-19 epidemic, including a decline in activity and movement. Gambir, which receives 95% of all origin and destination visits, is the most popular location. A definition of movement based on the method of transportation used or integration with public transportation is anticipated from the future follow-up study.

References

1. Smolak, K., Si\la-Nowicka, K., Delvenne, J.-C., Wierzbiński, M., Rohm, W.: The impact of human mobility data scales and processing on movement predictability. Scientific Reports. 11, 1–10 (2021).
2. Karamshuk, D., Boldrini, C., Conti, M., Passarella, A.: Human mobility models for opportunistic networks. IEEE Communications Magazine. 49, 157–165 (2011).
3. Gonzalez, M.C., Hidalgo, C.A., Barabasi, A.-L.: Understanding individual human mobility patterns. nature. 453, 779–782 (2008).
4. Becker, R., Cáceres, R., Hanson, K., Isaacman, S., Loh, J.M., Martonosi, M., Rowland, J., Urbanek, S., Varshavsky, A., Volinsky, C.: Human mobility characterization from cellular network data. Communications of the ACM. 56, 74–82 (2013).
5. Zhao, C., Zeng, A., Yeung, C.H.: Characteristics of human mobility patterns revealed by high-frequency cell-phone position data. EPJ Data Science. 10, 5 (2021).
6. Jiang, S., Guan, W., Zhang, W., Chen, X., Yang, L.: Human mobility in space from three modes of public transportation. Physica A: Statistical Mechanics and its Applications. 483, 227–238 (2017).
7. Huang, Z., Ling, X., Wang, P., Zhang, F., Mao, Y., Lin, T., Wang, F.-Y.: Modeling real-time human mobility based on mobile phone and transportation data fusion. Transportation research part C: emerging technologies. 96, 251–269 (2018).
8. Barbosa, H., Barthelemy, M., Ghoshal, G., James, C.R., Lenormand, M., Louail, T., Menezes, R., Ramasco, J.J., Simini, F., Tomasini, M.: Human mobility: Models and applications. Physics Reports. 734, 1–74 (2018). https://doi.org/10.1016/j.physrep.2018.01.001.
9. Human Mobility Analytics in Big Data Era – Geoinformation and Big Data Research Laboratory @ USC, http://gis.cas.sc.edu/gibd/human-mobility-analytics-in-big-data-era/, last accessed 2022/09/28.
10. Lokanathan, S., Lucas Gunaratne, R.: Mobile Network Big Data for Development: Demystifying the Uses and Challenges. (2015).
11. Andersson, A., Engelson, L., Börjesson, M., Daly, A., Kristoffersson, I.: Long-distance mode choice model estimation using mobile phone network data. Journal of Choice Modelling. 42, 100337 (2022). https://doi.org/10.1016/j.jocm.2021.100337.
12. Smith, A.: The best (and worst) of mobile connectivity. Pew Internet & American Life Project Washington, DC (2012).
13. Wang, Z., He, S.Y., Leung, Y.: Applying mobile phone data to travel behaviour research: A literature review. Travel Behaviour and Society. 11, 141–155 (2018).
14. Putriani, O., Priyanto, S.: Optimization Big Data Real-time Analytics Using Mobile Phone Data in Origin Destination National Transportation (ATTN) Survey. Presented at the 11th Asia Pacific Transportation and the Environment Conference (APTE 2018) October (2019). https://doi.org/10.2991/apte-18.2019.39.
15. Zona UTM Indonesia (Peta Zone Universal Transverse Mercator), https://www.asifah.com/zona-utm-indonesia/, last accessed 2022/10/01.
16. Shankar, V.G., Devi, B., Srivastava, S.: DataSpeak: Data Extraction, Aggregation, and Classification Using Big Data Novel Algorithm. In: Iyer, B., Nalbalwar, S.L., and Pathak, N.P. (eds.) Computing, Communication and Signal Processing. pp. 143–155. Springer, Singapore (2019). https://doi.org/10.1007/978-981-13-1513-8_16.