ISSN No:-2456-2165

Classification of Max Rider's Earning using Machine Learning

O.J. Timilehin
Pelliumlinks limited,
Lagos, Nigeria

O.A. Balogun, M.A. Dosunmu
Department of Computer Science and Information Technology, Bells University of Technology, Ota, Nigeria

Abstract:- This work looked into the ban of bike hailing activities which came into effect in the year 2020 by the Lagos state government. The work looked into the various angles that may have influenced this decision and what factors influenced the earning potential of bike riders. We made use of very limited data due to the reluctance of bike hailing companies to release more and initial analysis showed that there were a lot of discrepancies embedded in it. Using data pre-processing techniques, we were able to get more insight and carried out machine leaning classification operations consisting of Linear Regression, K-Nearest Neighbor, and Support Vector Classifiers to determine the most influencing earning factors. Results showed that all three methods performed averagely and it was recommended that more accurate and voluminous data will be required to predict better results.

Keywords:- *Transportation, Machine Learning, Information technology, Hailing services, Ban.*

I. INTRODUCTION

The advent of the internet and its penetration into every facet of human activity has enabled a shift from the traditional way of doing things. In today's world, businesses and customers do not need to interface physically to transact many types of business. Technological advancement has not only made it possible to make use of the internet to carry out transactions but new technologies are also making it possible for customers to experience the look and feel of the purchases. Today, customers can virtually experience and customize their product via augmented and virtual reality, and also tracking of a purchased product movement from the point of purchase to its destination.

The transportation system in Nigeria is plagued with numerous challenges ranging from societal to management. The Nigeria's transportation system reveals a sector suffering from warped or defective developmental approach [2], although year after year, a whopping amount of money is allocated to the transportation sector, 158 billion from Federal allocation in 2021 alone [13], there seem to be no end to the hassles confronting the effective management of road transportation which suggests that it is more of a policy and human problem than monetary. Transportation is so much important to the economic well-being of any society that we have seen a recent uptick in interest and investment for autonomous vehicles, re-usable and clean energy to power them, trains (Japan, South-Korea and China making use of maglev® technology), re-usable space shuttle transportation for space exploration (SpaceX, blue-origin, and Virgin), revival of trains for mass transportation in Nigeria (and Morocco being the first country in Africa to own a bullet train). The Lagos state government has also invested massively in light rail, BRT services, water transportation etc.

The traditional mode of making use of public transport facilities has also been impacted uponby information technology. Technologically inclined companies like Uber® and Lyft®, armed with the idea of "shared economy" where business practically own no asset but provides an information technology platform for the seamless running of the organization's operations, changed the landscape of transportation business worldwide. Before recent times, the transportation system in Lagos, Nigeriaconsisted mainly of various types of buses i.e., "Danfo", "LT"; Federal Mass Transit buses, Old Mercedes Benz buses popularly called "Molue",seven-seater mini-buses known as "Korope", Red Buses, Blue BRT (Bus Rapid Transit) buses, and other smaller taxi services in their variations, and motorcycle operators popularly known as "Okada or bike".Save for the blue BRT buses operated in conjunction with the state government, the other modes of transportation are usually not well maintained and serviced, and the customer relationship is toxic. [3] posits that lagos roads play host to over 5 million cars and 200,000 commercial vehicles; furthermore, in 2017 alone, Lagos State recorded an average of 227 vehicles per kilometre of the road on a daily basis.Despite this assertion, these authors observed that the population of people who make use of Lagos roads vis-à-vis the number of well-functioning buses is inadequate going by the number of stranded commuters especially during rush hour.

The introduction of information technology inclined transportation services in Lagos was to a larger extent, able to fill the gap of taking rides in a comfortable and well-maintained vehicle whose price was based on distance covered and time factors. Such lofty implementations gave some reprieve to passengers but unfortunately, the state of bad roads [5], [10] and the number of users especially during rush hour, coupled with human factors like failure to adhere to traffic rules subjected every means of transportation within the metropolis to traffic anomalies. To beat traffic, passengers often make use of motor-cycle services because they are faster, have easier maneuverability, and can easily access and pass through the most horrid of roads, although they are comparatively more expensive. At different times, the Lagos state Government had tried to regulate this mode of transportation as a result of their reckless endangerment of passengers and other road users, propensity to cause accidents and their use in robbery operations. The gap left was tried to be filled by bike hailing

ISSN No:-2456-2165

services such as Max®, ORide®, and Gokada® whose operations where similar to that of Uber where a passenger would order for a trip using the organization's application. It is pertinent to note here that this event was simultaneously taking place in other economically similar countries like Indonesia, Thailand, and India some of which took almost a decade to find a framework to properly regulate the sector [8].

In February of 2020, the Lagos state Government banned the operations of all bike hailing services within the Lagos metropolis citing that such services are unfit for the state's megacity status coupled with the argument of accidents caused by bikes. [8] also posits that the decision may have been as a result of the politically influential transport unions i.e., the Road Transport Employers' Association of Nigeria (RTEAN) and the National Union of Road Transport Workers (NURTW) wanting to force bike hailing companies under their umbrella and get such bikers to pay daily levies of ₦500. Expectedly, the ban affected the earning potential of such bikers within the state.This work seeks to understand the financial impact of the ban on the riders earning capacity and what factors determined riders earning potential. This work tries to identify and analyze how the ban of bike hailing services activities affected the riders. This will be achieved by carrying out a comprehensive literature review in order to understand transportation at a local and global scale and identifying the economic impact of bike hailing services and the implication of the ban using limited Max® data set between January and December, 2020.

II. LITERATURE REVIEW

Every human society has come to depend on transportation for the movement of people and products. Transportation is therefore very vital to the social and economic life of any society. It makes seamless, trade, manufacturing, tourism and serves the distribution needs of any geographical location via air, land, and sea.

A latest dimension that will be cogent to the future of transportation is the use of Artificial Intelligence algorithms for the purpose of classification, prediction, and identification with the transportation space. Machine Learning (ML), a subset of AI that uses statistical methods to enable machines to learn and improve with experience. Deep Learning is a subset of Machine Learning, which makes the computation of multilayer neural networks feasible. Machine Learning is seen as shallow learning while Deep Learning is seen as hierarchical learning with abstraction. One of the major challenges encountered in machine learning models is a process called feature extraction. The programmer needs to be specific and tell the computer the features to be looked out for. These features will help in making decisions. Entering raw data into the algorithm rarely works, so feature extraction is a critical part of the machine learning workflow.

For instance, [1] proposed a model which used deep CNN with auxiliary data. They used deep CNN for feature extraction and SVM for classification. This was applied on FHWA (The Federal Highway Administration) vehicle dataset and recorded 89% classification accuracy. [16] proposed a system which classified sedan, vans, pickups and trucks using AlexNet and inception model for classification. He used a manned aerial vehicle equipped with an infrared sensor to capture vehicles and recorded 80.3% classification accuracy.

Lagos state is especially challenged with road transportation problems ranging from bad roads, to indiscriminate parking, compromised traffic officials to disobedience of traffic regulations etc. These issues in turn transform into more cascading problems like traffic congestion, parking problems, accidents and environmental pollution and so on. Also, vehicles are seen crawling on the roads especially both in the morning and evening peak periods. This amounts to daily loss of time and money that could have been otherwise utilized to achieved better result-oriented goals. In the analysis of the cost of [11] shows that Lagos state lost N4trn annually, and =N=14.12m hours daily to traffic congestion alone.

The ban of bike hailing services in February 2020 came as a rude shock to the general populace and was seen in some quarters as a fight against progress, innovation, and entrepreneurship. This is because bike hailing companies were already required to pay =N= 181.8 million annually to continue their operations [6], their drivers were always well kitted and insisted on all safety measures before a ride would commence unlike what is still obtainable from state regulated bike services. Furthermore, the owners of such companies where young (within the youth age bracket) enterprising people who had checked all the boxes in order to do business legally in Nigeria and meeting all State's requirements. This gave credence to statements that saw the move against such services as politically motivated which in the long run would have an adverse effect on the eagerness of other entrepreneurs to pick Lagos, and maybe Nigeria as a whole as their business destination choice.

Another school of thought came from the angle that a Nation with very high youth population, that contains about 73% unemployed of which 64% falls within the youth age bracket i.e., between 15 and 29 years of age [9] and should not be toying with the means of livelihood for so many in order to curb other social vices that may emanate from restfulness.

These organizations made use of the latest technologies to run their business seamlessly as is obtainable in other parts of the world where such business exists. Texts show that these businesses thrive in other nations like Thailand and India because they are well run, and accountable within the ambit of laws and regulations. Despite the best efforts of the Lagos state governments, long queues and stressed-out commuters has become common sight within the metropolis.The idea that Lagos is a megacity and should not accommodate bike services [8] does not also stand well in some quarters because present day Lagos cannot stand toe-toe with the kind of road-network infrastructure that are obtainable in some of the countries where the bike hailing services exists. It is also

ISSN No:-2456-2165

pertinent to point out here that unlike the other nations that have well graded inter-roads, Lagos state to a very large extent, safe for some parts of the Island lack this sort of road network. This point is another cogent reason why commuters favor this mode of transport because it is not only safe, but its bikes are bigger and up to the standard Lagos state government recommended i.e., motorcycles or tricycles with engine capacity above 200cc as stipulated in the Lagos State Transport Sector Reform Law 2018 [4]. This gives the bikes the added advantage of maneuvering bad spots on the road and the ability to drop a passenger at the point where the passenger intends without having to board another vehicle thereby saving time and cost.

Perhaps the most important impact of the ban is on the livelihood of the drivers whose source of income was suddenly tampered with. [12] shows that some of these riders make as much =N=200,000 a month which is at par with the salaries of some entry level-mid level white-collar jobs. It gives room for flexibility and a chance for some of these riders to pursue other interests unlike a 9 a.m.-5 p.m. job. Asides for the loss of stable income, other investments that may have gone into personal developments in order to scale the selection process have gone down the drain due to the ban policy.

[15] posited that the rapid growth of development, makes road travel more convenient to travel by highway and rail transport which in turn promotes and influences the growth of the national economy. The relationship between the national economy and transportation seems complex because these two parameters form a dynamic system made up of a variety of factors, and these factors are interdependent and interact with each other when the system is operating. They show that current research is geared towards the view that the national economy and transportation complement each other; transportation is not just a condition of regional development, but also a result of it. In analyzing the relationship between transport and GDP of the economy, they made use of 14 variables showed in figure 2.12. and made use of GHRE (GDP-highway transportation-railway transportation-economic indicators of transportation) to show that their relationships issues are verycomplicated. Also, the Lotka-Volterra nonlinear system was used to capture the dynamic competitive relationships' principal features among them. A univariate linear model, the ARIMA (autoregressive integrated moving average) model was applied to predict the principal factors from 1985 to 2015, and to supply a benchmark to estimate the LV-COMSUD model's consequences. To evaluate the forecasting capability of the LV-COMSUD model, forecasting accuracy is examined by calculating the MAPE (mean absolute percentage error) statistic

$$MAPE = \frac{\sum_{i=1}^{n} |P_i - A_i|}{A_i} \ldots\ldots\ldots \text{Equation 1}$$

where A_i is the ith actual value, P_i is the ith predict value, and n is the number of forecasts.

The work concluded that understanding the different types of interactions among GDP and land transportation not only facilitates policy development to achieve a balance in the relationship between economy and transportation, but can also provide technical support and a scientific basis for the advancement of transportation development planning.

[14] posited that performance indicators are widely used to empirically assess the technical performance of different transport modes, namely their capacity to move passengers or freight around. They categorized these indicators into Network and Operational Indicators, Transport time / speed / turnover, Reliability, Punctuality, Load factor, Road Traffic Performance (Road conditions; Traffic conditions; Control conditions i.e., Attributes of the control structures and existing traffic laws such as speed limit, one-ways and priority, Economic Performance Indicators i.e., Output / Capital ratio is commonly used to measure the capital productivity of transportation, Output / Labor ratio performs the same productivity measurement but for the labor input. Capital / Labor ratio aims at measuring which factor predominates within the relationship between capital and labor productivity. They go further to show that more scale-specific indicators can also be used to evaluate the role of transport within the economy especially with freight transport, Output / GDP ratio which measures the relationship between economic activity and traffic intensity, Transport sector income / Local income ratio which evaluates the share of the transport industry in the local economy (e.g., municipal or state level), and Output / Local income ratio which is a measure of the relative production value of the transport industry output.

[7] made use of economic attributes to predict traffic indication that there is a correlation between them. They posited that an increase in traffic flow in a region reflects the development of transport infrastructure. Transport infrastructure development arises due to new investment in the region as it is a significant factor for assessing the social and economic viability of any highway project. Impact of economics on traffic is difficult to analyze traffic flow projection. They made use of investment and employment generated factors for the purpose of their work and made use of a regression model for predicting traffic with economic attributes stated above and found to be linear. Their precision of model was gauged and based on mean absolute percentage error (MAPE) rather only considering the coefficient of determination.

III. DATA COLLECTION AND IMPLEMENTATION

Based on the works of [7] and [15], this work combines the most appropriate parameters in the prediction model that covers investment, employment, passenger volume, vehicle mileage, and total number of rides parameters. Which will be used to answer the hypotheses questions.

The data given by Max spanned the record of a year i.e., January to December 2020 just before the ban of all operations containingRiderid, commission Paid, Positionid, State, Dob, Sex, Marital_desc, Employment status, Managerid, Av. Trips, Trips parameters in the first set of data i.e Jan-Feb, 2020 which contains 311 records.The

ISSN No:-2456-2165

second data set is made up also of 311 records RIDERID, COMMISION PAID PositionID, State, DOB, Sex, MaritalDesc, EmploymentStatus, ManagerID, TRIPS.

The Data show some similar records; this work therefore is carried out by the following steps: Data cleaning and analysisusing Microsoft Excel 2016, a spreadsheet application software for the first phase of data analysis i.e., cleaning and integration while the second phase i.e., classification will make use of Python. Three methods of classification i.e., Logistic Regression, K-nearest neighbor, and support vector machine classifiers will be used to measure and compare their accuracies.

The classification operation made use of SciPy, NumPy, Matplotlib, Pandas, Sklearn libraries. The change in the sex commission paid, and employment columns of the Max® dataset suggests that there was a change of rider after the second month of 2020 and so in order to capture all data, this work assumes that the RiderID stays unique to each rider and we therefore generate a set of unique new RiderIDs for the set of new data that exists between march and December, 2020. This was possible because the RiderID based on the available, has no bearing on the commission.

- **Hypothesis**
  - Ho: Commission paid to riders is not dependent on age, marital status, or sex (gender)
  - H_1: Commission paid to riders is dependent on age, marital status, or sex (gender)

IV. RESULT

After analysis of the data, the following information were obtained

Jan-Feb 2020		Mar-Dec 2020	
Row Labels	Count of Sex	Row Labels	Count of Sex
F	176	F	2
M	135	M	309
Grand Total	311	Grand Total	311
Row Labels	Count of MaritalDesc	Row Labels	Count of MaritalDesc
Divorced	30	Divorced	30
Married	124	Married	124
Separated	12	Separated	12
Single	137	Single	137
Widowed	8	Widowed	8
Grand Total	311	Grand Total	311

Table 1: Demographics of Max riders by Sex and Marital Description for the year 2020

Jan-Feb 2020		Mar-Dec 2020	
Row Labels	Sum of Trips	Row Labels	Sum of TRIPS
F	180600	F	23
M	137800	M	3161
Grand Total	318400	Grand Total	3184
Row Labels	Sum of Trips	Row Labels	Sum of TRIPS
Divorced	30000	Divorced	300
Married	135500	Married	1355
Separated	10300	Separated	103
Single	133900	Single	1339
Widowed	8700	Widowed	87
Grand Total	318400	Grand Total	3184

Table 2: Sum of trips by Marital Description and Gender for the year 2020

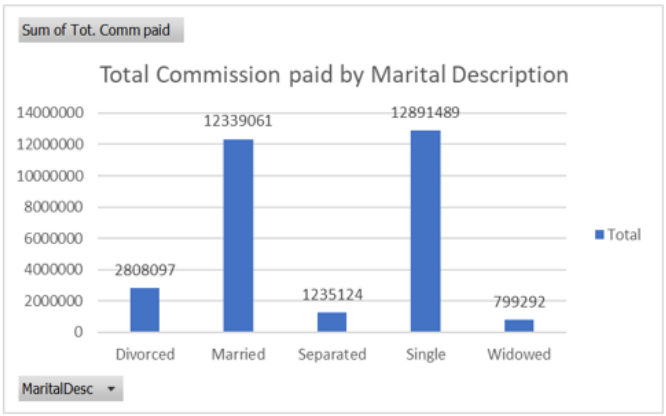

Table 3: Total Commission paid by Marital Description

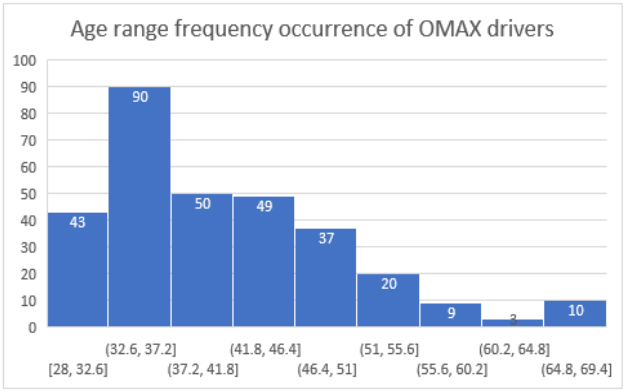

Table 4: Frequency Age range of Max riders

The analysis of the data showed the following observations:

RIDERID being the unique identifier for each record's unique identifier is repeated in both data sets. Analysis show disparity in the sex column of the data as the rider gender changed from male to female in some cases. Further analysis shows that the gender change coincided with a voluntary termination of work services between March and December of 2020. This suggests that the RIDERID is a unique identifier for the motorbikes and not the bike rider.

ISSN No:-2456-2165

Data shows there were 311 with 176 female and 135 male riders between January and February of 2020. Interestingly, we see that both sets show the same information with respect to marital demographics. This suggests that the data given was flawed especially with regard to the gender of the bike riders. The data also shows that there was a sharp decline in the number of female riders (from 176 to 2) within a two-month period. This may have been as a result of the altercations between bike riders and government sanctions road tax collectors and described in section 2 of this work. It may have also been a s a result of marriage, relocation etc.Analysis shows a total of 318, 400 trips within January and February and just a thousand fraction of the same figure for the rest the year. It also shows that female riders completed 42, 800 more trips than their male counterpart. in the reminder of the business year, their male counterpart completed 3138 more trips.

On the marital description front, the data shows that there are over 135,000 married riders with the singles figure showing 133,900. Interestingly, the youngest age of Max data is 28. This suggests the that are half-truths in the data collected.Analysis show that riders in the single marital description category made more money from completed trips and were commissioned a total of =N=12,89,1489 compared to that of married group which earned =N=12.330,061. We can deduce that that those in the singles category were able to take longer and probably more tedious trips which made it possible for them to earn more.

The large age range of bike riders suggests that it was a business that accepted as many as possible as long as criterium were met. This suggests that the earning potential of this service was at par in accordance with findings from the literature review of this work and so was used to sustain the expenses of families. Going further, with the assumption that other civil categories have nuclear families comprising spouse(s) and children, this would mean that the financial earning from this business for such families combined would amount to over 17 million Naira for the year 2020.

	MaritalDesc	Trips	PositionID	DOB	Sex
count	383.000000	383.000000	383.000000	383.000000	383.000000
mean	0.550914	0.000832	2.447898	0.041480	0.535248
std	0.498052	0.000661	14.938466	0.009158	0.499408
min	0.000000	0.000001	0.001000	0.028000	0.000000
25%	0.000000	0.000200	0.018000	0.034000	0.000000
50%	1.000000	0.000700	0.019000	0.039000	1.000000
75%	1.000000	0.001450	0.020000	0.047000	1.000000
max	1.000000	0.002000	152.058000	0.069000	1.000000

	EmploymentStatus	ManagerID	COMMISION PAID
count	383.000000	383.000000	383.000000
mean	0.812010	0.014018	75.694055
std	0.391215	0.008065	42.526407
min	0.000000	0.000000	2.957000
25%	1.000000	0.010000	57.699000
50%	1.000000	0.014000	67.878000
75%	1.000000	0.019000	99.775500
max	1.000000	0.039000	314.000000

Table 5: Dataset description showing count, mean, and standard deviation of each parameter in the Max dataset.

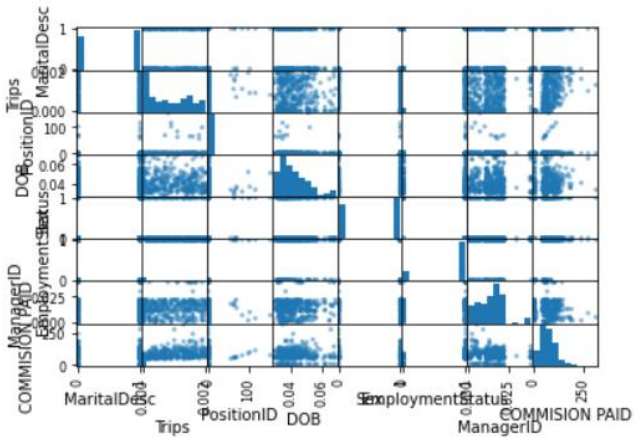

Table 6: Scatterplot showing dependencies between variables

Classification	Prediction
LR: 0.521746 (0.115194) KNN: 0.546190 (0.108497) SVM: 0.532540 (0.116213)	LR 0.5844155844155844 precision recall f1-score support 0.0 0.55 0.35 0.43 34 1.0 0.60 0.77 0.67 43 accuracy 0.58 77 macro avg 0.57 0.56 0.55 77 weighted avg 0.58 0.58 0.57 77 KNN 0.5844155844155844 precision recall f1-score support 0.0 0.53 0.50 0.52 34 1.0 0.62 0.65 0.64 43 accuracy 0.58 77 macro avg 0.58 0.58 0.58 77 weighted avg 0.58 0.58 0.58 77 SVM 0.5844155844155844 precision recall f1-score support 0.0 0.52 0.91 0.66 34 1.0 0.82 0.33 0.47 43 accuracy 0.58 77 macro avg 0.67 0.62 0.56 77 weighted avg 0.69 0.58 0.55 77

Table 7: Result of classification and prediction

The scatterplot (Table 6) shows that majority of the data points are independent of each other safe for a few which shows some level of linearity. This speaks to higher level degree of randomness of requested rides and also, that riders' ability to larger extent have no bearing on other factors. As expected, none of the classifiers was able to classify and predict the data to a higher precision. This led to a combination of the data-points with most relevance which landed on Commision_Paid, Age, and Sex variables. Table 7 shows that KNN classifier had the best result but all three performance was within the same range. The prediction result shows that for this particular dataset, they were all at par. The result indicates these classifiers may not be able to properly generalize new data to a higher precision and therefore shows that a lot of missing data and or wrong data exist which gives credence to earlier in this section. Based on the analysis and results of the Max® dataset, we therefore accept Ho.

V. CONCLUSION

This work made use of Max® dataset for the year 2020 to analyze the factors that influenced the earning potential of

riders. Initial analysis showed no dependencies of factors that influenced earning capacity and classification methods results showed a lower level of accuracy in both classification and predictions phases.Also, given the acceptance of Ho, we see that the ban had an adverse effect on the earning potential for all rider staff of the Max venture and can safely conclude that it has had a cascading effect on all other type of staff and investors.

In taking this work further, more accurate data that could show linear relationships between factors be used to further this work and enlarge the scope. Again, a factor indicating the number of dependents could throw more light on the effects of how the ban on these hailing services affects the economic capacity of stakeholders.

REFERENCES

[1.] Adu-Gyamfi, Y. O., Asare, S. K., Sharma, A., & Titus, T. (2017). Automated Vehicle Recognition with Deep Convolutional Neural Networks. *Transport Researb Board, 2645*(1), 113-122. doi:Automated Vehicle Recognition with Deep Convolutional Neural Networks

[2.] Badejo, B. (2011). Transportation, removing the clogs to Nigeria's development.

[3.] Biobaku, S. (2019). *20 Different Types of People in Lagos Public Vehicles*. Retrieved Aug 6, 2021, from https://blog.autochek.africa/lagos-public-vehicles/

[4.] Dakeje-Ibraheem, A. (2019). *Nigeria: Lagos State Transport Sector Reform Law 2018*. Retrieved from https://www.mondaq.com/nigeria/rail-road-cycling/853306/lagos-state-transport-sector-reform-law-2018

[5.] Editorial Board. (2020). *When will Lagos face inner-city roads?* Retrieved August 6, 2021, from https://guardian.ng/opinion/when-will-lagos-face-inner-city-roads/

[6.] Emmanuel, P. (2020). *Lagos Okada Ban: A policy expert walks us through some burning questions*. Retrieved August 7, 2021, from https://techpoint.africa/2020/02/04/okada-ban-burning-questions/

[7.] Galani, R., & Majithiya, D. P. (2018). Prediction of Traffic using Economic Attributes. *5*(1). doi:10.37591/ttea.v5i1.501

[8.] Idris, A. (2020). *The Asian example: how to regulate okada in Lagos*. Retrieved August 6, 2021, from https://techcabal.com/2020/02/19/asian-example-for-regulating-okada-lagos/

[9.] Maina, J. (2020). *EndSARS: How Nigeria can tap into its youthful population.* Retrieved 2021, from https://theconversation.com/endsars-how-nigeria-can-tap-into-its-youthful-population-148319

[10.] News Agency of Nigeria. (2021). *Lagos government upgrades 56 roads, others on the Island*. Retrieved August 6, 2021, from https://guardian.ng/news/lagos-government-upgrades-56-roads-others-on-the-island/

[11.] Nnorom, N. (2021). *Lagos loses N4trn annually, 14.12m hours daily to traffic congestion*. Retrieved August 7, 2021, from https://www.vanguardngr.com/2021/02/lagos-loses-n4trn-annually-14-12m-hours-daily-to-traffic-congestion/

[12.] Ola, A. (2020). *Opay oride | Earn N10K Every Day*. Retrieved from https://www.gistfocus.com/n10k-every-day-with-opay-oride/

[13.] Omidiji, O. (2021). *Federal Governemnt allcocation to Transport sector*. Retrieved August 7, 2021, from https://monitor.yafri.ca/federal-government-allocation-to-transportation-sector/

[14.] Sun, J., Li, Z., Lei, J., Teng, D., & Li, S. (2018). Study on the Relationship between Land Transport and Economic Growth in Xinjiang. *Sstainablity*, 1-17. doi:10.3390/su10010135

[15.] Sun, J., Li, Z., Lei, J., Teng, D., & Li, S. (2018). Study on the Relationship between Land Transport and Economic Growth in Xinjiang. *Sstainablity*, 1-17. doi:10.3390/su10010135

[16.] Tan, Y., Xu, Y., Das, S., & Chaudhry. (2018). Vehicle detection and classification in aerial imagery. *IEEE International Conference on Image Processing* (pp. 86-90). IEEE.