

Analyzing terrific traffic in urban areas: a small step towards bringing order into city roads

Md. Sakhawat Hossain¹, Md. Mosaddekul Islam², and Ashikur Rahman³

Abstract Accurate travel time information enables travellers to plan their journey more wisely and efficiently. This in turn, lessens traffic congestion and improves people's travel experience, particularly in urban areas. Open-source traffic data available from different sources and Google Map API have raised opportunities for analyzing and predicting the traffic more accurately. The purpose of this research work is to analyze bus or car travel time data and gain different insight and aspect of a society from its traffic pattern. Google Distance Matrix API, python programming language and machine learning algorithms have been applied in this study to automatically extract, analyze, and visualize traffic data and showcase analysis methodology to improve people's travel experience in Dhaka city and the city of New York. In particular, we apply data analytics to develop an oracle that will give answers to different queries about traffic, such as least congested period and/or least congested route within a day/week/month etc, which in turn would enable people to make informed decisions while travelling. The experimental results and detailed analyses show that there exists a wide fluctuation of travel time during the day in both cities. Furthermore, unlike other works, we accomplish various socio-cultural aspects and behaviour from traffic patterns in those two cities, perform the accessibility analysis and provide recommendations for further research.

Key words: Traveling Time Prediction, Crowdsourced data, Google Distance Matrix API, Polynomial Regression, scikit-learn

Md. Sakhawat Hossain
Department of CSE, Bangladesh Univ. of Engg. & Tech. (BUET), Dhaka-1000, Bangladesh,
e-mail: mdsakhawathossain05@gmail.com

Md. Mosaddekul Islam
Department of CSE, Bangladesh Univ. of Engg. & Tech. (BUET), Dhaka-1000, Bangladesh,
e-mail: mosaddek2020@gmail.com

Ashikur Rahman
Department of CSE, Bangladesh Univ. of Engg. & Tech. (BUET), Dhaka-1000, Bangladesh,
e-mail: ashikur@cse.buet.ac.bd

1 Introduction

Transportation in an urban area influences urban life to a great extent. Predicting travel time is an important area of research that can improve travel experiences in urban areas. In Intelligent Transportation Systems (ITS), the travel time prediction or traffic pattern prediction is a daunting task. It can be simplified using the regression technique because an approximation of the current travel time can be determined based on the previous travel time. However, collecting reliable data about traffic imposes an additional challenge. Many techniques exist for collecting travel time data such as global positioning system (GPS), cellular Geo-location, crowd-sourced travelling data, etc. Thus predicting travel time is an interesting blend of technology and machine learning. The technology part is responsible for accurately capturing and feeding the traffic data to a machine learning model that can analyze and predict the required travel time.

In this work, we use *polynomial regression* to derive machine learning models that can provide an estimation of the travel time based on past travel behaviours in the city of Dhaka and the city of New York. We have chosen these two cities as our testbeds because these are the heavily crowded zones in the world. To this extent, crowd sourced Google Distance Matrix API has been used for collecting traffic data in Dhaka city. On the other hand, publicly available taxi trip data from NYC Taxi and Limousine Commission (TLC) has been used for New York City. Machine Learning has been used to establish the initial framework as well as to extract, analyse, and visualize the data.

Experimental results show that our model can provide an aggregate view of the traffic with summary statistics based on previous observations. Given a particular ‘source-destination’ pair, our derived models can also provide several useful results such as a graph of weekdays’ relative travel time, expected travel time from a specific time in the day, least congested path within a day/week/month, least congested moment within a time interval etc. By using the built prediction models, it is also possible to infer social behavior from traffic patterns such as week-end behavior of the community, office hours, rush hours, holidays, night time travel habits etc. After successful implementation of the model it has been identified that the data provided by Google Distance Matrix API is highly reliable. The public authorities and transportation agencies may use this model to control traffic and effectively reduce the congestion.

The major contributions of this research work are summarized below:

- We utilize freely available traffic data sources and showcase analysis methodology to improve people’s travel experience in Dhaka city and the city of New York.
- We devise methodologies to access those data and apply data analytics to develop a model. The built model can provide traffic forecast ahead of time, for example, “what will be the right time or date to visit a place within a given day/week/month?”.

- Our built model can assist travellers/visitors to prepare a schedule considering the traffic so that they can visit a maximum (fixed) number of places within a given period.
- Our traffic model has the ability to assist government for city-wise traffic comparisons for making policies. Besides, it can assist government to take necessary measures to mitigate challenges related to traffic Jam. For example:
 - Suppose, the government wants to build a flyover or other infrastructure to reduce traffic Jam, then what will be the appropriate place to utilize the investment to the maximum.
 - Suppose, the government wants to build office or other infrastructure then what will be the appropriate place to build such premises so that the traffic density is equally sparse and it does not affect the surrounding traffic.
- We develop an oracle that will answer different queries about the traffic, such as
 - Traffic pattern
 - Least travel time
 - Least congested path
- Unlike other works we infer different Social behavior from traffic patterns and compare two societies based on their people's travel habit such as:
 - Week-end community behaviour
 - rush/office hours
 - holidays
 - night time travel habits
 -

2 Literature Review

A number of research works have been conducted for travel time prediction [1]. Siripanpornchan et al. [2] describes the concept and application of Deep Belief Networks (DBN) to predict travel times. In their model they use DBN as a stack of Restricted Boltzmann Machines (RBM) to automatically learn generic traffic features in an unsupervised mode, and then a sigmoid regression has been used to predict travel time in a supervised mode. In [3], the authors explore a deep learning model named the LSTM neural network model. They used real travel time dataset of 66 routes provided by Highways of England which was split into three parts: training set (80%), test set (10%) and the validation set (10%). In this dataset, the travel times are 15-minute interval average travel times for each route. Avigdor Gala et al. [4] investigate methods from Queueing Theory and Machine Learning in the prediction process. They implemented model based on segmentation of the travel time into stop-based segments using bus data of Dublin city. In [5], the researchers use Google Map's Distance Matrix API [6] to collect travel distances and time for

an origin-destination (OD) pair at a specified departure time. In their analysis, there are number of OD pairs, where $n^2 - n$ is the number of transport analysis zones. Additionally, they extracted each OD pairs' travel time at thirty-minute intervals during the day (from 3:00 to 21:30), which resulted in slightly more than 1.6 million requests [7]. *Ying Hay Chiu* [8] also use crowdsourced data to present a model that is used for identification and evaluation of opportunities to apply Transportation Demand Management strategies. *Deri* and *Moura* [9] present a spectral analysis of taxi movement based on the graph Fourier transform on "NYC Taxi Data Set", a historical repository of 750 million rides of taxi medallions over four years (2010-2013). They present the spectral decomposition of a large directed, sparse matrix. Important considerations toward handling this matrix are explained. Their Preliminary results show that their method can pinpoint locations of co-behavior for traffic in the Manhattan roads.

3 Data Collection and Preparation

We have worked on two data-sets: one is the New York City yellow taxi trip data and the other is the travel time data of the Dhaka city collected from Google API. Both are discussed in details below section 3.1 and section 3.2 respectively.

3.1 The TLC and Data

NYC Taxi Trip data-set is a publicly available data-set. NYC Taxi and Limousine Commission (TLC) has provided this public data-set which collects trips information for each taxi and other licensed vehicles. It collects different types of information for each trip including pickup and drop-off timestamp, pickup and drop-off locations, number of passengers travelled, payment method, distances of trips, fare amount and others. All this data are available at <https://www1.nyc.gov/site/tlc/about/tlc-trip-record-data.page>.

3.1.1 Data Description

The Yellow Taxi Trip Records contains the data regarding several taxi trips in New York City. In each trip record, one row represents a single trip made by a TLC-licensed yellow taxi that has different field such as:

- **Pickup Location ID:** This field contains pickup location ID of a trip. It is the TLC Taxi Zone in which the taximeter was started.
- **Drop off Location ID:** This field contains dropping location ID of a trip. It is the TLC Taxi Zone in which the taximeter was stopped.

- **Pickup Datetime:** This field contains pickup date-time of a trip. It is the date and time when the meter was initiated.
- **Drop off Datetime:** This field contains dropping date-time of a trip. It is the date and time when the meter was stopped.
- **Trip Distance:** The elapsed trip distance in miles was kept in this field which was reported by the taximeter.

3.1.2 Data Extraction

First, data was extracted from NY yellow taxi travel data-set taking ‘Starting_locationID’, ‘Dropping_locationID’, ‘Pickup_datetime’, ‘Dropoff_datetime’ and ‘Distance’ fields only. Three new fields - ‘Duration’, ‘Weekday’ and ‘Pickup_time_min’ were added. ‘Duration’ was calculated from pickup and drop-off date-time which is the difference between drop-off and pickup date-time. ‘Weekday’ is the day of the pickup date-time. Only time from pickup date-time was considered to keep input range between 00:00:00 to 23:59:59. The time was mapped to numerical value for polynomial regression and stored in the ‘Pickup_time_min’ field.

For example, suppose the pickup date-time is ‘2020-09-20 12:10:30’ and drop-off date-time is ‘2020-09-20 12:30:00’ So, ‘Duration’ = 19 minutes and ‘Pickup_time_min’ = $12 * 60 + 10 = 610$. ‘Weekday’ of the pickup date-time is Tuesday.

3.2 Travel Data-set for Dhaka City

The Dhaka city is divided into different zones and each zone is assigned with a unique ID. A lookup table is used to map source and destination name to their corresponding IDs(Table 1).

Table 1: Look up table for different zones.

Location	Location_ID	Area
Azimpur Bus Stop	1	Azimpur
Asad Gate	2	Mohammadpur
Mazar Road Bus Stop	3	Gabtohi
Uttara Sector 1	4	Uttara
Mohakhali Flyover	5	Mohakhali
Nikunja 1	6	Nikunja
Shahbagh Square	7	Shahbagh
Middle Badda	8	Badda
Mouchak Market	9	Malibagh
Turag Recreation World	10	Turag
...

We considered only some locations and their corresponding routes.

Google Maps Distance Matrix API [6] is a service that returns travel distance and time for a $\langle origin, destination \rangle$ (OD) pair at a specified departure time. The information received is based on the recommended route between the origin and destination, as calculated by the Google Maps Distance Matrix API. In our analysis, there was a number of OD (Origin-Destination) pairs. We extracted each OD pairs' travel time at thirty-minute intervals during the day (from 6:00 to 23:30), Every request was sent by an automated script, which constructed the Uniform Resource Locator (URL) by using the following parameters [6]

- Longitude of origin (required);
- Latitude of origin (required);
- Longitude of destination (required);
- Latitude of destination (required);
- Mode of travel (optional);
- Date and time of departure (optional).

Results of the Google Distance Matrix API were provided as a JavaScript Object Notation (JSON) objects which include

- Duration.
- Duration_in_traffic.

3.2.1 Data Description

The travel time data-set collected using Google Distance Matrix API contains the data regarding expected travel time in Dhaka City. Each row in the data-set represents a single trip that has different fields such as:

- **Starting_point_id**: The area mapped by an integer from which the trip started.
- **Dropping_point_id**: The area mapped by an integer from which the trip ended.
- **Pickup_datetime**: The date and time when the trip started.
- **Duration_in_traffic_value**: The length of time it takes to travel this route in minutes stored as a numerical value. This time is calculated based on current and historical traffic conditions.
- **Weekday**: The weekday on which the trip is made.

3.2.2 Data Extraction

Required data was extracted from the travel dataset of Dhaka City taking 'Starting_locationID', 'Dropping_locationID', 'Pickup_datetime', 'Dropoff_datetime', 'Distance', 'Duration' and 'Weekday' fields only. Only numerical values were taken for distance and travel duration where 'Duration_in_traffic' were taken as duration. The new field 'Pickup_time_min' was also added where the only time from pickup datetime was considered to keep input range between 00:00:00 to 23:59:59 and the times were mapped to numerical values.

4 Training and Tuning Model

After removing unexpected data, the data-set became ready and is split into three subsets:

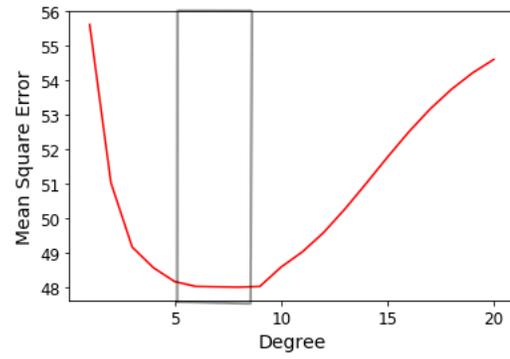
- Training data-set
- Validation data-set
- Testing data-set

The training data-set is used to train a machine learning model. A validation dataset has been used to validate the model and to overcome the overfitting and underfitting phenomenon. Finally, the testing data-set is used to test the final model.

Polynomial regression has been used to train the model. *PolynomialFeatures* class provided by scikit-learn is used for training purpose. Then, for every pair of locations, seven models for seven days of a week have been implemented to predict travel time for a given source, destination and pickup/drop-off date and time.

For each model, **Mean Square Errors** are calculated and compared for different degrees, and the degree for which the **Mean Square Error** is minimum has been selected. Fig. 1 shows one such instance of a degree of polynomial selection for New York City for the model used for Saturday traffics. After tuning the model, *Intercept* and *Co-efficients* for each model are saved in a *csv* file including starting and drop-off locations, weekday and degree of regression.

Fig. 1 Mean Square Errors for a model corresponding to the location pair (LaGuardia Airport, Midtown East) on Saturday shown for different degrees of polynomial regression. Here, MSE is lower for the degrees of polynomial ranging from 6 to 9, thus any degree from this range could be selected. In our model, we have chosen 8 as the degree of regression which generates minimal MSE.



5 Analysis

Once we fit the data into a regression model, we can proceed with the data analysis. The following sections provide our results, namely: (i) a graph of weekdays relative travel times, (ii) a matrix of mean weekday travel times, (iii) expected travel time from a specific time, and (iv) Seven models for seven days of a week.

5.1 Right Time to Travel

Once built, our model is ready to answer different questions such as “What will be the right time or date to visit a place within a given day/week/month”? A pickup time with a minimum required travel time will be recommended. This pickup time can be a time within a day or a week or even it can be a time interval in a day. Different time intervals as shown in Table 2 has been defined for generating outputs. Four intervals within the day time are considered as during night time the travel time is always minimum.

Table 2: Time intervals of a day.

Portion of the Day	Time rage
Morning	6:00 - 10:00
Midday	10:01 - 14:00
Afternoon	14:01 - 18:00
Evening	18:01 - 22:00
Night	22:01 - rest

5.1.1 New York City

We select a pair of locations (LaGuardia Airport, Midtown East) of New York City to analyse different queries. Some of the possible queries and the corresponding results are shown below.

(a) Day-wise query: Suppose, the user wants to know, for a given day, what is the minimum travel time and corresponding departure time within different segments of the day? The answer from the model that includes expected departure time and required travel time from LaGuardia Airport to Midtown East on Saturday is shown in Table 3.

Table 3: Expected departure time and Travel time

Portions of a Day	Departure Time	Travel Time (minutes)
Morning	06 : 01	18
Midday	10 : 01	23
Afternoon	17 : 41	25
Evening	21 : 41	20

(b) Minimum travel time and departure time for seven days of a week: Suppose the user wants to know what is the minimum travel time and corresponding departure time for all seven days of a week so that the user can plan ahead. The output generated by the model that includes expected departure time and required travel

time from LaGuardia Airport to Midtown East for Seven days is shown in Table 4 as a sample.

Table 4: Expected departure time and Travel time for seven days

Days of Week	Departure Time	Travel Time (minutes)	Part of Day
Sunday	07 : 01	16	Morning
Monday	18 : 41	26	Evening
Tuesday	18 : 41	27	Evening
Wednesday	18 : 41	29	Evening
Thursday	12 : 21	33	Midday
Friday	11 : 41	31	Midday
Saturday	07 : 01	19	Morning

(c) **Within given range, what is the minimum travel time and departure time for seven days of week:** For 08:00 to 14:00, the expected departure time and required travel time from LaGuardia Airport to Midtown East for Seven days, again generated by the built model, is shown in Table 5.

Table 5: Expected departure time and Travel time for seven days (8:00 - 14:00)

Days of Week	Departure Time	Travel Time (minutes)
Sunday	08 : 00	18
Monday	14 : 00	29
Tuesday	14 : 00	31
Wednesday	13 : 00	31
Thursday	12 : 20	33
Friday	11 : 40	31
Saturday	08 : 00	20

5.1.2 Dhaka City

Similar queries for Dhaka city can be performed using the built models which are described below.

For demonstration purpose, we select *Azimpur Bus Stand* as starting point and *Uttara Sector 1* as destination point because these two locations are two extreme opposite corners of the Dhaka city, *Azimpur Bus Stand* being on the extreme south side and *Uttara Sector 1* being on the extreme north side. Thus, anyone travelling from *Azimpur Bus Stand* to *Uttara Sector 1* needs to bisect the city by using a backbone road and overcome extreme traffic conditions of the city.

(a) The expected Departure time and required travel time for this $\langle Azimpur, Uttara \rangle$ source-destination pair for the seven days of a week are shown in Table 6.

Table 6: Expected departure time and Travel time for seven days (*Azimpur to Uttara Sector 1*)

Days of Week	Departure Time	Travel Time (minutes)
Sunday	13 :01	43
Monday	13 :01	43
Tuesday	13 :41	44
Wednesday	8 : 01	47
Thursday	12 :21	45
Friday	8 : 01	35
Saturday	8 : 01	37

(b) For a pair of source-destination, minimum travel time and departure time during the different segments of a given day can be queried from the model. The expected departure time and required travel time from Azimpur to Uttara on Saturday are shown in Table 7.

Table 7: Expected departure time and Travel time for Azimpur to Uttara Sector 1

Parts of Day	Departure Time	Travel Time (minutes)
Morning	07:01	40
Midday	13 :01	46
Afternoon	14 :01	46
Evening	19 : 41	54

(c) Within a given range, minimum travel time and departure time for which travel time is minimum can be shown for seven days of the week. For 08:00 to 14:00, expected departure time and required travel time from *Azimpur Bus Stand to Uttara Sector 1* for Seven days are shown in 8.

Table 8: Expected departure time and Travel time from Azimpur to Uttara Sector 1 (8:00 - 14:00).

Days of Week	Departure Time	Travel Time (minutes)	Part of the Day
Sunday	13:01	43	Midday
Monday	14 : 00	29	Midday
Tuesday	14 : 00	31	Midday
Wednesday	13 : 00	31	Morning
Thursday	12 : 20	33	Midday
Friday	11 : 40	31	Morning
Saturday	08 : 00	20	Morning

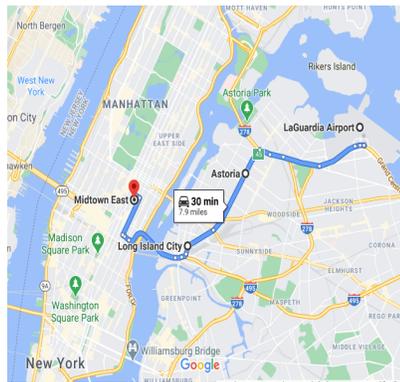
5.2 Least Congested Route

We define the route as having the least travel time among all the routes from a source to a destination for a specific time as Least Congested Route.

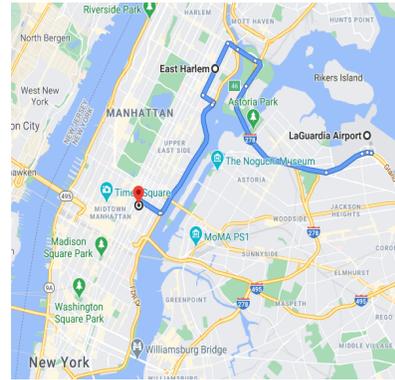
5.2.1 New York City

(a) When a preferred travel time and a source-destination pair are given, the model shows the route which requires minimal travel time among all the routes from the same source to the same destination. A least congested route from LaGuardia Airport to Midtown East on Wednesday, November 4, 2020, at 10:00 AM generated by the model is shown in the fig. 2a. In this route, the model predicts the 33 minutes required to complete the journey.

(b) When source-destination pair with an intermediate location is given, for a given time, the model shows a route among all the routes from source to destination that goes through the given intermediate location for which the expected travel time is minimum. A least congested path from LaGuardia Airport to Midtown East on Wednesday, November 4, 2020, at 10:00 AM that goes through East Harlem is shown in the fig. 2b. The model predicts 44 minutes to complete the journey.



(a) Least congested route (default)



(b) Least congested route with a given mid location

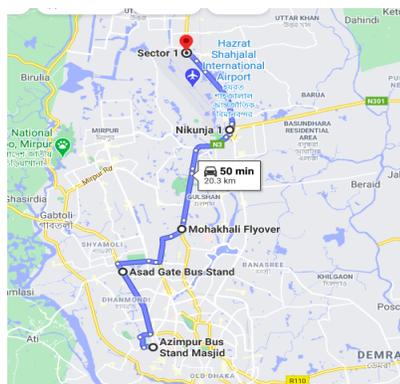
Fig. 2: Least Congested Route

5.2.2 Dhaka City

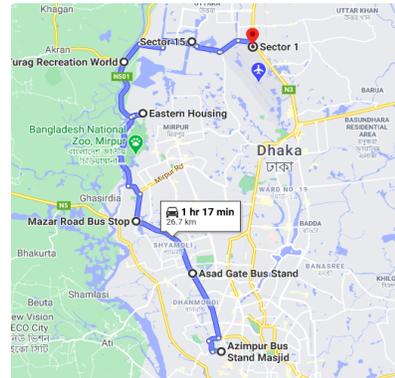
Similar queries for Dhaka city can be performed using the built models which are described below. (a) When a preferred travel time and a source-destination pair are given, the model shows the route which requires minimal travel time among all the

routes from the same source to the same destination. A least congested route from *Azimpur* to *Uttara Sector 1* on Saturday, February, 2021 at 8:30 PM generated by the model is shown in the fig. 3a. In this route, the model predicts 50 minutes required to complete the journey which is close to the prediction by Google Map (shown fig. 3a).

(b) When a source-destination pair with an intermediate location, in this case *Mazar Road*, is given, the model shows a route among all the routes from source to destination going through that intermediate location for which the expected travel time is minimum. A least congested route from *Azimpur* to *Uttara Sector 1* on Saturday, February, 2021 at 8:30 PM going through *Mazar Road* is shown in the fig. 3b. The model predicts 65 minutes to complete the journey which is very close to the prediction generated by the Google Map fig. 3b.



(a) Least congested route (default)



(b) Least congested route with a given mid point as *Mazar Road*

Fig. 3: Least Congested Route For Azimpur to Uttara Sector 1

5.3 Least Time to Travel

5.3.1 New York City

The time when travel time is minimum for a specific source-destination pair. The travel time required from a source to a destination is shown in Figure 4. Travel time is minimum during the night 16:00-22:00 (shown in red in Figure 4a).

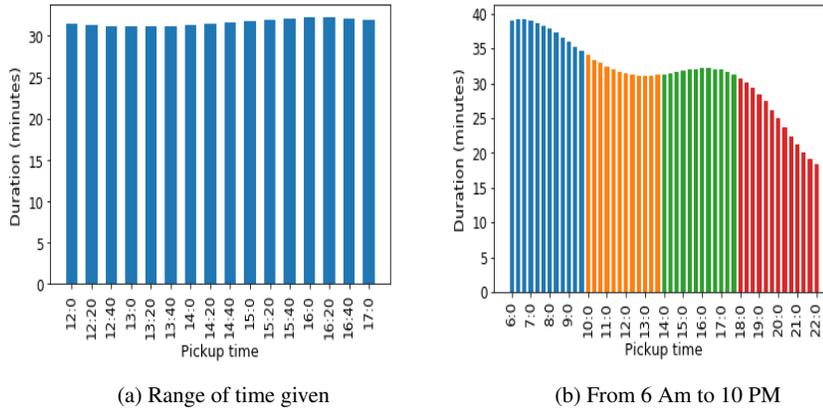


Fig. 4: Travel time comparison in a range of period in New York City

5.3.2 Dhaka City

The travel time required from a *Azimpur Bus Stand* to *Uttara Sector 1* is shown in Figure 5. Travel time is minimum during the early morning (shown in blue in Figure 5a).

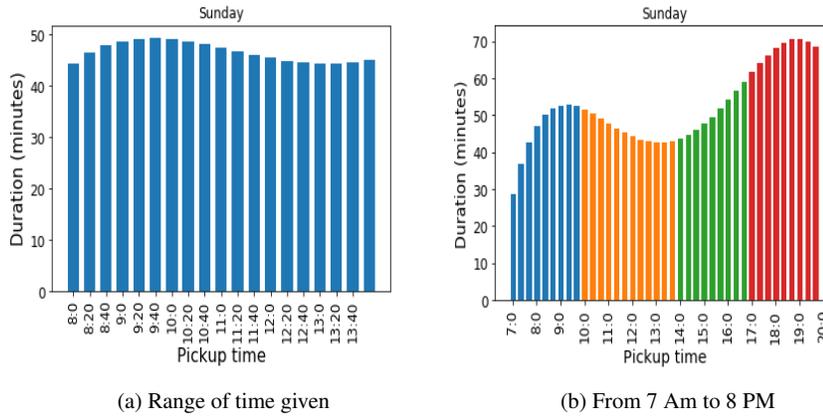


Fig. 5: Travel time comparison in a range of period in Dhaka City

5.4 Socio-cultural Aspects

Interestingly, different socio-cultural behaviour of society such as weekend behaviour, rush hours, nighttime travel habits etc. can be learned from its traffic pat-

tern. Different social behaviours are described in this section for the city of New York and Dhaka City respectively.

- **Traffic Pattern:** As we can see in the Figure 6 the traffic pattern is different from city to city. In the left (Figure 6 (a)) we show the traffic pattern of New York City while in the right (Figure 6 (a)) we show the traffic pattern of the Dhaka city. Observe that there are two pick-hours in Dhaka’s traffic pattern as well as in the New York city’s traffic pattern.

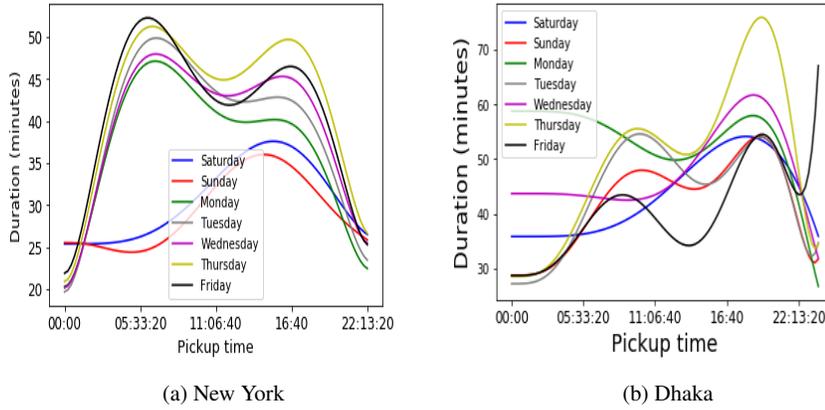


Fig. 6: Traffic pattern on different weekdays

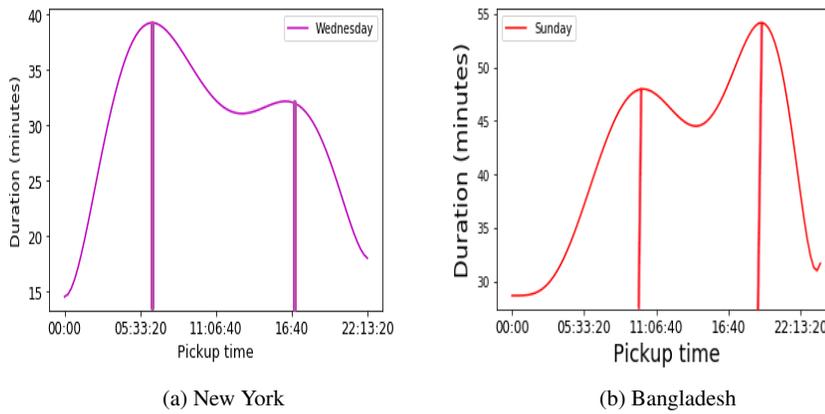


Fig. 7: Traffic pattern of a day

- **Rush Hours** As observed in Figure 7 during the morning and the late evening the traffic density increases. This is as expected because during these two times offices open and close respectively and triggers heavy traffic in the city.

- **Weekend:** We plotted a graph of weekdays relative travel times in Figure 6. We can compare the average traffic density between the different days of the week. As we can see in the left figure the traffic density relatively less on Saturday and Sunday as expected because these two days are weekends in New York City. On the other hand, one can easily see that on Friday traffic density is less in Dhaka city as it is the weekend for this part of the world.
- **Night time travel habit:** One can also observe the nighttime travel habit analysing the Figure 6. We found one common pattern between the two figures. The traffic increases at the night right before the weekends. For example, in the case of New York City, it is Friday night while it is Thursday night for Dhaka city.

6 Conclusions and Future Works

In this work, we showcase a machine learning technique to model road traffics of two populous cities of the world, one chosen from a developed country (the New York City) and the other chosen from a developing country (the Dhaka city). We provide some interesting insights including an impression of their socio-cultural aspects derived from road traffic patterns. In future, we plan to perform some rigorous analysis to assist the government in long term planning and designing flyovers, bridges and roads to improve their experience on road networks.

References

1. V. Agrawal, S. JN, and A. Negi, "Travel-time prediction: A short survey," *Recent Trends in Parallel Computing*, vol. 4, no. 3, 2017.
2. C. Siripanpornchana, S. Panichpapiboon, and P. Chaovalit, "Travel-time prediction with deep learning," in *2016 IEEE Region 10 Conference (TENCON)*, pp. 1859–1862, IEEE, 2016.
3. Y. Duan, Y. Lv, and F.-Y. Wang, "Travel time prediction with lstm neural network," in *2016 IEEE 19th Int. Conf. on Intelligent Transport. Systems (ITSC)*, pp. 1053–1058, IEEE, 2016.
4. A. Gal, A. Mandelbaum, F. Schnitzler, A. Senderovich, and M. Weidlich, "Traveling time prediction in scheduled transportation with journey segments," *Information Systems*, vol. 64, pp. 266–280, 2017.
5. V. Dumbliuskas, V. Grigonis, and A. Barauskas, "Application of google-based data for travel time analysis: Kaunas city case study," *Promet-Traffic & Transportation*, vol. 29, no. 6, 2017.
6. Google, "Google Distance Matrix API." <https://developers.google.com/maps/documentation/distance-matrix/intro>, 2020. [Online; accessed 10/03/2020].
7. S. P. Kumarage, *Use of crowdsourced travel time data in traffic engineering applications*. PhD thesis, 2018.
8. Y. H. Chiu, "Application of crowdsourced travel data in identifying potential opportunities for transportation demand management," Master's thesis, University of Waterloo, 2017.
9. J. A. Deri and J. M. Moura, "Taxi data in new york city: A network perspective," in *2015 49th Asilomar Conference on Signals, Systems and Computers*, pp. 1829–1833, IEEE, 2015.
10. E. P. Dennis, R. Wallace, and B. Reed, "Crowdsourcing transportation systems data," *Michigan Department of Transportation (MDOT) REQ*, no. 1259, 2015.