Ministère de l'Enseignement Supérieur
et de la Recherche Scientifique

Université de Carthage

Ecole Polytechnique de Tunisie

وزارة التعليم العالي و البحث العلمي
جامعة قرطاج
المدرسة التونسية للتقنيات

Project Report

for :

Infrastructure and transport engineering

Simulation and generation of traffic in the city of Chicago, modeling and forecasting speed

---

Elaborated by:

CHOUCHENE Skander

HOUES Naim

MAIRECH Mohamed Amine

MECHERGUI Malek

2nd year engineering students at EPT

Academic Year: 2019 - 2020

Rue Elkhawarezmi BP 743 La Marsa 2078
Tel: 71 774 611 -- 71 774 699 Fax: 71 748 843
Site Web: www.ept.rnu.tn

نهج الخوارزمي ص.ب 743 المرسى 2078
الهاتف: 611 774 71 -- 699 774 71 الفاكس: 843 748 71
موقع الواب: www.ept.rnu.tn

Contents

List of Tables

List of Figures

Introduction

During our Infrastructure and Transportation Engineering course, we learned about many aspects of this domain; mainly traffic modelling (the 4-step traffic model).

Being a group of students who are enthusiastic about artificial intelligence and its real-life solutions, we wanted to take a look on the traffic in a large city and try some AI techniques that can help further our understanding of mobility there. We chose the city of Chicago, the 5th largest north American city. During this project, we will look into Chicago's traffic patterns. We will create models that understand and that can forecast traffic indicators. In the final section, we will focus on a single area of Chicago, where we will study the zone, try to approximately generate traffic and finally apply our model on it. All our work, maps and python notebooks, can be found in our Github repository[4].

1. Overview and visualizing of Chicago traffic

Being a large North American city, Chicago has a very dense traffic. In this section, we will have an overview and visualize the traffic flow throughout time and regions to define the most congested areas.

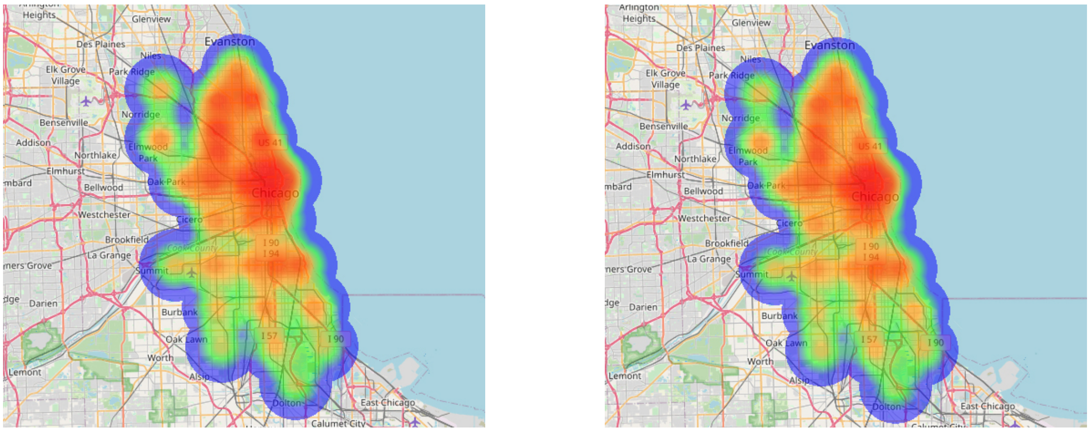
1.1 Tools used

To visualize and deal with the traffic data, we used the python libraries Pandas and Folium [3]. Pandas is the popular DataFrame tool, and Folium is a practical tool to deal with geographic data and to generate dynamic maps.
The Data that we obtained is called 'Chicago Traffic Tracker' and can be found at the 'city of Chicago' website[10]. It contains temporal entries for speed, bus count and number of reads for a certain zone. There are 29 zones altogether.

1.2 Visualizing general speed

In this part, we will have a general overview on the entire city's levels of traffic. We will compare the average speed between Chicago's areas and we will see its evolution with time. This will help us find out which areas are the most prone to extreme congestion at rush hour.
Let's start with a general view of traffic levels. The Package 'Folium' provides a dynamic heat map that makes for a good plotting tool for traffic speed levels. The figures below show the level of traffic speed on a Tuesday morning (7am vs 10am) and a Tuesday

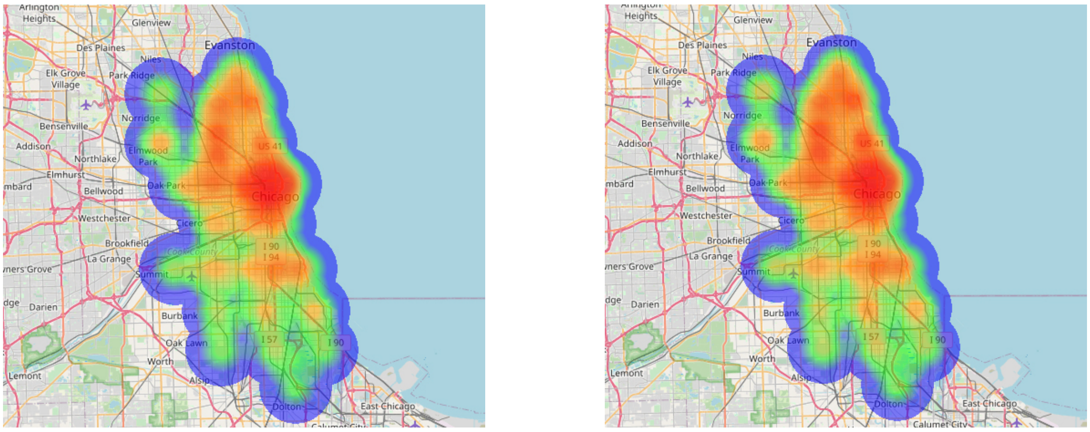
afternoon (4pm vs 6pm).



(a) 7 am

(b) 10 am

Figure 1.1: The level of traffic speed on a Tuesday morning



(a) 4 pm

(b) 6 pm

Figure 1.2: The level of traffic speed on a Tuesday afternoon

Rush hour begins at 7am in the morning and reaches its maximum point between 9 and 11am. In the afternoon, it starts at 4pm and reaches its highest point between 5 and 6pm.

For more animations you can consult our Github repository[6].

To know the most critical areas, we will calculate the median speed at rush hours for all the 29 regions.

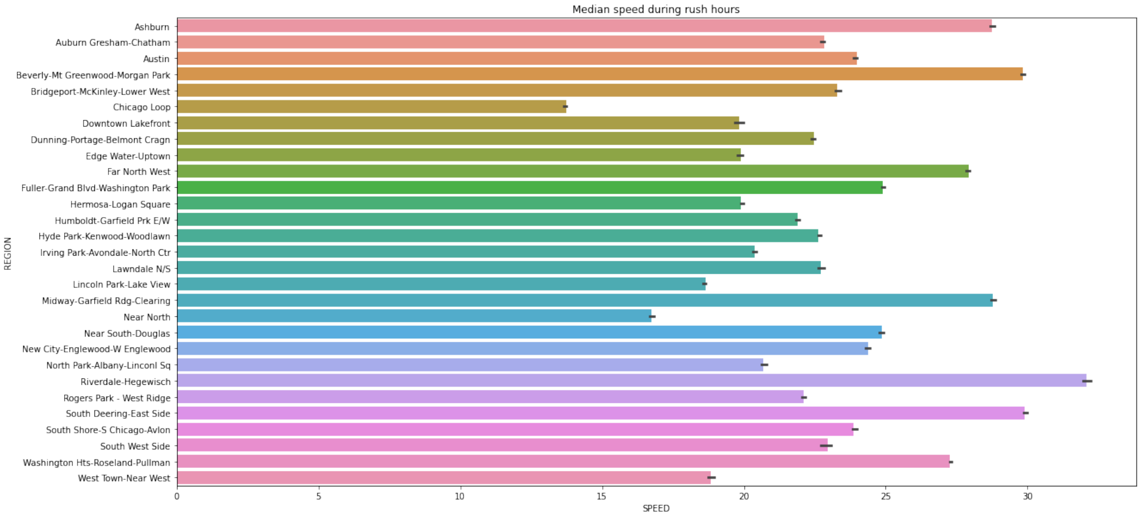


Figure 1.3: Median speed during rush hours[7]

From the graph we understand that the average rush hour speed varies from about 15mph to over 30mph. This graph clears things up: we can see that in rush hours, Chicago Loop is the worst in terms of median speed (less than 15mph). Followed by Near North and Lincoln Park/Lake View. Basically, and logically, the downtown area is the most congested during rush hours, with the area just north of it.

Some uptown areas are still prone to bad traffic as well (Rogers Park, Edge Water).

However, most areas on the south side have good speed levels (over 25mph).

Critical areas:

- Downtown (The Loop, Lakefront).
- North of downtown (Near North, Lincoln Park, Lake View).
- West Town (NW of downtown, bounded by Kennedy Expwy on the east).

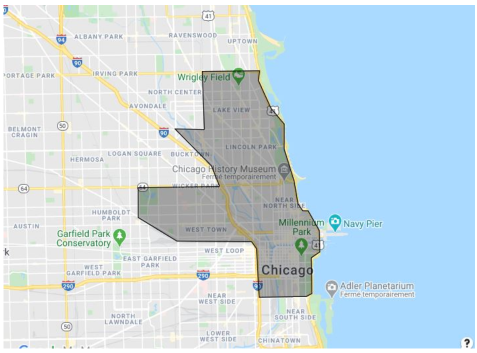


Figure 1.4: Most congested areas[2]

After this overview, we will discuss a tool that we developed to forecast future traffic in Chicago. We will use this tool to model and forecast speed levels in the congested areas.

2. Modeling Part

In this section we will be working on 2 aspects, using the provided dataset we'll build 3 models. The 1st two will be used to model the average speed from existing variables. The third model will be used to forecast the average speed of the different regions for the few upcoming months.
The source code for the models can be found in our Github repository[9].

2.1 First model

The 1st model will be used to predict the average speed in different locations based on the data collected previously, this model will help us detect the effect of the bus count(number of buses detected in a certain area at a certain time of the day) and the number of reads (Number of GPS probes received for estimating the speed) on the average speed. The number of reads is not the vehicle count, however it gives us an idea about the level of traffic. The first model based on Boosting machine learning algorithm (XGboost). We built some new features from the geographical variables (North, South, Est, West) such as the area of the region, length, width, etc.

Figure 2.1: Model 1 architecture

2.2 Second model

In the second model we used deep learning to build a neural network model capable of finding a pattern between the different variables (location, day, month, hour, year...) to predict the average speed given the needed information. In the training of this model we used the same features as the first model but we have used Entity Embedding technique to handle the categorical variables such as day, week_of_day, region_id, month, year, etc.

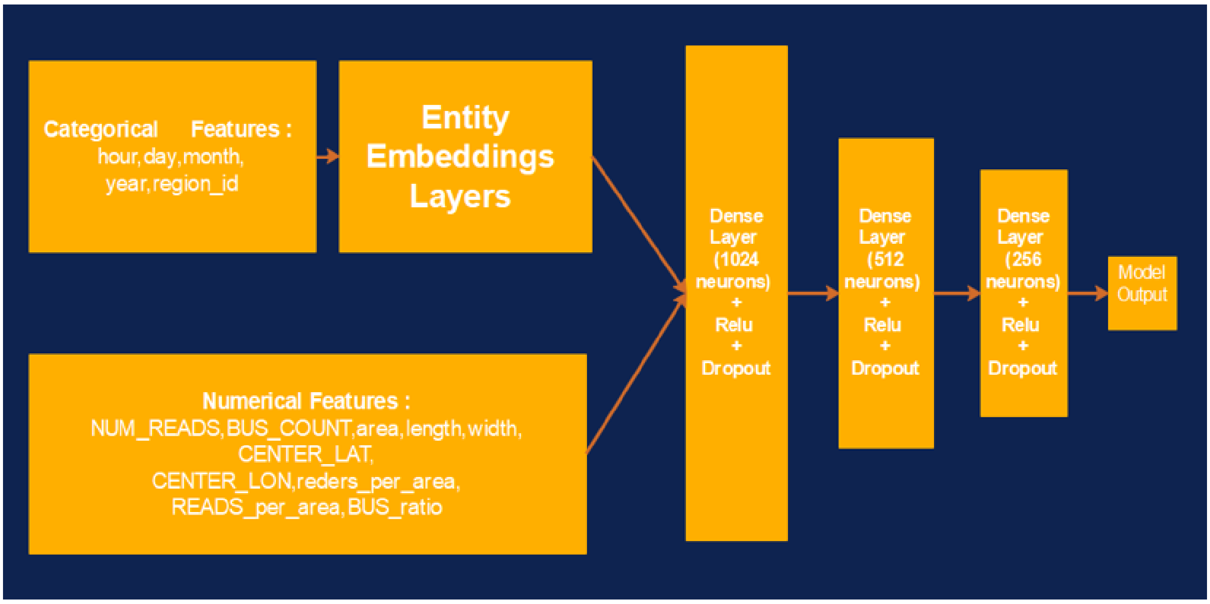


Figure 2.2: Model 2 architecture

2.3 Third model : Future speed prediction

The 3rd model will be mainly used to predict the average speed for future dates based on the previous average speed measured for the N last days. We'll be managing it this time as a time series problem. This model will help us predict the traffic behavior for future dates in order to anticipate the problems that might occur and rearrange the traffic flow and minimize risks of accidents and especially to gain time.

We'll be using Deep Learning this time as well to build a neural network model using the RNNs (Recurrent Neural Networks) and LSTM in particular (Long Short Term Memory). This technique is capable of detecting a pattern through time based on historic data to be able to forecast future values. We chose to work on the 'Chicago loop' location as an example for this model (Section III).

As further enhancement we added to the previous model a Recurrent Neural network layer, the role of this layer is to encode the information of the pervious hours' average speed then we have used this information with the other variables to predict the average speed. the idea that we assume that :

$$avg\_speed(t) = function(avg\_speed(t\text{-}24) \ , \ldots, avg\_speed(t), \ other\ variables\ )$$

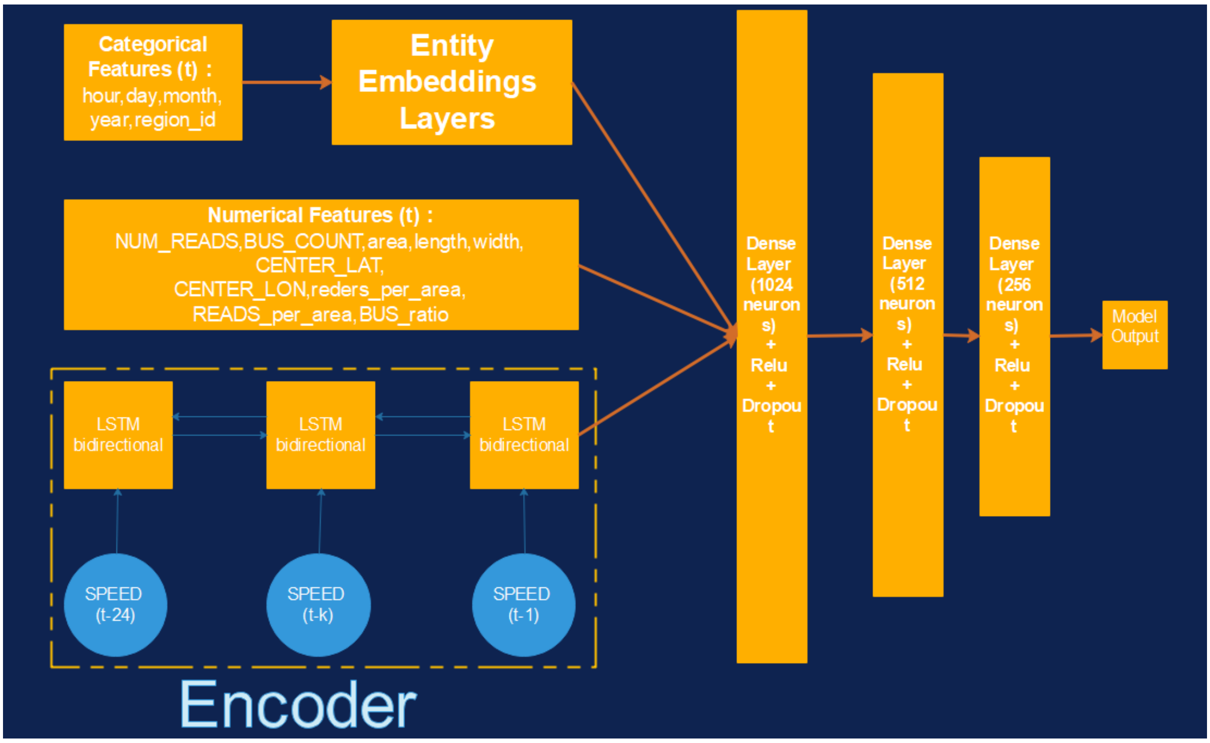


Figure 2.3: Model 3 architecture

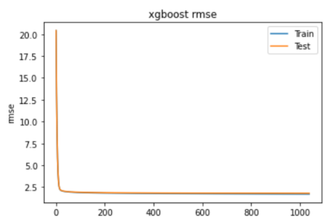
2.4 Results

To evaluate the performance of our model we split the data into training set, cross-validation and test set (train 70% validation 20% and Test set 10% as unseen data) and we find the best hyper-parameters in order to minimize the error on the cross-validation set. For this problem we'll be using the mean square error metric which helps us to penalize large errors that can affect the reliability of the predictions.

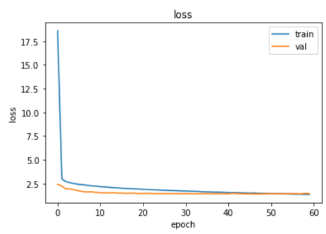
	Training time (min)	Training error	Validation error	Testing error
Model 1	0.463	1.7075	1.8027	1.7649
Model 2	102.34	1.1605	1.1652	1.16422
Model 3	0.315	3.74	3.97	3.95
Model 3 enhanced	236.5	1.1405	1.1330	1.1379

Table 2.1: Training time and errors

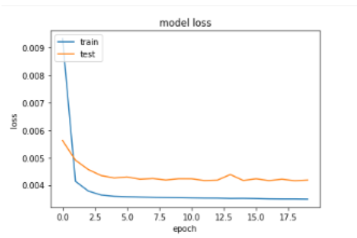
The figures below details the convergence of our models.



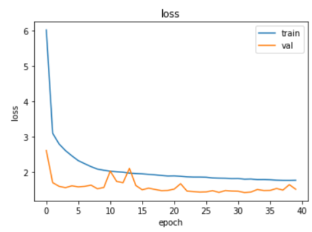
(a) Model 1



(b) Model 2



(c) Model 3



(d) Model 3 enhanced

Figure 2.4: Train vs. validation loss

As an example for the models' robustness we have randomly selected a period and a specified region and predicted the average speed from the selected data.

The example below is the average speed in a region called 'Hermosa-Logan Square'.

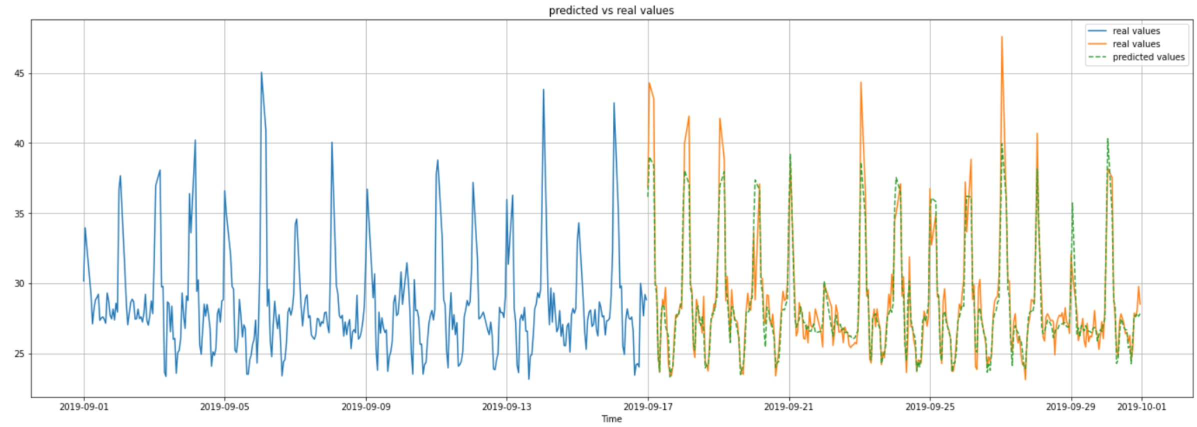


Figure 2.5: Forecasting example for 2 weeks in April

A relatively well fit!

3. Chicago loop

After locating the congested areas and creating models that track speed levels in these areas, now we focus on the downtown area of the city, called Chicago Loop and defined as follows:

- Bounded on the north and west by the Chicago River.
- Bounded on the east by Lake Michigan.
- Bounded on the south by Roosevelt Road.

It's Chicago's commercial business district and the 2nd largest in North America, after Downtown Manhattan. We chose to focus on this area as it's the liveliest place in the city. We will study the existing traffic there, either by speed or by vehicle count.

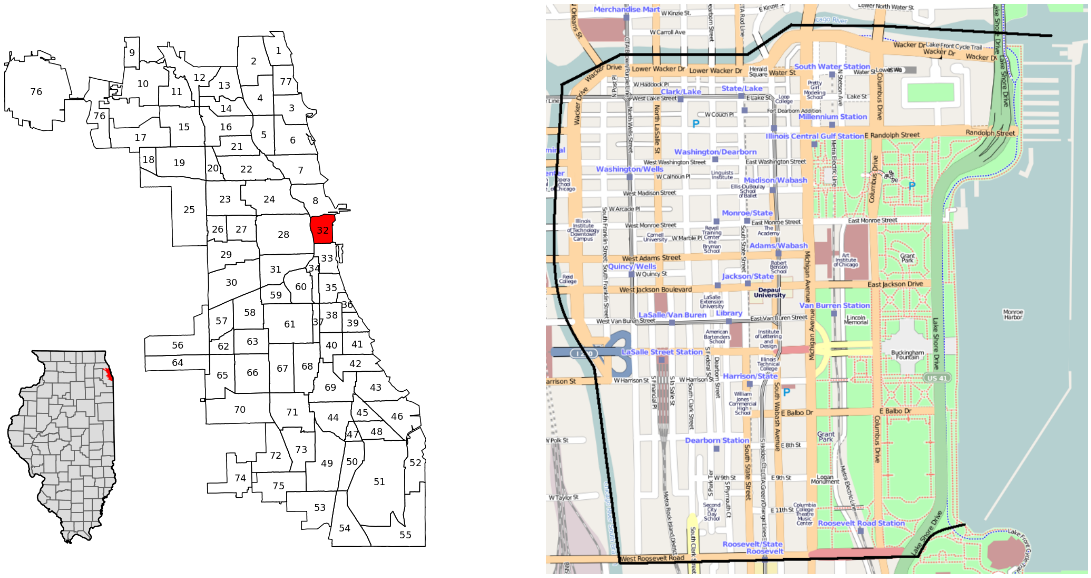


Figure 3.1: Chicago Loop

3.1 Transport network identification

3.1.1 Roads

Inside the area, the main roads are Michigan Avenue (runs parallel to Millennium Park), Congress Parkway (center of the business district) and Wacker Drive (runs with the Chicago river).

The main axis surrounding the area are Lake Shore Drive (on lake Michigan), Roosevelt Road (south of the area) and the Kennedy Expressway which runs parallel to The Loop and can be considered the most imposing road.

3.1.2 Rails and bus lines

There are several metro lines within the area, however they don't affect the traffic as the rails are elevated.

Loop Link[5] is a bus network that operates within the area, it links outer neighborhoods with The Loop and provides transportation within the area. Buses operate mostly on dedicated bus lanes.

3.1.3 Main intersections and interchanges

The main node is the Jane M. Byrne interchange that allows vehicles to enter downtown from the expressway. There are also 20 bridges that link the business district with the northern and western area, passing over the Chicago river; mainly on Roosevelt Road, Congress Parkway, N LaSalle Drive, Michigan Ave and the Outer Drive bridge on Lake Shore Dr (this one however can be considered in another area which is the Lakefront).

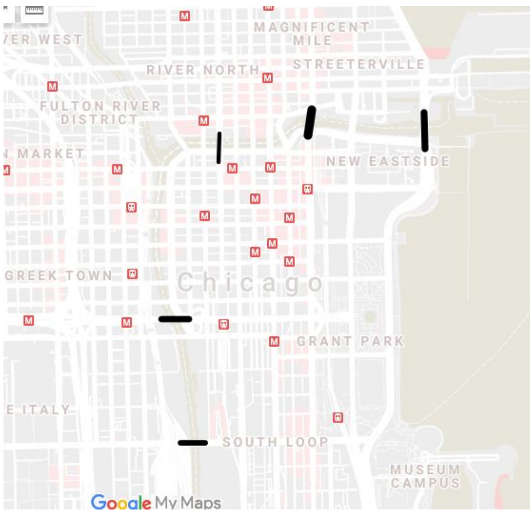


Figure 3.2: Main entrances

3.1.4 Zone delimitation

We can divide the area to 5 zones in terms of urban design and traffic level:

- Central Business District (purple): Contains most of the skyscrapers and is the city's commercial center.
- Southern Loop area (yellow)
- Grant Park (blue)
- Northeast of Kennedy Expwy (red)
- Southeast of Kennedy Expwy (orange)

The last 2 zones are not traditionally part of The Loop, however they're still parts of downtown, have a similar planning design and affect the traffic of the central zone. The main differences between the areas are dictated by density; the upper zones (purple and red) have the most density in terms of people and buildings. The southern areas are less dense, as well as the park area.

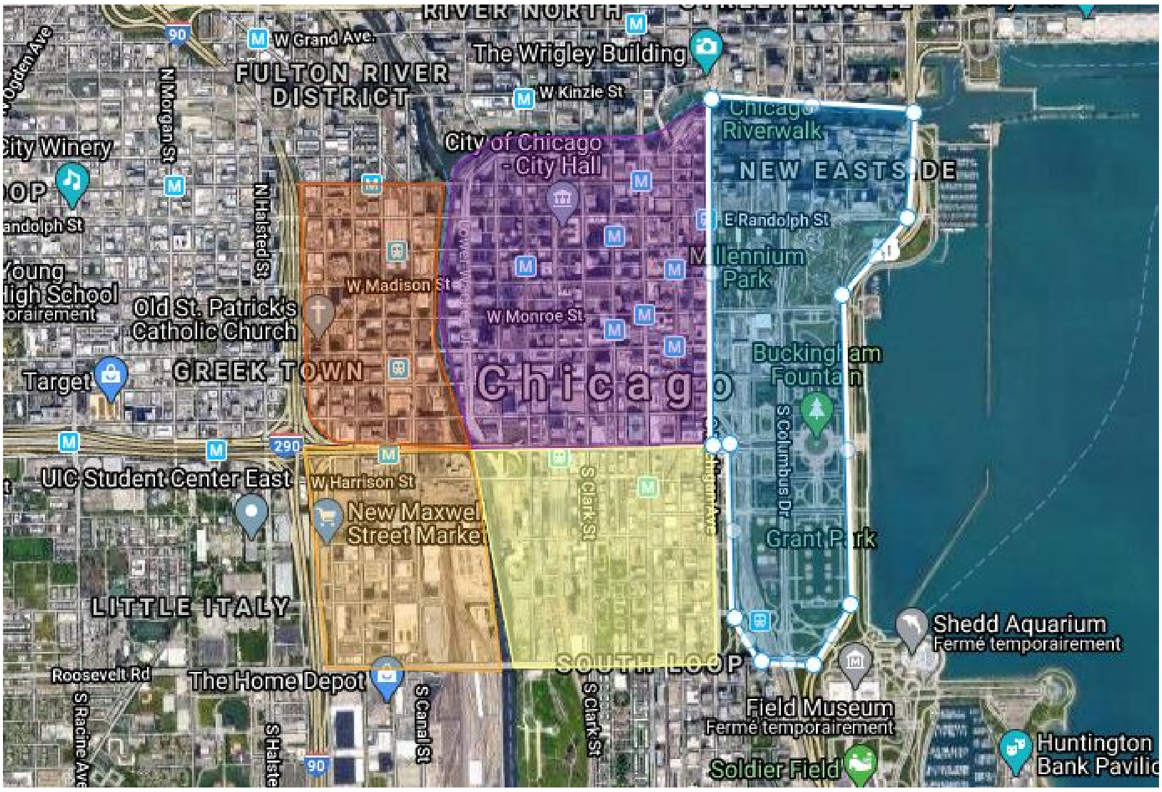


Figure 3.3: Zone delimitation[11]

3.2 Generating traffic

3.2.1 Vehicle count

Due to the fact that our data doesn't have the true number of vehicles but only a number of speed reads, we cannot truly estimate the number of cars passing through each area from this new data. We couldn't also find relevant and recent count data in a well-structured database (the most recent one is from 2006[8]). However, we were able to find a map on ArcGIS[1] with vehicle count that has values as recent as 2014.

The array below shows the number of vehicles going from zone to zone. The information at our disposal does not give us the direction, just the number of vehicles that exist on a certain street. Therefore, we can extract the number of vehicles going between two zones by summing all the vehicles that exist on the boundaries of these two areas, and adding the count difference in a street that traverses both areas.
If two areas are not adjacent or connected by a route, we will not attribute a vehicle count, as vehicles have to cross into an adjacent area to go to the 2nd zone.

Zones	Blue Zone	Purple zone	Yellow zone	Red zone	Orange zone
Blue zone	-	29 000	30 400	-	-
Purple zone	-	-	42 800	44 910	-
Yellow zone	-	-	-	-	34 300
Red zone	-	-	-	-	22 300
Orange zone	-	-	-	-	-

Table 3.1: Vehicle count

Below is the traffic count for the 5 major axis.

Main road	Kennedy Expwy	Roosevelt Road	Congress Pkwy	Michigan Avenue	N LaSalle Drive
Vehicle count (per day)	251 400	27 400	87 700	39 400	8 100

Table 3.2: Vehicle count of the 5 major axis

3.2.2 Visualizing speed

Below is a variation of the speed during a week (April 8th to 14th, 2019).

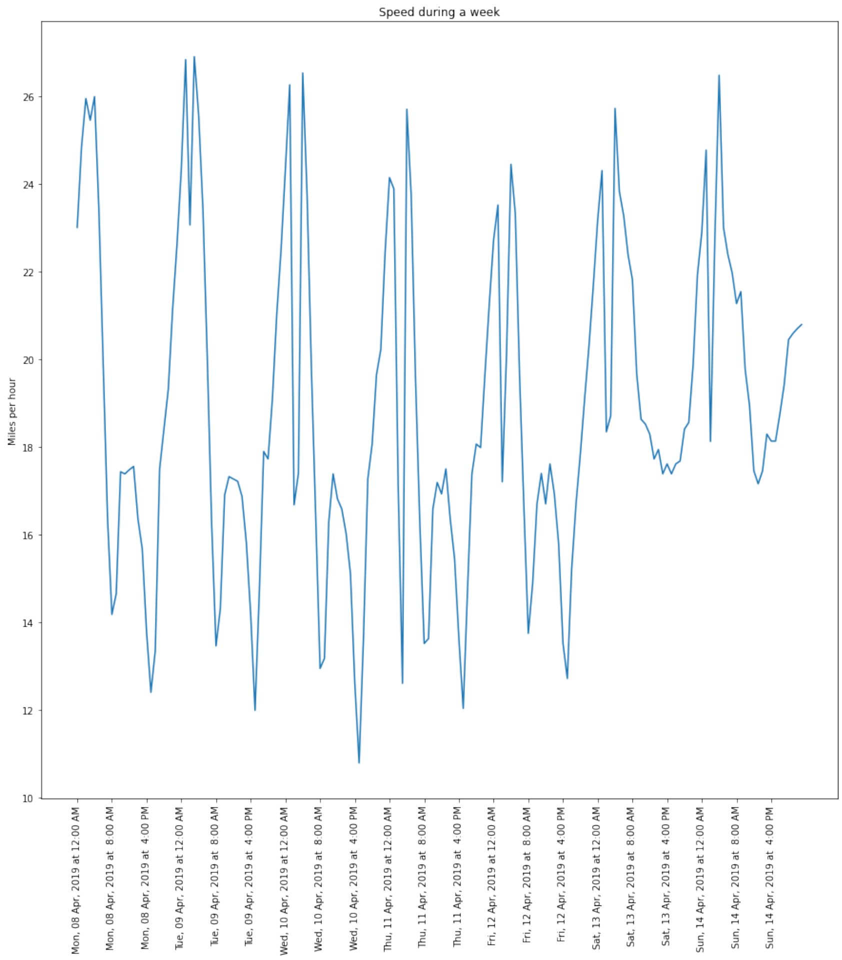


Figure 3.4: Speed during a week

We can see the general tendency of speed dropping at rush hours, especially in the morning.

3.3 Forecasting traffic

Using the DL models that we created, we can model the speed levels for The Loop, and even forecast it for the few upcoming months (This forecast obviously doesn't take in account the current pandemic state). We used the 3rd enhanced model to do this task because it is the best model that we created.

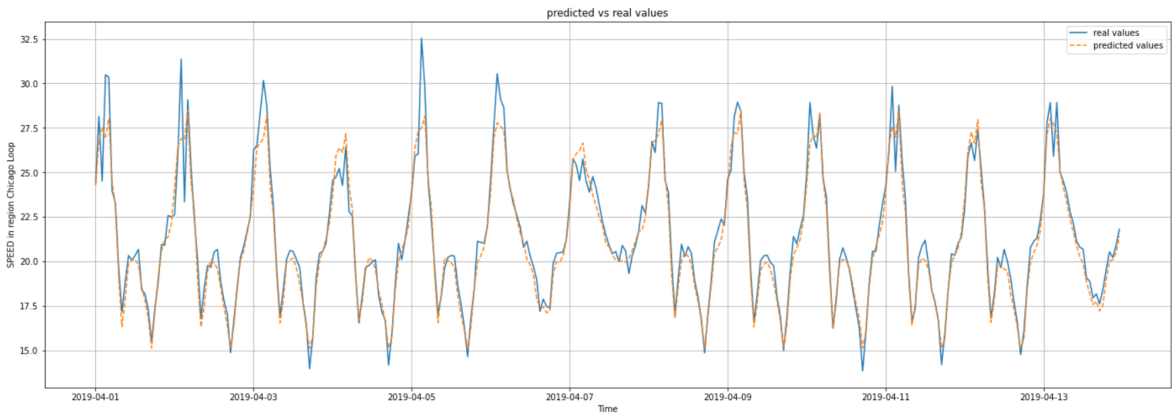


Figure 3.5: Real vs Predicted: Chicago loop average speed from (1/4/2019 to 14/4/2019)

The graph above shows the predicted speed levels laid on top of real speed levels.

The model's splendid accuracy mirrors its strength in predicting speed levels during the daytime. Such a tool can take streams of new real-time data and project new forecasts. These forecasts can alert traffic engineers to potential jams or bottlenecks, and can direct us to the exact congestion time of every day. Having predicted the speed, we can try to tweak some variables in order to see the effect on average speed.

For example, by reducing the number of reads, we get the speed levels below.

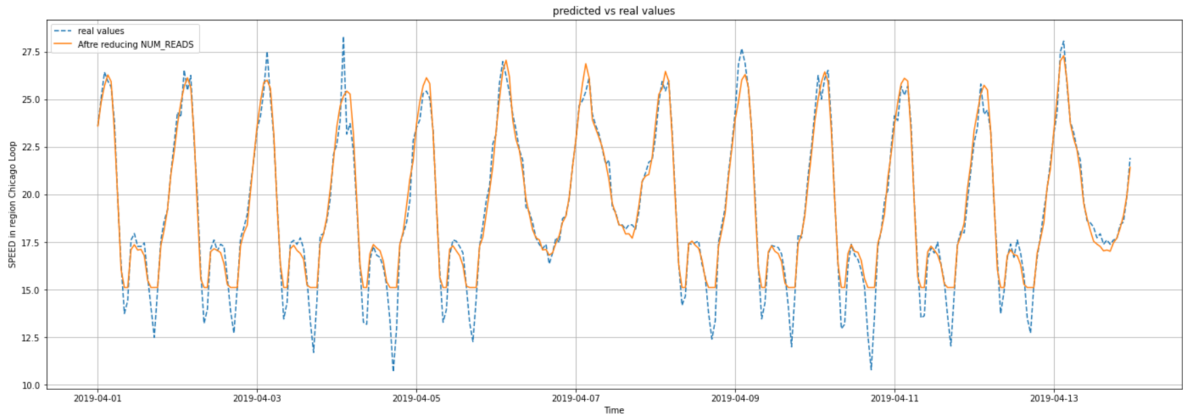


Figure 3.6: Speed levels after reducing num_reads

This graph shows that the minimum speed has increased. This is quite logical. Intuitively, the number of reads is correlated with the vehicle count. And by reducing the number of vehicles there will be less traffic jams! However, it's not that simple and obvious. There needs to be a wider and more extensive research into other qualitative factors that determine traffic jams. This is one of the shortcomings of our project, as qualitative data was scarce during our search for resources. And this didn't help us in our search for solutions.

Nevertheless, even if there are some very serious traffic jams, The Loop is a very organized downtown area, and its behavior is more flexible than other downtown areas in North America such as LA and Washington DC. The most critical area of Chicago is reportedly the interchange between I-90 and I-94 upstate, a few KMs away from The Loop.

Conclusion

During this project, we took a general look at the state of traffic in the city of Chicago. We identified the most congested areas that turned out to be around the downtown area, along with some uptown areas.

Later on, we discussed the AI models that we created from the Chicago traffic tracker dataset. We achieved very good accuracy results and outputted our own forecasts.

Finally, we took a closer look at the downtown area, also known as The Loop. We identified the main traffic elements of this zone (main sub-zones, delimitations…). We went on to generate some aspects of the traffic: vehicle count and speed levels. And finally, we tried our models to forecast speed levels and it achieved good results.

The biggest constraint of this project was the inexistence of a dataset that provides a total and relevant vehicle count. The dataset that we acquired only has speed as a target, however it was the most descriptive of the city's traffic levels. Nevertheless, we have got a good idea on the traffic levels in the city, including critical times and critical zones.

The only shortcoming of our project was the inability to find concrete solutions.

Bibliography

[1] *arcGIS US vehicle count map*. https://www.arcgis.com/home/item.html?id=ced1855778634da6b7

[2] *Congested areas*. https://drive.google.com/open?id=1oFH91rJxruN$_3G4FxaIkgm-w9Oa3I12Nusp = sharing$.

[3] *Folium documentation*. https://python-visualization.github.io/folium/quickstart.html.

[4] *Github repository*. https://github.com/nhoues/Infrastructure-Project.

[5] *Loop link*. https://www.transitchicago.com/looplink/.

[6] *More animated maps*. https://github.com/nhoues/Infrastructure-Project/blob/master/Animation

[7] *Notebook containing the plots*. https://github.com/nhoues/Infrastructure-Project/blob/master/A

[8] *Outdated vehicle count dataset 2006*. https://data.cityofchicago.org/Transportation/Average-Daily-Traffic-Counts/pfsx-4n4m.

[9] *Source code for the models*. https://github.com/nhoues/Infrastructure-Project/tree/master/Noteb

[10] *The dataset that we used*. https://data.cityofchicago.org/dataset/Chicago-Traffic-Tracker-Historical-Congestion-Esti/kf7e-cur8.

[11] *Zone delimitation map*. https://drive.google.com/open?id=15vpYJuB9gSKiHjrIUy4G0pTjctYq-i21usp=sharing.

To find all these links you can visit this google doc :

`https://docs.google.com/document/d/1EfNJ4K8yl1UB43rniFSsqyWbMeIIi9ra7cPtXkwYaKg/edit?usp=sharing`