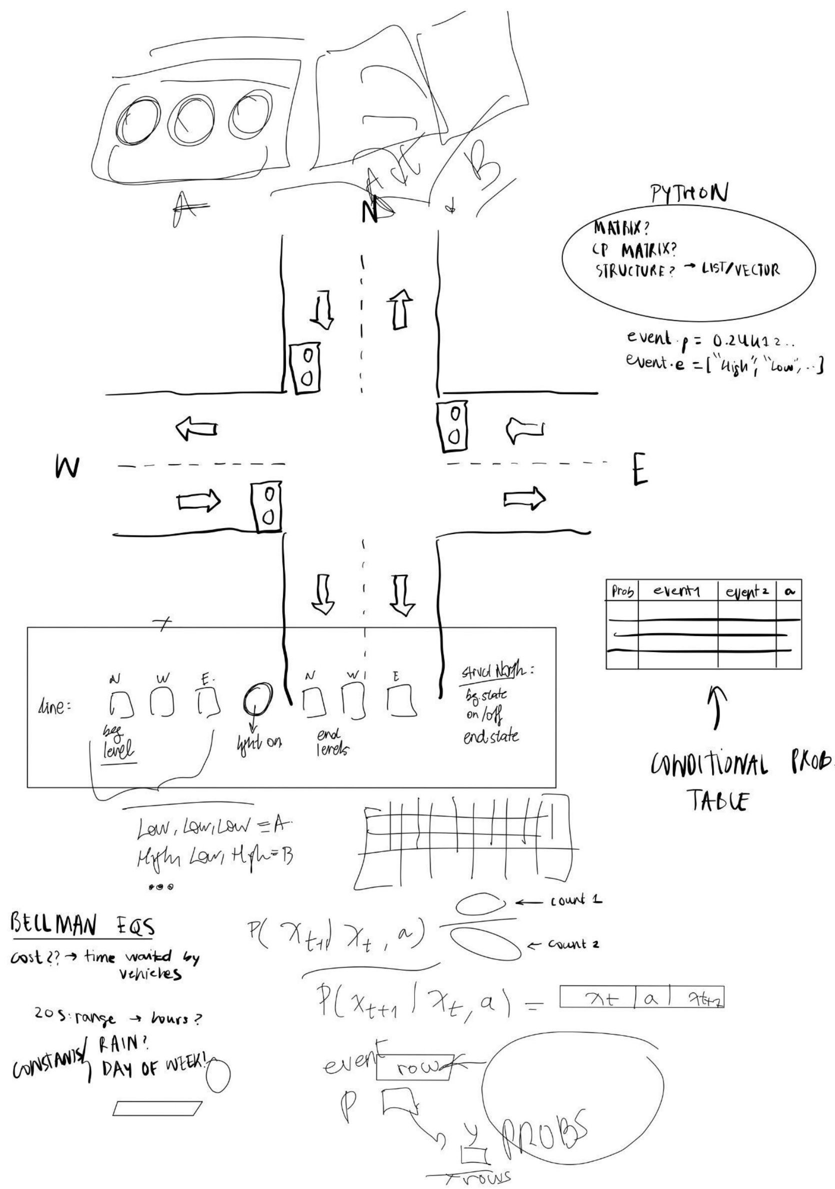
REPORT-AI 2022

Elena de Toledo Hernández & Carlos Carmona Ramos

GROUP 121 - ARTIFICIAL INTELLIGENCE

MAY 2022



INDEX

INTRODUCTION

This practice has put to test all of the knowledge we have acquired during the past few months regarding Artificial Intelligence and has also forced us to become our most creative selves to complete all of the tasks that we are not used to completing in most of the projects we have worked on so far.

Our main focus has been being original, so that both our program and solutions to the different problems that have arised during these weeks were different and interesting.

To develop this practice we have decided to take this situation as an actual real problem of an intersection in Madrid. According to this situation we have decided to gather real data about external facts involved. Apart from the traffic flow of each direction, we have decided to include some variables that may affect the traffic state, such as the day of the week, the month, rain or any special event (accidents, concerts, traffic cuts, etc).

EXECUTIVE SUMMARY

For the proposed problem we tried to approximate the situation to a real case, we took as reference an intersection in Madrid. The key point of our project is the definition of the reward function, as a cost that we want to minimize. We decided to relate the cost with the waiting time accumulated. To approach the most realistic model as possible we decided to involve more external variables to the traffic, such as the weather, and some additional traffic flow variables. We have done some research on the topic to focus on our specific example and collected data from official sources data that are provided in the bibliography. As a result, the cost of an action is the waiting time that it causes to the other two directions blocked by the traffic lights and conditioned by those external factors that we introduced.

Our solution is based on a Markov Decision Process, having as states all the possible combinations of High and Low in the directions North, East and West, being the actions either turning on the traffic light N, E or W.

The methodology applied has been through a python environment and the research we did is based on some official data found in the internet. Our problem resolution can be applied in two modes, a general mode with constant costs, and a specific one that can be adapted to certain conditions and give a more exact result.

We propose the following chronology to this project. Two months for the implementation, design, computation and research, and after that a few months until the final installation. During this period, sensors and traffic lights would be installed and tested in order to gather more data and check its correct functioning. After this period of tests the project would conclude with the final installation.

We provide a unique service due to the quality of the method and the different features our program offers.That is why we have set the budget to a total of 210K for all the work, distributed in: materials(sensors), study and data gathering, intellectual property and additional costs of electricity and others

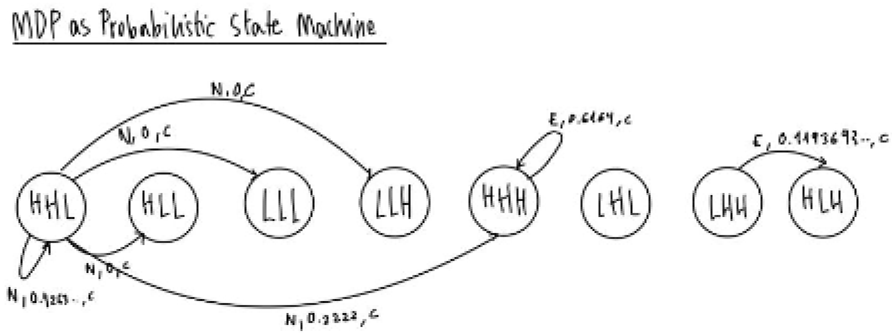
OBJECTIVES

The objectives of this practice have been the following:

- Making a model that accurately represents the situation we have been presented with.
- Study the different events that could alter the traffic flow so that our model represents real life as closely as possible.
- Create a Python program that leads us to solid optimal policies.
- Related to the point above, implementing all the different ideas we had into the program in the form of constants that could be altered easily for different outcomes.
- Create a good cost function that takes into account real life traffic variables.
- Reduce the average wait time for all vehicles while reaching our goal: having a low traffic flow in each direction.

FORMAL DESCRIPTION OF OUR PROJECT

This is the corresponding graphical representation. We didn’t include all the probabilities because that would have been extremely time consuming, but this is the idea behind it (c indicates the cost):



MDP = {S,A,P,R}

For this Markov Decision Process, we have decided to identify 8 states.

These set of states is:

-High, High, High

-High,High,Low

-High,Low,High

-Low,High,High

-High,Low,Low

-Low,High,Low

-Low,Low,High

-Low,Low,Low → GOAL STATE

S={HHH,HHL,HLH,LHH,HLL,LHL,LLH,LLL}

The set of actions will correspond to the different possibilities we have given a state. We can opt to turn on the traffic light of either the North, East or West, taking into account that by turning one on, the two other ones will be turned off.

A = {N,E,W}

The transition function is the function that indicates the probability of having Xt+1 as our next state, knowing the current state Xt and the action performed, A.

T(s,a,s’) = P(Xt+1 = s’ | Xt = s, At = a)

We stored all these probabilities in three vector lists in Python, but you can see here an example:

x_t A x_{t+1} $P(x_{t+1}|A,x_t)=$

```
['High', 'High', 'Low', 'E', 'High', 'High', 'Low'] -- 0.41133004926108374
['High', 'High', 'Low', 'E', 'High', 'Low', 'Low'] -- 0.22413793103448276
['High', 'High', 'Low', 'E', 'Low', 'Low', 'Low'] -- 0.0
['High', 'High', 'Low', 'E', 'Low', 'Low', 'High'] -- 0.0
['High', 'High', 'Low', 'E', 'High', 'High', 'High'] -- 0.24630541871921183
['High', 'High', 'Low', 'E', 'Low', 'High', 'Low'] -- 0.0
['High', 'High', 'Low', 'E', 'Low', 'High', 'High'] -- 0.0
['High', 'High', 'Low', 'E', 'High', 'Low', 'High'] -- 0.11822660098522167
N 1   0
['High', 'Low', 'Low', 'E', 'High', 'High', 'Low'] -- 0.0
['High', 'Low', 'Low', 'E', 'High', 'Low', 'Low'] -- 0.6190476190476191
['High', 'Low', 'Low', 'E', 'Low', 'Low', 'Low'] -- 0.0
['High', 'Low', 'Low', 'E', 'Low', 'Low', 'High'] -- 0.0
['High', 'Low', 'Low', 'E', 'High', 'High', 'High'] -- 0.0
['High', 'Low', 'Low', 'E', 'Low', 'High', 'Low'] -- 0.0
['High', 'Low', 'Low', 'E', 'Low', 'High', 'High'] -- 0.0
['High', 'Low', 'Low', 'E', 'High', 'Low', 'High'] -- 0.38095238095238093
N 2   0
['Low', 'Low', 'Low', 'E', 'High', 'High', 'Low'] -- 0
['Low', 'Low', 'Low', 'E', 'High', 'Low', 'Low'] -- 0
['Low', 'Low', 'Low', 'E', 'Low', 'Low', 'Low'] -- 0
['Low', 'Low', 'Low', 'E', 'Low', 'Low', 'High'] -- 0
['Low', 'Low', 'Low', 'E', 'High', 'High', 'High'] -- 0
['Low', 'Low', 'Low', 'E', 'Low', 'High', 'Low'] -- 0
['Low', 'Low', 'Low', 'E', 'Low', 'High', 'High'] -- 0
['Low', 'Low', 'Low', 'E', 'High', 'Low', 'High'] -- 0
```

The reward function in this type of problem is seen as the cost, and the objective is to take actions which minimize said cost. For us, the cost is the waiting time corresponding to the cars blocked by the light in red. Instead of opting for a fixed cost we decided to take into consideration different factors that make an impact on the traffic. Some of the factor we included are, the day of the week, the month, special events such as accidents, traffic cuts, concerts, etc. Also we included the fact of the weather, mainly if there is rain, since this causes many people to move in cars. All these variables are considered as constants that may be modified and influence the decision of the process. All this facts are considered in a way they are dependent of the flow rate of each road, the traffic is different at each direction and not all are at the same level of traffic density.

Here you can see the part of the code in which you can alter the different values to take into account the different real life possibilities.

Because the statement of the practice asked for a general code, we included the option of MODE, which can be general or specific. In the general mode, the variables r, m, d and e are set to 0, so that they don't alter the cost function.

```
MODE= "general" #CHANGE TO "specific" to alter the value of the variables involved in traffic flow
#HERE WE DEFINE OUR CONSTANTS
RAIN = True #Change here to True or False depending on what you wish to study
#We study the possibility that there's an event in one of the highways. Change as you please
EVENTN = False
EVENTE = False
EVENTW = True
DAYOFWEEK = "Friday" #Write here the day of the week that it's being studied
MONTH = "May" #Write here the month of the year being studied
#AVERAGE FLOW RATE ON EACH DIRECTION
FRATEN = 1
FRATEE = 1.02
FRATEW = 1.03
```

```
cost=[(r2+m2+d2+e2+20*FRATEW)+ (r3+m3+d3+e3+20*FRATEE) +
3,(r1+m1+d1+e1+20*FRATEN)+(r3+m3+d3+e3+20*FRATEE) +
3,(r2+m2+d2+e2+20*FRATEW) + (r1+m1+d1+e1+20*FRATEN) +3]
```

The main concept of our cost function that we want to minimize is the following. The cost of an action is the waiting time of the blocked directions, that is to say, given all the conditions of each road, the cost of an action is the sum of the other two roads' waiting time given their situation. As seen above, apart from the variables mentioned above, we included some constants, such as the number 20, which represents the waiting time of the vehicles, as 20s is the length of one cycle of data. We multiply this waiting time by the variables of FLOWRATE, so that the amount of cars that pass on average through these highways is also taken into account. The constant 3 represents the electricity consumption of each action.

METHODOLOGY

TASK	METHOD	TIME SPENT
Understand the project and apply all concepts to a markov decision process	Read project´s info and relate concepts with the materials worked in class	Half an hour approx
Come up with the concept of reward and cost of our process. Relate it with the real situation and make it realistic.	This was the key to understanding the project. By working together, we concluded that a good approximation to this real problem is considering the time waiting as the cost. We came up with this by associating it with a daily real situation in a city.	One hour
Import the data given and store the information in a data structure	Working with python to process all the information and all the operations of the process resolution. Import data from csv format to a python matrix	One hour
Estimate probabilities given the conditions and store them in a list	We applied a method based on the Bayes theorem, and the probabilities of the intersections are calculated with the Laplace formula. To store the data we created a structure that will be appended to a list	Two hours
Iterate Bellman equations	We followed an example of an exercise given in class to relate it to our case. And apply the formula given a stopping criteria.	30 minutes
Given an expected value once the Bellman equations converge. Estimate the optimal	Iterate one more time and come up with the optimal result.	One hour

policy		
Analyze results and contrast its reliability	We compared our job with some classmates and checked that all the logic behind correspond with the results	One hour
Adapt our program to several situations that affect the cost of the traffic Flow	Once a general version worked, we wanted to improve our code by making it more realistic. We created new variables that affect the traffic and research about their impact on the traffic based on an intersection of the city of Madrid	One hour
Prepare a small slide presentation for the video explanation	We thought that a project presentation with slides would distinguish our job and make it more professional. Using google presentation and exposing it in a monitor	Half an hour
Write report	At this point, all of our ideas were very clear and it went more fluent than the code	Two hours
Record video	We used a monitor to present our slides and we briefly explain our idea with a video of ourselves	45 minutes

RESOURCES USED

To develop our code we decided to use python, as it is a familiar environment for us. To write the code we had to go to the internet to research new concepts we were not taught in programming classes.

These include the module pandas, which we had not used ever before, the function fabs() or the reader from the module csv.

DEVELOPMENT

One of the main tasks that took us more time is code development. We were familiar with the python programming language but writing this code and programming a model to solve the markov decision process was quite a challenge for us.

We started importing the data table to python and reading the csv format. Once we were able to read the file, we store all of its content in a matrix as a python variable. The fact of it being csv format made the reading of the file easier. This matrix enables us to compute all our desired probabilities.

We have made a structure for each probability, such as the fact that the probability is stored and also its numerical value. We created an auxiliary function to compute all the conditional probabilities, this function is based on the Bayes theorem, and the probabilities of the intersections are estimated using the Laplace formula.

In the result below we can check that our probabilities are well computed due to the Law of Total Probability. The result 1 is the sum of the corresponding probabilities.

```
C:\Users\elena\PycharmProjects\EDA\venv\Scripts\python.exe C:/Users/elena/PycharmProjects/IA/finalprojectIA2022.py
['High', 'High', 'Low', 'N', 'High', 'High', 'Low']
['High', 'High', 'Low', 'N', 'High', 'Low', 'Low']
['High', 'High', 'Low', 'N', 'Low', 'Low', 'Low']
['High', 'High', 'Low', 'N', 'Low', 'Low', 'High']
['High', 'High', 'Low', 'N', 'High', 'High', 'High']
['High', 'High', 'Low', 'N', 'Low', 'High', 'Low']
['High', 'High', 'Low', 'N', 'Low', 'High', 'High']
['High', 'High', 'Low', 'N', 'High', 'Low', 'High']
N 1   1.0
['High', 'Low', 'Low', 'N', 'High', 'High', 'Low']
['High', 'Low', 'Low', 'N', 'High', 'Low', 'Low']
['High', 'Low', 'Low', 'N', 'Low', 'Low', 'Low']
['High', 'Low', 'Low', 'N', 'Low', 'Low', 'High']
['High', 'Low', 'Low', 'N', 'High', 'High', 'High']
['High', 'Low', 'Low', 'N', 'Low', 'High', 'Low']
['High', 'Low', 'Low', 'N', 'Low', 'High', 'High']
['High', 'Low', 'Low', 'N', 'High', 'Low', 'High']
N 2   1.0
['Low', 'Low', 'Low', 'N', 'High', 'High', 'Low']
['Low', 'Low', 'Low', 'N', 'High', 'Low', 'Low']
['Low', 'Low', 'Low', 'N', 'Low', 'Low', 'Low']
['Low', 'Low', 'Low', 'N', 'Low', 'Low', 'High']
['Low', 'Low', 'Low', 'N', 'High', 'High', 'High']
['Low', 'Low', 'Low', 'N', 'Low', 'High', 'Low']
['Low', 'Low', 'Low', 'N', 'Low', 'High', 'High']
['Low', 'Low', 'Low', 'N', 'High', 'Low', 'High']
```

Once we have all the conditional probabilities stored, we proceed to iterate the Bellman equations. To control this loop we establish a tolerance for convergence and if it is not the case we have decided a stopping criteria to set a maximum number of iterations. To decide which number will be the maximum iteration we have considered making at least 200 iterations, that is a significant amount given the data provided. In most of the cases the solution will converge at early iterations.

The resulting expected values of the Bellman equations are used to estimate the optimal policies of each state. With the formula we estimate the better option and assign the corresponding optimal action to the state.

After many errors and versions of the code, once finished we contrasted our results with some of our classmates and shared ideas and different views about the project. This was helpful for us and we realized how subjective the interpretations of the projects are.

To conclude our code, we included some constant variables that can be adapted to a determined situation. These constants are external factors that affect the traffic flow. As can be the day of the week, a determined month, special events or the existence of rain. All these facts are involved in the calculation of the cost. As explained, we considered the cost of its action as the time waiting of the directions blocked. To be able to relate the waiting time with a priority, we consider the flow rate of each direction, since each of them has its own rate of traffic and they are not comparable in terms of traffic.

To make the code as realistic as possible, and considering the situation proposed as a real intersection of the city of Madrid, we have made some research to develop our conditions and their effect on the traffic. We researched about the months of most traffic in Madrid, the flow given a certain day of the week, the amount of accidents and events throughout the year.

The table below is one example of an official source that we have collected, and is one of the facts we have taken into account. In the bibliography we expose some of our sources.

	A	B	C	D	E	F	G	H	I	J
1	Año	Mes	Conjunto	Interior 1er cint	En el 1er cintu	Entre 1er y 2º cintu	En el 2er cintu	Entre 2º cinturón y	M-30	Entre M-30 y M-4
2	2021	Enero - 12	1.256.393	74.730	156.336	245.694	124.815	284.763	121.532	229.084
3	2021	Febrero -10	1.679.471	87.692	216.794	316.431	167.985	391.636	175.112	299.229
4	2021	Marzo -7	1.759.087	91.020	224.400	336.439	177.010	416.263	172.148	315.432
5	2021	Abril -8	1.792.404	93.416	226.234	344.172	180.541	423.275	181.425	317.344
6	2021	Mayo -1	2.045.884	119.160	266.690	389.193	224.524	460.850	204.013	353.684
7	2021	Junio -6	1.865.638	95.163	235.135	361.688	188.613	437.599	181.476	338.675
8	2021	Julio -9	1.745.562	98.560	223.691	348.275	168.194	392.413	172.713	317.407
9	2021	Agosto -11	1.353.794	92.286	175.129	259.818	137.151	286.719	148.147	234.933
10	2021	Septiembre -5	1.956.426	112.192	256.995	370.634	209.856	445.432	202.155	332.095
11	2021	Octubre -2	2.040.332	115.608	264.641	383.211	224.842	460.117	218.328	345.557
12	2021	Noviembre -3	2.033.947	114.354	264.060	382.801	223.398	457.277	221.280	342.587
13	2021	Diciembre -4	2.000.329	111.172	257.406	377.905	225.869	456.919	200.787	342.873
14										
15										

Once we wrote the code, we started with the report. At this point our idea and interpretation of the project was even more clear. The development of the report was more fluent since we had a more clear idea of our model to solve this problem.

To present the executive summary in the video we decided to make a slides google presentation as support to make it more visual and attractive to the audience.

RESULTS

Q1. The input data does not include any cases where the starting situation was low traffic level in all three directions. Is this normal? If you had this kind of data, what would happen or what should we have done?

This is since the state (Low, Low, Low) is the goal state and it is an ideal situation, what usually happens in real life is that this ideal condition is never fully satisfied, that is why the objective is to reach an optimal solution that approaches our goal in the most efficient way.

In the case this data were collected, the probability of the goal condition given any other state would be ignored in the Bellman equation as it would be multiplied by zero (knowing that it is the goal state). And the probability of reaching any other state given the ideal goal is null since, once reached that condition the problem resolution is solved.

Q2. The statement does not say anything about the cost of the actions. What reasonable assumption could we make?

From our point of view, we should see the cost of each action as the average waiting time for each car in the highways that are not being freed. This is why in our project we have taken into account many variables, like the possible weather conditions, events or even the day of the week, which are all factors that affect the traffic flow.

We understand that a reasonable assumption would have been to take a constant as the cost of every action and argue that the cost must be the same for all of them, but we wanted to give another point of view to distinguish our work and make it more realistic.

Q3. What are the expected values of the states? Provide the values to six decimal places. The precision must be greater than one thousand.

Having set the constant MODE to “general”, that is, not taking into account the specifications explained above, these are the expected values:

Our precision is 10^-6:

V(HHL) = 101.4665856

V(HLL) = 72.989011

V(LLL) = 0

V(LLH) = 129.149442

V(HHH)= 280.291855

V(LHL)=120.104235

V(LHH)= 185.504221

V(HLH)=112.415997

Q4.What is the optimal policy?

Setting as above the constant MODE to general, this is the optimal policy:

```
Policy for H H L -> E
Policy for H L L -> N
Policy for L L L -> Goal
Policy for L L H -> W
Policy for H H H -> E
Policy for L H L -> E
Policy for L H H -> E
Policy for H L H -> W
```

We wanted to include another example setting the constant MODE to specific, so that it could be seen how it affects the optimal policies:

```
#-------MAIN PROGRAM--------
# open file in read mode

MODE= "specific" #CHANGE TO "specific" to alter the value of the variables involved in traffic flow
#HERE WE DEFINE OUR CONSTANTS
RAIN = True #Change here to True or False depending on what you wish to study
#We study the possibility that there's an event in one of the highways. Change as you please
EVENTN = False
EVENTE = False
EVENTW = True
DAYOFWEEK = "Friday" #Write here the day of the week that it's being studied
MONTH = "May" #Write here the month of the year being studied
```

```
Policy for H H L -> N
Policy for H L L -> N
Policy for L L L -> Goal
Policy for L L H -> W
Policy for H H H -> N
Policy for L H L -> E
Policy for L H H -> E
Policy for H L H -> N
```

Q5. If we had also had incoming traffic from the South and for each direction we had measured traffic at 5 levels instead of 2, what would have changed in the problem?

The set of states instead of having size 8 = 2^3 would have had size 625 = 5^4.

We assume that adding another highway would also mean adding one more traffic light , so we suppose the set of actions would have also grown in size and become A = {N,E,W,S}.

Also, the transition function would involve more variables and the conditional probability tables would be bigger.

CONCLUSIONS

As for the conclusions of the project, we have several facts to stand out. First of all we would like to underline the difficulty of the code, mainly due to the big amount of data we had to deal with and how to focus the code in order to have effective data structure and develop the algorithms as best as possible. This caused us many problems since it is not easy to deal with these amounts of data.

The problems encountered while programming were basically solved with hours of work and some research on the internet. The development of the report was more entertaining, since we had a clear idea of our model and it was just expressing ourselves in an efficient way that would highlight and distinguish our job.

The results of this particular project in a general case with a unique and equal cost to each action, was shocking for us in some details we did not expect. All the outcomes were more or less expected, however, in situations of conflict, we can not deduce which is the most optimal action and the resulting policy follows a criteria based on data. In the case we specify the conditions we created, we can check that those variables affect the result and influence the problem resolution. We are satisfied with this variant that we created as we think it approaches the problem to a real situation.

With this practice, we have learned in a practical way the purpose of artificial intelligence and how it is applied to real situations. We consider a key factor on how to model the problem in order to develop a certain decision process, depending on the view of the problem and above all, the reward, the objective, it will result with an optimal and logical solution that would focus on that goal.

BUDGET

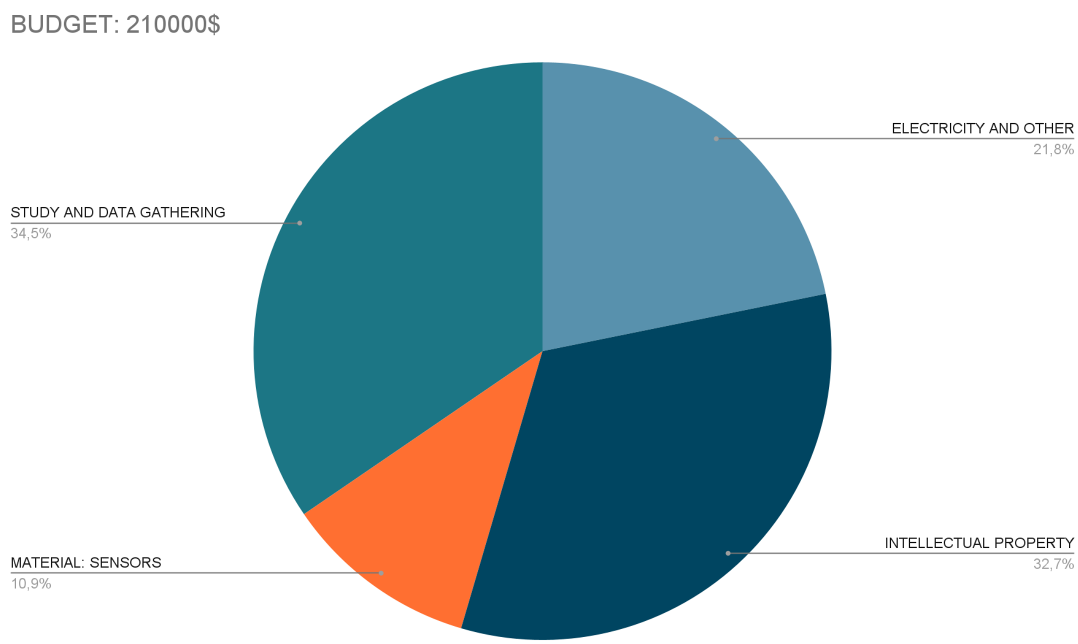
We have divided the budget of this project into 4 categories.

Firstly, the study and data gathering. This has been the greatest focus of our project (34.5% of budget). We have obtained the data from the sensors that measure the level of traffic in each direction, but also, we have had to understand how the different variables we implemented in our project, like the rain, the month or even the day of the week affect traffic. Therefore, much of the budget has been destined for research purposes to make our model as solid as possible.

We have also reserved a great percentage (32,7%), to the intellectual property. In order to be able to develop this model, we have had to acquire different knowledge throughout many years, and that should be taken into account and compensated. Also, we believe we have ended up with a very complete and original proposal, very distinguished from the rest. This, without doubt, adds value to our project.

21,8% has been destined to electricity and other resources, like for example, the memory it has taken in our computers, the time invested and CPU utilization.

Finally, the different sensors used to analyze traffic or weather conditions should also be added to this budget.



It's because of all this that we are asking for a budget of 21000$.

STUDY AND DATA GATHERING: 72450$

ELECTRICITY AND OTHERS: 45780$

MATERIAL : SENSORS AND TRAFFIC LIGHTS 22890$

INTELLECTUAL PROPERTY: 68670$

BIBLIOGRAPHY

"Portal De Datos Abiertos Del Ayuntamiento De Madrid." *Tráfico. Intensidad Media En Días Laborables*, Tráfico. Intensidad media en días laborables - Portal de datos abiertos del Ayuntamiento de Madrid

Departamento Tecnologías Del Tráfico Informe Hora Punta - Madrid.es. https://www.madrid.es/UnidadesDescentralizadas/UDCMovilidadTransportes/Quizas/Informes_CGM/ficheros/Mayo_2022/Informe_diario_CGM_20220513.pdf