2019

Cycling in Mexico City: ECOBICI as the solution to mobility.

#Pytheam#

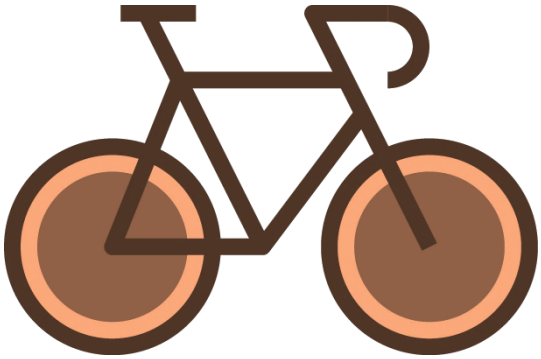
<Blanca Chavarría>

<Lizette González>

<Martha Meses>

<Ricardo Moreno>

<Diana Pardo>



INDEX

1. PROJECT DEFINITION

A. BRAINSTORMING

The definition of a Project can be very challenging, particularly when the team is formed with such a wide variety of formations, beliefs and opinions. In order to get a better picture of our interests and define the best possible Project for a lot of us, we developed a mind map, with the Help of a tool called "MindMeister".

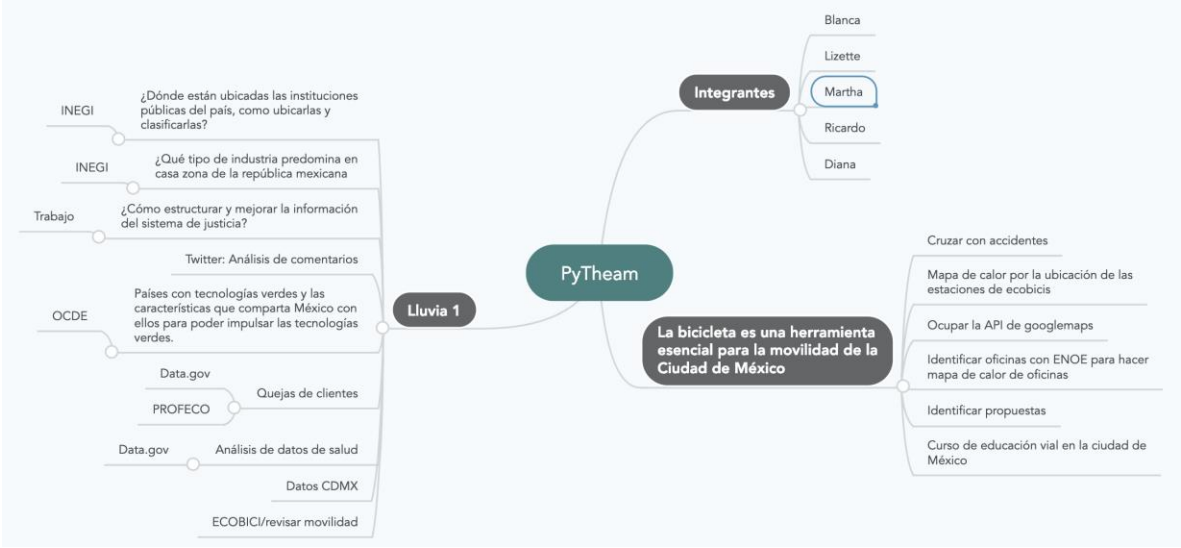


Image 1. Brainstorming result.

As a result, a common interest surfaced: *Ecobici and its impact on mobility.*

B. DATA SEARCH: IS THE PROJECT POSSIBLE?

Once the Project outline was defined, it was needed to stablish if there was enough information to develop an extensive research. During the exploration of datasets, it was determined the Project could be reasonably developed, since several sets of information for Mexican's bicycle-sharing system (ECOBICI) were available, both in csv format and trough an API.

C. PROJECT DESCRIPTION:

<*The scope*>

Assess whether Mexican's bicycle-sharing system (ECOBICI) is the solution for mobility for Mexico City.

<*Project Description/Outline*>

Establish whether Mexican's bicycle-sharing system (ECOBICI) is the solution for mobility for Mexico City, through the analysis of several sources of information.

<Our hypotheses>

$$H_0 = ECOBICI\ is\ a\ mobility\ solution\ for\ Mexico\ City$$

$$H_1 = ECOBICI\ is\ NOT\ a\ mobility\ solution\ for\ Mexico\ City$$

<Research Questions to Answer>

// Is ECOBICI and alternative to move around the city through efficacy?

[] Serve area coverage

[] Operating indicators

// Is ECOBICI complementary to the massive public transportation system network?

[] At least one ECOBICI station near each public transportation stop

// Is ECOBICI available for people in the city?

[] Number of bicycles and docks

[] Availability in peak hours (6 AM to 10 AM, 12 PM to 4 PM and 6PM to 10 PM)

// Is ECOBICI accessible?

[] Subscription and payment mechanisms

[] Accessibility by low income users

< Data Sets to be Used>

IN CDMX		
SOURCE	**DATABASES**	**LINK**
ECOBICI	Datos Abiertos, Estadísticas	https://www.ecobici.cdmx.gob.mx/es
DATOS ABIERTOS (CDMX)	Cicloestaciones	https://datos.cdmx.gob.mx/explore/dataset/estaciones-de-ecobici/export/
INEGI	Directorio Estadístico de Unidades Económicas	https://www.inegi.org.mx/app/mapa/denue/
DATOS ABIERTOS (CDMX	Conteo ciclista	https://datos.cdmx.gob.mx/explore/dataset/estudio-de-conteo-ciclista-2018/information/

AS A COMPARISSON-OTHER COUNTRIES		
SOURCE	**DATABASES**	**LINK**
CITI BIKE NYC	Citi Bike System Data | Citi Bike NYC	https://www.citibikenyc.com/system-data
ECOBICI-BUENOS AIRES	Recorridos realizados	https://datos.cdmx.gob.mx/explore/dataset/estaciones-de-ecobici/export/

2. PROJECT DEVELOPMENT

A. ECOBICI: A BRIEFING

< *History, success and growth*>

ECOBICI started operations in February 2010 with 84 bicycle stations and 1,200 bicycles.

In 8 years, the demand has boosted the growth of the system with 480 bicycle stations and more than 6,800 bicycles, of which 28 stations and 340 bicycles are part of the new system of electric bicycles.

ECOBICI currently has more than 170 thousand active registered users and the service is available in 55 neighborhoods in Mexico City[1].

B. ANSWERING THE RESEARCH QUESTIONS

What is mobility? According to the Oxford Dictionary, mobility is the ability to move or be moved freely and easily, other studies and organizations give this term the extension of accessibility, comprising it as they see it as the final goal to mobility. Accesibility is understood as the capacity to access different areas of a city, an obvious peculiarity to Mexico City, where ECOBICI is concentrated into three municipalities.

< *Is ECOBICI and alternative to move around the city through efficacy?*>

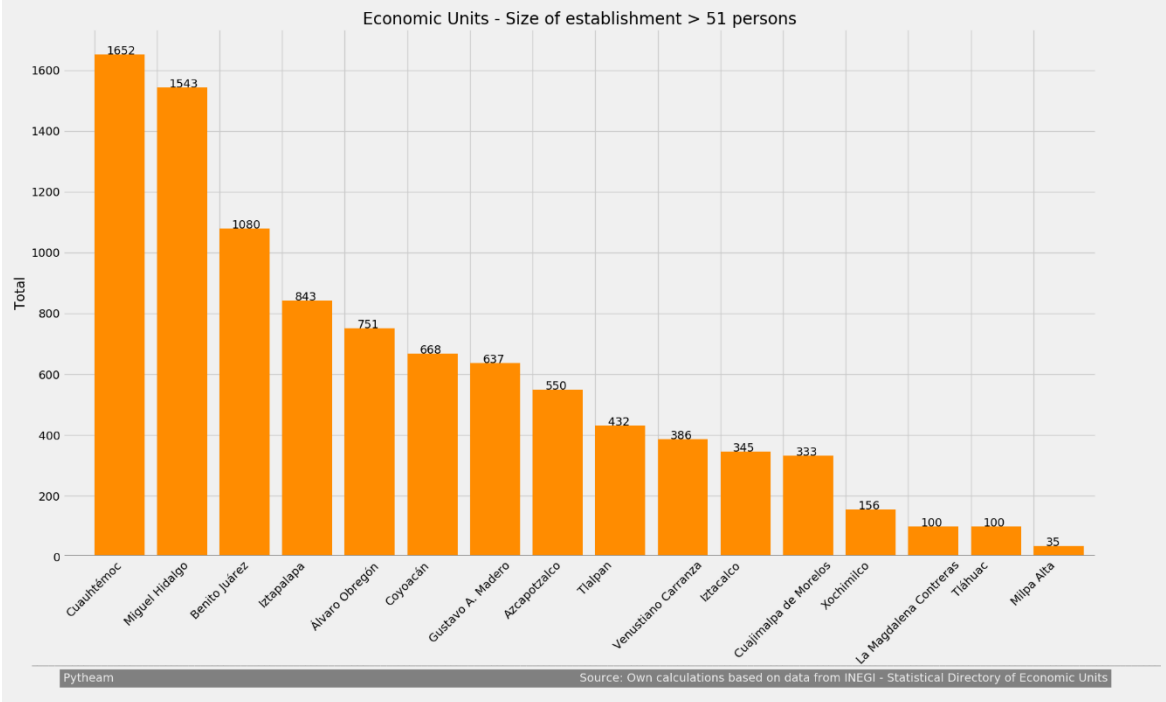
Municipality	Ecobici stations
Benito Juárez	168
Cuauhtémoc	199
Miguel Hidalgo	113
Total	**480**

Table 1. *Ecobici stations by municipality*

[1] Source: https://www.ecobici.cdmx.gob.mx/es/informacion-del-servicio/que-es-ecobici

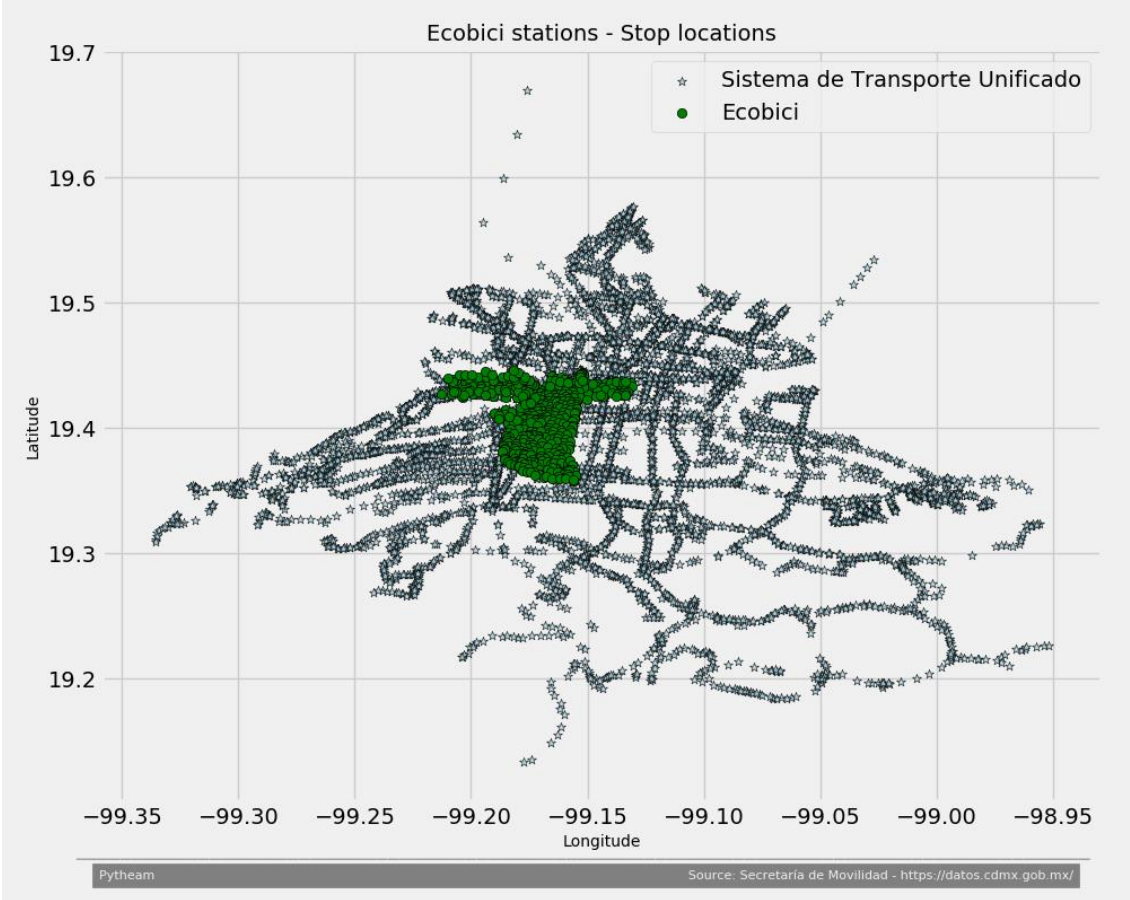
Statistical Directory of Economic Units-DNUE			
Municipality	**Economic units**	**Minimum of occupied persons**	**Maximum of occupied persons (at least)**
Cuauhtémoc	1,652	201,002	304,890
Miguel Hidalgo	1,543	183,593	285,768
Benito Juárez	1,080	123,530	195,974
Iztapalapa	843	92,843	149,626
Álvaro Obregón	751	91,051	141,062
Coyoacán	668	82,318	126,878
Gustavo A. Madero	637	68,187	110,768
Azcapotzalco	550	69,950	104,988
Tlalpan	432	52,932	79,004
Venustiano Carranza	386	48,286	71,984
Iztacalco	345	33,695	58,160
Cuajimalpa de Morelos	333	42,233	62,458
Xochimilco	156	16,506	26,914
La Magdalena Contreras	100	11,400	17,252
Tláhuac	100	10,400	17,388
Milpa Alta	35	3,285	5,760

Table 2. *Statistical Directory of Economic Units. INEGI*



Graph 1. *Economic units by municipality INEGI*

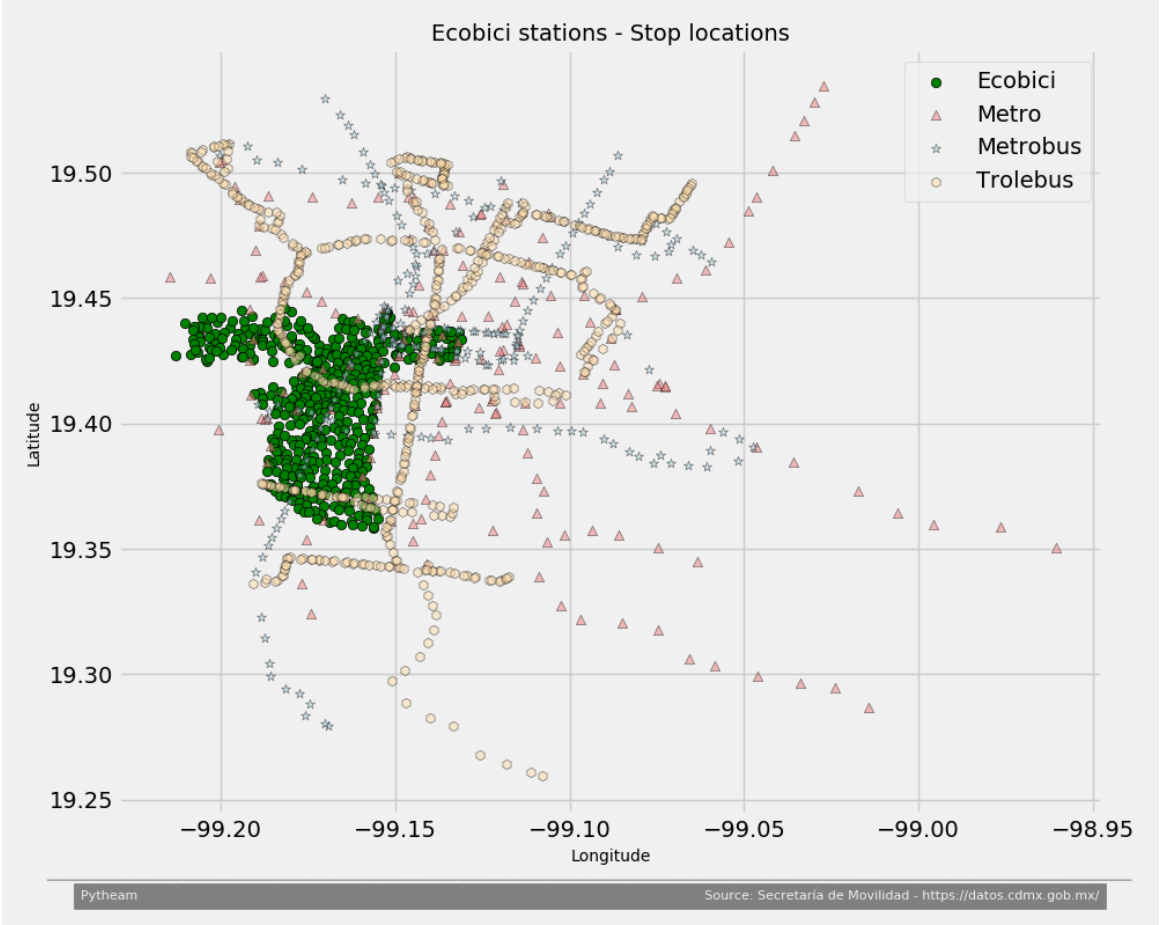
<Is ECOBICI complementary to the massive public transportation system network?>



Graph 2. *Ecobici stations – Stops locations*

System	Stops
Transporte Colectivo - Metro	195
Metrobus	234
Transporte Eléctrico – Trolebus	596
Other	4,996
Transporte Unificado de la Ciudad de México	**6,021**

Table 3. *Stops by system of transportation*



Graph 3. *Ecobici stations – Stops locations (Metro, Metrobus, Trolebus)*

Municipality	Approx. walking distance 100 meters from	Ecobici stations
Benito Juárez	Metro stop	6
	Metrobus stop	3
	Trolebus stop	18
	Subtotal	**27**
Cuauhtémoc	Metro stop	7
	Metrobus stop	26
	Trolebus stop	4
	Subtotal	**37**
Miguel Hidalgo	Metro stop	1
	Metrobus stop	1
	Trolebus stop	3
	Subtotal	**5**
Total		**69**

Table 4. *Walking distance between an Ecobici station and a stop location (100 mts)*

Municipality	Approx. walking distance between 100 and 200 meters from	Ecobici stations
Benito Juárez	Metro stop	13
	Metrobus stop	22
	Trolebus stop	39
	Subtotal	**74**
Cuauhtémoc	Metro stop	26
	Metrobus stop	69
	Trolebus stop	30
	Subtotal	**125**
Miguel Hidalgo	Metro stop	5
	Metrobus stop	3
	Trolebus stop	19
	Subtotal	**27**
Total		**226**

Table 5. *Walking distance between an Ecobici station and a stop location (100-200 mts)*

< *Is ECOBICI available for people in the city?*>

User age	
mean	35.2
min	16
max	92

Table 6. *Ecobici trips – User age*

Gender	Trips	%
Male	510,546	75
Female	173,469	25
Total	**684,015**	

Table 7. *Ecobici trips - Gender*

As of February 2019 (one month)

Trips	684,015
Bicycles used	6,073
Departure from	477 stations
Return to	477 stations
The departure station with more trips was	271 Av. Jesús García - J. Meneses, Located in Buenavista, Cuauhtémoc with 7,232 trips. 25% of those trips to stations located in Juárez suburb, 16% to stations located in Cuauhtémoc suburb, and 9% to stations located in Tabacalera suburb

Table 8. *Ecobici trips - Totals*

Departure from	Trips	%
Cuauhtémoc	**404,216**	**59**
Peak Hours: 6-10	119,991	
Peak hours: 12-16	119,118	
Peak hours: 18-22	112,604	
Non-Peak hours	52,503	
Benito Juárez	**140,397**	**21**
Peak Hours: 6-10	43,285	
Peak hours: 12-16	38,373	
Peak hours: 18-22	40,999	
Non-Peak hours	17,740	
Miguel Hidalgo	**139,402**	**20**
Peak Hours: 6-10	37,870	
Peak hours: 12-16	39,824	
Peak hours: 18-22	41,104	
Non-Peak hours	20,604	
Total	**684,015**	

Table 9. *Ecobici trips - Peak hours by municipality*

This question would have presented a better answer if there was more information available. Other possible questions to be answered are the following:

- What is the availability of units at critical times for all users?

When analyzing the dataset, we found that there are no adequate indicators to calculate the percentage of availability in critical hours. Some data that are not available are:

- number of available bicycles
- number of users
- number of stations in service
- number of requests for bicycles from the demand not met
- the time it takes to provide bicycles to stations

Therefore, we propose that the Public Bicycle System (Ecobici) provides the following base measurement data:

- number of bicycles not available and the time it takes to be available
- number of stations not available and the time it takes to be available
- the time to travel the bike paths during non-critical hours
- the time to travel the bike paths during critical hours
- the condition of the bicycles (available, in use, in maintenance, low)

In this way we will be able to measure the availability of the demanded service in critical hours.

< *Other facts: User's perception* >

As an attempt to understand cyclist's behavior, a profound analysis was made through the "Cyclist Count Research 2018", a survey conducted to more than 3,000 cyclists in the Mexico City area.

Results are as follows:

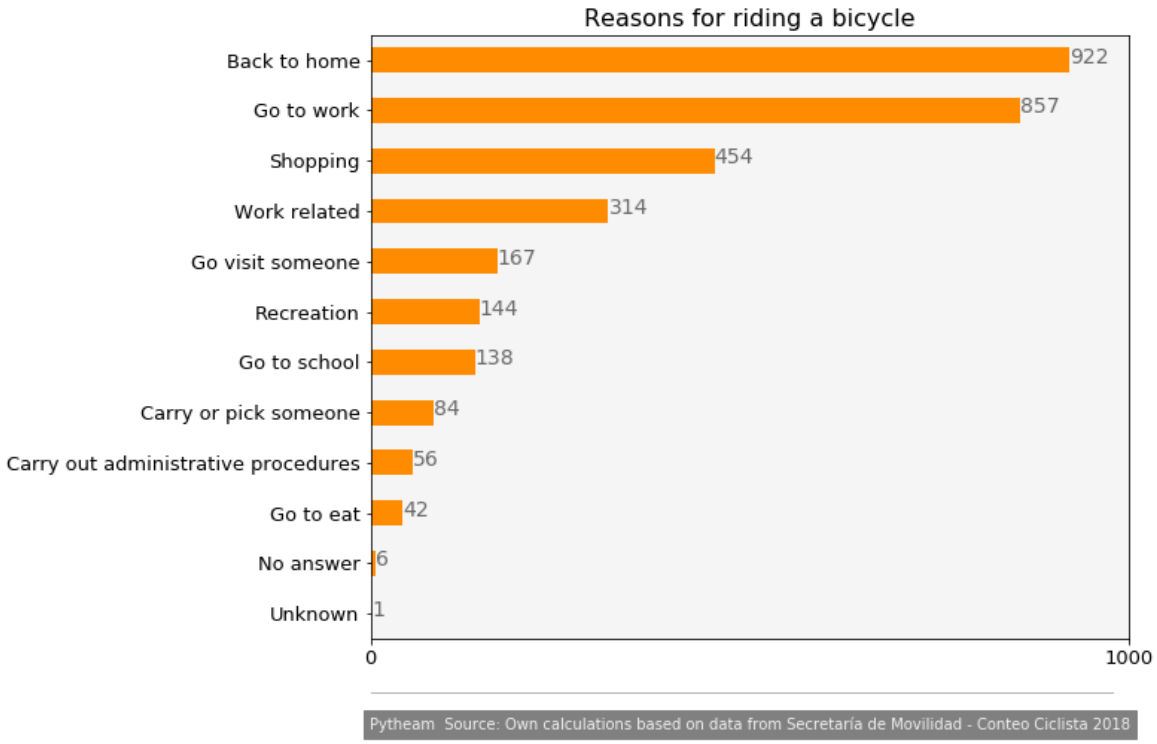
Bicycle type	n
Own	2,968
Other	91
Ecobici	65
Mobike	39
V-Bike	20
Dezba	1
Unknown	1
Total	**3,185**

Table 10. *Cyclist Count – Type of bicycle used*

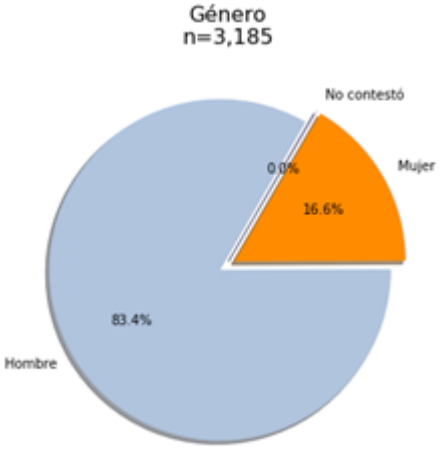
57.11% of **3,185** persons used the bicycle for work related activities.

From	To	Count
Home	Work	761
Work	Home	573
Work	Work	181
Work	Mall, store, market	84
Work	Other	82
Work	Another house	29
Other	Work	27
Work	School	19
Mall, store, market	Work	18
Work	Restaurant, bar, coffee shop	14
Another house	Work	11
Restaurant, bar, coffee shop	Work	9
School	Work	9
Work	Unknown	2
	Total	**1,819**

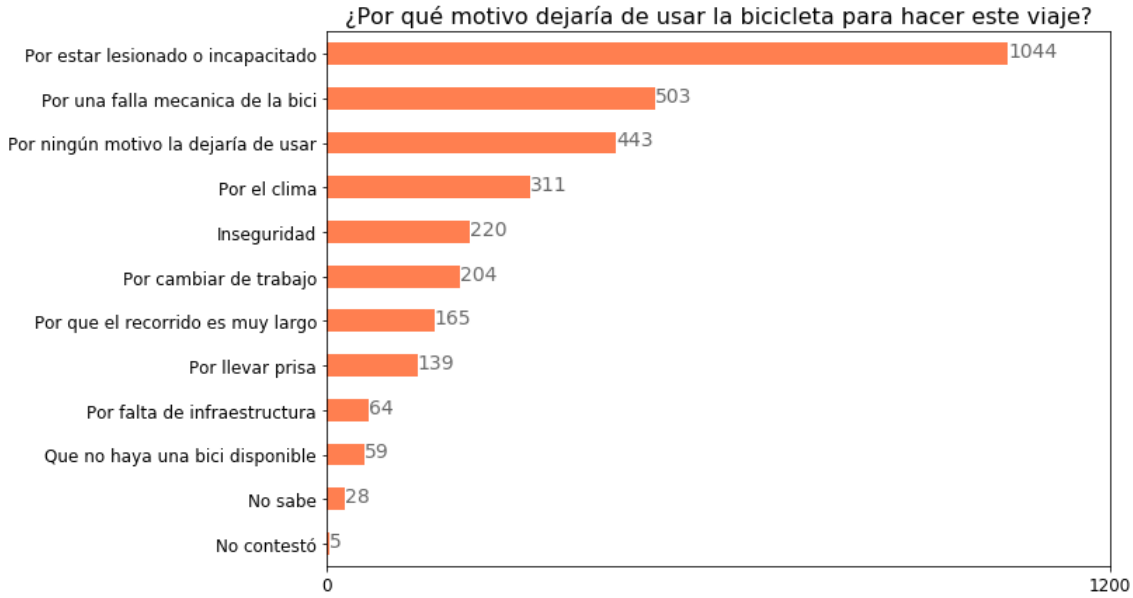
Table 11. *Cyclist Count – Use of bicycle for work related activities*



Graph 4. Cyclist Count –Reasoning for riding a bicycle



Graph 5. *Cyclist Count – Gender*



Graph 6. Cyclist Count – Reasoning for stop using a bicycle

C. RESULTS AND FINDINGS

<Our hypothesis>

$$H_0 = ECOBICI\ is\ a\ mobility\ solution\ for\ Mexico\ City$$

//Efficacy was proved for the users in the service area. Although it benefits to the people employed by the 45% of the formal companies, it is necessary to extend the benefit to the whole population.

//Requirements allow to access only to a minimum portion of the population.

//2 of the 3 municipalities where ECOBICI is located are in the list of the high income of the city.

//The installed bicycle infrastructure complements the existing public transportation network, but the coverage is very limited and centralized in only 7% of the total Mexico City area.

//The information shows that the main part of the trips is done in peak hours, however there is not information about the unsatisfied demand.

//In order to solve the previous results, the Strategic Plan for Mexico City's Secretaría de Movilidad set **three main goals for 2019**: Integrate, improve and protect, being the first the most important to complete, since ECOBICI operates under a different Ministry (Medio Ambiente). This plan foresees an ambitious unification: Ecobici and biciestacionamientos fully integrated into public transport strategy.

For the previously mentioned reasons, we can reject the null hypothesis.

3. PROCESS FOR CLEANING AND ANALYZING THE DATASETS

A. Process for cleaning the datasets

A brief summary of the steps for cleaning the datasets and some examples follows.

1. Select the tool to clean and explore the data.

   One of the project requirements was to use Pandas to clean and format our data sets, however we found sources to answer the research question about accessibility specifically by low income users, which had the information in PDF files or was presented in a web page. First, we copied it to Excel, then saved it as csv file and then we did the cleaning and analysis processes as described below or we included that information in our presentation because we didn't need to process it.

   Another optional project requirement was to use at least one API, if we could find an API with data pertinent to our primary research questions. We did find an API to obtain Ecobici information in https://www.ecobici.cdmx.gob.mx/es/informacion-del-servicio/open-data, but we couldn't complete the registration process in time for the project, because the authorization to use the data was not received.

2. Try to read the files in Pandas.

   a. For csv files we used the method *read_csv* with the default parameter (file location and name), stored the information in a data frame and displayed its head or its tail. If something went wrong, we added the following parameters and values: *low_memory=False*, *delimiter=","*.

      We set the combination of *low_memory=False* for dataset with more than 400,000 rows to avoid the warning and we set the combination of *delimiter=","* when the source used a semicolon to separate the different values in a row.

      In one dataset (Ecobici stations – 480 rows) we also used Excel and Google Maps to complete the information since it did not contain all the zip codes to make a group by municipalities and later in the analysis, we found that for many rows a default value in the column "Colonia" was assigned, so we had to search the correct station location and correct the file.

   b. For JSON we used the function *open()* and the method *load()*, displayed the JSON file content, stored the needed information in a data frame and displayed its head or its tail.

3. Print the data frame information with the method *info()* and length to find the name of the columns, the data type and the number of rows with information (not null).

4. Print the content of the data frame and make a general visual exploration to gain insight about what we will do in the next steps.

5. Select the appropriate columns to decide which part of the data frame is useful for our analysis and if it is necessary to transform the information included in those columns.

6. Set the conditions to filter the information.

   For the Statistical Directory of Economic Units, we focus the cleaning and posterior analysis, on those units with more than 51 occupied persons.

7. Create additional columns to store the information in the required format to begin with the analysis.

   a. Bins. Define the bins and the appropriate labels.

      We divided Ecobici trips by peak hours (6-10, 12-16, 18-22) and non-peak hours (0-6, 11, 17, 23).

      In the Economic Units we divided the information using the values in the catalog "Sistema de Clasificación Industrial de América del Norte 2018".

      For the Cyclist count we defined bins to divide the information in age groups.

   b. Transformations. Depending on the value or the data type of a column, we performed the required transformation to ease the calculations or to give the appropriate representation to the information.

      To make the division of the Ecobici trips by peak hours, we had to add a column with a representation of the hour as integer.

      For the questions in the Cyclist count, we had to transform all the integer values to their corresponding label. For example, if the question one had the value equals to 1, we had

to look in a catalog what label was associated to the value 1, we found that it was "Home" and then we assigned this label to the new column.

For the walking distance between a massive transportation stop and an Ecobici station, we defined a function to calculate the distance between two locations and add the result to a new column.

8. Make the necessary merges and verify if the resultant data frame had any variations with respect to the sources.

   For the Ecobici trips we need the location of the station, we merged the trips with the catalog and discovered that the length of the resultant data frame was different from the total trips. Then we verified the trips and found that the source reported trips to stations that where not in the catalog and we had no way to know where these stations were located, therefore we made the decision to omit these rows.

It is important to mention that even though the steps were described in a certain order, sometimes was necessary to do a step more than once or to return to a previous step later in the process.

Also, during the analysis we had to return to the cleaning process as deemed necessary (e.g. we found that for many rows of Ecobici stations a default value in the column "Colonia" was assigned, so we had to search the correct station location and correct the file).

Finally, in the project repository, folder *Notebooks,* you can find several Jupyter Notebooks with comments that describe the data exploration and cleanup process for each data set, along with the final data analysis.

B. Process for analyzing the datasets

A brief summary of the steps for analyzing the datasets and some examples follows.

1. Perform data frames functions to know more about the composition of the data.

   describe(). To display a statistical overview of a specific column and include this information in our final report (e.g. the user age of the Cyclist count).

   nunique(). To display the count of unique values of a specific column (e.g. number of bicycles used or stations in use during a period on Ecobici trips).

sum().To display the sum of values in a specific column (e.g. the minimum and maximum of occupied persons by Economic Unit or in the Cyclist count to find which percent of trips were made for work related activities).

value_counts(). To display the different values of a specific column order from largest to smallest. The information obtained in this way, was included in our final report (e.g. the gender of the Cyclist count or the Ecobici trips) or was the base for further analysis (e.g. origin station with more Ecobibi trips or the number of occupied persons by Economic Unit).

2. Obtain information to answer our research questions.

   a. Grouping. Select the columns to form the group and define which columns to extract from the source data frame and whether specific conditions must be met.

      To answer the research question about the Efficacy, specifically for the Service area coverage, we grouped the data frame by municipality. In this way, we discovered that the municipalities that has presence of Ecobici stations were also the municipalities with the greatest number of Economic Units with more than 51 occupied persons.

      To answer the research question about Complementing Public Transportation, specifically for the Bicycle station location, we grouped the data frame by municipality and type of transport. Previously, we had to set two conditions: Ecobici stations within a 100-meters walking distance from a Metro, Metrobus or Trolebus stop and Ecobici stations between 100- and 200-meters walking distance from a Metro, Metrobus or Trolebus stop.

      To answer the research question about Availability, specifically in peak hours, we grouped the data frame to find the trip routes and the most popular trip route by peak hour.

   b. Order. Present the information from largest to smallest to extract the top 3 observations or to continue the analysis.

3. Decide which information to present in tables and which information in graphs.

   To report the findings from our research and support the arguments of our answers we used:

a. Tables. For individual values, to compare individual values but not entire series of values to one another, to communicate quantitative information in different units of measure, and when both summary and detail values were included.

b. Graphs. To communicate messages contained in the shape of the values such as patterns, trends or exceptions, or to reveal relationships among whole sets of values.

4. Plot the information.

   In the previous step we defined which information to plot, in this step we selected an appropriate graph to communicate our findings.

   We chose bar charts for categoric, ordered, and discrete variables, pie charts for comparing the size of relative parts and scatter plot to show the relationship between two variables.

   In general, for the graphs we set the title, named the axes, units of measurement to use, the values to plot, the range and interval of these values and the appropriate scale to use. Then we plotted the information and discussed the findings within the team.

   Additionally, for the scatter plots we searched for ways to improve the appearance and found that we could use styles in Matplotlib. We used the style *fivethirtyeight* for the bar and scatter plots used in the presentation.

   Also, when we were discussing the ways to show the locations of the Economic Units versus the Ecobici Stations we made an implementation of a Heat Map using the Google's APÍ, however the visual representation was hard to understand for the viewer and the navigation into the map was not useful, so after showing the output to our instructional team, we decided against that implementation and used a scatter plot to communicate the relationships within the variables.

5. Decide which tables and graphs represented the significant features and findings of our investigation.

   Considering the project requirements about the use of Matplotlib to create a total of 6-8 visualizations of our data (ideally, at least 2 per research question) and save PNG images of our visualizations for inclusion in our presentation, we had to decide of all graphs created on our Jupyter Notebooks which ones represented the significant features and findings of our investigation. The following table lists the reasons for including a table or graph in our presentation.

Table	Include in presentation?
Research question: Efficacy. Service area coverage	
1. Ecobici stations by municipality	Yes, because it shows the municipalities with Ecobici stations and if we add the graph that shows the municipalities with more Economic Units (top 3), we can establish a relationship between the data.
2. Statistical Directory of Economic Units. INEGI	No, although it provides the Economic Units, the minimum of occupied persons and the maximum of occupied persons (at least) it is better to include the bar chart *Economic units by municipality* and include the data of occupied persons in a text box in the presentation.
Research question: Complement massive transportation	
1. Ecobici stations by municipality	Yes, as a remainder of the presence of Ecobici.
3. Stops by system of transportation	Yes, because it shows the number of stops for each system of transportation and the total of stops in the Unified System of Transportation for Mexico City.
4. Walking distance between an Ecobici station and a stop location (100 meters)	Yes, because it shows the number of Ecobici stations by municipality within a walking distance from a stop location of 100 meters, and we can make the distinction between these information and the Daily average distance people walk by trip from the source: https://moovitapp.com/insights.
5. Walking distance between an Ecobici station and a stop location (100-200 meters)	Yes, because it shows the number of Ecobici stations by municipality between 100- and 200-meters walking distance from a stop location, and we can make the distinction between these information and the Daily average distance people walk by trip from the source: https://moovitapp.com/insights.
Research question: Availability	
6. Ecobici trips - User age	Yes, it shows the average, minimum and maximum age of the users that made the trips during the month.
7. Ecobici trips - Gender	Yes, it shows the distribution of gender for the users that made the trips during the month.
8. Ecobici trips - Totals	Yes, it shows the total trips, the stations used and the departure station with most trips.
9. Ecobici trips - Peak hours by municipality	Yes, it shows the trips by municipality done in peak and non-peak hours.
10. Cyclist Count – Type of bicycle used	No, but we include some information in a text box in the presentation because it shows that 9 of 10 bicycles belong to the user.

Table	Include in presentation?
Research question: Availability	
11. Cyclist Count – Use of bicycle for work related activities	Yes, because it shows that the bicycle is widely used to go to work.

Graph	Include in presentation?
Research question: Efficacy. Service area coverage	
1. Economic units by municipality INEGI	Yes, because it shows the municipalities with more Economic Units (top 3) and if we add the table that shows the municipalities with Ecobici stations, we can establish a relationship between the data.
Research question: Complement massive transportation	
2. Ecobici stations – Stops locations	Yes, because it shows the concentration of Ecobici in one part of the city and the rest of the stops without a nearer Ecobici station.
3. Ecobici stations – Stops locations (Metro, Metrobus, Trolebus)	Yes, because it provides a deeper view of the systems of transportation that are near the concentration of Ecobici stations and the number of stops without a nearer Ecobici station.
Ecobici stations – Metro stops locations)	No, although it provides a deeper view of the Metro stops locations that are near the concentration of Ecobici stations and the number of stops without a nearer Ecobici station, it is better to show Metro, Metrobus, and Trolebus together.
Ecobici stations – Stops locations Metro, Metrobus)	No, although it provides a deeper view of the Metro and Metrobus stops locations that are near the concentration of Ecobici stations and the number of stops without a nearer Ecobici station, it is better to show Metro, Metrobus, and Trolebus together.
Research question: Accesibility	
Type of Credits acquired by Mexico City's Population	Yes, because it shows the percent of the population with No formal credit, Credit Cards and other type of credits.
Research question: Availability	
4. Cyclist Count – Reasoning for riding a bicycle	Yes, because it shows that the bicycle is widely used to go to work and back home.

Graph	Include in presentation?
Research question: Availability	
5. Cyclist Count – Gender	No, but we chose to show the summary data in a text box and translate to English the results.
6. Cyclist Count – Reasoning for stop using a bicycle	No, but we will include it in our final report, in Spanish to preserve the sense of the sample's opinions.

6. Write the documentation with the findings and the conclusions.

   The required documentation for the project was a presentation and a write-up summarizing our major findings, that include a heading for each research question we asked of our data, and under each heading, a short description of what we found and any relevant plots.

   The presentation and the document can be found in our repository (https://github.com/bkachava/project1) under the folder *Docs*.

7. Save the resulting data frames for later use.

   Because some of the analyzed data sets had more information that was not required to answer our research questions and we made a relevant view of the data that we wanted to show to our instructional team, we saved the resulting data frames for later use using the Pandas method *to_csv()*.

Finally, in the project repository, folder *Notebooks,* you can find several Jupyter Notebooks with comments that describe the data exploration and cleanup process for each data set, along with the final data analysis.