CASES IN URBAN MOBILITY

Master of Science in Urban Mobility

Course REPORTS 2024-2025

R23. BICYCLE TRAFFIC IN BARCELONA

Luca Moritz Liborio Liebscht

Barcelona, November 24, 2024

1. INTRODUCTION AND OBJECTIVES

The aim of this project is to investigate the potential of permanent inductive loop bicycle counters by analyzing data collected in Barcelona from 2017 to 2023. The received data is aggregated in 15 minutes intervals and comes with an error estimation that indicates what percentage of the counter was not working during the observation. The observations have been complemented with the data about the station location and typology available in the municipal open data portal.

One of the key objectives of this study is to perform comprehensive data cleaning to ensure the accuracy and reliability of the bicycle count dataset. The analysis revealed a variety of errors in the data, including observations classified as "Invalid," "Unknown," and "Partial," which differ significantly in their characteristics and occurrence patterns compared to "Valid" data. These discrepancies highlight the need to remove or impute certain types of erroneous data to reduce noise and enhance the quality of the dataset. By addressing these data issues, the study aims to improve the precision of subsequent analyses and the robustness of the model outputs.

Another objective is to calculate standard bicycle count metrics commonly referenced in the literature, ensuring both adequate data coverage and consistency in observed patterns. To achieve this, data has been aggregated at hourly and daily levels, with daily averages calculated for each year. Eight validity tests were applied to detect any atypical observations and aggregate only when enough valid data was available.

The last objective of this study is to analyze the temporal patterns in bicycle traffic data, capturing long-term trends as well as seasonal (yearly), weekly, and daily variations. To accomplish this, we used Prophet, a Python library specifically designed for forecasting time series data. Prophet was chosen due to its ability to model complex, non-linear trends while handling strong seasonal effects, irregular data, and outliers. This approach allows us to explore not only general trends over the years but also recurring patterns within weeks and days, enhancing our understanding of cycling behavior over different time scales.

The rest of the report is structured in the following way. First, we will go over the literature, the description of the dataset and explore what has been done in the context of Barcelona, then thought the methodology, the results and conclusions.

2. LITERATURE REVIEW

One of the most commonly used units to measure bicycle volume is the annual average daily bicycle traffic (AADBT) (Miah et al., 2024; Nordback et al., 2013) inspired in the annual average daily traffic (AADT) (Esawey, 2014). Some other metrics are the monthly average daily bicycle traffic (MADBT) and the daily bicycle traffic (DBT) (Miah et al., 2024).

The estimation of the units of measure has been traditionally done using permanent automatic counters (Miah et al., 2024) and short-duration counts generalized used expansion factors generated from the permanent counters (Roll & Proulx, 2018). There have also been other approaches using Crowdsourced data (Dadashova et al., 2020) or some approaches using short-duration counters and statistical models (Roll & Proulx, 2018).

3. DATA DESCRIPTION

The available data set is comprised of 54.037.979 15-minute observations collected using permanent counters in various locations in Barcelona between 2017 and 2024. The number of stations varies during the years and does not match on both datasets, as seen in A1. The counters are in both unidirectional and bidirectional bicycle lanes and treat each travel direction separately. Each observation contains the id of the counter, the date and time of the observation, the bicycle count, and the associated error. Each station contains the id, a string describing the counter, the type of vehicle that it counts, the number of lanes, the district and neighborhood number, the type of equipment and the coordinates. A map of the counters in has been made and is available in A2.

4. SIMILAR STUDIES IN BARCELONA

The student has searched for reports and studies done in Barcelona or it's metropolitan area that involve counting of cyclists or use the data obtained from counting stations. The goal has been to find studies that were performed on a recurrent basis, and to determine what data is being collected.

Three main studies have been found, the bicycle barometer done by the *Real Automóvil Club de Catalunya* (RACC) between 2018 to 2024 (RACC, 2024), the Radiografia Ciclista made by the *Bicicleta Club of Catalunya* (BACC) between 2020 and 2023 (BACC, 2024), and the counting's on the metropolitan cycling network done by the Metropolitan Area of Barcelona (AMB) in 2019 and 2021 (AMB, 2022).

The studies done by the RACC often are composed of a survey and counts done in multiple locations. The student found six different editions published on the website, the first three focusing on urban cycling, and the last three on the accesses to the city. The studies talk about the demographics of the cyclists, focusing on gender and age, as well as the characteristics of the bicycle, wherever it's shared or private, and electric vs mechanical. It is unclear what methodology/technology has been used for the counts.

The studies done by the BACC consist in recording serval locations using cameras during peak hours, with recording during the morning and during the evening. The recordings are later analyzed, with the focus of determining wherever the user was using a scooter or a bicycle (With a detailed bicycle subclassification into seven categories, the propulsion of the vehicle and the identified gender of the user.

Regarding the studies done by the AMB, the studies collected data using four different methodologies: automatic pneumatic tubes installed during a full week, recording using cameras on intersections during 24h, manual counts on intersections during peak hours in the mornings and evenings, and fixed automatic counters installed in the network. The studies have estimated the AADBT for the different studied locations. A very interesting conclusion of the second edition, is that the increase in the use of the bicycle between 2019 and 2021 has been higher more in the metropolitan municipalities (49%) than in the city of Barcelona (5%)

As a general overview, most of the recent studies do focus on characterizing who uses bicycle by type of vehicle and gender. All the studies indicate that there is a considerable gender gap, that has been slowly diminishing, and that the use of the scooter in the bicycle lanes has been growing during the last years.

5. METHODOLOGY

The first step has been to use the associated error to distinguish the valid observations of the erroneous ones. The error is the percentage of the counter that was not working during the observation. The observations have been classified in four categories based on the error, described in the Table 1. Once classified, some metrics have been estimated for each category, such as the count of observations, the percentage of total observations, the number of days the category was present, and the highest daily error percentage.

Category	Situation	Treatment ideas
Valid	Error is equal to 0	It is good so it should be used.
Partial	Error is 0<E<100	It can be used, if relevant.
Invalid	Error is equal to 100	It is wrong, it should not be used.
Unknown	Error is missing (NAN)	More information is needed to evaluate.

Table 1: Classification of observations based on the error

The second step involved analyzing the observations to identify outliers and periods with potentially anomalous patterns. To facilitate this, observations were aggregated at the hourly level, requiring a minimum of 30 minutes of valid data per hour, a standard threshold in similar studies. Aggregated data was then subjected to a series of tests to detect irregularities in individual counter behavior:

1. No site should show a prolonged zero count exceeding 90 hours.
2. No site should have six or more identical values above 5.
3. Sudden increases over 10 times, with an initial count above 15, should be flagged.
4. Unusual hourly patterns (e.g., 3 AM > 3 PM) should be reviewed.
5. Hourly counts should not exceed 885 bicycles.
6. Daily counts should not exceed 11,225 bicycles.
7. Each valid day needs at least 22 hours of data.
8. Each valid month needs at least three weeks of data.

Then, the DBT, MBT and AADBT have been estimated. In the case of the AADBT, the number of days with valid data had to be higher than 120 (Three months).

The third step involved analyzing temporal patterns in the dataset using Prophet, a forecasting tool that builds an additive model with non-linear trends. The model is represented as $y^{(t)} = \mathrm{Trend(t)} + \mathrm{Yearly(t)} + \mathrm{Weekly(t)} + \mathrm{Daily(t)} + \epsilon$, where ϵ accounts for random noise. Prophet is designed to handle time series data with strong seasonality and irregularities, such as missing data, trend shifts, and outliers, making it an ideal choice for analyzing bicycle traffic patterns.

Prophet decomposes the time series into key components: trend, capturing long-term changes; seasonality, identifying recurring patterns at yearly, weekly, and daily scales; and holiday effects, reflecting temporary changes due to events. The model outputs include the overall trend, visualizations of yearly, weekly, and daily seasonal patterns, and future forecasts with confidence intervals. These components allow us to identify not only recurring patterns but also anomalies, such as sudden declines or peaks, which may be linked to external factors like weather or policy changes.

6. RESULTS

6.1. Error Analysis and Data Quality Assessment

In the Figure 1 we can observe the evolution in the number of observations as well as the presence of errors of varying seriousness. The number of observations increases overtime, with some notable drops that can last various days or may not recover at all.

Figure 1: Daily Count of Observations Categorized by Error Type

Focusing on the categories of observations, we can see that most of the data falls under the "Valid" category, followed by "Invalid" and "Unknown," as shown in Table 2. The "Valid" category accounts for most observations (89.9%) and is present on almost every day (99.9%). In contrast, "Unknown" errors are relatively rare, making up only 2.5% of the total, and occur on fewer days (15.1%). The "Partial" category, though less frequent overall (1.4%), appears on most days (99.1%) but never exceeds 25.5% of the daily observations. "Invalid" observations are more common than "Unknown" (6.1%) and occur on nearly as often as "Valid," with a significant maximum daily ratio of 61.6%.

It is interesting to note that the metrics for "Partial" and "Invalid" observations are quite similar, while the "Unknown" category shows markedly different characteristics. The absence of the "Valid" category on some days and the occurrence of a 100% ratio for "Unknown" suggest that there is at least one day where all observations are classified as "Unknown." This is confirmed by Figure 1, which shows that this happened only once, on August 14, 2023. Additionally, there was a period from November 2 to November 30, 2022, when most observations were classified as "Unknown".

Category	Observations	Percentage	Days present	Max ratio per day
Valid	48623146	89.9%	99.9%	100%
Unknown	1363399	2.5%	15.1%	100%
Partial	754720	1.4%	99.1%	25.5%
Invalid	3296714	6.1%	98.9%	61.6%

Table 2: Summary of Observation Categories and Error Statistics

The bicycle counts for observations categorized as “Invalid” and “Unknown” are always zero, requiring their deletion and imputation. “Partial” observations also show lower intensities than “Valid,” with some subcategories (e.g., 60% and 80% error) consistently at zero. Therefore, these partial observations will also be deleted and imputed, as they differ statistically from valid data and may introduce noise that reduces accuracy. See Figure 2 to observe the various subcategories of “Partial”, and their relative importance.

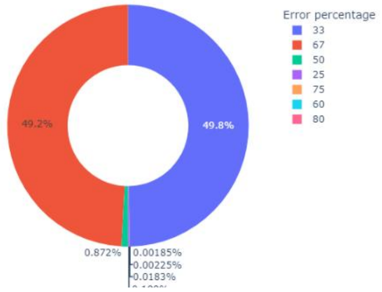


Figure 2: Partial observation by error percentage

6.2. Unit of analysis generation

The data is reduced to almost ¼ of the initial size during the grouping by hour, and as can be seen in Figure 3, only a 0.87% of the observations are deleted as they have less than ½ hour of valid data. The number of observations after the aggregation is 12.26 million. This number is reduced when we apply the various tests described in the methodology, with relevant decreases during the test 1 (0.24M), the test 5 (0.13M), the test 7 (1.91M) and test 8 (0.98M) as can be seen in A3.

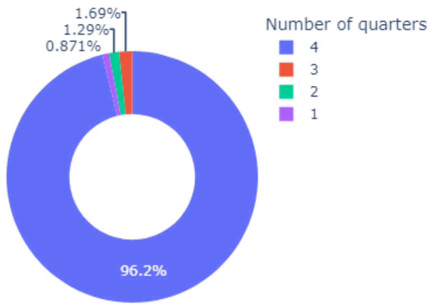


Figure 3: Results of grouping by hour

Then, we have computed the DBT and the AADBT for each of the years, the number of observations for each of the units of analysis are described in Table 3. We can observe a significant increase between 2018 and 2019, that was already visible in Figure 2.

Unit	2017	2018	2019	2020	2021	2022	2023
Hours	143.136	368.712	1.359.312	1.233.240	1.794.000	1.813.176	2.267.136
DBT	5.964	15.363	56.638	51.385	74.750	75.549	94.464
AADBT	23	33	196	213	235	261	306

Table 3: Units of analysis

Figure 4 shows the annual average daily bicycle traffic (AADBT) distribution per year, with boxplots summarizing the spread of values. The data highlights a decrease in median AADBT from 2017 to 2019, followed by more stable and higher values in later years, likely due to increased data coverage. The wider spread in recent years suggests greater variability in bicycle counts across different counters.

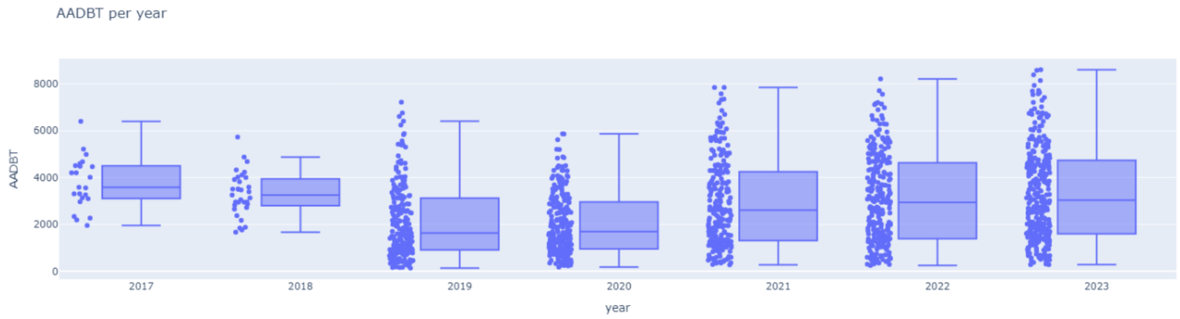


Figure 4: Boxplots of the AADBT generated for each year

6.3 Temporal patterns analysis

After preprocessing the data, we retained 8,978,712 valid hourly observations, which were used to train the Prophet model. Model training took approximately 100 minutes and 36 seconds. Once trained, the model generated outputs for the first six months of 2024, including the overall trend, seasonal patterns (yearly, weekly, and daily), and corresponding confidence intervals. These visualizations provide insight into both long-term trends and recurring temporal patterns.

6.3.1 Trend Analysis (Figure 5):

The trend component highlights key changes over the historical period. A sharp decline in March 2020 coincides with the onset of lockdown measures, resulting in a significant reduction in bicycle traffic. This is followed by a gradual recovery, indicating a return to pre-lockdown levels in the subsequent months. Additionally, an initial decline from 2017 to 2019 can likely be attributed to a lower number of active counters during those years, while the increase after 2019 reflects more comprehensive data coverage (see Table 3).

Figure 5: Trend of the observations generated with Prophet

6.3.2 Yearly Seasonality (Figure 6):

The yearly seasonality plot reveals a distinct cyclical pattern, with bicycle counts peaking in June and reaching their lowest in January. This variation corresponds to weather conditions, where warmer months favor cycling activity, and colder months see reduced usage. Seasonal climate changes and potential holiday schedules also likely contribute to these fluctuations.

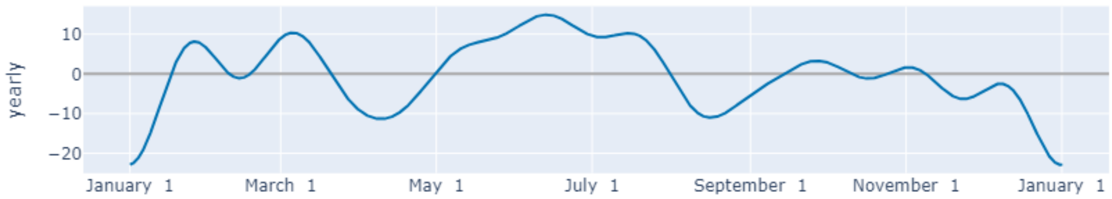


Figure 6: Yearly patterns observed with Prophet

6.3.3 Weekly Seasonality (Figure 7):

The weekly pattern shows the highest cycling counts occurring from Tuesday to Friday, while Sundays and Mondays exhibit the lowest activity. This behavior is consistent with weekday commuting patterns, where cycling is more common for work or school travel. Lower counts on weekends, particularly Sundays, may indicate reduced commuting activity.

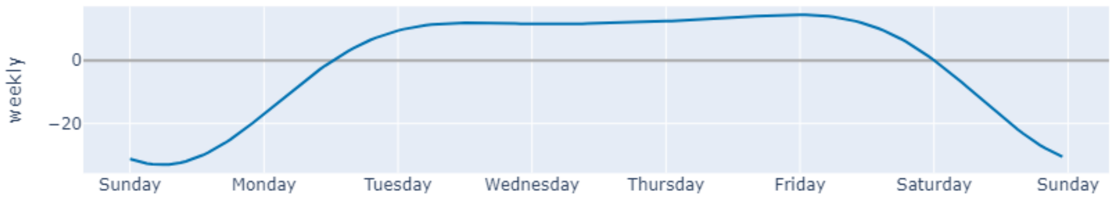


Figure 7: Weekly patterns observed with Prophet

6.3.4 Daily Seasonality (Figure 8):

Daily patterns display a bimodal distribution, with peaks occurring at 9 AM and 7 PM. These peaks align with typical commuting hours, suggesting a high volume of cyclists traveling to and from work or school. The more pronounced evening peak may reflect both commuters and recreational cyclists, who are likely to ride more after work hours.

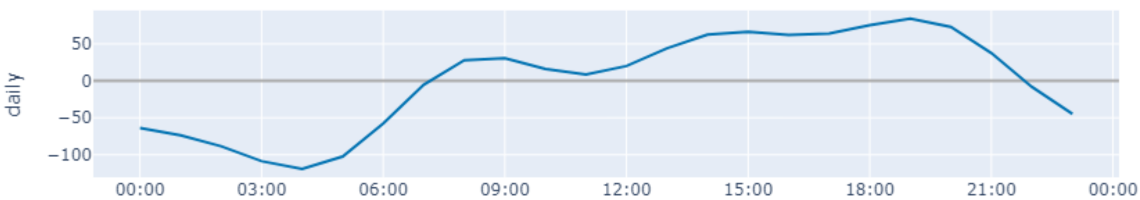


Figure 8: Daily patterns observed with Prophet

7. CONCLUSIONS AND RECOMMENDATIONS

This study has provided a comprehensive analysis of bicycle traffic patterns in Barcelona from 2017 to 2023 using data from permanent inductive loop counters. Key findings and implications include:

- **Data Quality and Error Analysis**: The majority of observations (89.9%) were valid, ensuring reliable insights. However, the presence of invalid and unknown data highlights the need for robust data-cleaning mechanisms. Techniques like imputation and deletion of erroneous data were essential to maintain the integrity of the analysis.
- **Temporal Patterns in Bicycle Traffic**: Long-term trends revealed a steady growth in bicycle use after 2019, potentially linked to improved counter coverage and increased cycling infrastructure. Seasonal, weekly, and daily patterns reflected expected variations, such as higher counts in summer and during weekday commuting hours. The impact of external events, like the sharp decline during March 2020 due to lockdowns, underscores the sensitivity of cycling patterns to external factors.
- **Use of Advanced Modeling Tools**: Prophet's capability to handle non-linear trends and seasonalities proved valuable in analyzing complex temporal patterns. Its ability to highlight anomalies and predict future trends makes it a crucial tool for urban mobility studies.
- **Key Implications for Urban Planning**: Encouraging gender diversity in cycling, addressing safety concerns, and managing the coexistence of bicycles and scooters in shared lanes remain critical for fostering inclusive cycling culture.

8. REFERENCES

AMB. (2022, febrero 27). Aforaments a la xarxa Bicivia. https://shorturl.at/49I3V

BACC. (2024, febrero 14). Radiografia Ciclista de Barcelona 2023. https://bacc.cat/radiografia-ciclista-de-barcelona-2023-resultats-de-lestudi/

Dadashova, B., Griffin, G. P., Das, S., Turner, S., & Sherman, B. (2020). Estimation of Average Annual Daily Bicycle Counts using Crowdsourced Strava Data. Transportation Research Record: Journal of the Transportation Research Board, 2674(11), 390-402. https://doi.org/10.1177/0361198120946016

Esawey, M. E. (2014). Estimation of Annual Average Daily Bicycle Traffic with Adjustment Factors. Transportation Research Record: Journal of the Transportation Research Board, 2443(1), 106-114. https://doi.org/10.3141/2443-12

Miah, M. M., Hyun, K. K., & Mattingly, S. P. (2024). A review of bike volume prediction studies. Transportation Letters, 1-28. https://doi.org/10.1080/19427867.2024.2310831

Nordback, K., Marshall, W. E., Janson, B. N., & Stolz, E. (2013). Estimating Annual Average Daily Bicyclists: Error and Accuracy. Transportation Research Record: Journal of the Transportation Research Board, 2339(1), 90-97. https://doi.org/10.3141/2339-10

RACC. (2024, septiembre 26). Sexto Barómetro RACC de la movilidad ciclista en Barcelona y en sus accesos. https://movilidad.racc.es/sexto-barometro-racc-de-la-movilidad-ciclista-en-barcelona-y-en-sus-accesos/

Roll, J. F., & Proulx, F. R. (2018). Estimating Annual Average Daily Bicycle Traffic without Permanent Counter Stations. Transportation Research Record: Journal of the Transportation Research Board, 2672(43), 145-153. https://doi.org/10.1177/0361198118798243

ANNEXES

A1. Evolution number of counters in Barcelona

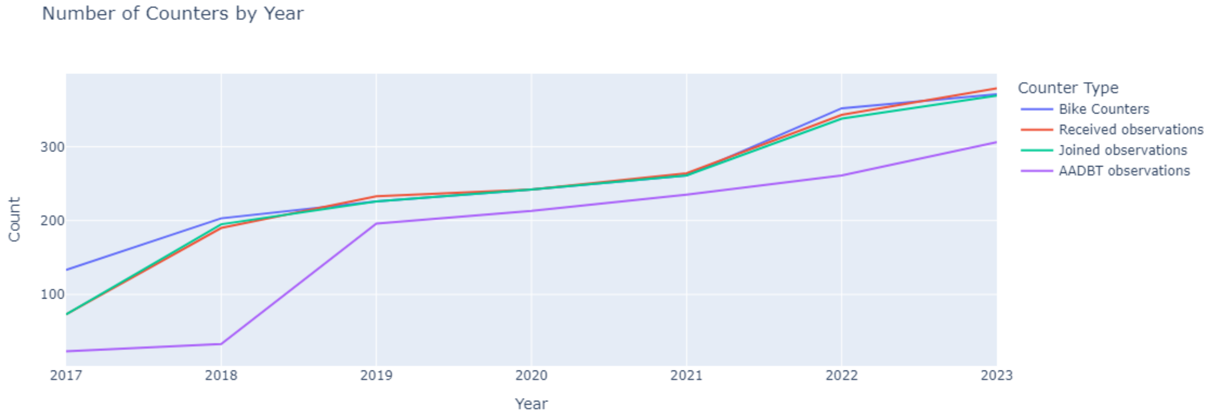
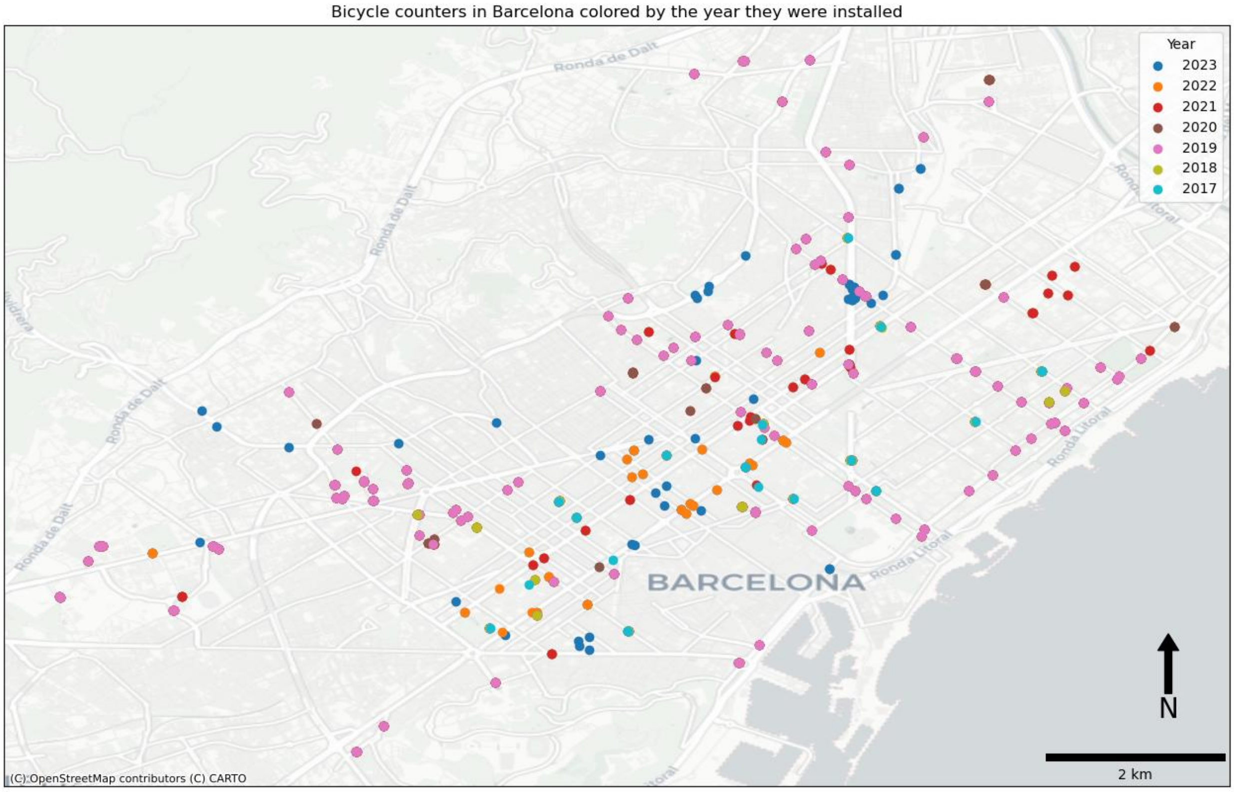


Figure 9: Evolution of number of bike counters in Barcelona

A2. Map of bicycle counters



A3. Reduction of Data After Each Test

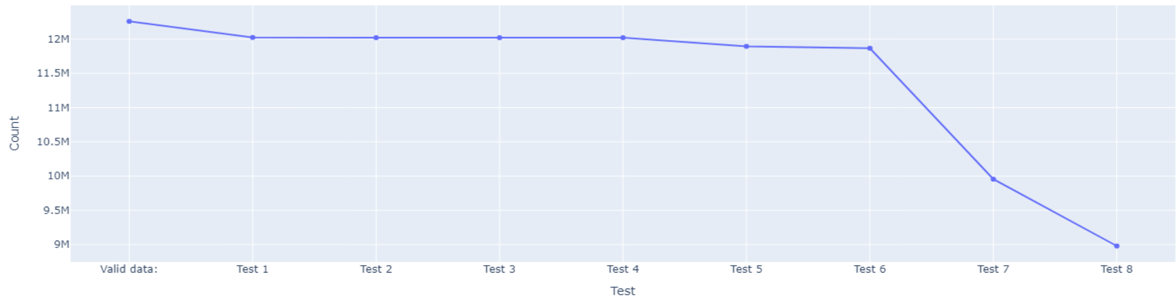
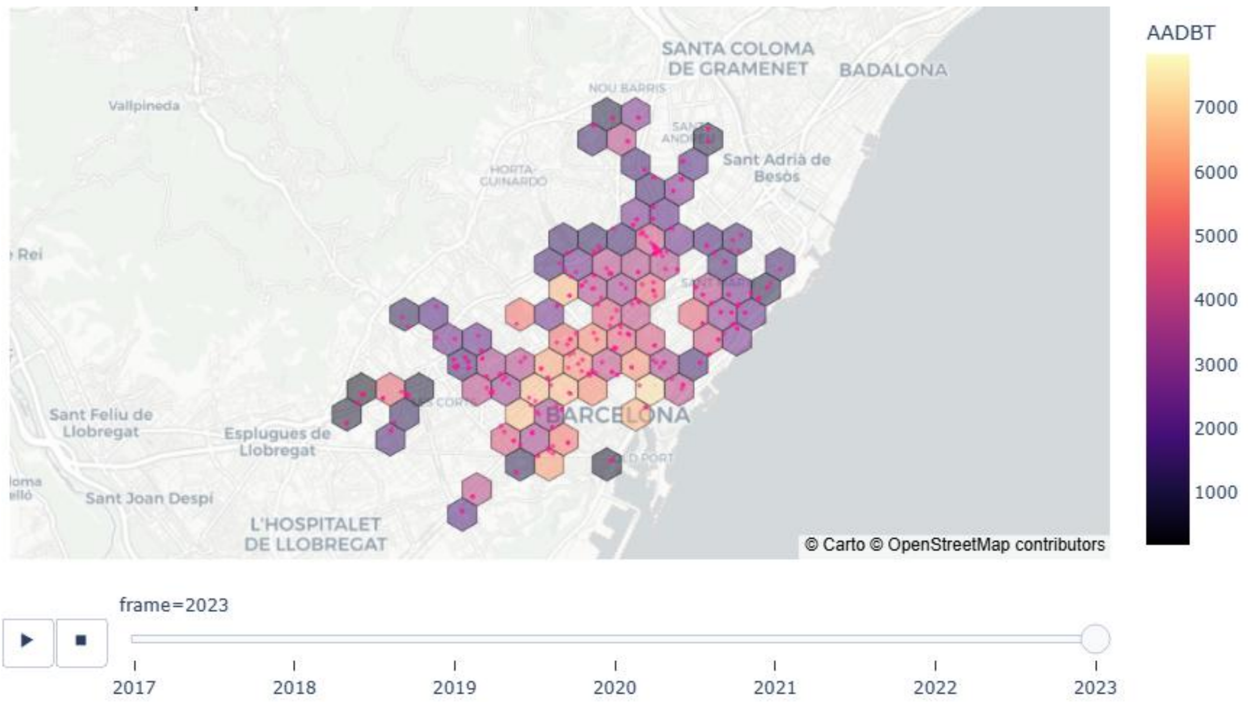


Figure 10: Reduction of Data After Each Test

A4. Interactive visualization of the installation of the counters and the average AADBT by hexbins

Click on the image to open the interactive visualization made using hexbins. The interactive version of some of the other figures is available in the [following link](https://cyclingmoritz.github.io/250MUM006/)[1].



[1] https://cyclingmoritz.github.io/250MUM006/