
Realizado por:
Isanagui Rojas



PyCing

Tutorizado por:
Joan Gasull

➔ Introducción ➔

En este trabajo de análisis de datos se investigó el uso del sistema de bicicletas **Bicing** de la ciudad de Barcelona, utilizando datos desde el **01-01-2020** hasta el **31-05-2024**:

- Uno de los mayores desafíos del proyecto fue trabajar con los más de **222 millones de registros** que componen el dataset.
- Para optimizar el manejo, se compiló el dataset en un archivo **.parquet**, reduciendo el tamaño de los datos de unos **18GB** a **1.3GB** aprox.
- El proyecto se ha desarrollado integralmente en **Python**.

Hipótesis de Investigación

→ HIPÓTESIS₀

El horario de inicio de la jornada laboral no muestra una diferenciación de clase en el uso de las estaciones del Bicing.

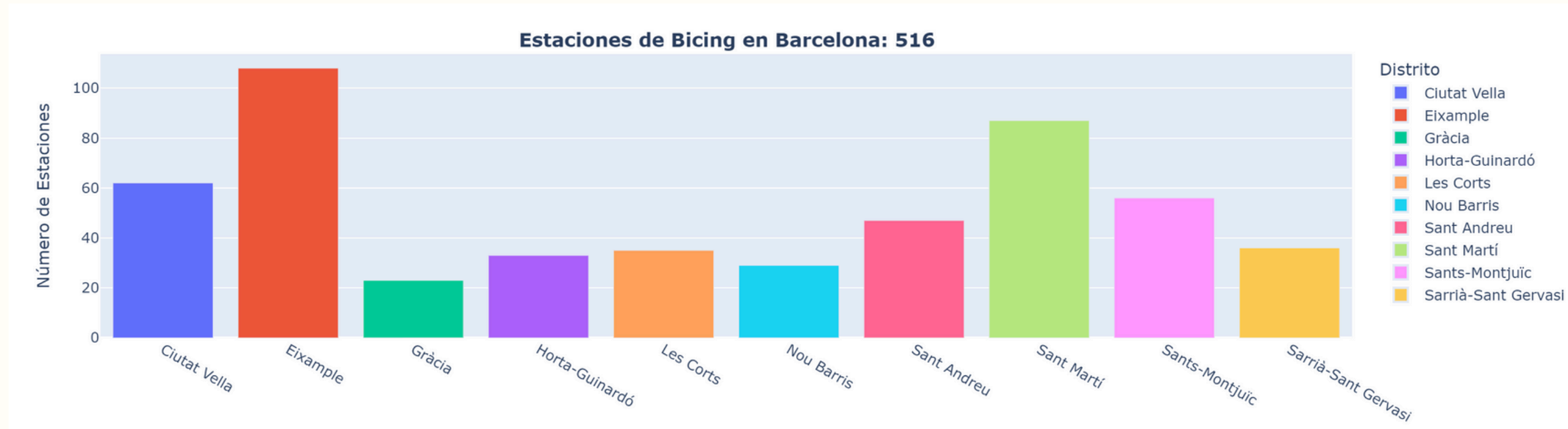
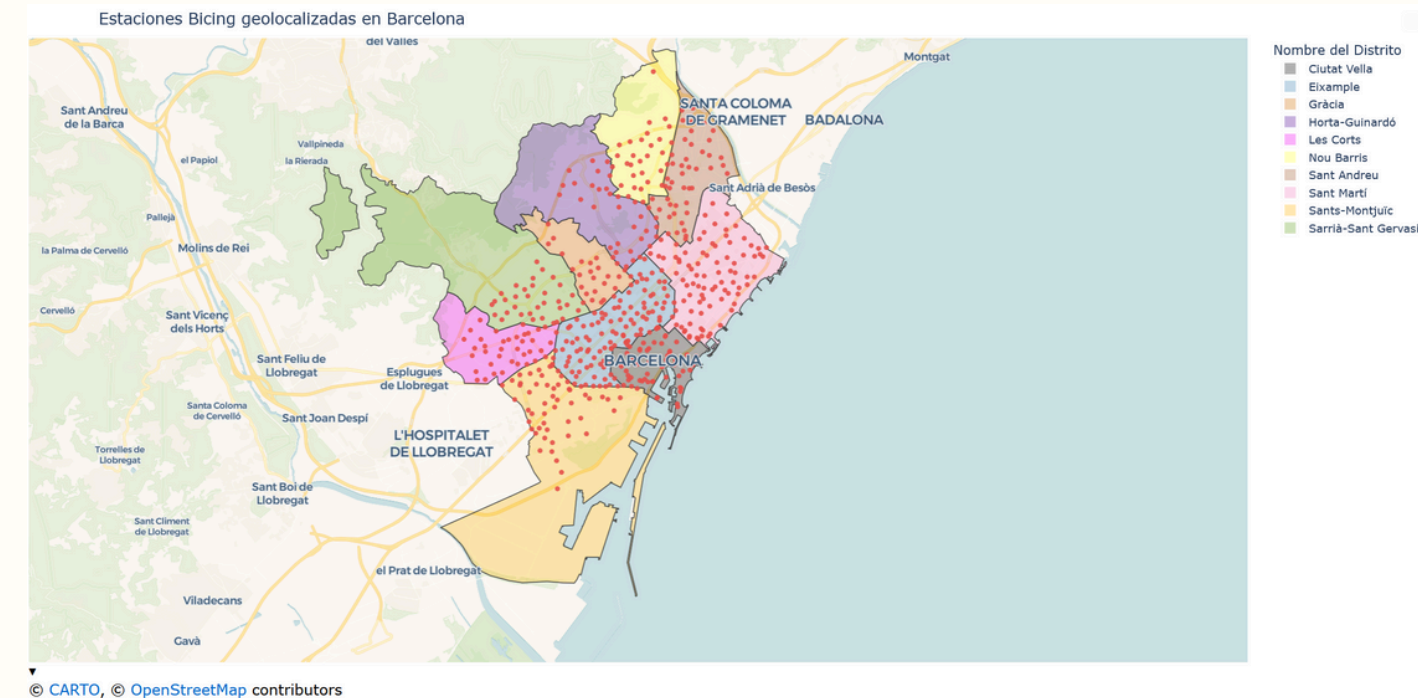
→ HIPÓTESIS₁

El horario de inicio de la jornada laboral muestra una diferenciación por clase en el uso de las estaciones del Bicing.

→ Datos:

Geolocalización de las estaciones de Bicing:

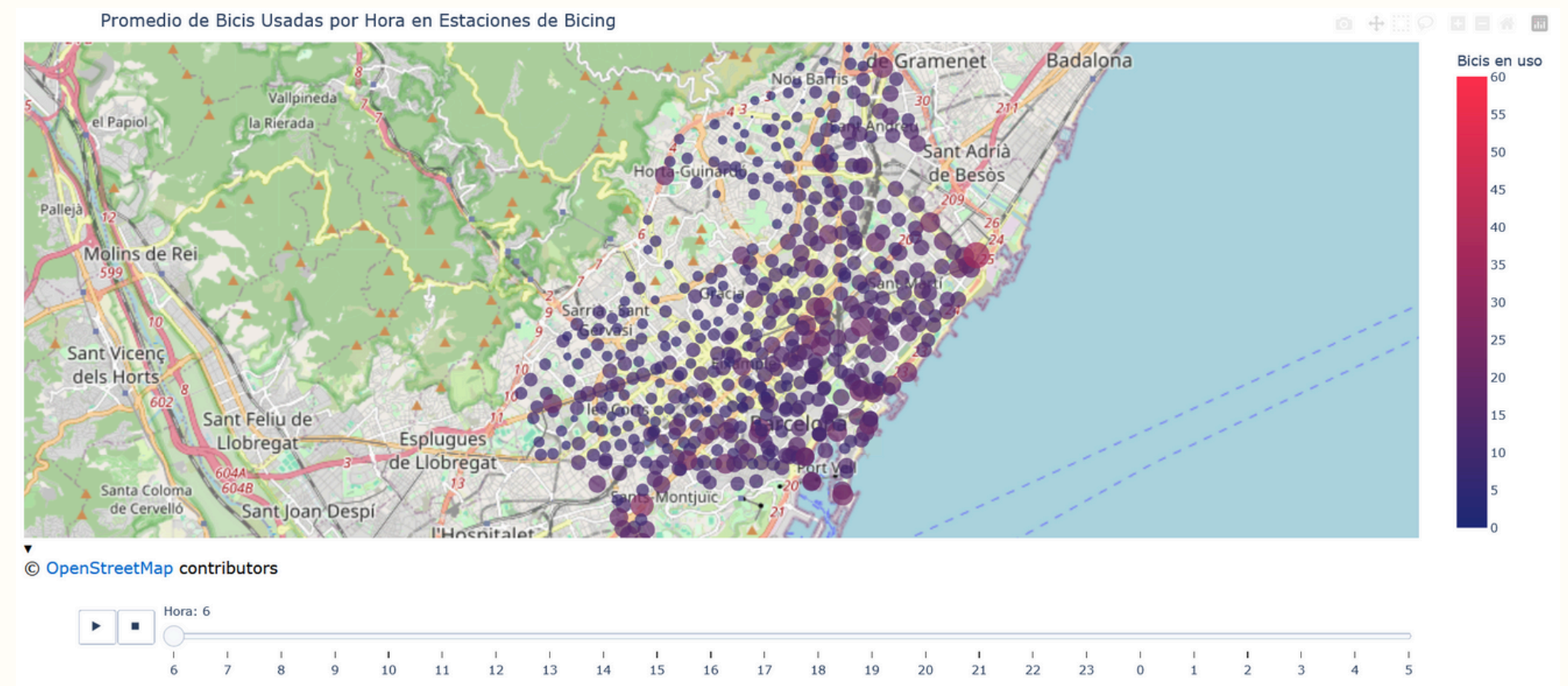
Fuente de datos Bicing 2025



→ Datos:

"Inicialmente, se planteó la generación de mapas de calor con un filtro anual, para poder observar las diferencias entre años. No obstante, debido a limitaciones técnicas, se optó por calcular un promedio del uso de las estaciones por hora del día para la totalidad del conjunto de datos."

Para graficar el uso de las bicicletas se tuvo que asumir un supuesto sobre el dataset. Hay una variable que cuenta cuantas bicicletas faltan en la estación. Se ha asumido que las bicicletas que faltan en la estación es debido a que se están usando, pero puede deberse también al hecho de que no se han cargado previamente.

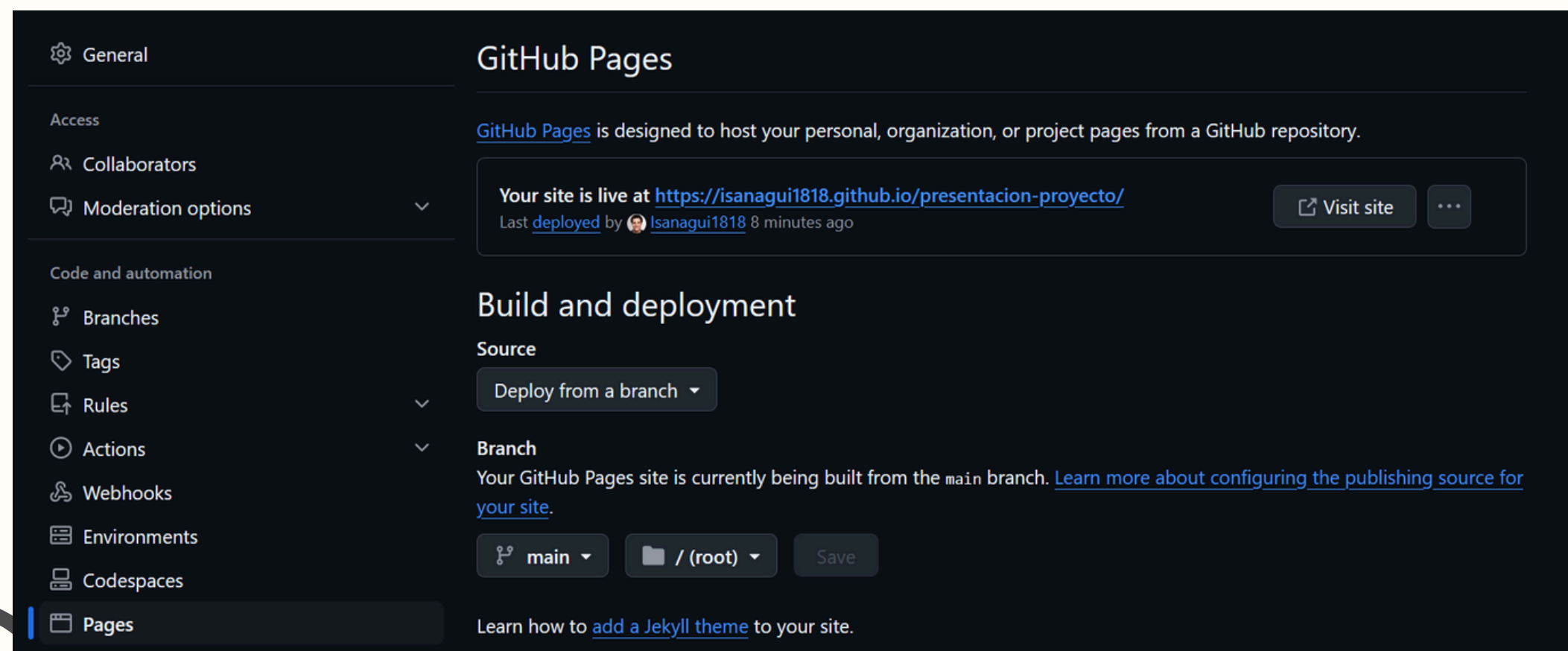




EL PROYECTO

Para poder usar los gráficos interactivos de **plotly** hubo que afrontar varios retos:

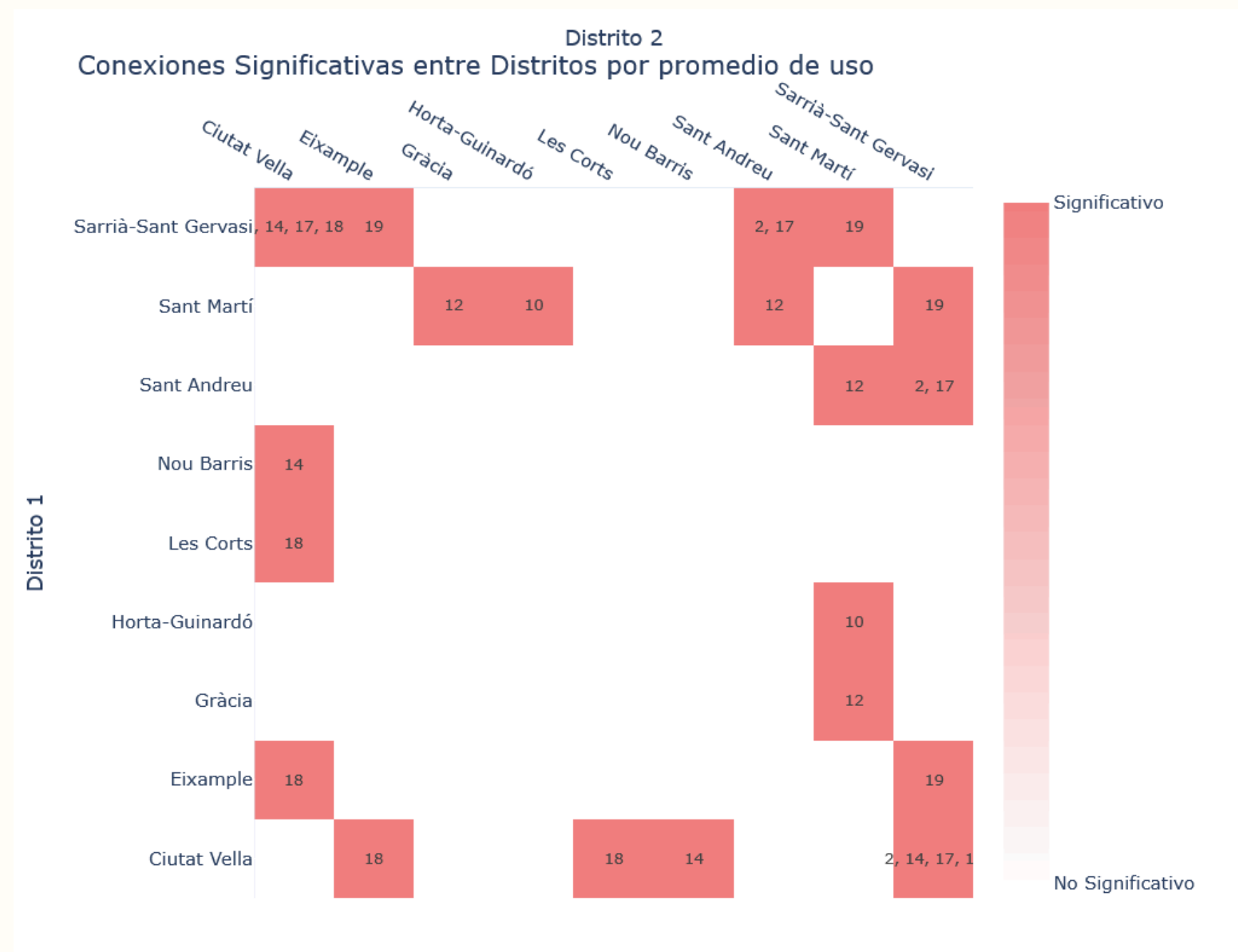
- Exportación de los gráficos en **.html**.
- Configurar el repositorio de **GitHub Page** para poder abrir el archivo **.html** desde el repositorio.



→ Datos:

Prueba gráfico de correlación de la prueba Dunn por pares

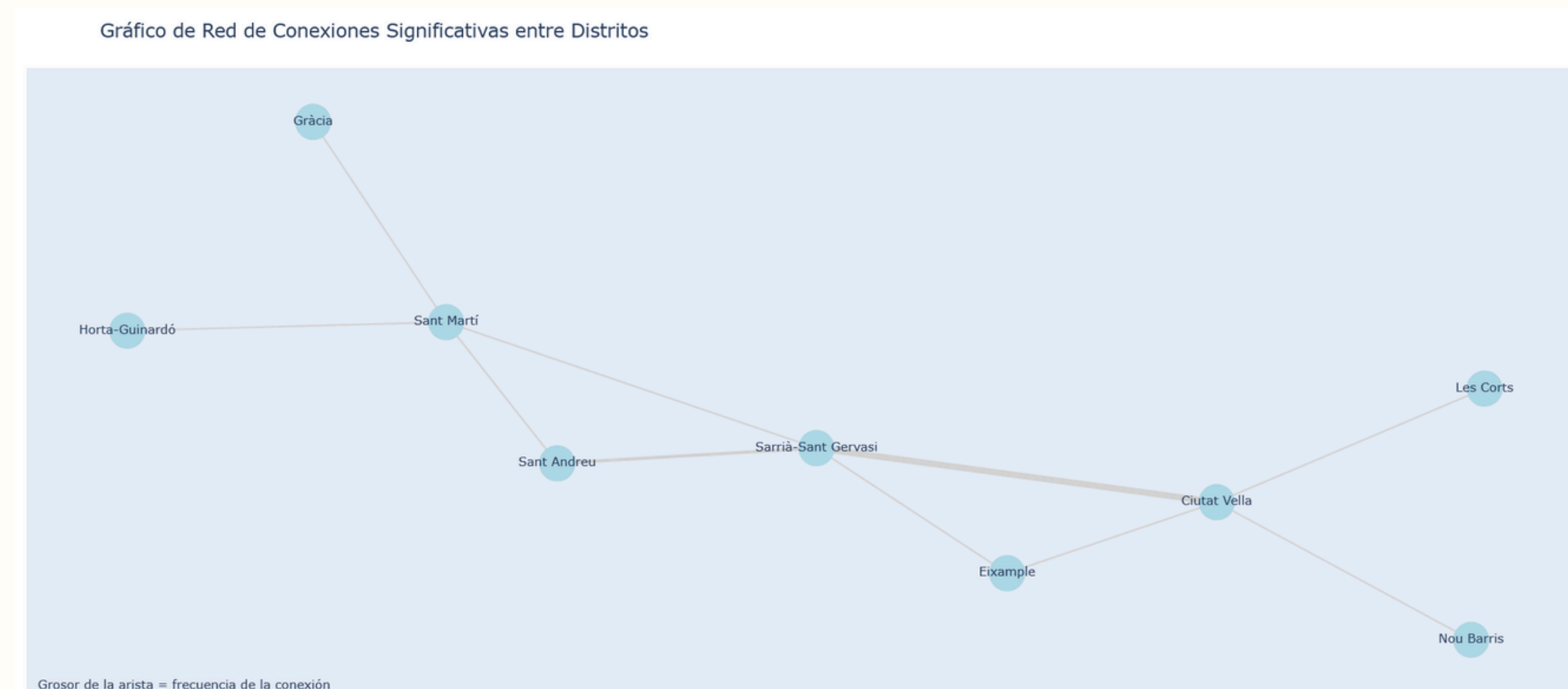
Advertencia: Estos cálculos no se hicieron con todo el dataset, se hicieron con un sample de 10k de datos del año 2024



→ Datos:

Gráfico de red mostrando la relación entre los distritos de la prueba Dunn por pares

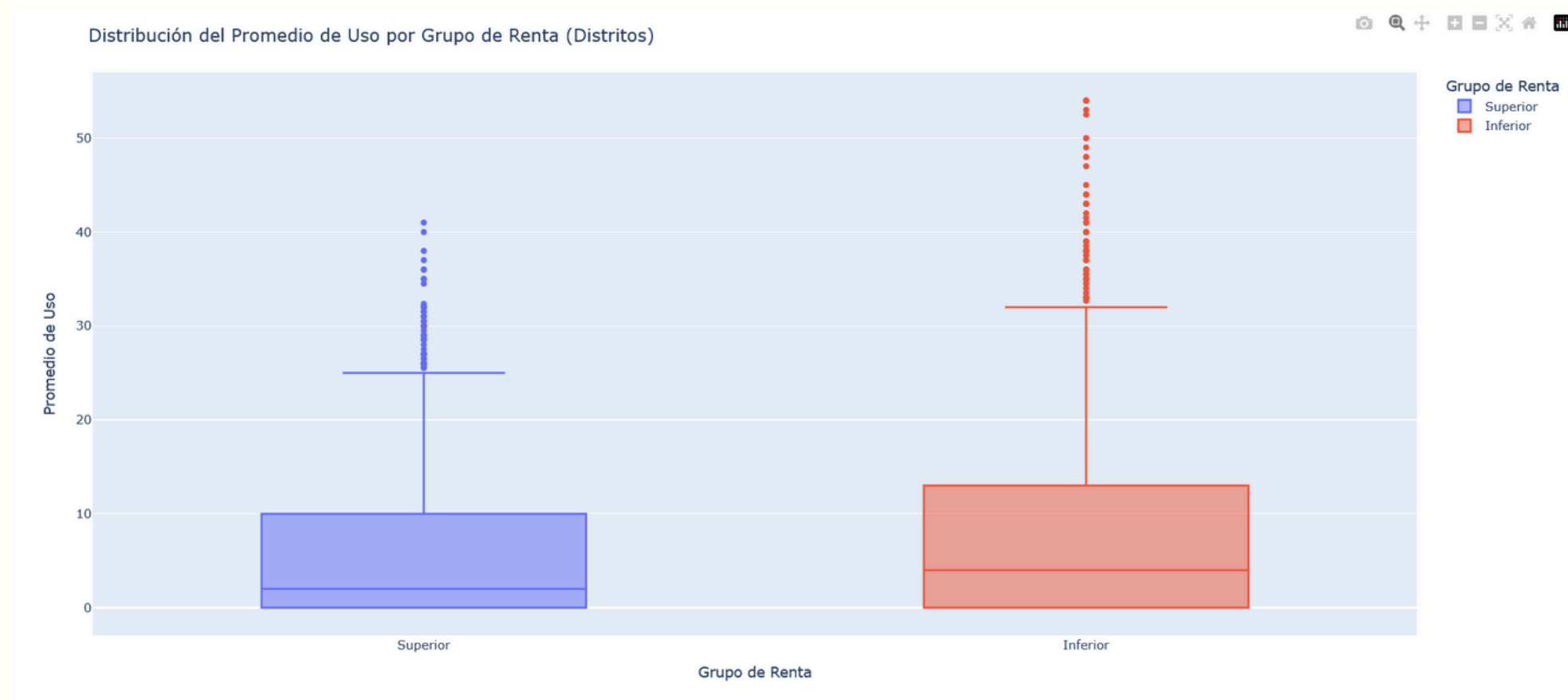
Advertencia: Estos cálculos no se hicieron con todo el dataset, se hicieron con un sample de 10k de datos del año 2024



→ Datos:

Boxplot prueba Dunn por pares usando los distritos, el uso promedio de las bicis y diviendo la renta promedio de los barrios a través de la mediana en dos grupos (superior e inferior)

Advertencia: Estos cálculos no se hicieron con todo el dataset, se hicieron con un sample de 10k de datos del año 2024



Conclusiones

- Podemos afirmar H_1 , tras la realización de la prueba **ANOVA (Kruskal-Wallis)** se observa diferencias entre los distintos grupos.
- Posteriormente se ha realizado la prueba **Dunn por pares** para poder observar dichas diferencias.
- Debido al elevado número de comparaciones múltiples (hecho que incrementa la posibilidad falsos positivos) **no se puede dar un nivel de confianza del 95,5%** (teorema de Bonferroni)



(In)conclusiones

Análisis que faltaron por realizar en la investigación:

- Analizar la diferencia entre bicis mecánicas y eléctricas.
- Analizar la altitud de las estaciones del Bicing y observar la relación con el tipo de bicicletas y la geolocalización.