



Segmentación de estaciones de alquiler de bicicletas mediante clusterización

Julio Avila

Guillermo Furlan



ECOBICI



ECOBICI

Ecobici es el sistema de bicicletas públicas de la Ciudad de México que ha integrado a la bicicleta como parte esencial de la movilidad.

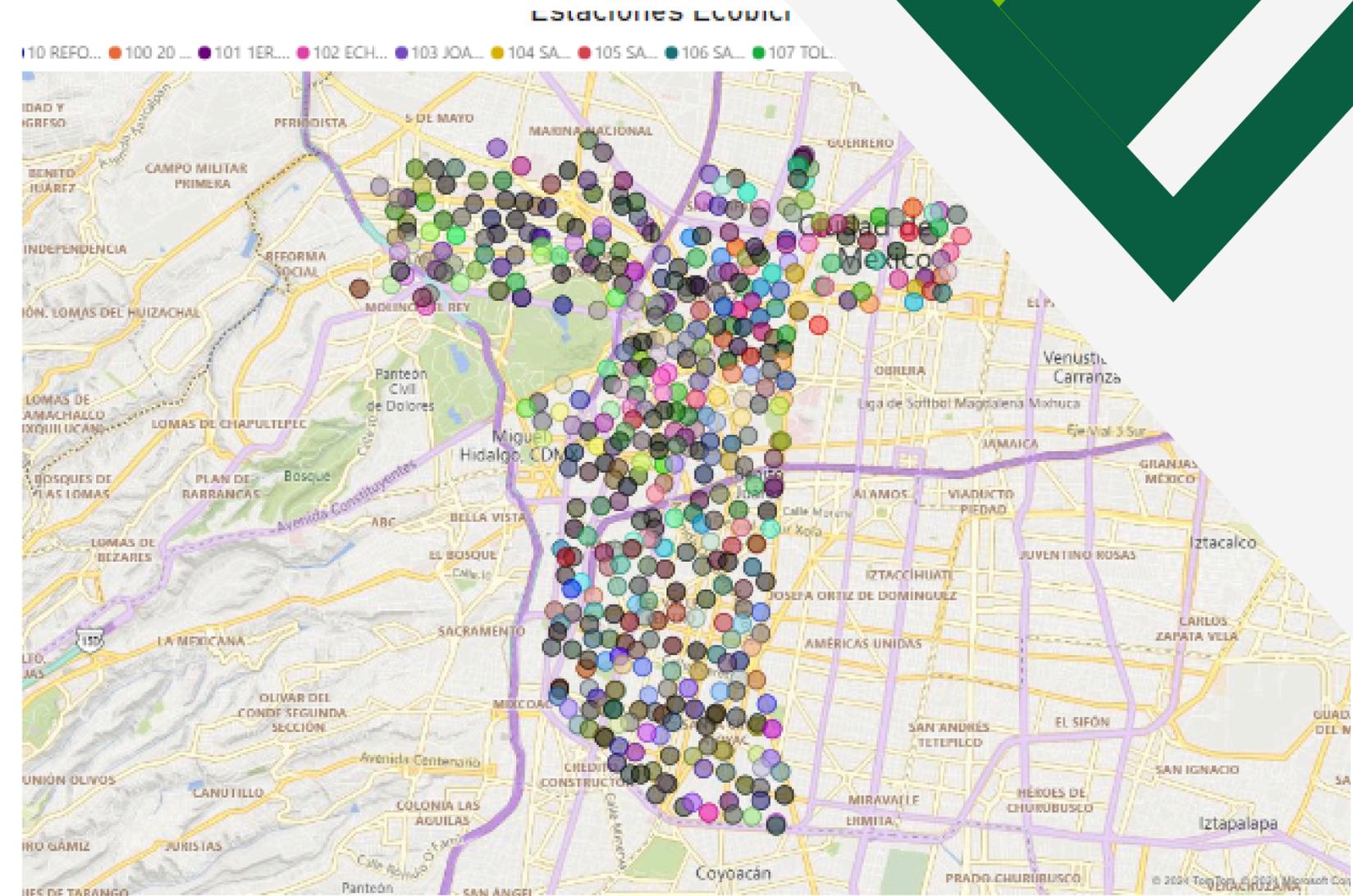
Permite a los usuarios registrados tomar una bicicleta de cualquier cicloestación y devolverla en la más cercana a su destino en trayectos ilimitados de 45 minutos.

Los datos

Las bases de datos consistían en información histórica de los viajes con los siguientes datos:

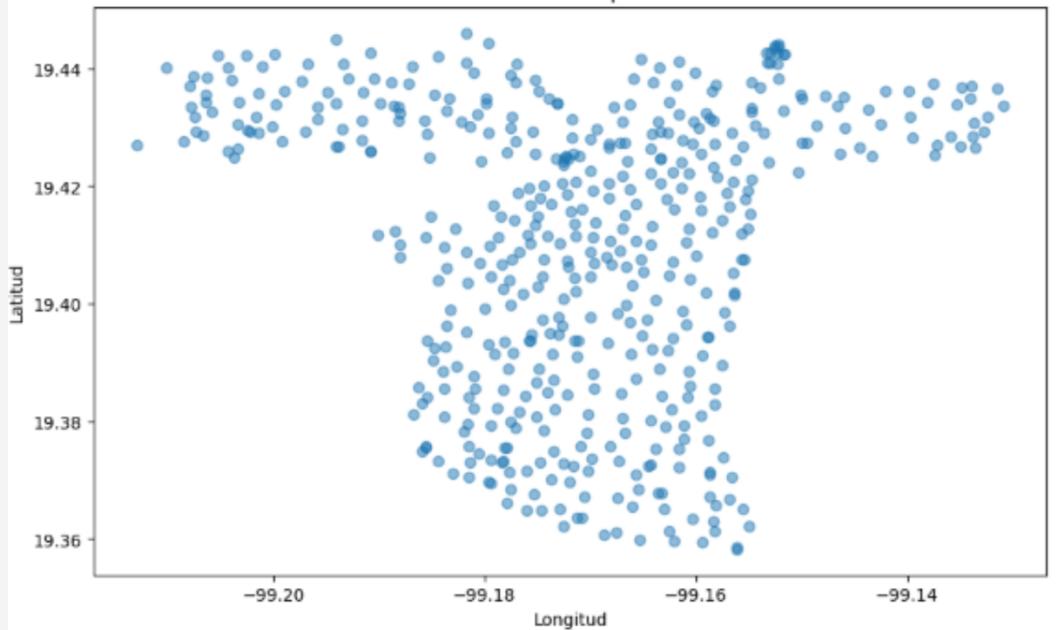
- Genero_Usuario
- Edad_Usuario
- Bici
- Ciclo_Estacion_Retiro
- Fecha_Retiro
- Hora_Retiro
- Ciclo_Estacion_Arribo
- Fecha_Arribo
- Hora_Arribo

También se proporcionó una base con información de la locación de las estaciones

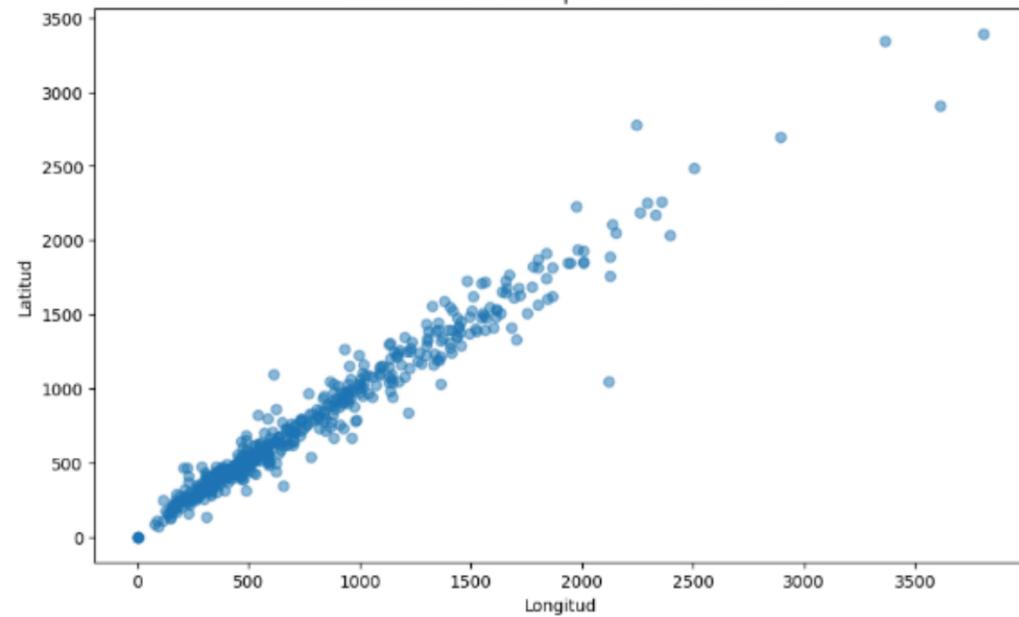


Analisis exploratorio

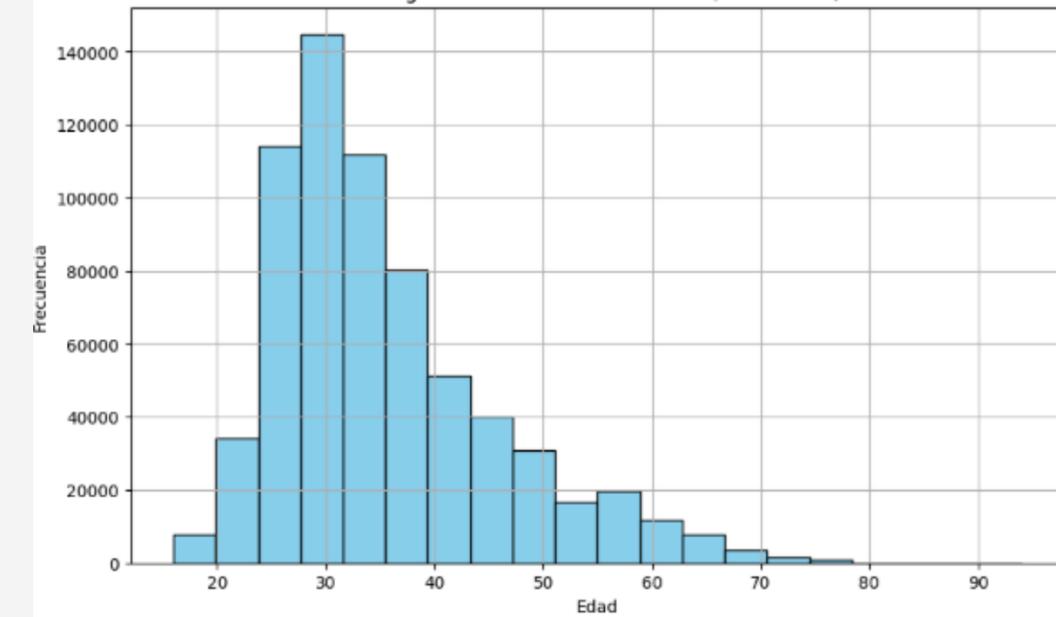
Distribucion de datos por distancias



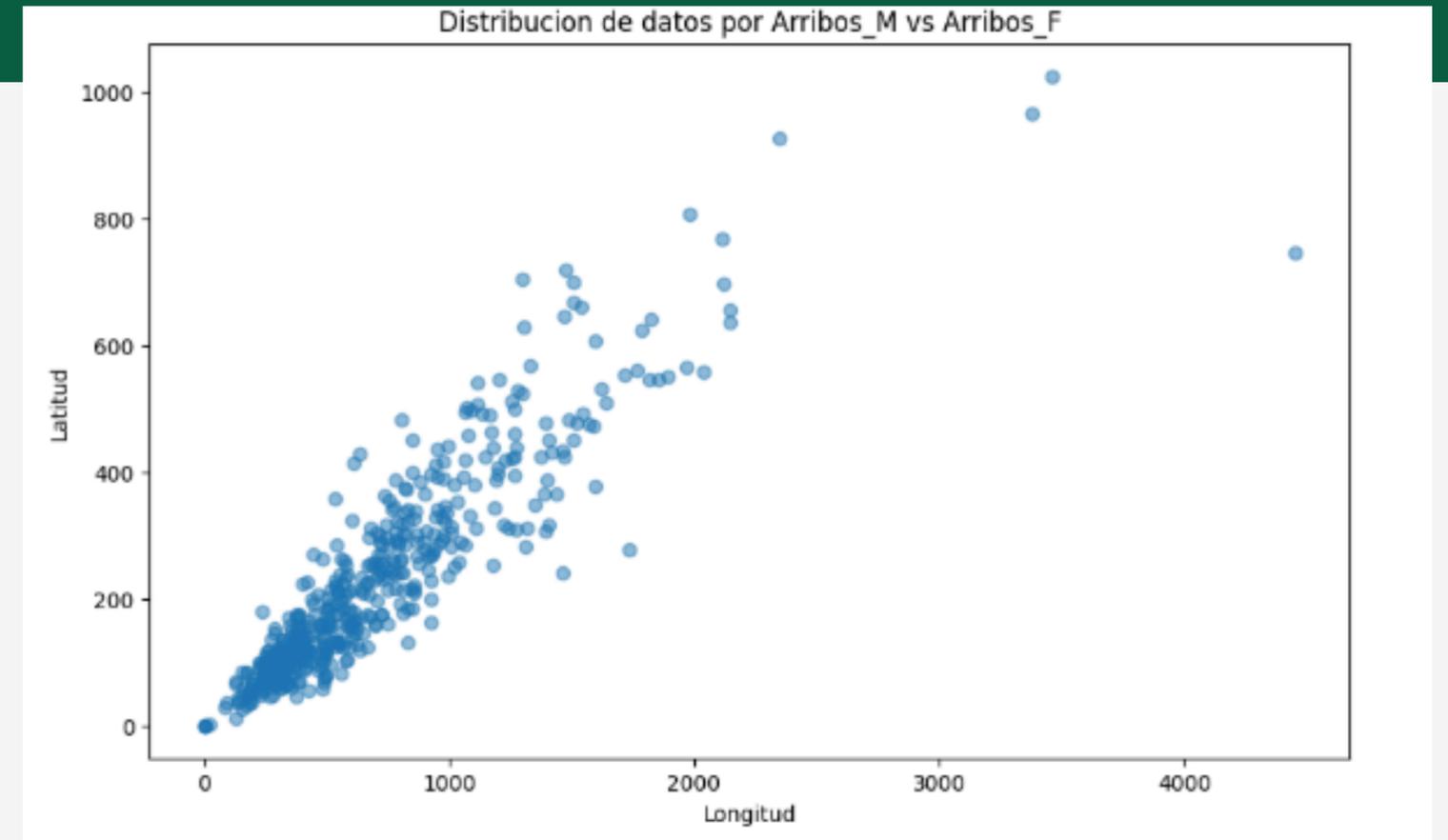
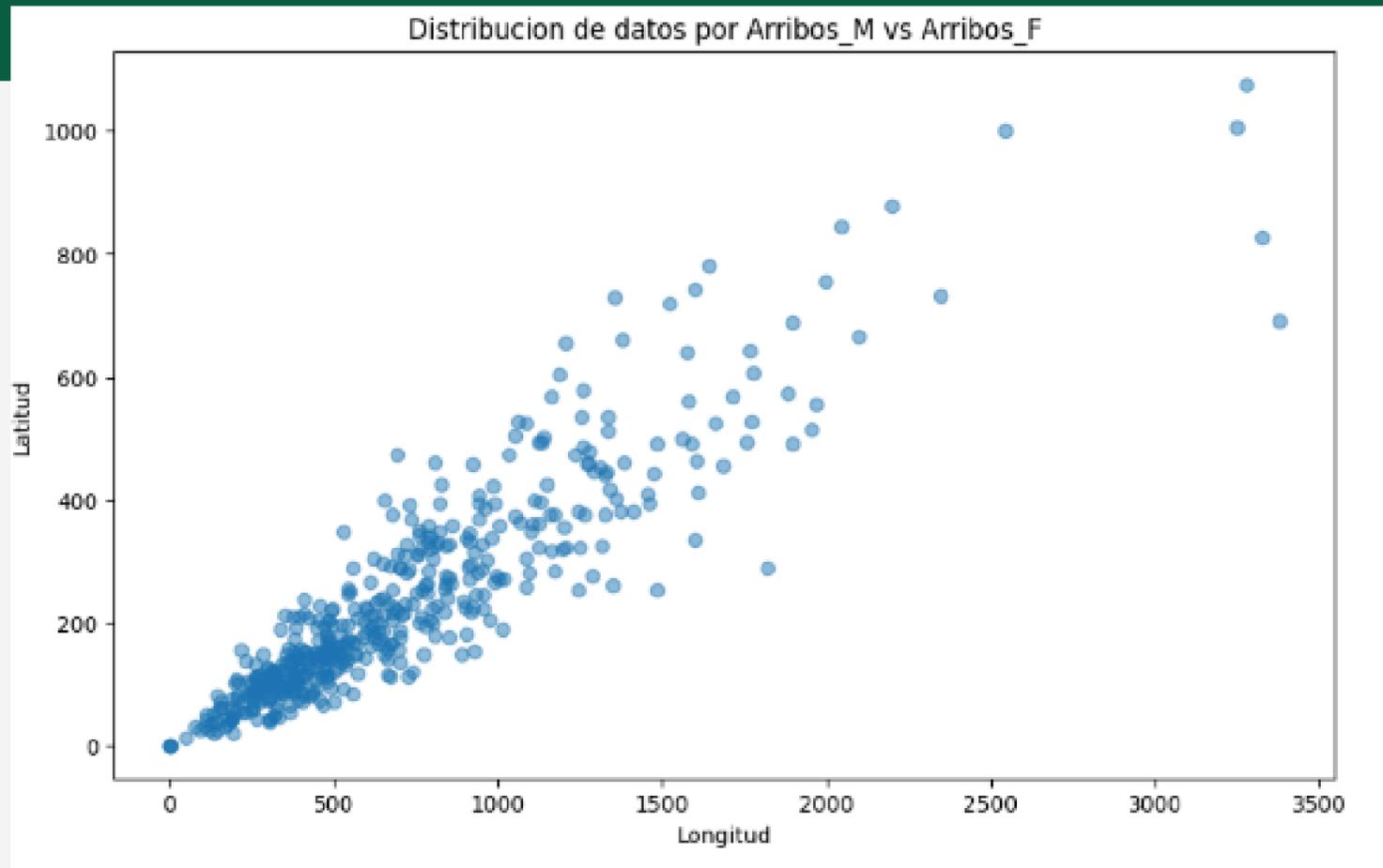
Distribucion de datos por Arribos vs Retiros



Histograma de Edades de Usuarios (Enero 2020)



Analisis exploratorio



Hipótesis



La posición geográfica afecta la afluencia de personas de la estación.



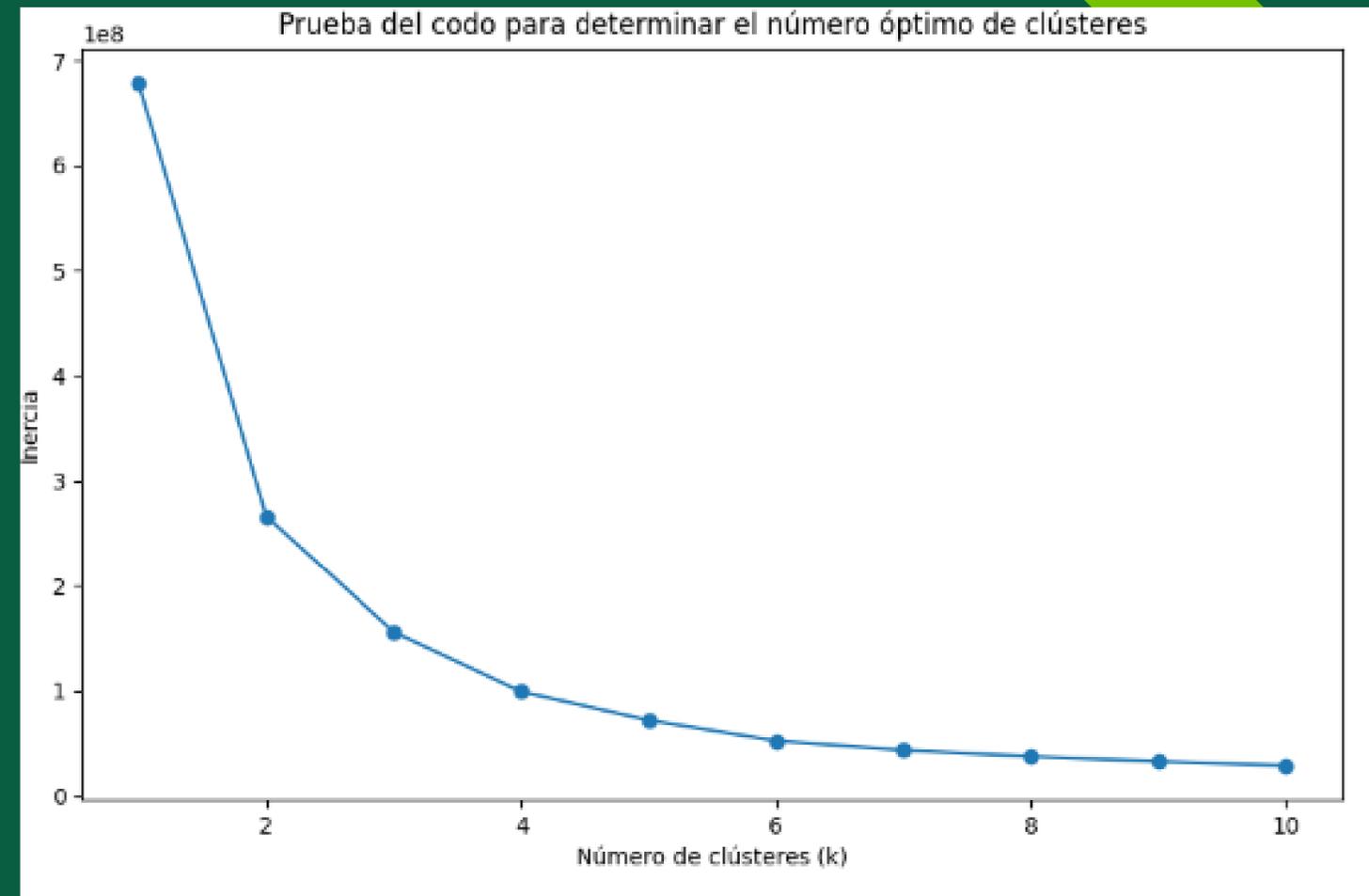
La temporada del año afecta la afluencia de uso de bicicletas en 2022



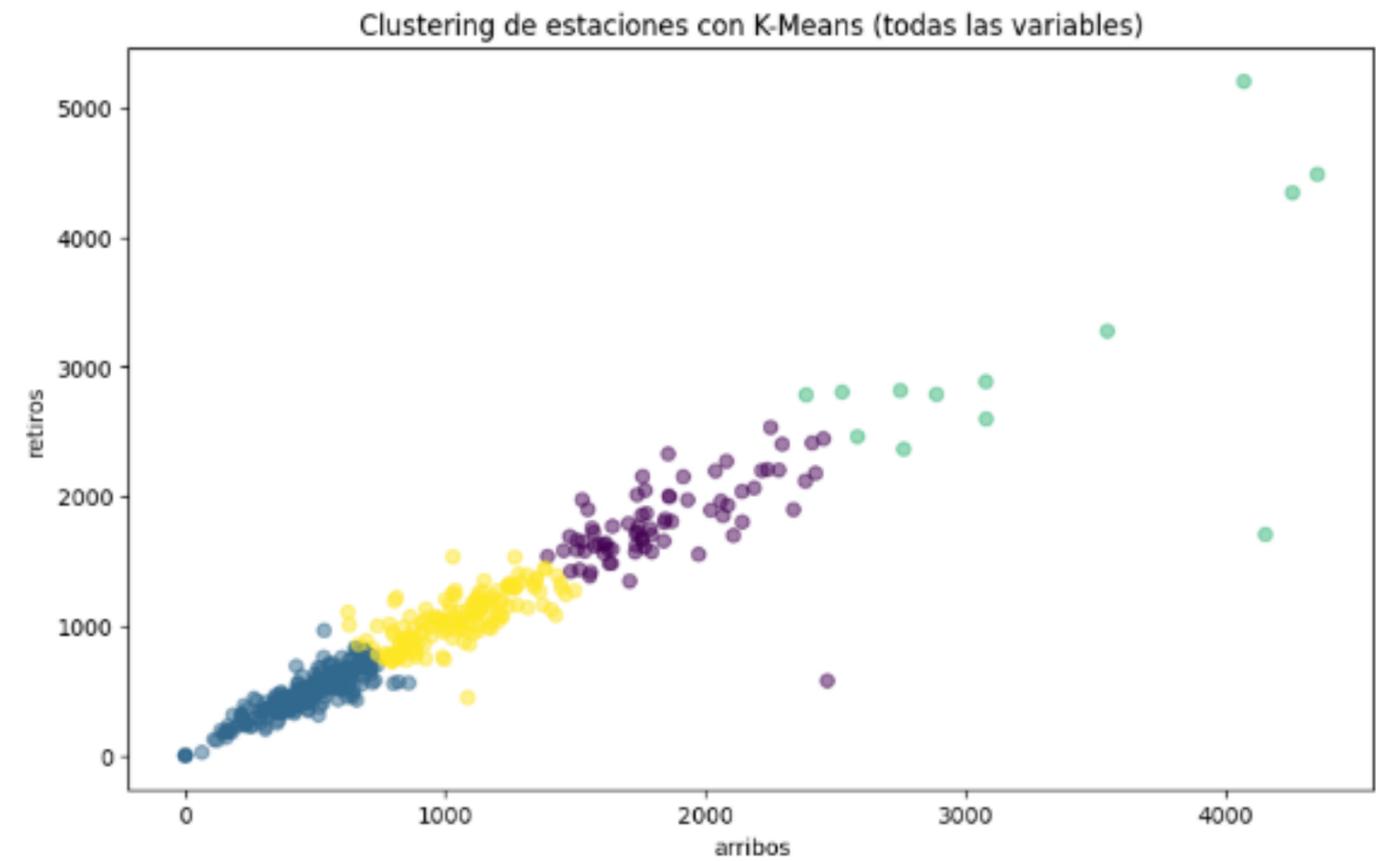
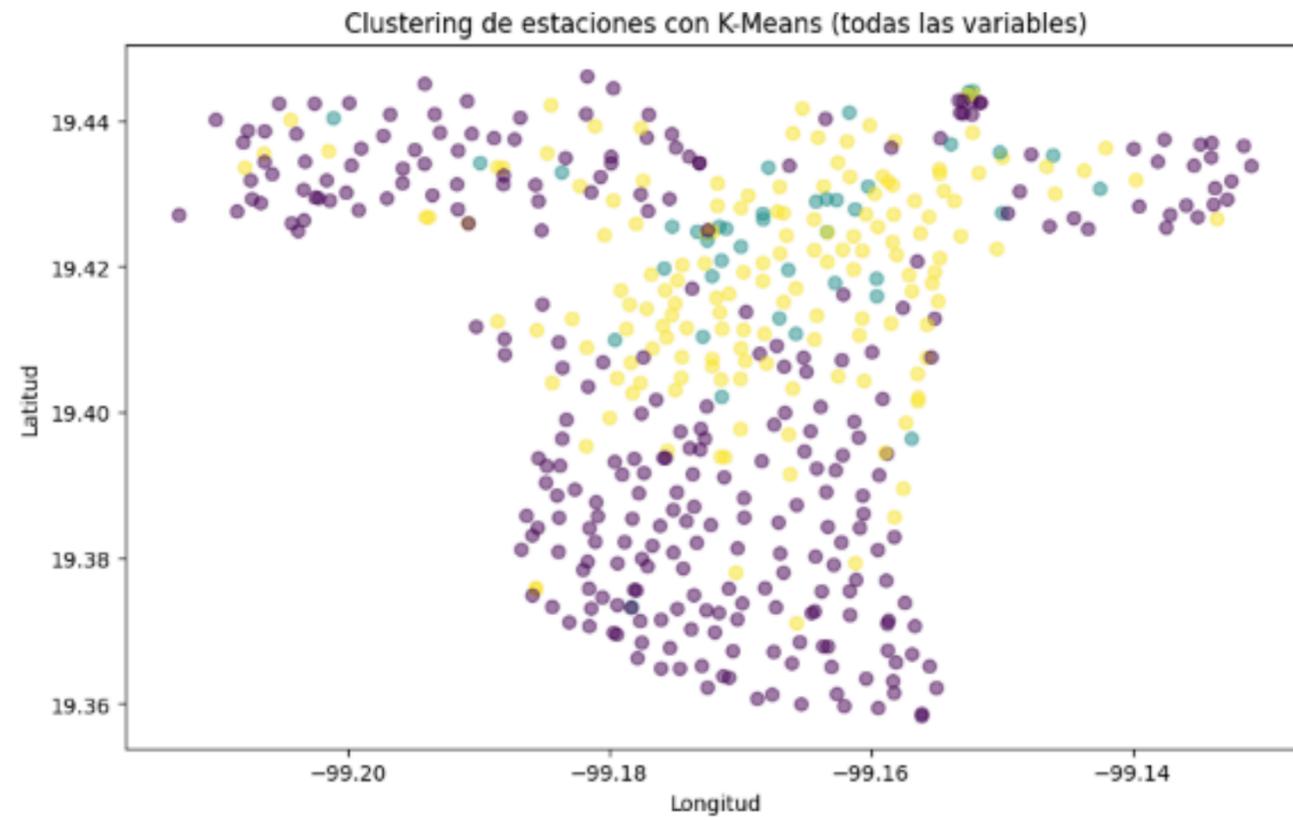
Es posible identificar grupos claros que separen las estaciones por su afluencia.

Metodología

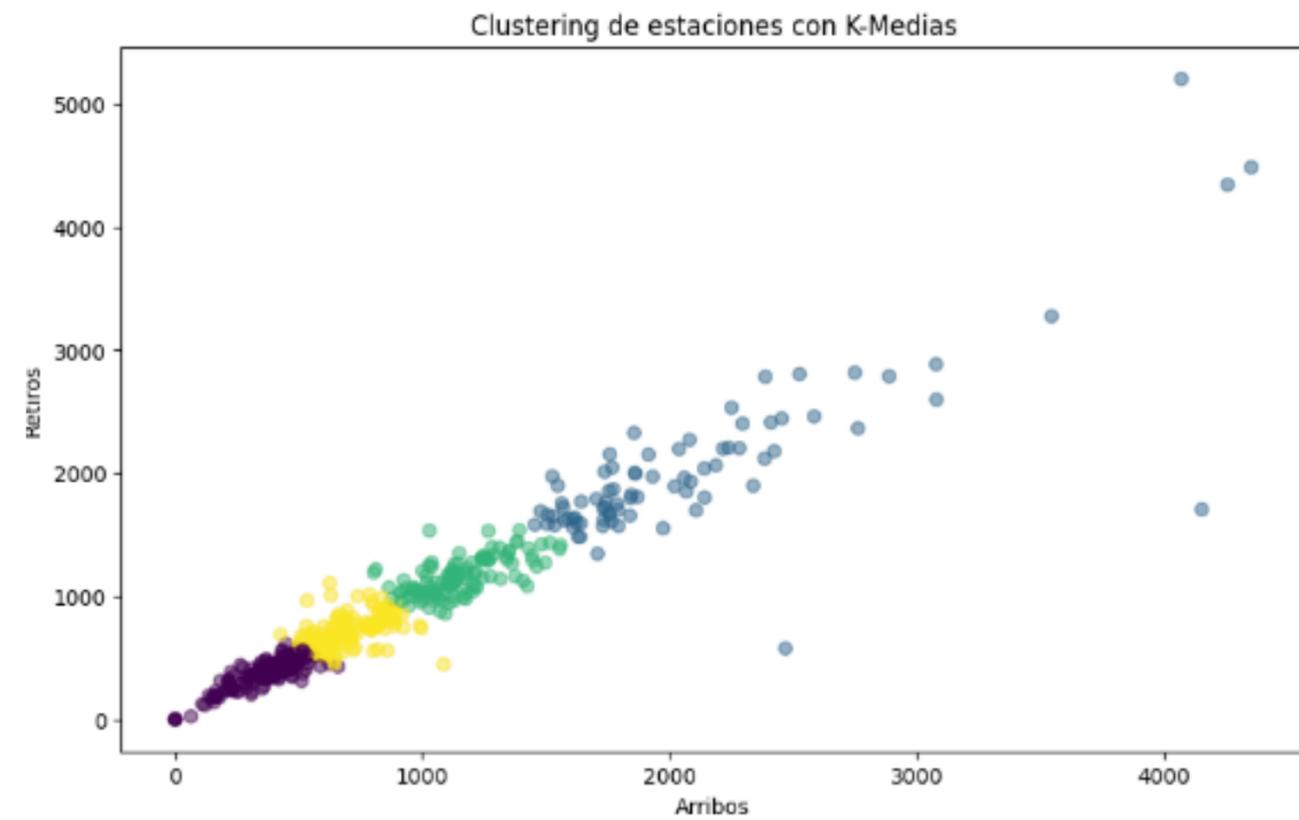
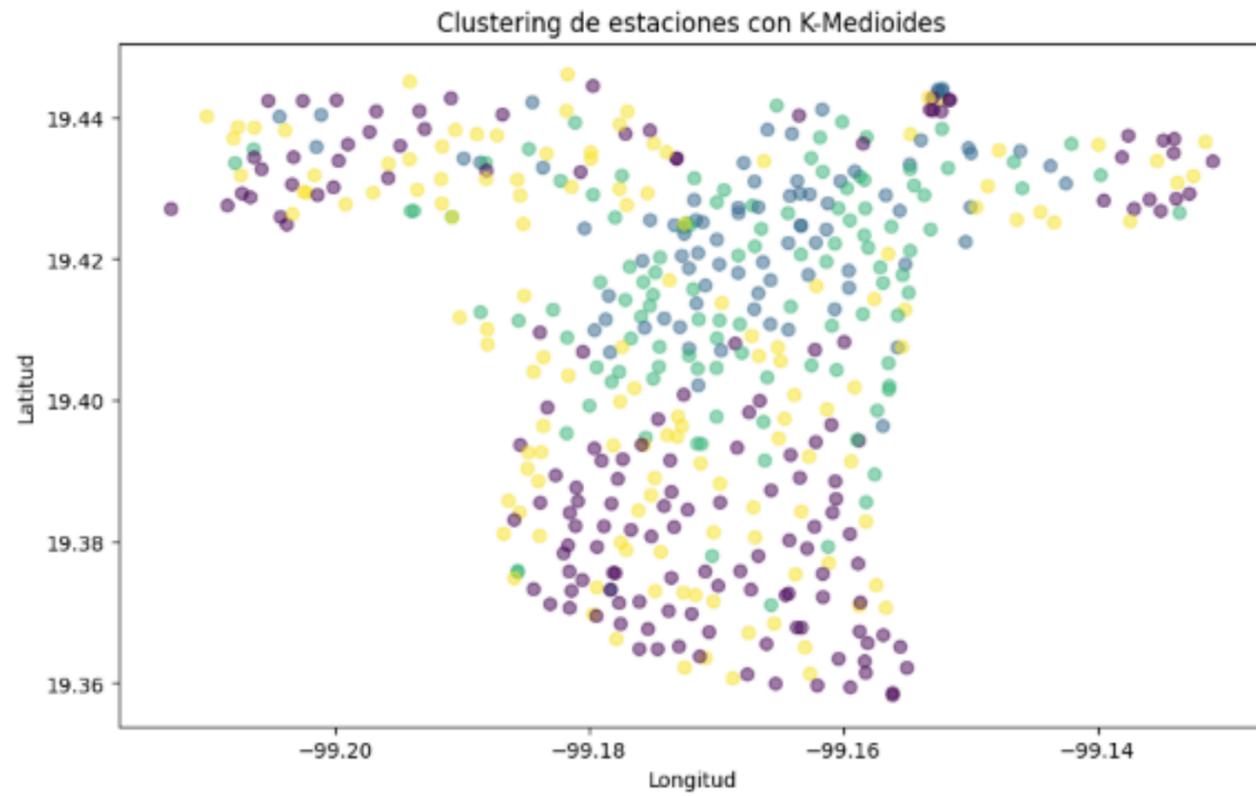
Tomando como muestra la información de enero 2020, mediante cinco diferentes formas de clusterización (K-means, K-mediodes, Mezclas Gaussianas, OPTICS, Jerarquicos) se buscó analizar las maneras en que se segmentaban los grupos para identificar patrones de comportamiento que permitieran agrupar las diferentes estaciones y generar un modelo que pudiera analizar, segmentar y brindar información de las estaciones para cualquier otro mes que se busque



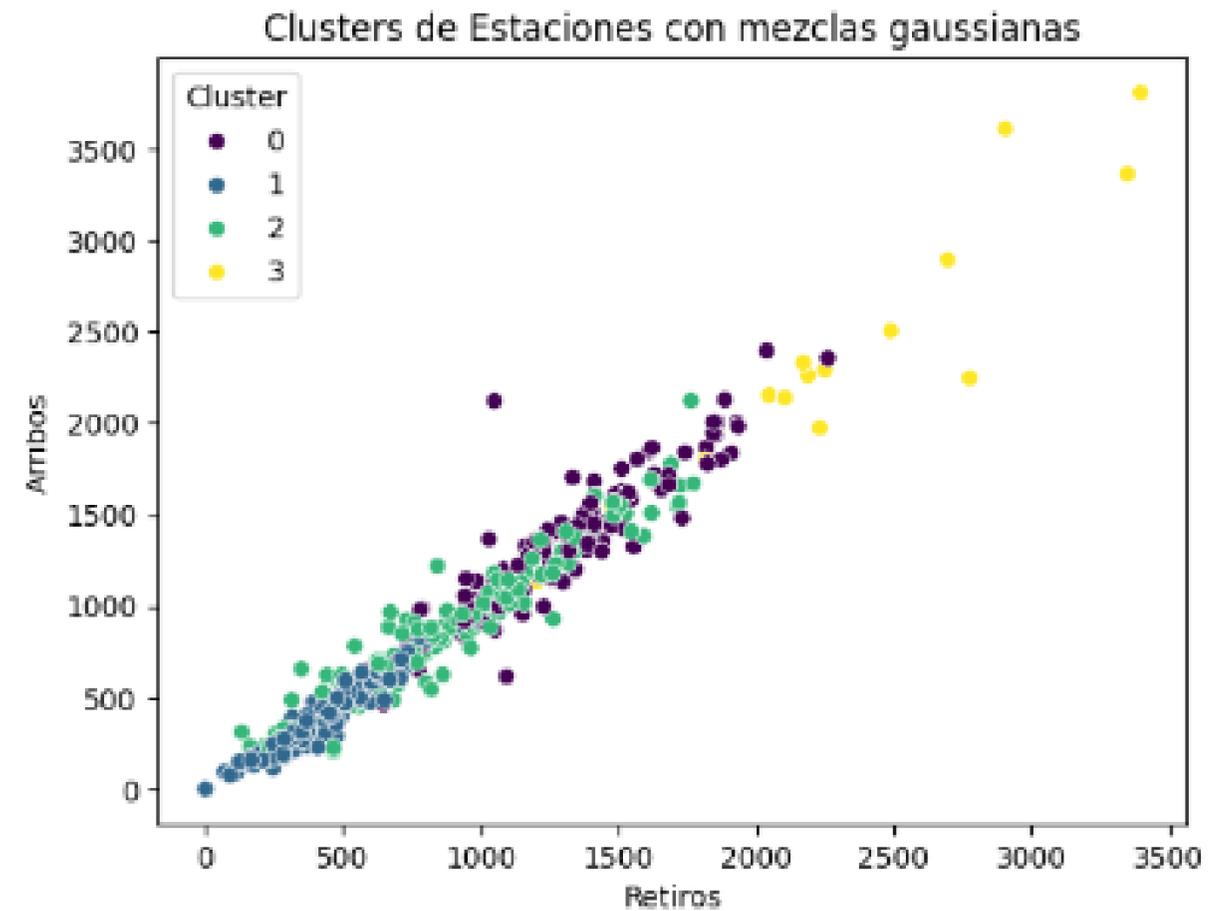
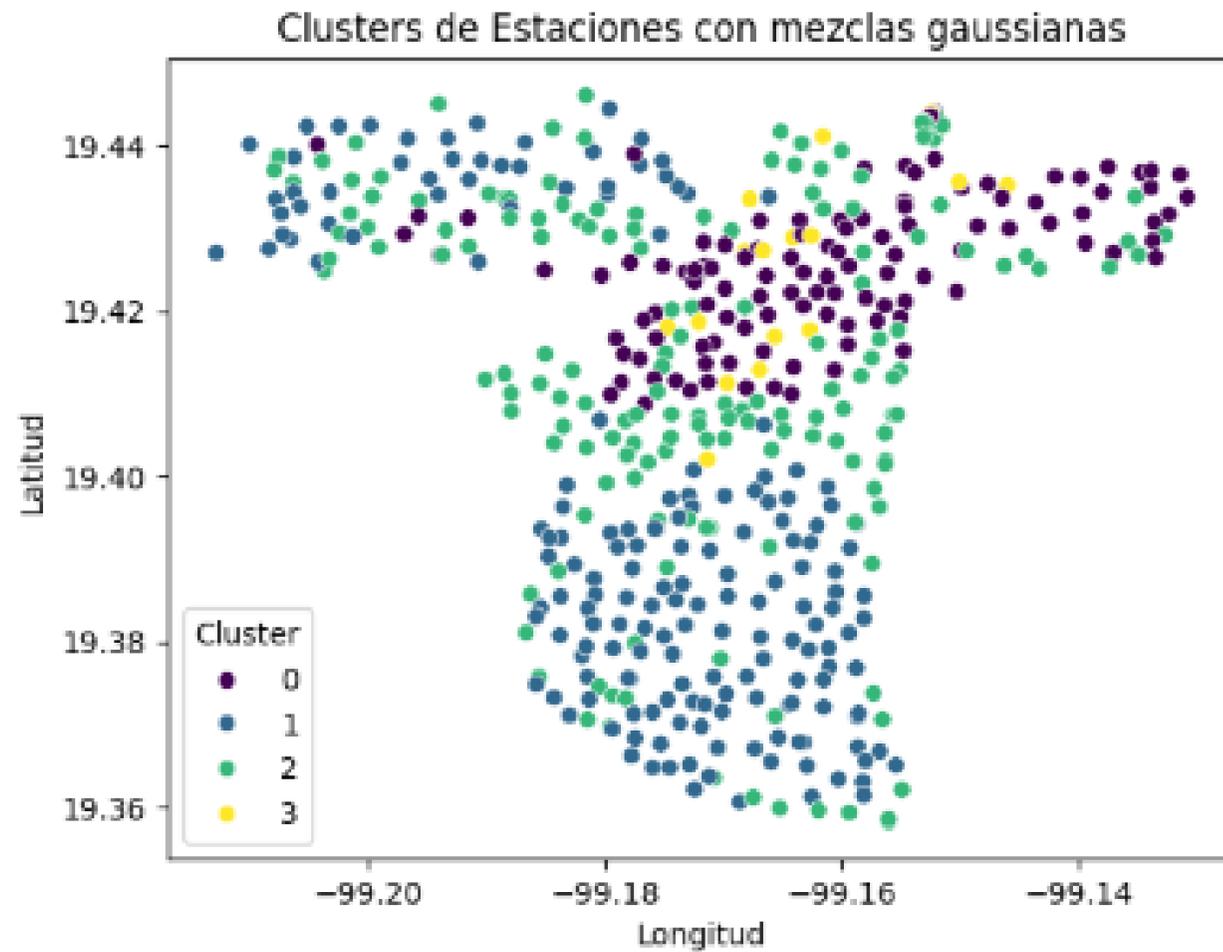
K-Means



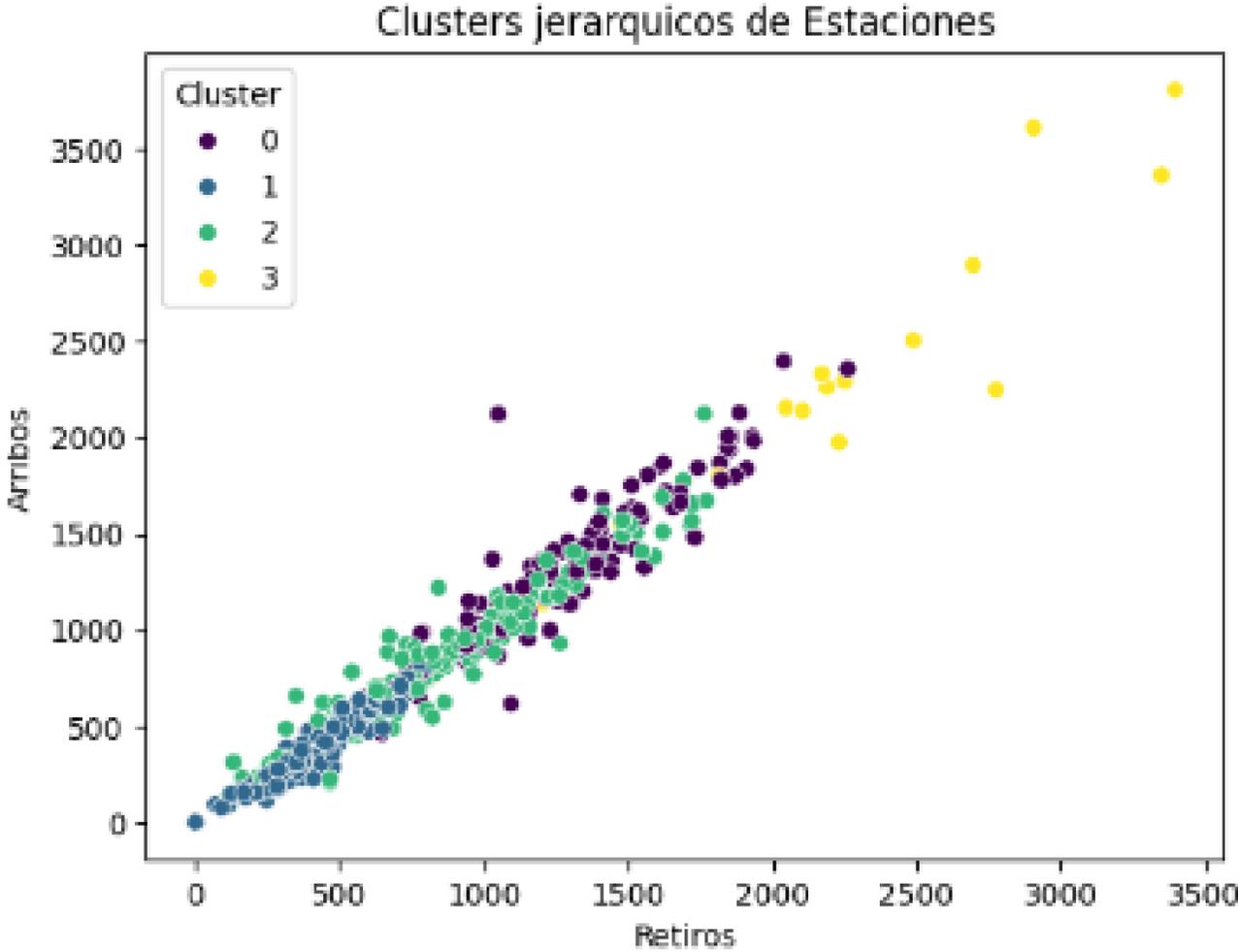
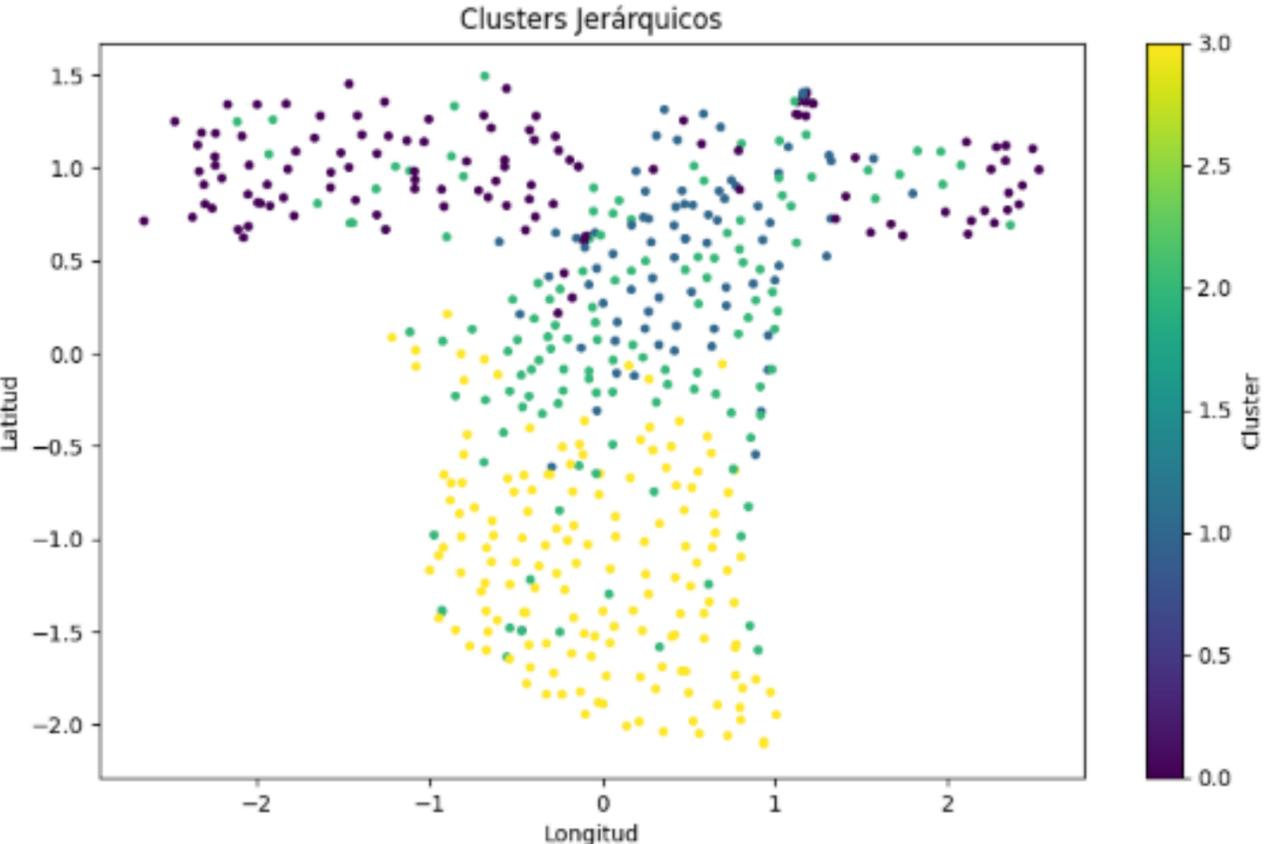
K-Mediodes



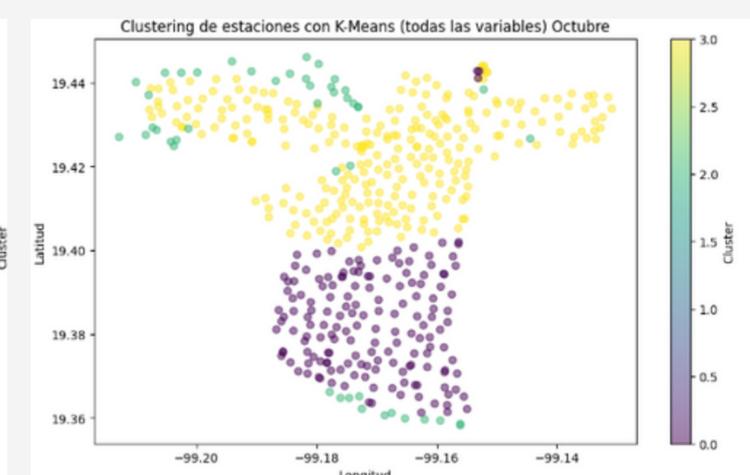
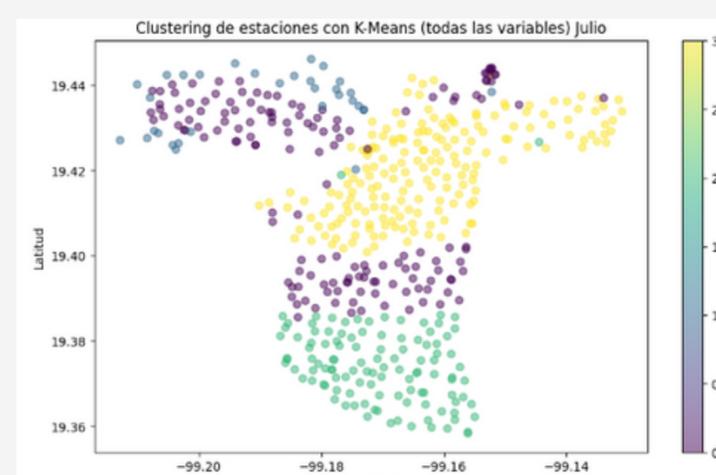
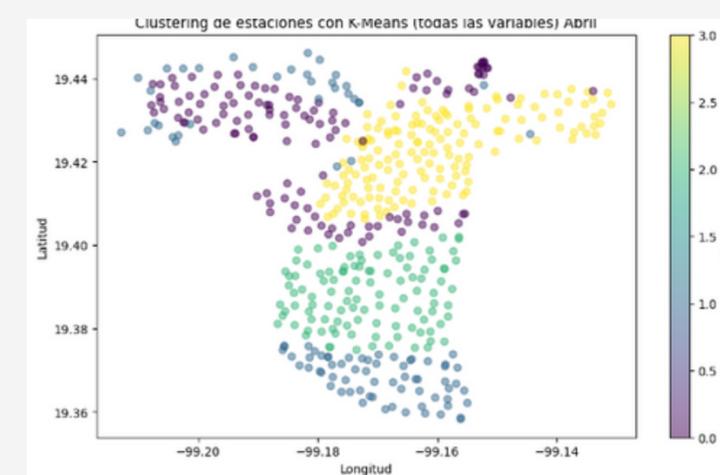
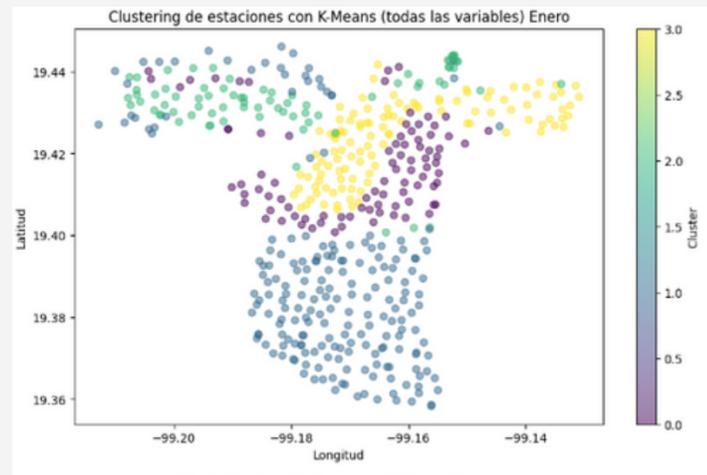
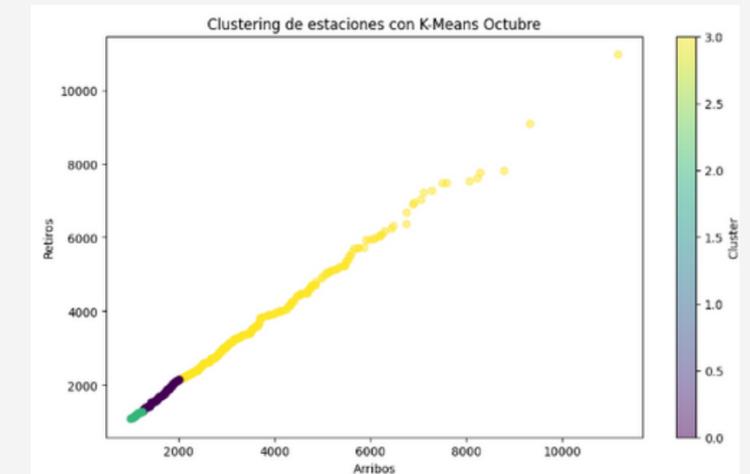
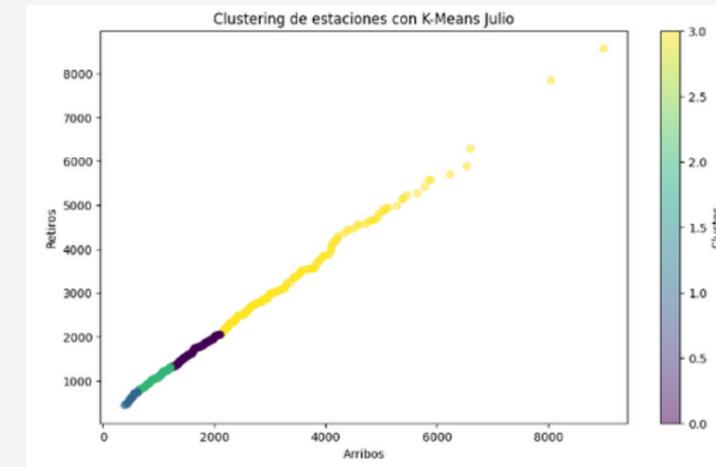
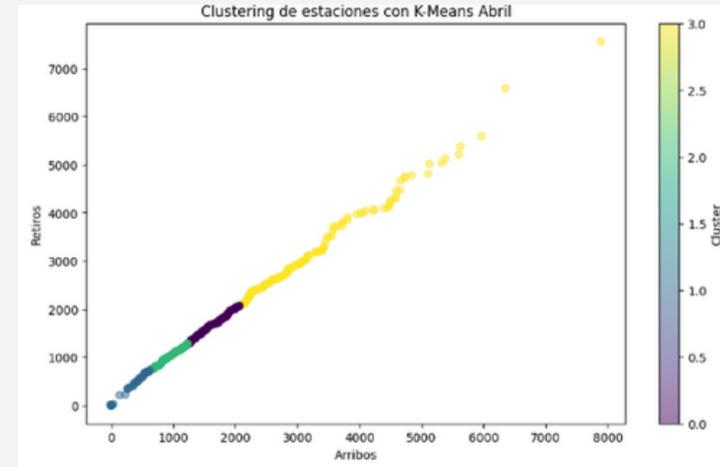
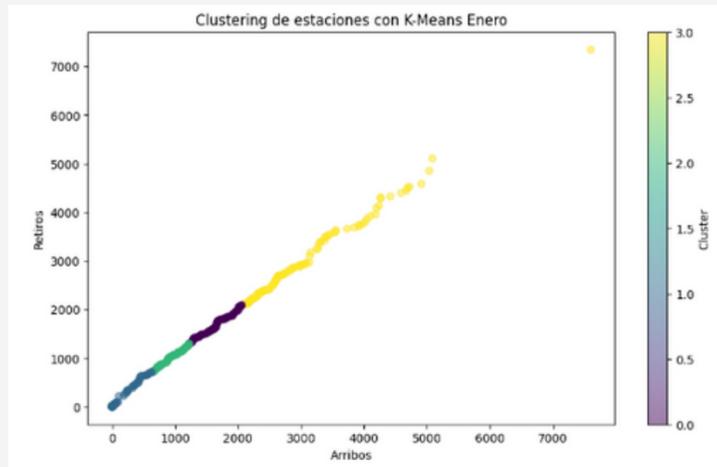
Mezclas Gaussianas



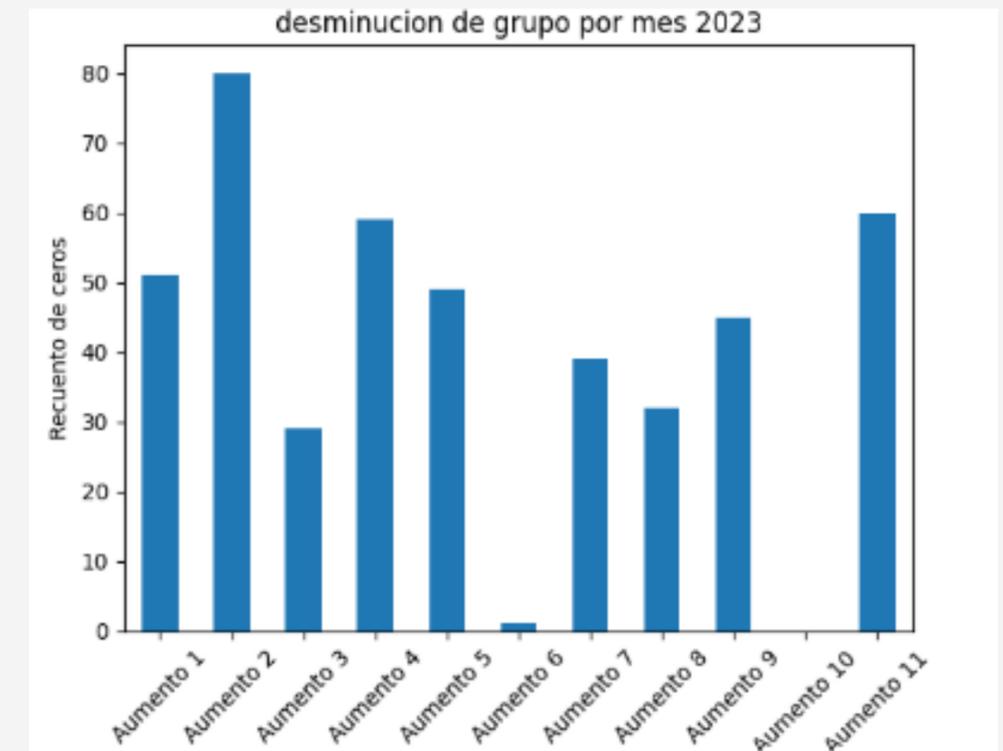
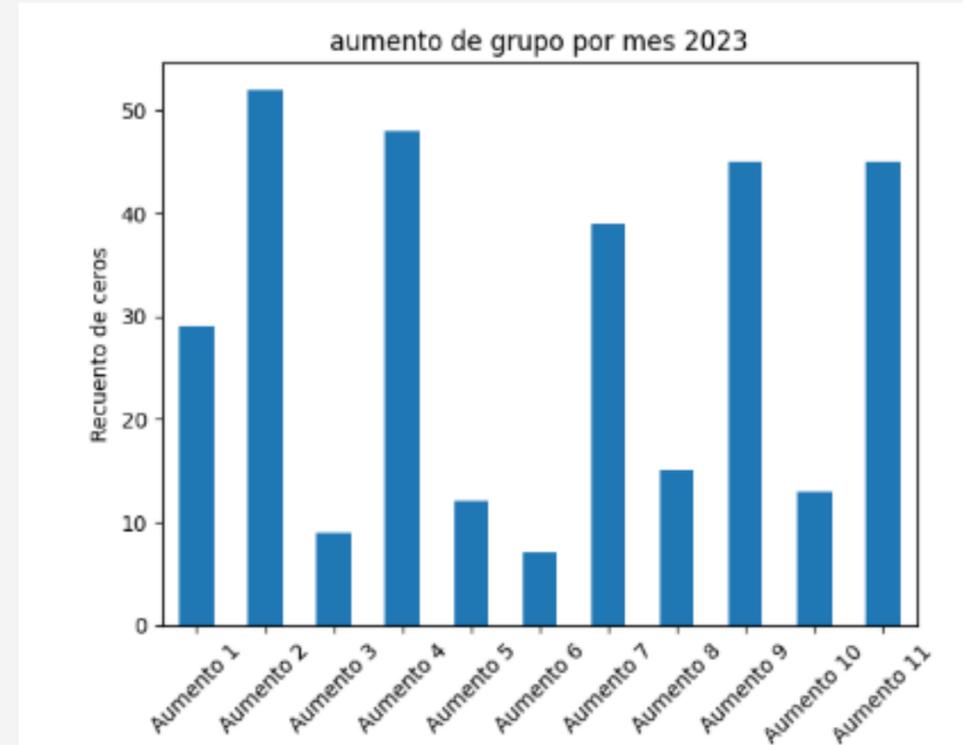
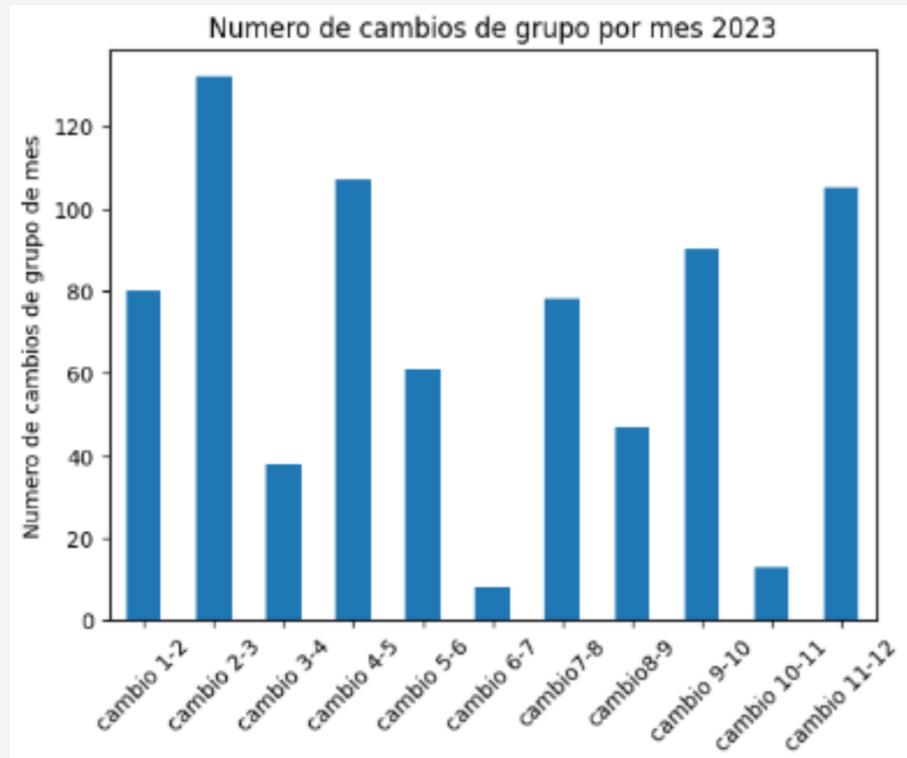
Jerarquicos



Modelo K-means utilizado en los meses de 2023



Modelo K-means utilizado en los meses de 2023





Conclusiones

- El cluster K-Means fue el que dio mejores resultados para esta información.
- La posición geográfica si afecta la afluencia de personas en las estaciones
- La temporada del año si afecta el uso de las estaciones
- Si es posible categorizar las estaciones de biciletas en grupos claros según afluencia

Conclusiones y Recomendaciones

Recomendaciones

- Se recomienda a la empresa tener presenta la disribucion para poder aprovechar para mantenimientos
- Tener mayor disponibilidad de bicicletas en la estaciones del centro