

¿Dónde es mejor poner una estación de ECOBICI en la CDMX?

Rodrigo Morales 17027



Introducción

ECOBICI es un servicio de bicicletas públicas de la Ciudad de México. Es una alternativa ecológica de transporte. Inicio operaciones en 2010, con 84 estaciones y 1200 bicicletas. Actualmente tienen 480 estaciones y más de 6800 bicicletas, de las cuales algunas de las estaciones y bicicletas forman parte del sistema de bicicletas eléctricas. Los usuarios registrados de ECOBICI tienen accesos a viajes ilimitados de 45 minutos entre estaciones.

Con esto en mente, se planteó un proyecto que respondiera a la pregunta de donde es la mejor ubicación para otra estación de ECOBICI. La solución contempla 17 propuestas de puntos de interés, elegidos bajo los criterios de distintas variables y fuentes de datos, tales como:

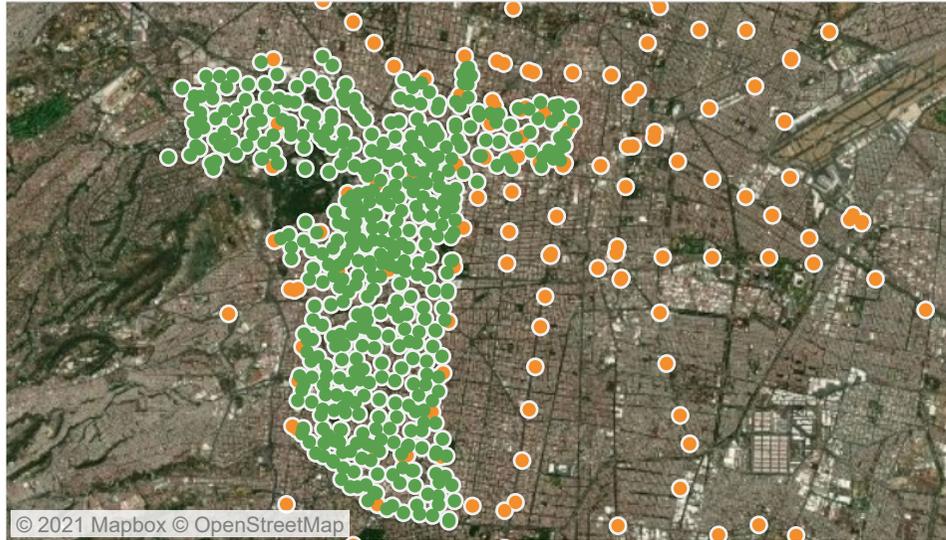
- Población flotante.
- Espacios vacíos entre estaciones.
- Gasto en transporte público.
- Población Flotante Joven
- Áreas donde las estaciones de ECOBICI tienen mayor afluencia de bicicletas.
- Áreas de más presencia de personas por los distintos niveles socio económicos.

Para cada una de las propuestas, se utilizaron tres modelos diferentes:

- Clusterización con KNN medias, con el método: Hartigan -Wong en R
- Árbol de Regresión en R
- Red Neural usando Keras en Python

Fuentes de Datos Públicas

Mapa de las estaciones de ECOBICI y el metro de CDMX



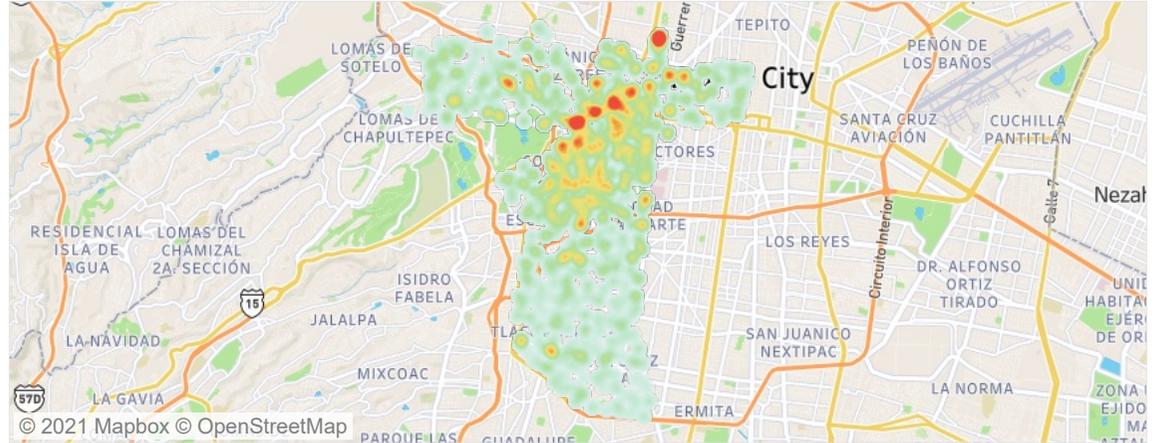
■ ECOBICI ■ METRO

Gasto en Trasnorte público

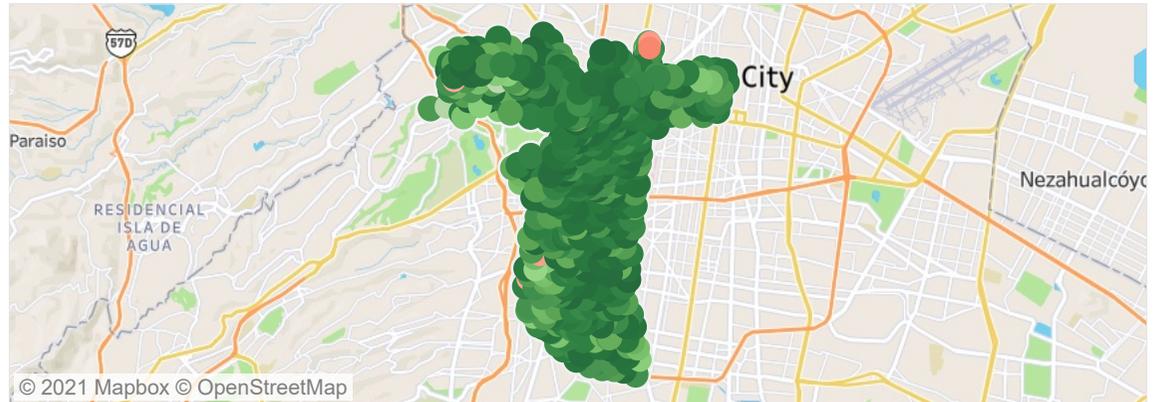
NSE	Población	% Poblacion Gas..	Gasto Per Capita	Gasto Total Sem..
AB	1.27M	73%	\$53	\$49M
C+	1.83M	83%	\$81	\$123M
CDE	5.68M	88%	\$69	\$346M
Total g..	8.78M	82%	\$68	\$519M

Fuente: INEGI

Mapa de calor de las estaciones de ECOBICI sobre el Total de Bicicletas



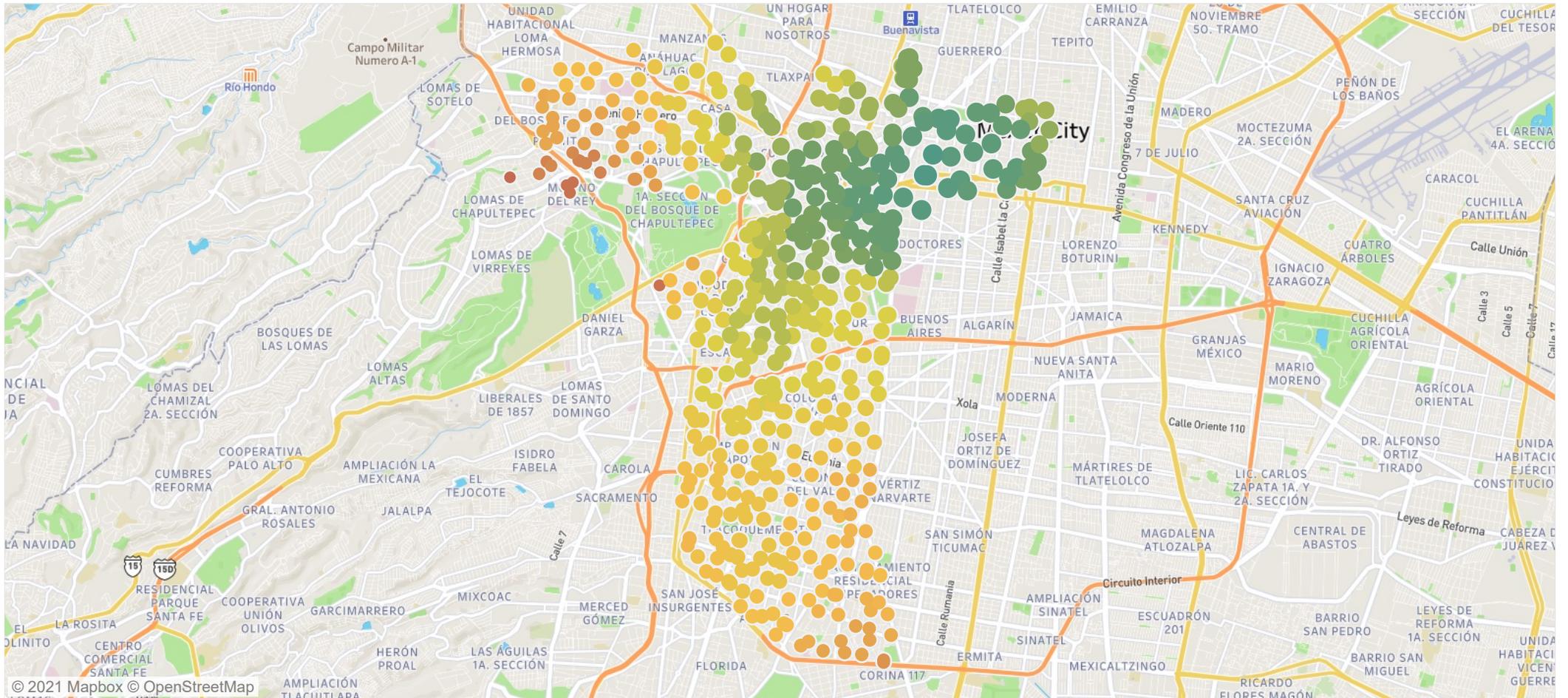
Porcentaje de Rotación por Estaciones de ECOBICI



56.62% 99.96%

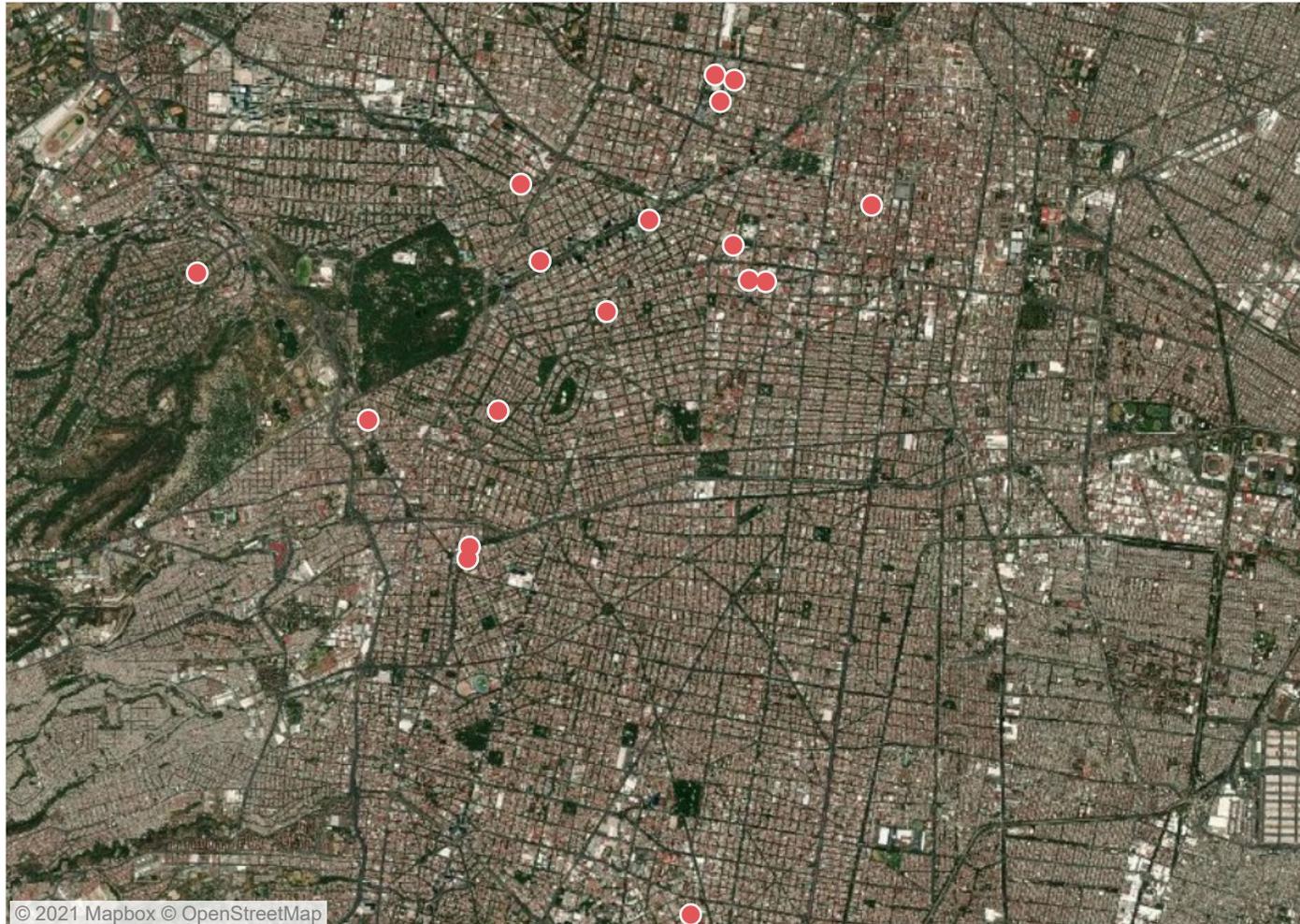
Población Flotante

Métrica de Población
Total



Propuestas de Nuevas Ubicaciones de Estaciones

Mapa de las Propuestas

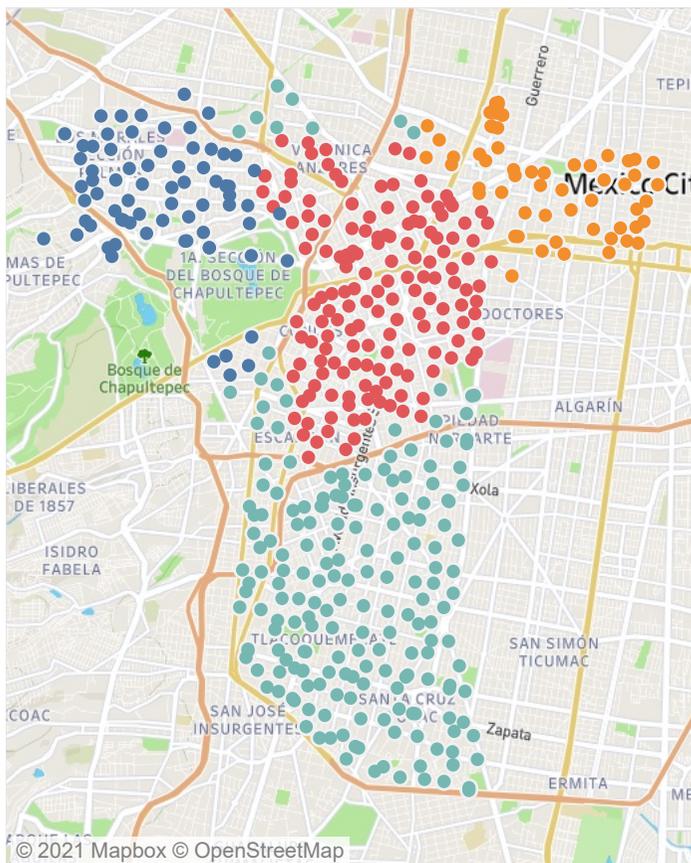


Lista de Propuestas

P_1	Vacio entre estaciones
P_2	Vacio entre estaciones
P_3	Mayor Población Flotante
P_4	Mayor Población Flotante
P_5	Mayor Gasto en Transporte publico
P_6	Mayor Gasto en Transporte publico
P_7	Menor Gasto en Transporte Publico
P_8	Menor Gasto en Transporte Publico
P_9	Poblacion Mas Joven
P_10	Poblacion Menos Joven
P_11	Mayor Número de Bicicletas por estación
P_12	Mayor Número de Bicicletas por estación
P_13	Mayor Promedio de Rotacion
P_14	Menor Promedio de Rotacion
P_15	Mayor Poblacion AB
P_16	Mayor Poblacion C+
P_17	Mayor Poblacion CDE

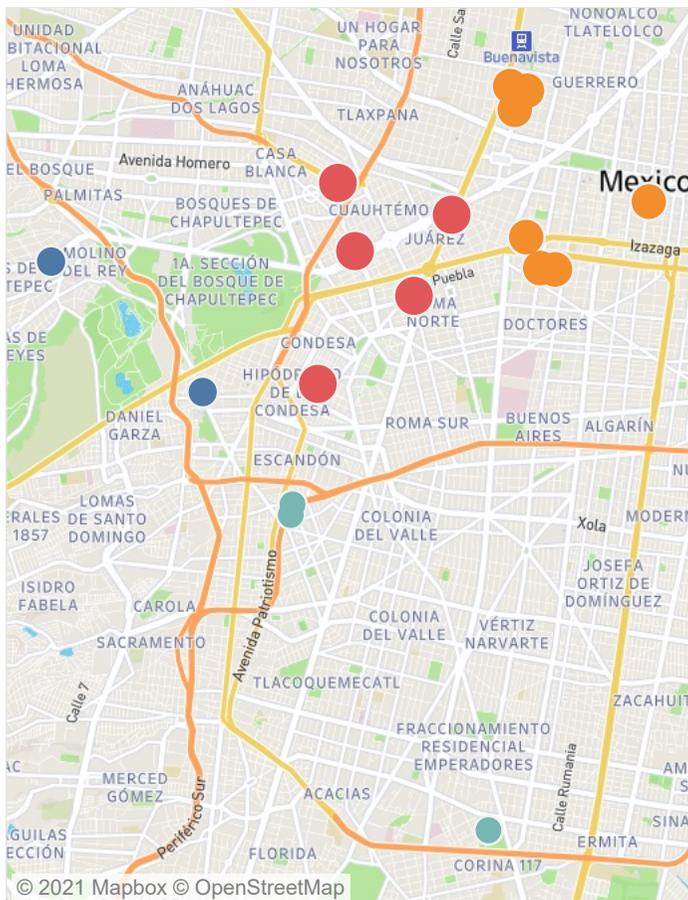
Primer Método: Clusterización

Mapa de todas las estaciones de ECOBICI con su cluster



Cluster ■ 1 ■ 2 ■ 3 ■ 4

Clusters de las Estaciones Propuestas

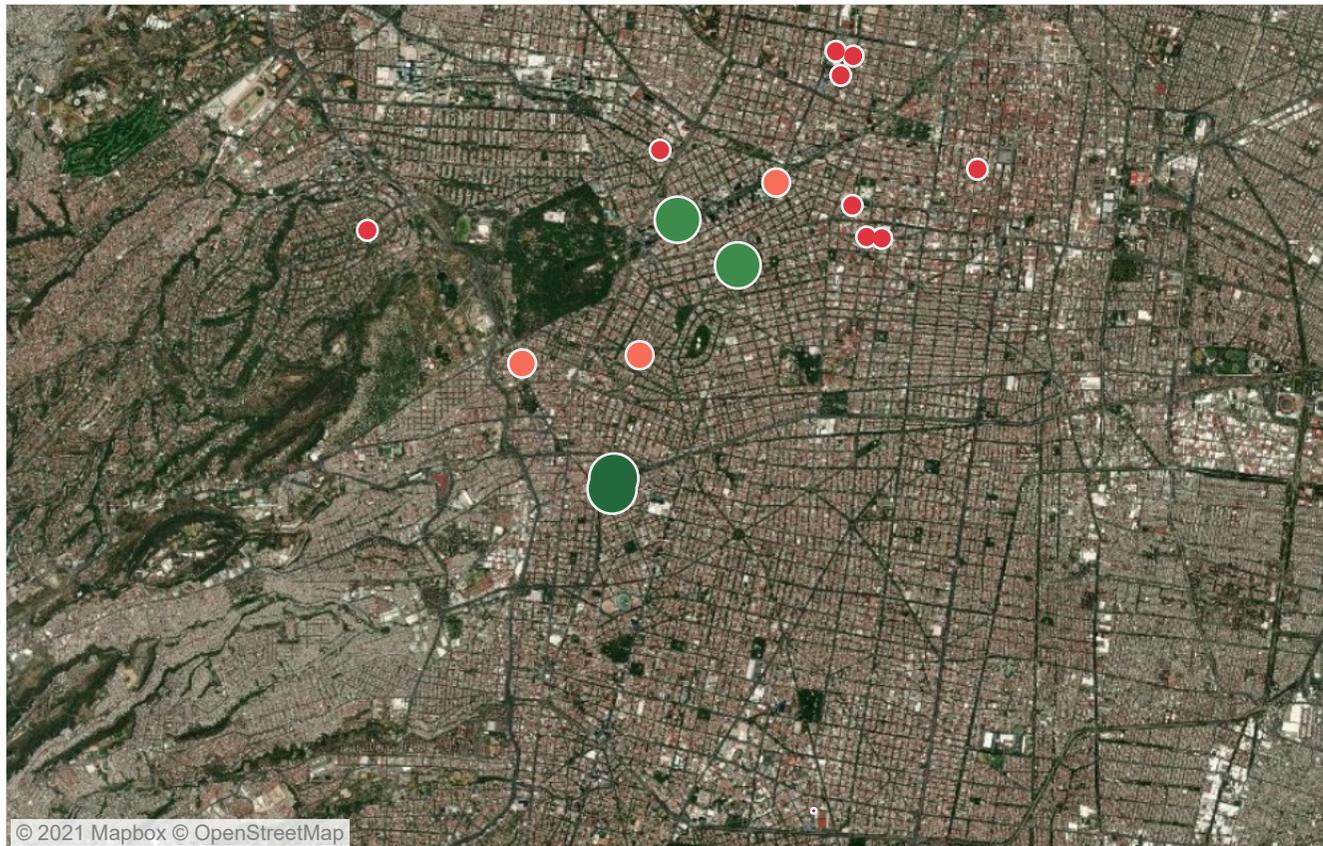


Resultados de las predicciones del método

Id	Name	Cluster	Rotacion	Total
P_1	Vacio entre esta..	4	94%	1,285
P_2	Vacio entre esta..	4	94%	1,285
P_3	Mayor Población..	3	96%	2,759
P_4	Mayor Población..	2	93%	2,282
P_5	Mayor Gasto en ..	3	96%	2,759
P_6	Mayor Gasto en ..	2	93%	2,282
P_7	Menor Gasto en ..	1	92%	1,541
P_8	Menor Gasto en ..	4	94%	1,285
P_9	Poblacion Mas J..	2	93%	2,282
P_10	Poblacion Menos..	1	92%	1,541
P_11	Mayor Número d..	2	93%	2,282
P_12	Mayor Número d..	3	96%	2,759
P_13	Mayor Promedio ..	3	96%	2,759
P_14	Menor Promedio..	2	93%	2,282
P_15	Mayor Poblacion..	3	96%	2,759
P_16	Mayor Poblacion..	2	93%	2,282
P_17	Mayor Poblacion..	2	93%	2,282

Tercer Modelo: Red Neural

Mapa de las estaciones propuestas con sus predicciones



El modelo con Deep Learning y una red neural que consiste en usar distintas capas de atributos que se construyen a partir de regresiones lineales entre las variables iniciales de las estaciones existentes de ECOBICI. Este es un modelo de clasificación por lo que se construyeron 10 intervalos de misma longitud sobre la variable de Total de bicicletas, esto con el objeto de que el modelo prediga a que intervalo pertenece cada propuesta. Esta red neural tuvo un 75% de exactitud.

Resultados del modelo

Identifi..	Name	Total	Probabili..	Rango
P_1	Vacio entre esta..	6156 - 7036	90.92%	7
P_2	Vacio entre esta..	6156 - 7036	94.18%	7
P_3	Mayor Población..	5277 - 6156	99.97%	6
P_4	Mayor Población..	879 - 1759	99.65%	1
P_5	Mayor Gasto en ..	879 - 1759	58.82%	1
P_6	Mayor Gasto en ..	879 - 1759	100.00%	1
P_7	Menor Gasto en ..	879 - 1759	99.99%	1
P_8	Menor Gasto en ..	0 - 879	100.00%	0
P_9	Poblacion Mas J..	879 - 1759	100.00%	1
P_10	Poblacion Menos..	1759 - 2638	100.00%	2
P_11	Mayor Número d..	879 - 1759	37.79%	1
P_12	Mayor Número d..	1759 - 2638	88.45%	2
P_13	Mayor Promedio ..	1759 - 2638	68.93%	2
P_14	Menor Promedio..	879 - 1759	38.95%	1
P_15	Mayor Poblacion..	5277 - 6156	47.95%	6
P_16	Mayor Poblacion..	879 - 1759	100.00%	1
P_17	Mayor Poblacion..	879 - 1759	55.89%	1

Resumen de los modelos

Resumen de todas las variables de predicción de cada modelo

		Cluster		Red Neural			Regression Tree
		Rotacion	Total de Bicile..	Intervalo de To..	Probabilidad	Rango	Total de Bicile..
P_1	Vacio entre esta..	94%	1285	6156 - 7036	91%	7	1,681
P_2	Vacio entre esta..	94%	1285	6156 - 7036	94%	7	1,681
P_3	Mayor Población..	96%	2759	5277 - 6156	100%	6	3,346
P_4	Mayor Población..	93%	2282	879 - 1759	100%	1	2,820
P_5	Mayor Gasto en ..	96%	2759	879 - 1759	59%	1	1,681
P_6	Mayor Gasto en ..	93%	2282	879 - 1759	100%	1	2,299
P_7	Menor Gasto en ..	92%	1541	879 - 1759	100%	1	1,359
P_8	Menor Gasto en ..	94%	1285	0 - 879	100%	0	2,041
P_9	Poblacion Mas J..	93%	2282	879 - 1759	100%	1	2,904
P_10	Poblacion Menos..	92%	1541	1759 - 2638	100%	2	2,041
P_11	Mayor Número d..	93%	2282	879 - 1759	38%	1	2,820
P_12	Mayor Número d..	96%	2759	1759 - 2638	89%	2	2,904
P_13	Mayor Promedio ..	96%	2759	1759 - 2638	69%	2	2,820
P_14	Menor Promedio..	93%	2282	879 - 1759	39%	1	2,820
P_15	Mayor Poblacion..	96%	2759	5277 - 6156	48%	6	2,904
P_16	Mayor Poblacion..	93%	2282	879 - 1759	100%	1	2,904
P_17	Mayor Poblacion..	93%	2282	879 - 1759	56%	1	1,259

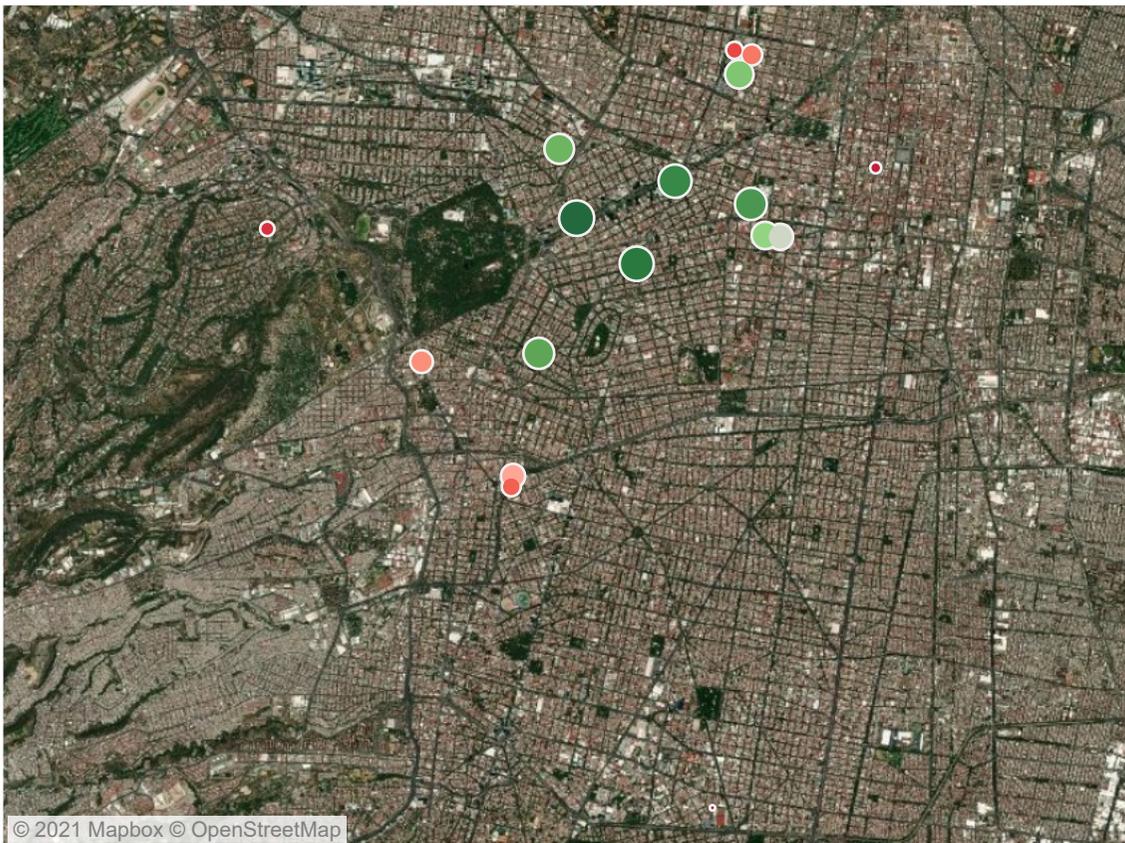
Métricas de Exactitud de cada modelo

Modelo	Metrica Exactitud	
K-means: Hartigan-Wong Clusters	Maximize Between Distance and Minimize Within Distance	62%
	Accuracy	76%
Keras Neural Network	Mean Difference %	40%
	RMSE	816

Resultados

Al tener más de un solo método, hay dos opciones. Una es decantarse por completo por un modelo y la otra es usar los tres para decir cual sería la mejor estación propuesta. Entonces, lo que se hizo fue ordenar de mayor a menor las predicciones de los tres modelos y se promedió el lugar en que quedaron.

Aquellos con el promedio más bajo son lo primeros lugares con los tres modelo.



Mapa de las estaciones propuestas con sus rankings finales

Ranking por los distintos métodos

Rango	Cluster	Neural Ne..	Tree
1	P_5	P_2	P_3
2	P_3	P_1	P_9
3	P_15	P_3	P_16
4	P_13	P_15	P_15
5	P_12	P_13	P_12
6	P_9	P_12	P_4
7	P_6	P_10	P_14
8	P_4	P_9	P_13
9	P_17	P_7	P_11
10	P_16	P_6	P_6
11	P_14	P_5	P_8
12	P_11	P_4	P_10
13	P_7	P_17	P_5
14	P_10	P_16	P_2
15	P_8	P_14	P_1
16	P_2	P_11	P_7
17	P_1	P_8	P_17

Ranking Total

Identifi..	Name	Ranki..
P_3	Mayor Población..	1
P_15	Mayor Poblacion..	2
P_12	Mayor Número d..	3
P_9	Poblacion Mas J..	4
P_13	Mayor Promedio ..	5
P_5	Mayor Gasto en ..	6
P_4	Mayor Población..	7
P_16	Mayor Poblacion..	8
P_6	Mayor Gasto en ..	9
P_2	Vacio entre esta..	10
P_10	Poblacion Menos..	11
P_14	Menor Promedio..	12
P_1	Vacio entre esta..	13
P_11	Mayor Número d..	14
P_7	Menor Gasto en ..	15
P_17	Mayor Poblacion..	16
P_8	Menor Gasto en ..	17

Conclusiones y Recomendaciones

1. Se le recomienda a ECOBICI colocar su nueva estación donde haya más concentración de población en general.
2. La segunda estación que se coloque deberá estar donde haya más población de nivel socioeconómico alto.
3. La data de población flotante y de movilidad en general es la más valiosa del proyecto.

Referencias

- Data de ECOBICI: <https://www.ecobici.cdmx.gob.mx/es/informacion-del-servicio/open-data>
- Data Metro CDMX: <https://www.metro.cdmx.gob.mx/>
- Data del INEGI: <https://www.inegi.org.mx/programas/enigh/nc/2018/>
- Data Población Flotante: <https://www.marketdatamexico.com/es/static/home>