

PREDICIENDO LA VELOCIDAD

ECOBICI

Jose Ramos

Universidad del Valle de Guatemala

La seguridad de los usuarios es una prioridad, si van muy rápido, puede ser un problema.

1. Intentar crear un modelo de clasificación que nos indique si un usuario se va a mover rápido.
2. Mostrar el proceso.
3. Comparar los modelos con dos datasets, el de marzo y el de enero de 2021.

CREANDO EL MAPA



Se tiene un archivo de excel con la información de las estaciones.

	id	name	address	addressNumber	zipCode	districtCode	districtName	altitude	nearbyStations_001	location_lat	location_lon	station
0	448	448 DR. ANDRADE - ARCOS DE BELÉN	DR. ANDRADE ARCOS DE BELÉN	S/N	NaN	NaN	NaN	NaN	448	19.426611	-99.144470	BIKE
1	443	443 BRUNO TRAVEN-GOLONDRINAS	443 - Bruno Trave-Avenida México-Coyoacan	S/N	3340.0	1	Ampliación Granada	NaN	432	19.359655	-99.162025	
2	463	463 LAGO ANDRÓMACO-LAGO ZURICH	463 LAGO ANDRÓMACO-LAGO ZURICH	S/N	NaN	1	Ampliación Granada	NaN	197	19.442338	-99.202492	
3	416	416 RECREO-PARROQUÍA	416 - Recreo-Parroquia	S/N	3104.0	1	Ampliación Granada	NaN	390	19.371308	-99.177607	
4	437	437 MORAS-COMUNAL	437 - Moras-Comunal	S/N	3240.0	1	Ampliación Granada	NaN	423	19.364808	-99.176022	

Figura 1: Dataframe de las estaciones

Fuente: elaboración propia

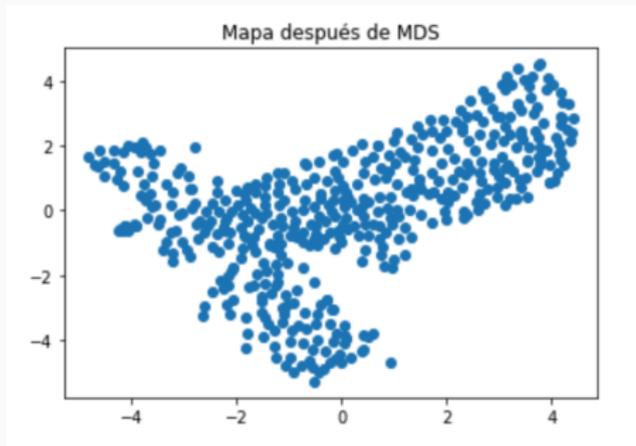


Figura 2: Mapa recreado
Fuente: elaboración propia

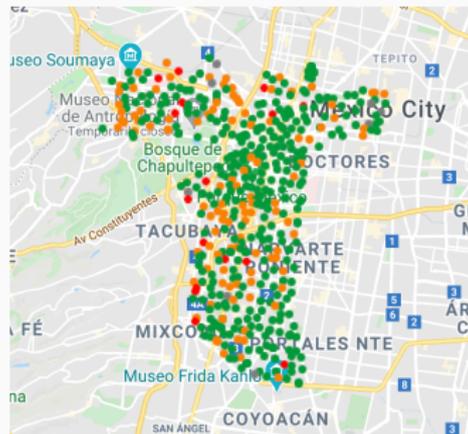


Figura 3: Mapa real
Fuente: [1]

VARIABLES Y DATOS FALTANTES



Creación de nuevas variables

1. Distancia
2. Duración del viaje
3. Velocidad
4. Hora, día y minuto

```
Genero_Usuario      28659
Edad_Usuario        4
Bici                 4
Ciclo_Estacion_Retiro 4
Fecha_Retiro        4
Hora_Retiro         4
Ciclo_Estacion_Arribo 0
Fecha_Arribo        0
Hora_Arribo         0
dtype: int64
```

Figura 4: Enero

Fuente: elaboración propia

```
Genero_Usuario      37244
Edad_Usuario        0
Bici                 0
Id_Usuario          0
Ciclo_Estacion_Retiro 0
Fecha_Retiro        0
Hora_Retiro         0
Ciclo_EstacionArribo 0
Fecha_Arribo        0
Hora_Arribo         0
dtype: int64
```

Figura 5: Marzo

Fuente: [1]

```
In [20]: df.Duracion.min()  
Out[20]: Timedelta('-323 days +00:04:51')
```

Figura 6: Valor mínimo de la variable duración
Fuente: elaboración propia

```
In [32]: len(df.loc[df['Duracion'] < pd.Timestamp.now() - pd.Timestamp.now()])  
Out[32]: 49
```

Figura 7: Número de elementos con variable duración negativa
Fuente: elaboración propia

MODELO



Los usuarios utilizan el camino más corto para moverse de estación en estación.

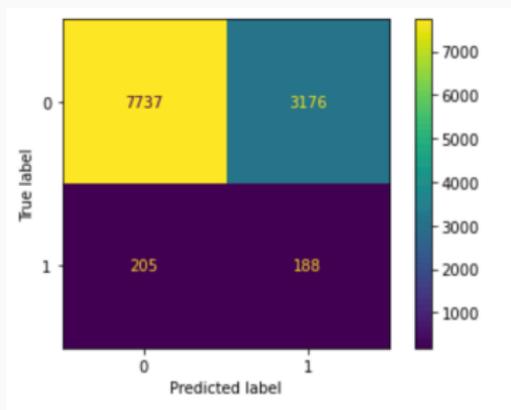


Figura 8: Enero

Fuente: elaboración propia

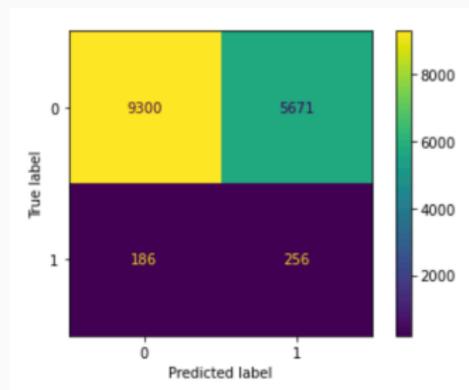


Figura 9: Marzo

Fuente: [1]



Gobierno de la Ciudad de México.
Mapa de cicloestaciones.