



# EXAMEN PARCIAL



- Báez Herrera Samuel
- Giddings Soto Lorrain
- Hernández Jimenez  
Luis Daniel
- Sosa Medellín Adrián

# INTRODUCCIÓN



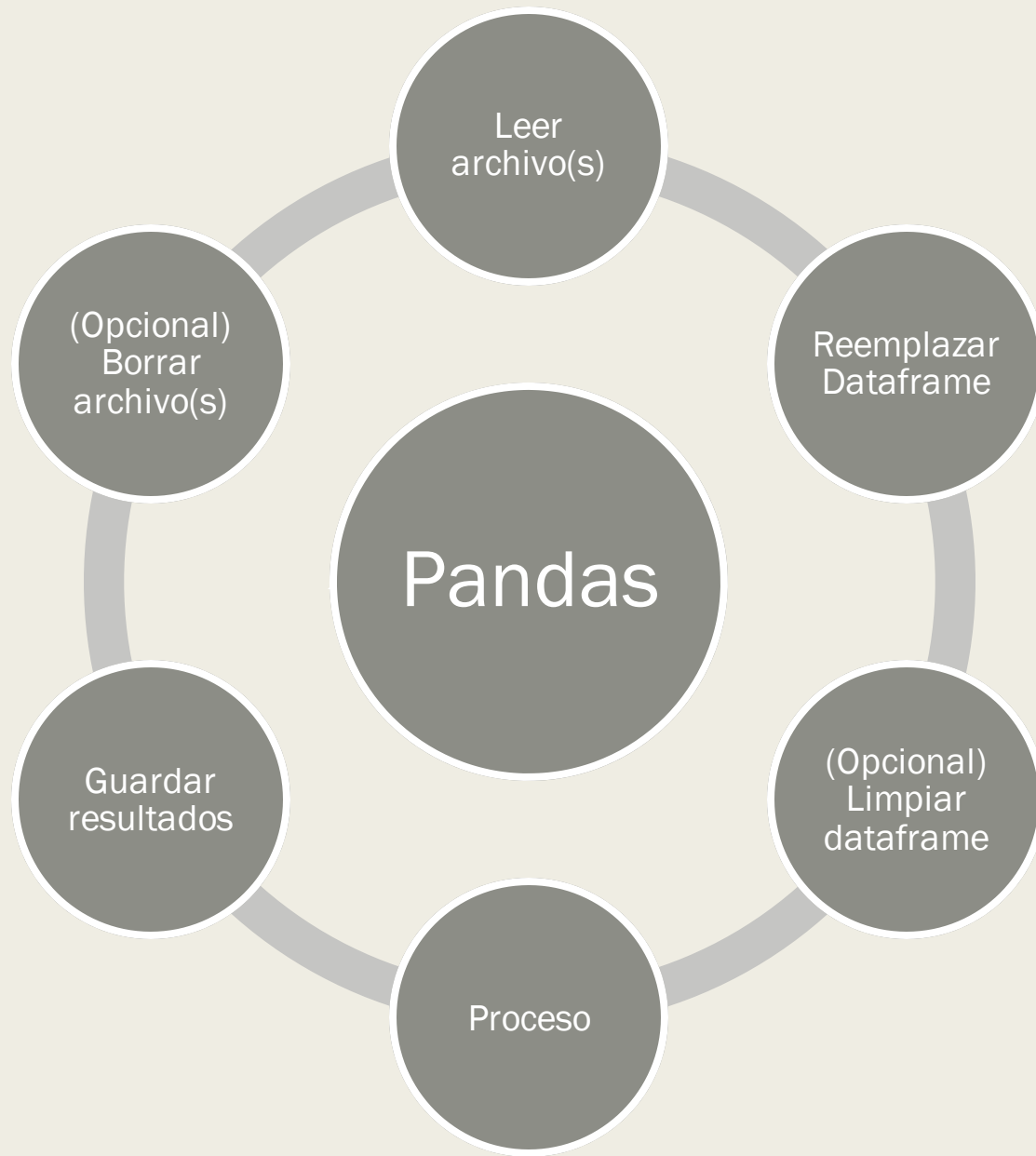
# Introducción

El examen consistió en trabajar con archivos de manera remota para responder lo siguiente:

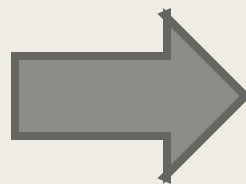
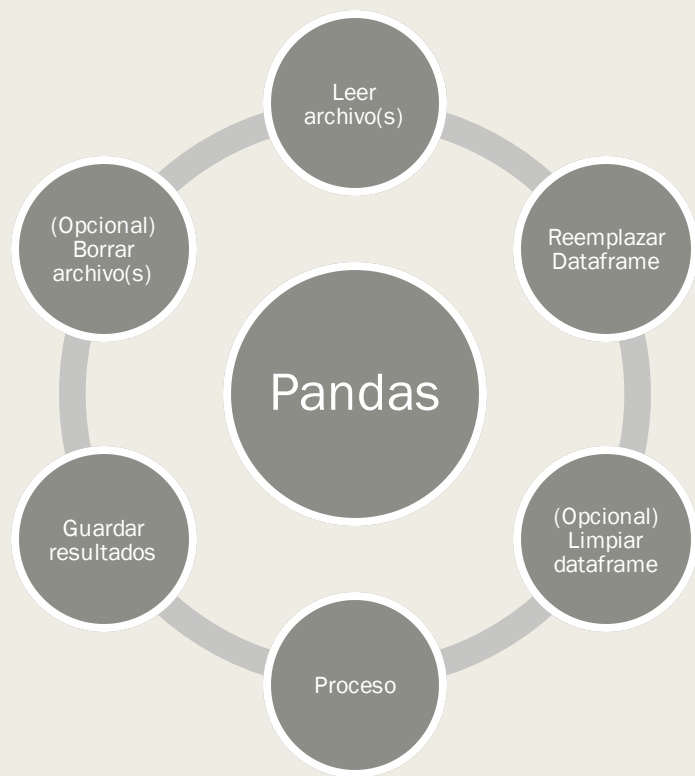
- Con Pandas:
  - Forma de los archivos, nombres de las columnas, tiempo que tarda en cargar .csv a dataframe, huella de memoria de cada uno de los archivos, detectar registros erróneos, tipos de datos en columna, calcular distancia media, y tiempo que tarda en calcularla.
- Con Dask:
  - *Calcular distancia media , y tiempo que tarda dask en calcularla, diferencia por archivos de la columna trip\_time\_in\_secs y la columna calculada, 12 taxis con más viajes, taxi con más viajes, cuenta de viajes largos, suma de pasajeros por hora y día de la semana.*
- Graficas de pasajeros por hora y día de la semana.
- Graficar puntos de pickup y dropoff para día de la semana y rango de horas.

PANDAS

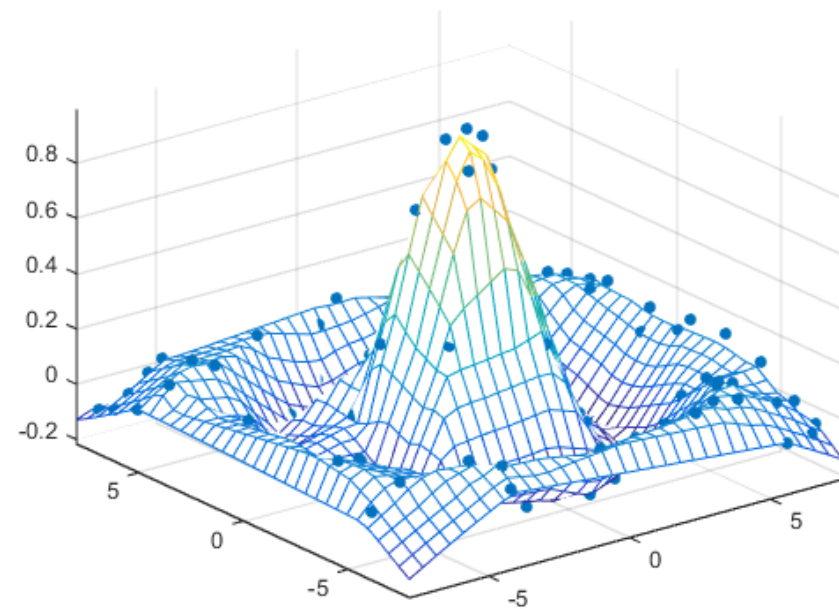




## Flujo de trabajo Pandas



## Mathplotlib



Estadística  
de  
resultados  
de operación

Mes	Dimensiones del Dataframe		Memoria	Tiempo	
	Registros	Columnas		Tiempo de carga de DataFrame	Operación
1	14776615	10	1701.85	38.35	0.03
2	13990176	10	465.50	36.13	0.02
3	15749228	10	975.77	40.45	0.03
4	15100468	10	829.02	39.23	0.03
5	15285049	10	798.10	38.66	0.03
6	14385456	10	508.19	36.33	0.03
7	13823840	10	585.87	35.22	0.02
8	12597109	10	516.43	33.23	0.02
9	14107693	10	1242.54	36.51	0.02
10	15004556	10	1186.45	38.74	0.02
11	14388451	10	801.95	37.61	0.02
12	13971118	10	497.66	35.05	0.02

DASK



# Obtención de promedios

trip\_distance

511.94 segundos

8.53

MINUTOS

trip\_time\_in\_secs

513.56 segundos

8.56

MINUTOS

# Diferencia entre *trip\_time\_in\_secs* y *duración*

- Existen **173176305** registros en los que la columna *duración* y la columna *trip\_time\_in\_secs* difieren.
- De los **173179759** registros, **3454** coinciden.
- Media de diferencia tiempos por archivo:
  - 2.196776054597078
  - 1.6487864770250216
  - 3.267570385037286
  - 0.9756265170059633
  - 3.8272250223077515
  - 0.9058830668975674
  - 0.8176269401266221
  - 799.3146654522081
  - 1.8182331441434116
  - 1.4549623461034102
  - 1.1991220597686296
  - 1.2428704703517666

# Cantidad de viajes largos

Mes	Viajes largos	Porcentaje
1	1715237	12%
2	2545155	18%
3	2373698	15%
4	2428514	16%
5	1768887	12%
6	2111850	15%
7	2264701	16%
8	2552684	20%
9	2400513	17%
10	2084036	14%
11	1919857	13%
12	2417638	17%

# Taxis con más viajes para cada uno de los archivos

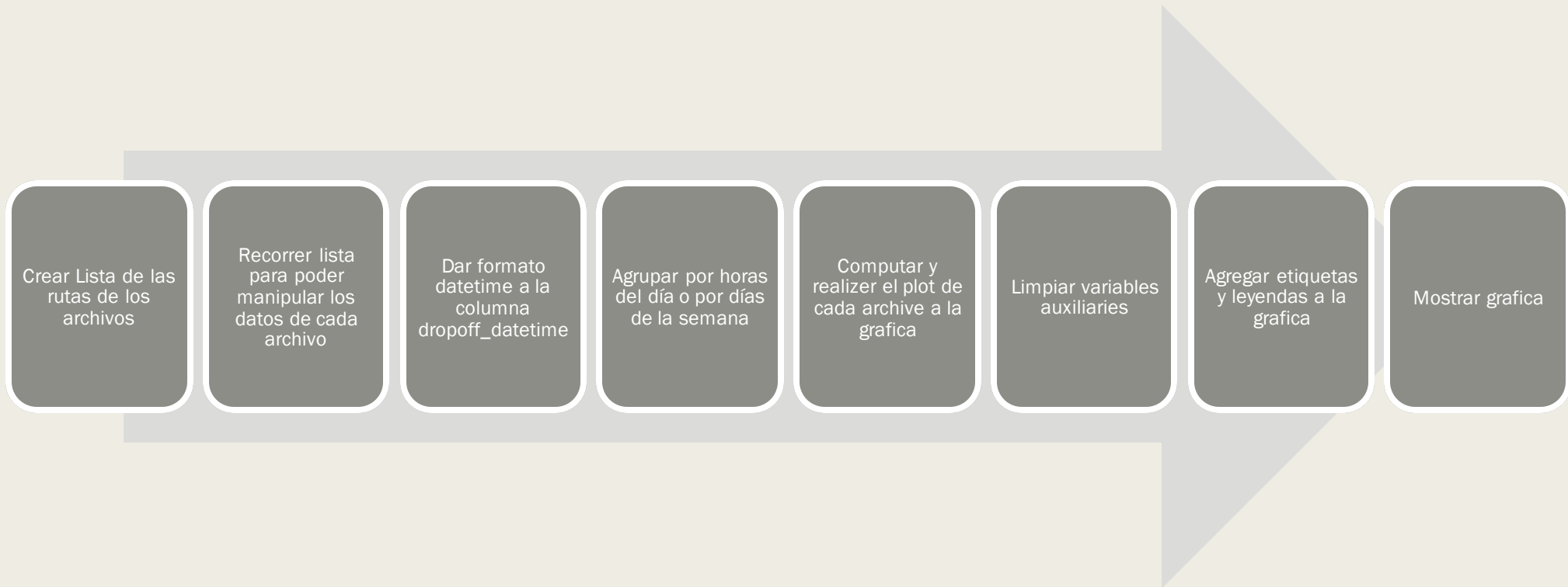
- 1 - DAF57CF25F00457CC6077CD628EC71AC 272
- 2 - DB51F659A90E9457DD2A2F4D5A2A3186 262
- 3 - 4E834DFB7A8831D0A5B6F9B80092A61F 287
- 4 - 1E7C1EB194CCFD58634305DBE0588B85 296
- 5 - 4E834DFB7A8831D0A5B6F9B80092A61F 320
- 6 - 698A6074D905BC18FE001032FA0B2048 315
- 7 - 5466D714601371299033C01FB08BB93B 309
- 8 - 6BD1B641A1CD55803A21560299B985A7 308
- 9 - 570D50E20C4E20D4428EBF94F11DF190 328
- 10- 7E5A55739A6EFA325F650FD7739135BF 340
- 11- 1239300107099B26BF07526F66C30BAF 326
- 12- BDF61165DAA42F17D35F5875F01B5C7A 329

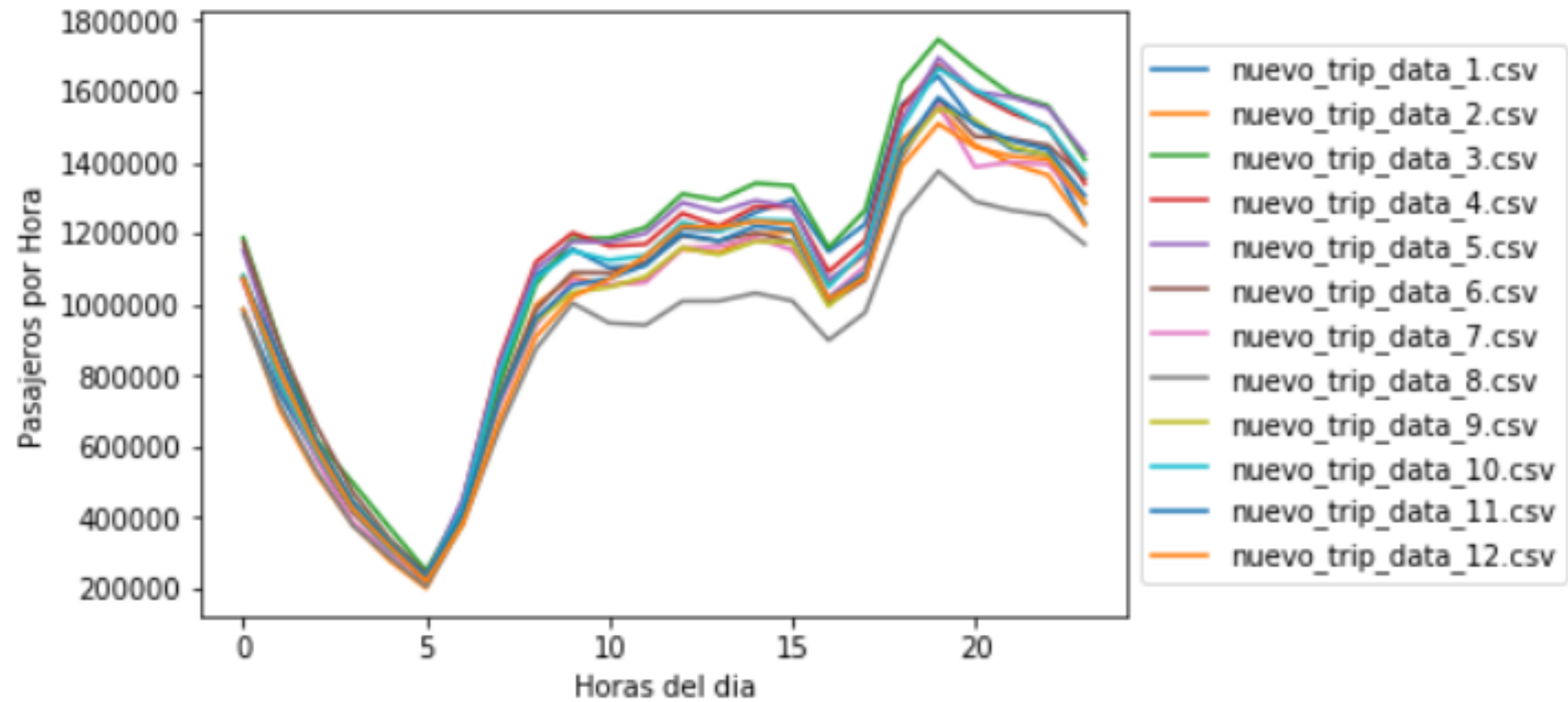
GRAFICAS

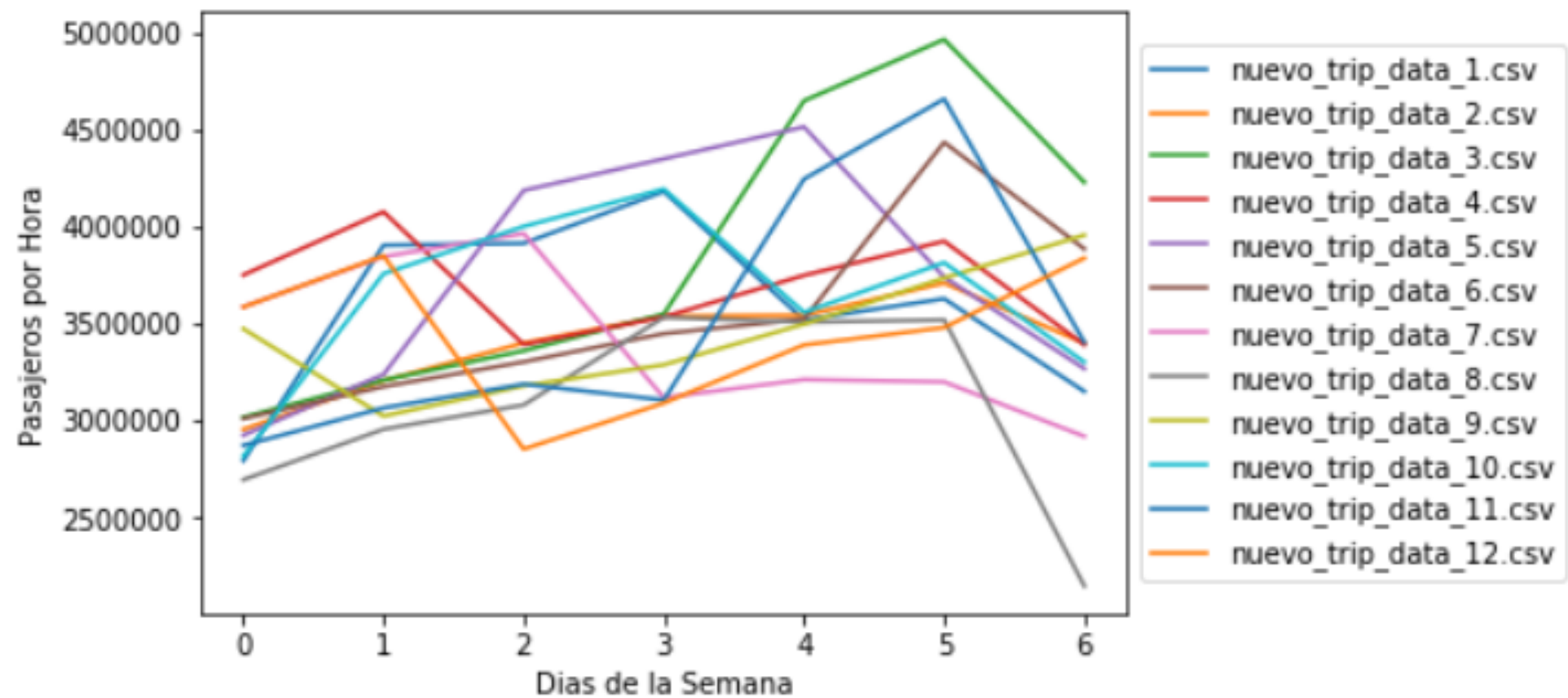


# Flujo de trabajo

## Graficas





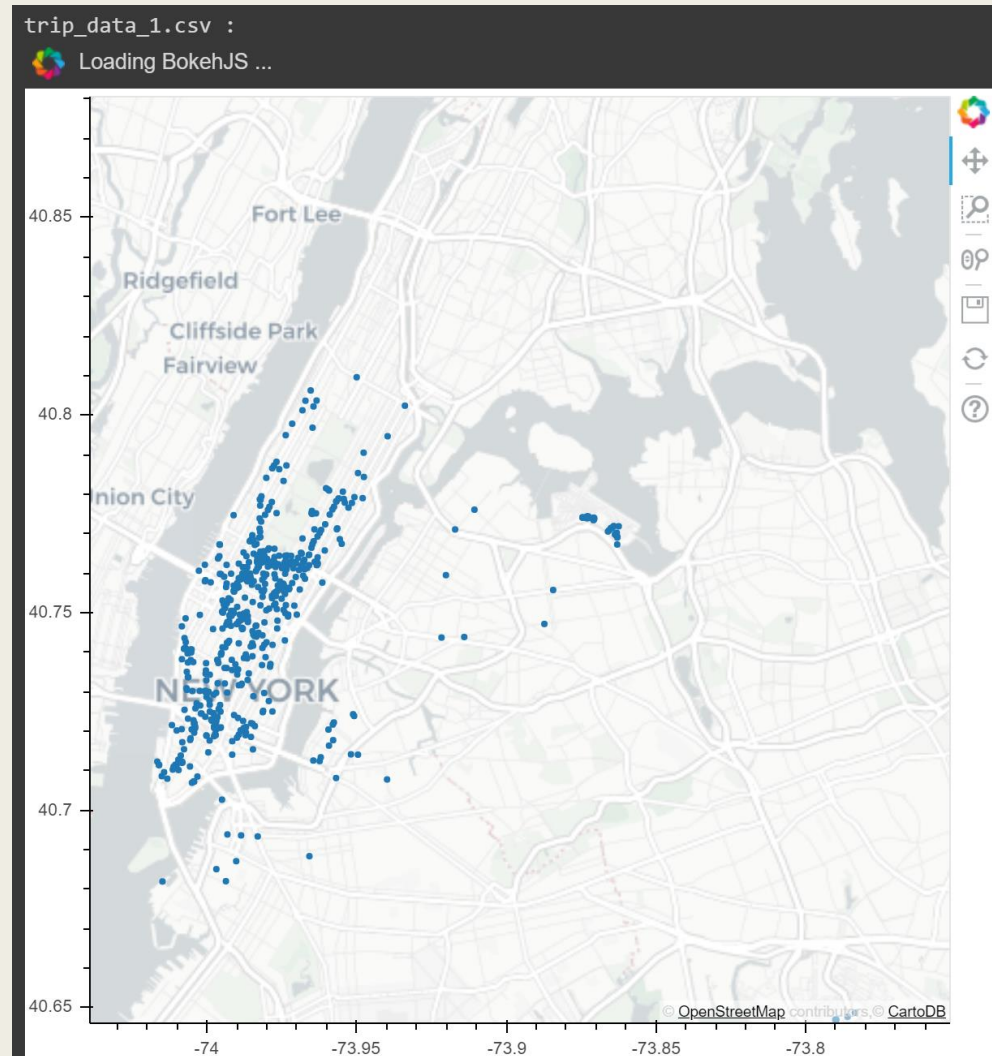




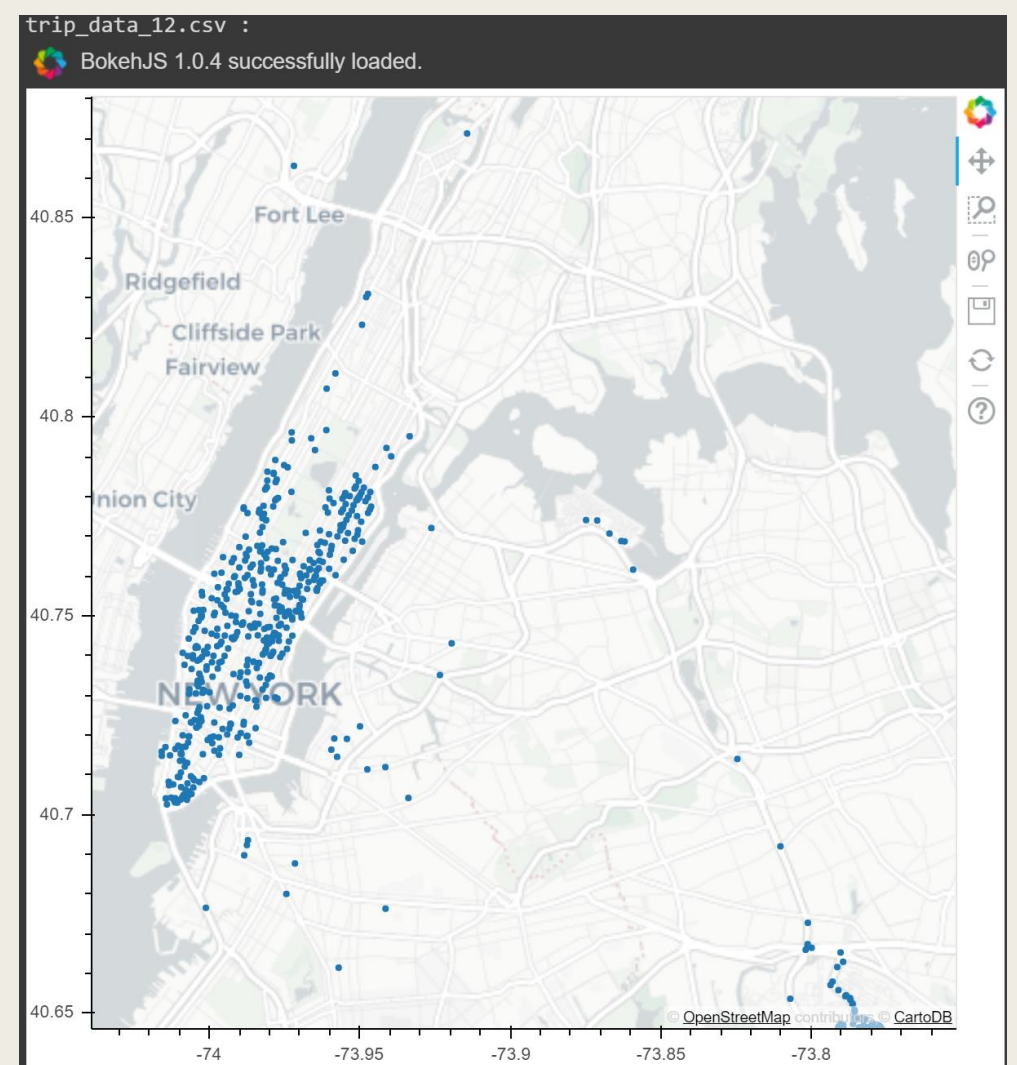
MAPAS



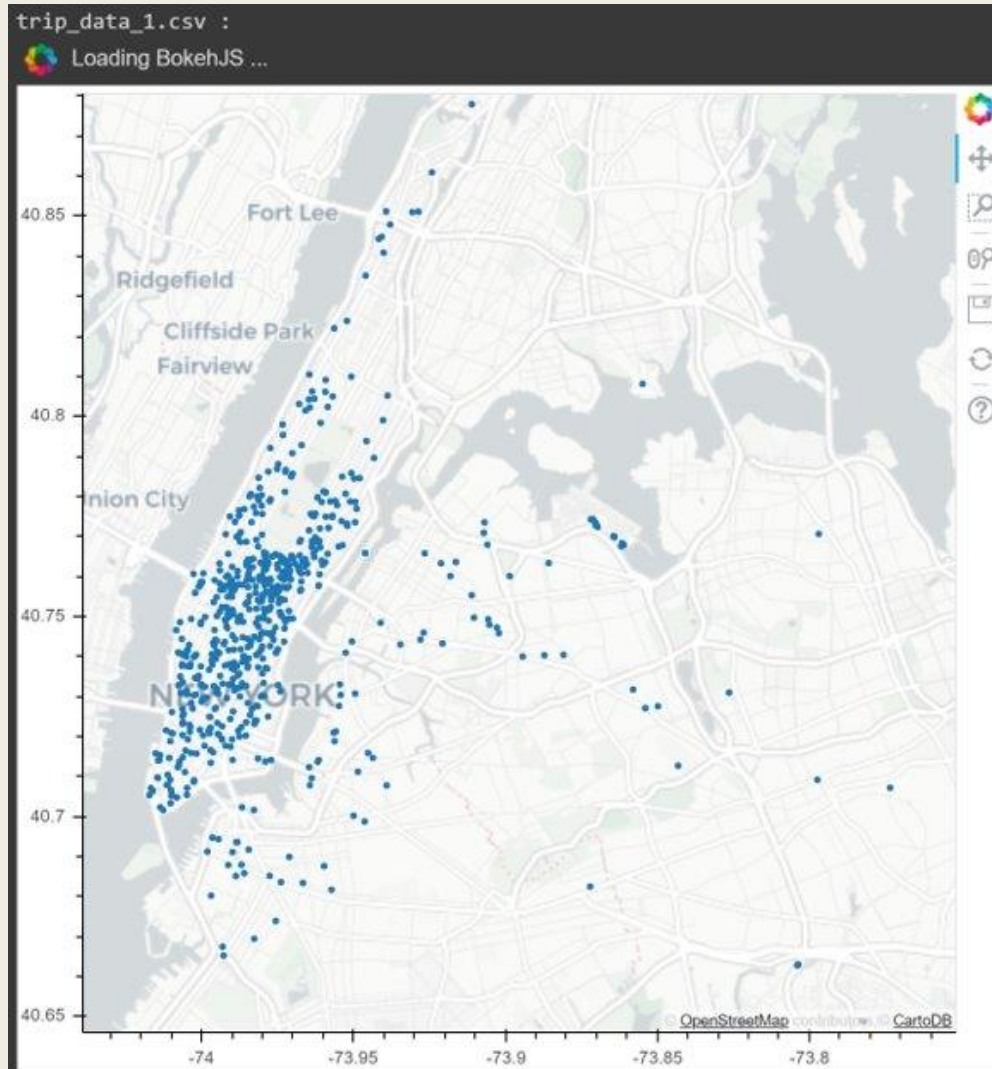
# Pickup



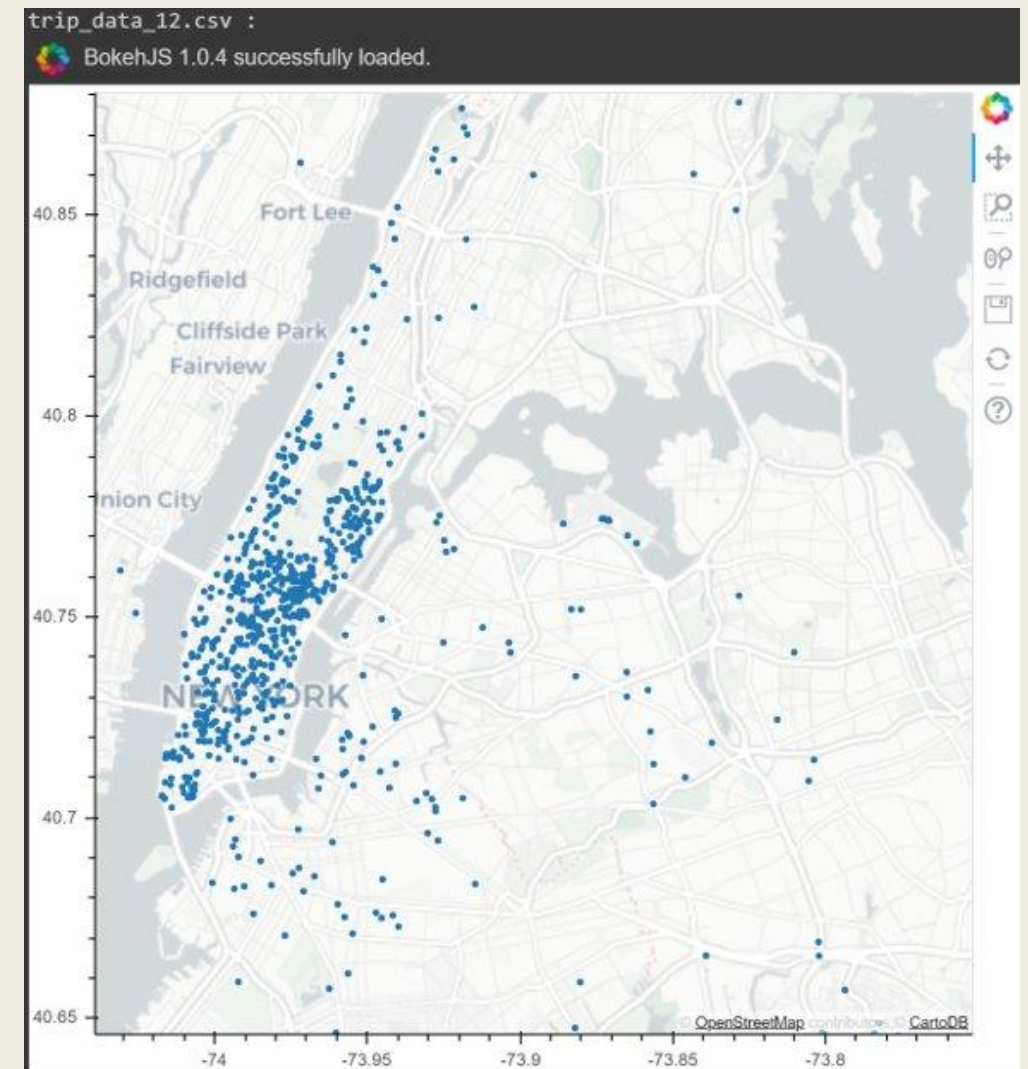
...



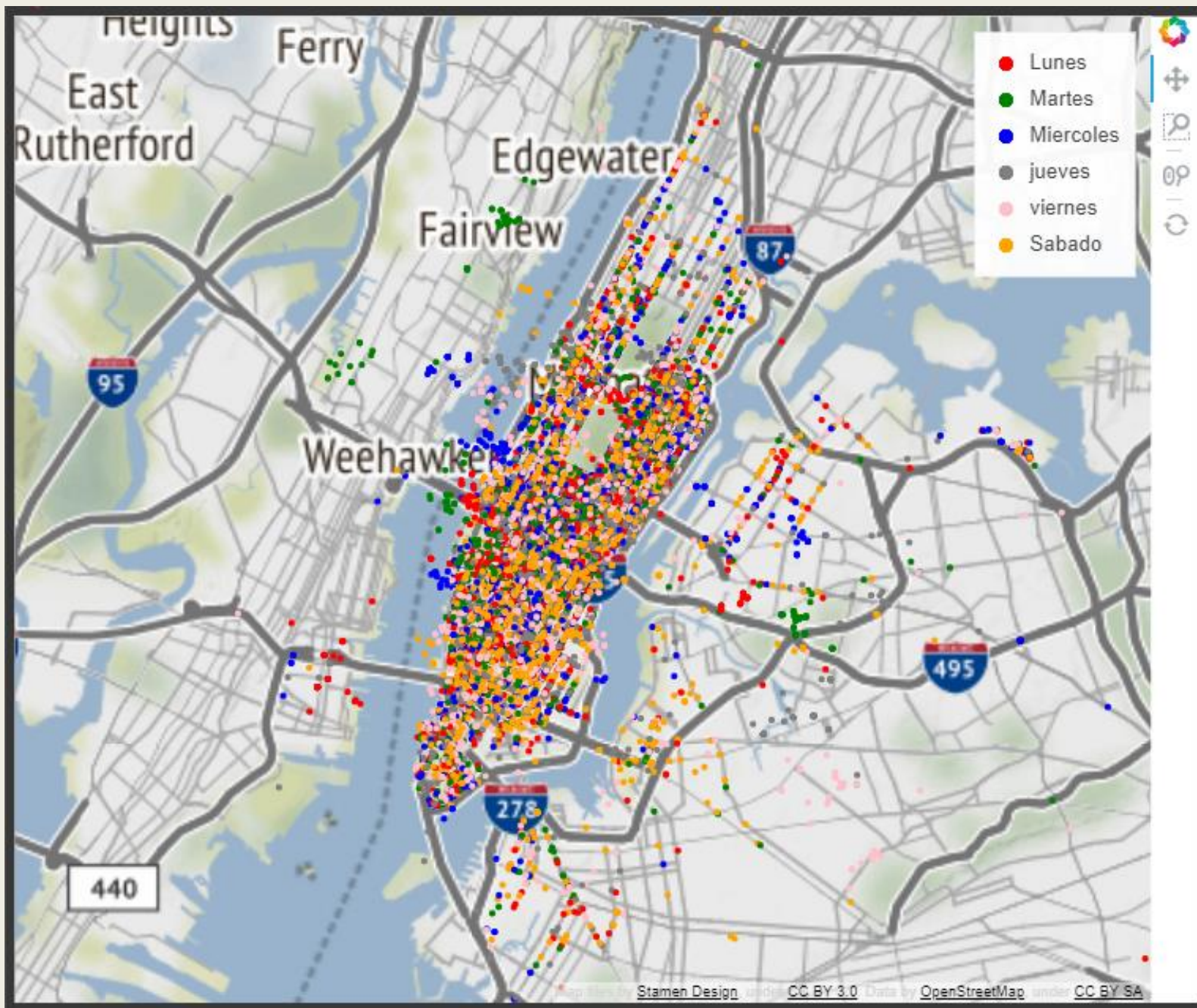
# Dropoff



...







CLASIFICADO