



Proceso creativo en un proyecto de
Data Science

Luis Valdeavellano
Data Scientist



Proyecto de Site Location en Guatemala y El Salvador

Proceso creativo

02.

V

Abordar el problema

Formular hipótesis

Data Engineering

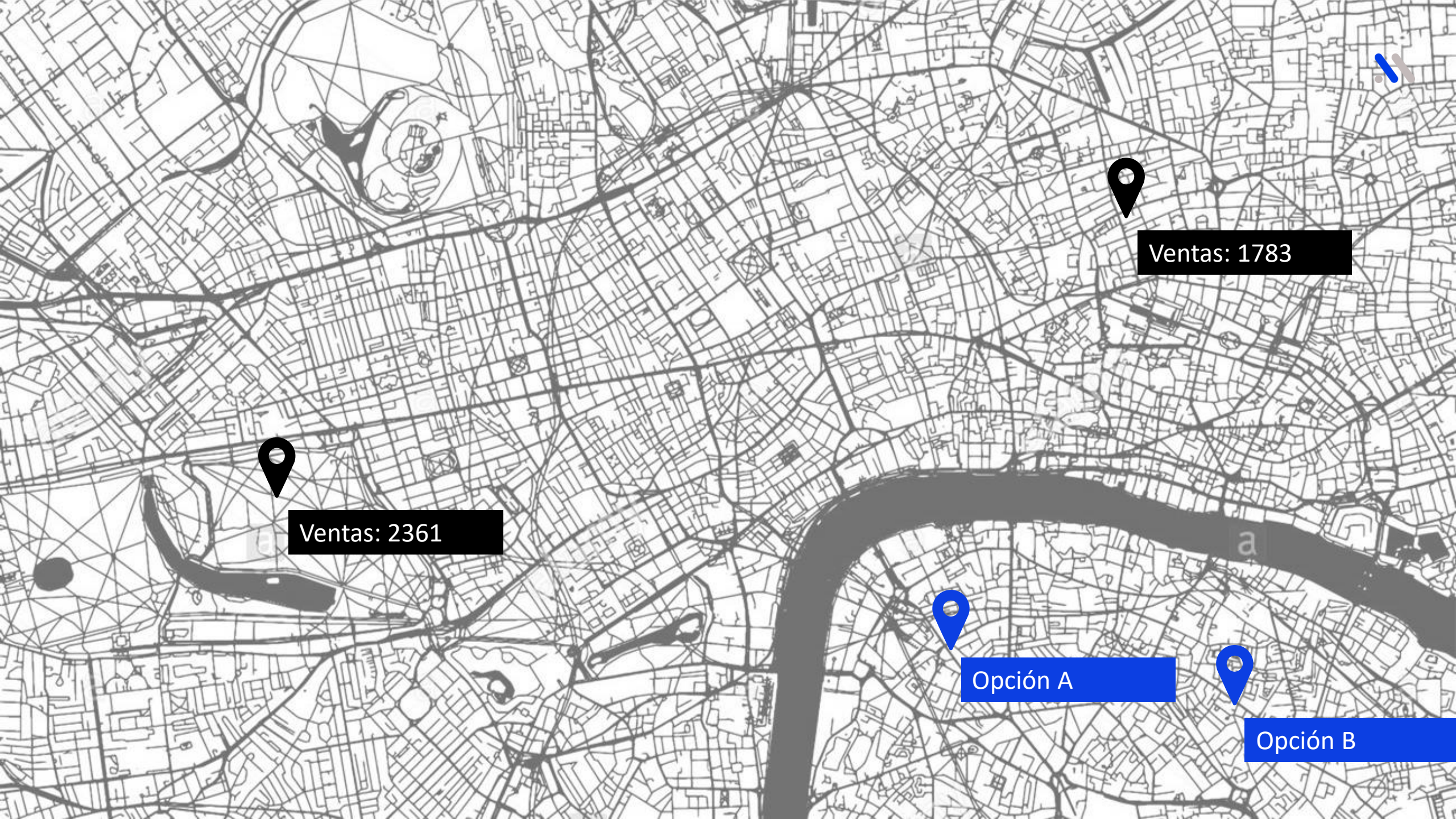
Identificar fuentes
disponibles

Convertir datos etéreos
en variables

+

+

+



Ventas: 2361



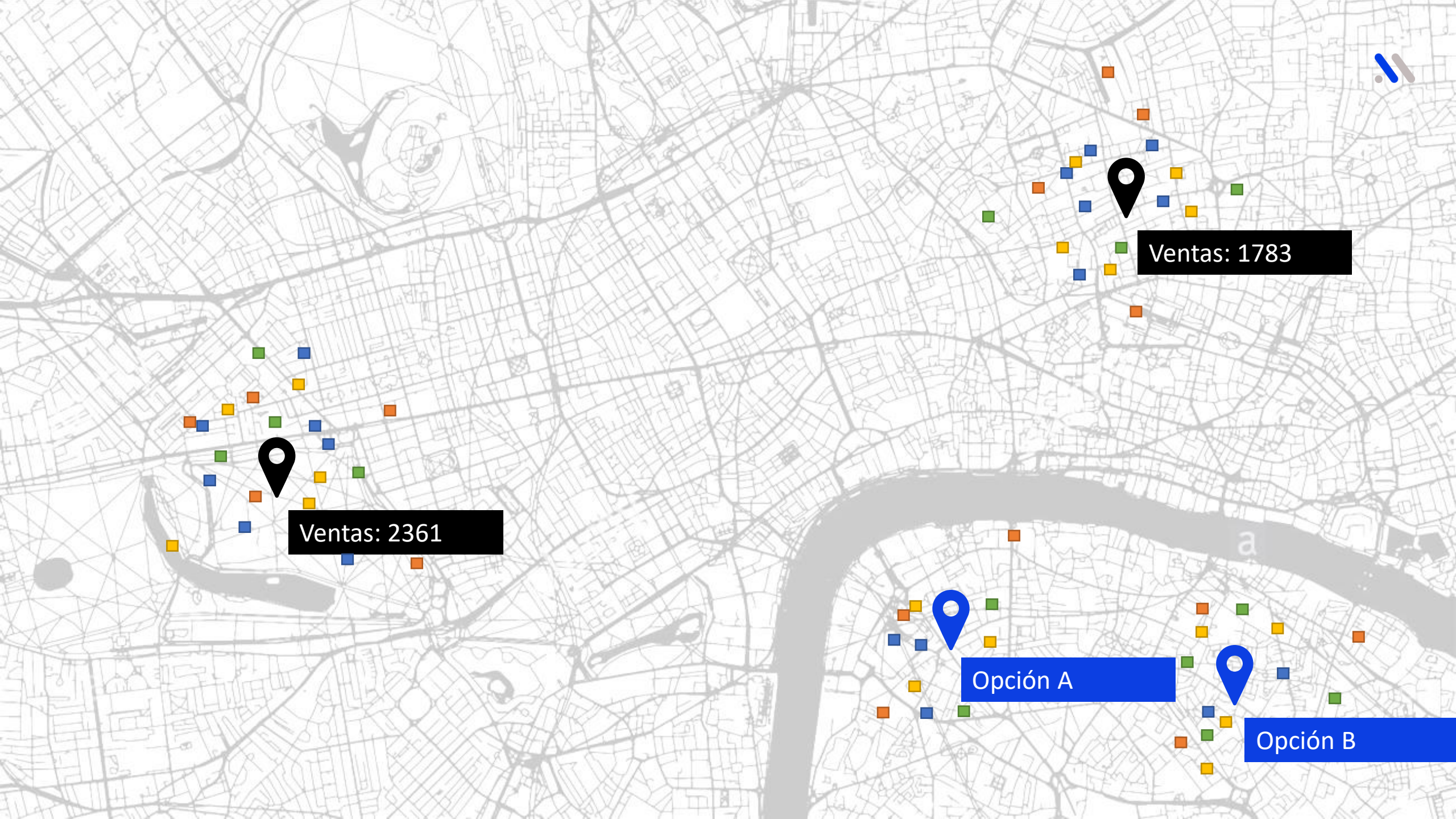
Ventas: 1783



Opción A



Opción B

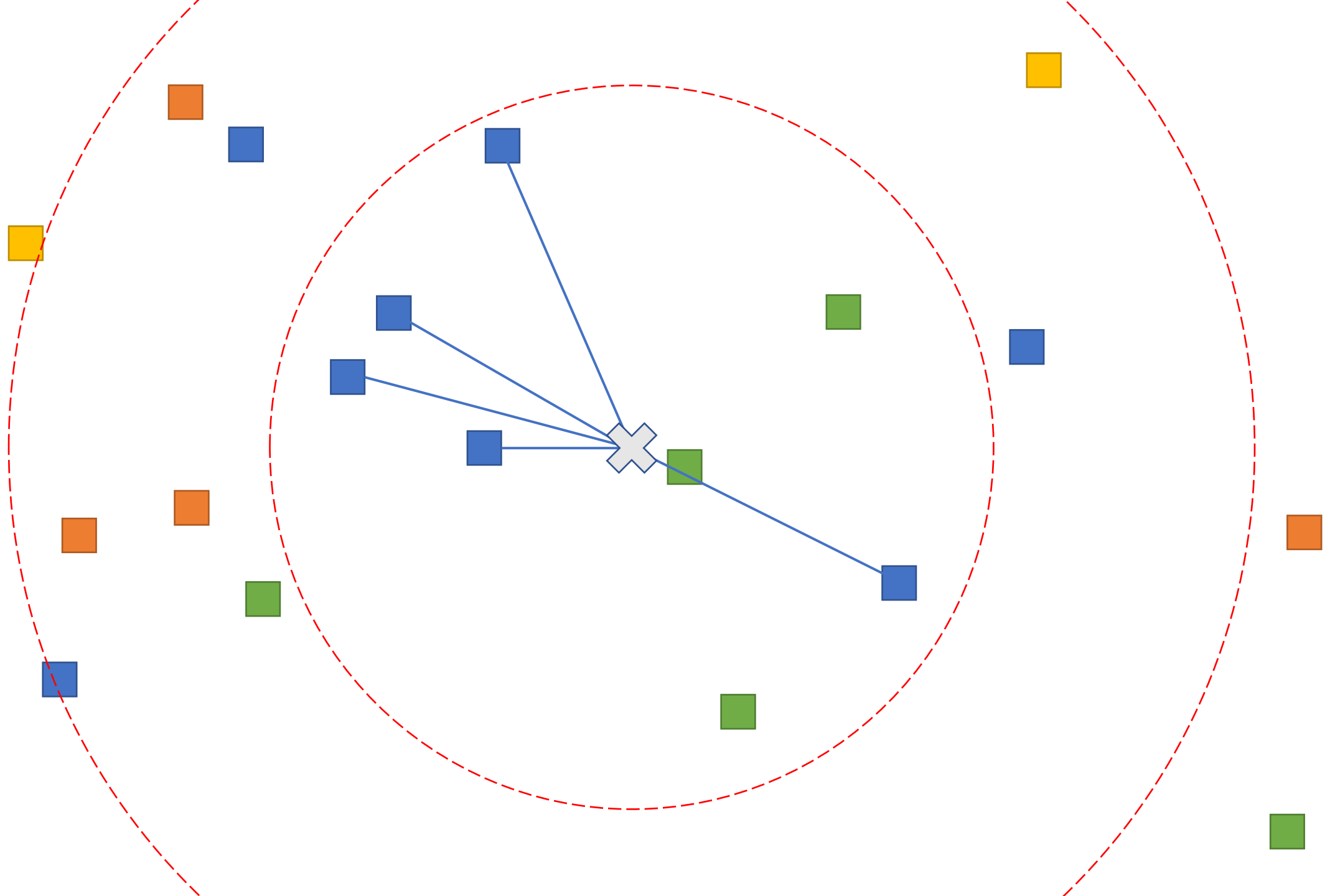


Ventas: 1783

Ventas: 2361

Opción A

Opción B

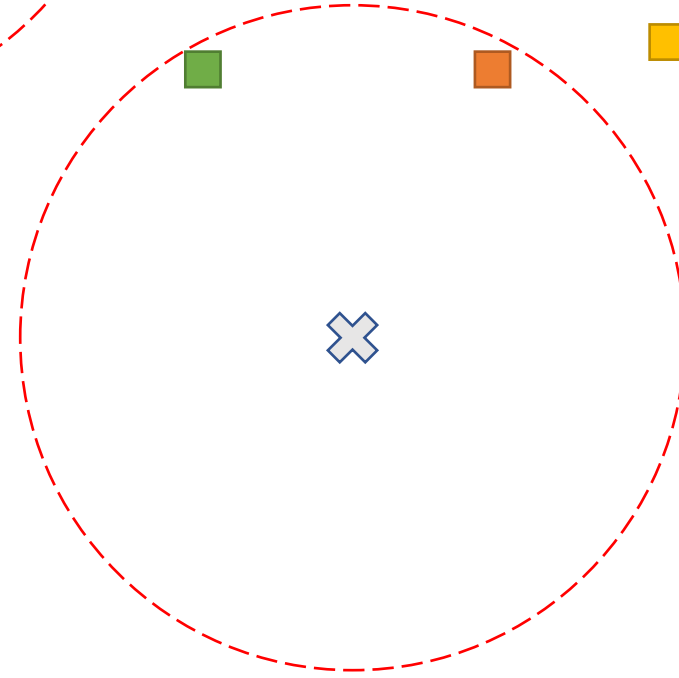
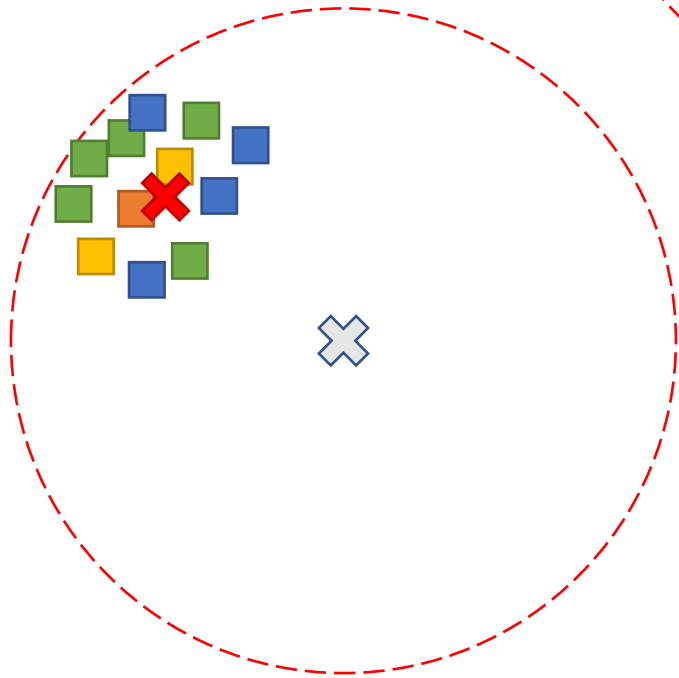
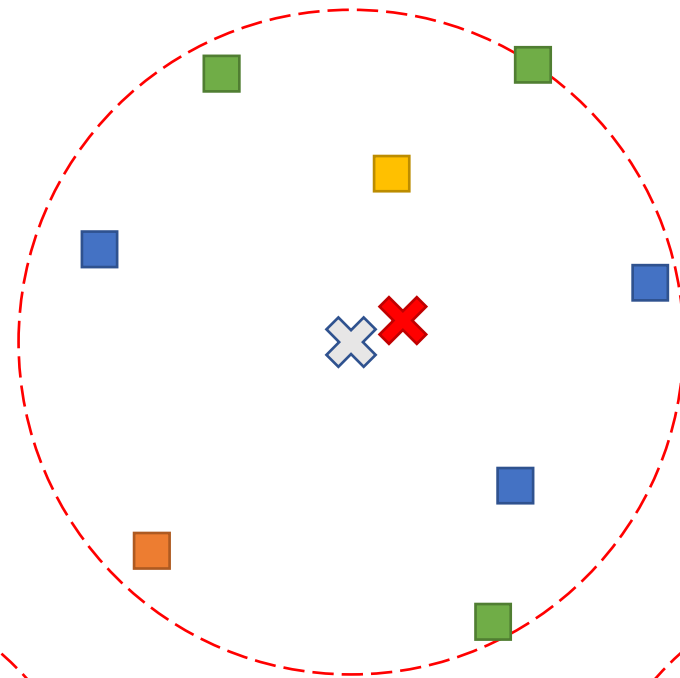
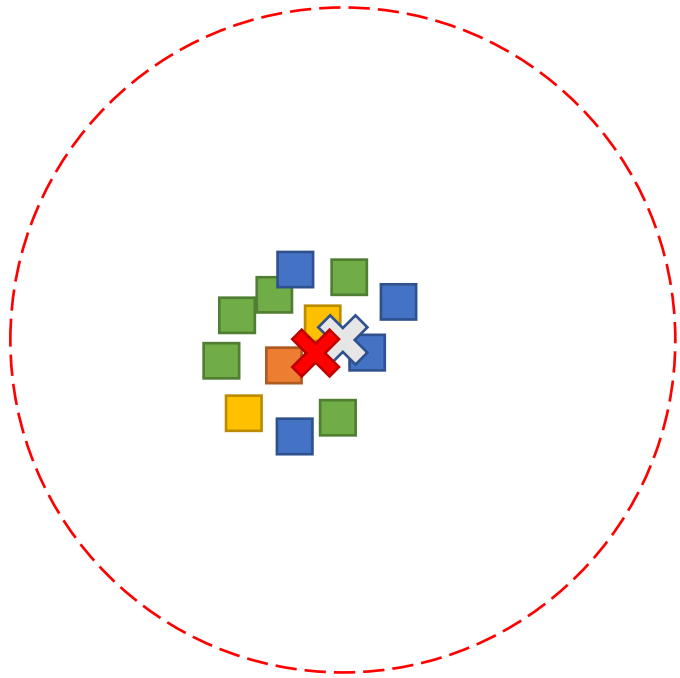




FORMULAR HIPÓTESIS

Trazar una Ruta





Promedio de distancias
de establecimientos a
centroide

vs.

Distancia del punto
evaluado al centroide



Índice de Concentración Económica

ICE

ICE

BABY

$\rho =$

Promedio de distancias
de establecimientos a
centroide

$d =$

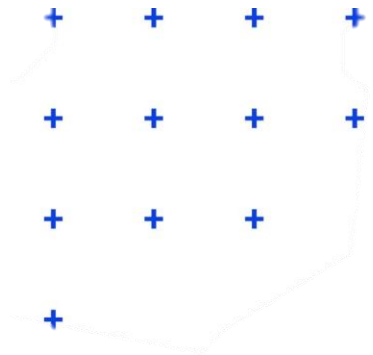
Distancia del punto
evaluado al centroide

$$ice = \left(1 - \frac{1}{\log(\rho)}\right) \cdot d$$



~~*nada*~~

ejecución





Proyecto de Site Location en Guatemala y El Salvador

Data Engineering

11.

V

Conocer la fuente

Definir una estrategia

Implementar

Conocer el output

Diseñar un algoritmo

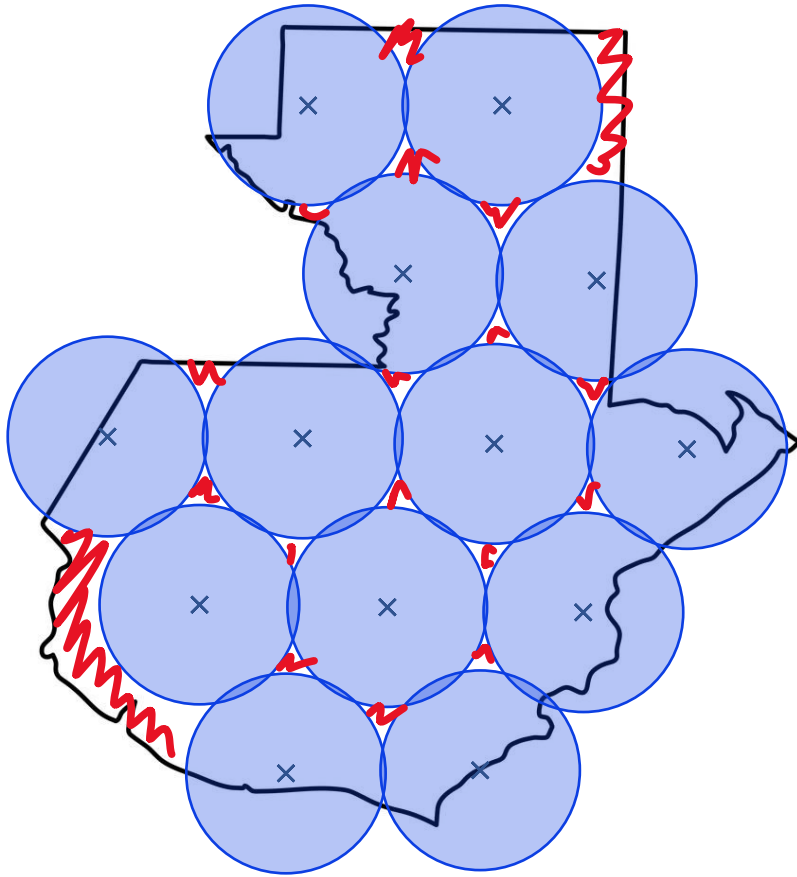




Máximo radio: 5km

Estrategia de recolección de data

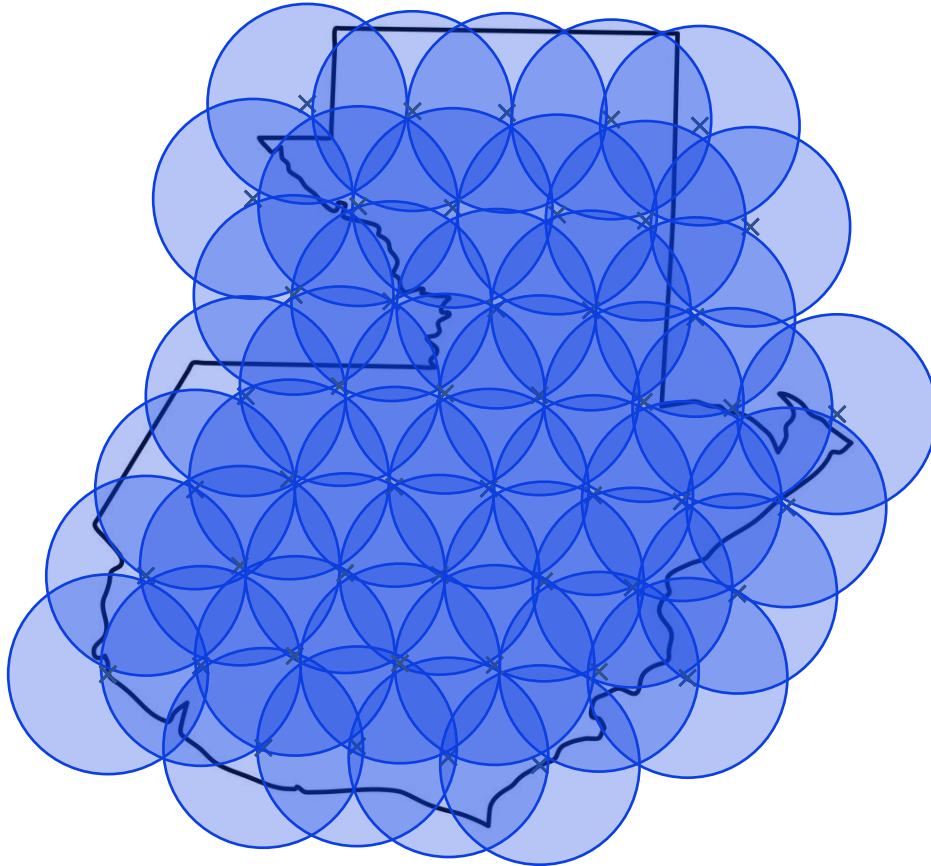




Máximo radio: 5km

Estrategia de recolección de data





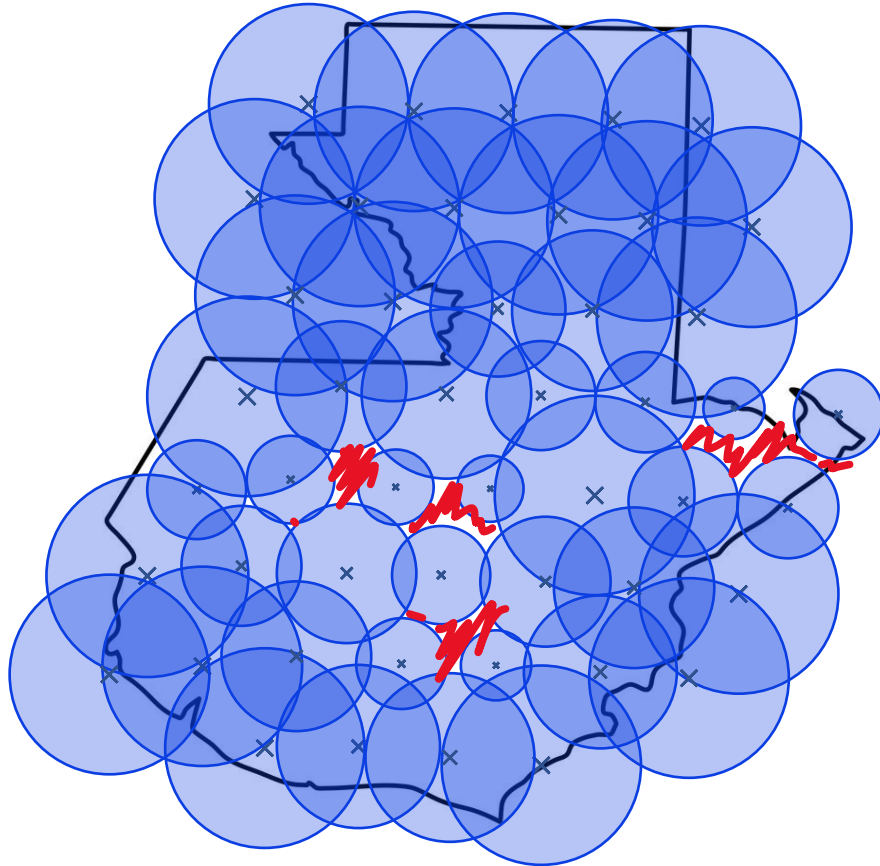
Máximo radio: 5km



**Máximo número de
establecimientos en
resultado: 60**

Estrategia de recolección de data





Máximo radio: 5km



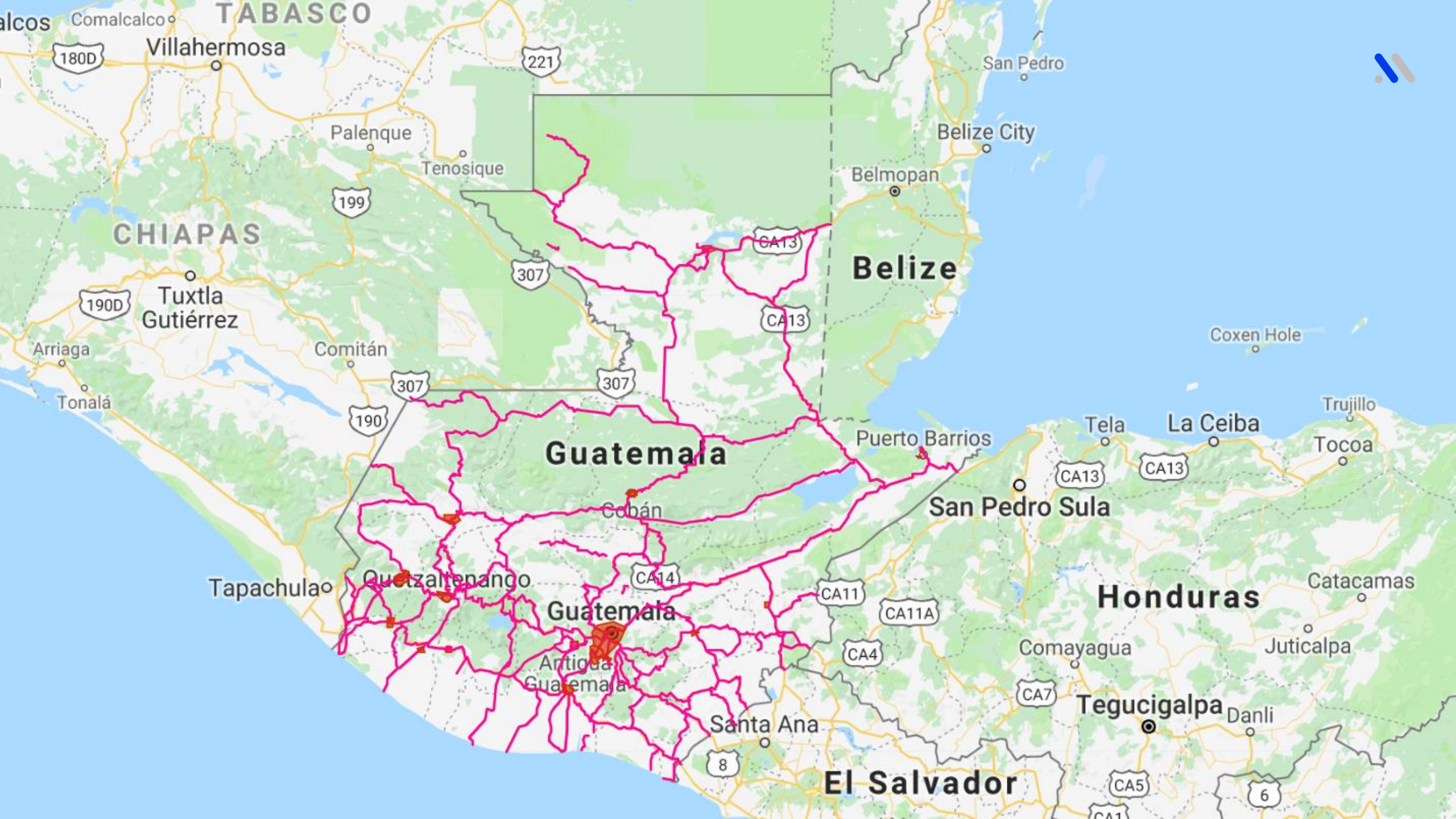
**Máximo número de
establecimientos en
resultado: 60**



**Costo por consulta:
\$0.035**

Estrategia de recolección de data







Puntos por explorar:

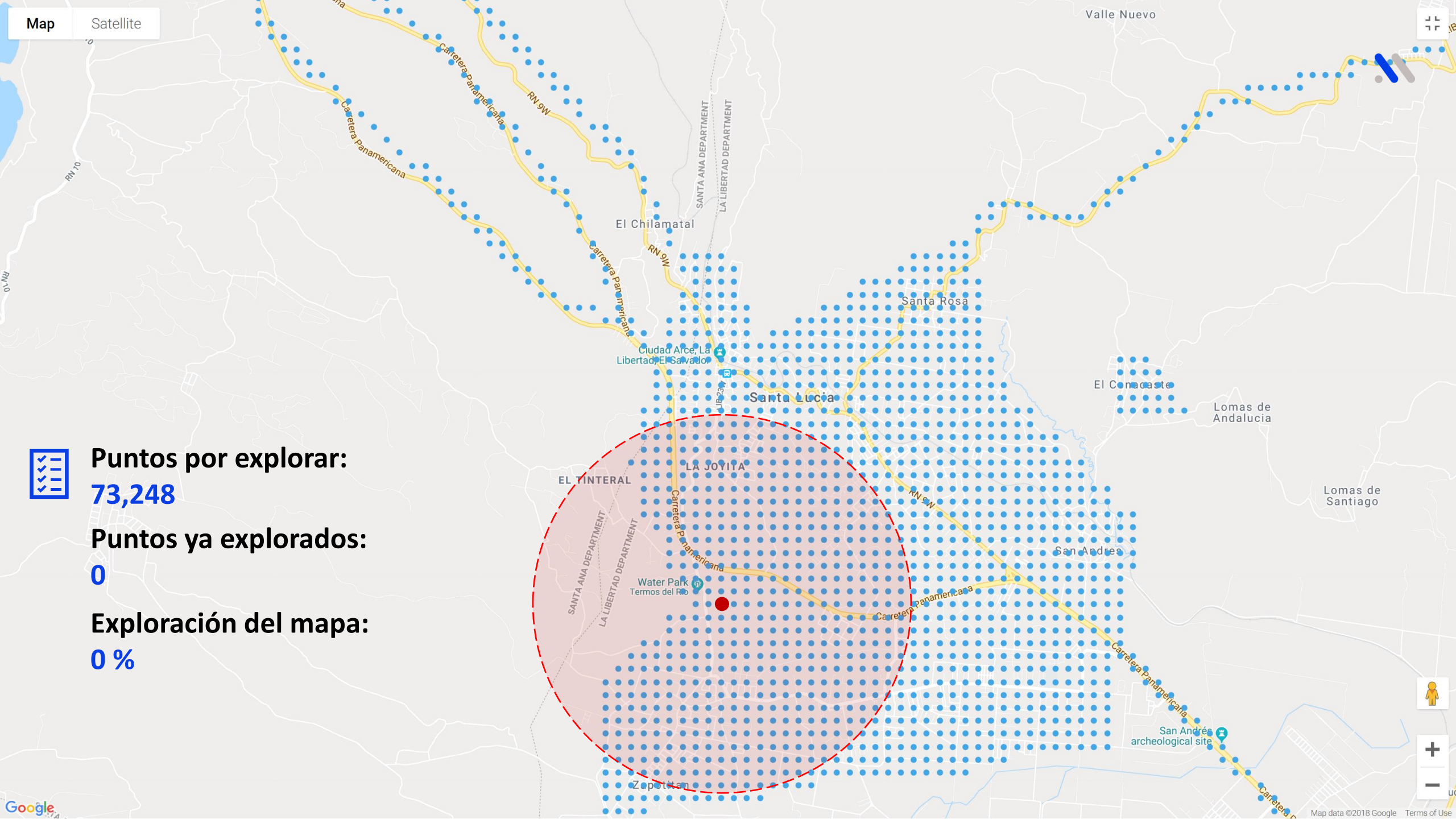
73,248

Puntos ya explorados:

0

Exploración del mapa:

0 %



Map

Satellite



Puntos por explorar:

72,710

Puntos ya explorados:

538

Exploración del mapa:

0.7 %



Puntos por explorar:

72,217

Puntos ya explorados:

1,031

Exploración del mapa:

1.4 %

+

+

+

+

Resultados Google Places

Fetch de data

- 73,248 puntos de exploración en el mapa de Guatemala
- Threshold de 95% en exploración del mapa
- 41,883 establecimientos en 24 categorías
- 2,690 consultas a Google Places API

05.



+

+

+

+

+

+

+

+

+

+

+

+



Continuación: Construcción de un modelo

Stephano Bohl
Data Scientist



bigdata@martinexsa.com
www.martinexsa.com/bigdata

