

# Brewing Data Cup

Datatón, 27-29 de noviembre, 2020

Juan Santos y Juvenal Campos  
Databuesos

2020-11-29

# Descripción del problema



El objetivo es generar un modelo **original, escalable y con lógica de negocio** que permita proponer un **esquema de logística para la repartición de producto proveniente de un Centro de Distribución Regional a lo largo de una semana**, satisfaciendo la demanda local de manera óptima y eficiente.

## Planteamiento del modelo.

- **Primer intento:** Generar un modelo de Programación lineal puro. ❌
  - (Problema: Muchas restricciones, difícil de computar).
- **Segundo intento:** Generar un modelo de *Machine Learning* para generar las zonas de reparto. ❌
  - (Problema: Buena segmentación de los datos, pero mal balanceo)
- **Tercer intento:** Mezclar *ML* con *Programación lineal*. ✅
  - **Resultado:** Buen modelo, aprovechando lo mejor de ambos mundos.

# Lógica del modelo.

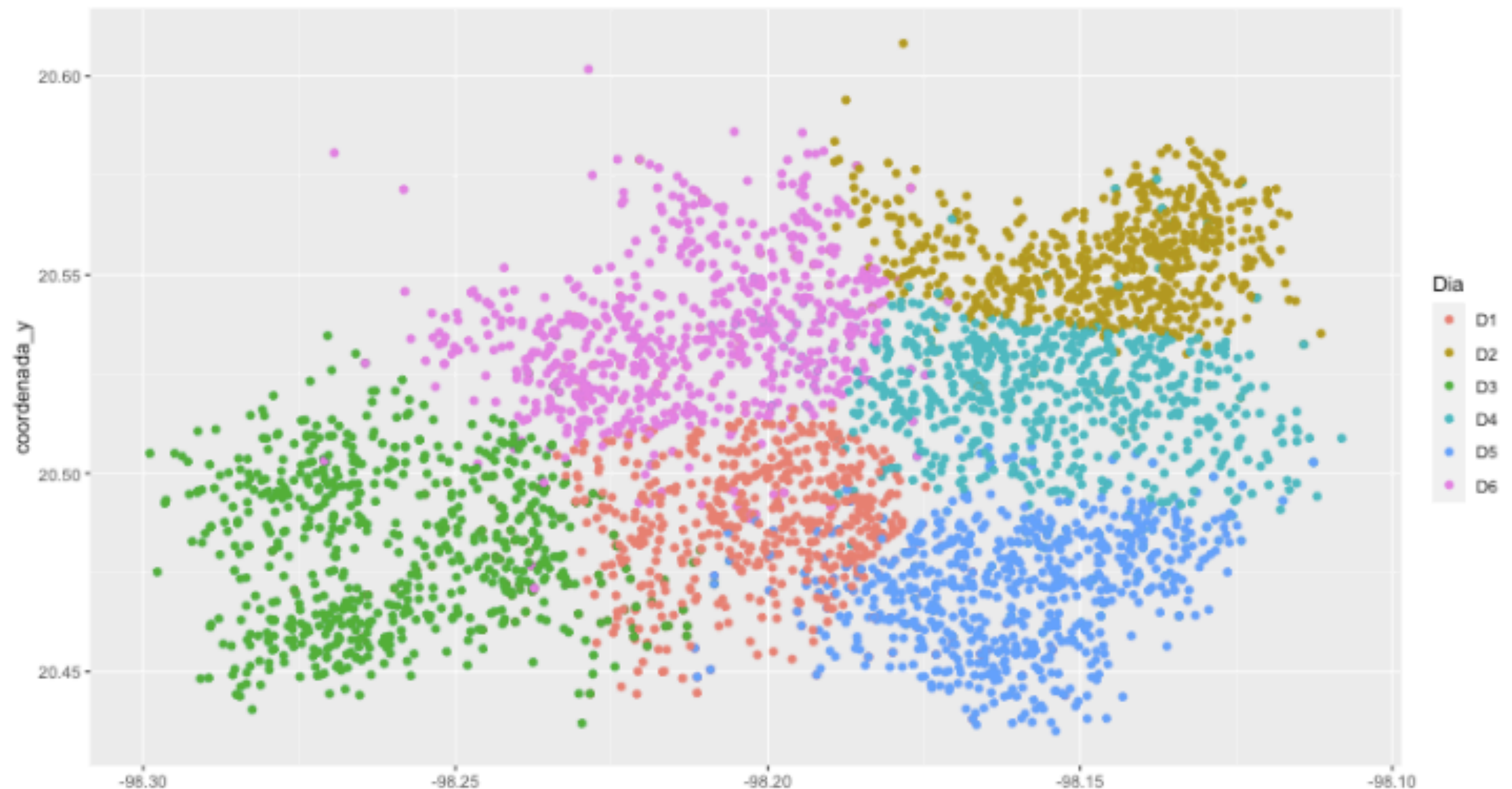
A continuación se muestra el proceso de cálculo seguido por los modelos utilizados para obtener la agrupación en días de reparto.



# Resultados.



El método de solución puede consultarse en el [repositorio de Github](#).



# Resultados.



El método de solución puede consultarse en el [repositorio de Github](#).

	Dia	No. Paradas	Volumen repartido	Distancia
1	D1	666	9,247	7,124
2	D2	656	8,886	5,799
3	D3	650	9,246.67	7,234
4	D4	676	8,886.83	7,493
5	D5	653	9,247.5	6,019
6	D6	676	8,886	7,607

Variable	Desviación estándar	Media	Desviación como pctje de la media
No. Paradas	11.53	662.83	1.73%
Volumen de reparto	197.6	9066.67	2.20%
Distancia entre puntos	774.47	6879.33	11.20%



# Ventajas del Modelo

- **Rápido** (tiempo de cálculo de  $< 2$  min).
- **Baja demanda de CPU y RAM**
- **Escalamiento** a grandes conjuntos de datos
- **Construido con herramientas *Open Source*** (Python: `sklearn`, `pulp`, `pandas`).
- Flexibilidad para que **cada centro personalice** sus zonas de reparto.

# Potencial a futuro.

- Posibilidad de montar en una **aplicación web** o **elaborar un servicio centralizado** para calcular zonas de reparto regionales.
- Eventualmente, se podría robustecer el modelo incorporando en las *restricciones de programación lineal*:
  - Datos de capacidad y número de camiones repartidores.
  - Datos de almacenamiento en centros de reparto.
  - Ubicación de centros de reparto.
  - Cálculo de rutas de calle.
  - Clusterizado con **datos socioeconómicos**, etc.

Entre otros.

# Gracias por su atención

