

# CÓDIGOS DE MINERÍA DE TEXTO PARA COINCIDENCIA DE NOMBRES EN LAS CUENTAS SUBNACIONES DE COSTA RICA

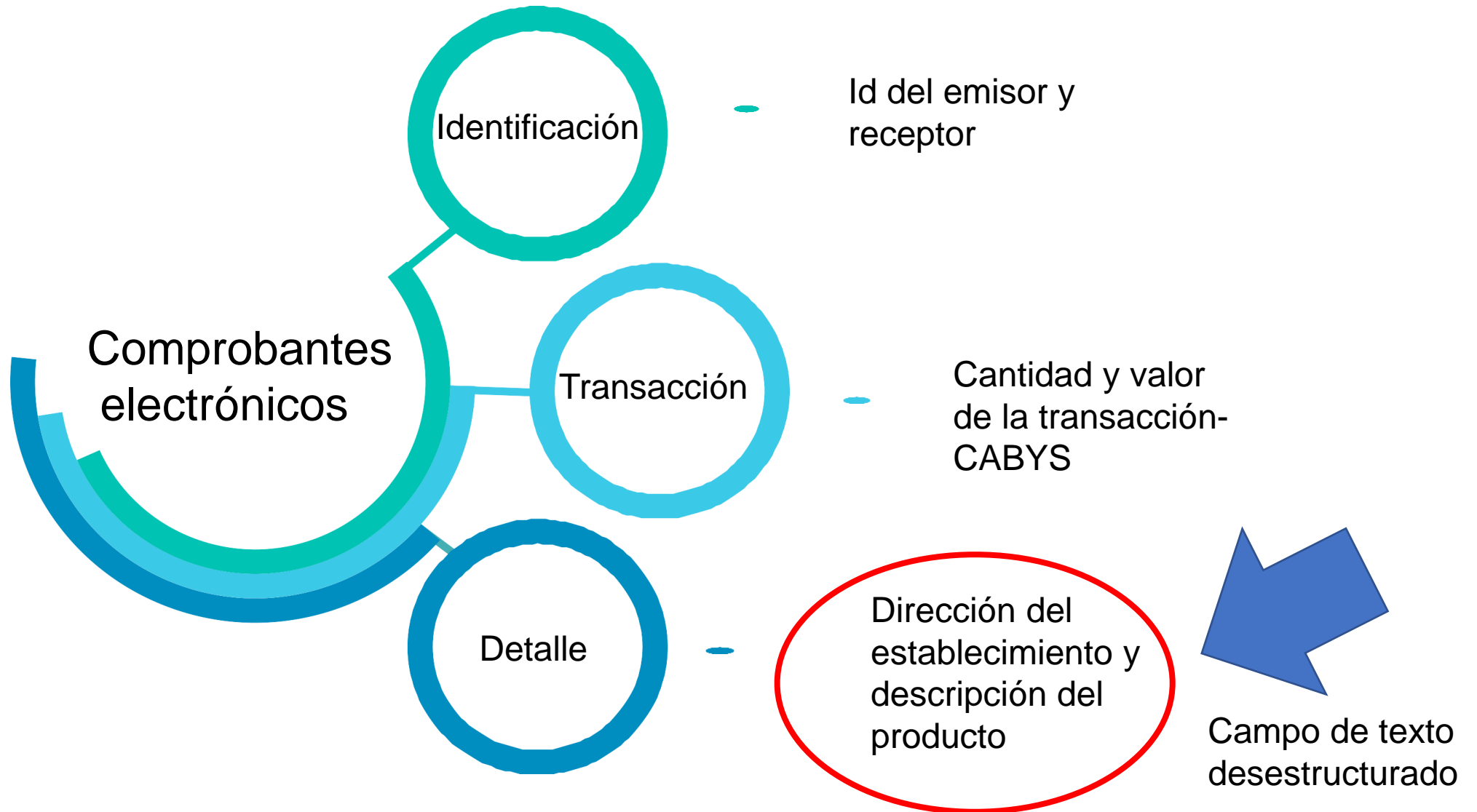
División de Análisis de Datos y Estadísticas

Setiembre 2024

# Contenido

1. Antecedentes
2. Código
3. Etapas
4. Lecciones aprendidas







Indicadores  
subnacionales  
para el PIB  
cantonal



Corroborar  
descripción de  
producto con  
CABYS

## 7 provincias

CodProvincia	Provincia
1	San José
2	Alajuela
3	Cartago
4	Heredia
5e	Guanacast
6	Puntarenas
7	Limón

## 83 cantones

CodCanton	Canton
101	San José
102	Escazú
103	Desamparados
104	Puriscal
105	Tarrazú
106	Aserrí
107	Mora
108	Goicoechea
109	Santa Ana
110	Alajuelita
111	Vázquez de Coronado
112	Acosta
113	Tibás
114	Moravia
115	Montes de Oca
116	Turrubares
117	Dota
n	Canton n

## 490 distritos

CodDistrito	Distrito
60114	Acapulco
60803	Aguabuena
30105	Aguacaliente o San Francisco
21302	Aguas Claras
21004	Aguas Zarcas
20101	Alajuela
11001	Alajuelita
70306	Alegría
20209	Alfaro
60712	Ángeles
11303	Anselmo Llorente
60116	Arancibia
50807	Arenal
10601	Aserrí
20501	Atenas
50401	Bagaces
60504	Bahía Ballena
60506	Bahía Drake
n	Distrito n

<b>Año</b>	<b>Cantidad de registros</b>	<b>Crec. %</b>	<b>Teras</b>
2020	11 818 715 661		1,25
2021	13 740 475 496	16,3%	1,45
2022	14 824 125 783	7,9%	1,57
2023	16 010 584 373	8,0%	1,70



## Prueba de concepto (Poc)

El objetivo es desarrollar un modelo personalizado para extraer información del campo de texto libre y asignar el catálogo de ubicación.

## Requisitos



Acceso al grupo de recursos



Creación de recursos



Acceso a los datos



Análisis de coincidencia



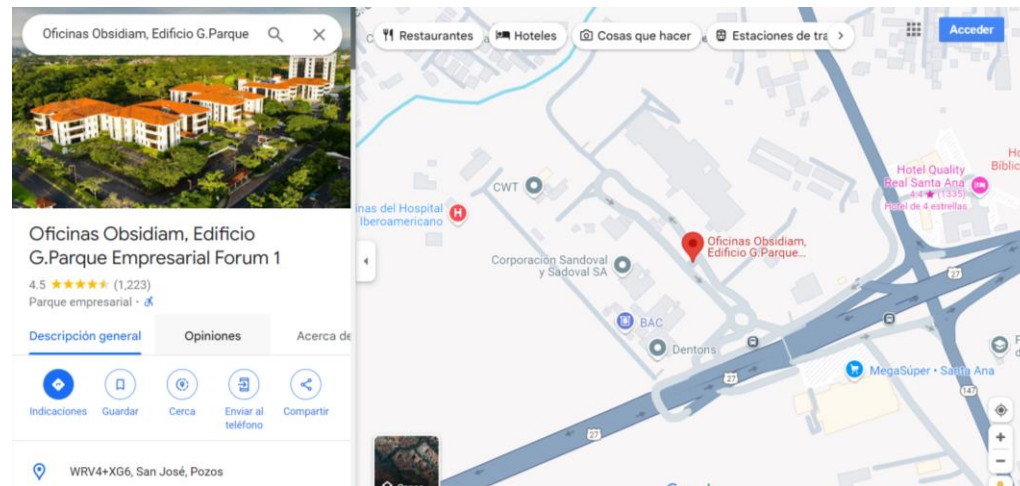
Otras_Senas	provincia	canton	distrito	codigo_provin...	codigo_canton	codigo_distrito	error
parque empresarial forum ii	San José	Santa Ana	Pozos	1	109	10903	
costado oeste de la municipalidad de san ...	San José	San José	Carmen	1	101	10101	
costado norte plaza de peaje ruta 27	San José	Escazú	San Rafael	1	102	10202	
san José guachipelin de escazú oficentro ...	San José	Escazú	San Rafael	1	102	10202	
parque empresarial forum ii	San José	Santa Ana	Santa Ana	1	109	10901	
san jose 250 metros norte torre mercedes	San José	San José	Mercedes	1	101	10102	
san vito centro	Puntarenas	Coto Brus	San Vito	6	608	60801	
interior plaza paraíso	San José	Escazú	San Rafael	1	102	10202	
parque empresarial forum ii	San José	Santa Ana	Santa Ana	1	109	10901	
de la embajada de cuba cien metros este ...							Lo siento
costado sur del parque de santa barbara	Heredia	Santa Bárbara	Santa Bárbara	4	404	40401	
100 oeste entrada de tierra blanca	San José	Aserri	San Gabriel	1	106	10604	

Lo siento



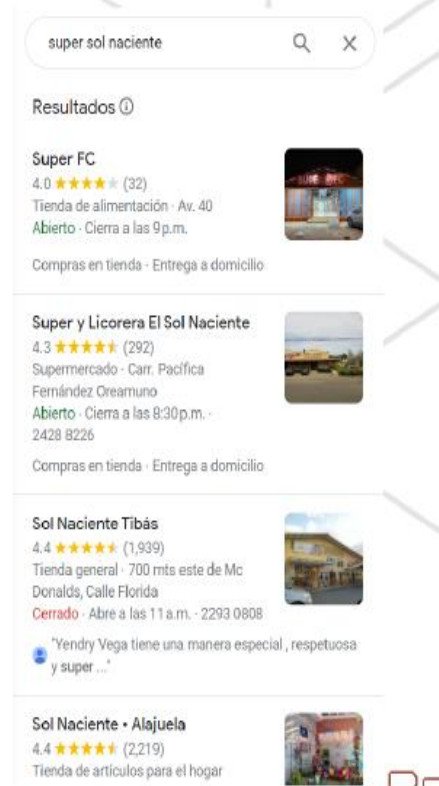
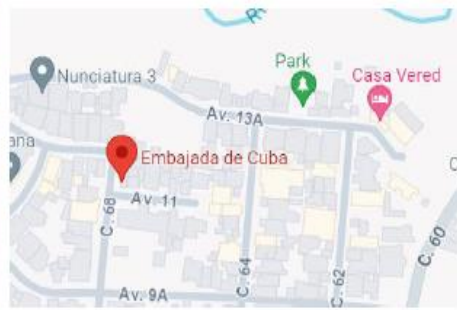
En caso de campos en blanco o descripción muy reducida.

Hay puntos de referencia que el código busca en internet.

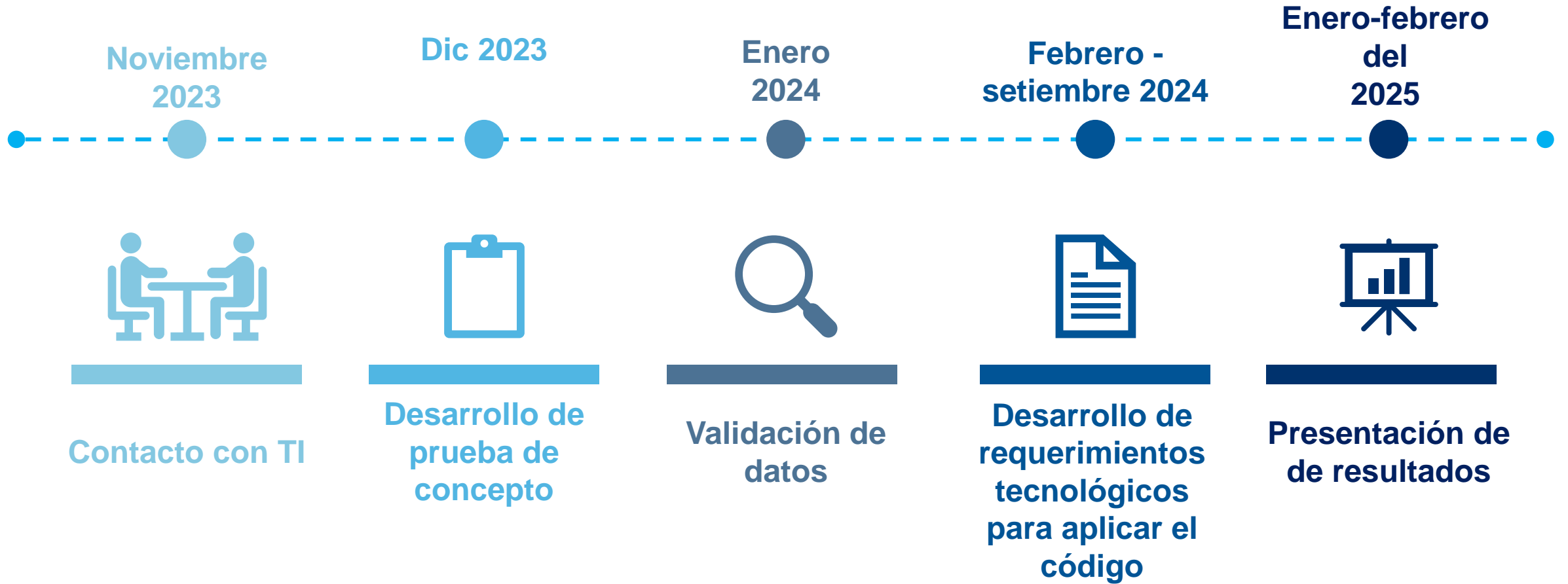


# Aspectos a considerar

- Para mejorar la precisión es necesario crear una base de datos de puntos de interés conocidos por los costarricenses.



# Etapas



# 4

## Lecciones aprendidas

- Proveedor de servicios en la nube
- Arquitectura y almacenamiento de la base de datos
- Elección de software más apropiado y capacitación del recurso humano
- Buena conexión entre el equipo interdisciplinario
- Políticas claras de gobernanza de los datos
- Laboratorio de datos



# CÓDIGOS DE MINERÍA DE TEXTO PARA COINCIDENCIA DE NOMBRES EN LAS CUENTAS SUBNACIONES DE COSTA RICA

División de Análisis de Datos y Estadísticas

Setiembre 2024