



# Seminario de alto nivel sobre la integración de fuentes de datos no tradicionales en los sistemas estadísticos nacionales

## Web Scraping Masivo de Alquileres de Viviendas

**Giovanni Savio, María Paz Collinao, Bruno Lana y Rodrigo Lara**

Unidad de Estadísticas Económicas y Ambientales, División de Estadísticas  
Comisión Económica para América Latina y el Caribe (CEPAL)



- ▶ Los alquileres / arriendo de vivienda son parte importante de varios indicadores económicos de uso habitual:
  - Estimación del PIB
    - enfoque de la producción
    - enfoque del gasto
  - Estimación del IPC
  - Estimación de las PPAs
  
- ▶ Los alquileres los podemos dividir en:
  - Efectivos
  - Imputados



# Tipos de alquileres

	Efectivos	Imputados
Método de Estimación	Encuestas de gastos de los hogares	<ul style="list-style-type: none"> <li>• Método de estratificación/ renta equivalente.</li> <li>• Método de los costos.</li> </ul>
Indicadores	PIB	PIB
	PPA -2017	PPA -2017
	PPA-2011	
	IPC	

Fuente: Elaboración propia



► Estimaciones de los alquileres efectivos e imputados de los países OCDE.

Consumo final de los servicios de la vivienda 2011, como % del PIB

País	Actual Rent	Imputed Rents	Total Rents
Canada	2,7	8,1	10,8
Czech Republic	2,0	6,4	8,4
Estonia	1,2	5,9	7,0
France	2,3	7,9	10,2
Germany	3,8	5,3	9,0
Hungary	0,4	6,2	6,6
Poland	0,8	3,5	4,3
Slovak Republic	0,6	4,4	4,9
Slovenia	0,4	6,1	6,5
Spain	1,5	7,0	8,5
United Kingdom	3,3	9,7	13,0
United States	2,6	8,1	10,7

Fuente: OECD



CEPAL



# ¿Por qué incursionamos en web scraping de precios?

- ▶ Aprovechar los avances técnicos existentes
- ▶ Existe evidencia, experiencia y documentación al respecto
- ▶ Hay países en la región que han avanzado en la recolección web
- ▶ Existen países interesados en colaborar con nuestra investigación



Centraal Bureau  
voor de Statistiek



UNITED NATIONS

ECLAC



NACIONES UNIDAS

CEPAL



Se busca constatar la **factibilidad** y la **efectividad** de realizar una recolección de precios en línea.

- ▶ **Factibilidad:** Se busca recolectar precios de arriendos de inmuebles (casas y apartamentos) en diferentes países de la región
- ▶ **Efectividad:** Se pretende comparar los precios obtenidos con los recolectados mediante métodos tradicionales



## 3 países de estudio:

- ▶ Brasil (BR-SP, BR-RJ)
- ▶ Ecuador (EC-QT, EC-GY)
- ▶ Perú (PE-LM)

## 13 productos del PCI:

- ▶ 7 tipos de casas
- ▶ 6 tipos de apartamentos

## 5 semanas de recolección:

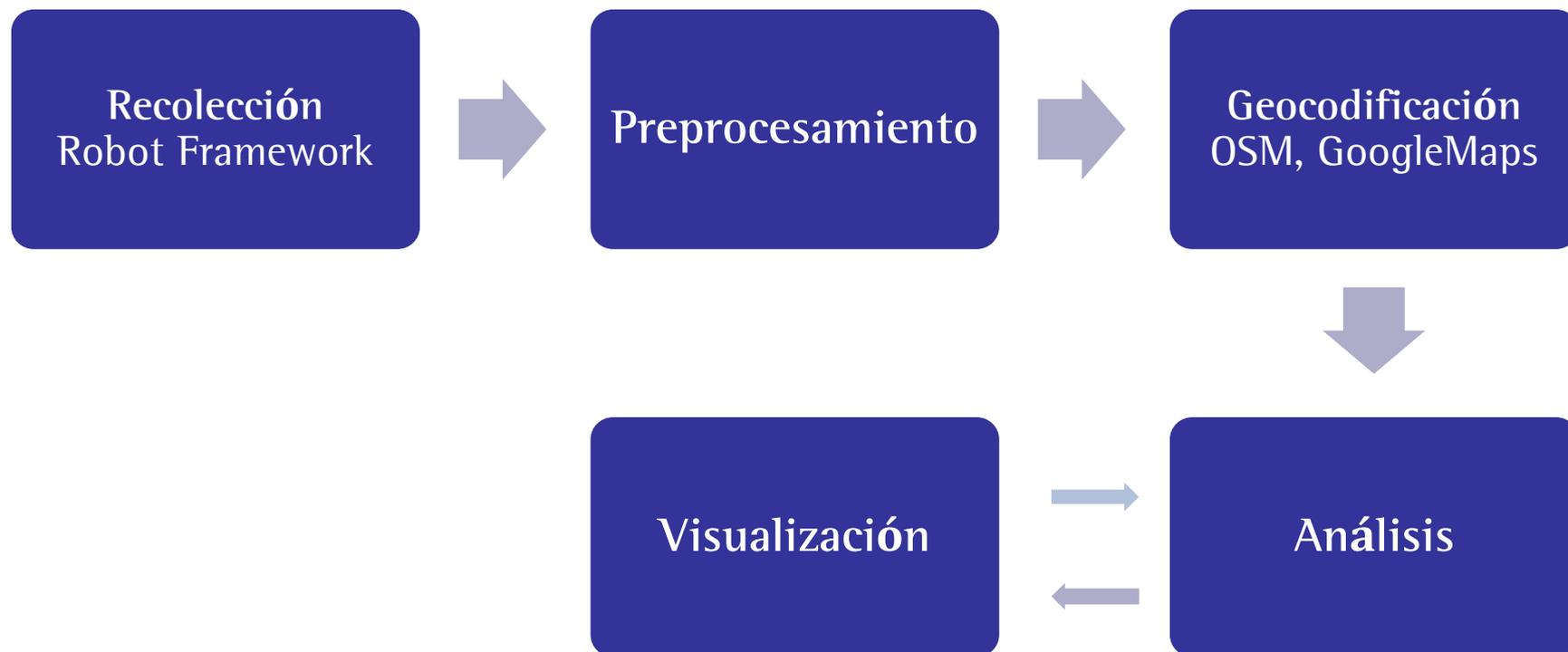
- ▶ Desde la semana del 22 de julio,
- ▶ Hasta la semana del 20 de agosto de 2018.



NACIONES UNIDAS

CEPAL

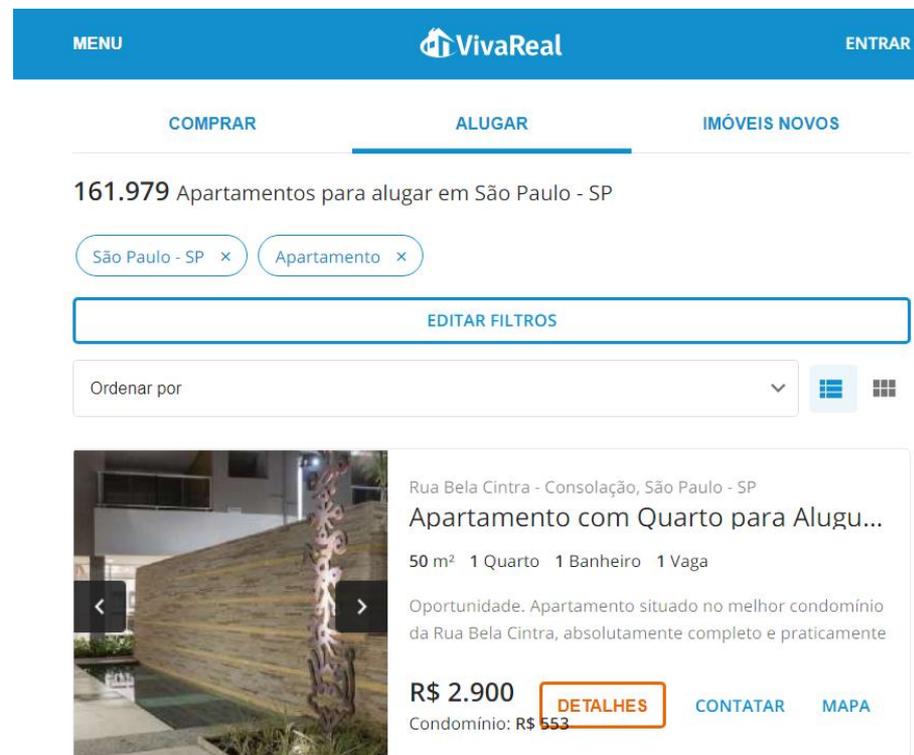




Etapas del estudio

# Robot Framework

- ▶ Desarrollado en Node.js por CBS (Central Bureau of Statistics - Holanda)
- ▶ La configuración es a través de un archivo de texto JSON
- ▶ Genera un archivo CSV con la información recolectada



The screenshot displays the VivaReal website interface. At the top, there is a blue navigation bar with 'MENU', the VivaReal logo, and 'ENTRAR'. Below this, three tabs are visible: 'COMPRAR', 'ALUGAR' (which is selected and highlighted), and 'IMÓVEIS NOVOS'. The main content area shows '161.979 Apartamentos para alugar em São Paulo - SP'. There are two filter buttons: 'São Paulo - SP' and 'Apartamento'. Below the filters is a search bar with the text 'EDITAR FILTROS'. A dropdown menu for 'Ordenar por' is visible. The main listing features a photograph of a modern building at night. To the right of the image, the text reads: 'Rua Bela Cintra - Consolação, São Paulo - SP', 'Apartamento com Quarto para Alugu...', '50 m² 1 Quarto 1 Banheiro 1 Vaga', and 'Oportunidade. Apartamento situado no melhor condomínio da Rua Bela Cintra, absolutamente completo e praticamente'. The price is listed as 'R\$ 2.900' with a 'DETALHES' button next to it. Below the price, it says 'Condomínio: R\$ 553'. There are also 'CONTATAR' and 'MAPA' buttons.



CEPAL



## Análisis

- ▶ Se eliminan datos innecesarios
- ▶ Se filtran los datos según los requerimientos del PCI:
  - Apartamentos de 1 y 2 dormitorios, con superficie entre 15 y 80 m<sup>2</sup>
  - Casas con superficie entre 80 y 460 m<sup>2</sup>

## Geocodificación

- ▶ Consultas a las API's de Open Street Map y de Google Maps
- ▶ Los puntos obtenidos se intersectan con las entidades geoespaciales de interés



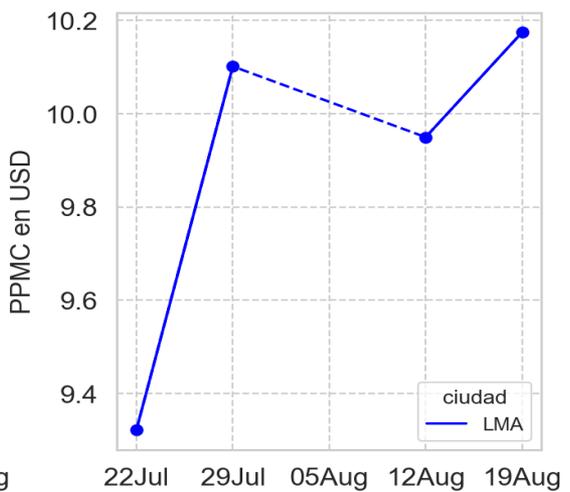
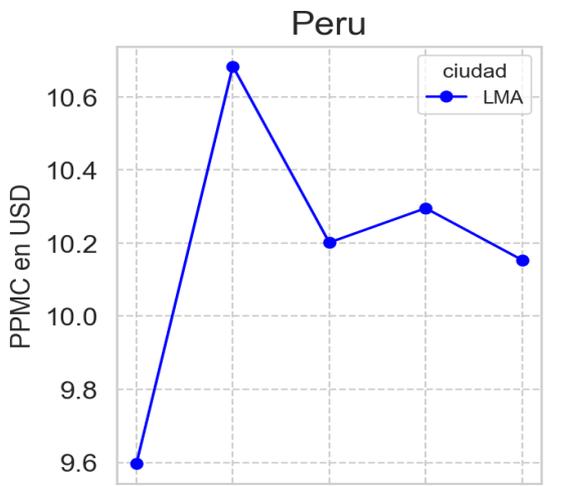
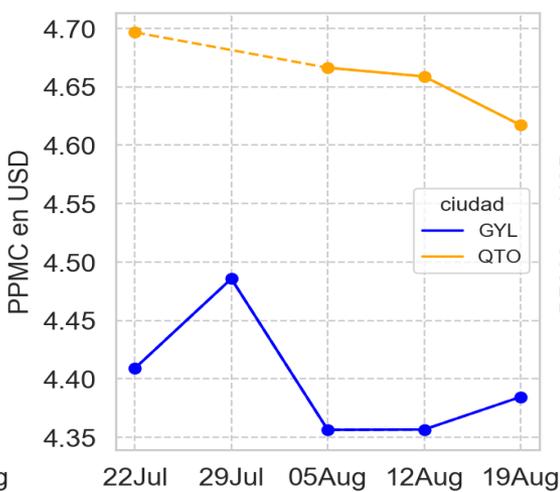
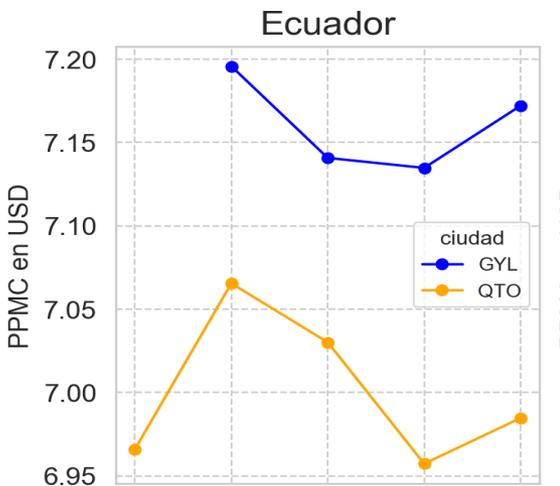
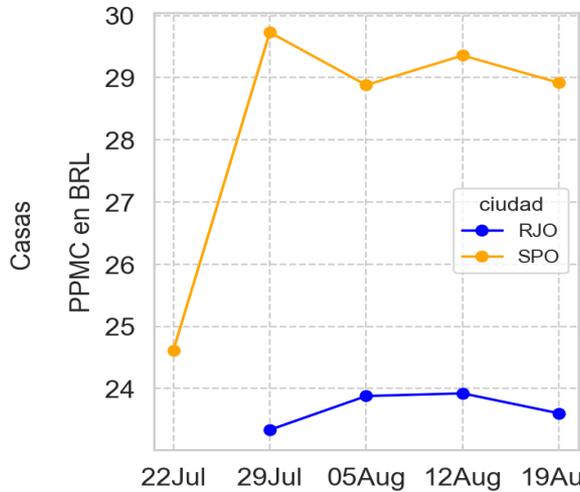
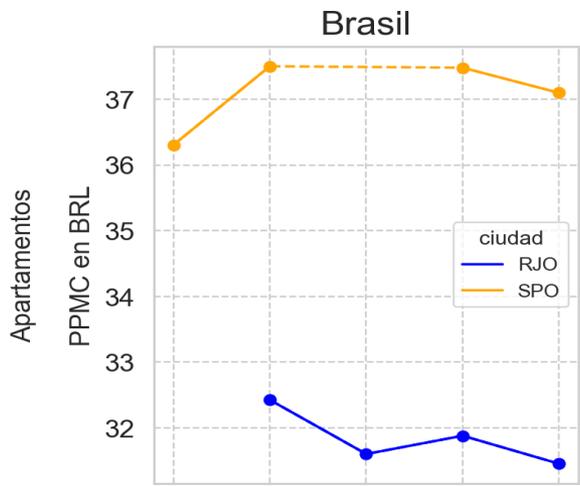
NACIONES UNIDAS

CEPAL



- ▶ Obtención del precio por metro cuadrado de cada inmueble y eliminación de valores anómalos (por ejemplo, ventas de propiedades)
- ▶ Agregación por zonas geográficas para el cálculo de estadísticas descriptivas
- ▶ Categorización de tipo de producto en base a:
  - Tipo de vivienda: casa o apartamento
  - Superficie
  - Número de dormitorios
  - Número de ambientes: uno o multi-ambiente



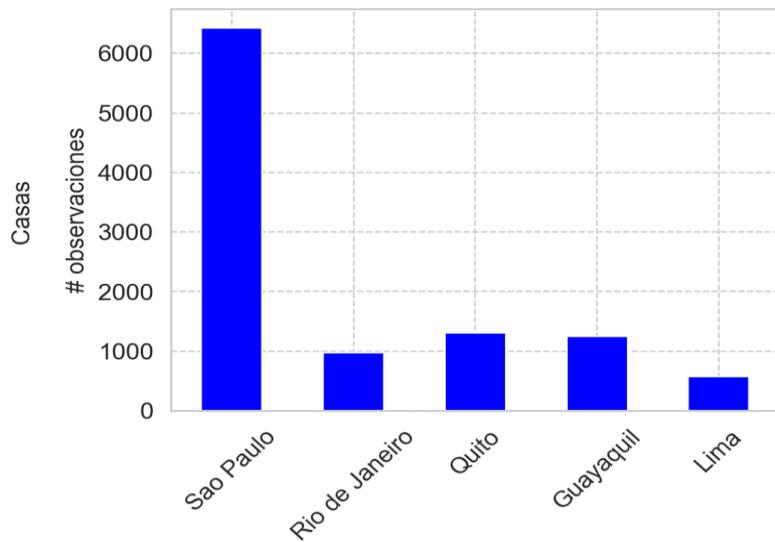
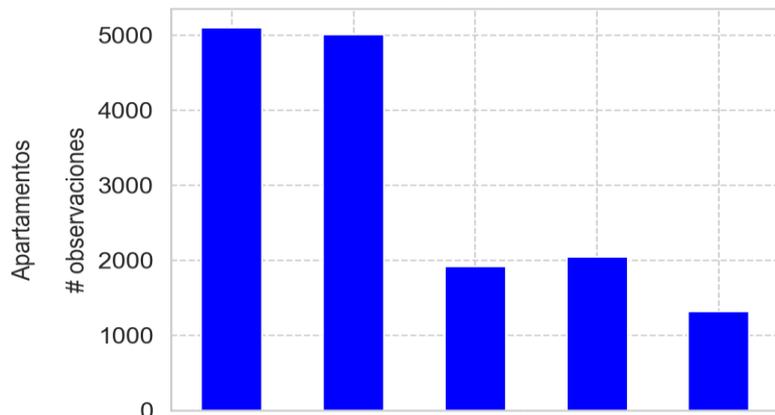


# Análisis temporal de precios

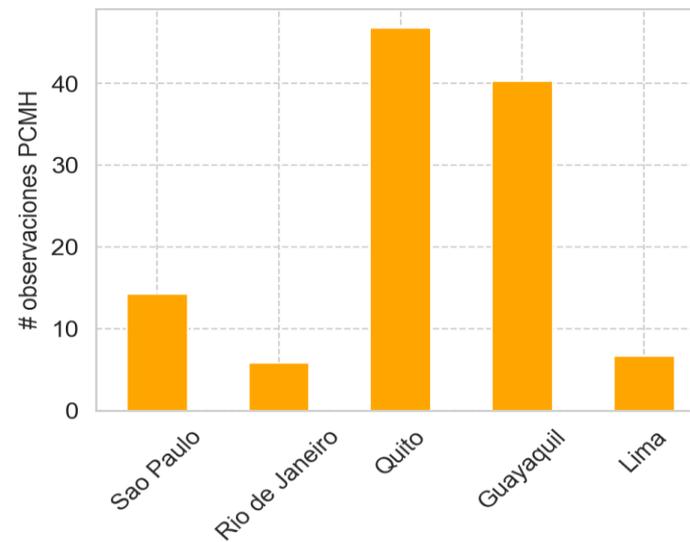
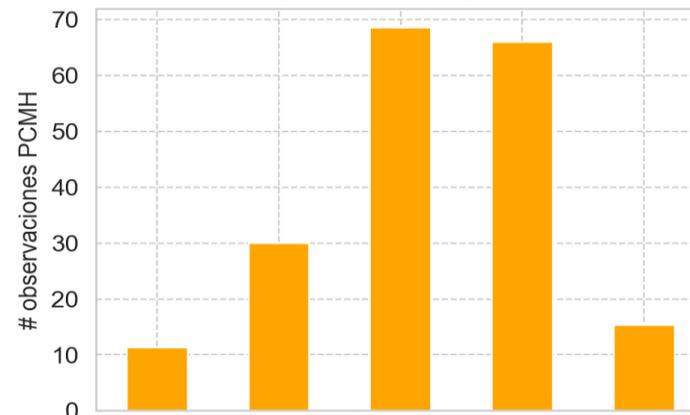




Promedio de observaciones

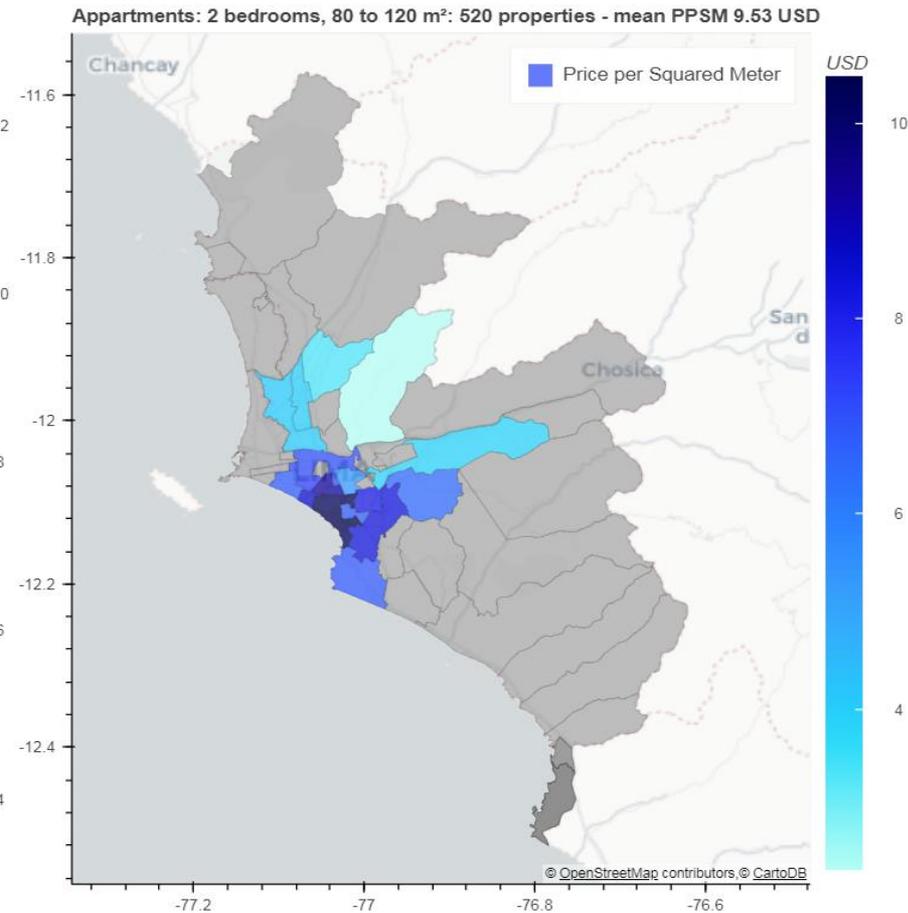
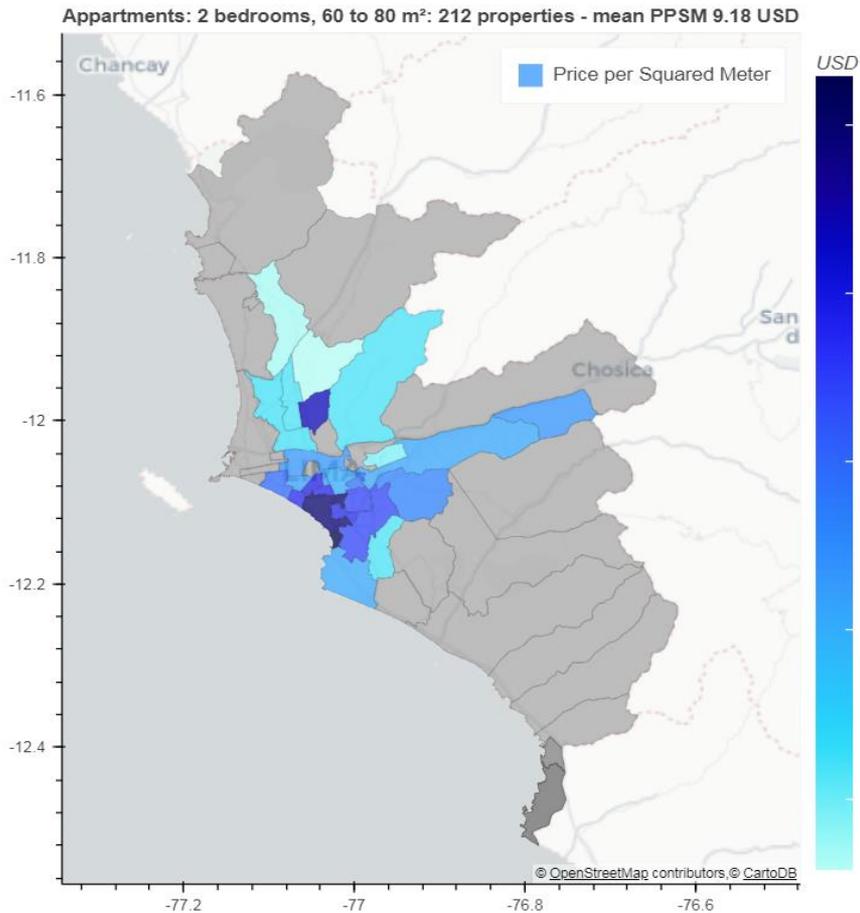


Promedio de observaciones por cada 100 mil habitantes



# Análisis temporal de número de observaciones





## Análisis espacial - Lima



CEPAL



## Ventajas

- ▶ Información en tiempo real
- ▶ Reducción en los costos de recolección y procesamiento de información
- ▶ Mayor cobertura y cantidad de información
- ▶ Análisis geográfico

## Desventajas

- ▶ No hay muestreo asociado a la recolección de datos
- ▶ Los valores obtenidos son precios ofertados no alquileres reales
- ▶ La recolección no es trivial y se necesita un **perfil profesional adecuado** que elabore y supervise el proceso de recolección



CEPAL



Muchas gracias por su atención.

