

Trabajo Final ///

Análisis de correlación entre precio de venta (por m2) de viviendas y cercanía a estaciones de subte en la Ciudad Autónoma de Buenos Aires.



Julieta Giovannoni
Rodrigo Hermoso
Josefina Tamis
Javier Yanzón

Scope de proyecto ///

Desafíos Open Data / Urbanismo ///

¿Existe una correlación entre precio de venta (por m²) de viviendas y la cercanía a estaciones de subte en Ciudad Autónoma de Buenos Aires?



Problema a Resolver //

En el presente proyecto queremos abordar el siguiente interrogante: ¿cómo se relaciona la presencia de estaciones próximas de subte con el precio de venta (por m²) de una vivienda en la Ciudad Autónoma de Buenos Aires? A priori suponemos que es muy probable que cuanto más cercano se encuentre de un medio masivo de transporte (como el subte) el precio de venta de una propiedad debería ser mayor.

Beneficios del proyecto //

El beneficio con este análisis es conocer cómo influye en una determinada zona la presencia de un medio masivo de transporte y la plusvalía que genera una determinada infraestructura en zonas cercanas debido a la mayor accesibilidad.

Notebook resumen // [Link.](#)

Scope de proyecto // [Link.](#)

EANT

Python Data Analytics

Lenguaje de programación // Python.

Para los análisis realizados en el trabajo utilizamos los siguientes paquetes:

Pandas / Manejo de dataframes.

Numpy / Funciones adicionales.

Folium / Generación de mapas interactivos.

Geopandas / Manipulación y visualización de GeoDataframes.

Geopy / Cálculo de distancias (con lat y long).

Plotly / Armado de gráficos interactivos.

Seaborn / Armado de gráficos.

BeautifulSoup / Web scraping.

Request / Consultas a API.

Pingouin / Análisis de correlación.

WordCloud / Nube de palabras.

GeoCoder / Cálculo de latitud y longitud.

1 / Análisis de los datos - Workflow ///

1. Integración de los Datos



 DB **Properati**
Scraping **Argenprop**

GeoDataframes del
Gobierno de la Ciudad:

 **Subtes**

 **Barrios**


 **Ecobicis**


 **Metrobus**


 **Trenes**

2. Preparación y transformación de los Datos



 Limpieza de la base de datos:
eliminación de datos faltantes y
corrección de datos. Tratamiento de
Outliers.


 Limitación del análisis a los
departamentos en venta entre 1 y 3
amb en Capital Federal publicados en
el 1Q2020.


 Creación de variables:


- Distancia al subte más cercano
- Estación de subte más cercana
- Precio m2


3. Visualización y Análisis de los Datos



 Visualización de la información.

 Creación de mapas.

 Análisis de Correlación.

 Nube de palabras.

4. Conclusiones



1. Integración de los datos ///

EANT

Python Data Analytics

Datos //

A /// Estaciones de Subte de la ciudad.



**Datos
abiertos
GCBA //**

. **Subte:** Estaciones. (Archivo .geojson)
*Ubicación geográfica de las líneas y
estaciones de subte.*
Shape = (90,5)

. **Polígonos de los barrios de CABA**
. **GeoJson de Ecobicis, Metrobús, Trenes**
. **Cantidad de viajes en Subte 2019**

B /// Precio m².



Argenprop //

Realizamos un scrapping de la web de ArgenProp datos con las siguientes características:

- Localizados en Capital Federal
- En venta entre 1 y 3 ambientes
- Con información del precio
- Precio de venta entre USD 5.000 y USD 1.000.000

A esta base se le obtuvo la latitud y longitud utilizando OSM utilizando la librería Geocoder.

Luego, haremos parte del análisis que veremos más adelante pero sin llegar a muchas conclusiones, por lo que decidimos utilizar Properati.

B /// Precio m².



Properati //

Tomaremos la base que publica Properati para venta (alrededor de 1 millón de registros).

Es una base mucho más ordenada y con información más precisa y que nos conviene para el análisis a realizar.

Notebook.

2. Preparación y transformación de datos ///

EANT

Python Data Analytics

B /// Precio m².



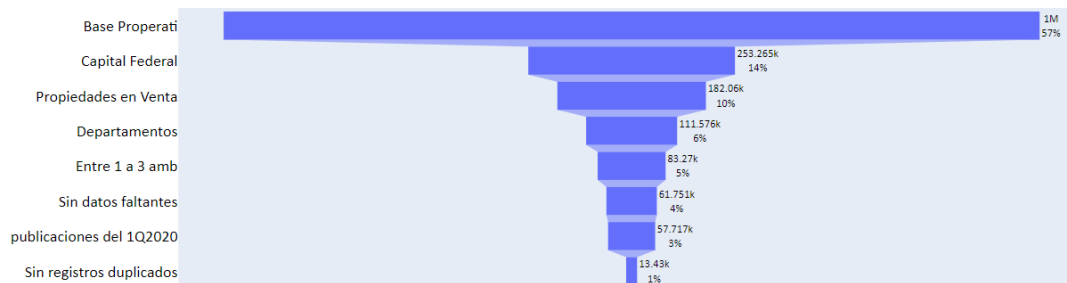
Properati //

Limpieza de la base Partimos de 1 millón de datos y realizamos las siguientes transformaciones:

- Filtramos las publicaciones de Argentina (983.764)
- Filtramos las publicaciones de CABA (253.265)
- Filtramos únicamente propiedades en venta (182.060)
- Filtramos únicamente departamentos (111.576)
- Filtramos únicamente deptos entre 1 y 3 ambientes (83.270)
- Eliminamos registros sin precio publicado (81.034)
- Eliminamos registros sin el dato de cantidad de m2 de la propiedad (62.473)
- Filtramos únicamente propiedades con precios en dólares (61.751)
- Calculamos el precio del m2 como Precio/Superficie cubierta
- Renombramos Columnas
- Filtramos los datos con información de la latitud y longitud de la propiedad (57.717)
- Para evitar demoras en el procesamiento de datos, trabajamos únicamente con las publicaciones del 1Q2020 (13.430) y eliminamos las publicaciones duplicadas (11.829)



Proceso Limpieza DB Properati






2. Preparación y transformación de datos ///

EANT

Python Data Analytics

C /// Cruce de bases.



-  a) Utilizando Geopy **calculamos la distancia** de cada propiedad a la **estación de subte más cercana** y adicionalmente **generamos una columna** con el nombre de la estación más cercana
-  b) Trabajamos únicamente con las publicaciones del **1Q2020 (13.430)** y eliminamos las publicaciones duplicadas (11.829). De esta manera acotamos un poco el análisis y **evitamos demoras** en el procesamiento
-  c) Realizamos la **transformación de los siguientes datos**:
 - **Renombramos los barrios** para llamarlos de la misma forma que figuran en la base de "Barrios" del gobierno de la ciudad
 - **Reemplazo de outliers** (percentiles 2 y 95)
- d) Agrupación líneas N-S y E-O.

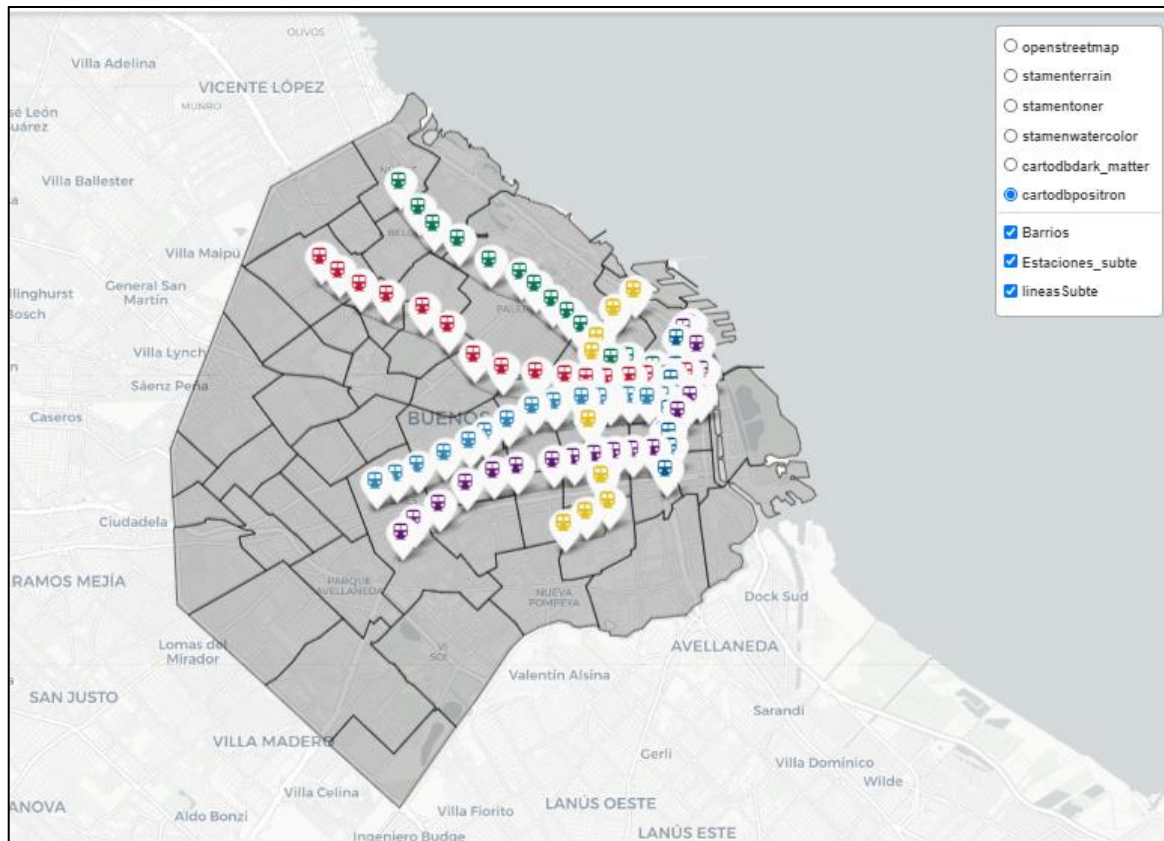
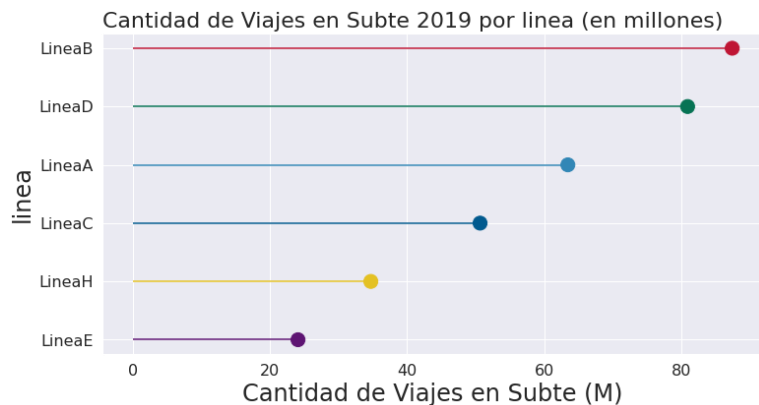
3. Visualización y análisis de los datos ///

EANT

Python Data Analytics

1 / Estaciones de Subte de la Ciudad ///

Línea	Año Inauguración	Longitud (km)	Estaciones
A	1913	9,8	17
B	1930	11,9	17
C	1934	4,5	9
D	1937	10,4	16
E	1944	12	18
H	2007	8,8	12



Python Data Analytics

Nube de palabras de descripciones.

The map displays the city of Buenos Aires, Argentina, with its various neighborhoods and districts. The map is color-coded based on the 'Precio Promedio m2' (Average Price per m2) variable, with a color scale ranging from 1,823 (light yellow) to 4,872 (dark red). The map also shows the distribution of subway lines, with points representing the 'Cant. Publicaciones' (Number of Publications) variable. The legend on the right lists the subway lines and the variables being mapped.

Variables:

- ☐ Cercanos_Linea_A
- ☐ Cercanos_Linea_B
- ☐ Cercanos_Linea_C
- ☐ Cercanos_Linea_D
- ☐ Cercanos_Linea_E
- ☐ Cercanos_Linea_H
- ☒ Linea A
- ☒ Linea B
- ☒ Linea C
- ☒ Linea D
- ☒ Linea E
- ☒ Linea H
- ☒ lineasSubte
- ☐ precio_prom_m2
- ☒ cant_publicaciones

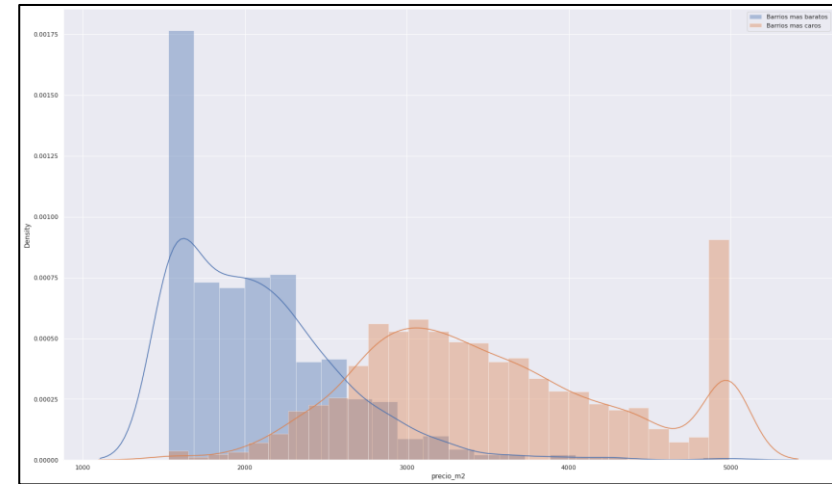
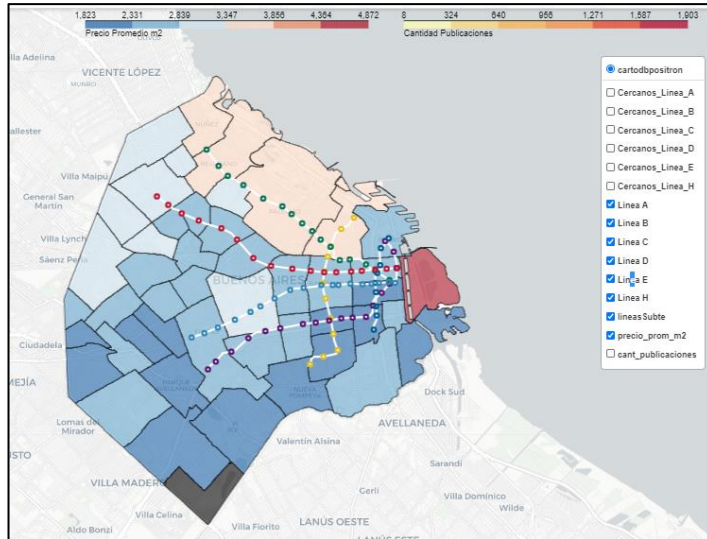
A horizontal bar chart titled 'Cantidad Publicaciones' on the x-axis and 'precio_prom_m2' on the y-axis. The x-axis ranges from 0 to 2000 with major ticks at 0, 500, 1000, 1500, and 2000. The y-axis lists 15 neighborhoods. Each bar is labeled with its value at the end. The bars are color-coded according to the average price per square meter, with a color scale on the right ranging from 2400 (dark blue) to 3600 (dark red). The data is as follows:

Neighborhood	Cantidad Publicaciones	precio_prom_m2 (approx.)
PALERMO	1900	3600
ALMAGRO	970	2800
VILLA CRESPO	950	2800
BELGRANO	930	3600
CABALLITO	860	2800
RECOLETA	840	3200
VILLA URQUIZA	550	3000
BALVANERA	460	2400
COLEGIALES	400	3000
NUÑEZ	330	3400
PARQUE CHACABUCO	270	2400
FLORES	260	2400
SAN NICOLAS	220	2600
VILLA DEL PARQUE	190	2800
MONSERRAT	190	2400

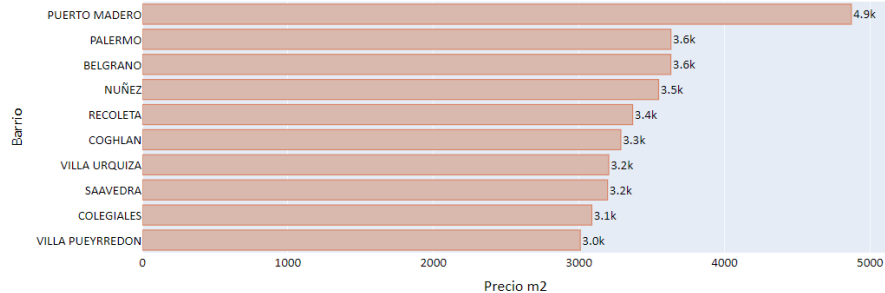


3. Visualización y análisis de los datos ///

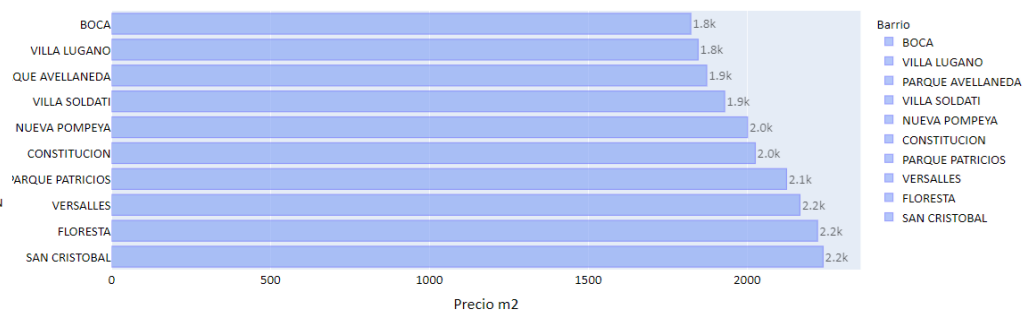
Distribución acorde a precios de m2 por barrio.



Los 10 Barrios mas caros CABA



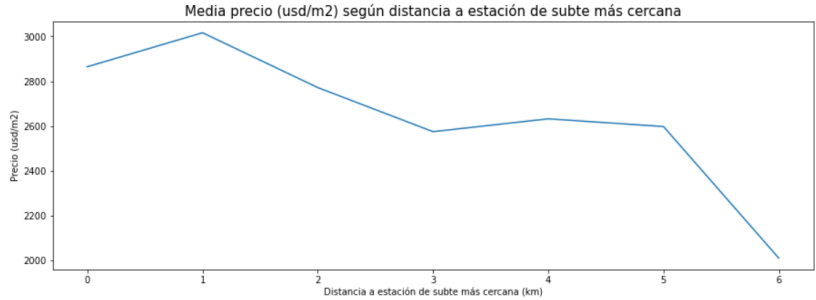
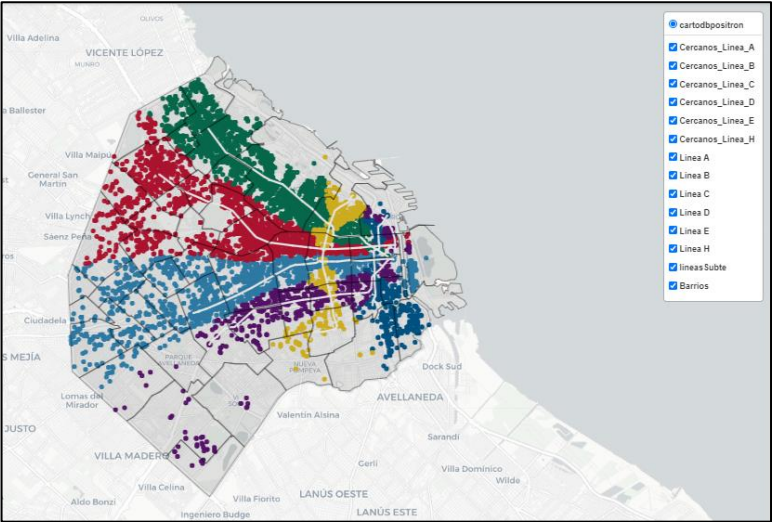
Los 10 Barrios mas baratos CABA



3. Visualización y análisis de los datos ///

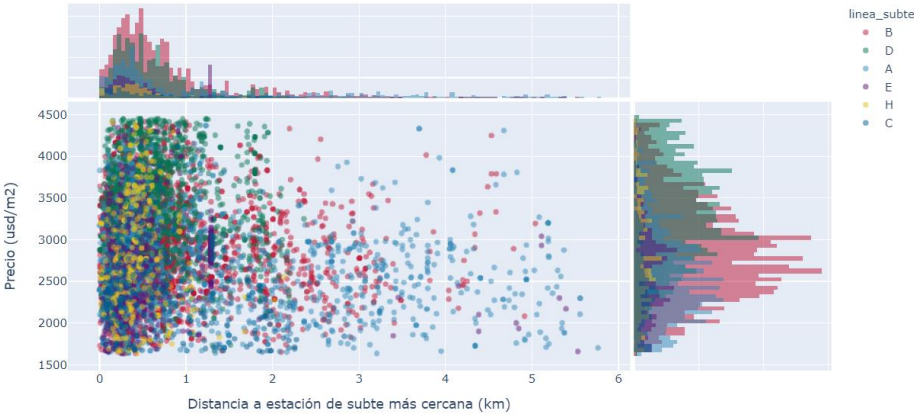
3 / Cruce de bases ///

Propiedades según línea de subte más cercana.



Precio promedio y monto propiedad por cercanía subte

Línea	# Publicaciones	Precio m2		Monto	
		Media	Desvío	Media	Desvío
A	2,104	2,724	831	135,047	68,909
B	4,164	2,898	650	139,504	57,947
C	560	2,420	688	125,341	70,417
D	3,462	3,569	807	187,329	77,215
E	926	2,673	800	141,972	76,745
H	613	3,009	916	158,715	81,693

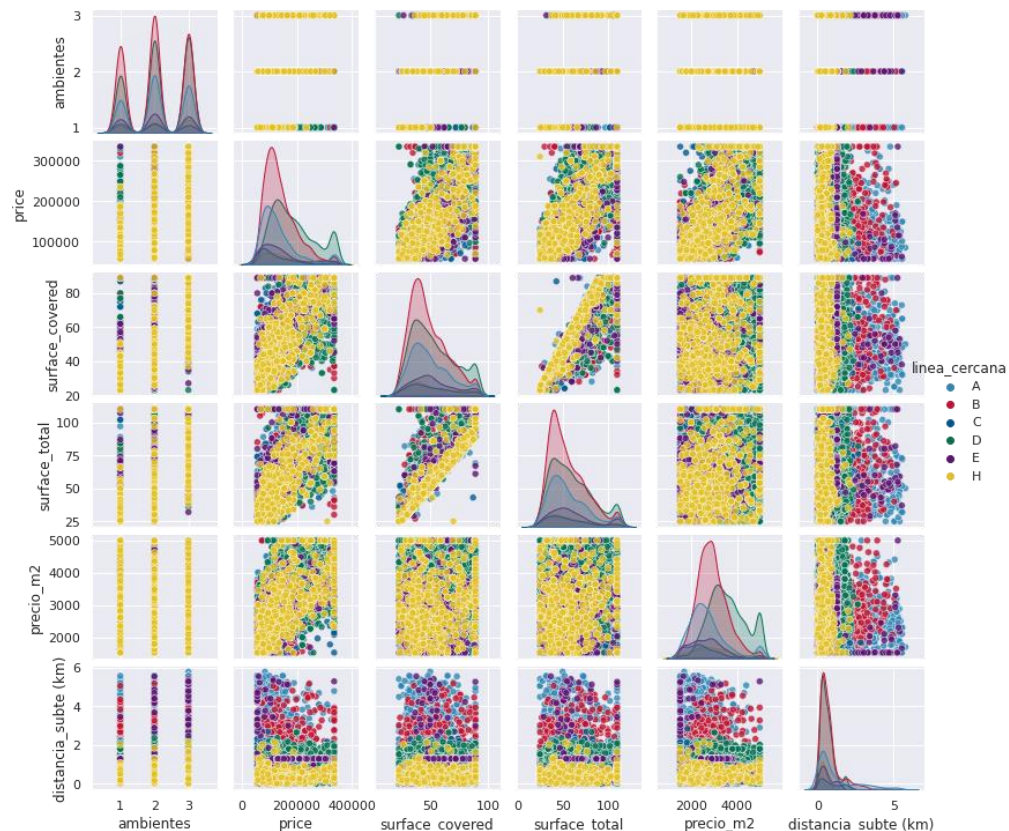


3. Visualización y análisis de los datos ///

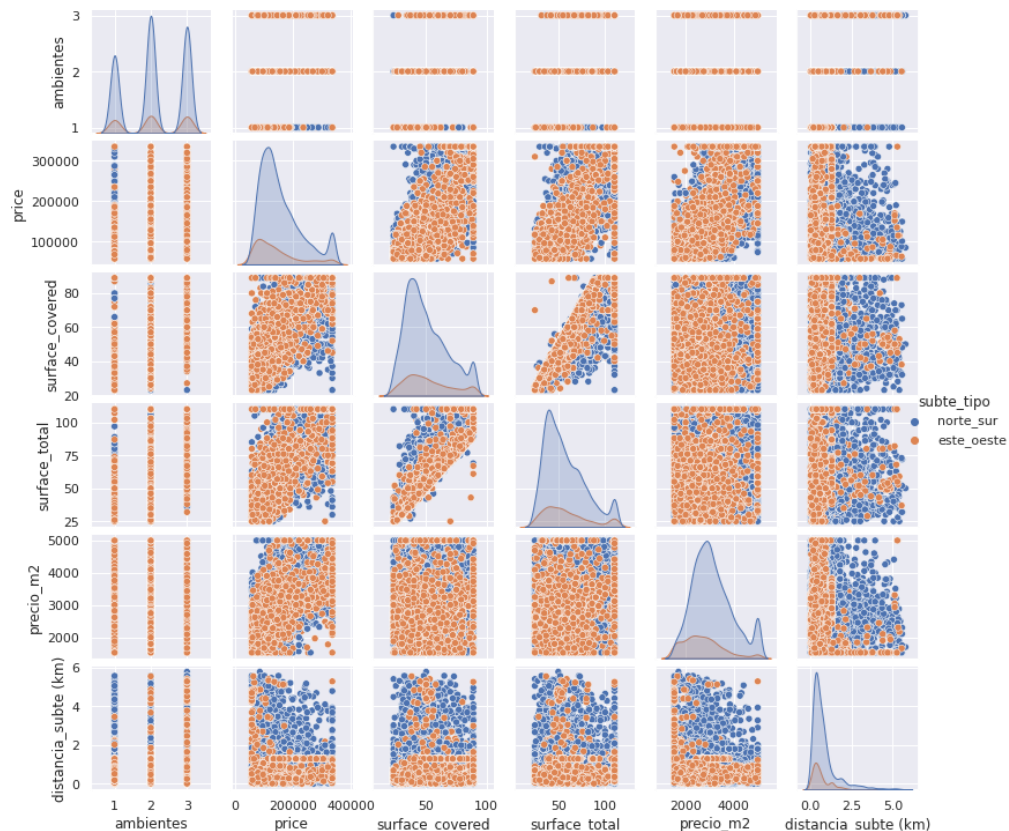
EANT

Python Data Analytics

Pairplot // Agrupación por línea de subte.

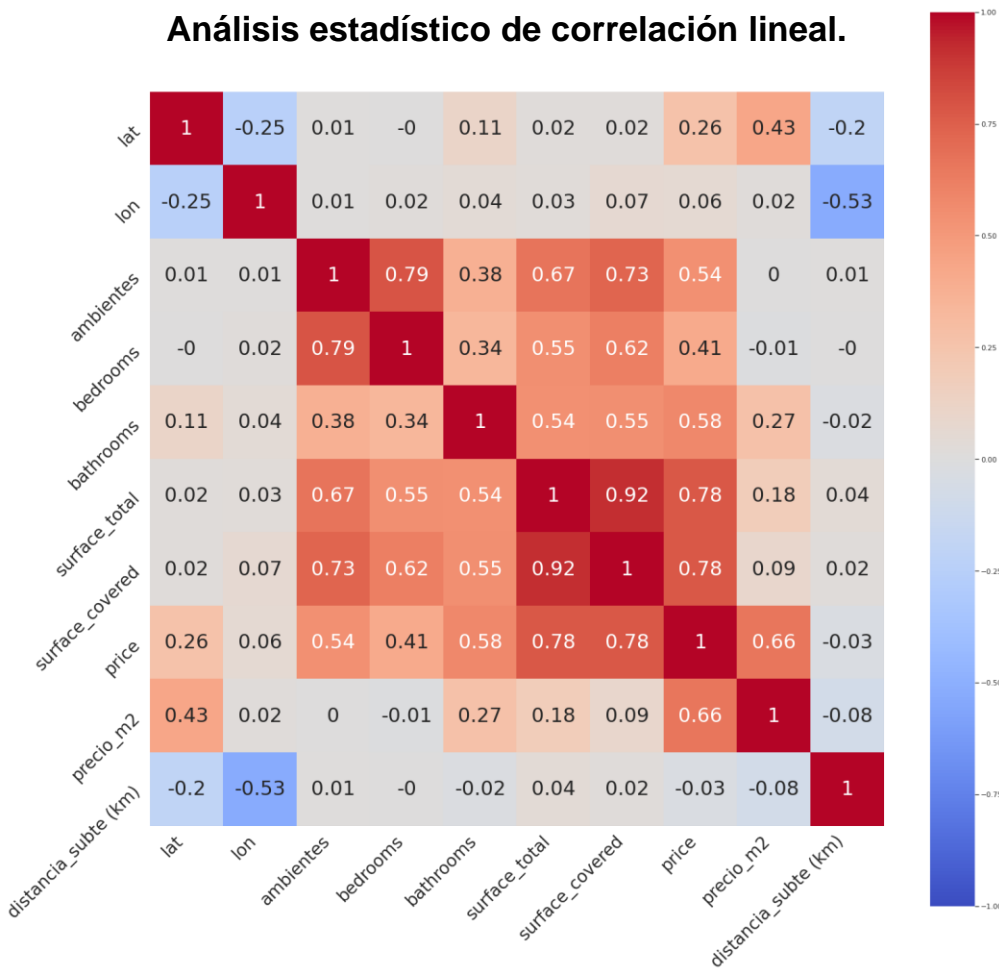


Agrupación por N-S y E-O



3. Visualización y análisis de los datos

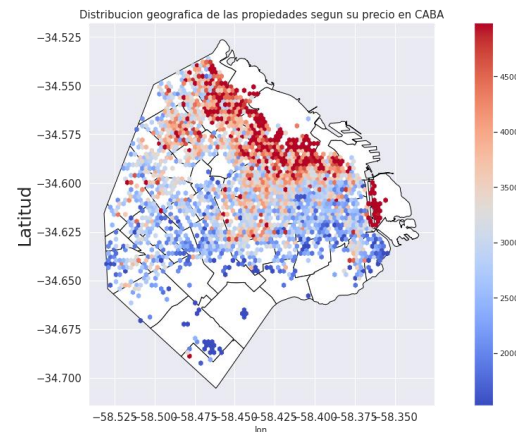
Análisis estadístico de correlación lineal.



Si analizamos el precio por m2, las variables con mayor correlación lineal son:

- Latitud de la ubicación de la propiedad ($r=0.43$)
- Cantidad de baños de la propiedad ($r=0.27$)
- Superficie total (0.18)
- Distancia al subte ($r=-0.08$)

Al observar correlación del precio m2 con latitud y longitud de la base se observa mayor correlación entre latitud y precio m2 que long y precio m2, es decir, la ubicación Norte/Sur de la propiedad en la ciudad tiene un mayor peso en el precio del metro cuadrado que la ubicación Este/Oeste relación con el precio del m2



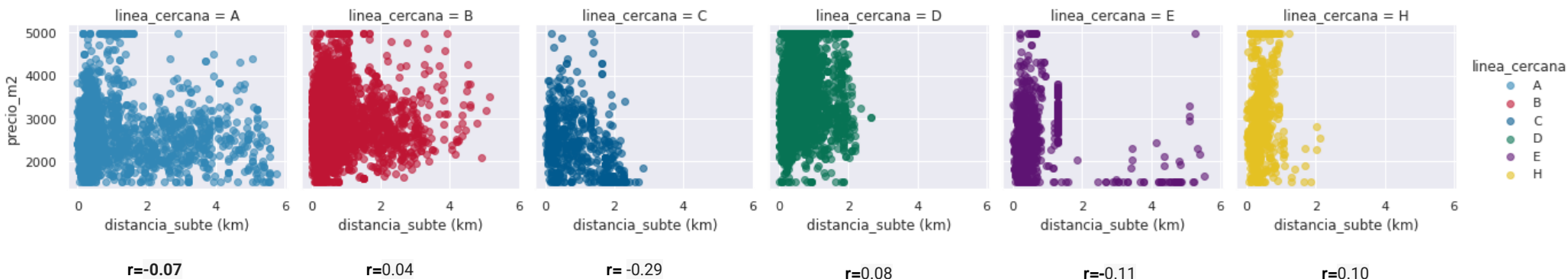
	n	r	CI95%	r2	adj_r2	p-val	BF10	power
pearson	11829	-0.083364	[-0.1, -0.07]	0.00695	0.006782	1.069958e-19	9.303e+15	1.0

3. Visualización y análisis de los datos ///

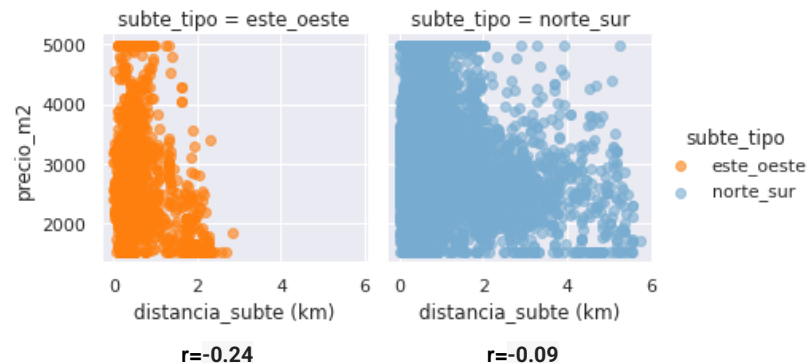
EANT

Python Data Analytics

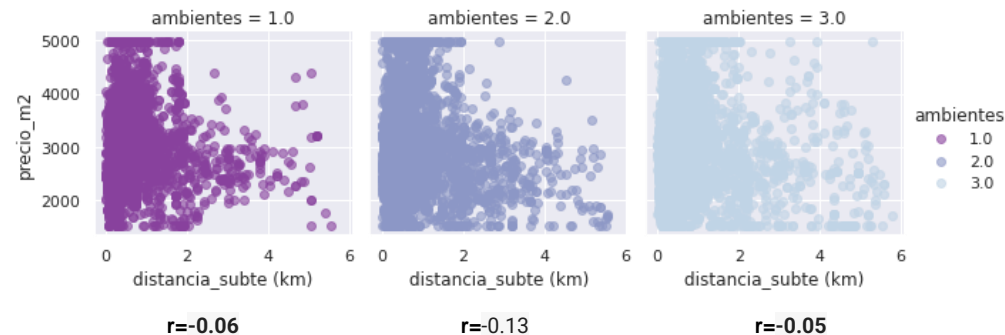
La relación entre las variables precio m2 y cercanía al subte varía de acuerdo a la línea que queda más cercana a la propiedad:



Agrupación de subtes entre norte-sur(A, B y D) y este-oeste (C, E y H)



Se observan también diferentes niveles de correlación entre las variables distancia al subte y precio m2 de acuerdo a la cantidad de ambientes de la propiedad:



- Si bien observamos que **existe relación entre el precio del m2** y la distancia a la boca de subte más cercana, **el coeficiente de correlación es bajo**, lo que indicaría que existen otro conjunto de atributos que influyen en el precio final de una propiedad.
- El coeficiente de correlación entre el precio del m2 y la distancia al subte varía de acuerdo a la línea analizada y al agrupamiento de los datos.
- **“Distancia Óptima”**: Visualizamos que las distancias menores a 200m del subte influyen negativamente en el precio del m2 y la localización ideal es entre 0,4 km y 1 km.
- **Mayor correlación entre latitud y precio m2 que longitud y precio m2**, es decir, la ubicación Norte/Sur de la propiedad en la ciudad tiene un mayor peso en el precio del metro cuadrado que la ubicación Este/Oeste. Esto se evidencia en la línea H, la cual tiene un recorrido transversal y presenta el mayor desvío en el precio por m2. Así como en la línea D, con los precios totales promedio más elevados.

Scope de proyecto ///

Limitaciones //

1. Cabe destacar que los procesos que acontecen en las ciudades no suelen ser unicausales, sino que se producen en general por muchas causas y variables debido a la propia complejidad. Entendemos que **el precio de venta puede estar influido también por muchas otras variables** que sería interesante abordar en posibles futuros estudios:

- Otros medios de transporte (Buses, Ecobici, Tren).
- Espacios verdes.
- Percepciones de seguridad.
- Turismo (Airbnb).
- Áreas comerciales, gastronomía.
- Cercanía a Hospitales / Nodos universitarios-educativos.
- Características propias del inmueble.
- Otras características.

2. El estudio fue realizado en base a datos del primer trimestre del año 2020. **Sería interesante analizar la variación del precio de las propiedades en el tiempo.**

3. El estudio fue realizado analizando únicamente la correlación lineal entre las variables. Este análisis se podría extender a otro tipo de estudios de **relaciones no lineales**

Reflexiones sobre futuros estudios //

- **Diferenciar causa y efecto de la cercanía de una propiedad y su precio.** Generar escenarios contrafácticos: ¿Cuánto cambiaría el precio de una propiedad en Palermo sin la línea D? ¿Cuanto una línea de subte en las propiedades de La Boca?
- **Escenario post- covid:** Teletrabajo y Virtualidad: ¿En qué medida continuará la necesidad de trasladarse de forma rápida y cómoda al microcentro o a zonas aledañas?

EANT

Python Data Analytics

Incorporación de más variables.

