

3) Public transport

August 8, 2023

1 1) Preparación previa

1.0.1 Carga de librerías

```
[1]: import requests
from pandas import json_normalize
import mpu
import pandas as pd
import numpy as np
import folium
```

1.0.2 Estaciones a usar luego

```
[2]: # Dirección web para acceder a los datos del gobierno de la ciudad:
url = 'https://cdn.buenosaires.gob.ar/datosabiertos/datasets/subte-estaciones/
↳subte_estaciones.geojson'
response = requests.get(url)
dictr = response.json()

# Hacemos referencia a las estaciones dentro del conjunto de datos:
estaciones = dictr['features']
df = json_normalize(estaciones)

# Creamos la columna Longitud:
df['Longitud'] = df['geometry.coordinates'].apply(lambda x:x[0])

# Creamos la columna Latitud:
df['Latitud'] = df['geometry.coordinates'].apply(lambda x:x[1])

# Creamos la máscara para filtrar los de Linea A:
mascara_lineas_elegidas = df['properties.LINEA'].isin(['A', 'B', 'C', 'E', 'H'])
lineas_elegidas = df.loc[mascara_lineas_elegidas, :]
lineas_elegidas = lineas_elegidas.drop(['type', 'properties.ID', 'properties.
↳LINEA', 'geometry.type', 'geometry.coordinates'], axis = 1)
lineas_elegidas
```

```
[2]:
properties.ESTACION  Longitud  Latitud
0                    CASEROS -58.398928 -34.635750
1      INCLAN - MEZQUITA AL AHMAD -58.400970 -34.629376
2                    HUMBERTO 1° -58.402323 -34.623092
3                    VENEZUELA -58.404732 -34.615242
4      ONCE - 30 DE DICIEMBRE -58.406036 -34.608935
..
85      SANTA FE - CARLOS JAUREGUI -58.402376 -34.594525
86  FACULTAD DE DERECHO - JULIETA LANTERI -58.391019 -34.583036
87                    RETIRO -58.375850 -34.592114
88                    CATALINAS -58.371700 -34.596597
89      CORREO CENTRAL -58.370413 -34.603014
```

[74 rows x 3 columns]

1.0.3 Funciones a usar luego

```
[3]: # La siguiente función nos permite definir la distancia en latitud y longitud
      ↪ respecto de las líneas de subte:
def distancia(lat, long):
    aux = []
    for index, row in lineas_elegidas.iterrows():
        aux.append(
            mpu.haversine_distance(
                (row['Latitud'], row['Longitud']),
                (lat, long)
            )
        )
    return min(aux)

# La siguiente función nos permite limpiar la superficie según las
      ↪ inconsistencias entre la total y la cubierta.
# Se eligió dicha columna con respecto a 'surface_total_in_m2' ya que tenía
      ↪ menor cantidad de registros nulos
def limpieza_superficie(sup_total, sup_cubierta):
    if sup_total is not None and sup_cubierta is not None:
        if sup_total >= sup_cubierta:
            superficie = sup_total
        else:
            superficie = np.NaN
    elif sup_total is not None:
        superficie = sup_total
    elif sup_cubierta is not None:
        superficie = sup_cubierta
    else: superficie = np.NaN

    return superficie
```

1.0.4 Lectura del dataset original de Properati

```
[4]: df_properati = pd.read_csv("https://media.githubusercontent.com/media/
↳Agustin-Bulzomi/Projects/main/Programming/Digital%20House%20(Python)/
↳Support%20Files/Project%201/Properati.csv")
```

2 2) Limpieza del dataset

2.0.1 Generación de DF solo para Capital

```
[5]: df_capital = df_properati[df_properati["state_name"] == "Capital Federal"]
df_capital.shape
```

```
[5]: (32316, 26)
```

2.0.2 Corrección de ortografía de barrios y mapeo de barrios inexistentes

```
[6]: %%capture output
df_capital["place_name"].replace(['San Cristobal', 'Barrio Norte', 'Centro /_
↳Microcentro'], ['San Cristóbal', 'Recoleta', 'Montserrat'], inplace=True)
```

2.0.3 Lectura el archivo de Mapping de Barrios por Comuna

```
[7]: df_barrios_comuna = pd.read_csv("https://raw.githubusercontent.com/
↳Agustin-Bulzomi/Projects/main/Programming/Digital%20House%20(Python)/
↳Support%20Files/Project%201/Barrios_por_comuna.csv")
```

2.0.4 Merge del DF de Capital con los barrios según comunas

```
[8]: df_capital_comuna = pd.merge(df_capital, df_barrios_comuna, how='left',_
↳left_on='place_name', right_on='Barrio')
del df_capital_comuna["description"]
del df_capital_comuna["title"]
del df_capital_comuna["properati_url"]
del df_capital_comuna["image_thumbnail"]
```

```
[9]: df_capital_comuna["Comuna"].isna().sum()
```

```
[9]: 3612
```

2.0.5 Enumeración de los barrios por los que atraviesa la Línea A, elegida como la principal para analizar.

```
[10]: barrios_linea_a = pd.Series(['Montserrat', 'Balvanera', 'Almagro', 'Caballito'])
barrios_linea_a
```

```
[10]: 0    Monserrat
      1    Balvanera
      2     Almagro
      3    Caballito
      dtype: object
```

2.0.6 Filtro del DF de Capital según los barrios definidos

```
[11]: df_capital_barrios_subte = df_capital_comuna[df_capital_comuna['Barrio'].
      ↪isin(barrios_linea_a)]
      df_capital_barrios_subte.shape
```

```
[11]: (4866, 24)
```

2.0.7 Eliminación de los registros sin coordenadas

```
[12]: df_latlon_notna = df_capital_barrios_subte[df_capital_barrios_subte['lat-lon'].
      ↪notna()]
```

```
[13]: df_latlon_notna.shape
```

```
[13]: (4031, 24)
```

2.0.8 Eliminación de los registros sin precios

```
[14]: df_price_notna = df_latlon_notna[df_latlon_notna['price_usd_per_m2'].notna()]
      df_price_notna
```

```
[14]:      Unnamed: 0  operation  property_type  place_name  \
14              48      sell      apartment  Balvanera
16              52      sell      apartment  Caballito
17              53      sell      apartment  Caballito
18              54      sell      apartment  Caballito
19              55      sell      apartment  Caballito
...           ...      ...      ...      ...
31793          119970      sell      apartment  Balvanera
31817          120075      sell      apartment  Caballito
32198          120984      sell      apartment  Caballito
32267          121059      sell      apartment  Monserrat
32273          121075      sell      apartment  Caballito

      place_with_parent_names  country_name  state_name  \
14  |Argentina|Capital Federal|Balvanera|  Argentina  Capital Federal
16  |Argentina|Capital Federal|Caballito|  Argentina  Capital Federal
17  |Argentina|Capital Federal|Caballito|  Argentina  Capital Federal
18  |Argentina|Capital Federal|Caballito|  Argentina  Capital Federal
19  |Argentina|Capital Federal|Caballito|  Argentina  Capital Federal
```

...
31793	Argentina Capital Federal Balvanera	Argentina	Capital Federal
31817	Argentina Capital Federal Caballito	Argentina	Capital Federal
32198	Argentina Capital Federal Caballito	Argentina	Capital Federal
32267	Argentina Capital Federal Monserrat	Argentina	Capital Federal
32273	Argentina Capital Federal Caballito	Argentina	Capital Federal

	geonames_id	lat-lon	lat	...	\
14	6693228.0	-34.6096748,-58.3986588	-34.609675	...	
16	3435874.0	-34.6222111,-58.4391283	-34.622211	...	
17	3435874.0	-34.6238943143,-58.4468606529	-34.623894	...	
18	3435874.0	-34.6209278,-58.4458736	-34.620928	...	
19	3435874.0	-34.6209278,-58.4458736	-34.620928	...	

...
31793	6693228.0	-34.6096441,-58.3964932	-34.609644	...
31817	3435874.0	-34.6221978289,-58.4291563359	-34.622198	...
32198	3435874.0	-34.6204033093,-58.4410224908	-34.620403	...
32267	3430570.0	-34.613408336,-58.3845426597	-34.613408	...
32273	3435874.0	-34.627466548,-58.4386614795	-34.627467	...

	price_aprox_usd	surface_total_in_m2	surface_covered_in_m2	\
14	240000.0	141.0	124.0	
16	110000.0	50.0	50.0	
17	170500.0	100.0	46.0	
18	220410.0	81.0	75.0	
19	222780.0	81.0	75.0	

...
31793	118400.0	45.0	38.0
31817	160000.0	78.0	70.0
32198	147000.0	77.0	77.0
32267	85500.0	37.0	34.0
32273	315000.0	91.0	83.0

	price_usd_per_m2	price_per_m2	floor	rooms	expenses	Barrio	\
14	1702.127660	1935.483871	NaN	5.0	NaN	Balvanera	
16	2200.000000	2200.000000	NaN	3.0	NaN	Caballito	
17	1705.000000	3706.521739	NaN	2.0	NaN	Caballito	
18	2721.111111	2938.800000	NaN	3.0	NaN	Caballito	
19	2750.370370	2970.400000	NaN	3.0	NaN	Caballito	

...
31793	2631.111111	3115.789474	NaN	1.0	NaN	Balvanera
31817	2051.282051	2285.714286	NaN	NaN	2300.0	Caballito
32198	1909.090909	1909.090909	NaN	NaN	3800.0	Caballito
32267	2310.810811	2514.705882	NaN	NaN	1200.0	Monserrat
32273	3461.538462	3795.180723	NaN	NaN	3000.0	Caballito

Comuna

```

14      Comuna-3
16      Comuna-6
17      Comuna-6
18      Comuna-6
19      Comuna-6
...
31793  Comuna-3
31817  Comuna-6
32198  Comuna-6
32267  Comuna-1
32273  Comuna-6

```

[3024 rows x 24 columns]

2.0.9 Limpieza de la superficie

```
[15]: # Se debe corregir inconsistencias entre la superficie total y la cubierta
```

```
[16]: %%capture output
df_price_notna["superficie"] =
    ↪df_price_notna[['surface_total_in_m2', 'surface_covered_in_m2']].apply(lambda
    ↪df_price_notna:
    ↪limpieza_superficie(df_price_notna['surface_total_in_m2'], df_price_notna['surface_covered_i
df_price_notna.head(4)
```

3 3) Análisis de hipótesis

3.0.1 Cálculo de las distancias entre cada propiedad y las 17 estaciones

```
[17]: %%capture output
df_price_notna['distancia'] = df_price_notna.apply(lambda x:
    ↪distancia(x['lat'], x['lon']), axis=1)
df_price_notna
```

3.0.2 Muestreo aleatorio para obtener una visualización más prolija

```
[18]: df_randomizado = df_price_notna.sample(n=100, random_state = 1)
```

3.0.3 Definición de cuantiles

```
[19]: df_randomizado['quantile'] = pd.qcut(df_randomizado['price_usd_per_m2'], 4,
    ↪labels=False)
df_randomizado
```

```
[19]: Unnamed: 0 operation property_type place_name \
22423      81835      sell      apartment  Caballito
17309      64627      sell      apartment  Caballito
```

13749	52061	sell	store	Montserrat
10967	42365	sell	apartment	Caballito
1231	6751	sell	house	Almagro
...
20847	76673	sell	apartment	Caballito
29288	111790	sell	apartment	Montserrat
695	4643	sell	apartment	Caballito
29947	113701	sell	apartment	Almagro
18	54	sell	apartment	Caballito

	place_with_parent_names	country_name	\
22423	Argentina Capital Federal Caballito	Argentina	
17309	Argentina Capital Federal Caballito	Argentina	
13749	Argentina Capital Federal Centro / Microcentro	Argentina	
10967	Argentina Capital Federal Caballito	Argentina	
1231	Argentina Capital Federal Almagro	Argentina	
...	
20847	Argentina Capital Federal Caballito	Argentina	
29288	Argentina Capital Federal Montserrat	Argentina	
695	Argentina Capital Federal Caballito	Argentina	
29947	Argentina Capital Federal Almagro	Argentina	
18	Argentina Capital Federal Caballito	Argentina	

	state_name	geonames_id	lat-lon	lat	\
22423	Capital Federal	3435874.0	-34.6316825892,-58.3574144339	-34.631683	
17309	Capital Federal	3435874.0	-34.6040302276,-58.4323198061	-34.604030	
13749	Capital Federal	3435548.0	-34.607924,-58.374815	-34.607924	
10967	Capital Federal	3435874.0	-34.6117495,-58.441553	-34.611750	
1231	Capital Federal	3436397.0	-34.603771,-58.381587	-34.603771	
...	
20847	Capital Federal	3435874.0	-34.6277891,-58.4523659	-34.627789	
29288	Capital Federal	3430570.0	-34.6197919041,-58.3813379478	-34.619792	
695	Capital Federal	3435874.0	-34.6209278,-58.4458736	-34.620928	
29947	Capital Federal	3436397.0	-34.599744,-58.419144	-34.599744	
18	Capital Federal	3435874.0	-34.6209278,-58.4458736	-34.620928	

	...	price_usd_per_m2	price_per_m2	floor	rooms	expenses	Barrio	\
22423	...	3065.306452	3878.551020	NaN	NaN	NaN	Caballito	
17309	...	2205.882353	2272.727273	NaN	NaN	2150.0	Caballito	
13749	...	3033.707865	3033.707865	NaN	NaN	NaN	Montserrat	
10967	...	3560.000000	3955.555556	NaN	2.0	NaN	Caballito	
1231	...	1078.125000	1642.857143	NaN	NaN	NaN	Almagro	
...	
20847	...	3125.000000	3289.473684	6.0	3.0	1375.0	Caballito	
29288	...	1955.555556	2155.102041	NaN	NaN	1000.0	Montserrat	
695	...	1687.500000	1862.068966	NaN	5.0	NaN	Caballito	
29947	...	978.260870	3214.285714	NaN	2.0	NaN	Almagro	

18	...	2721.111111	2938.800000	NaN	3.0	NaN	Caballito
----	-----	-------------	-------------	-----	-----	-----	-----------

	Comuna	superficie	distancia	quantile
22423	Comuna-6	62.0	2.243691	3
17309	Comuna-6	34.0	0.228711	1
13749	Comuna-1	89.0	0.086576	3
10967	Comuna-6	50.0	0.864416	3
1231	Comuna-5	640.0	0.081200	0
...
20847	Comuna-6	80.0	0.416598	3
29288	Comuna-1	54.0	0.206999	1
695	Comuna-6	160.0	0.384825	0
29947	Comuna-5	138.0	0.415206	0
18	Comuna-6	81.0	0.384825	3

[100 rows x 27 columns]

3.0.4 Visualización en un mapa

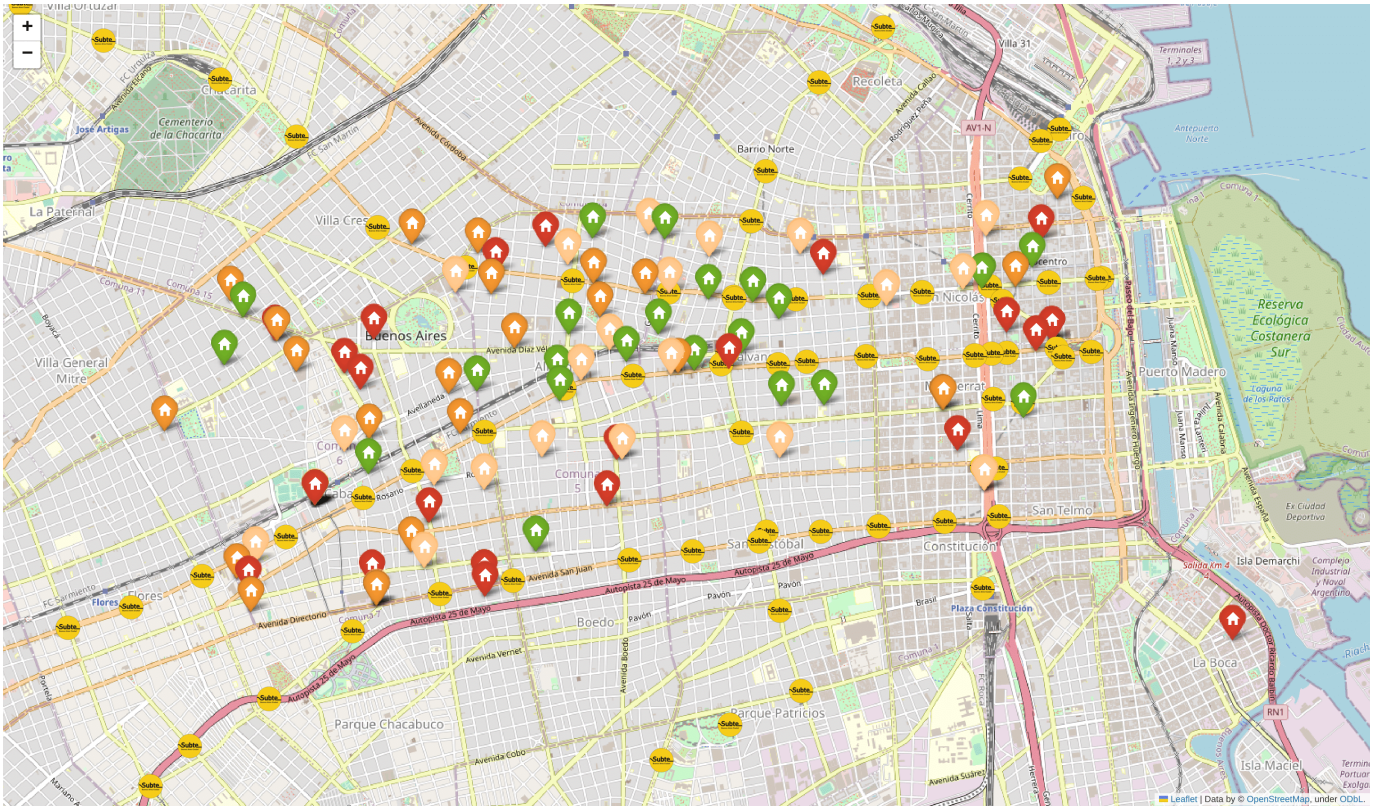
```
[20]: url = 'https://upload.wikimedia.org/wikipedia/commons/thumb/b/bf/Subte-logo.svg/
↳1200px-Subte-logo.svg.png'

map = folium.Map(location=[-34.613, -58.41], zoom_start=14)

for index, row in lineas_elegidas.iterrows():
    marker_subte = folium.Marker([row['Latitud'], row['Longitud']], icon=folium.
↳features.CustomIcon(url, icon_size=(28, 28))).add_to(map)

# Se agregan los íconos para las casas
for index, row in df_randomizado.iterrows():
    if row['quantile'] == 0:
        folium.Marker([row['lat'], row['lon']], icon=folium.Icon(icon='home',
↳color='green')).add_to(map)
    elif row['quantile'] == 1:
        folium.Marker([row['lat'], row['lon']], icon=folium.Icon(icon='home',
↳color='beige')).add_to(map)
    elif row['quantile'] == 2:
        folium.Marker([row['lat'], row['lon']], icon=folium.Icon(icon='home',
↳color='orange')).add_to(map)
    else:
        folium.Marker([row['lat'], row['lon']], icon=folium.Icon(icon='home',
↳color='red')).add_to(map)

display(map)
```

3.0.5 Exportación del .CSV

```
[21]: df_randomizado.to_csv("df_randomizado_CABA.csv", index=False)
```