TAREA DE BBDD NOSQL CON MONGODB Y PYMONGO.

Marilin Patricia Lanche Peña

Análisis de datos de los distritos, los barrios y la población de Barcelona

Fuente de datos: https://opendata-ajuntament.barcelona.cat/data/es/dataset/est-densitat

Servidor: MongoDB

Cliente: Python con el conector pymongo, y ejecutado en Jupyter.

Se adjunta en la misma carpeta el notebook de Jupyter y los CSVs.

Introducción:

Para el presente trabajo he cogido los archivos CSV de la página del ayuntamiento de Barcelona, que contiene datos sobre los indicadores demográficos.

Los archivos CSV están publicados por años, así que he creado mi dataset uniendo 5 de estos archivos, contando así con la información de 2017 - 2021.

He optado por hacer las distintas queries con Python en un notebook de Jupyter para poder ejecutar cada query por separado, y he usado la librería pandas para poder tener una visualización más clara de los resultados de las queries.

En Jupyter he ido planteando las queries de menor a mayor dificultad. Y en este documento las traeré en el orden según sea necesario para las conclusiones.

Los campos de mi dataset me aportan información sobre el año, distrito, barrio, población, superficie general en hectáreas, superficie residencial en hectáreas, la densidad general y la densidad neta de la ciudad de Barcelona. Los nombres de los campos están en catalán.

He optado en gran parte por utilizar pipelines de agregaciones que en su mayoría arrojan datos sobre la población y superficie por barrios y distritos.

La estructura de mi dataset con pandas:

```
cursor = barcelona.find({}, {"_id":0}).limit(10)

pd.DataFrame.from_dict(cursor)
```

	Any	Codi_Districte	Nom_Districte	Codi_Barri	Nom_Barri	Població	Superfície (ha)	Superfície Residencial (ha)	Densitat (hab/ha)	Densitat neta (hab/ha)
0	2021	1	Ciutat Vella	1	el Raval	47228	110.0	49.6	429	952
1	2021	1	Ciutat Vella	2	el Barri Gòtic	21827	81.6	34.1	268	639
2	2021	1	Ciutat Vella	3	la Barceloneta	14643	117.9	13.5	124	1081

La estructura de mi datas en formato Json:

```
db.barcelona.find({})
    .projection({_id:0})
    .sort({_id:-1})
    .limit(2)
```

arcelona 0.019 s 2 Docs

```
/* 1 */
{
    "Any" : 2017,
    "Codi_Districte" : 10,
    "Nom_Districte" : "Sant Martí",
    "Codi_Barri" : 73,
    "Nom_Barri" : "la Verneda i la Pau",
    "Població" : 28691,
    "Superfície (ha)" : 112.3,
    "Superfície Residencial (ha)" : 40.1,
    "Densitat (hab/ha)" : 256,
    "Densitat neta (hab/ha)" : 716
```

Conexión a mongodb desde Python en Jupyter:

Instalé pymongo e importé las librerías y módulos que iba a necesitar, me conecté a mongodb utilizando MongoClient. Finalmente creé la base de datos y la colección con la que trabajaré.

```
: !pip install pymongo
  import pandas as pd
  from pymongo import UpdateMany
  from pymongo import MongoClient

  client = MongoClient('mongodb://localhost:27017/')

  Requirement already satisfied: pymongo in c:\users\maril\anaconda3\lib\site-packages (4.3.3)
  Requirement already satisfied: dnspython<3.0.0,>=1.16.0 in c:\users\maril\anaconda3\lib\site-packages (from pymongo) (2.2.1)

: barcelona = client['nosql']['barcelona']
```

Importar los CSV e insertarlos en la colección:

Creo una función para cargar los CSV y convertirlos a diccionario.

Utilizo el método insert_many() y le paso como argumento la función con el archivo CSV, así cargo los datos en la colección “barcelona”.

```
def csv_to_json(filename, header=None):
    data = pd.read_csv(filename)
    return data.to_dict('records')

barcelona.insert_many(csv_to_json('./data/2021_densitat.csv'))
barcelona.insert_many(csv_to_json('./data/2020_densitat.csv'))
barcelona.insert_many(csv_to_json('./data/2019_densitat.csv'))
barcelona.insert_many(csv_to_json('./data/2018_densitat.csv'))
barcelona.insert_many(csv_to_json('./data/2017_densitat.csv'))

<pymongo.results.InsertManyResult at 0x19577143b20>
```

Compruebo que se han cargado los archivos y veo la estructura de la colección:

```
cursor = barcelona.find({}, {"_id":0}).limit(10)

pd.DataFrame.from_dict(cursor)
```

	Any	Codi_Districte	Nom_Districte	Codi_Barri	Nom_Barri	Població	Superficie (ha)	Superficie Residencial (ha)	Densitat (hab/ha)	Densitat neta (hab/ha)
0	2021	1	Ciutat Vella	1	el Raval	47228	110.0	49.6	429	952
1	2021	1	Ciutat Vella	2	el Barri Gòtic	21827	81.6	34.1	268	639
2	2021	1	Ciutat Vella	3	la Barceloneta	14643	117.9	13.5	124	1081
3	2021	1	Ciutat Vella	4	Sant Pere, Santa Caterina i la Ribera	22131	111.0	32.4	199	683
4	2021	2	Eixample	5	el Fort Pienc	31262	92.9	33.6	336	930
5	2021	2	Eixample	6	la Sagrada Família	51319	104.2	51.0	493	1007
6	2021	2	Eixample	7	la Dreta de l'Eixample	44675	212.0	113.9	211	392
7	2021	2	Eixample	8	l'Antiga Esquerra de l'Eixample	41990	122.8	66.9	342	627
8	2021	2	Eixample	9	la Nova Esquerra de l'Eixample	57054	134.1	65.2	426	876
9	2021	2	Eixample	10	Sant Antoni	38053	80.4	41.3	473	921

Estructura de los barrios de Barcelona en 2021:

Utilizando el método distinct() y el len() sabemos que en el año 2021 Barcelona cuenta con 73 barrios, cuyos nombres podemos ver en la tabla, aunque por defecto se muestran los 5 primeros y los 5 últimos del total.

El Número de Barrios de Barcelona que existen en 2021 y el nombre de cada barrio.

In [122]:

```
print(f'El número de barrios de Barcelona es: {len(barcelona.distinct("Nom_Barri", { "Any": 2021 }))}')

cursor = barcelona.distinct("Nom_Barri", { "Any": 2021 })

pd.DataFrame.from_dict(cursor)
```

```
El número de barrios de Barcelona es: 73
```

Out[122]:

	0
0	Baró de Viver
1	Can Baró
2	Can Peguera
3	Canyelles
4	Ciutat Meridiana
...	...
68	la Vila Olímpica del Poblenou
69	la Vila de Gràcia
70	les Corts
71	les Roquetes
72	les Tres Torres

73 rows × 1 columns

Estructura de los Distritos de Barcelona en 2021:

Si filtramos por año y agrupamos por distritos vemos que Barcelona cuenta con 10 distritos, en la tabla podemos ver el nombre de los distritos y también a simple vista sabemos que el distrito con más barrios es *Nou Barris* y el que menos barrios tiene es el distrito *Les Corts*.

El número de distritos en Barcelona y cuantos barrios tiene cada distrito en 2021. Ordenados según el número de barrios de mayor a menor.

In [123]:

```
print(f'El número de distritos  de Barcelona es: {len(barcelona.distinct("Nom_Districte", { "Any": 2021 }))}')

cursor = barcelona.aggregate([
    {"$match": {"Any":2021}},
    {"$group": { "_id": "$Nom_Districte",
                 "Nº_Barrios":{"$sum": 1} }
    },
    {"$sort": {"Nº_Barrios": -1}}

    ])


pd.DataFrame.from_dict(cursor)
```

```
El número de distritos  de Barcelona es: 10
```

Out[123]:

	_id	Nº_Barrios
0	Nou Barris	13
1	Horta-Guinardó	11
2	Sant Martí	10
3	Sants-Montjuïc	8
4	Sant Andreu	7
5	Eixample	6
6	Sarrià-Sant Gervasi	6
7	Gràcia	5
8	Ciutat Vella	4
9	Les Corts	3

Densidad neta de Barcelona en el año 2021:

En este apartado obtenemos que la media de la densidad neta de Barcelona en 2021 es aproximadamente *689* habitantes por hectárea, para ello he filtrado por año y ya que para la media necesito un group al que le he dejado con el id null. Posteriormente con un find() filtramos aquellos documentos que tengan una densidad neta mayor o igual a la media. Podemos ver que *Provençals del Poblenou* es el barrio más alejado de la media y que el barrio más cercano a la media es el *Parc i la Llacuna del Poblenou*. Y ambos Barrios pertenecen al mismo distrito.

Cuál es la media de la densidad neta de Barcelona en 2021 y qué barrios están por encima de la media, y a qué distrito pertenecen. Ordenado de mayor a menor densidad.

In [126]:

```
mediaDensidad = list(barcelona.aggregate([
    {"$match":{"Any":2021}},
    {"$group": {"_id": "null",
                "Media_Densidad_Neta":{"$avg": "$Densitat neta (hab/ha)"}
               }
    }])) [0]["Media_Densidad_Neta"]

print(f" La media de la densidad es : {mediaDensidad}")

cursor = barcelona.find( {"Any":2021, "Densitat neta (hab/ha)": {"$gte": mediaDensidad}, ""},
                         {"_id":0,"Nom_Districte" :1, "Nom_Barri" : 1, "Densitat neta (hab/ha)" : 1}
                       ).sort("Densitat neta (hab/ha)", pymongo.DESCENDING)

pd.DataFrame.from_dict(cursor)
```

```
La media de la densidad es : 689.931506849315
```

Out[126]:

	Nom_Districte	Nom_Barri	Densitat neta (hab/ha)
0	Sant Martí	Provençals del Poblenou	1368
1	Nou Barris	el Turó de la Peira	1204
2	Sant Martí	el Clot	1131
3	Sant Martí	Sant Martí de Provençals	1111
4	Ciutat Vella	la Barceloneta	1081
5	Horta-Guinardó	el Baix Guinardó	1079
6	Sants-Montjuïc	el Poble Sec	1033
7	Eixample	la Sagrada Família	1007
8	Nou Barris	la Prosperitat	985
9	Sant Andreu	Navas	984
10	Sants-Montjuïc	Sants - Badal	980
11	Ciutat Vella	el Raval	952
12	Eixample	el Fort Pienc	930
13	Sants-Montjuïc	la Bordeta	928
14	Eixample	Sant Antoni	921
15	Sant Martí	el Besòs i el Maresme	913
16	Gràcia	el Camp d'en Grassot i Gràcia Nova	907
17	Nou Barris	les Roquetes	897
18	Nou Barris	Verdun	894
19	Sant Martí	el Camp de l'Arpa del Clot	894
20	Eixample	la Nova Esquerra de l'Eixample	876
21	Nou Barris	Vilapicina i la Torre Llobeta	875
22	Sants-Montjuïc	la Font de la Guatlla	873
23	Horta-Guinardó	el Carmel	822
24	Sant Andreu	la Trinitat Vella	817
25	Horta-Guinardó	la Vall d'Hebron	813
26	Sants-Montjuïc	la Marina de Port	789
27	Sant Andreu	la Sagrera	770
28	Nou Barris	Porta	768
29	Sant Andreu	el Bon Pastor	763
30	Sant Andreu	Sant Andreu	758
31	Sant Martí	el Poblenou	755
32	Sant Andreu	el Congrés i els Indians	753
33	Les Corts	la Maternitat i Sant Ramon	741
34	Nou Barris	Ciutat Meridiana	729
35	Sants-Montjuïc	Sants	726
36	Sants-Montjuïc	Hostafrancs	717
37	Sant Martí	la Verneda i la Pau	714
38	Horta-Guinardó	el Guinardó	710
39	Horta-Guinardó	la Teixonera	707
40	Les Corts	les Corts	702
41	Sant Martí	el Parc i la Llacuna del Poblenou	699

Mayor y menor densidad neta en 2021:

Si hacemos un filtrado por año, ordenamos (asc y desc) y limitamos la devolución a un documento, obtenemos:

- El Barrio con menor densidad neta en 2021 es *el Tibidabo i les Planes Vallvidrera* que pertenece al distrito *Sarrià-Sant Gervasi.*

- El Barrio con mayor densidad neta en 2021 es *Provençals del Poblenou* que pertenece al distrito *Sant Martí*

El barrio que tiene una densidad neta menor en el año 2021, y a que distrito pertenece. ¶

In [124]:

```
cursor = barcelona.aggregate([
    {"$match": {"Any":2021}},
    {"$sort": {"Densitat neta (hab/ha)":1}},
    {"$limit": 1},
    {"$project": {"_id":0,"Any":1, "Nom_Districte":1,"Nom_Barri":1,"Densitat neta (hab/ha)":1}}
])

pd.DataFrame.from_dict(cursor)
```

Out[124]:

	Any	Nom_Districte	Nom_Barri	Densitat neta (hab/ha)
0	2021	Sarrià-Sant Gervasi	Vallvidrera, el Tibidabo i les Planes	19

El barrio que tiene la mayor densidad neta en el año 2021, y a que distrito pertenece.

In [125]:

```
cursor = barcelona.aggregate([
    {"$match": {"Any":2021}},
    {"$sort": {"Densitat neta (hab/ha)":-1}},
    {"$limit": 1},
    {"$project": {"_id":0,"Any":1, "Nom_Districte":1,"Nom_Barri":1,"Densitat neta (hab/ha)":1}}
])

pd.DataFrame.from_dict(cursor)
```

Out[125]:

	Any	Nom_Districte	Nom_Barri	Densitat neta (hab/ha)
0	2021	Sant Martí	Provençals del Poblenou	1368

Porcentaje de habitantes por barrio en Barcelona en el año 2021:

He utilizado una agregación para obtener el total de habitantes de Barcelona en 2021, luego he casteado/convertido la agregación a una lista, para mediante posición acceder únicamente al valor numérico del total de habitantes, valor que se guarda en la variable “total_habitantes”.

Esta variable la utilizo para las dos agregaciones a continuación.

En este apartado calculo en el $project el porcentaje de población por barrio y ordeno el resultado de mayor a menor porcentaje. Y vemos que *Sant Andreu* es el barrio con mayor porcentaje de población, mientras que *la Clota* es el barrio con el porcentaje menor de población.

Porcentaje de habitantes por barrio en el año 2021, en orden descendiente.

In [128]:

```
total_habitantes = list(barcelona.aggregate([
    {"$match": {"Any":2021}},
    {"$group": { "_id": "$Any",
                 "Total_habitantes": {"$sum": "$Població"}
               }
    },
    {"$project": {"_id" : 0, "Total_habitantes":1 }}

    ]))[0]["Total_habitantes"]

cursor = barcelona.aggregate([

    {"$match": {"Any":2021}},

    {"$project": {
                 "_id":0,
                 "Nom_Districte": 1,
                 "Nom_Barri" : 1,
                 "Any": 1,
                 "Porcentaje_Población_Barrio": {"$trunc":
                                                 [{"$multiply": [{"$divide": ["$Població", total_habitantes ]}, 100]}, 2]
                                                 }
                 }
    },

    {"$sort": {"Porcentaje_Población_Barrio": -1}}

    ])

pd.DataFrame.from_dict(cursor)
```

Out[128]:

	Any	Nom_Districte	Nom_Barri	Porcentaje_Población_Barrio
0	2021	Sant Andreu	Sant Andreu	3.54
1	2021	Eixample	la Nova Esquerra de l'Eixample	3.48
2	2021	Eixample	la Sagrada Família	3.13
3	2021	Gràcia	la Vila de Gràcia	3.02
4	2021	Ciutat Vella	el Raval	2.88
...	...	...	...	...
68	2021	Sant Andreu	Baró de Viver	0.16
69	2021	Nou Barris	Can Peguera	0.13
70	2021	Sants-Montjuïc	la Marina del Prat Vermell	0.08
71	2021	Nou Barris	Vallbona	0.08
72	2021	Horta-Guinardó	la Clota	0.04

73 rows × 4 columns

Porcentaje de habitantes por distrito en Barcelona en el año 2021:

Este apartado es similar al anterior, pero me parece que aporta una información más llamativa.

Hago una agregación para filtrar por el año, agrupar por distrito y proyectar el distrito y su porcentaje de población; para ello utilizo la variable creada en el apartado anterior "total_habitantes". Finalmente ordeno descendientemente y vemos que los distritos con mayor y menor porcentaje de población son *Eixample* y *Les Corts*, respectivamente.

Porcentaje de población en cada distrito, en el año 2021, en orden descendiente.

In [129]:

```
cursor = barcelona.aggregate([
    {"$match": {"Any":2021}   },
    {
     "$group": { "_id": "$Nom_Districte",
                 "Total_población": {"$sum": "$Població"}
                }
    },
    {"$project": {"_id":0,
                  "Distrito": "$_id",
                  "Porcentaje_población": {"$trunc":
                                           [ {"$multiply": [{"$divide": ["$Total_población", total_habitantes]}, 100]}, 2 ]
                                          }
                  }
    },
    {"$sort": {"Porcentaje_población": -1}}
    ])


pd.DataFrame.from_dict(cursor)
```

Out[129]:

	Distrito	Porcentaje_población
0	Eixample	16.15
1	Sant Martí	14.55
2	Sants-Montjuïc	11.26
3	Horta-Guinardó	10.52
4	Nou Barris	10.50
5	Sant Andreu	9.18
6	Sarrià-Sant Gervasi	9.00
7	Gràcia	7.41
8	Ciutat Vella	6.46
9	Les Corts	4.92

Mayor y menor población en cada distrito, Barcelona 2021:

En este apartado filtro por año, ordeno de mayor a menor población y luego agrupo por distrito. En el $group y gracias a que tengo ordenado por el campo población, puedo obtener el $min y el $max de población. Ahora bien para saber el nombre del barrio al que corresponden los valores min y max, utilizo los operadores $last y $first.

El resultado es el nombre del distrito, el nombre del barrio con mayor población y su población correspondiente, y el nombre del barrio de menor población con su población correspondiente.

En cada distrito qué barrio tiene la mayor y la menor población en el año 2021.

In [130]:

```
cursor =  barcelona.aggregate([
    {"$match": {"Any": 2021}},
    {"$sort": {"Població": 1}},
    {"$group": { "_id": "$Nom_Districte",

                "Barrio_Mayor_población":{"$last": "$Nom_Barri"},

                "Mayor_Población":{"$max": "$Població"},

                "Barrio_Menor_población":{"$first": "$Nom_Barri"},

                "Menor_Población":{"$min": "$Població"}

                }
    }
    ])

pd.DataFrame.from_dict(cursor)
```

Out[130]:

	_id	Barrio_Mayor_población	Mayor_Población	Barrio_Menor_población	Menor_Población
0	Sarrià-Sant Gervasi	Sant Gervasi - Galvany	46816	Vallvidrera, el Tibidabo i les Planes	4697
1	Les Corts	les Corts	45422	Pedralbes	11581
2	Horta-Guinardó	el Guinardó	36939	la Clota	689
3	Gràcia	la Vila de Gràcia	49492	el Coll	7590
4	Sant Martí	el Camp de l'Arpa del Clot	38360	la Vila Olímpica del Poblenou	9240
5	Nou Barris	Porta	27418	Vallbona	1384
6	Ciutat Vella	el Raval	47228	la Barceloneta	14643
7	Eixample	la Nova Esquerra de l'Eixample	57054	el Fort Pienc	31262
8	Sants-Montjuïc	Sants	43082	la Marina del Prat Vermell	1350
9	Sant Andreu	Sant Andreu	57954	Baró de Viver	2645

Población de Barcelona a través de los años:

Aquí hago una agrupación por año y calculo el total de población y la media por barrio.

Así podemos ver las fluctuaciones en la población a través de los años, en este caso vemos que de 2017 a 2018 hay un decrecimiento en el total de habitantes, pero de 2018 a 2020 la población va aumentando, hasta que hay un notable descenso de 2020 a 2021, y se podría suponer que se debe a la pandemia por coronavirus.

El total de habitantes en Barcelona y la media de habitantes por barrio, a lo largo de los años (2017 -2021)

In [127]:

```
cursor = barcelona.aggregate([

    {"$group": { "_id": "$Any",
                 "Total_habitantes": {"$sum": "$Població"},
                 "Media_por_barrio": {"$avg": "$Població"}
                }
    },
    {"$sort": {"_id":1}}
])


pd.DataFrame.from_dict(cursor)
```

Out[127]:

	_id	Total_habitantes	Media_por_barrio
0	2017	1620809	22202.863014
1	2018	1620343	22196.479452
2	2019	1636762	22421.397260
3	2020	1664182	22797.013699
4	2021	1636732	22420.986301

Población por distritos de Barcelona a través de los años:

Aquí agrupo por dos campos (año y distrito) y obtengo el total y la media de la población, luego selecciono los campos que voy a mostrar en la tabla y modifico sus nombres a conveniencia, finalmente ordeno por los mismos campos por los que agrupé. Así puedo ver la evolución de la población a través de los años en cada distrito.

Es decir, es un análisis más granular que el apartado anterior. Pero se observa el mismo patrón en la mayoría de los distritos: de 2017 a 2018 un descenso de la población, de 2018 a 2020 la población va aumentando, hasta que comparamos 2020 y 2021 y se observa un notable descenso en la población, que se podría suponer que se debe a la pandemia por coronavirus.

Cambio en la media de la población y en el total de la población en cada distrito a través de los años (2017-2021)

In [131]:

```
cursor = barcelona.aggregate([

    {"$group": { "_id": {"Nom_Districte":"$Nom_Districte", "Any": "$Any"},

                "Media_Población": {"$avg": "$Població"},

                "Total_habitantes": {"$sum": "$Població"}

                 }
    },
    {"$project":{ "_id":0,
                  "Nom_Districte":"$_id.Nom_Districte",
                  "Any": "$_id.Any",
                  "Media_Población": 1,
                  "Total_habitantes":1
                 }
    },

    {"$sort": {"Nom_Districte":1, "Any":1}}

    ])

pd.DataFrame.from_dict(cursor)
```

Out[131]:

	Media_Población	Total_habitantes	Nom_Districte	Any
0	25346.750000	101387	Ciutat Vella	2017
1	25178.500000	100714	Ciutat Vella	2018
2	25857.250000	103429	Ciutat Vella	2019
3	26964.500000	107858	Ciutat Vella	2020
4	26457.250000	105829	Ciutat Vella	2021
5	44402.666667	266416	Eixample	2017
6	44145.666667	264874	Eixample	2018
7	44318.333333	265910	Eixample	2019
8	45055.166667	270331	Eixample	2020
9	44058.833333	264353	Eixample	2021
10	24269.400000	121347	Gràcia	2017
11	24201.000000	121005	Gràcia	2018
12	24359.600000	121798	Gràcia	2019
13	24722.800000	123614	Gràcia	2020
14	24287.600000	121438	Gràcia	2021
15	15341.000000	168751	Horta-Guinardó	2017
16	15410.909091	169520	Horta-Guinardó	2018
17	15590.454545	171495	Horta-Guinardó	2019
18	15858.818182	174447	Horta-Guinardó	2020
19	15654.454545	172199	Horta-Guinardó	2021
20	27344.333333	82033	Les Corts	2017
21	27251.666667	81755	Les Corts	2018
22	27324.666667	81974	Les Corts	2019
23	27510.666667	82532	Les Corts	2020
24	26859.000000	80577	Les Corts	2021
25	12813.769231	166579	Nou Barris	2017
26	12912.923077	167868	Nou Barris	2018
27	13128.384615	170669	Nou Barris	2019
28	13374.153846	173864	Nou Barris	2020
29	13221.769231	171883	Nou Barris	2021
30	21084.857143	147594	Sant Andreu	2017
31	21176.000000	148232	Sant Andreu	2018
32	21403.000000	149821	Sant Andreu	2019
33	21708.571429	151960	Sant Andreu	2020
34	21486.714286	150407	Sant Andreu	2021
35	23551.300000	235513	Sant Martí	2017
36	23596.000000	235960	Sant Martí	2018
37	23831.500000	238315	Sant Martí	2019
38	24126.300000	241263	Sant Martí	2020
39	23824.500000	238245	Sant Martí	2021
40	22738.750000	181910	Sants-Montjuïc	2017
41	22662.875000	181303	Sants-Montjuïc	2018
42	23011.375000	184091	Sants-Montjuïc	2019
43	23428.125000	187425	Sants-Montjuïc	2020
44	23045.875000	184367	Sants-Montjuïc	2021
45	24879.833333	149279	Sarrià-Sant Gervasi	2017
46	24852.000000	149112	Sarrià-Sant Gervasi	2018
47	24876.666667	149260	Sarrià-Sant Gervasi	2019
48	25148.000000	150888	Sarrià-Sant Gervasi	2020
49	24572.333333	147434	Sarrià-Sant Gervasi	2021

Superficie de Barcelona a través de los años:

Aquí utilizo un $project para seleccionar los campos que voy a necesitar, también creo un nuevo campo que será una resta de la superficie general y la residencial para obtener la superficie no residencial.

El resultado es una tabla con dos campos complementarios entre sí, es decir la superficie no residencial y la residencial desde 2017 hasta 2021.

Se observan los cambios en la superficie, así tenemos que de 2017 a 2018 hubo un descenso en la superficie no residencial, es decir que se habrá más suelo considerada residencial. De 2018 hasta 2021 hay cambios muy pequeños.

Cómo ha ido cambiando el total de superficie residencial y no residencial a través de los años (2017-2021)

In [133]:

```
cursor = barcelona.aggregate([
    {"$project": {"superficie_No_residencial":{"$subtract": ["$Superfície (ha)", "$Superfície Residencial (ha)"]},
                  "Superfície (ha)":1,
                  "Superfície Residencial (ha)":1,
                  "Any": 1
                 }
    },

    {"$group": { "_id": "$Any",
                  "Total_sup_NO_residencial": {"$sum": "$superficie_No_residencial"},
                  "Total_sup_Residencial": {"$sum": "$Superfície Residencial (ha)"}
              }
    },
    {"$sort": {"_id":1}}

    ])


pd.DataFrame.from_dict(cursor)
```

Out[133]:

	_id	Total_sup_NO_residencial	Total_sup_Residencial
0	2017	7625.9	2590.0
1	2018	7387.8	2747.4
2	2019	7389.6	2746.1
3	2020	7424.6	2745.8
4	2021	7424.5	2745.9

Porcentajes de Superficie por Distritos, a través de los años:

Utilizamos una agregación similar a la anterior, pero en este caso hacemos un $group por dos campos y con las operaciones necesarias para el posterior $project en donde calculamos los porcentajes, finalmente ordeno por el campo *distrito* y *año*, para poder observar el cambio de 2017 a 2021 de cada distrito. Este es un análisis un poco más granular, podemos ver que de 2017 a 2018 la superficie residencial aumento notablemente en el distrito *Sarrià-Sant Gervasi*. En el resto de los distritos no se aprecian cambios significativos.

En cada distrito cómo a ido cambiando el porcentaje de superficie residencial y no residencial a través de los años (2017-2021)

In [132]:

```
cursor = barcelona.aggregate([

    {"$project": {"superficie_No_residencial":{"$subtract": ["$Superfície (ha)", "$Superfície Residencial (ha)"]},
                  "Nom_Districte": 1,
                  "Nom_Barri" : 1,
                  "Superfície (ha)":1,
                  "Superfície Residencial (ha)":1,
                  "Any": 1
                 }
    },

    {"$group": { "_id": {"Nom_Districte":"$Nom_Districte", "Any": "$Any"},

                  "Superficie_total_Distrito": {"$sum": "$Superfície (ha)"},

                  "Superficie_Residencial_por_distrito": {"$sum": "$Superfície Residencial (ha)"},

                  "Superficie_no_residencial": {"$sum": "$superficie_No_residencial"}
                }
    },

    {"$project": {"_id":0,
                  "Distrito": "$_id.Nom_Districte",
                  "Año": "$_id.Any",
                  "%Sup_Residencial": {"$trunc": [ {"$multiply" :

                                        [{"$divide":
                                          ["$Superficie_Residencial_por_distrito","$Superficie_total_Distrito" ]}, 100]  }, 2]
                                       },
                  "%Sup_NO_Residencial": {"$trunc": [{"$multiply":

                                           [{"$divide":["$Superficie_no_residencial","$Superficie_total_Distrito" ]},100]

                                          }, 2]}

                  }
    },

    {"$sort": {"Distrito":1, "Año":1}}

    ])

pd.DataFrame.from_dict(cursor)
```

Out[132]:

	Distrito	Año	%Sup_Residencial	%Sup_NO_Residencial
0	Ciutat Vella	2017	29.71	70.28
1	Ciutat Vella	2018	31.52	68.47
2	Ciutat Vella	2019	31.44	68.55
3	Ciutat Vella	2020	30.82	69.17
4	Ciutat Vella	2021	30.82	69.17
5	Eixample	2017	49.77	50.22
6	Eixample	2018	49.82	50.17
7	Eixample	2019	49.85	50.14
8	Eixample	2020	49.82	50.17
9	Eixample	2021	49.82	50.17
10	Gràcia	2017	47.96	52.03
11	Gràcia	2018	48.10	51.89
12	Gràcia	2019	47.91	52.08
13	Gràcia	2020	47.91	52.08
14	Gràcia	2021	47.96	52.03
15	Horta-Guinardó	2017	24.63	75.36
16	Horta-Guinardó	2018	24.35	75.64
17	Horta-Guinardó	2019	24.32	75.67
18	Horta-Guinardó	2020	24.33	75.66
19	Horta-Guinardó	2021	24.32	75.67
20	Les Corts	2017	29.94	70.05
21	Les Corts	2018	29.62	70.37
22	Les Corts	2019	29.60	70.39
23	Les Corts	2020	29.60	70.39
24	Les Corts	2021	29.65	70.34
25	Nou Barris	2017	28.93	71.06
26	Nou Barris	2018	28.91	71.08
27	Nou Barris	2019	28.90	71.09
28	Nou Barris	2020	28.88	71.11
29	Nou Barris	2021	28.86	71.13
30	Sant Andreu	2017	29.27	70.72
31	Sant Andreu	2018	28.91	71.08
32	Sant Andreu	2019	28.91	71.08
33	Sant Andreu	2020	28.89	71.10
34	Sant Andreu	2021	28.89	71.10
35	Sant Martí	2017	26.57	73.42
36	Sant Martí	2018	27.81	72.18
37	Sant Martí	2019	27.78	72.21
38	Sant Martí	2020	27.65	72.34
39	Sant Martí	2021	27.65	72.34
40	Sants-Montjuïc	2017	10.86	89.13
41	Sants-Montjuïc	2018	11.08	88.91
42	Sants-Montjuïc	2019	11.09	88.90
43	Sants-Montjuïc	2020	10.99	89.00
44	Sants-Montjuïc	2021	10.98	89.01
45	Sarrià-Sant Gervasi	2017	22.84	77.15
46	Sarrià-Sant Gervasi	2018	30.70	69.29
47	Sarrià-Sant Gervasi	2019	30.67	69.32
48	Sarrià-Sant Gervasi	2020	30.65	69.34
49	Sarrià-Sant Gervasi	2021	30.65	69.34

Crear otra colección y cambiar los nombres de las columnas al español.

Este apartado es un ejemplo de un update, para ello creo una nueva colección en el cuál solo subo dos CSV, ya que el propósito es demostrar que se podía haber cambiado los nombres de las columnas al castellano para que sea comprensible para las otras comunidades. He utilizado el método update_many() y el operador $rename.

```
In [134]: barcelona_castellano = client['nosql']['barcelona_castellano']

In [135]: def csv_to_json(filename, header=None):
              data = pd.read_csv(filename)
              return data.to_dict('records')

          barcelona_castellano.insert_many(csv_to_json('./data/2021_densitat.csv'))
          barcelona_castellano.insert_many(csv_to_json('./data/2020_densitat.csv'))

Out[135]: <pymongo.results.InsertManyResult at 0x195771173a0>

In [137]: cursor = barcelona_castellano.find({},{"_id":0}).limit(1)

          pd.DataFrame.from_dict(cursor)
```

Out[137]:

	Any	Codi_Districte	Nom_Districte	Codi_Barri	Nom_Barri	Població	Superfície (ha)	Superfície Residencial (ha)	Densitat (hab/ha)	Densitat neta (hab/ha)
0	2021	1	Ciutat Vella	1	el Raval	47228	110.0	49.6	429	952

```
In [138]: cursor = barcelona_castellano.update_many({},
                                        { "$rename": {   "Any": "Anio",
                                                         "Codi_Districte": "Cod_Distrito",
                                                         "Nom_Districte": "Distrito",
                                                          "Codi_Barri" : "Cod_Barrio",
                                                          "Nom_Barri" : "Barrio",
                                                          "Población" :  "Habitantes",
                                                          "Superfície (ha)" : "Superficie general(ha)",
                                                          "Superfície Residencial (ha)" :  "Superficie Residencial (ha)",
                                                          "Densitat (hab/ha)" : "Densidad general(hab/ha)",
                                                          "Densitat neta (hab/ha)":  "Densidad neta(hab/ha)"
                                                     }
                                            }
                                        )

          cursor = barcelona_castellano.find({},{"_id":0}).limit(1)

          pd.DataFrame.from_dict(cursor)
```

Out[138]:

	Anio	Barrio	Cod_Barrio	Cod_Distrito	Densidad general(hab/ha)	Densidad neta(hab/ha)	Distrito	Habitantes	Superficie Residencial (ha)	Superficie general(ha)
0	2021	el Raval	1	1	429	952	Ciutat Vella	47228	49.6	110.0

Conclusiones

- Barcelona en el año 2021 es una ciudad que cuenta con una superficie residencial de 2745.9 hectáreas que se distribuyen en 73 barrios organizados en 10 distritos y con 1.636.732 habitantes. Obtenido la siguiente información:

  1. El distrito con más barrios es *Nou Barris* (13) y el que menos barrios tiene es el distrito *Les Corts* (3).
  2. En cuanto a porcentaje de población por distritos *Eixample* (16.15), es el distrito con mayor porcentaje y el de menor es *Les Corts* (4.92).
  3. Si hablamos de densidad neta, la media de Barcelona es aproximadamente 689 habitantes/hectárea. Pero a nivel barrios, el barrio con menor densidad neta es el *Tibidabo i les Planes Vallvidrera* (19) ,que pertenece al distrito *Sarrià-Sant Gervasi* y el barrio con mayor densidad neta es *Provençals del Poblenou* (1368), que pertenece al distrito *Sant Martí*.

- Población de barcelona desde 2017 hasta 2021:
  1. De 2017 a 2018 se observa un descenso en la población en orden de cientos.
  2. De 2018 a 2020 la tendencia es el crecimiento de la población en decenas de miles.
  3. De 2020 a 2021 también se observa un descenso en la población, pero está vez es en orden de decenas de miles (en torno a unas 27 mil personas), un factor a tener en cuenta es la pandemia.

  Esto se puede observar en la totalidad de Barcelona, pero también a nivel de distritos.

- La superficie residencial y no residencial de barcelona desde 2017 hasta 2021:
  1. De 2017 a 2018 se observa hubo un descenso en la superficie no residencial, es decir que se habrá más hectáreas que han pasado a ser suelo residencial.
  2. Si el análisis se hace a nivel de distritos, podemos ver que hay poca variación de porcentajes
  3. Si se analiza el porcentaje de superficie residencial y no residencial, pero a nivel de distritos observamos una escasa variación de porcentajes a lo largo de los años, pero hay un distrito en concreto que de 2017 a 2018 tiene un gran aumento de porcentaje (casi un 8%) de superficie residencial y es Sarrià-Sant Gervasi.