

# Padron\_Madrid

March 31, 2021

## 1 Datos socioeconómicos de Madrid

En este cuaderno se cruzan los datos del padrón 2020 con la renta media 2016 y los datos de los servicios sociales 2020.

@Author: Javier Cebrián Casado

La finalidad del cuaderno es mostrar el trabajo con la librería Pandas y Plotly: - Realizar agrupaciones (groupby) - Intersecciones de dataframes (merge) - Modificación de datos (apply) y otras operaciones de transformación - Representación - Correlación

@Source:

- [Padrón municipal Open Data](#)
- [Urgencias atendidas desde los Servicios Sociales del Ayuntamiento de Madrid](#)
- [Superficie por distritos](#)
- [Renta medida por distrito 2016](#)

```
[1]: import pandas as pd
import numpy as np
import plotly.express as px

import plotly.io as pio
pio.renderers.default = "notebook+pdf"
```

### 1.1 Carga datos de padrón Dic. 2020

Contiene los datos por distrito, barrio, edad, sexo y procedencia (extranjero o expaño)

```
[2]: padron=pd.read_csv('Rango_Edades_Seccion_202012.csv',delimiter=';')
print(padron.shape)
padron.head()
```

(237675, 12)

```
[2]:
```

	COD_DISTRITO	DESC_DISTRITO	COD_DIST_BARRIO	DESC_BARRIO	\
0	1	CENTRO	101	PALACIO	
1	1	CENTRO	101	PALACIO	
2	1	CENTRO	101	PALACIO	

3	1	CENTRO	101	PALACIO
4	1	CENTRO	101	PALACIO

	COD_BARRIO	COD_DIST_SECCION	COD_SECCION	COD_EDAD_INT	EspanolesHombres \
0	1	1006	6	103	NaN
1	1	1007	7	0	1.0
2	1	1007	7	1	2.0
3	1	1007	7	2	1.0
4	1	1007	7	3	4.0

	EspanolesMujeres	ExtranjerosHombres	ExtranjerosMujeres
0	1.0	NaN	NaN
1	1.0	NaN	3.0
2	3.0	NaN	NaN
3	4.0	NaN	NaN
4	NaN	NaN	NaN

### 1.1.1 Limpia espacios

El dataset original tiene espacios al final de los nombres de barrio y distrito. Con esta función elimina los espacios.

```
[3]: #Quita los espacios sobrantes que vienen en el archivo original
padron['DESC_DISTRITO']=padron['DESC_DISTRITO'].apply(lambda x: x.rstrip())
padron['DESC_BARRIO']=padron['DESC_BARRIO'].apply(lambda x: x.rstrip())
```

### 1.1.2 Crea un padrón con los index nombre del distrito y barrio

```
[4]: padronByDistNeig=padron[['DESC_DISTRITO', 'DESC_BARRIO', 'EspanolesHombres',
                             'EspanolesMujeres', 'ExtranjerosHombres',
                             'ExtranjerosMujeres']].
↪set_index(['DESC_DISTRITO', 'DESC_BARRIO'])
```

### 1.1.3 Agrupa la información por distrito y barrio

```
[5]: infoDistrito=padronByDistNeig.groupby('DESC_DISTRITO').sum()
infoBarrio=padronByDistNeig.groupby('DESC_BARRIO').sum()
infoDistrito.head()
```

```
[5]:
```

	EspanolesHombres	EspanolesMujeres	ExtranjerosHombres \
DESC_DISTRITO			
ARGANZUELA	64566.0	73947.0	7896.0
BARAJAS	21955.0	23156.0	2302.0
CARABANCHEL	94274.0	108948.0	27018.0
CENTRO	52732.0	52699.0	19507.0
CHAMARTIN	59872.0	71553.0	6119.0

	ExtranjerosMujeres
DESC_DISTRITO	
ARGANZUELA	8878.0
BARAJAS	2872.0
CARABANCHEL	29908.0
CENTRO	17005.0
CHAMARTIN	9035.0

#### 1.1.4 Calcula población total por distrito y barrio

Se obtiene para el daset de distritos y de barrios los totales de población por sexo, procedencia y total.

```
[6]: # Creación de nuevas columnas combinando datos
infoDistrito['TotalEspanoles']=infoDistrito[['EspanolesHombres',
                                              'EspanolesMujeres']].sum(axis=1)
infoDistrito['TotalExtranjeros']=infoDistrito[['ExtranjerosHombres',
                                              'ExtranjerosMujeres']].
    ↪sum(axis=1)
infoDistrito['TotalHombres']=infoDistrito[['EspanolesHombres',
                                              'ExtranjerosHombres']].sum(axis=1)
infoDistrito['TotalMujeres']=infoDistrito[['EspanolesMujeres',
                                              'ExtranjerosMujeres']].sum(axis=1)
infoDistrito['Total']=infoDistrito[['EspanolesHombres','EspanolesMujeres',
                                     'ExtranjerosHombres',
                                     'ExtranjerosMujeres']].sum(axis=1)
# Creación de nuevas columnas combinando datos
infoBarrio['TotalEspanoles']=infoBarrio[['EspanolesHombres',
                                              'EspanolesMujeres']].sum(axis=1)
infoBarrio['TotalExtranjeros']=infoBarrio[['ExtranjerosHombres',
                                              'ExtranjerosMujeres']].sum(axis=1)
infoBarrio['TotalHombres']=infoBarrio[['EspanolesHombres',
                                              'ExtranjerosHombres']].sum(axis=1)
infoBarrio['TotalMujeres']=infoBarrio[['EspanolesMujeres',
                                              'ExtranjerosMujeres']].sum(axis=1)
infoBarrio['Total']=infoBarrio[['EspanolesHombres','EspanolesMujeres',
                                     'ExtranjerosHombres','ExtranjerosMujeres']].
    ↪sum(axis=1)

infoDistrito.head()
```

```
[6]:
```

	EspanolesHombres	EspanolesMujeres	ExtranjerosHombres	\
DESC_DISTRITO				
ARGANZUELA	64566.0	73947.0	7896.0	
BARAJAS	21955.0	23156.0	2302.0	
CARABANCHEL	94274.0	108948.0	27018.0	
CENTRO	52732.0	52699.0	19507.0	

CHAMARTIN	59872.0	71553.0	6119.0
-----------	---------	---------	--------

	ExtranjerosMujeres	TotalEspañoles	TotalExtranjeros \
DESC_DISTRITO			
ARGANZUELA	8878.0	138513.0	16774.0
BARAJAS	2872.0	45111.0	5174.0
CARABANCHEL	29908.0	203222.0	56926.0
CENTRO	17005.0	105431.0	36512.0
CHAMARTIN	9035.0	131425.0	15154.0

	TotalHombres	TotalMujeres	Total
DESC_DISTRITO			
ARGANZUELA	72462.0	82825.0	155287.0
BARAJAS	24257.0	26028.0	50285.0
CARABANCHEL	121292.0	138856.0	260148.0
CENTRO	72239.0	69704.0	141943.0
CHAMARTIN	65991.0	80588.0	146579.0

## 1.2 Carga dataset de servicios sociales de urgencia 2020

```
[7]: servsoc=pd.read_csv('servicios_sociales_urgencias_2020.csv',delimiter=';')
print(servsoc.shape)
servsoc.head()
```

(24351, 18)

```
[7]:
```

	Secuencia	Código Centro	Centro	Código Distrito	Centro \
0	1	69	CSS José Villarreal		2
1	2	69	CSS José Villarreal		2
2	3	69	CSS José Villarreal		2
3	4	69	CSS José Villarreal		2
4	5	99	CSS Teresa de Calcuta		21

	Distrito Centro	Código Distrito	Distrito	Código Barrio	\
0	ARGANZUELA	2	ARGANZUELA	205	
1	ARGANZUELA	2	ARGANZUELA	205	
2	ARGANZUELA	2	ARGANZUELA	201	
3	ARGANZUELA	2	ARGANZUELA	206	
4	BARAJAS	21	BARAJAS	2101	

	Barrio	Sección Censal	Tramo Edad	Nacionalidad	Sexo \
0	DELICIAS	50	50 - 54	Española	H
1	DELICIAS	62	>= 85	Española	M
2	IMPERIAL	90	60 - 64	Extranjera	H
3	PALOS DE MOGUER	64	75 - 79	Española	H
4	ALAMEDA DE OSUNA	11	40 - 44	Española	M

	Código	Tipo Atención	Tipo Atención	Código	Tipo	Supuesto Urgente	\
0		1	Primera Atención			50.0	
1		2	Zona / Historia Social			40.0	
2		2	Zona / Historia Social			30.0	
3		1	Primera Atención			40.0	
4		2	Zona / Historia Social			50.0	

		Tipo Supuesto Urgente	Fecha Cita
0	Desprotección socio-familiar		02/01/2020
1	Desproteccion social problema salud		02/01/2020
2	Alojamiento/Vivienda		02/01/2020
3	Desproteccion social problema salud		02/01/2020
4	Desprotección socio-familiar		02/01/2020

### 1.2.1 Cálculo de servicios sociales de urgencia por distrito

```
[8]: servsoc['numeroServ']=np.ones(servsoc.shape[0])
servsocDistrict=servsoc[['Distrito','numeroServ']].groupby('Distrito').sum()
servsocDistrict.head()
```

```
[8]:
      numeroServ
Distrito
ARGANZUELA      225.0
BARAJAS          13.0
CARABANCHEL     595.0
CENTRO          334.0
CHAMARTÍN        57.0
```

### 1.3 Intersección de los dataset de padrón y servicios sociales

```
[9]: #Hay que corregir algunos nombres para que coincidan en ambos dataset (Quitar
      ↪acentos y algún espacio)
servsocDistrict.index=['ARGANZUELA', 'BARAJAS', 'CARABANCHEL', 'CENTRO',
      ↪'CHAMARTIN',
      'CHAMBERI', 'CIUDAD LINEAL', 'FUENCARRAL-EL PARDO', 'HORTALEZA',
      'LATINA', 'MONCLOA-ARAVACA', 'MORATALAZ', 'PUENTE DE VALLECAS',
      'RETIRO', 'SALAMANCA', 'SAN BLAS-CANILLEJAS', 'TETUAN', 'USERA',
      'VICALVARO', 'VILLA DE VALLECAS', 'VILLAVERDE']
infoDistrito=pd.
      ↪merge(infoDistrito,servsocDistrict,how='outer',left_index=True,right_index=True
      ↪)
infoDistrito.head()
```

```
[9]:
      EspanolesHombres  EspanolesMujeres  ExtranjerosHombres  \
DESC_DISTRITO
ARGANZUELA            64566.0            73947.0            7896.0
BARAJAS               21955.0            23156.0            2302.0
```

CARABANCHEL	94274.0	108948.0	27018.0
CENTRO	52732.0	52699.0	19507.0
CHAMARTIN	59872.0	71553.0	6119.0

	ExtranjerosMujeres	TotalEspaños	TotalExtranjeros \
DESC_DISTRITO			
ARGANZUELA	8878.0	138513.0	16774.0
BARAJAS	2872.0	45111.0	5174.0
CARABANCHEL	29908.0	203222.0	56926.0
CENTRO	17005.0	105431.0	36512.0
CHAMARTIN	9035.0	131425.0	15154.0

	TotalHombres	TotalMujeres	Total	numeroServ
DESC_DISTRITO				
ARGANZUELA	72462.0	82825.0	155287.0	225.0
BARAJAS	24257.0	26028.0	50285.0	13.0
CARABANCHEL	121292.0	138856.0	260148.0	595.0
CENTRO	72239.0	69704.0	141943.0	334.0
CHAMARTIN	65991.0	80588.0	146579.0	57.0

#### 1.4 Añadido de los datos de renta y superficie por distrito 2016

```
[10]: renta2016={'ARGANZUELA':17306,'BARAJAS':17641,'CARABANCHEL':10872,'CENTRO':
→16147,'CHAMARTIN':25969,'CHAMBERI':22499,
      'CIUDAD LINEAL':15048,'FUENCARRAL-EL PARDO':18573,'HORTALEZA':
→18277,'LATINA':12232,'MONCLOA-ARAVACA':22152,
      'MORATALAZ':13944,'PUENTE DE VALLECAS':9545,'RETIRO':
→21504,'SALAMANCA':24433,'SAN BLAS-CANILLEJAS':13404,
      'TETUAN':14970,'USERA':9395,'VICALVARO':11695,'VILLA DE VALLECAS':
→11925,'VILLAVERDE':9756}

superficie={'ARGANZUELA':646.22,'BARAJAS':4192.28,'CARABANCHEL':1404.
→83,'CENTRO':522.82,'CHAMARTIN':917.55,'CHAMBERI':467.92,
      'CIUDAD LINEAL':1142.57,'FUENCARRAL-EL PARDO':23783.84,'HORTALEZA':
→2741.98,'LATINA':2542.72,'MONCLOA-ARAVACA':4653.11,
      'MORATALAZ':610.32,'PUENTE DE VALLECAS':1496.86,'RETIRO':546.
→62,'SALAMANCA':539.24,'SAN BLAS-CANILLEJAS':2229.24,
      'TETUAN':537.47,'USERA':777.77,'VICALVARO':3526.67,'VILLA DE
→VALLECAS':5146.72,'VILLAVERDE':2018.76}

infoDistrito['RentaMedia2016']=pd.Series(renta2016)
infoDistrito['SuperficieHa']=pd.Series(superficie)

infoDistrito.head()
```

```
[10]:          EspanolesHombres  EspanolesMujeres  ExtranjerosHombres \
DESC_DISTRITO
```

ARGANZUELA	64566.0	73947.0	7896.0
BARAJAS	21955.0	23156.0	2302.0
CARABANCHEL	94274.0	108948.0	27018.0
CENTRO	52732.0	52699.0	19507.0
CHAMARTIN	59872.0	71553.0	6119.0

	ExtranjerosMujeres	TotalEspaños	TotalExtranjeros \
DESC_DISTRITO			
ARGANZUELA	8878.0	138513.0	16774.0
BARAJAS	2872.0	45111.0	5174.0
CARABANCHEL	29908.0	203222.0	56926.0
CENTRO	17005.0	105431.0	36512.0
CHAMARTIN	9035.0	131425.0	15154.0

	TotalHombres	TotalMujeres	Total	numeroServ \
DESC_DISTRITO				
ARGANZUELA	72462.0	82825.0	155287.0	225.0
BARAJAS	24257.0	26028.0	50285.0	13.0
CARABANCHEL	121292.0	138856.0	260148.0	595.0
CENTRO	72239.0	69704.0	141943.0	334.0
CHAMARTIN	65991.0	80588.0	146579.0	57.0

	RentaMedia2016	SuperficieHa
DESC_DISTRITO		
ARGANZUELA	17306	646.22
BARAJAS	17641	4192.28
CARABANCHEL	10872	1404.83
CENTRO	16147	522.82
CHAMARTIN	25969	917.55

## 1.5 Agrupación de la población de los distritos por edad

```
[11]: #Métodos para agrupar de 5 en 5 o 10 en 10
def groupAgesQuinquenal(x):
    if(x==0 or x<=4):
        return '0-4'
    elif(x==5 or x<=9):
        return '5-9'
    elif(x==10 or x<=14):
        return '10-14'
    elif(x==15 or x<=19):
        return '15-19'
    elif(x==20 or x<=24):
        return '20-24'
    elif(x==25 or x<=29):
        return '25-29'
    elif(x==30 or x<=34):
```

```

        return '30-34'
    elif(x==35 or x<=39):
        return '35-39'
    elif(x==40 or x<=44):
        return '40-44'
    elif(x==45 or x<=49):
        return '45-49'
    elif(x==50 or x<=54):
        return '50-54'
    elif(x==55 or x<=59):
        return '55-59'
    elif(x==60 or x<=64):
        return '60-64'
    elif(x==65 or x<=69):
        return '65-69'
    elif(x==70 or x<=74):
        return '70-74'
    elif(x==75 or x<=79):
        return '75-79'
    elif(x==80 or x<=84):
        return '80-84'
    elif(x==85 or x<=89):
        return '85-89'
    elif(x==90 or x<=94):
        return '90-94'
    elif(x==95 or x<=99):
        return '95-99'
    elif(x>=100):
        return '>=100'

def groupAgesDecanal(x):
    if(x==0 or x<=4):
        return '0-4'
    elif(x==5 or x<=9):
        return '5-9'
    elif(x==10 or x<=14):
        return '10-14'
    elif(x==15 or x<=19):
        return '15-19'
    elif(x==20 or x<=29):
        return '20-29'
    elif(x==30 or x<=39):
        return '30-39'
    elif(x==40 or x<=49):
        return '40-49'
    elif(x==50 or x<=59):

```



```

        return '50-59'
    elif(x==60 or x<=65):
        return '60-65'
    elif(x==66 or x<=69):
        return '60-69'
    elif(x==70 or x<=79):
        return '70-79'
    elif(x==80 or x<=89):
        return '80-89'
    elif(x==90 or x<=99):
        return '90-99'
    elif(x>=100):
        return '>=100'

def groupAgesSimples(x):
    if(x==0 or x<=4):
        return '0-4'
    elif(x==5 or x<=14):
        return '5-14'
    elif(x==15 or x<=24):
        return '15-24'
    elif(x==25 or x<=64):
        return '25-64'
    elif(x>=65):
        return '>=65'

```

```

[31]: padronByEdadDecanal=padron[['COD_EDAD_INT', 'DESC_DISTRITO', 'DESC_BARRIO',
                                'EspanolesHombres', 'EspanolesMujeres',
                                'ExtranjerosHombres', 'ExtranjerosMujeres']].
    ↪set_index(['DESC_DISTRITO',
    ↪    'DESC_BARRIO'])
padronByEdadDecanal['COD_EDAD_INT']=padronByEdadDecanal['COD_EDAD_INT'].
    ↪apply(groupAgesDecanal)
padronByEdadDecanal=padronByEdadDecanal.
    ↪groupby(['DESC_DISTRITO', 'COD_EDAD_INT']).sum()
padronByEdadDecanal['Total']=padronByEdadDecanal[['EspanolesHombres',
                                'EspanolesMujeres',
                                ↪    'ExtranjerosHombres', 'ExtranjerosMujeres']].sum(axis=1)
padronByEdadDecanal=pd.pivot_table(padronByEdadDecanal, values='Total',
    ↪index=['DESC_DISTRITO'],
                                columns=['COD_EDAD_INT'],aggfunc=np.sum)
padronByEdadDecanal['Total']=padronByEdadDecanal[['0-4', '10-14',
                                '15-19', '20-29',
                                '30-39', '40-49',
                                '5-9', '50-59',

```

```
'60-65','60-69',
'70-79','80-89',
'90-99','>=100']] .sum(axis=1)
```

padronByEdadDecanal

[31]:	COD_EDAD_INT	0-4	10-14	15-19	20-29	30-39	40-49	\
	DESC_DISTRITO							
	ARGANZUELA	5853.0	5766.0	5870.0	17300.0	24901.0	24848.0	
	BARAJAS	2408.0	2962.0	2761.0	4580.0	6370.0	9315.0	
	CARABANCHEL	10051.0	12723.0	12079.0	32004.0	36297.0	44372.0	
	CENTRO	3558.0	3257.0	3346.0	21226.0	31059.0	25997.0	
	CHAMARTIN	6390.0	5937.0	5819.0	16220.0	21200.0	21040.0	
	CHAMBERI	4793.0	4409.0	4810.0	18380.0	22665.0	19276.0	
	CIUDAD LINEAL	7515.0	8841.0	9645.0	25106.0	29729.0	32178.0	
	FUENCARRAL-EL PARDO	12566.0	13975.0	11434.0	23300.0	32110.0	42484.0	
	HORTALEZA	9685.0	10939.0	9748.0	18529.0	25469.0	33739.0	
	LATINA	8277.0	9626.0	10207.0	27662.0	31952.0	36666.0	
	MONCLOA-ARAVACA	4775.0	5668.0	6296.0	14662.0	15177.0	17376.0	
	MORATALAZ	2902.0	3701.0	4297.0	10763.0	10635.0	12375.0	
	PUENTE DE VALLECAS	9726.0	10598.0	11361.0	31890.0	34893.0	36467.0	
	RETIRO	4341.0	4569.0	4581.0	12269.0	15803.0	17240.0	
	SALAMANCA	5239.0	4860.0	5485.0	19216.0	22266.0	20114.0	
	SAN BLAS-CANILLEJAS	6056.0	8372.0	9464.0	18545.0	20312.0	25256.0	
	TETUAN	6032.0	5488.0	5974.0	21686.0	29097.0	24760.0	
	USERA	5795.0	7383.0	7533.0	18193.0	19956.0	24633.0	
	VICALVARO	2976.0	5147.0	4554.0	8477.0	8890.0	14114.0	
	VILLA DE VALLECAS	6718.0	6315.0	5147.0	12047.0	19000.0	23332.0	
	VILLAVERDE	6924.0	8643.0	8131.0	18992.0	21224.0	25892.0	
	COD_EDAD_INT	5-9	50-59	60-65	60-69	70-79	80-89	\
	DESC_DISTRITO							
	ARGANZUELA	5864.0	23371.0	12568.0	6676.0	12369.0	7427.0	
	BARAJAS	2905.0	7369.0	3039.0	1843.0	4366.0	1946.0	
	CARABANCHEL	11559.0	38326.0	16564.0	8495.0	19384.0	14562.0	
	CENTRO	3204.0	20064.0	9477.0	4706.0	8825.0	5370.0	
	CHAMARTIN	6316.0	19887.0	10968.0	6381.0	14172.0	9123.0	
	CHAMBERI	4589.0	18010.0	10580.0	6388.0	13697.0	8711.0	
	CIUDAD LINEAL	8406.0	32796.0	16546.0	8601.0	19697.0	14975.0	
	FUENCARRAL-EL PARDO	14986.0	31864.0	16382.0	9963.0	24113.0	12665.0	
	HORTALEZA	11270.0	27793.0	12669.0	6720.0	16257.0	9159.0	
	LATINA	9125.0	35283.0	16699.0	8682.0	24937.0	18369.0	
	MONCLOA-ARAVACA	5409.0	17795.0	9066.0	4990.0	11158.0	6981.0	
	MORATALAZ	3296.0	15368.0	7512.0	3935.0	9674.0	8189.0	
	PUENTE DE VALLECAS	10025.0	37754.0	17945.0	7910.0	15748.0	12900.0	
	RETIRO	4635.0	16411.0	9731.0	6012.0	13225.0	7958.0	
	SALAMANCA	5018.0	19983.0	11269.0	6452.0	14160.0	9371.0	

SAN BLAS-CANILLEJAS	6983.0	27944.0	11679.0	5299.0	10374.0	8493.0
TETUAN	5651.0	21973.0	11285.0	5724.0	11964.0	8383.0
USERA	6586.0	21672.0	9074.0	4316.0	8879.0	7330.0
VICALVARO	4017.0	12900.0	4345.0	1889.0	4467.0	3323.0
VILLA DE VALLECAS	7567.0	14941.0	6588.0	3342.0	6132.0	3420.0
VILLASVERDE	7596.0	23818.0	9640.0	4426.0	10326.0	8297.0

COD_EDAD_INT	90-99	>=100	Total
DESC_DISTRITO			
ARGANZUELA	2390.0	84.0	155287.0
BARAJAS	412.0	9.0	50285.0
CARABANCHEL	3629.0	103.0	260148.0
CENTRO	1784.0	70.0	141943.0
CHAMARTIN	3016.0	110.0	146579.0
CHAMBERI	3201.0	163.0	139672.0
CIUDAD LINEAL	3999.0	111.0	218145.0
FUENCARRAL-EL PARDO	3038.0	91.0	248971.0
HORTALEZA	2087.0	61.0	194125.0
LATINA	4030.0	133.0	241648.0
MONCLOA-ARAVACA	2074.0	91.0	121518.0
MORATALAZ	1764.0	51.0	94462.0
PUENTE DE VALLECAS	3166.0	65.0	240448.0
RETIRO	2415.0	94.0	119284.0
SALAMANCA	3209.0	137.0	146779.0
SAN BLAS-CANILLEJAS	2295.0	34.0	161106.0
TETUAN	2727.0	79.0	160823.0
USERA	1899.0	53.0	143302.0
VICALVARO	534.0	15.0	75648.0
VILLA DE VALLECAS	728.0	22.0	115299.0
VILLASVERDE	1645.0	24.0	155578.0

```
[13]: padronByEdadSimplif=padron[['COD_EDAD_INT', 'DESC_DISTRITO', 'DESC_BARRIO',
                                'EspanolesHombres', 'EspanolesMujeres',
                                'ExtranjerosHombres', 'ExtranjerosMujeres']].
    ↪set_index(['DESC_DISTRITO', 'DESC_BARRIO'])
padronByEdadSimplif['COD_EDAD_INT']=padronByEdadSimplif['COD_EDAD_INT'].
    ↪apply(groupAgesSimples)
padronByEdadSimplif=padronByEdadSimplif.
    ↪groupby(['DESC_DISTRITO', 'COD_EDAD_INT']).sum()
padronByEdadSimplif['Total']=padronByEdadSimplif[['EspanolesHombres',
                                                    'EspanolesMujeres',
                                                    'ExtranjerosHombres',
                                                    'ExtranjerosMujeres']].
    ↪sum(axis=1)
padronByEdadSimplif=pd.pivot_table(padronByEdadSimplif, values='Total',
    ↪index=['DESC_DISTRITO'], columns=['COD_EDAD_INT'],
```

```

                                aggfunc=np.sum)
padronByEdadSimplif['Total']=padronByEdadSimplif[["0-4", '5-14',
                                                    '15-24', '25-64',
                                                    '>=65']].sum(axis=1)

padronByEdadSimplif

```

```

[13]: COD_EDAD_INT      0-4    15-24    25-64    5-14    >=65    Total
DESC_DISTRITO
ARGANZUELA           5853.0  12866.0  94094.0  11630.0  30844.0  155287.0
BARAJAS              2408.0   5036.0  27974.0   5867.0   9000.0   50285.0
CARABANCHEL         10051.0  26460.0 150817.0  24282.0  48538.0  260148.0
CENTRO               3558.0  10285.0  99499.0   6461.0  22140.0  141943.0
CHAMARTIN            6390.0  12368.0   81112.0  12253.0  34456.0  146579.0
CHAMBERI             4793.0  11489.0   80494.0   8998.0  33898.0  139672.0
CIUDAD LINEAL        7515.0  20679.0 122970.0  17247.0  49734.0  218145.0
FUENCARRAL-EL PARDO 12566.0  22209.0 132736.0  28961.0  52499.0  248971.0
HORTALEZA            9685.0  18676.0 107409.0  22209.0  36146.0  194125.0
LATINA               8277.0  22712.0 133287.0  18751.0  58621.0  241648.0
MONCLOA-ARAVACA      4775.0  13200.0   65764.0  11077.0  26702.0  121518.0
MORATALAZ            2902.0   9432.0   50446.0   6997.0  24685.0   94462.0
PUENTE DE VALLECAS   9726.0  26074.0 141854.0  20623.0  42171.0  240448.0
RETIRO               4341.0   9667.0   64735.0   9204.0  31337.0  119284.0
SALAMANCA            5239.0  13088.0   83488.0   9878.0  35086.0  146779.0
SAN BLAS-CANILLEJAS  6056.0  18503.0   93155.0  15355.0  28037.0  161106.0
TETUAN               6032.0  14264.0   98833.0  11139.0  30555.0  160823.0
USERA                5795.0  16145.0   83686.0  13969.0  23707.0  143302.0
VICALVARO            2976.0   8902.0   43874.0   9164.0  10732.0   75648.0
VILLA DE VALLECAS    6718.0  10440.0   69621.0  13882.0  14638.0  115299.0
VILLAVERDE           6924.0  17079.0   89321.0  16239.0  26015.0  155578.0

```

## 2 Representaciones gráficas

### 2.1 Ordenación de los dataset en función de la población total

```

[14]: infoDistrito=infoDistrito.sort_values('Total')
      infoBarrio=infoBarrio.sort_values('Total')

```

### 2.2 Corrección de codificación en ñ

```

[15]: infoBarrio.index=['BARRIOS EN EDIF. BDC', 'ATOCHA', 'ATALAYA', 'AEROPUERTO', 'EL_
↳ PLANTIO', 'FUENTELARREINA', 'EL PARDO', 'EL CAÑAVERAL',
'CUATRO VIENTOS', 'COLINA', 'HORCAJO', 'PALOMAS', 'VALDEMARIN', 'LOS_
↳ JERONIMOS', 'CASCO H.BARAJAS', 'CORRALEJOS', 'SOL', 'AMPOSTA',
'PAVONES', 'HELLIN', 'CORTES', 'EL SALVADOR', 'SAN JUAN BAUTISTA', 'CASA DE_
↳ CAMPO', 'TIMON', 'ORCASUR', 'ZOFIO', 'PIOVERA', 'NIÑO JESUS',

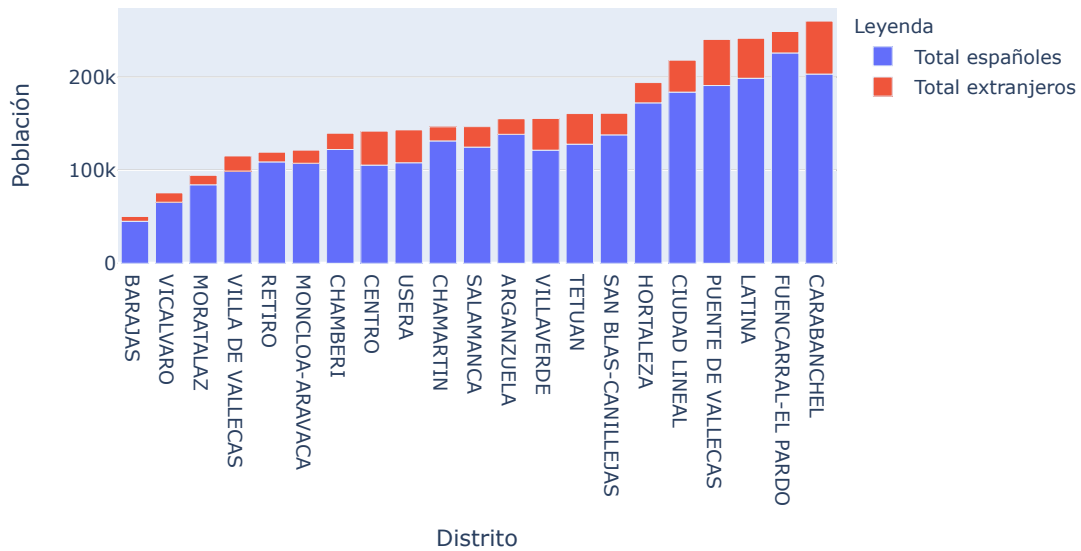
```

```
'APOSTOL SANTIAGO','RECOLETOS','CIUDAD UNIVERSITARIA','VINATEROS','SAN_
↳CRISTOBAL','CASTILLA','CASTELLANA','REJAS','FONTARRON',
'EL VISO','VALDEBERNARDO','LOS CARMENES','MEDIA LEGUA','SAN_
↳PASCUAL','PRADOLONGO','VALDERRIVAS','JUSTICIA','CIUDAD JARDIN',
'ADELFAS','EL GOLOSO','ALMAGRO','CAMPAMENTO','LEGAZPI','ALAMEDA DE_
↳OSUNA','CHOPERA','VALLEHERMOSO','CASTILLEJOS',
'BUTARQUE','CONCEPCION','LISTA','FUENTE DEL_
↳BERRO','IBIZA','COSTILLARES','ALMENDRALES','IMPERIAL','COMILLAS','ALMENARA',
'GAZTAMBIDE','ESTRELLA','ORCASITAS','PALACIO','SAN FERMIN','ARGUELLES','SANTA_
↳EUGENIA','ARAPILES','ARCOS','TRAFALGAR','QUINTANA',
'NUEVA ESPAÑA','BERRUGUETE','PALOS DE_
↳MOGUER','VALDEACEDERAS','MOSCARDO','MARROQUINA','RIOS_
↳ROSAS','ARAVACA','DELICIAS',
'SIMANCAS','PORTAZGO','CANILLEJAS','GOYA','BELLAS_
↳VISTAS','VALDEZARZA','ROSAS','ABRANTES','HISPANOAMERICA','LOS ANGELES',
'LA PAZ','PACIFICO','UNIVERSIDAD','OPAÑEL','MIRASIERRA','CUATRO_
↳CAMINOS','ENTREVIAS','CASCO H.VICALVARO','PUERTA BONITA','ACACIAS',
'PROSPERIDAD','LUCERO','LOS ROSALES','SAN ISIDRO','CASCO H.
↳VALLECAS','CANILLAS','PALOMERAS BAJAS','GUINDALERA','PUERTA DEL ANGEL',
'PALOMERAS SURESTE','SAN DIEGO','PEÑA GRANDE','EL PILAR','VILLAVERDE ALTO C.H.
↳','VISTA ALEGRE','EMBAJADORES','BUENAVISTA',
'NUMANCIA','VENTAS','ENSANCHE DE VALLECAS','LAS AGUILAS','PINAR DEL_
↳REY','PUEBLO NUEVO','VALDEFUENTES','VALVERDE','ALUCHE']
```

## 2.3 1- Población española y extranjera por distrito

```
[16]: fig = px.bar(infoDistrito, x=infoDistrito.index,
↳y=["TotalEspañoles","TotalExtranjeros"],
title="Población española y extranjera por distrito"
)
fig.layout.xaxis.title.text='Distrito'
fig.layout.yaxis.title.text='Población'
fig.layout.legend.title.text='Leyenda'
fig.data[0].name='Total españoles'
fig.data[1].name='Total extranjeros'
fig.show()
```

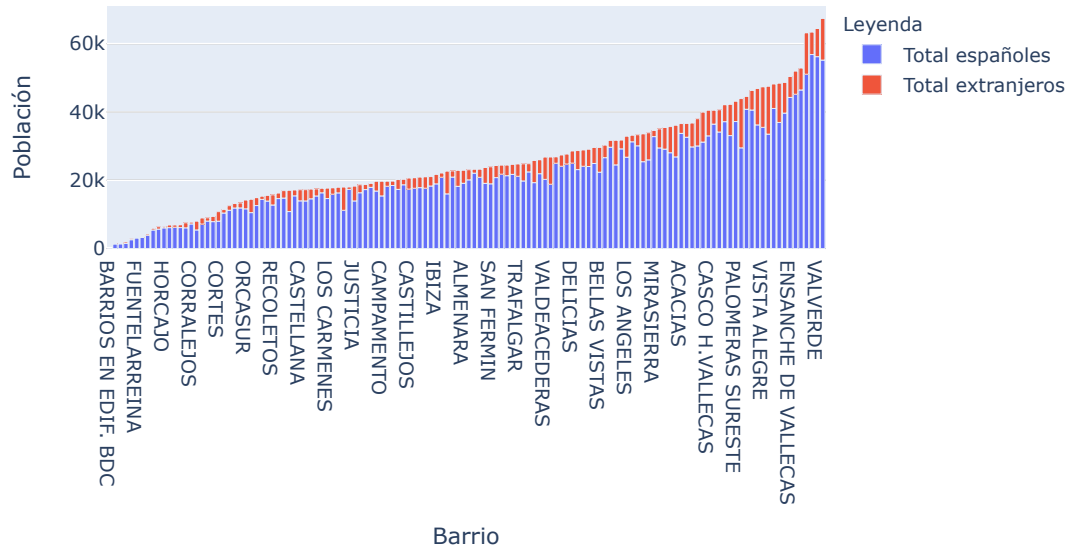
## Población española y extranjera por distrito



## 2.4 2- Población española y extranjera por barrio

```
[17]: fig = px.bar(infoBarrio, x=infoBarrio.index,
    ↪y=["TotalEspanoles", 'TotalExtranjeros'],
        title="Población española y extranjera por barrio")
fig.layout.xaxis.title.text='Barrio'
fig.layout.yaxis.title.text='Población'
fig.layout.legend.title.text='Leyenda'
fig.data[0].name='Total españoles'
fig.data[1].name='Total extranjeros'
fig.show()
```

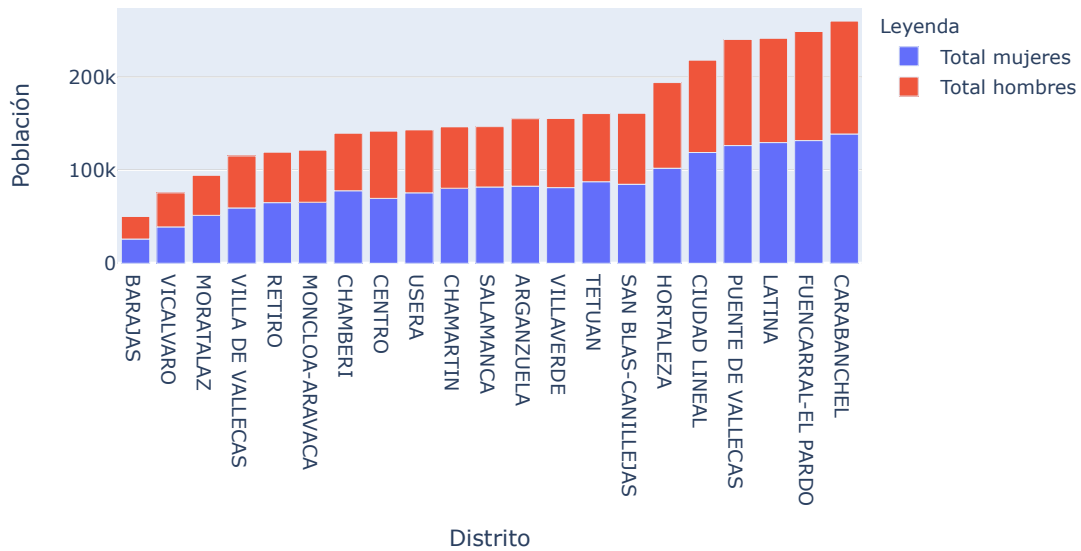
## Población española y extranjera por barrio



## 2.5 3- Población de hombres y mujeres por distrito

```
[18]: fig = px.bar(infoDistrito, x=infoDistrito.index,
    →y=["TotalMujeres", 'TotalHombres'],
        title="Población de hombres y mujeres pors distrito")
fig.layout.xaxis.title.text='Distrito'
fig.layout.yaxis.title.text='Población'
fig.layout.legend.title.text='Leyenda'
fig.data[0].name='Total mujeres'
fig.data[1].name='Total hombres'
fig.show()
```

## Población de hombres y mujeres pors distrito

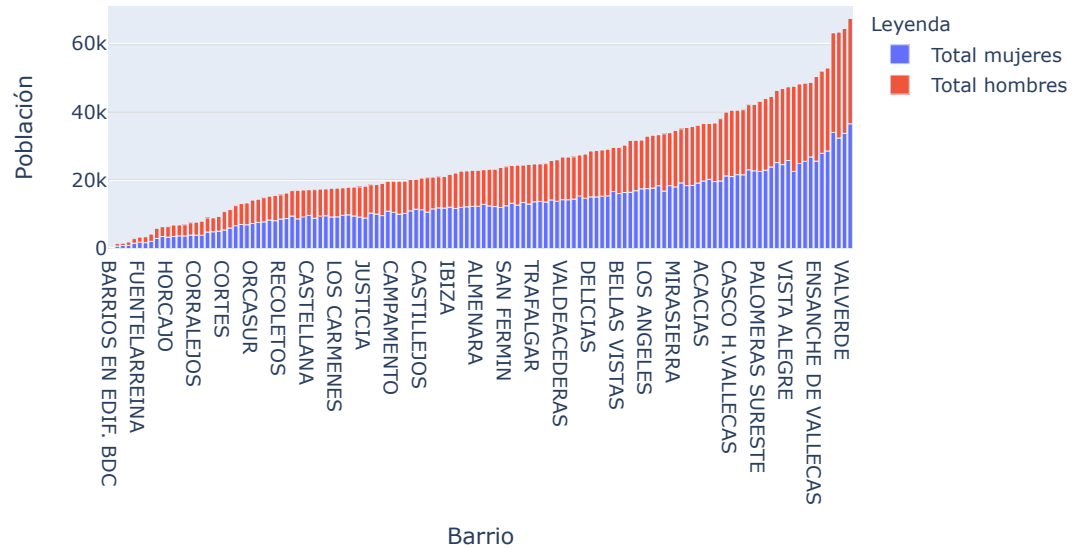


## 2.6 4- Población de hombres y mujeres por barrio

```
[19]: fig = px.bar(infoBarrio, x=infoBarrio.index, y=["TotalMujeres", 'TotalHombres'],
                title="Población de hombres y mujeres por barrio")
fig.layout.xaxis.title.text='Barrio'
fig.layout.yaxis.title.text='Población'
fig.layout.legend.title.text='Leyenda'
fig.data[0].name='Total mujeres'
fig.data[1].name='Total hombres'
fig.show()
```



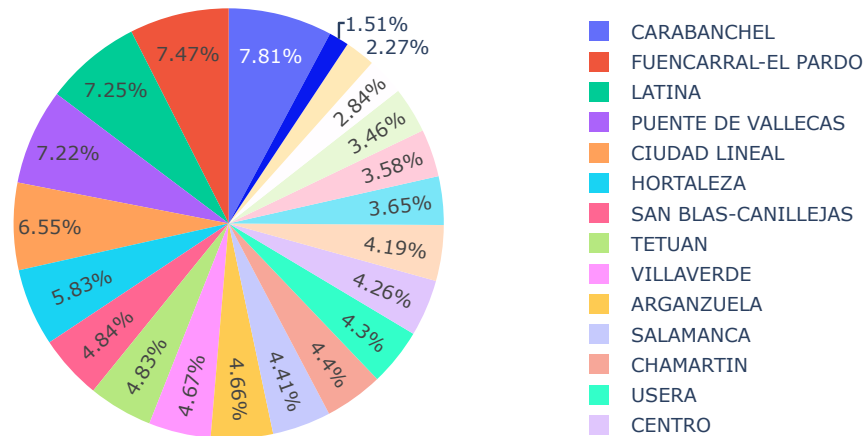
## Población de hombres y mujeres por barrio



## 2.7 5- Población total por distrito

```
[20]: fig = px.pie(infoDistrito, values='Total', names=infoDistrito.index,
    ↪title='Población de Madrid por distrito')
fig.show()
```

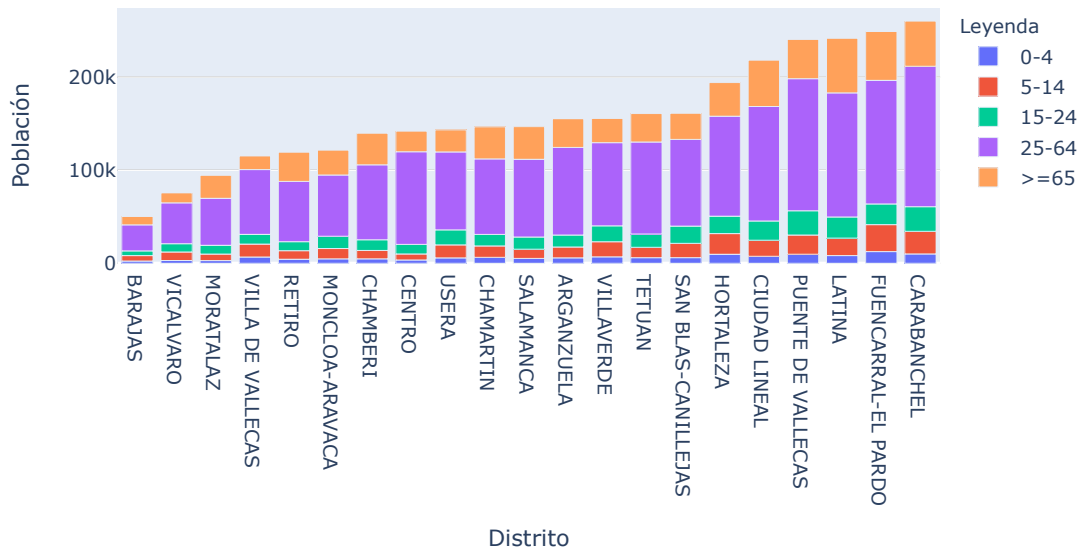
## Población de Madrid por distrito



## 2.8 6- Población por grupo de edad y por distrito

```
[21]: padronByEdadSimplif=padronByEdadSimplif.sort_values('Total')
fig = px.bar(padronByEdadSimplif, x=padronByEdadSimplif.index,
            y=["0-4", '5-14', '15-24', '25-64', '>=65'],
            title="Población por grupo de edad y por distrito"
        )
fig.layout.xaxis.title.text='Distrito'
fig.layout.yaxis.title.text='Población'
fig.layout.legend.title.text='Leyenda'
fig.show()
```

Población por grupo de edad y por distrito

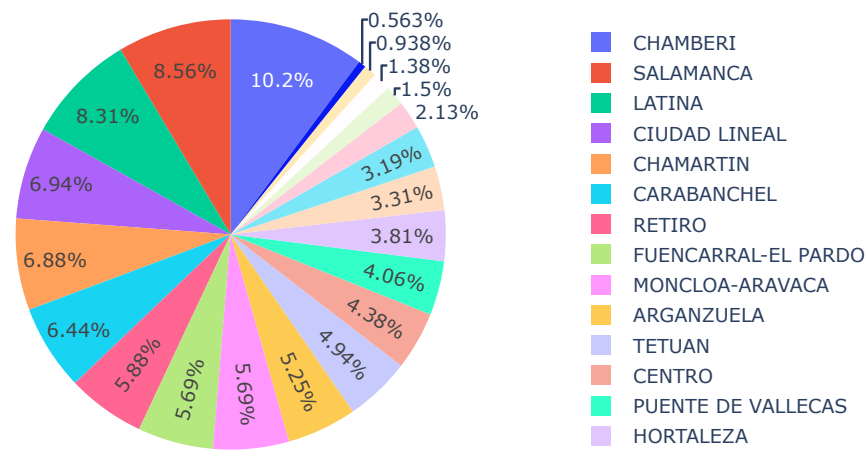


## 2.9 7- Población centenaria por distrito

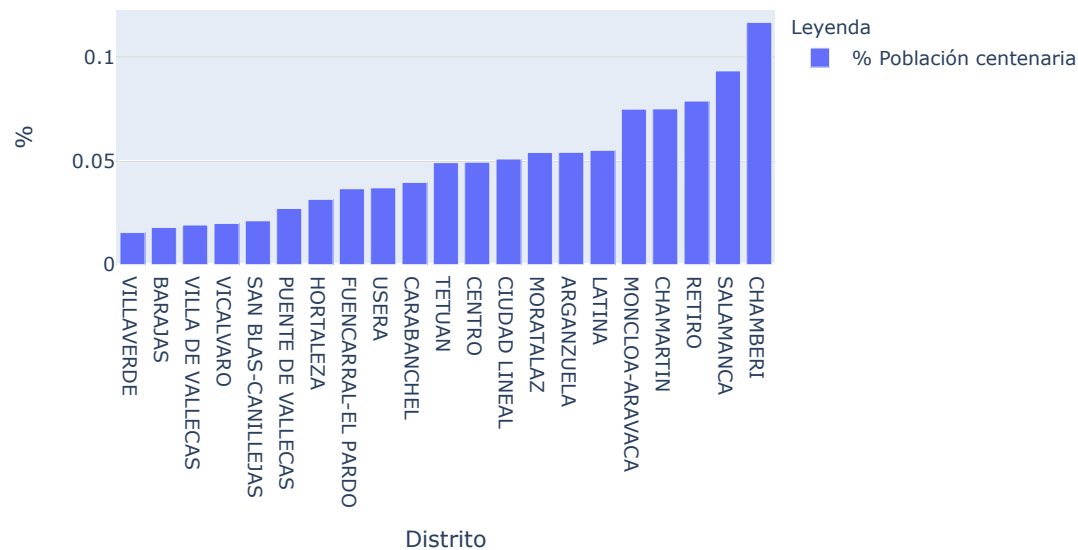
```
[22]: fig = px.pie(padronByEdadDecanal, values=padronByEdadDecanal['>=100'],
    names=padronByEdadDecanal.index, title='Total de la población centenaria por
    distrito')
fig.show()

padronByEdadDecanal['%CentenarioDistrito']=(padronByEdadDecanal['>=100']*100)/
    padronByEdadDecanal['Total']
padronByEdadDecanal=padronByEdadDecanal.sort_values(['%CentenarioDistrito'])
fig = px.bar(padronByEdadDecanal, x=padronByEdadDecanal.index,
    y=['%CentenarioDistrito'],
    title="Porcentaje de la población centenaria respecto al población
    de cada distrito")
fig.layout.xaxis.title.text='Distrito'
fig.layout.yaxis.title.text='%'
fig.layout.legend.title.text='Leyenda'
fig.data[0].name='% Población centenaria'
fig.show()
```

Total de la población centenaria por distrito

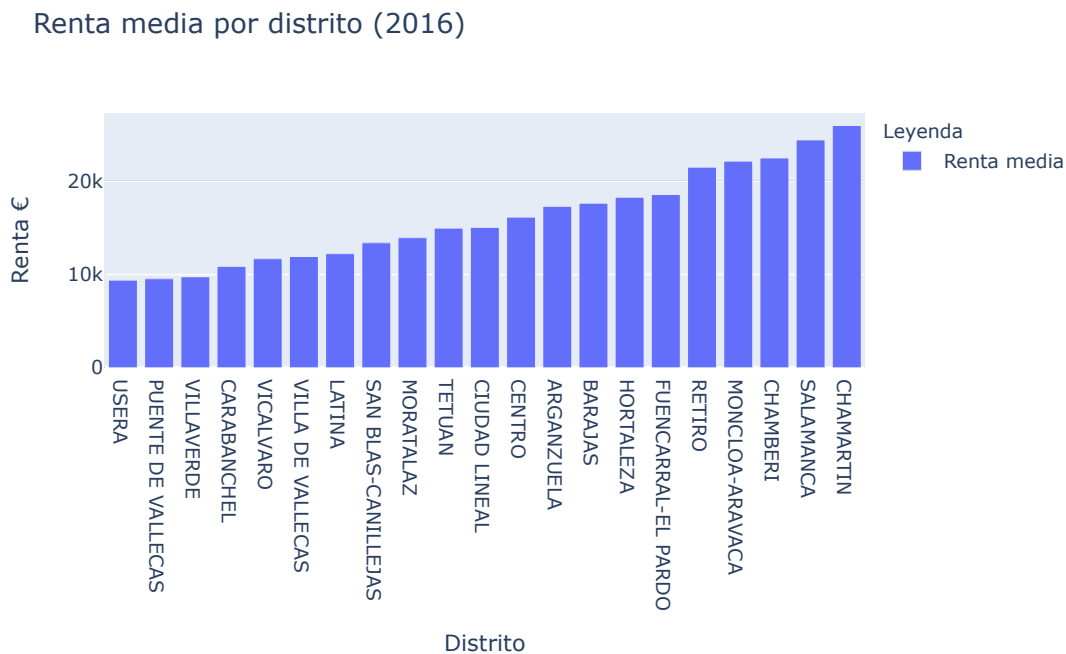


Porcentaje de la población centenaria respecto al población de cada distrito



## 2.10 8- Renta media 2016 por distrito

```
[23]: infoDistrito=infoDistrito.sort_values(['RentaMedia2016'])
fig = px.bar(infoDistrito, x=infoDistrito.index, y=['RentaMedia2016'],
             title="Renta media por distrito (2016)")
fig.layout.xaxis.title.text='Distrito'
fig.layout.yaxis.title.text='Renta €'
fig.layout.legend.title.text='Leyenda'
fig.data[0].name='Renta media'
fig.show()
```

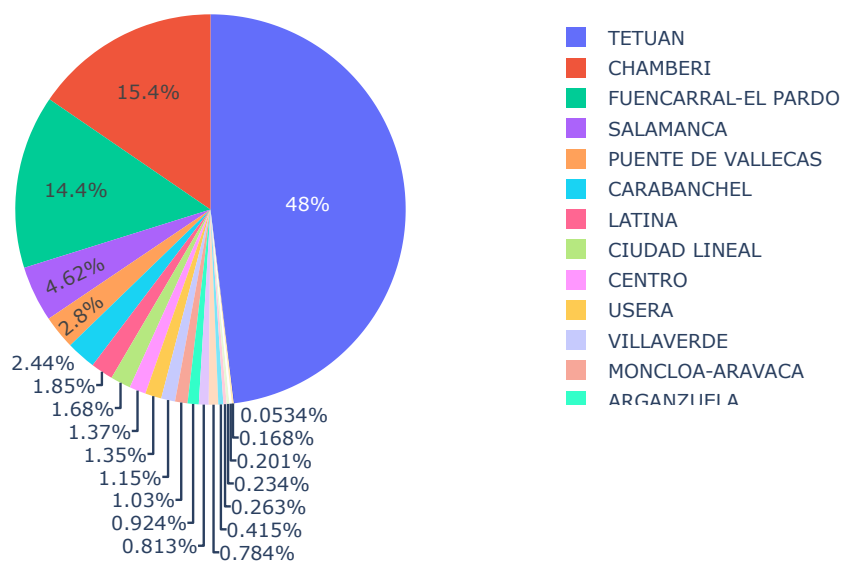


## 2.11 9- Servicios sociales de urgencia por distrito en 2020

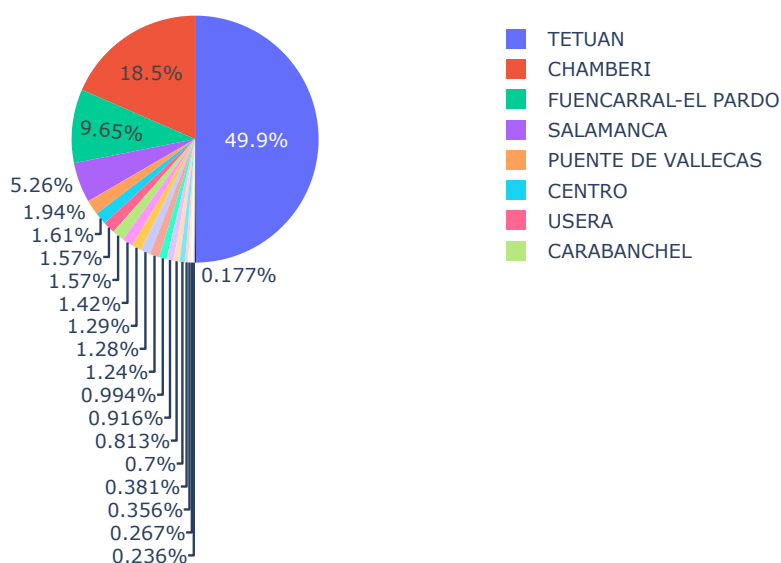
```
[24]: fig = px.pie(infoDistrito, values=infoDistrito['numeroServ'],
                 ↪names=infoDistrito.index, title='Distribución de servicios sociales de_
                 ↪urgencia por distrito en datos absolutos')
fig.show()

fig = px.pie(infoDistrito, values=infoDistrito['numeroServ']/
                 ↪infoDistrito['Total'], names=infoDistrito.index, title='Distribución del_
                 ↪ratio servicios sociales de urgencia / población para cada distrito')
fig.show()
```

Distribución de servicios sociales de urgencia por distrito en datos absolutos



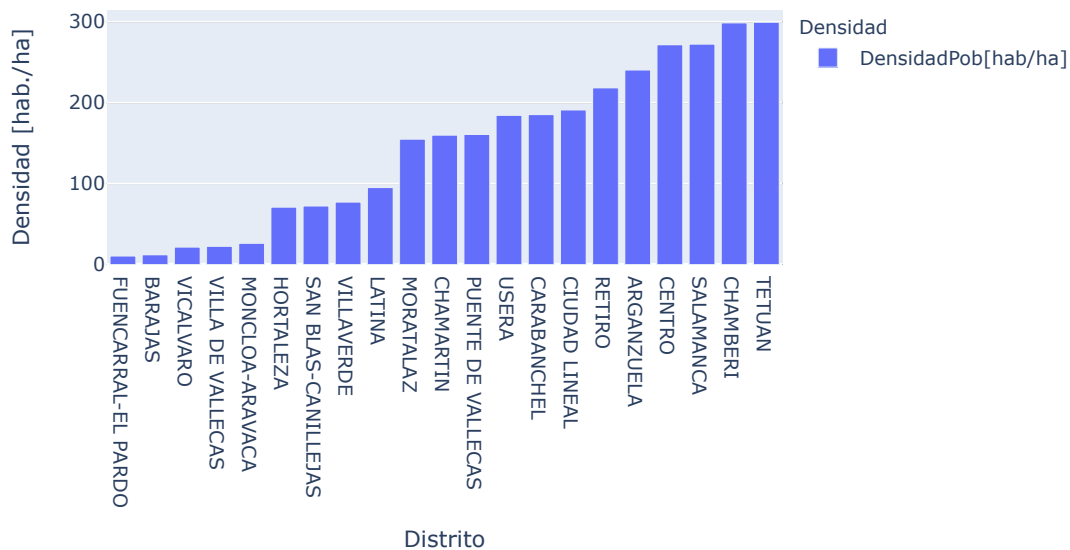
Distribución del ratio servicios sociales de urgencia / población para cada distrito



## 2.12 10- Densidad de población por distrito

```
[25]: infoDistrito['DensidadPob[hab/ha]']=infoDistrito['Total']/  
      ↪infoDistrito['SuperficieHa']  
infoDistrito=infoDistrito.sort_values(['DensidadPob[hab/ha]'])  
fig = px.bar(infoDistrito, x=infoDistrito.index, y=['DensidadPob[hab/ha]'],  
             title="Densidad de población por distrito [hab./ha]")  
fig.layout.xaxis.title.text='Distrito'  
fig.layout.yaxis.title.text='Densidad [hab./ha]'  
fig.layout.legend.title.text='Densidad'  
fig.show()
```

Densidad de población por distrito [hab./ha]



## 3 Correlaciones

### 3.1 1- Correlación ratio $\frac{\text{Población por edad}}{\text{Población total de cada distrito}}$ y renta por distrito 2016

```
[26]: for i in iter(padronByEdadDecanal.columns):  
      # print(infoDistrito['RentaMedia2016'].corr(padronByEdad[i]/  
      ↪infoDistrito['Total']))  
      if abs(infoDistrito['RentaMedia2016'].corr(padronByEdadDecanal[i]/  
      ↪infoDistrito['Total']))>=0.7 and i!='%CentenarioDistrito':
```

```
print('Correlación >= 0,7 para franja de edad:'+i+' es_
↳'+str(round(infoDistrito['RentaMedia2016'].corr(padronByEdadDecanal[i]/
↳infoDistrito['Total']),5)))
```

Correlación >= 0,7 para franja de edad:60-69 es 0.80103

Correlación >= 0,7 para franja de edad:>=100 es 0.74789

### 3.2 2- Correlación ratio $\frac{\text{Servicios sociales de urgencia 2020}}{\text{Población total de cada distrito}}$ y renta por distrito 2016

```
[27]: (infoDistrito['numeroServ']/infoDistrito['Total']).
↳corr(infoDistrito['RentaMedia2016'])
```

[27]: 0.08397805704376078

### 3.3 3- Correlación ratio $\frac{\text{Población extranjera}}{\text{Población total de cada distrito}}$ y renta por distrito 2016

```
[28]: (infoDistrito['TotalExtranjeros']/infoDistrito['Total']).
↳corr(infoDistrito['RentaMedia2016'])
```

[28]: -0.6134603214140901

### 3.4 4- Correlación densidad de población y renta por distrito 2016

```
[29]: infoDistrito[['DensidadPob[hab/ha]', 'RentaMedia2016']].corr()
```

```
[29]:
```

	DensidadPob[hab/ha]	RentaMedia2016
DensidadPob[hab/ha]	1.000000	0.233196
RentaMedia2016	0.233196	1.000000

## 4 Store created CSV files

```
[30]: infoDistrito.to_csv(r'infodistrito.csv', sep=',')
padronByEdadSimplif.to_csv(r'padronByEdadSimplif.csv', sep=',')
padronByEdadDecanal.to_csv(r'padronByEdadDecanal.csv', sep=',')
```