

Data analyst case

Ventas y Clientes Enero 2020

01.



VTEX



Data

● Datos de ventas y clientes

Datos de los clientes

```
In [2]: 1 clientes = pd.read_csv('Clientes.csv', sep=';')  
        2 clientes.head()
```

Out[2]:

	Cliente	Nombre	Sexo	Zona
0	9001	Ana_1	F	Buenos Aires
1	9002	Juan_1	M	Capital Federal
2	9003	Pedro_1	M	Capital Federal
3	9004	Romina_1	F	Buenos Aires
4	9005	Ana_2	F	Buenos Aires

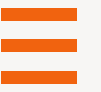
Datos de las ventas

```
In [3]: 1 ventas = pd.read_csv('Ventas.csv')  
        2 ventas.head()
```

Out[3]:

	Cliente	Fecha	Importe	Estado
0	9452	01/01/2020	51356	Completado
1	9567	01/01/2020	96994	Completado
2	10551	01/01/2020	33280	Completado
3	9320	01/01/2020	18045	Completado
4	9043	01/01/2020	2286	Completado

Procesamiento



● Procesamiento de los datos

Combinamos ambos datasets

```
In [4]: 1 data = ventas.merge(clientes, on='Cliente')
        2 data.head()
```

Out[4]:

	Cliente	Fecha	Importe	Estado	Nombre	Sexo	Zona
0	9452	01/01/2020	51356	Completado	Romina_113	F	Capital Federal
1	9452	23/01/2020	79480	Completado	Romina_113	F	Capital Federal
2	9452	01/02/2020	3701	Completado	Romina_113	F	Capital Federal
3	9452	24/03/2020	9367	Completado	Romina_113	F	Capital Federal
4	9452	12/07/2020	34220	Completado	Romina_113	F	Capital Federal

Pasamos la fecha a datetime

```
In [20]: 1 data['Fecha'] = pd.to_datetime(data['Fecha'], format='%d/%M/%Y', errors='coerce')
```

```
In [21]: 1 data.Fecha.dtypes
```

Out[21]: dtype('<M8[ns]')

Análisis

- Análisis inicial de los datos del dataset

Cantidad de compras que hizo cada cliente

```
In [56]: 1 data.Cliente.value_counts()
```

```
Out[56]: 9359      12
          9827      11
          9869      10
          9504      10
          9136      10
          ..
          9054       1
          10033       1
          9947       1
          10232       1
          9297       1
          Name: Cliente, Length: 1624, dtype: int64
```

Cantidad de clientes

```
In [54]: 1 data.Cliente.nunique()
```

```
Out[54]: 1624
```

Estado del viaje

```
In [55]: 1 data.Estado.value_counts()
```

```
Out[55]: Completado      6833
          En Viaje        27
          Name: Estado, dtype: int64
```

```
In [73]: 1 estado_enviaje = data[data['Estado'] == 'En Viaje']
          2 estado_enviaje
```

```
Out[73]:
```

	Cliente	Fecha	Importe	Estado	Nombre	Sexo	Zona
14	9567	2020-01-07 00:12:00	25752	En Viaje	Pedro_142	M	Capital Federal
66	9456	2020-01-07 00:12:00	54804	En Viaje	Romina_114	F	Capital Federal
1079	10260	2020-01-07 00:12:00	57856	En Viaje	Romina_315	F	Buenos Aires
1152	10336	2020-01-07 00:12:00	75090	En Viaje	Romina_334	F	Capital Federal
1257	10575	2020-01-08 00:12:00	44888	En Viaje	Pedro_394	M	Buenos Aires

Algunas métricas interesantes



Gasto promedio por cliente

Gasto promedio por cliente por mes

```
In [64]: 1 (data.Importe.sum()) / (data.Cliente.nunique())  
Out[64]: 213519.6379310345
```



Compras promedio por cliente

Compras promedio por cliente por mes

```
In [57]: 1 (data.Cliente.value_counts().sum()) / (data.Cliente.nunique())  
Out[57]: 4.224137931034483
```



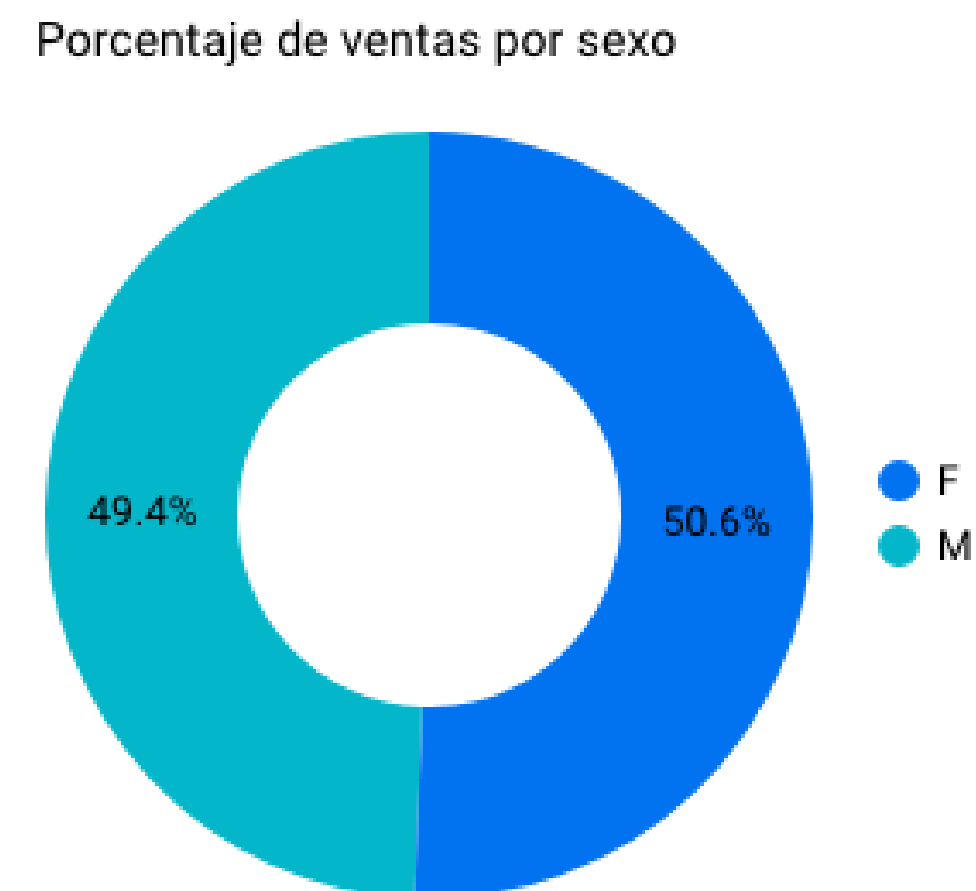
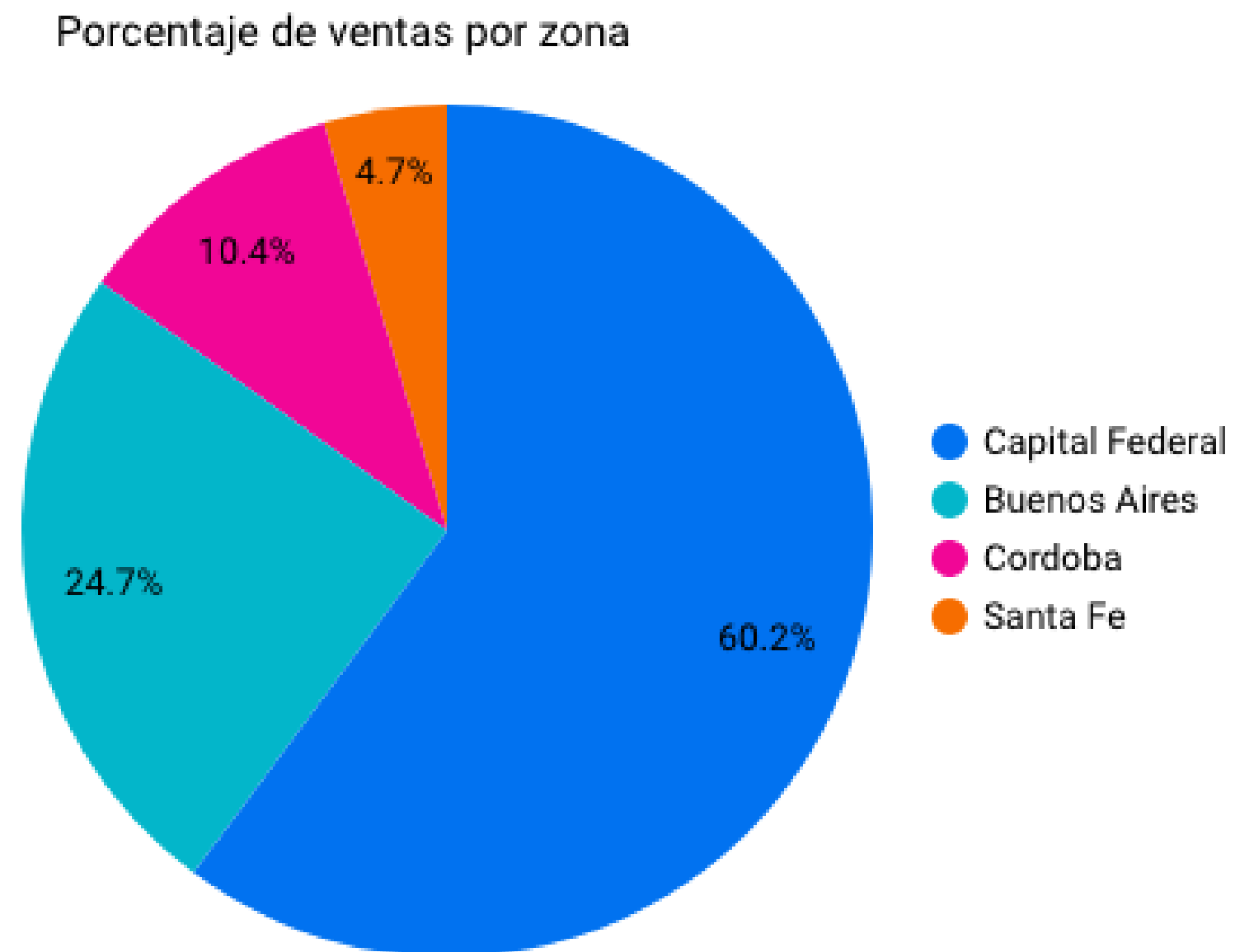
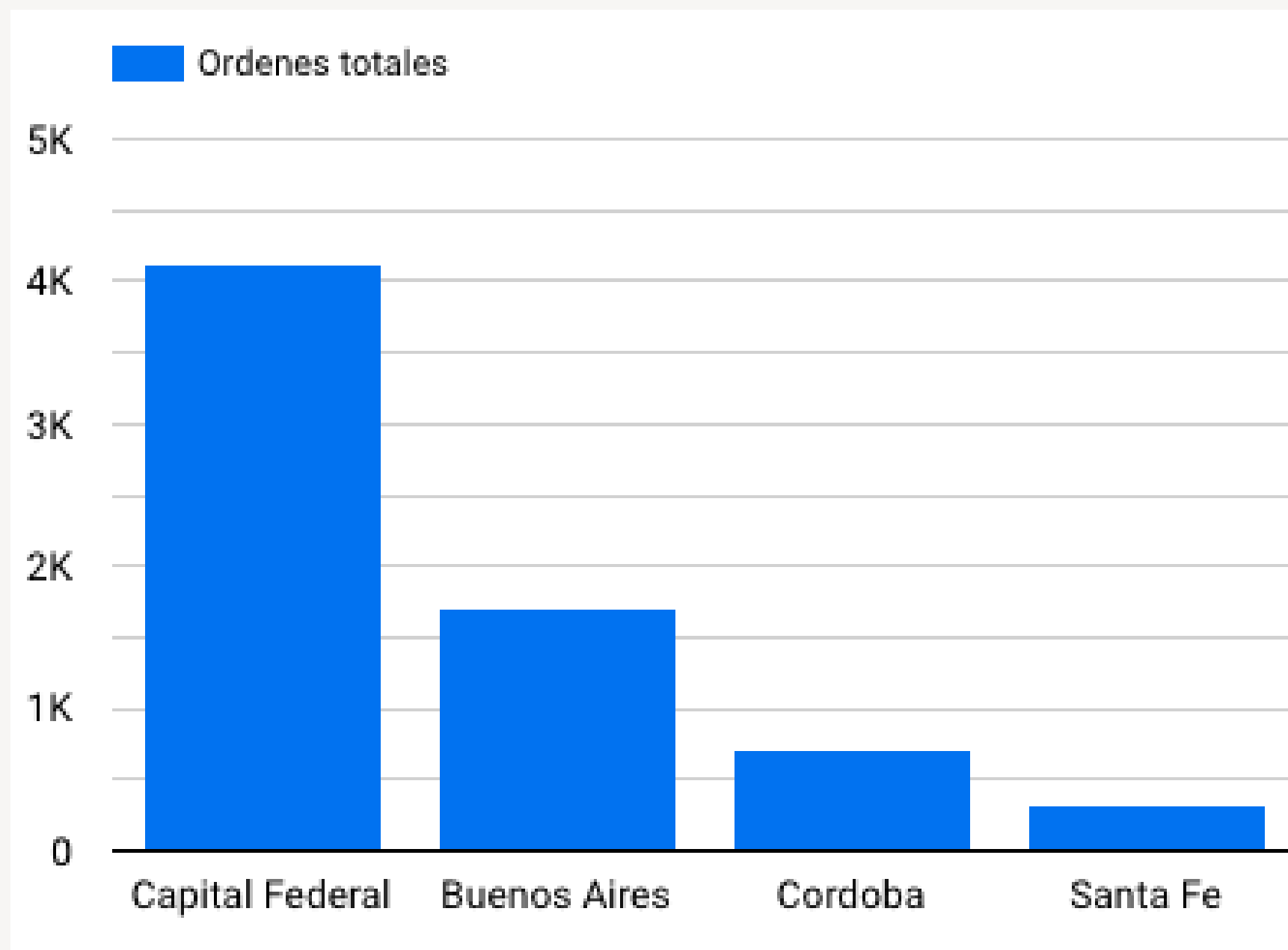
Cientes con mayores gastos

Cientes que más gastaron y el porcentaje del total de ventas mensuales

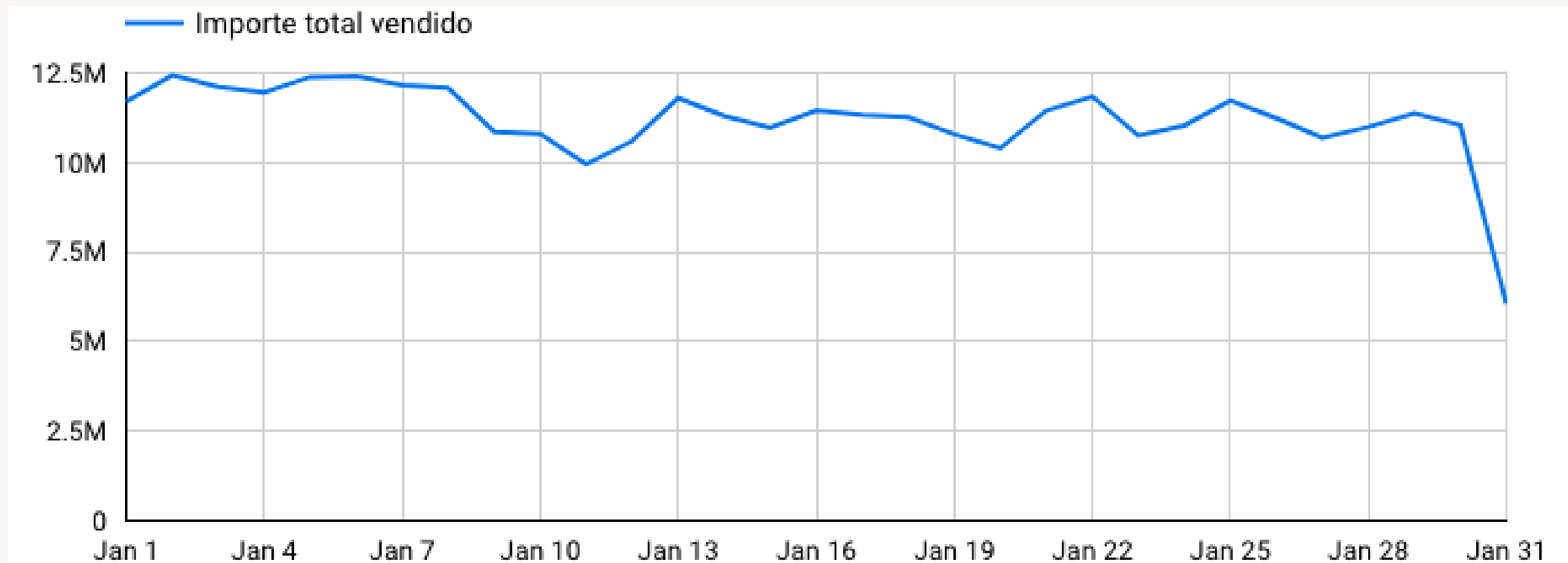
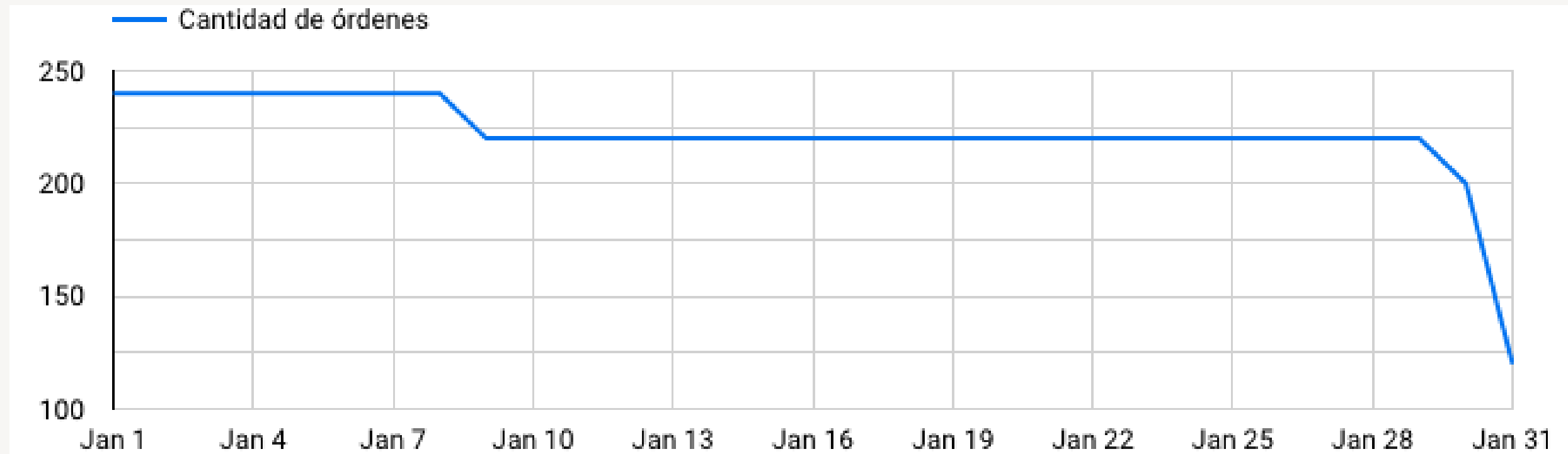
```
In [46]: 1 clientes_caros = pd.DataFrame(data.Importe.groupby(by=data['Cliente']).sum())  
         2 clientes_caros = clientes_caros.sort_values(by='Importe', ascending=False)  
  
In [48]: 1 clientes_caros['Porcentaje del total'] = (clientes_caros['Importe']*100)/(clientes_caros['Importe'].sum())  
  
In [58]: 1 clientes_caros.head(20)  
Out[58]:
```

	Importe	Porcentaje del total
Cliente		
9869	624944	0.180226
10377	614929	0.177338
10000	607107	0.175168

Visualizaciones



Visualizaciones



Visualizaciones



Cientes que más gastaron en todo el mes

	Cliente	Zona	Sexo	Órdenes totales	Importe ▾
1.	9869	Capital Federal	F	10	624,944
2.	10377	Santa Fe	F	10	614,929
3.	10373	Buenos Aires	F	10	607,407
4.	9576	Capital Federal	F	10	593,188
5.	10168	Capital Federal	F	9	586,896
6.	9625	Santa Fe	F	9	555,205
7.	10443	Capital Federal	M	9	553,345
8.	9590	Capital Federal	M	6	548,018
9.	9023	Santa Fe	M	9	531,153
10.	9827	Capital Federal	M	11	530,030
1 - 100 / 1624					< >



Conclusiones

09.

- Mayor cantidad de compras en Capital Federal, seguido de Buenos Aires, luego Córdoba y finalmente Sante Fé.
- Compras distribuidas en ambos sexos de forma similar
- Cantidad de órdenes constante a lo largo del mes. Se observa una fuerte caída el último día del mes.
- Hubo algún inconveniente los días 06, 07 y 08 de enero. Algunas órdenes realizadas a las 00:12 no pudieron completarse y figuran 'En viaje'.

