



# Productos de supermercados

---

Dossier técnico

## Descripción

Productos con su precio y descripción en los **principales supermercados de España**.

Las características de este dataset son las siguientes:

- Frecuencia de actualización: actualizado cada 12h
- Volumen estimado: 50.000 registros cada día
- Histórico: desde julio de 2020

El dataset completo se puede adquirir en [DataMarket](#), plataforma de referencia de datos externos en España.

## Muestra

La muestra contiene 500.000 filas, las cuales se han seleccionado desde mitad de julio de 2020 hasta principios de febrero de 2021. Se puede acceder a dicha muestra en el enlace mostrado a continuación:

<https://datamarket.es/media/samples/productos-de-supermercados-sample.csv>

## Columnas

A continuación se muestran las columnas de las que consta el dataset:

nombre	tipo	descripción	ejemplo
category	str	Categoría del producto.	marisco_y_pescado_pescado_congelado
description	str	Información adicional del producto (formato de empaquetado, etc.).	Pack-2
insert_date	datetime	Fecha de extracción de la información.	2020-09-23 14:15:00
name	str	Nombre del producto.	Anguriñas de surimi Pescanova
price	float	Precio absoluto del producto. En caso de existir algún tipo de descuento aparecerá el menor precio disponible.	1.99
reference_price	float	Precio unitario (por unidad de medida del producto, €/Kg, €/L, etc.).	7.96
reference_unit	str	Unidad de referencia del producto (Kg, L, etc.).	kg
supermarket	str	Supermercado al que pertenece el producto.	433463D95980C252B92C204E3655BB81

## Casos de uso

En esta sección se describen cinco casos de uso reales para este dataset. Estos casos de uso se estructuran en forma de cinco preguntas que se pueden responder con el mismo.


### | Variedad total de productos por supermercado, desde el inicio del histórico

#### Descripción

Número de productos diferentes por supermercado desde el inicio del histórico de datos.

#### Solución

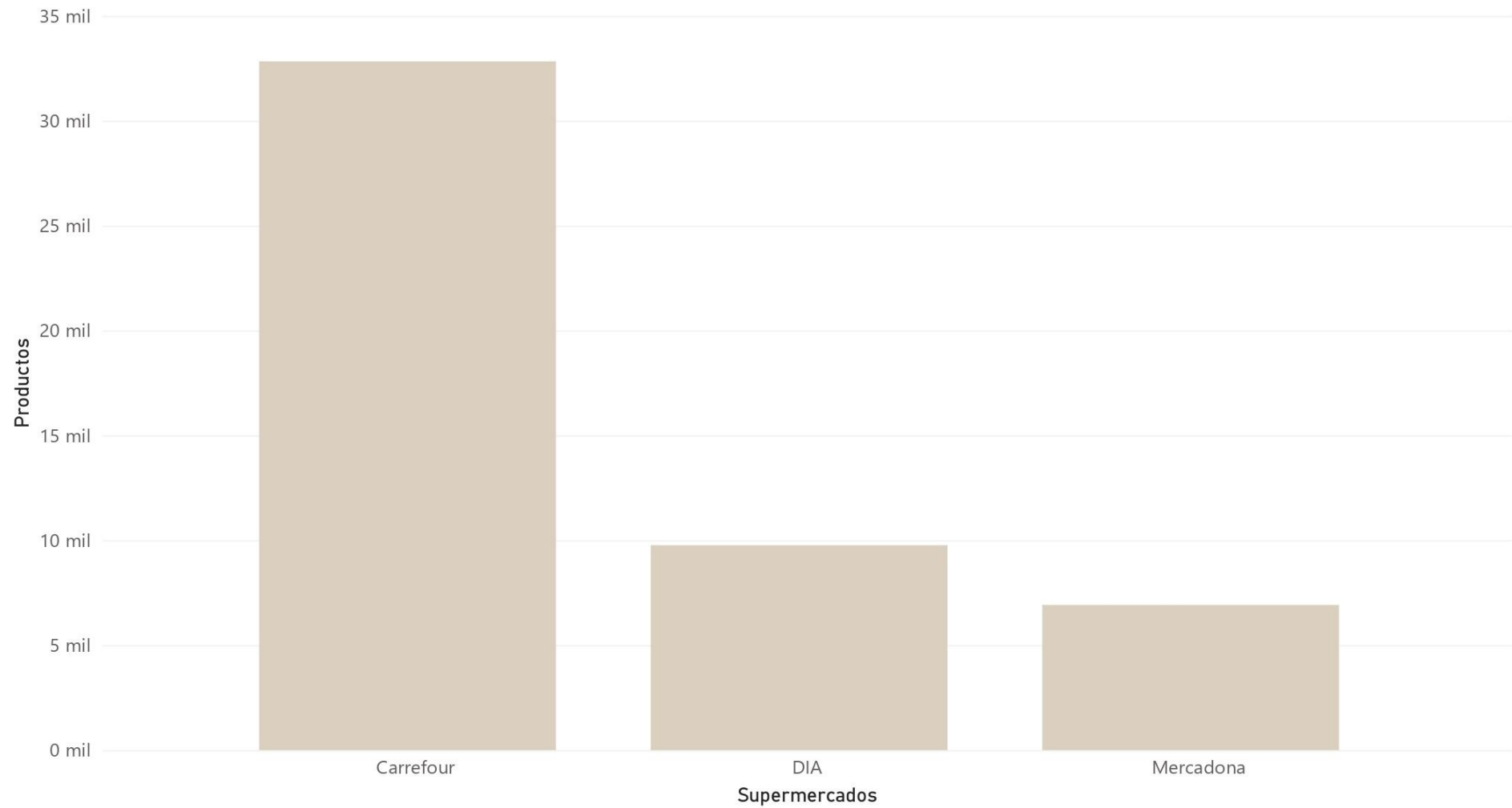
Consulta SQL



```
select
    supermarket,
    count(url) as n_products
from
    productos_de_supermercados
group by
    supermarket
```

## Gráfico

## Variedad de productos por supermercados



## | Histórico de precios de las mascarillas higiénicas en los diferentes supermercados

### Descripción

Evolución del precio de las mascarillas quirúrgicas en los principales supermercados de España.

### Solución

## Consulta SQL



```
with surgical_mask_price_dia as (  
  select  
    insert_date as date_dia,  
    reference_price as masks_dia  
  from  
    productos_de_supermercados  
  where  
    name = 'MASK4U mascarilla quirúrgica de protección 3 capas pack de 10 unidades'  
    and supermarket = 'dia-es'  
),  
surgical_mask_price_mercadona as (  
  select  
    insert_date as date_mercadona,  
    reference_price as masks_mercadona  
  from  
    productos_de_supermercados  
  where  
    name = 'Mascarillas higiénicas no reutilizables Deliplus'  
    and supermarket = 'mercadona-es'  
  order by  
    insert_date  
)
```



```
surgical_mask_price_carrefour as (  
  select  
    insert_date as date_carrefour,  
    reference_price as masks_carrefour  
  from  
    productos_de_supermercados  
  where  
    name = 'Mascarilla facial higiénica no reutilizable colores adulto Bergner 10 ud.'  
    and supermarket = 'carrefour-es'  
  order by  
    insert_date desc  
)  
select  
  date(insert_date) as date,  
  masks_dia,  
  masks_mercadona,  
  masks_carrefour  
from  
  surgical_mask_price_dia  
join surgical_mask_price_mercadona on  
  surgical_mask_price_mercadona.date_mercadona = surgical_mask_price_carrefour.date_dia  
join surgical_mask_price_carrefour on  
  surgical_mask_price_dia.date_carrefour = surgical_mask_price_carrefour.date_dia  
join productos_de_supermercados on  
  productos_de_supermercados.insert_date = surgical_mask_price_carrefour.date_dia
```

Gráfico

Evolución histórica del precio unitario de las mascarillas quirúrgicas por supermercado





| ¿Cuáles son los 10 productos de la marca Hacendado más baratos y más caros?

#### Descripción

Los 10 productos más caros y más baratos por precio unitario de la marca Hacendado más recientes.

#### Solución

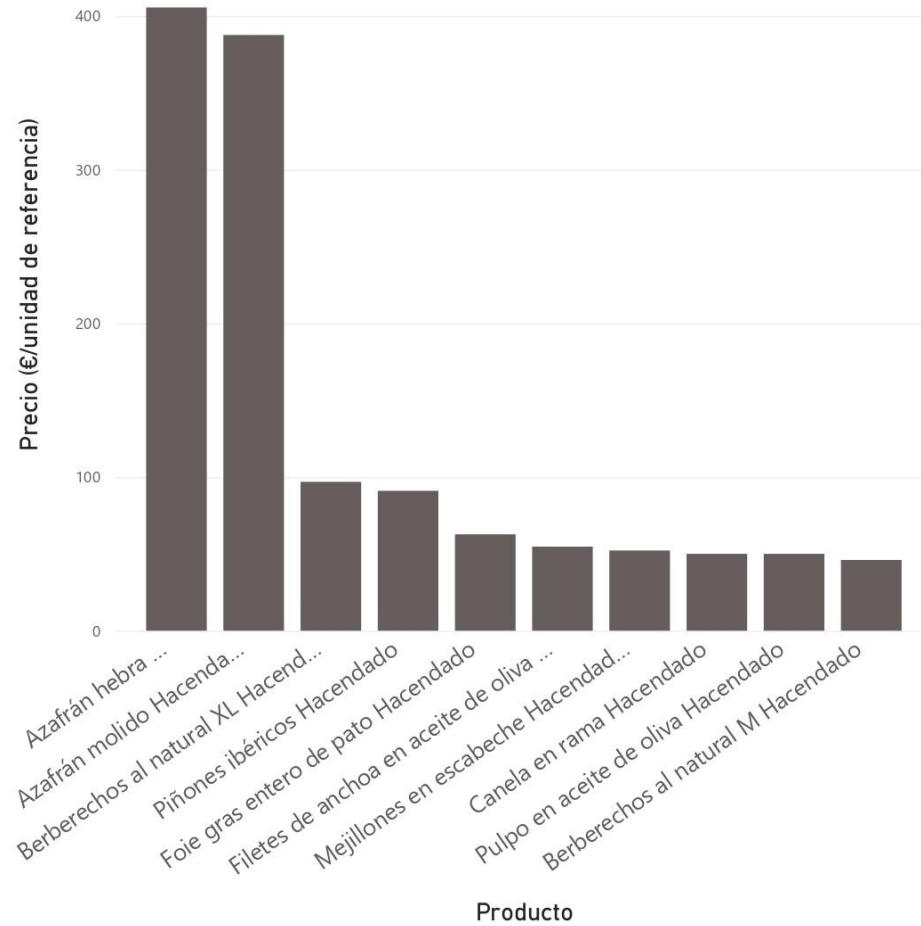
## Consulta SQL

```
with last_appearance_date as(  
  select  
    max(insert_date) as last_date, url, name  
  from  
    coches_de_segunda_mano  
  where  
    lower(name) like '%hacendado%' and url is not null  
  group by  
    url, name  
)  
,  
top_10_product_cheap as (  
  select  
    last_appearance_date.*, reference_price  
  from  
    last_appearance_date  
  join coches_de_segunda_mano on  
    coches_de_segunda_mano.insert_date = last_appearance_date.last_date  
    and coches_de_segunda_mano.url = last_appearance_date.url  
  order by  
    reference_price  
  limit 10  
)
```

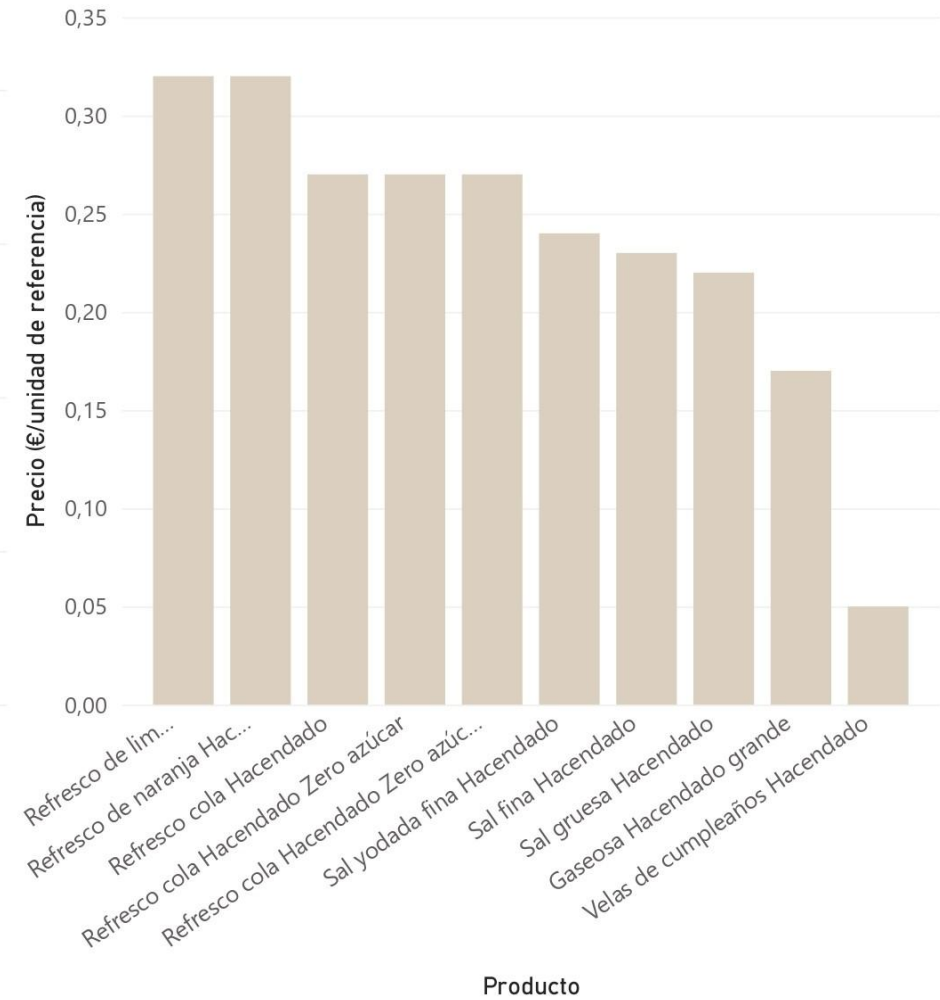
```
top_10_product_expensive as (  
  select  
    last_appearance_date.*, reference_price  
  from  
    last_appearance_date  
  join coches_de_segunda_mano on  
    coches_de_segunda_mano.insert_date = last_appearance_date.last_date  
    and coches_de_segunda_mano.url = last_appearance_date.url  
  order by  
    reference_price desc  
  limit 10  
)  
select  
  last_date, url, name, reference_price  
from  
  top_10_product_cheap  
union  
select  
  last_date, url, name, reference_price  
from  
  top_10_product_expensive
```

## Gráfico

Top 10 productos de la marca Hacendado más caros



Top 10 productos de la marca Hacendado más baratos



| Productos con mayores cambios de precios (en %) en las últimas N semanas de Mercadona, top 10 que más suben y top 10 que más bajan

#### Descripción

Los 10 productos que más han subido y los 10 que más han bajado, en porcentaje, en las últimas N semanas, para este ejemplo se ha tomado una semana como referencia.

#### Solución

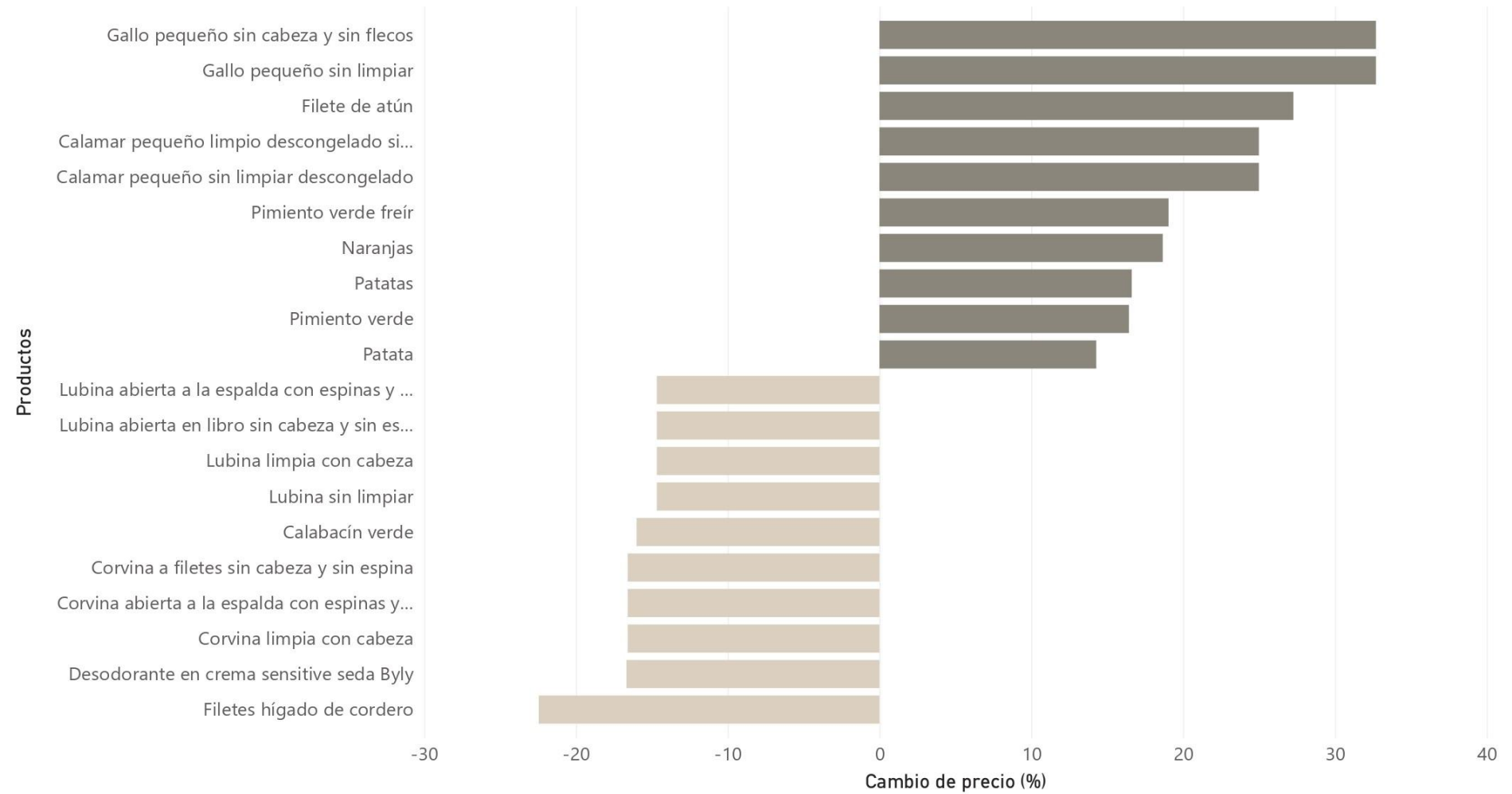
## Consulta SQL

```
with last_price_product as (  
  select  
    url, name, price as last_price  
  from  
    coches_de_segunda_mano  
  where  
    insert_date = (select max(insert_date) from coches_de_segunda_mano) and supermarket = 'mercadona-es'  
),  
old_price_product as (  
  select  
    url, name, price as old_price  
  from  
    coches_de_segunda_mano  
  where  
    insert_date = (select max(insert_date) - interval '1 week' from coches_de_segunda_mano)  
    and supermarket = 'mercadona-es'  
),  
top_10_products_most_fall as (  
  select  
    old_price_product.name, last_price, old_price,  
    (last_price - old_price) / old_price * 100 as percentage_diff_price  
  from  
    old_price_product  
  join last_price_product on  
    last_price_product.url = old_price_product.url  
  order by  
    percentage_diff_price desc  
  limit 10  
)
```

```
top_10_products_most_up as (  
  select  
    old_price_product.name, last_price, old_price,  
    (last_price - old_price) / old_price * 100 as percentage_diff_price  
  from  
    old_price_product  
  join last_price_product on last_price_product.url = old_price_product.url  
  order by  
    percentage_diff_price  
  limit 10  
)  
select * from top_10_products_most_up  
union  
select * from top_10_products_most_fall  
where  
  percentage_diff_price != 0  
order by  
  percentage_diff_price
```

Gráfico

Productos con mayores cambios de precios del supermercado Mercadona





## | Predicción (serie temporal) del precio de las naranjas

### Descripción

Predecir el precio que tendrán las naranjas en un período determinado de tiempo. Se ha cogido el supermercado Mercadona como referencia para la realización de este modelo predictivo.

### Solución

#### Modelo predictivo simple

```
import pandas as pd
import numpy as np
import matplotlib.pyplot as plt
from fbprophet import Prophet

#Cargamos el dataset
productos_df = pd.read_csv('productos_de_supermercados.csv')

#Cogemos las naranjas y las agrupamos los diferentes tipos diariamente
productos_df = productos_df.loc[productos_df.name.str.contains('Naranja') &
productos_df.category.str.contains('fruta')]
productos_df = productos_df.loc[productos_df.supermarket == 'mercadona-es']
productos_df.loc[:, 'insert_date'] = pd.to_datetime(productos_df.loc[:, 'insert_date'], dayfirst=True)
productos_df = productos_df.groupby([pd.Grouper(key='insert_date', freq='D')]).mean()
productos_df = productos_df.reset_index()[['insert_date', 'reference_price']]
productos_df.columns = ['ds', 'y']
```



```
#Cargamos el algoritmo y predecimos 15 días en el futuro
pt = Prophet(yearly_seasonality=False, daily_seasonality=False)
pt.fit(productos_df)
future = pt.make_future_dataframe(periods=15)
forecast = pt.predict(future)[['ds', 'yhat']]

#Representamos los resultados
fig, ax = plt.subplots(ncols = 1, nrows= 1, figsize=(20,8))
ax.plot(productos_df.ds, productos_df.y, c='#675e5e')
ax.plot(forecast.ds, forecast.yhat, c='#c0b283')
ax.legend(['Precio real', 'Precio predicho'], prop={'size' : 12})
ax.set_xlabel('Fecha', fontdict={'size' : 14})
ax.set_ylabel('Precio medio naranjas (€/Kg)', fontdict={'size' : 14})
ax.set_title('Evolución precio naranjas Mercadona', fontdict={'size' : 20}, loc='left')
ax.spines['right'].set_visible(False)
ax.spines['top'].set_visible(False)
plt.setp(ax.get_xticklabels(), rotation=45, ha="right", rotation_mode="anchor")
ax.tick_params(axis='both', which='major', labelsize=12)
plt.tight_layout()
plt.show()
```

## Gráfico

