

Análisis de modificación de Precios en Surtidores

Marcos Ingani

Francisco Chedufau

Alexander Lopez

Abstract

En este análisis se busca evaluar el comportamiento de precios de productos en surtidores según las variables que presenta el dataset y la correlación del precio respecto a estas y la cotización del dólar.

1. Introducción

En este trabajo se abordará un dataset acerca de la modificación y actualización de precios del gobierno de la nación, con herramientas de Exploratory Data Analysis y Machine Learning.

El dataset utilizado se encuentra disponible en la página (Link: <https://datos.gob.ar/dataset/energia-precios-surtidor---resolucion-3142016>) del Ministerio de Modernización.

“Datos del Sistema en línea de información de precios de combustibles en surtidor, a través del cual los titulares de bocas de expendio de combustibles líquidos inscriptos en el registro creado por la resolución n° 1.102/2004 presentan la información relativa a precios de comercialización minorista dentro de las 8 (ocho) horas de producida una modificación en el precio en surtidor.”

2. Dataset

En el dataset cómo ya se mencionó anteriormente, se encuentran las actualizaciones de precios por surtidor, informándose en las 8 hs posteriores a la modificación por parte de las empresas.

Los campos para actualizar son:

```
'indice_tiempo', 'idempresa',  
'cuit', 'empresa',  
'direccion', 'localidad',  
'provincia', 'region',  
'idproducto', 'producto',  
'idtipohorario',  
'tipohorario', 'precio',  
'fecha_vigencia',  
'idempresabandera',  
'empresabandera', 'latitud',  
'longitud', 'geojson'
```

Al cual le sumamos un dataset de la cotización del dólar respecto de cada fecha.

Fuente:

<https://es.investing.com/currencies/usd-ars-historical-data>

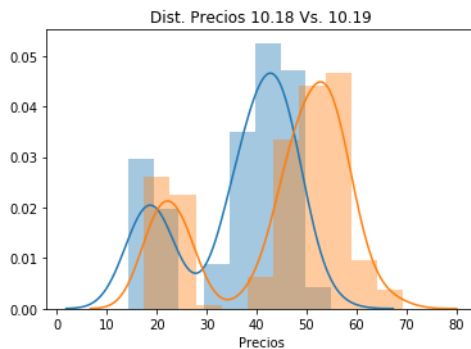
El cual presenta los campos: Fecha, Último, Máximo, Mínimo, %Var.

Para poder analizar la correlación entre los precios promedios de la nafta en este caso con la cotización promedio del dólar a la fecha.

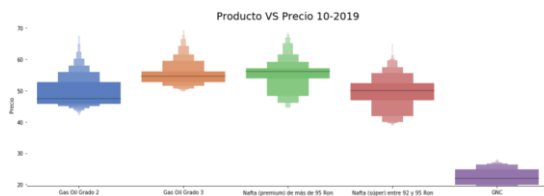
3. Exploratory Data Analysis

Respecto del análisis de datos logramos visualizar el comportamiento del precio de todos los tipos de productos en venta, su distribución y aumento a través del tiempo.

En el caso de distribución se pudo visualizar la comparación del avance de precios en un año:



También la media y la distribución del precio respecto de cada producto:



Finalmente, con la cotización del dólar se analizó la correlación entre el precio promedio de la nafta y la cotización del dólar promedio por mes, la cual resultó siendo de 0.97.

4 Materiales y métodos

Respecto a los materiales, fueron nombrados anteriormente, se utilizó la agrupación por fecha, producto, provincia, precio promedio y cotización promedio del dólar a la fecha.

4.1 Regresión

Los métodos utilizados para la regresión son:

1. Regresión Lineal
2. KNN Regression
3. Support Vector Regression (SVR)

Se calcularon los errores, MAE y MSE, y resultados respecto a cada modelo.

$$MAE = \frac{|\sum (\hat{y}_t - y_t)|}{n}$$

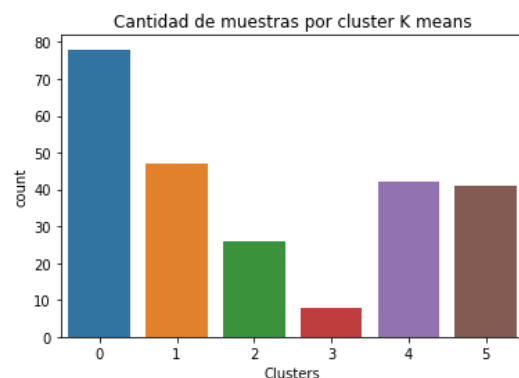
$$MSE = \frac{\sum (\hat{y}_t - y_t)^2}{n}$$

	KNN	SVR	Regresión Lineal
RMSE	8,084	4,252	4,133
MSE	17,085	18,084	17,085
MAE	3,285	3,297	3,53
Score	0,606	0,892	

Por lo que se recomienda utilizar de acuerdo con los parámetros seleccionados el modelo SVR.

4.2 Clustering

Para el método de Clustering se utilizó el modelo de clasificación KNN generando 6 clusters del producto GNC en el año 2017. Siendo 242 muestras y obteniendo una clasificación:



El silhouette score, que mide la similaridad (a) entre muestras de cada clase y la disimilaridad (b) respecto de otras clases, resultó ser de 0.645, siendo 1 la máxima.

$$a(x_i) = \frac{1}{n_k - 1} \sum_{x_j \in C_k, x_j \neq x_i} d(x_i, x_j)$$

$$b(x_i) = \min_{v=1, \dots, K, v \neq k} \left[\frac{1}{n_v} \sum_{x_j \in C_v} d(x_i, x_j) \right]$$

$$S(x_i) = \frac{b(x_i) - a(x_i)}{\max[b(x_i), a(x_i)]} \quad S_X = \frac{1}{K} \sum_{k=1}^K \left[\frac{1}{n_k} \sum_{x \in C_k} S(x_i) \right]$$

5 Discusión y Conclusiones

Respecto al manejo de datos, se podría generar un dataset respecto de los precios vigentes por fecha de cada producto y empresa por provincia/región. De esta manera se lograrían hacer comparaciones y modelos de series de tiempos con mayor valor de predicción a futuro para dar utilidad ya sea a particulares o a personas jurídicas, para una planificación de costos logísticos o cualquiera sea la utilidad del producto en cuestión. Teniendo en cuenta la volatilidad de la cotización del dólar en este país y su traslado al precio de productos respaldados por el valor de este.

Respecto a los modelos se utilizaron modelos utilizados en la cursada, siendo limitado a los recursos dados por la catedra y con la apertura a poder desarrollar mejores predicciones de precios con otros parámetros u otros modelos.

Las herramientas utilizadas son una manera de poder demostrar tendencias e información que no se alcanza a simple vista y con la aplicación de estas herramientas y modelos, logran ser de gran utilidad para analizar el contexto de los precios de surtidor.

Se puede comparar las actualizaciones de precios respecto a la legislación Nacional que la regula y poder analizar de manera más intrínseca los cambios de precios.

Se tomó el valor del dólar como un recurso que impacta de manera directa el precio y su variación se traslada al corto plazo. No siendo así el precio de barril el cual está sujeto a relaciones políticas comerciales internacionales las cuales no se puede transferir su variabilidad de manera directa al precio de los productos abordados a nivel país.

5 Referencias

1. Fuentes:

Ministerio de Modernización [Precios en Surtidor](#).

Conclusiones

En este trabajo se evaluó la capacidad de predecir el valor de los productos mediante distintos modelos teniendo en cuenta variables como cotización del dólar, provincia y tipo de producto, alcanzando niveles significativos de accuracy en el caso de Regresión.

Respecto del modelo de clustering se logró clasificar en 6 grupos con alta similaridad entre muestras de mismos clusters y disimilaridad respecto de otros, obteniendo así parámetros aceptables de clasificación.

Investing.com [dólar historico](#).

2. Paper petroleum product prices [Petroleum products](#).
3. The effects of fuel Price [Fuel price effects](#).