Actividad: Análisis exploratorio de datos y tratamiento de series temporales.

La base de datos Avocado prices (<https://www.kaggle.com/timmate/avocado-prices-2020>) contiene datos del precio de distintos tipos de aguacate y del mercado donde se vende en Estados Unidos. Más concretamente, esta base está compuesta por un total de 13 variables:

* Fecha (date). En frecuencia semanal desde inicios de 2015.
* Precio por unidad (average\_price).
* Volumen de ventas (total\_volume).
* Código de producto (4046,42254770).
* Formato de ventas (total\_bags, small\_bags,large\_bags,xlarge\_bags).
* Tipo, orgánico o conventional (type)
* Año (year)
* Lugar de ventas (geography)

En un intento de aplicar las nuevas técnicas o herramientas analíticas frente a la analítica tradicional de negocios, nos ha encargado realizar un pequeño análisis de dicha base de datos, así como la realización de una serie de ejercicios para explorar la información existente en este conjunto de datos:

* ¿De qué tipo de variables se compone mi base de datos? Realice un análisis exploratorio de las variables numéricas (media, varianza, diagrama de cajas, etc) ¿Qué conclusiones se pueden extraer de la muestra de datos?
* Extraiga de la base de datos el precio de venta (la variable) de los aguacates orgánicos vendidos en Albany y de Boston.
* Como paso previo al modelado, calcule la covarianza y la matriz de correlación de del precio de los aguacates orgánicos, convencionales y su volumen de ventas. ¿Qué conclusiones se pueden extraer?
* Determine la posible relación existente entre dichos precios y su volumen de ventas. Si tomáramos logaritmos, ¿Cómo sería dicha relación?
* Realice una predicción de precio de venta de los aguacates orgánicos vendidos en Albany a 3 meses.

**Rúbrica:**

|  |  |  |  |
| --- | --- | --- | --- |
|  | Descripción | Puntuación máxima  (puntos) | Peso  % |
| Criterio 1 | Es capaz de identificar los distintos tipos de variables. Aporta conclusiones válidas de los estadísticos descriptivos. | 2 | 20% |
| Criterio 2 | Es capaz de crear submuestras. | 2 | 20% |
| Criterio 3 | Aporta conclusiones razonables, aunque no haya obtenido la matriz de correlación correcta. | 2 | 20% |
| Criterio 4 | Es capaz de realizar un análisis de regresión y extraer información de dicho análisis. | 2 | 20% |
| Criterio 5 | Es capaz de realizar una predicción de una variable. | 2 | 20% |
|  |  | **10** | **100 %** |