

Nombre: Victor Miguel Barrera Peña
Fecha: 15/05/2024
Equipo: 10
Proyecto: 3.4_series_de_tiempo

Actividad

Realizar pronósticos de Ventas (importe) utilizando Series de tiempo:

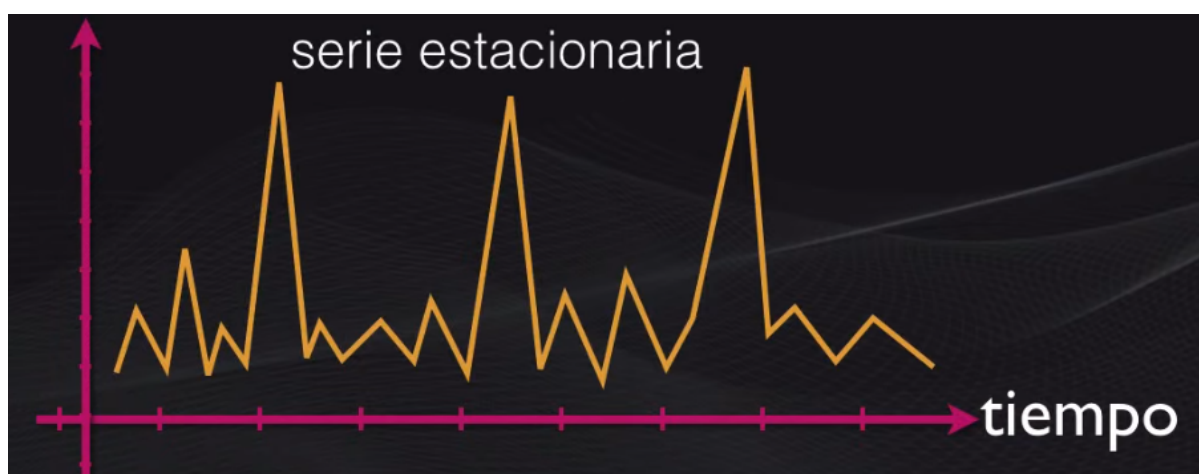
- Triple Exponential Smoothing y
- ARIMA

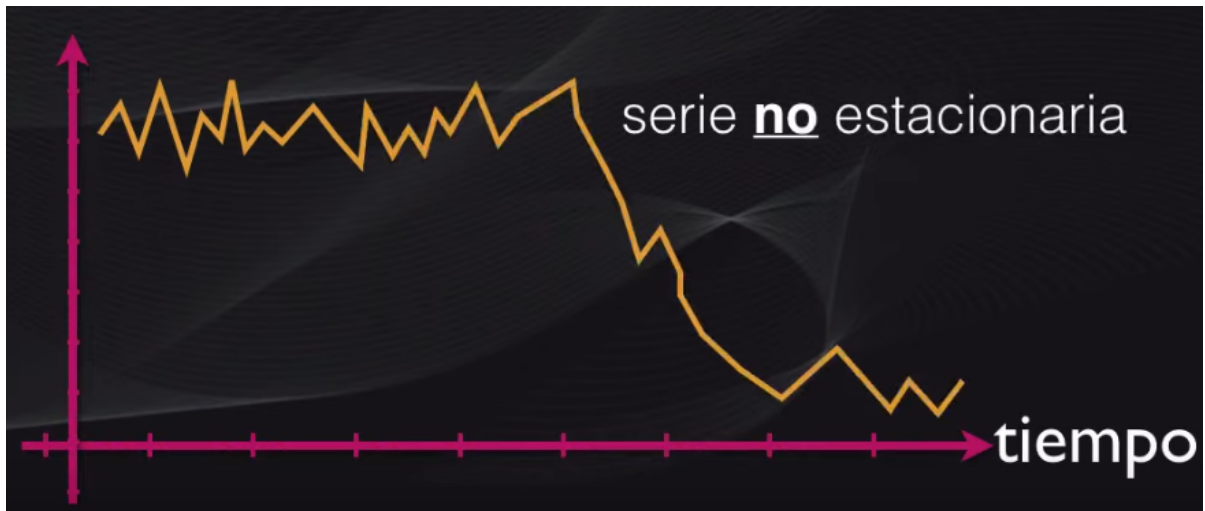
utilizando Python para los siguientes escenarios:

1. Seleccionar 2 variables objetivo que consideren entre los datos de: Sector, Marca, Línea, Sublínea o Producto.
2. Para cada variable seleccionada:
 1. Describir el caso de negocio, del porque se desea realizar un pronóstico
 2. En el caso de Triple Exponential Smoothing descomponer cada elemento como:
Tendencia, Ciclicidad, Estacionalidad y el componente Irregular
 3. Para el caso de ARIMA obtener ACF y PACF y Ruido Blanco
 4. Detectar valores atípicos (Outliers)
 5. Dar sus comentarios de los resultados obtenidos

Conceptos

Estacionalidad





Autocorrelación



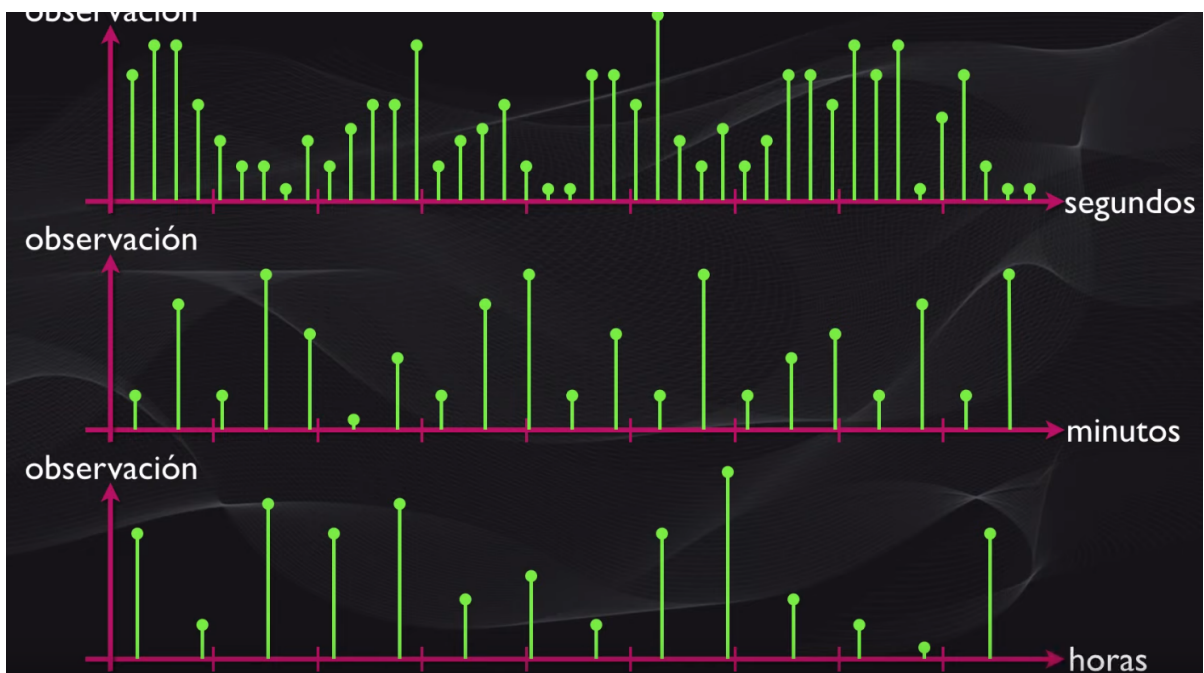
Datos Faltantes

- Conocidos como huecos
 - Tenemos dos métodos imputación y eliminación

Resolución temporal

- Tiene que tener un distanciamiento lineal uniforme (que tenga el mismo espacio en el tiempo los datos representados).

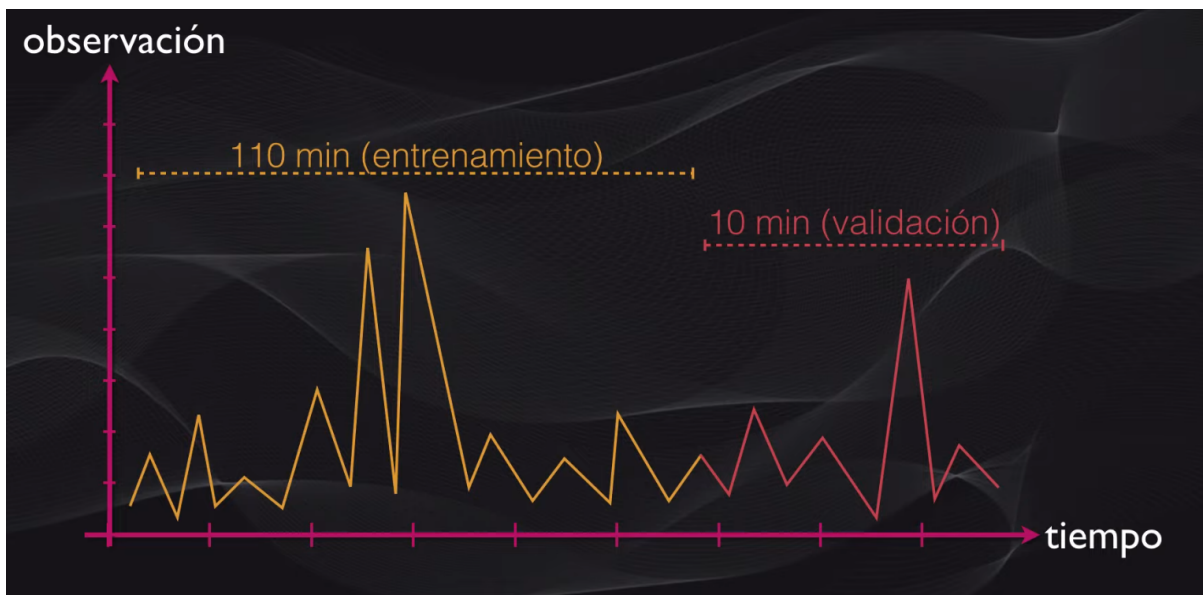
- Se puede usar técnicas de sub o sobre muestrear para ajustar.



Horizonte de tiempo

- Es decir desde cuando y hasta cuando se realiza el análisis.

Partición de los datos



Predictivo

- Basados en datos
 - ARIMA
 - Depende de la periodicidad
- Basados en modelos.

Solución

1.) Seleccionar 2 variables objetivo que consideren entre los datos de: Sector, Marca, Línea, Sublínea o Producto.

El CSV proporcionado contiene datos de transacciones de bebidas isotónicas. A continuación, te explico cada una de las variables incluidas en el archivo:

1. **Sector:** Sector al que pertenece el producto.
2. **Tipo:** Tipo de producto.
3. **Línea:** Línea de producto.
4. **Sublínea:** Subcategoría dentro de la línea de producto.
5. **Presentación:** Forma en que se presenta el producto (por ejemplo, individual).
6. **Gramaje:** Peso del producto.
7. **Empresa:** Empresa que produce o distribuye el producto.
8. **Año:** Año de la transacción.
9. **Fecha:** Fecha específica de la transacción.
10. **Cliente:** Identificación del cliente.
11. **Transacción:** Identificación de la transacción.
12. **Pedido:** Identificación del pedido.
13. **Control:** Código de control interno.
14. **Producto:** Identificación del producto.
15. **Descripción:** Descripción del producto.
16. **Unidad:** Unidad de medida (por ejemplo, pieza).
17. **Cantidad:** Cantidad de producto comprada.
18. **Precio:** Precio unitario del producto.
19. **Dummy_1:** Variable dummy (su propósito no está claro sin contexto adicional).
20. **Precio_Max:** Precio máximo registrado para el producto.
21. **Dif_PrecioMax:** Diferencia entre el precio actual y el precio máximo.
22. **Importe:** Importe total de la transacción.

23. **Clave_Cliente:** Identificación del cliente.
24. **No._Hijos:** Número de hijos del cliente.
25. **Antigüedad:** Antigüedad del cliente.
26. **Edad:** Edad del cliente.
27. **Edad_Rango:** Rango de edad del cliente.
28. **Escolaridad:** Nivel educativo del cliente.
29. **Estado_Civil:** Estado civil del cliente.
30. **Estado:** Estado de residencia del cliente.
31. **Sexo:** Género del cliente.
32. **Función:** Función o rol del cliente.
33. **Grupo:** Grupo al que pertenece el cliente.
34. **Fecha_2:** Segunda fecha (posiblemente repetida para conveniencia de análisis).
35. **Year:** Año (repetido).
36. **Quarter:** Trimestre del año.
37. **Month:** Mes.
38. **Day:** Día.
39. **Dia_Num_Sem:** Número del día en la semana.
40. **Dia_Semana:** Día de la semana.

Para seleccionar las variables objetivo en una serie de tiempo, es importante elegir aquellas que puedan ser predictoras de la evolución temporal del fenómeno que se quiere estudiar. En este caso, seleccionaré dos variables objetivo:

1. **Precio:** Es crucial en series de tiempo ya que puede mostrar tendencias, estacionalidades y fluctuaciones en los precios de los productos a lo largo del tiempo.
2. **Cantidad:** La cantidad de productos vendidos puede proporcionar información sobre la demanda y su variabilidad a lo largo del tiempo, permitiendo identificar patrones de compra y estacionalidades.

Estas dos variables son fundamentales para entender el comportamiento del mercado y las ventas a lo largo del tiempo. Analizar cómo varían estas variables puede ayudar a predecir futuras ventas y ajustar estrategias de precios y marketing.

Eliminar redundancia de los datos

Basándome en la estructura de las columnas y en su potencial redundancia, he identificado algunas columnas que podrían ser redundantes o repetitivas. A continuación, enumero las posibles columnas redundantes y la razón de su redundancia:

1. **Fecha** y **Fecha_2**: Ambas columnas parecen contener la misma información, ya que en las primeras filas los valores son iguales.
2. **Año** y **Year**: Ambas columnas contienen el año de la transacción.
3. **Cliente** y **Clave_Cliente**: Ambas columnas parecen contener información sobre la identificación del cliente.
4. **Edad_Rango**: Podría derivarse de la columna **Edad**.
5. **Precio_Max** y **Dif_PrecioMax**: Aunque no son directamente redundantes, **Dif_PrecioMax** puede calcularse a partir de **Precio** y **Precio_Max**.
6. **Month**, **Quarter**, **Day**, **Dia_Num_Sem**, **Dia_Semana**: Estas columnas pueden derivarse de la columna **Fecha**.

Voy a realizar una comprobación adicional para confirmar la redundancia de estas columnas en el conjunto de datos proporcionado.

El análisis de redundancia de las variables:

1. **Fecha** vs **Fecha_2**: Redundantes. Ambas columnas son iguales.
2. **Año** vs **Year**: Redundantes. Ambas columnas son iguales.
3. **Cliente** vs **Clave_Cliente**: No redundantes. Las columnas no son iguales.
4. **Edad_Rango**: Redundante. Puede derivarse de la columna **Edad**.
5. **Month**: Redundante. Puede derivarse de la columna **Fecha**.
6. **Quarter**: Redundante. Puede derivarse de la columna **Fecha**.
7. **Day**: Redundante. Puede derivarse de la columna **Fecha**.
8. **Dia_Num_Sem**: Redundante. Puede derivarse de la columna **Fecha**.
9. **Dia_Semana**: No redundante. Las columnas no son iguales (posiblemente debido a diferencias en la configuración regional).

Con base en estos resultados, las siguientes columnas pueden considerarse redundantes y potencialmente eliminables:

- Fecha_2
- Year
- Edad_Rango
- Month
- Quarter
- Day
- Dia_Num_Sem

Paso 2: convertir a variables numéricas aquellas que sean ya categóricas, o derivadas de otras y eliminando algunos valores.

2.1)) En el caso de Triple Exponential Smoothing descomponer cada elemento como:

- Tendencia, Ciclicidad, Estacionalidad y el componente Irregular

Esta implementado en código.

2.2)) Para el caso de ARIMA obtener ACF y PACF y Ruido Blanco

Esta implementado en código.

2.3)) Detectar valores atípicos (Outliers)

Esta implementado en código.

2.4)) Dar sus comentarios de los resultados obtenidos

Los resultados obtenidos fue correcto a lo que se necesitaba predecir para cada producto, ya que en el tiempo lo que se busca es predecir el precio o cantidad de un producto.

- Los errores no son grandes con respecto a los valores que se tienen, el error esta alrededor de 8% con error cuadrático y por ello puedo decir que esta perfectamente, sin embargo pudo realizarse más adecuaciones de las que se tenían, por ejemplo una campana que comprueba la regularidad necesaria para las series de tiempo, pero ello todavía no puedo expresar

Referencias

- Shakeel Ahmed. (2020, 20 noviembre). *SPSS LEARNING TUTORIAL 25: TIME SERIES ANALYSIS (EXPONENTIAL SMOOTHING AND ARIMA MODELS)* [Vídeo]. YouTube. <https://www.youtube.com/watch?v=CXPuOxh5Ozw>