

Reporte de Análisis Inicial

Inteligencia Artificial Avanzada para la Ciencia de Datos II (Gpo 101)

Eryk Elizondo González - A01284899

Jacobo Hirsch Rodríguez - A00829679

Juan Pablo Bernal Lafarga - A01742342

Cleber Gerardo Pérez Galicia - A01236390

1. Planteamiento del Problema

Arca Continental quiere conocer el desempeño de nuevos productos con sus clientes históricos, tomando en consideración un historial de ventas de años anteriores.

2. Pregunta de Investigación

¿Cómo se comportan las ventas de galones totales de Arca Continental a lo largo de los meses del año, en los últimos 3 años?

3. Hipótesis

La suma de las ventas de galones mensuales presentan una variación estacional cada año. Además, en los meses de verano se venden más galones en el año, y en los meses de invierno se venden menos galones.

4. Metodología

Analizando los datos disponibles, se encontró que el registro de ventas en el año del 2019 está incompleto, ya que solo incluye ventas del mes de septiembre en adelante. Con el objetivo de no sesgar el análisis, se eliminaron los registros de este año, dejando únicamente aquellas ventas del año 2020 a 2022.

En búsqueda de la estacionalidad propuesta, se realizó un agrupamiento de las ventas acumuladas por mes en el paso de 3 años. De manera que si las ventas en enero de los años 2020, 2021 y 2022 fueron 50000, 60000, 70000 registros cada uno, entonces el “Total de ventas en Enero” sería un total de 180 mil registros de venta.

Finalmente, validamos la estacionalidad de las ventas con la prueba de Dickey-Fuller Aumentada (ADF Test), la cual verifica si una serie de tiempo es estacionaria evaluando si existe una raíz unitaria en la serie de tiempo. La aplicación de la prueba llevó a los siguientes resultados:

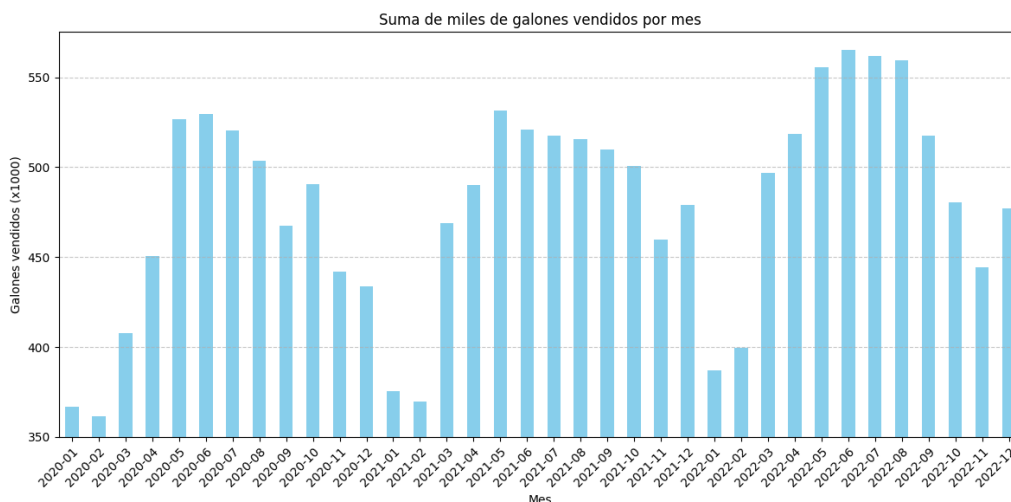
Hipótesis de estacionalidad de las ventas:

$H_0 : \beta = 0$ Las ventas no presentan estacionalidad

$H_1 : \beta \neq 0$ Las ventas son estacionales

Dado un valor de significancia estándar de $\alpha = 0,05$ y que el p-value obtenido en la prueba ADF es de 0.01, contamos con suficiente evidencia para rechazar la hipótesis inicial, por lo que las ventas son, en efecto, estacionales.

Una vez agrupadas las ventas y confirmada la estacionalidad, se escalaron los datos a miles de galones y se generó una gráfica de barras donde en el eje horizontal se encuentran los meses de los 3 años y en el eje vertical el total de galones vendidos donde cada unidad representa mil galones.



Con base en la gráfica anterior, es posible identificar que los meses con mayores ventas en el 2020 fueron Junio, Mayo y Julio, en 2021 fueron Mayo, Junio y Julio y en 2022 fueron Junio, Julio y Agosto. Con esto, el conjunto de meses que se vendió más entre los 3 años fueron Mayo, Junio, Julio y Agosto.

Por otra parte, los meses con menores ventas en el 2020 fueron Enero, Febrero y Marzo, en 2021 fueron Enero, Febrero y Noviembre y en 2022 fueron Enero, Febrero y Noviembre. Con esto, el conjunto de meses que se vendió más entre los 3 años fueron Enero, Fe-

brero, Marzo y Noviembre.

5. Discusión

Recopilando el análisis anterior, es posible validar la hipótesis que las ventas presentan una variación estacional dada el resultado de la prueba de Dickey-Fuller y analizando la gráfica fue posible confirmar que los meses con mayores ventas fueron en los meses de verano y las menores ventas en los meses de invierno.