

Evaluación Final

Natalie Julian

Delivery de productos de supermercado

El e-commerce (venta de productos por internet) se ha transformado en tendencia en el último tiempo, la idea es simple: Pides por internet los productos que necesites y el delivery (o despacho) se realiza a la comodidad de tu casa, te ahorras tiempo, filas, y el traslado desde tu hogar al supermercado.

El archivo `delivery` contiene información de productos de un supermercado que provee este servicio de despacho a casa. Los clientes piden por la aplicación los productos deseados, pagan su pedido por internet y todo llega rápidamente a su hogar. La información es la siguiente:

- **ID producto:** Código identificador del producto
- **Nombre producto:** Nombre del producto
- **Pasillo:** Pasillo del supermercado donde se encuentra dicho de producto (categoría)
- **NPedidos:** Cantidad de pedidos en los que se pidió el producto
- **RepPedidos:** Porcentaje de veces que los usuarios vuelven a pedir este producto. Un valor 1 indica que los usuarios **siempre** vuelven a pedir este producto, un valor de 0 indica que los usuarios nunca volvieron a pedir este producto.
- **Prioridad:** Indica la prioridad del producto en la canasta, es decir, el orden en el que se pidió. Si aparece 1 indica que es el primer producto del pedido.

Análisis previo

- a.1)** Cargue los datos en R. Observe el formato de las variables, ¿es el adecuado para todas? Realice las modificaciones pertinentes. Explique brevemente por qué ocurrió este problema de lectura en los datos.
- a.2)** Verifique que en los datos cada registro corresponde a un producto y que ningún producto se repite. ¿Por qué es importante hacer esta verificación previo al análisis?

Análisis Estadístico

- b.1)** Realice un análisis exploratorio de los datos para cada variable de interés, ¿cuántos productos hay? ¿cuántas categorías de productos (pasillos) hay? ¿qué rango de valores recorren las variables cuantitativas? ¿qué variables presentan mayor variabilidad? Comente.
- b.2)** Determine cuáles son los 20 productos que más y menos han sido solicitados para delivery y compare la cantidad de veces que han sido solicitados. ¿A qué categoría suelen pertenecer estos productos? ¿Cree usted que tienen sentido estos resultados? ¿Por qué?

Evaluación Final

Natalie Julian

- b.3)** Compare estadísticas del porcentaje de veces que los productos fueron pedidos nuevamente en las siguientes categorías:

- eggs
- milk
- coffee
- beauty
- cleaning products
- vitamins supplements

En base a las cifras, ¿en cuáles de estas categorías esperaría que los clientes vuelvan a pedir productos? ¿Tiene sentido? Usted como cliente, ¿compraría con regularidad estos productos? Comente.

- b.4)** Obtenga por categoría el promedio del porcentaje de veces que los productos fueron pedidos nuevamente. Realice un histograma de estos promedios, ¿qué observa? Comente sobre el histograma (forma, rango de valores, inclinación, etcétera). Sobreponga una curva de una distribución normal, puede probar con distintos parámetros (μ y σ^2). Comente.
- b.5)** Defina como variable aleatoria X tal que $X = 1$ indica que el producto tiene un porcentaje de veces que fue pedido nuevamente (*RepPedidos*) mayor o igual a 0.6 y es $X = 0$ si es menor a 0.6. ¿Qué distribución tendría X ? Utilizando los datos, ¿cuál sería el parámetro que propondría utilizar?. Si se le pide determinar la cantidad de productos que tienen un porcentaje de veces que fue pedido nuevamente mayor o igual a 0.6, ¿qué distribución utilizaría? ¿con qué parámetros?. Finalmente, ¿cuál sería el valor esperado de la cantidad de productos que tienen un porcentaje de veces que fue pedido nuevamente mayor o igual a 0.6?
- b.6)** Obtenga por departamento el promedio de la prioridad de los productos y el promedio del porcentaje de veces que los productos fueron pedidos nuevamente. Comente sobre el comportamiento entre estas dos variables. Interprete la correlación de Pearson. ¿El valor es positivo o negativo? ¿Tiene sentido? Interprete. A su vez, comente sobre la magnitud de la correlación, ¿es alta? ¿baja? Complemente con un gráfico de dispersión.
- b.7) Conclusión:** Explique, con sus palabras, por qué cree que es interesante y útil realizar este análisis. ¿De qué manera los resultados obtenidos podrían conllevar alguna mejora en (por ejemplo) problemas de abastecimiento (stock) del supermercado, estrategias de venta, sugerencias y ofertas para los clientes u otro? Proponga aplicaciones.