

Droguerías Los Alpes

Droguerías Los Alpes es una cadena de droguerías con más de mil sucursales en diferentes países de Latinoamérica y El Caribe. Con el objetivo de lograr una mejor gestión de la cadena de suministros, el área operativa de Droguerías Los Alpes desea explorar nuevos mecanismos para estimar las ventas en las diferentes sucursales, las cuales están influenciadas por muchos factores como el tipo de la sucursal, el nivel de surtido de sus productos, la competencia, el día de la semana, si es día festivo, las promociones y la estacionalidad.

Como científico de datos de Droguerías Los Alpes, su tarea es validar el uso de técnicas de machine learning que apoyen la toma de decisiones del equipo de operaciones a partir de las ventas que se puedan predecir para una sucursal en un día particular del año.

Para este propósito le han compartido dos conjuntos de datos: (1) el histórico de ventas de las diferentes sucursales para el periodo de enero del 2013 a julio del 2015, y (2) las características correspondientes a cada sucursal. Adicionalmente, le han compartido el diccionario de datos correspondiente.

Dentro del alcance del proyecto se deben considerar los siguientes entregables:

1. **[20%]** Un análisis exploratorio y de calidad de los datos suministrados. En la medida de lo posible, incluya técnicas de análisis univariado, bivariado, gráficas y no gráficas. Responda la pregunta: **¿Qué tanto varían las ventas dependiendo del mes o del día de la semana, así como del nivel de surtido de la sucursal?**
2. **[20%]** Realice la limpieza y preparación de datos correspondiente para que puedan ser consumidos por los algoritmos de machine learning que va a utilizar. **Adjunte un diagrama de bloques que evidencie los diferentes procesos de limpieza y preparación implementados.**
3. **[40%]** Muestre los resultados obtenidos para los conjuntos de entrenamiento y prueba utilizando el mejor modelo obtenido. Utilice las técnicas de transformación de datos, selección de atributos y regularización que considere convenientes. *Tenga en cuenta que cuando se cuenta con información temporal es común hacer la separación de los conjuntos de datos por ventanas de tiempo, en lugar de hacerla de forma aleatoria. Por ejemplo, si cuenta con información de solo un año, podría utilizar los datos de enero a septiembre para el entrenamiento y de octubre a diciembre para prueba.* **Explique si los resultados del modelo serían, de acuerdo a su criterio, suficientes para el uso de dicho modelo por parte de la empresa.** Recuerde argumentar sus decisiones con base en la interpretación cuantitativa y cualitativa de los resultados.
4. **[20%]** Responda a la pregunta: Asumiendo que el mejor modelo obtenido es suficiente en términos de capacidad de predicción y validación de supuestos, **¿cuál**

es el impacto promedio en la ventas de una sucursal si se decide establecer una promoción y adicionalmente si dicha promoción se aplica un fin de semana?

Condiciones de entrega:

La entrega debe hacerse por BloqueNeón antes de la hora estipulada. Adjunte el notebook **ejecutado** con el desarrollo de cada uno de los puntos planteados de forma clara y ordenada.

¡Éxitos!