Pasos del proceso de Precio Óptimo

El proceso se divide en 6 pasos que se tienen que realizar para cada producto que vende la estación. Antes de empezar con este proceso se calculan los principales competidores de la estación.

1. **Análisis Exploratorio de Datos:** Construir un análisis exploratorio de la información de la estación, así como de sus 10 competidores más cercanos. También se hace una lectura de precios de las estaciones que están en el mismo municipio para hacer un resumen de los precios y un comparativo. Para este paso se usa la información de precios del archivo XML de la CRE.

Se inserta la estación al catálogo de estaciones que tienen el módulo de precio recomendado para poder añadirla posteriormente al proceso automático que actualiza la información de precios para la sección de precio óptimo. Se hace una lectura de las ventas de la estación y se cambia al formato necesario para ocuparse en los siguientes pasos.

1. **Cálculo de correlaciones:** Se hace un análisis de correlaciones lineales entre las ventas y los cocientes de precios del precio de la estación contra el promedio de precios de competidores.

Primero se calculan las correlaciones de ventas contra precios de cada uno de los competidores individualmente y después se toma la correlación contra los *n* competidores más cercanos (donde *n* se toma de 1 a 10). Esto se hace para delimitar las opciones de la combinación de competidores que se elegirá ya que de no hacerlo el costo computacional sería muy alto y así se descartan los competidores cuyos cambios en precios no parecen influir en las ventas de la estación.

Ya que se tiene la combinación óptima se guardan las correlaciones encontradas y asignamos una ponderación a cada competidor seleccionado tomando en cuenta la correlación, si el competidor está en la misma calle, si está antes o después de la estación y la cercanía que tiene a la estación. Las ponderaciones para este cálculo son las siguiente:

* Correlación de cociente de ventas – 30 %
* Cercanía a la estación – 50 %. Toma valores de:

0 si el competidor está a más de 2 km de distancia.

.25 si el competidor está entre 1.5 y 2 km de distancia.

.5 si el competidor está entre 1 y 1.5 km de distancia.

.75 si el competidor está entre .5 y 1 km de distancia.

1 si el competidor está a menos de .5 km de distancia.

* Misma calle – 20%. Toma valores de:

1 si el competidor está en la misma calle y se encuentra antes de la estación.

.3 si el competidor se encuentra en la misma calle y después de la estación.

0 en otro caso.

Se suma el total de las ponderaciones nuevas calculadas para los competidores y se dividen las ponderaciones entre este número para que la suma nos de 100 % otra vez.

Nota: Para calcular las correlaciones se usa el coeficiente de correlación de Pearson que mide la correlación **lineal**entre dos variables, casos en los que haya una fuerte correlación entre las ventas de la estación con los precios de algún competidor o cociente pero que no sea un comportamiento lineal no va a resultar en una correlación tan significativa.

1. **Ajuste lineal ventas/cociente:** Se usa el cociente de precios contra los competidores seleccionados en el paso anterior para construir un ajuste lineal de las ventas contra el cociente. También se escogen los días sobre los cuales se considerará la información en caso de que se decida no tomar en cuenta algunos días por temas de estacionalidad o que sea necesario aplicar alguna transformación a los datos.

Se analiza el comportamiento del ajuste lineal encontrado en el paso anterior para ver si la demanda es muy inelástica o si resulta haber una pendiente positiva para ver si conviene meter algún quiebre o cambiar de ajuste para evitar que se recomiende aumentar el precio a niveles muy altos.

Se construye una primera curva de ganancias usando el margen de la estación para el producto en cuestión (contra el precio de pipas o el precio en TAR) para ver en que nivel de cociente estaría la recomendación.

1. **Selección de variables para modelo de Regresión Lineal Múltiple:** Se calculan distintas variables (variables de estacionalidad, información de ventas de otros productos de la misma estación, variables de precios de competidores además del precio promedio de los competidores seleccionados, variables de cambios en ventas de la misma estación) para construir diferentes modelos de regresión lineal múltiple y después hacer una parte de selección de variables para quedarnos con el mejor modelo.

La selección de variables se realiza mediante el método de *Selección Hacia Adelante* en el cual se empieza con un modelo de regresión lineal simple (solamente tomando el cociente como variable independiente) y se van agregando variables de las que se calcularon en el paso anterior. Se selecciona una variable adicional si el modelo resultante puede explicar de mejor forma la variación en las ventas del producto que se esté analizando (un nivel de R^2 mayor y mayor R^2 ajustada), sin que al hacer esto se rompan los supuestos de los modelos de regresión lineal (que no haya multicolinealidad entre las variables y que se tenga homocedasticidad).

Se escoge esta forma de elegir variables para partir del cociente de precios como una primera variable dependiente que es la principal variable de interés. De esta forma se pueden considerar distintas combinaciones del resto de las variables sin que el costo computacional de realizar la selección sea tan alto y además se tiene como resultado un modelo que no es tan complejo y por lo tanto más fácil de interpretar.

Se revisan los niveles de R^2 y R^2 ajustada ya que, si solamente tomáramos en cuenta la primera, siempre se tendrían mejores resultados al agregar más variables ya que de esta forma no estaríamos tomando en cuenta la complejidad del modelo y puede llevar a un modelo que explica mejor la variabilidad de las ventas en los datos de entrenamiento, pero que es peor en generalizar a los datos de prueba que se estudian en el siguiente paso.

1. **Entrenamiento y validación del modelo:** Se dividen los datos en una parte para entrenar el modelo (70%) y otra parte (30%) para poder hacer pruebas y ver qué tanto es el error en el pronóstico de ventas que genera el modelo.

Al separar los datos de esta forma se evita que el modelo resultante sea malo en generalizar las predicciones de ventas cuando se tienen niveles de variables independientes que no se conocen.

1. **Simulación del proceso de recomendación de precios en producción:** Se hace una simulación de cómo quedarían los resultados al mandarlos a producción para poder revisar las aproximaciones que se tienen de ganancia y de ventas en distintos niveles de cociente.

Se usan los parámetros del modelo calculado anteriormente para obtener la curva de ganancias y el ajuste a las ventas.

Se revisa el nivel de cociente para el precio recomendado. Después se revisa si la recomendación es menor al precio mínimo o mayor al precio máximo de los 10 competidores más cercanos a la estación y en caso de serlo la recomendación se cambia al precio mínimo o máximo respectivamente para evitar cambios de precios muy drásticos.

Finalmente se envían los parámetros del modelo a las tablas que alimentan el proceso automático para actualizar la información cada hora con para ajustar la recomendación a los nuevos niveles de precio de la estación y de los competidores.