**Elasticidad precio cruzada de la demanda para el mercado de motos nuevas en Colombia**

**Grupo 3**

*Cristian Castro Arias, Javier Patiño Serna, Sandra Marcela Díaz Cordero, Simón Madrid Álvarez*

**Problema del negocio.**

El mercado de motos nuevas en Colombia vende alrededor de 750 mil unidades al año; en este participan diferentes marcas (Auteco, Yamaha, Akt, Hero, Honda, Suzuki, Uma) que compiten por ganar share (porcentaje de las ventas).

El problema que se quiere resolver con este proyecto integrador es: Para cada una de las referencias de moto de la marca Auteco, hallar la variación de las unidades vendidas (variable dependiente) como resultado de las variaciones de precios las referencias de motos de sus competidores (Variables explicativas).

**Impacto de la solución**

Con la implementación de la solución, la ensambladora podría identificar como le afectan los cambios de precios de sus competidores en las unidades a vender de sus referencias.

Este efecto en las ventas puede no se inmediato sino tener un rezago de días o semanas, es decir, la ensambladora podrá saber que, si el competidor X decidió mover los precios de una referencia de moto específica hoy, sus unidades se van a ver afectadas en un periodo t+k (siendo k el rezago simulado por el modelo).

Con esta información, la empresa podrá hacer un presupuesto dinámico (de ser posible diario) de las ventas e incluso hacer una intervención temprana de precios si así lo requiere.

Adicionalmente, a través de la matriz de coeficientes, la empresa tendrá la posibilidad de identificar cuáles son los competidores más directos

**Descripción de datos a utilizar**

***Base de datos 1:***

RUNT motocicletas.

*Descripción:*

Esta información es pública (no confidencial), pero con acceso limitado a través de un pago mensual.

La base RUNT de motocicletas contiene las ventas diarias de motos nuevas en Colombia discriminadas por referencias.

*Variables*

1. Periodo: Día de la venta (la información es diaria)
2. Ensambladora: Compañía que ensambla la motocicleta en Colombia
3. Marca: Marca de la motocicleta (para la mayoría de los casos, la marca es igual a la ensambladora)
4. Modelo: Corresponde al nombre técnico de la referencia de la motocicleta, tal y como está registrado en el RUNT
5. Combustible: Muestra si la moto es eléctrica o a gasolina
6. Color: Color de la moto vendida
7. Departamento: Departamento en el cual se matriculó la motocicleta
8. Ciudad: Ciudad de Colombia en la cual se matriculó la motocicleta
9. Cantidad total: Unidades de motos vendidas

*Anonimización*

SI\_\_\_ | NO\_X\_\_

*¿Esta base de datos es pública?*

SI\_X\_\_ | NO\_\_\_

***Base de datos 2:***

Seguimiento de precios de referencias de motos por ensambladora

*Descripción:*

Esta información es pública (no confidencial) y proviene de las páginas web de cada una de las ensambladoras que pertenecen al mercado de motos en Colombia. La base de datos se genera en formato csv a través de un web scraping que consulta la información en la web.

Esta base contiene los precios diarios de las referencias de motos que se venden en Colombia y está discriminada por cada ensambladora.

*Variables*

1. Periodo: Día de la consulta
2. Ensambladora: Compañía que ensambla la motocicleta en Colombia
3. Marca: Marca de la motocicleta (para la mayoría de los casos, la marca es igual a la ensambladora)
4. Modelo agrupado: Corresponde al nombre de la moto en la página web
5. Precio: Precio de la moto publicado en la página web de su distribuidor

*Anonimización*

SI\_\_\_ | NO\_X\_\_

*¿Esta base de datos es pública?*

SI\_X\_\_ | NO\_\_\_

***Base de datos 3:***

Diccionario de homologación de nombres de referencias.

*Descripción:*

Esta información es de construcción interna de la empresa y de actualización a demanda (cada vez que se crea un nuevo modelo de motocicleta)

Contiene una homologación del nombre la moto en el RUNT, el nombre de la moto en la página web y un nombre homologado de fácil interpretación interna

*Variables*

1. Nombre RUNT: Corresponde al nombre técnico de la referencia de la motocicleta, tal y como está registrado en el RUNT
2. Nombre página Web: Corresponde al nombre de la moto en la página web
3. Nombre homologado: Corresponde al nombre construido internamente par fácil interpretación (en la mayoría de los casos es igual al nombre de la página web)

*Anonimización*

SI\_\_\_ | NO\_X\_\_

*¿Esta base de datos es pública?*

SI\_\_\_ | NO\_X\_\_

***Base de datos 4:***

Listado de referencias de motos por segmento

*Descripción:*

Esta información es de construcción interna. Contiene los segmentos a los que pertenecen cada una de las referencias de motos del mercado (incluyendo las motos propias y de los competidores).

Estos segmentos son motos de trabajo, deportivas, deluxe, scooter y moped.

*Variables*

1. Nombre homologado: Es el nombre resultante del diccionario de la base 3
2. Nombre del segmento: Es el nombre del segmento al cual pertenece la referencia de moto (trabajo, deportivas, deluxe, scooter y moped)

*Anonimización*

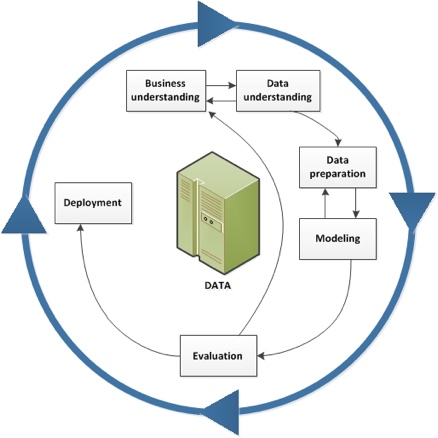
SI\_\_\_ | NO\_X\_\_

*¿Esta base de datos es pública?*

SI\_\_\_ | NO\_X\_\_

**Metodología a emplear.**

El modelo a utilizar es CRISP-DM. Los pasos principales de CRISP-DM son:



Los pasos a seguir son:

1. Entendimiento del negocio, problema o hipótesis: Las ensambladoras necesitan identificar el impacto en su participación de mercado de cada referencia de moto (demanda) de acuerdo con los cambios que sus competidores realizan diariamente en los precios. El objetivo principal es calcular y proyectar una elasticidad precio cruzada de la demanda que les permitirá reaccionar oportunamente a los cambios en su entorno.
2. Entendimiento de los datos: Los datos provienen del RUNT. Esta es una base de datos pública. Se obtendrá información como cantidad de motos vendida y los precios de las motos en distintos periodos de tiempo.
3. Preparación de los datos: En esta etapa se realizará un análisis exploratorio de datos con el objetivo de identificar outliers. Adicionalmente, se identificarán las variables relevantes para predecir correctamente la variable objetivo, de manera que no se tenga multicolinealidad entre las columnas del dataframe.
4. Modelado: En esta etapa se utilizarán diferentes modelos para resolver el problema propuesto. Entre los modelos que se han considerado hasta el momento tenemos: redes neuronales, regresión lineal, modelo Lasso, modelo VAR.
5. Evaluación: En esta etapa, se evaluará el desempeño de cada algoritmo mediante diferentes métricas. Con base en estas se escogerá el modelo que mejor prediga la demanda de las motos.
6. Despliegue: El despliegue se realizará un archivo .exe del código de python, un ETL leerá dicho archivo, y finalmente se publicará en una herramienta de inteligencia de negocios (Power BI).

**Métodos y aplicación en cada una de las materias del proyecto integrador**

*Modelos a usar:*

*Curso 1: ST1800 Almacenamiento y Recuperación*

los conceptos aprendidos en esta materia buscamos darle un correcto ciclo de vida a los datos, es decir, ser muy organizado con el proceso desde el almacenado en los datalakes(AWS), hasta la extracción de la data necesaria para el aprendizaje de los modelos, usando los moteres de búsqueda proporcionados por AWS.

*Modelos a usar:*

*Curso 2: CM0888 Algebra en ciencia de los datos*

Las tecnicas vistas en esta materia, serán muy útiles para el proceso de análisis de los datos, antes de comenzar a entrenar cualquiera de los modelos que serán implementados por nosotros. Estas tecnicas nos van a permitir identificar outliers, identificar colinealidad entre las variables de entrada de los modelos, elegir las métricas más adecuadas para el aprendizaje de nuestros modelos, entre otras tecnicas vistas.

*Modelos a usar:*

*Curso 3: EC1801 Estadística en Analítica*

Con los conceptos y tecnicas que vamos a ver en esta materia, esperamos elegir de manera adecuada el modelo correcto para nuestro problema en cuestión. Una vez tengamos un modelo entrenado, necesitamos medir el accuracy del mismo, y para ello necesitamos herramientas para tomar la decisión si nuestro modelo es adecuado o no. (Herramientas que esperamos profundizar en Estadística Analítica)

**Avances:**

Para el desarrollo de este proyecto, esperamos tener avances semanales secuenciales de acuerdo a los pasos de la metodología CRISP.

Versión 1: (15 de noviembre)

1. Problema/Contexto: hallar la variación de las unidades vendidas (variable dependiente) como resultado de las variaciones de precios las referencias de motos de sus competidores (Variables explicativas)
2. Datos: preparación de datos. Datos de entrenamiento. (preparado, identificación de outlier, preprocesamiento)
3. Análisis: Probar los diferentes modelos
4. Visualización: Python

*Versión 2: (30 de noviembre)*

1. Problema/Contexto: hallar la variación de las unidades vendidas (variable dependiente) como resultado de las variaciones de precios las referencias de motos de sus competidores (Variables explicativas)
2. Datos: Los datos ya han sido procesados y preparados.
3. Análisis: Identificar el modelo que mejor se ajusta a la solución esperada para el problema planteado
4. Visualización: Power Bi

*Versión 3: Entrega Final*

(formato específico para la entrega)

**Cronograma.**

|  |  |
| --- | --- |
| Fecha | ENTREGABLE |
| 06 de octubre | Entrega preliminar proyecto |
| 21 de octubre | Construcción de la propuesta definitiva |
| 28 de octubre | Arquitectura y almacenamiento de los datos. Exploración de modelos de aplicables |
| 02 de noviembre | Implementación código de modelos |
| 12 de noviembre | Definición de métricas e indicadores para seleccionar el modelo óptimo |
| 22 de noviembre | Visualización de la implementación |
| 29 de noviembre | Elaboración para entrega del documento final del proyecto |
| 30 de noviembre | **Documento final del proyecto** |
| 10 de diciembre | **Presentación** |