# Proyecto Data Science: AUTOMOTORA ANACONDA

### Integrantes:

- Paulo González
- Jorge Guerrero
- Daniel Mardones
- Gonzalo Rojas
- Esteban Sánchez

### **Profesor:**

Daniel Beffermann

#### **Tutor:**

- Jhon Poma





# Equipo de trabajo

## Roles para el desarrollo del proyecto



### Lider de equipo

Paulo González y Jorge Guerrero



### Analista de datos

Esteban Sánchez y Gonzalo Rojas



# Control de calidad y validación de datos

Esteban Sánchez y Gonzalo Rojas



### Ingeniero de modelamiento

Daniel Mardones y Jorge Guerrero



# Visualización de datos

Daniel Mardones y Jorge Guerrero



### **Documentador**

Daniel Mardones y Paulo González



# PLANIFICACIÓN DE LA INVESTIGACIÓN REQUERIMIENTOS



Generar una herramienta que ayude al equipo de compras a encontrar las mejores oportunidades.

 Modelo de clasificación con técnicas de Machine Learning para encontrar mejor rentabilidad



Definir propuesta de inversión para la Automotora.

 Definición de propuesta de inversion relacionada a los vehículos que estén por debajo del precio de venta promedio, estableciendo un margen de venta



Definir estrategia de venta por gama de vehículos

 Creación de columna binaria que clasifique la marca del vehículo por generalista (0) o premium (1), para posterior análisis del comportamiento en la venta de vehículos y desarrollar estrategia



Encontrar oportunidades de compra y venta entre ciudades/estados.

 Uso de EDA para determinar ubicación de las mejores oportunidades de compra y así encontrar mejores ciudades / estados para vender estos vehículos, maximizando la utilidad.



## Planificación de la investigación



### Hipótesis

Determinar mediante un modelo clasificatorio si es rentable la compra de un vehículo, de acuerdo a sus especificaciones, para una posterior venta



## Vector objetivo

Variable 'Price'



# Estrategias analíticas descriptivas

Análisis Exploratorio de Datos (EDA), para visualizar la distribución de vehículos de acuerdo: marca, modelos, año, ciudad, estado, millas y valores de vehículos



# Control de calidad, validación y análisis de datos

Control de calidad, validación y análisis de datos

Enriquecimiento del data set a través de webscrapping para obtener la mayor cantidad de información del vehículo (realizado mediante columna VIN)



## Planificación de la investigación

# Control de calidad, validación y análisis de datos

Creación de base de datos para almacenamiento de información

Normalización y validación de datos

Creación de columna binaria relacionada a la gama del vehículo: generalista o premium



## Planificación de la investigación:

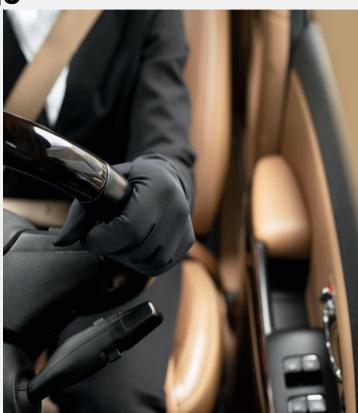
Modelación y predicción de trabajo

Implementación estándar de librería LazyPredict

 para conocer preliminarmente los mejores clasificadores Selección de los 5 mejores modelos a partir del paso anterior e implementación de ellos con hiperparámetros a definir

Las métricas a utilizar para seleccionar el mejor modelo serán: precisión, recall y f1-score

Utilización de base de datos Postgres para almacenar resultados, permitiendo realizar querys y obtener convenciencia de compra de vehículo



## Consideraciones del proyecto

