



**2022**

## **PROYECTO FINAL**

**VENTA DE CELULARES**

**PREPARADO POR  
RAFAEL HERNANDEZ**

**DATA SICIENCE**

**BOOTCAMP**  
Especialidad en Ciencia de Datos

# Preguntas planteadas

## Objetivo del proyecto

Una empresa libero datos sobre celulares estas características están atadas al un rango de precio, lo que nos pide este proyecto es predecir el rango de un celular con características que nos dan. La primera pregunta que me hice "Cuales son las características que afectan en el precio de un celular?".

## Exploracion de los datos

Al estar en esta etapa, pude preguntarme algunos puntos importantes es mas lo pregunte sin haber explorado el Data set . Lo primordial era verificar que el data set no tenga datos nulos porque me perjudicara después, en este caso no tuve, entonces procedí a ver la consistencia de datos que sean equivalentes o que estén completos. Todo iba excelente cuando me topo con lo que me preocupo desde le inicio el cual fue la correlación entre las características y el precio, empecé a graficar las variables que no eran categóricas y si note que no tenían mucho impacto el precio en estas variables. Analice las variables categóricas que solo eran de 0 y 1 el cual nos decía que 1 es que tenia esa función en caso contrario es que no la tenia.

Me decidí a sacar las variables que si tenían mas impacto en el rango de precio, me pregunte como puedo hacer un grafico donde pueda ver esta correlaciones y entonces hice un mapa de calor 'HeatMap' es aquí donde pude observar las variables que si afectan al precio. Observe que 4 variables predominantes pero la que sobresalía era la 'RAM' , viendo esas 4 no me quise ir directo hacer un modelo ML ya que no estaba seguro si en verdad hice un buen análisis entonces recurrí a una herramienta que nos enseñaron el bootcamp que es 'ProfileReport' el cual me ayudo hacer un análisis rápido y ojo me tope con muchos 0, los cuales me iban afectar en mi modelo. En el ancho de la pantalla 'sc\_w' en esta variable había la inconsistencia de 180 ceros.



# Aproximación a la solución de los planteamientos

## Limpieza de datos

Al notar esos 0, tuve que eliminarlos ya que es mejor para el modelo. Al hacer esto hicimos un análisis de todos los datos para ver si estaban equilibrados y en efecto si lo estaban.

## Dividimos el data set

En nuestro data set principal lo dividimos en dos, en una parte tenemos las características y en la otra el rango de precios.

Ahora cada una de esas en la función de nuestra navaja suiza 'train\_test\_split' lo dividimos el 80% para entrenamiento y el resto para validar. En este caso lo dividimos en dos ya que nuestros datos no son bastantes son 2000. Entonces esta decisión fue tomada por la escases de datos.

## Variables dependientes e independientes

'mutual\_info\_classif' Nos va ayudar a ver las dependencias que existen con respecto al valor objetivo si es 0 son independientes y en el caso contrario son dependientes.

Con el paso anterior ya podemos ver las variables las cuales van mejor para la predicción, ahora filtramos los valores que sea mayores a 0 ya que buscamos las variables dependientes esto lo logramos recorriendo los la función mutual la cual ya tiene los valores el cual le ponemos una condición que sean mayor a 0 y lo vamos agregando a una nueva variable.



## Escalar los valores

Usando la función de nuestra navaja suiza 'StandardScaler' ya que nos ayuda a normalizar los datos ya que nuestro data set tiene valores de diferente escala.

Al ejecutar esto no se lo que pasaba pero los nombres de las columnas me las ponía en números. Esto lo solucionamos cambiando las columnas con pandas tomando las columnas de los valores que filtramos con las variables dependientes .

## Preparando Modelo

Ya teniendo nuestras features ya entrenadas en el scaler y nuestras variables. Comenzamos a preparar el modelo en este caso usamos regresión lineal ya que se presta en este caso entrenamos y predecimos con nuestro entrenamiento el accuracy fue de 0.9752 el cual nos dice que no hubo underfitting.



# Hallazgos y conclusiones

## Modelo en validación

Al ejecutar nuestro modelo con los datos en validación, nos dio una buena precisión que desde mi punto de vista es eficaz en la predicción en rango de precios , uní las tablas de características y de la precisión para ver los resultados y ver mejor lo que quiero mostrar.

## Modelo en test

Ahora recurrimos con los datos de prueba que nos da esta compañía lo procesamos con los parámetros ya dichos y predecimos. Me sorprendí a ver los valores fue un éxito este proyecto.

## Conclusión

Pudimos solucionar el problema de Bob y su compañía ahora podemos ingresar características con las features que sacamos con dependencia al rango de precio ya ahora si ya podemos sacar los precios según sus características.



# Datos utilizados y GitHub



Mobile Price  
Classification

kaggle



Repositorio

