



DMC PERU

Estrategia de mercado en el sector retail

APLICACION DE MODELO DE REGRESION (ML SUPERVISADO)
APLICACION DE MODELO DE ARBOL DE DECISION

PRUEBA AZZ - LVV Somos + Podemos +

Edison Enrique Velazco Bautista
Business Analytics Jr

Indice



Tip: Use links to go to a different page inside your presentation.

How: Highlight text, click on the link symbol on the toolbar, and select the page in your presentation you want to connect.

Entendimiento del Negocio

Analisis Exploratorio de Datos (EDA)

ITratamiento y Preprocesamiento de Datos

Feature Engineering

Desarrollo del Modelo

Entrenamiento y Validacion

Resultados

Conclusiones

Implications and Recommendations

References

Whiteboard

Brainstorm

Q&A

[Back to Overview](#)



Entendimiento del Negocio

BigMart es una cadena de tiendas minoristas que opera en el sector de retail, ofreciendo una amplia variedad de productos a los consumidores en sus zona de venta. Para comprender mejor su negocio, es esencial analizar tanto los productos que ofrecen como su estrategia de zona de ventas y asesor por zona.



Statement of the Problem

¿Donde estamos?

[Back to Overview](#)

Objectives

- Tratar de extender de la gama de productos a gran escala en cada zona de venta particular
- Establecer medidas de zona de venta competitivos y con mejor experiencia al usuario con atencion con asesores
- Obtener mayor cuota de mercado en el sector retail a traves del posicionamiento estrategico.

[Back to Overview](#)

Metas para el desarrollo estrategico

Planteo de modelo estrategico

Desarrollar modelo supervisado a traves de una solucion de mercado como variable objetivo a las ventas que se expliquen a traves de las variables de precio nacional , zona, asesoramiento y cantidades estimadas

Planteamiento de operacion

Mostrar una relacion significativa entre las ventas y producto ,y ubicacion para hacer predecir el flujo sobre las ventas futuras y tomar decisiones informadas sobre estrategias de Precio y Area.

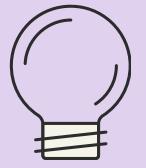
Ciclo de vida del modelo

- Exploracion empirica multidisciplanario del negocio.
- Tratamiento y Desarrolllo de Analitica de Datos



Source: ¿COMO LO HARE?

Hypothesis



Area Estrategica

Verificar la predicción de zonas
para incrementar datos



Planificacion estrategica de ventas

Determinar insights para un
mayor rendimiento comercial



Experiencia del Usuario

Lograr el alcance y sincronizacion
con clientes y mercados
potenciales

[Back to Overview](#)

ANALISIS EXPLORATORIO DE DATOS

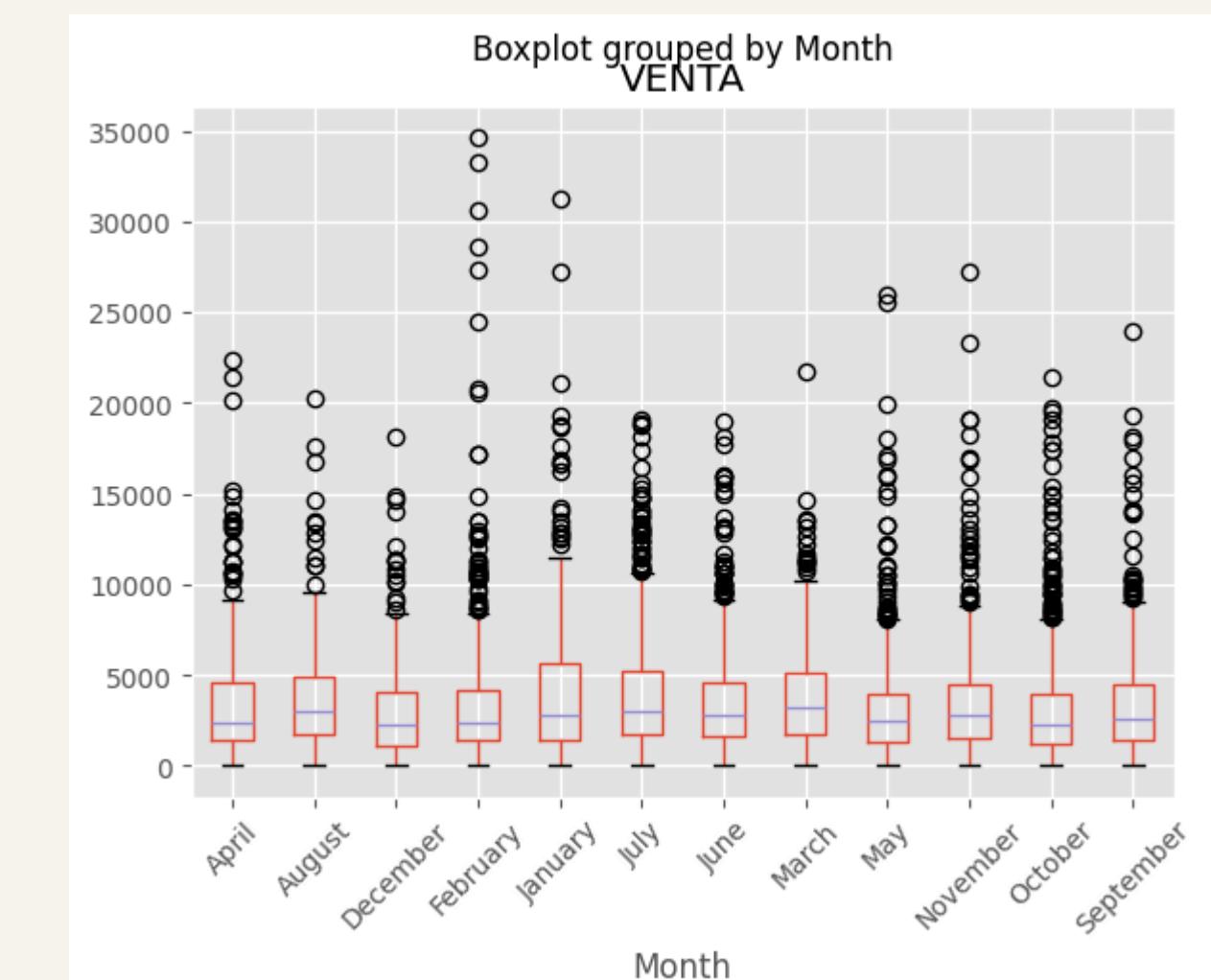
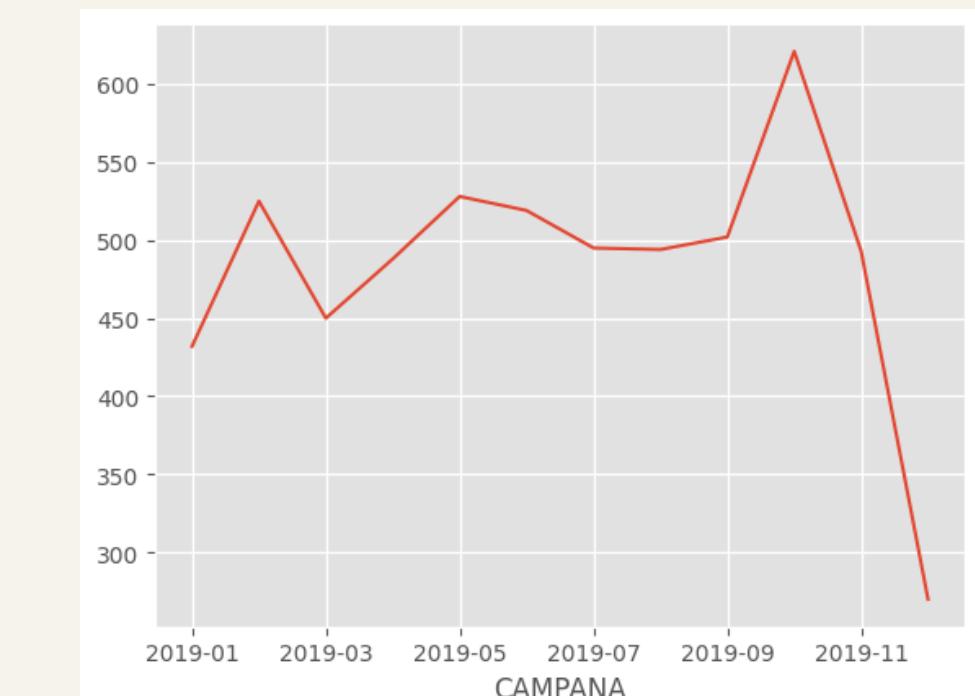
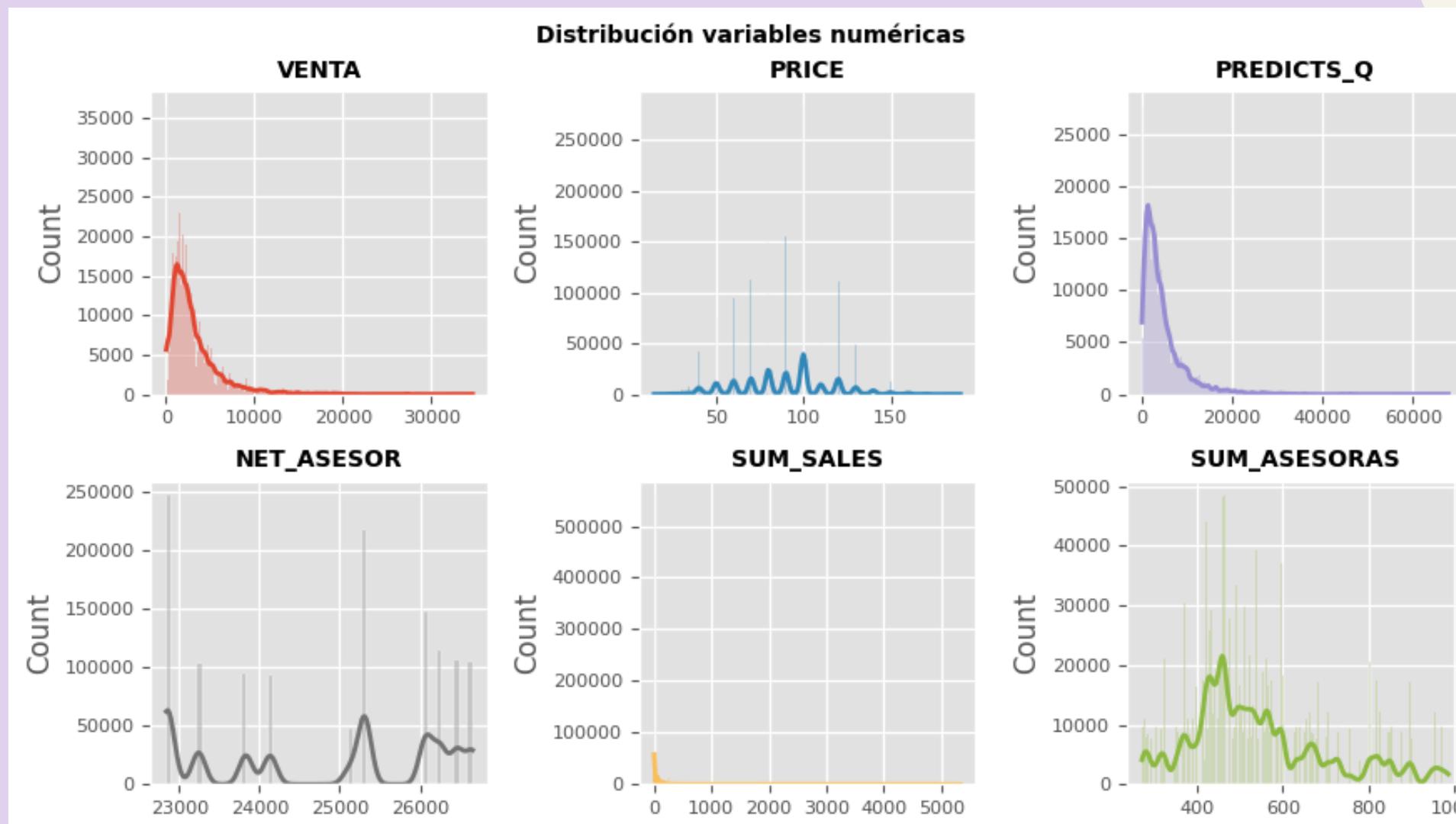
[Back to Overview](#)

	CAMPANA	VENTA	SUBGRUPO	SIZE	PRICE	PREDICTS_Q	NET_ASESOR	ZONA	SUM_SALES	ASESORA	SUM_ASESORAS
0	2019-01	41	VESTIDOS	T-L	84.9	86	24133	101	0	101	903
1	2019-01	41	VESTIDOS	T-L	84.9	86	24133	101	0	102	679
2	2019-01	41	VESTIDOS	T-L	84.9	86	24133	101	0	103	554
3	2019-01	41	VESTIDOS	T-L	84.9	86	24133	101	0	104	289
4	2019-01	41	VESTIDOS	T-L	84.9	86	24133	101	0	107	542
...
1273099	2019-12	18111	JEANS	T-8	49.9	36221	25126	119	254	111	492
1273100	2019-12	18111	JEANS	T-8	49.9	36221	25126	119	254	112	417
1273101	2019-12	18111	JEANS	T-8	49.9	36221	25126	119	254	115	437
1273102	2019-12	18111	JEANS	T-8	49.9	36221	25126	119	254	116	710
1273103	2019-12	18111	JEANS	T-8	49.9	36221	25126	119	254	119	347

Analisis Exploratorio de Datos

[Back to Overview](#)

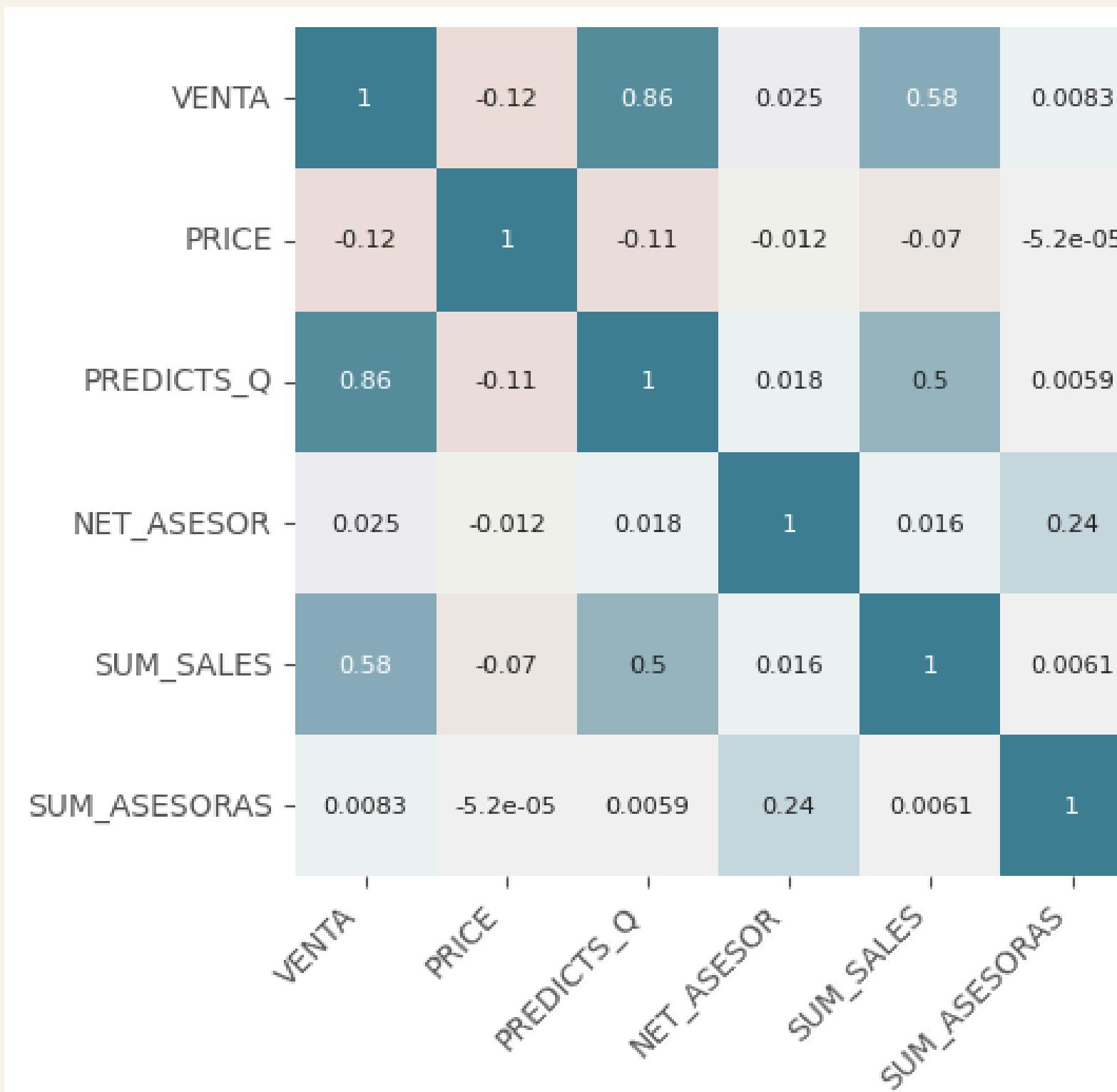
Realizamos un analisis a cada variable a nivel univariado seleccionando una buena distribucion de los datos



Analisis Exploratorio de Datos

[Back to Overview](#)

Matriz de correlacion



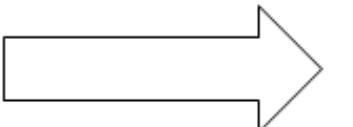
Existe mayor correlacion entre las ventas ,precio nacional ,estimacion de ventas,neto de asesores para obtener una vision amplia del negocio

Desarrollo del Modelo

[Back to Overview](#)

Se ha convertido las variables object en dummies para preprocesar la base de datos categoricas

Precio	area	Piscina
12	3	grande
15	4	sin
25	1	pequeña
11	6	pequeña
16	5	sin
7	3	grande



Precio	area	PisPeq	PisGra
12	3	0	1
15	4	0	0
25	1	1	0
11	6	1	0
16	5	0	0
7	3	0	1

Variables dummy

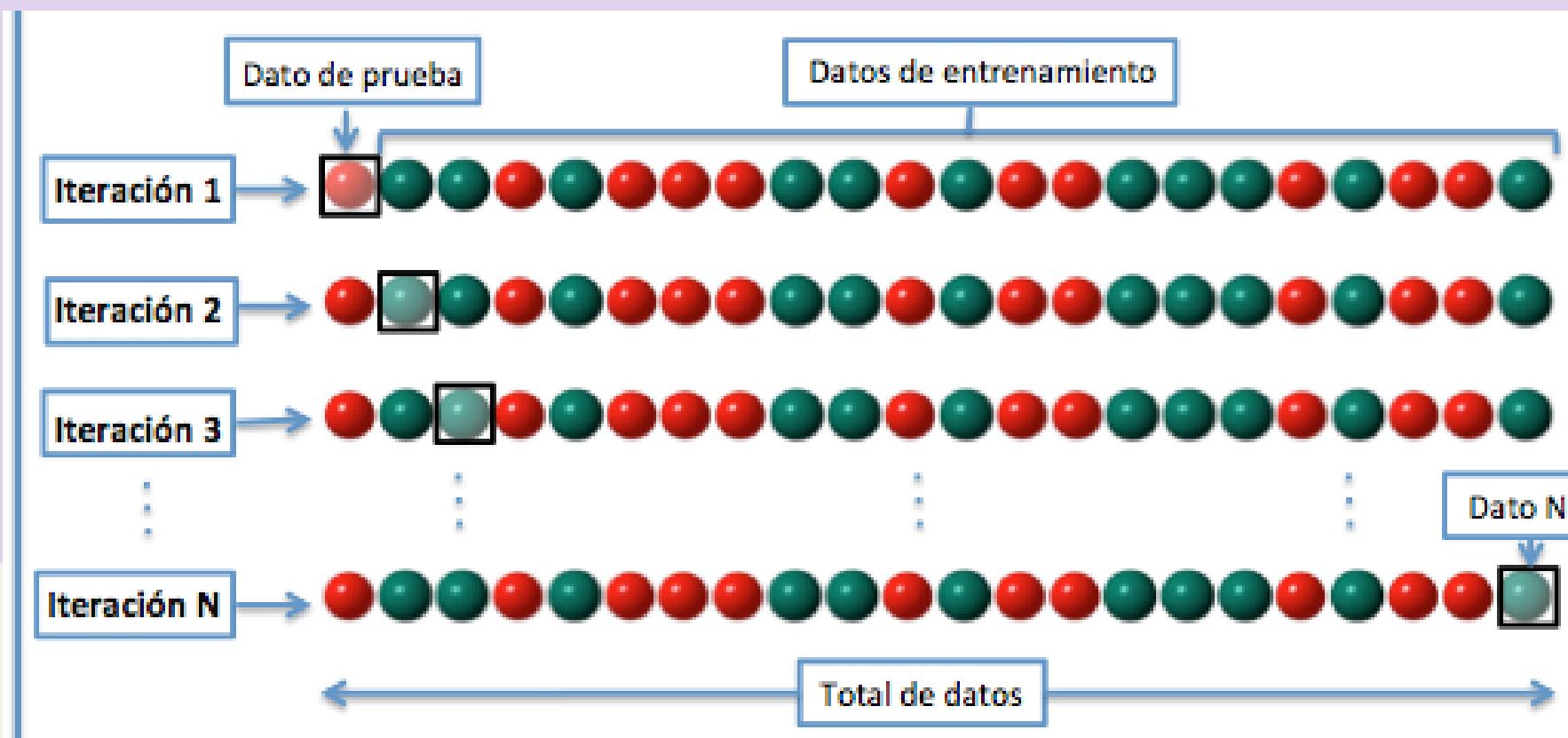
```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 1273104 entries, 0 to 1273103
Data columns (total 11 columns):
 #   Column           Non-Null Count  Dtype  
--- 
 0   CAMPANA          1273104 non-null  object 
 1   VENTA             1273104 non-null  int64  
 2   SUBGRUPO          1273104 non-null  object 
 3   SIZE              1273104 non-null  object 
 4   PRICE              1273104 non-null  float64
 5   PREDICTS_Q         1273104 non-null  int64  
 6   NET_ASESOR         1273104 non-null  int64  
 7   ZONA               1273104 non-null  object 
 8   SUM_SALES          1273104 non-null  int64  
 9   ASESORA            1273104 non-null  object 
 10  SUM_ASESORAS       1273104 non-null  int64  
dtypes: float64(1), int64(5), object(5)
memory usage: 106.8+ MB
```

Desarrollo del Modelo

Se ha dividido la base de datos 2 dataframes para probar la asistencia y rendimiento del modelo a través de una dataset de prueba y test



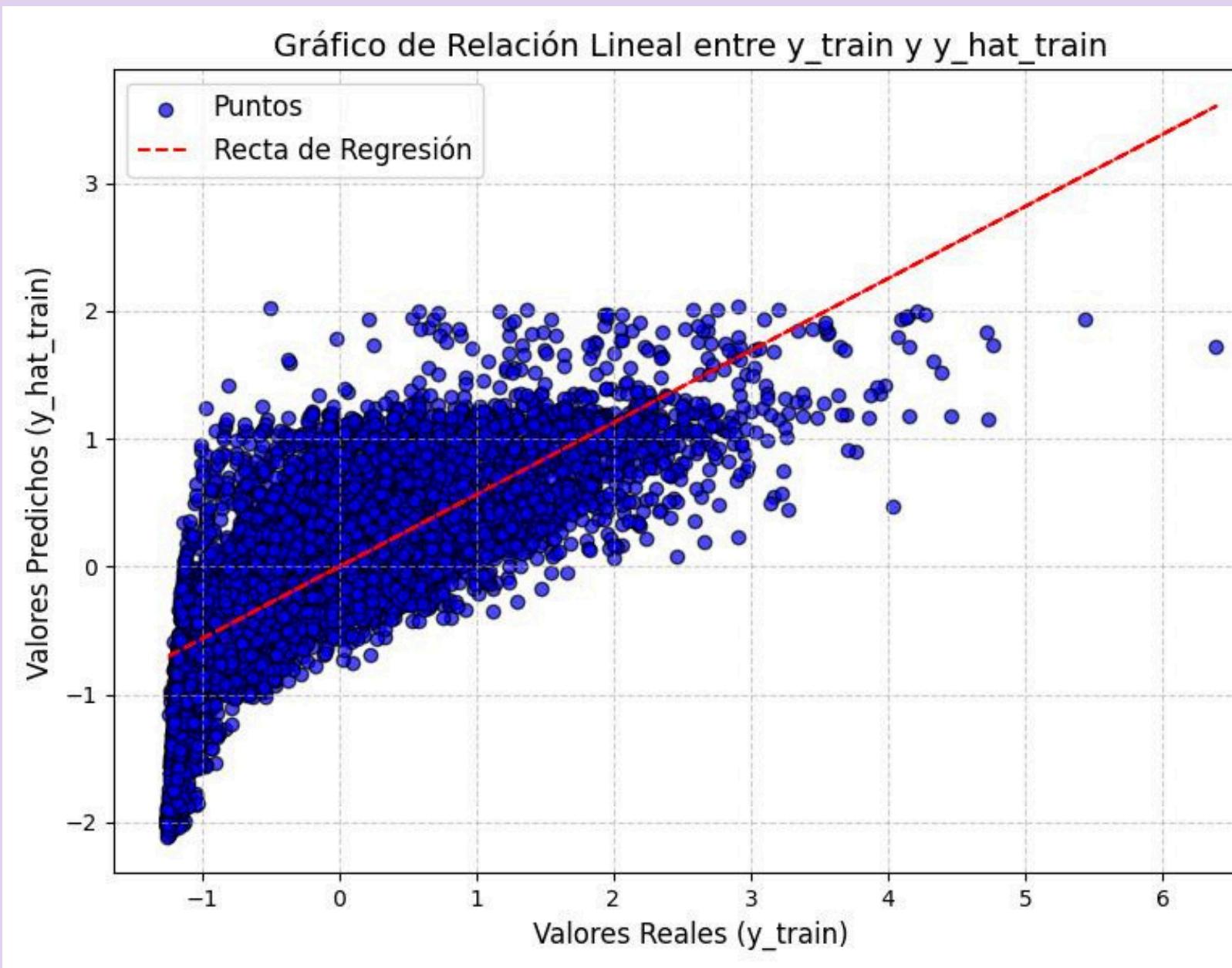
Se ha ajustado la cantidad de cada dataset de acuerdo a un 80% y 20% de datos correspondiente dataset a partir de la división de campañas



```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 1273104 entries, 0 to 1273103
Data columns (total 11 columns):
 #   Column           Non-Null Count  Dtype  
--- 
 0   CAMPAÑA          1273104 non-null  object  
 1   VENTA             1273104 non-null  int64  
 2   SUBGRUPO          1273104 non-null  object  
 3   SIZE              1273104 non-null  object  
 4   PRICE              1273104 non-null  float64 
 5   PREDICTS_Q         1273104 non-null  int64  
 6   NET_ASESOR         1273104 non-null  int64  
 7   ZONA               1273104 non-null  object  
 8   SUM_SALES          1273104 non-null  int64  
 9   ASESORA            1273104 non-null  object  
 10  SUM_ASESORAS       1273104 non-null  int64  
dtypes: float64(1), int64(5), object(5)
memory usage: 106.8+ MB
```

Desarrollo del Modelo

Se ha decidido de armonizar un modelo de regresion lineal,pues tendremos a ajustar los datos a partir de una recta que se ajuste de bondad con los datos



Método

El Método de Mínimos Cuadrados Ordinarios es una técnica fundamental en la estadística y la econometría y se utiliza para ajustar modelos de regresión lineal simple y múltiple a datos observados. Su objetivo es encontrar la mejor línea (o hiperplano) que representa la relación entre las variables, minimizando la suma de los errores al cuadrado. Esto facilita la interpretación de las relaciones y la predicción de valores futuros.

$$\sum_{i=1}^n e_i^2 = \sum_{i=1}^n (y_i - \hat{y}_i)^2$$



El modelo presenta una buena predicción entre las ventas reales y las predección de las ventas.

Se necesitan pruebas de ajuste que muestre la bondad de linealiadad de los datos,como la composicion de sus variables

Entrenamiento y Evaluacion

[Back to Overview](#)

p-value = 0.00

R2-score: 0.769

1. Relación Lineal entre las variables predictoras y la variable objetivo.
 2. Variable de Salida Continua La regresión lineal se utiliza para predecir una variable de salida continua. .
 3. Interpretación Simple: La regresión lineal proporciona coeficientes que son fácilmente interpretables.
 4. Simplicidad: La regresión lineal es un modelo simple y fácil de entender.
 5. Requisitos de Datos: En comparación con algunos modelos de aprendizaje profundo o no supervisados, la regresión lineal a menudo requiere menos datos para entrenar con precisión.

```

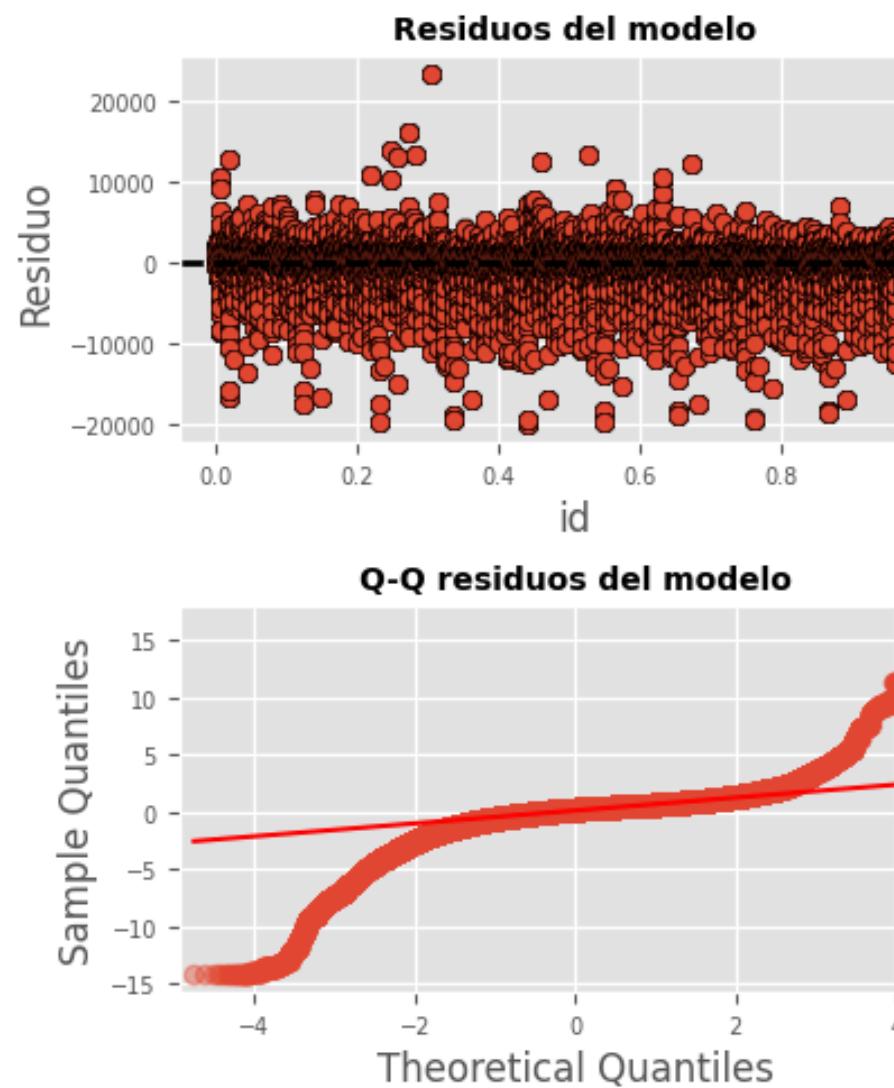
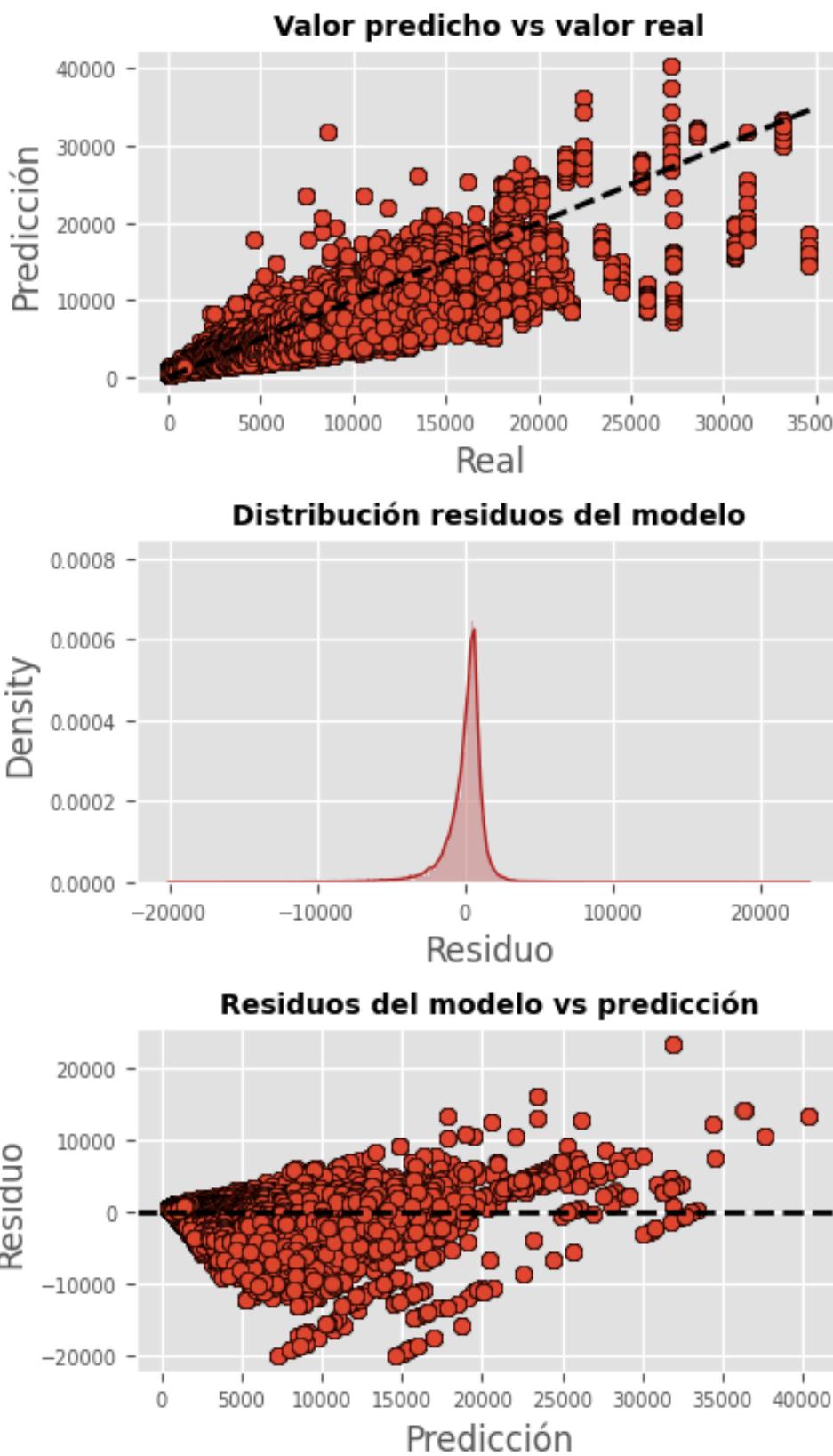
OLS Regression Results
=====
Dep. Variable: VENTA   R-squared:      0.769
Model:          OLS    Adj. R-squared:  0.769
Method:         Least Squares F-statistic:   6.797e+05
Date:           Mon, 15 Apr 2024 Prob (F-statistic): 0.00
Time:            11:45:50 Log-Likelihood: -8.8334e+06
No. Observations: 1018484 AIC:             1.767e+07
Df Residuals:    1018478 BIC:             1.767e+07
Df Model:        5
Covariance Type: nonrobust
=====
              coef    std err       t      P>|t|      [0.025      0.975]
-----
const      549.5345   25.286     21.733     0.000    499.975    599.094
PRICE      -3.2505    0.053    -61.100     0.000    -3.355    -3.146
NET_ASESOR  0.0172    0.001     16.798     0.000     0.015     0.019
SUM_SALES   4.9172    0.014     356.846    0.000     4.890     4.944
SUM_ASESORAS 0.0119    0.009     1.334     0.182    -0.006     0.029
PREDICTS_Q  0.4512    0.000    1383.135    0.000     0.451     0.452
-----
Omnibus:            683637.048 Durbin-Watson:      0.115
Prob(Omnibus):      0.000  Jarque-Bera (JB): 23881695.135
Skew:                2.757  Prob(JB):        0.00
Kurtosis:             26.073 Cond. No. 4.56e+05
-----
Notes:
[1] Standard Errors assume that the covariance matrix of the errors is correctly specified.
[2] The condition number is large, 4.56e+05. This might indicate that there are
strong multicollinearity or other numerical problems.

```

Entrenamiento y Evaluacion

[Back to Overview](#)

Diagnóstico residuos



Predicciones

	mean	mean_se	mean_ci_lower	mean_ci_upper	obs_ci_lower	obs_ci_upper
0	738.914561	3.962930	731.147352	746.681771	-2032.509924	3510.339047
1	736.257887	2.479173	731.398792	741.116982	-2035.159974	3507.675748
2	734.775368	2.064505	730.729008	738.821728	-2036.641187	3506.191923
3	731.632427	2.905307	725.938124	737.326731	-2039.787024	3503.051879

Summary and Conclusion

[Back to Overview](#)



Atraccion de mayores proveedores.

Planear una estrategia de conseguir un red de proveedores que compitan por un precio minorista maximo.



Planeamiento de area estrategica

Redefinir la atencion al usuario a traves de la creacion de algun otro tipo de tienda.



Diseño Operativo de productos

Actuar sobre una seleccion de producto que manifiesten un poder de mercado