

# Experiencia de clientes en la compra de autos: Análisis de datos

---

INGENIERO LUIS DE LOS REYES VIDAL

# INDICE

---

- Contexto y Audiencia
- Hipótesis
- Análisis exploratorio
- Análisis bivariado
- Modelo de machine learning
- Cross validation
- Modelo final
- Conclusiones

# CONTEXTO Y AUDIENCIA

---

Analizar una base de datos de ventas de autos y reparaciones de autos es fundamental para obtener información valiosa que puede impulsar decisiones comerciales estratégicas y mejorar la eficiencia operativa en el sector automotriz.

El análisis de las ventas de autos proporciona información sobre los modelos y marcas más populares. Esto ayuda a los concesionarios y fabricantes a gestionar de manera más eficiente sus inventarios, enfocándose en los vehículos que tienen una mayor demanda.

La experiencia del cliente durante el proceso de compra proporciona información sobre qué aspectos impactan positiva o negativamente en la satisfacción del cliente. Esto permite implementar mejoras específicas para maximizar la satisfacción del cliente.

# HIPOTESIS

---

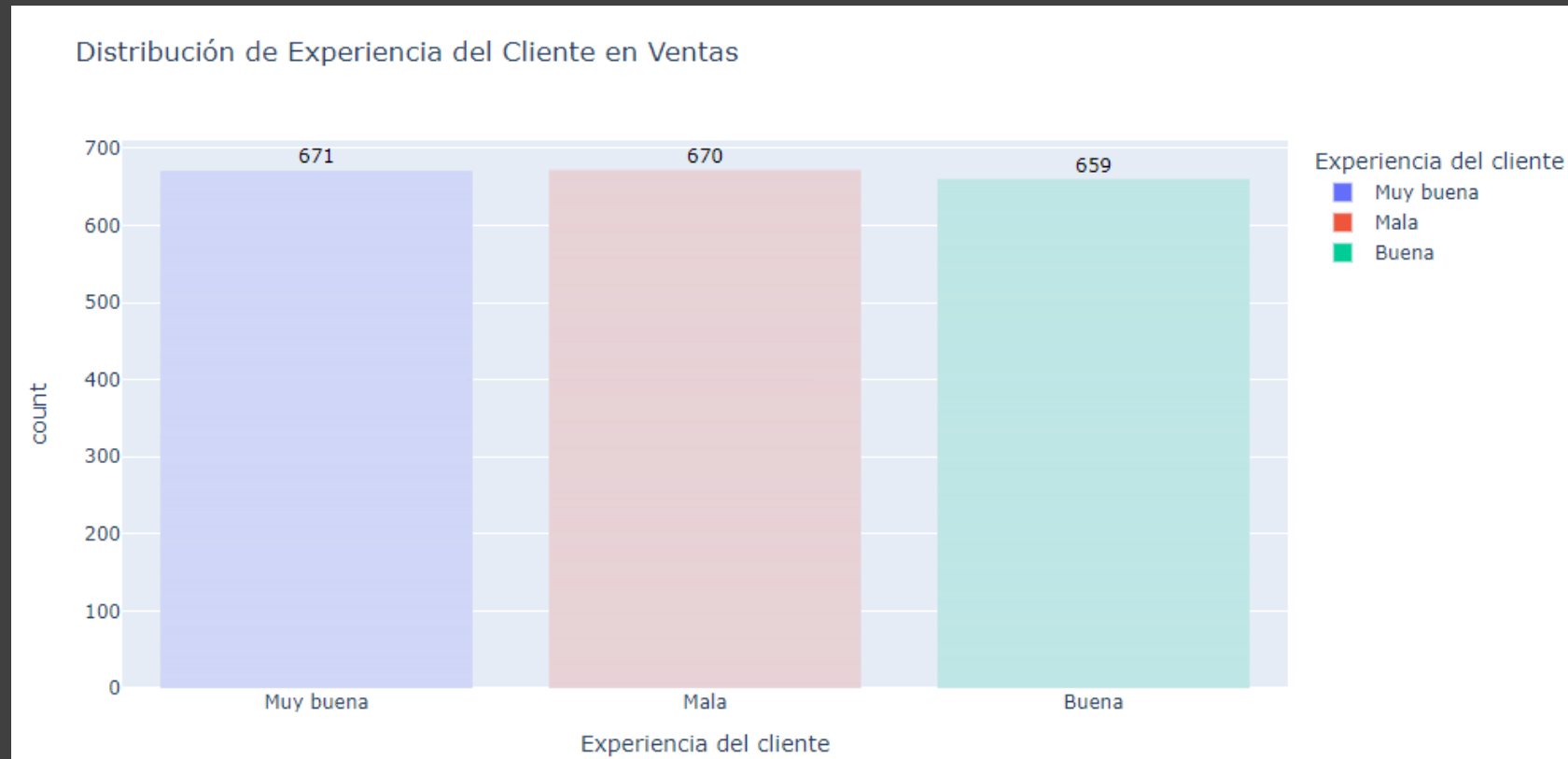
*Hipótesis nula ( $H_0$ ):* No hay diferencia en la probabilidad de que un nuevo cliente tenga una buena experiencia en el mercado automotriz.

*Hipótesis alternativa ( $H_1$ ):* Existe una diferencia en la probabilidad de que un nuevo cliente tenga una buena experiencia en el mercado automotriz.

La hipótesis nula sugiere que el modelo de predicción no tiene un impacto significativo en la capacidad de predecir si un nuevo cliente tendrá una buena experiencia en el mercado automotriz. Mientras que la hipótesis alternativa sugiere lo contrario: que el modelo de predicción tiene un impacto significativo en la capacidad de predecir la experiencia de los nuevos clientes.

# ANALISIS EXPLORATORIO

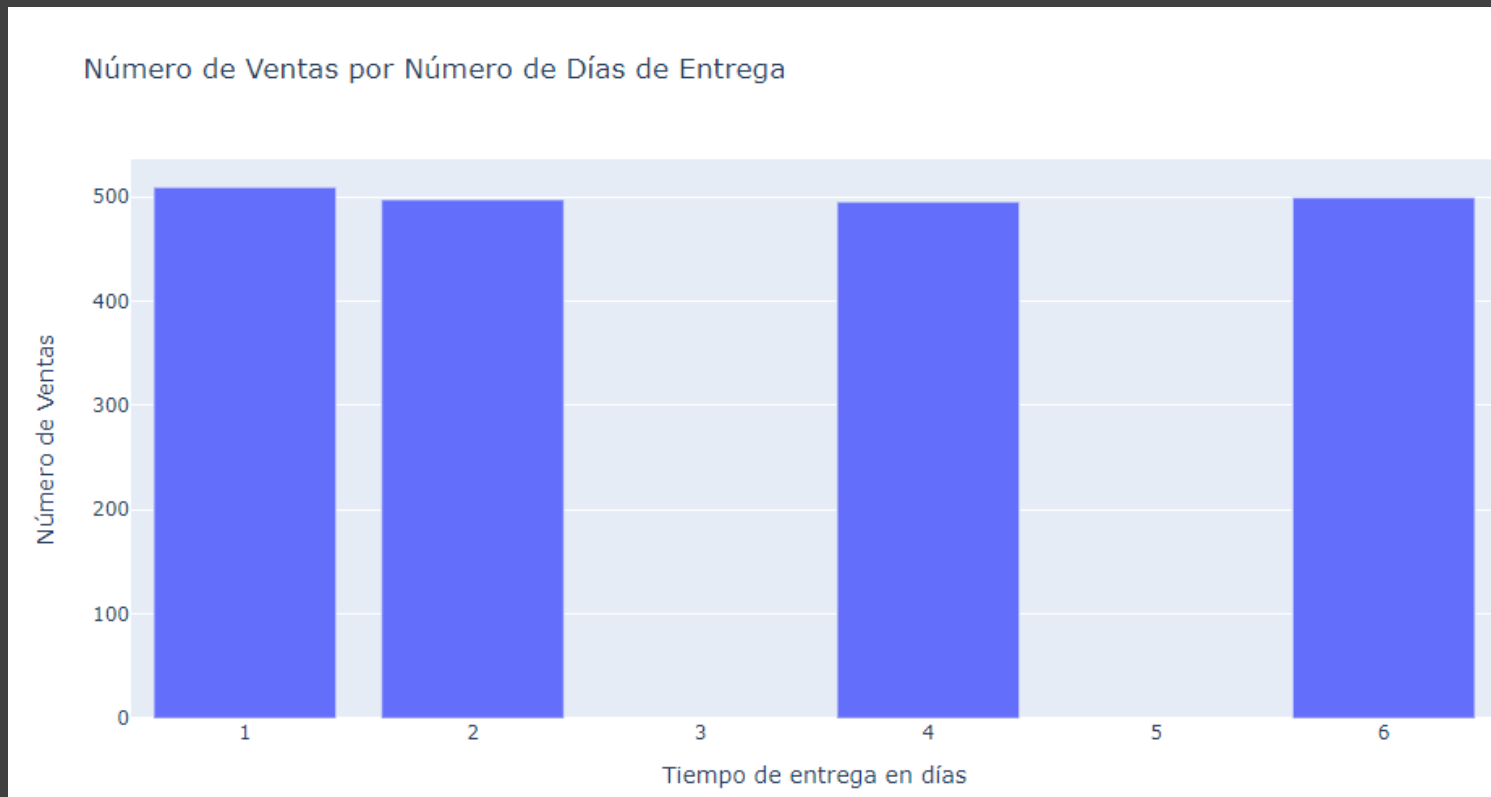
Experiencias obtenidas en la base de datos



# ANALISIS EXPLORATORIO

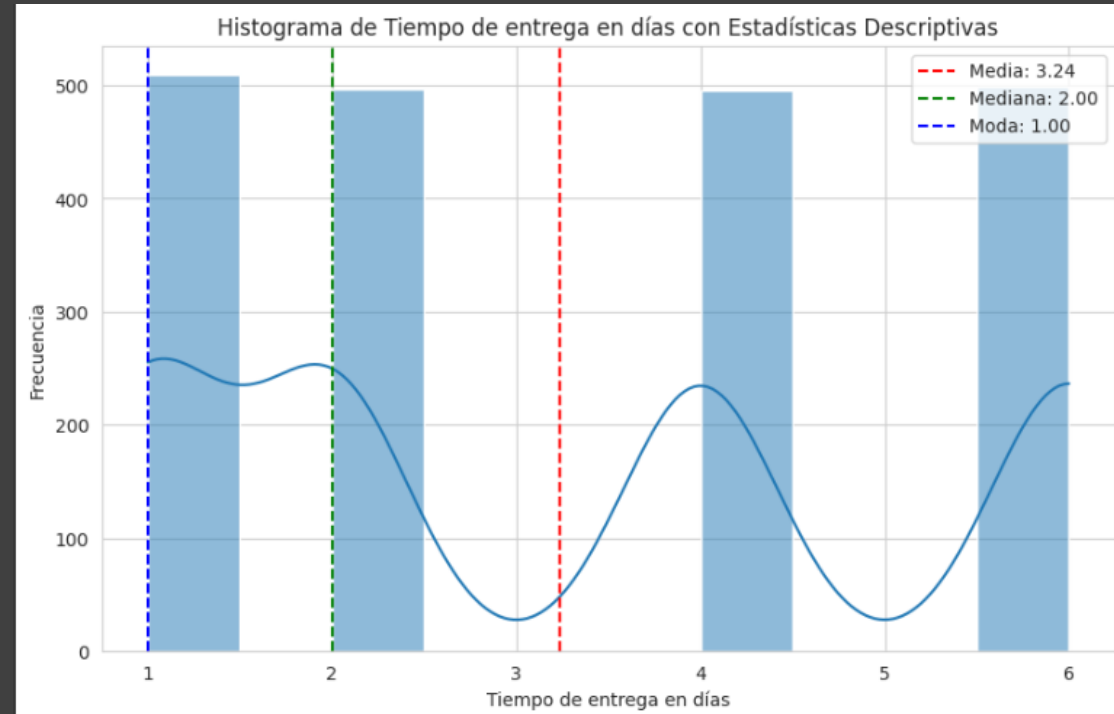
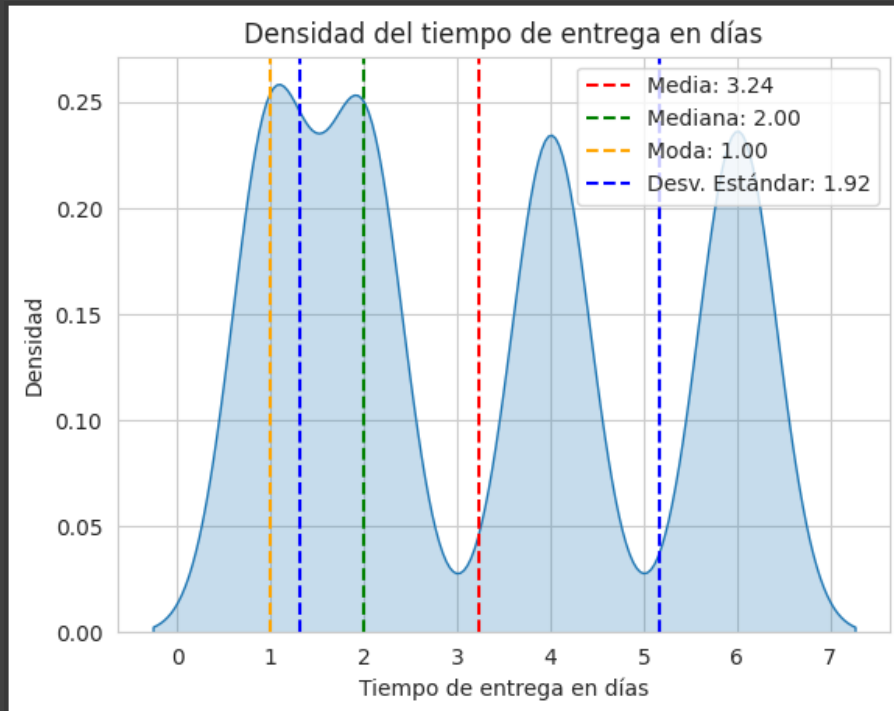
*Tiempo de entrega en días*

---



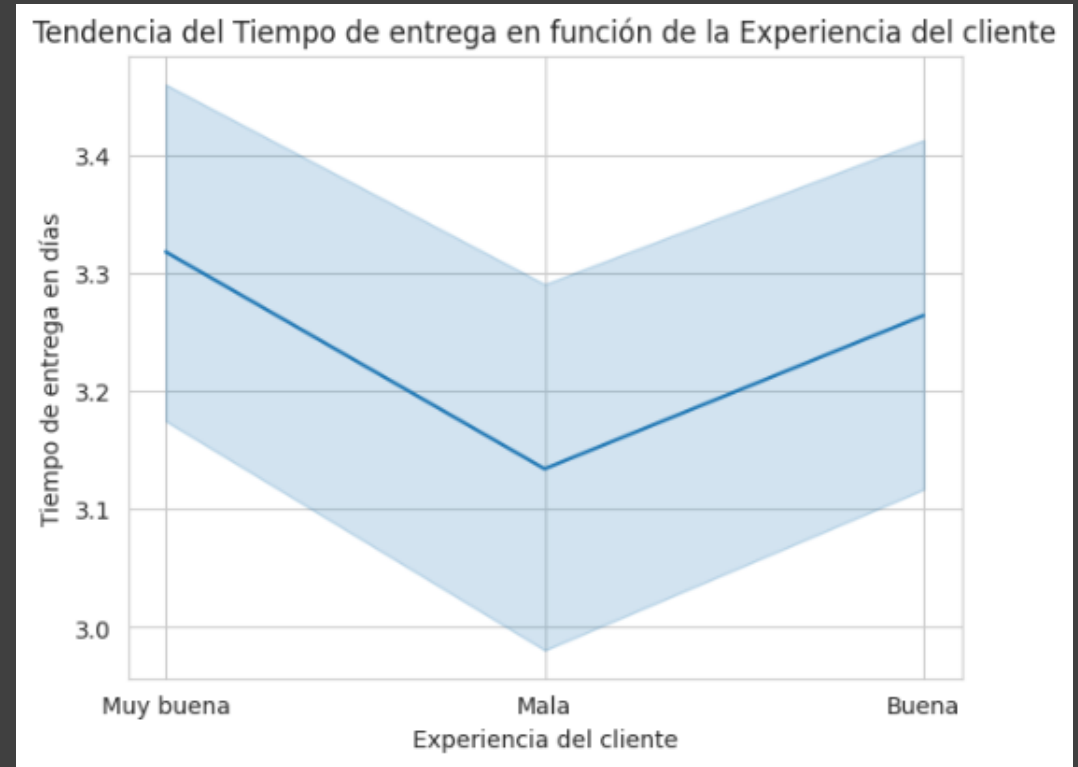
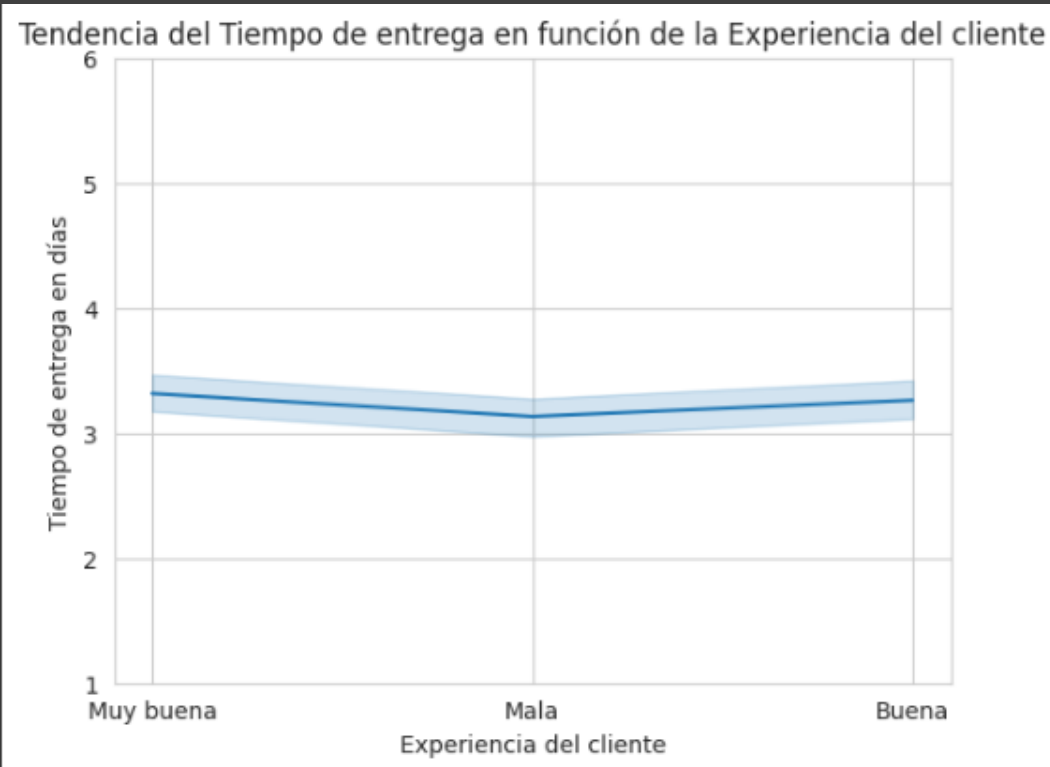
# ANÁLISIS EXPLORATORIO

*Análisis de la distribución de los datos de tiempo de entrega en días*



# ANALISIS EXPLORATORIO

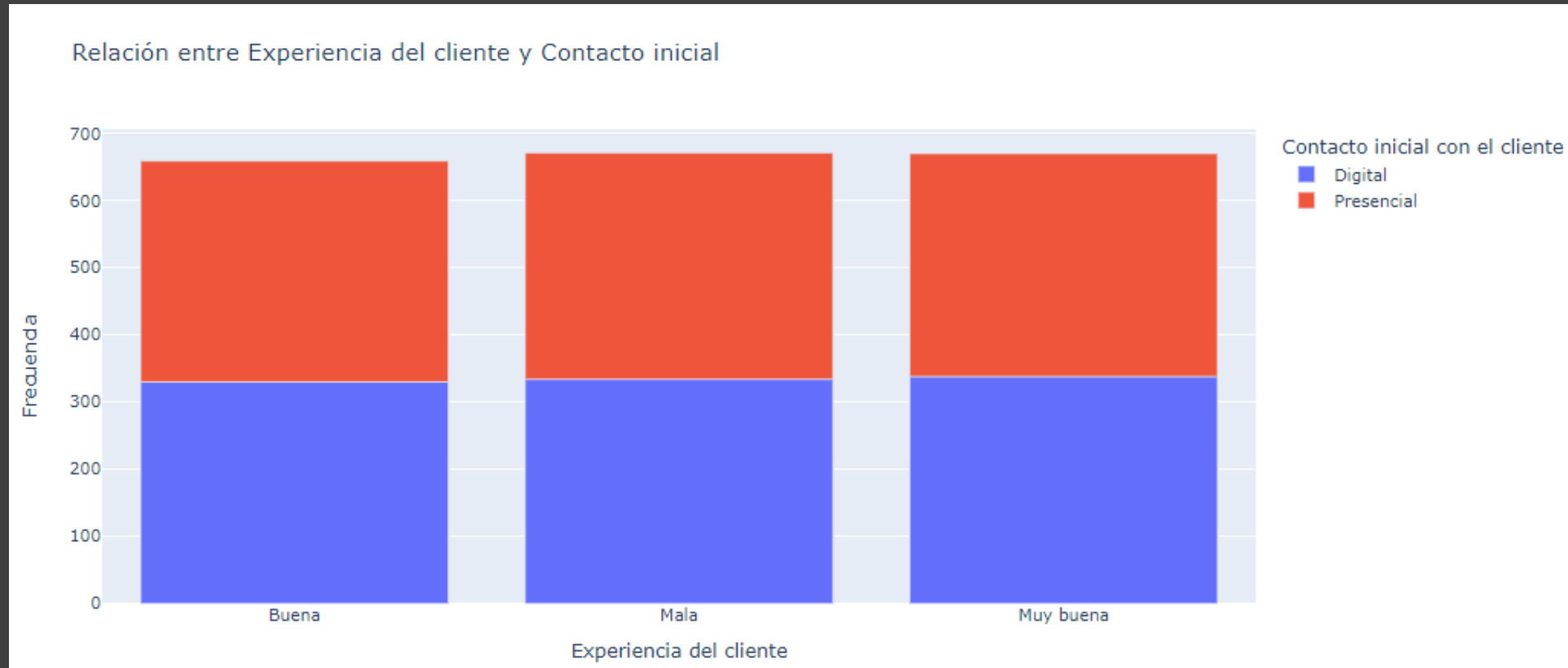
*Relación entre el tiempo de entrega en días y la experiencia del cliente*





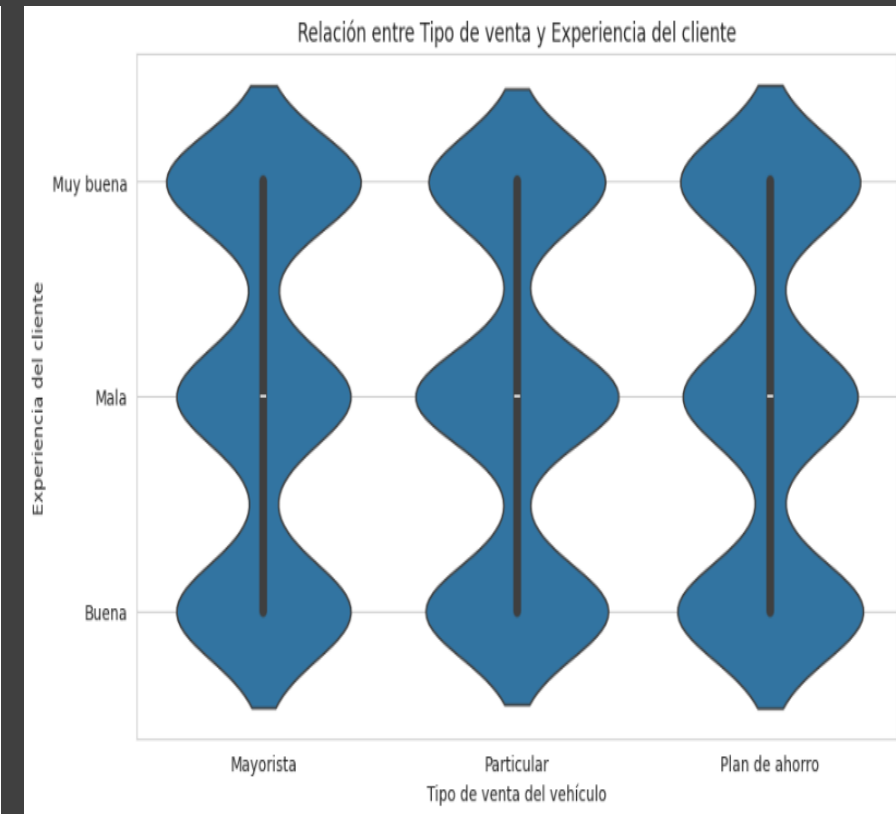
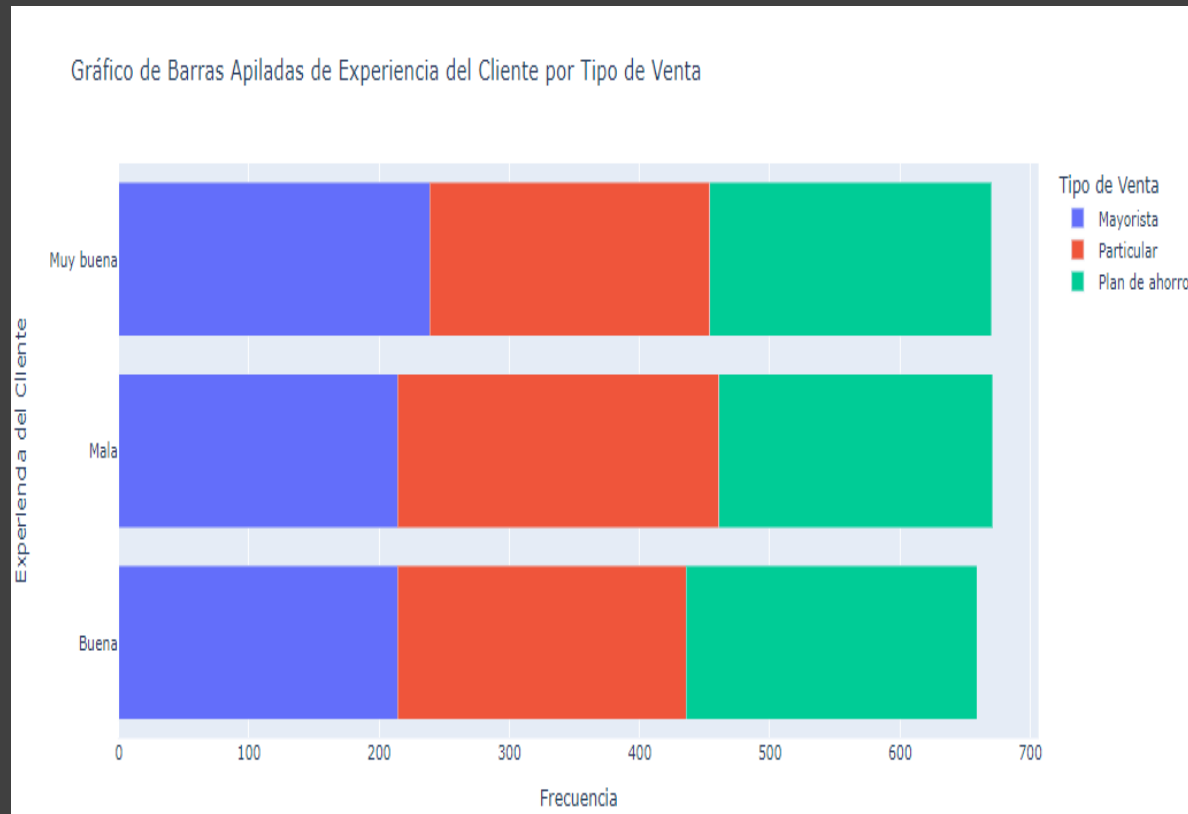
# ANALISIS EXPLORATORIO

*Relación entre experiencia del cliente y el primer contacto con el cliente, si fue presencial o digital*



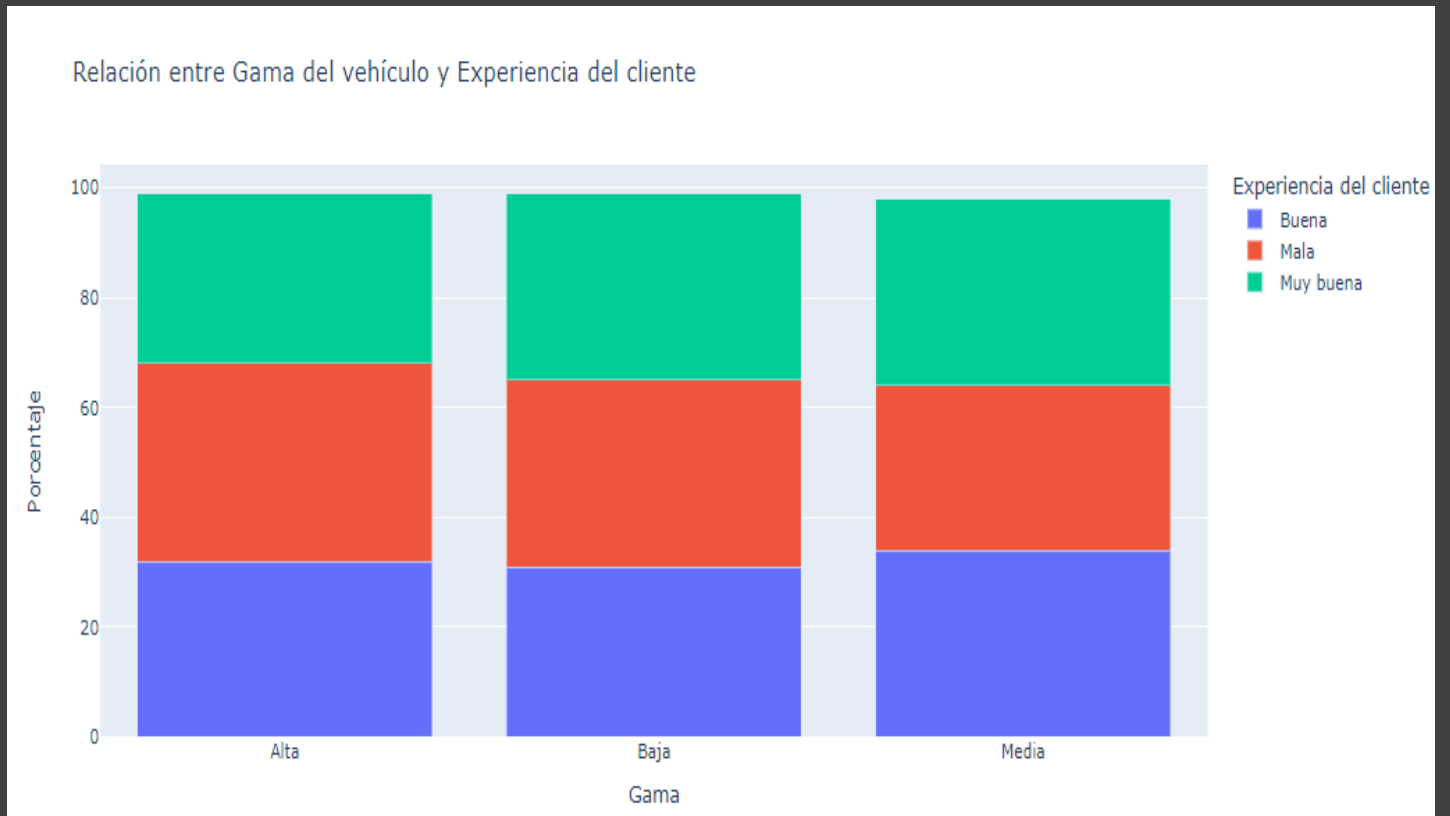
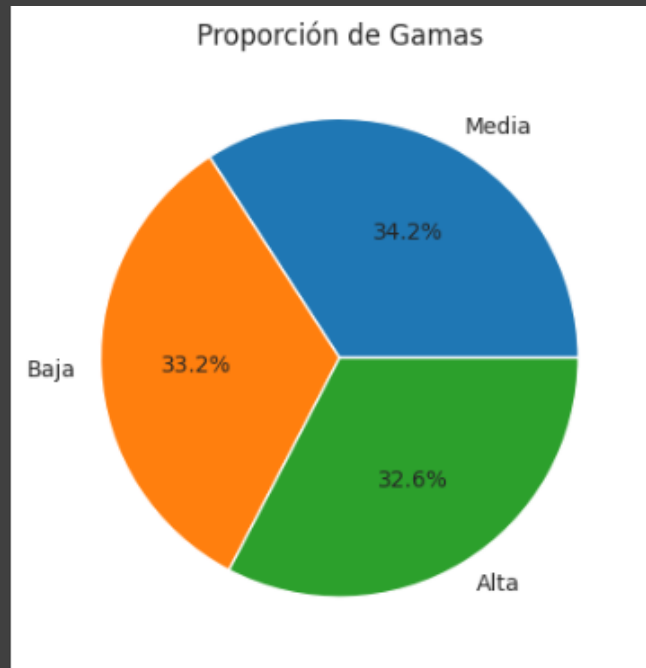
# ANALISIS EXPLORATORIO

*Relación entre el Tipo de venta y la Experiencia del cliente*

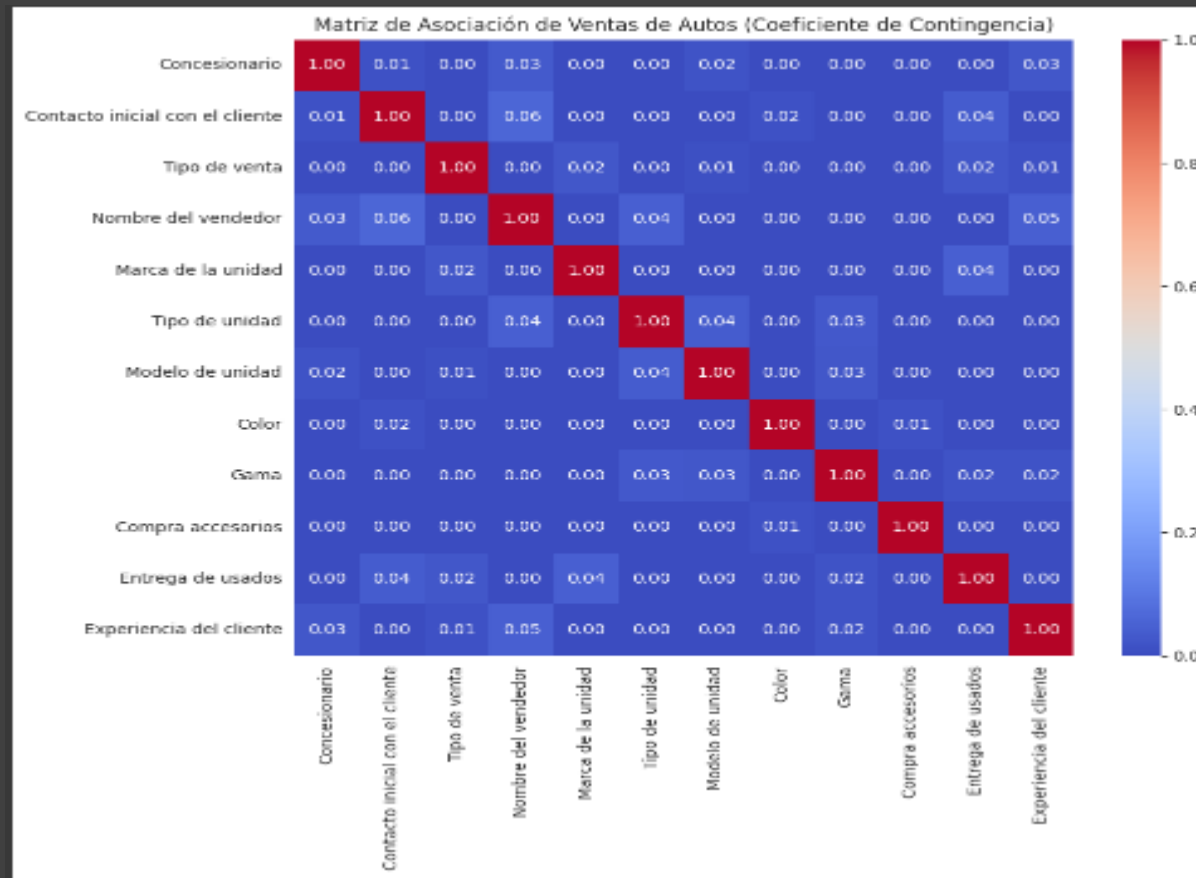


# ANALISIS EXPLORATORIO

*Relación entre el Tipo de venta y la Experiencia del cliente*

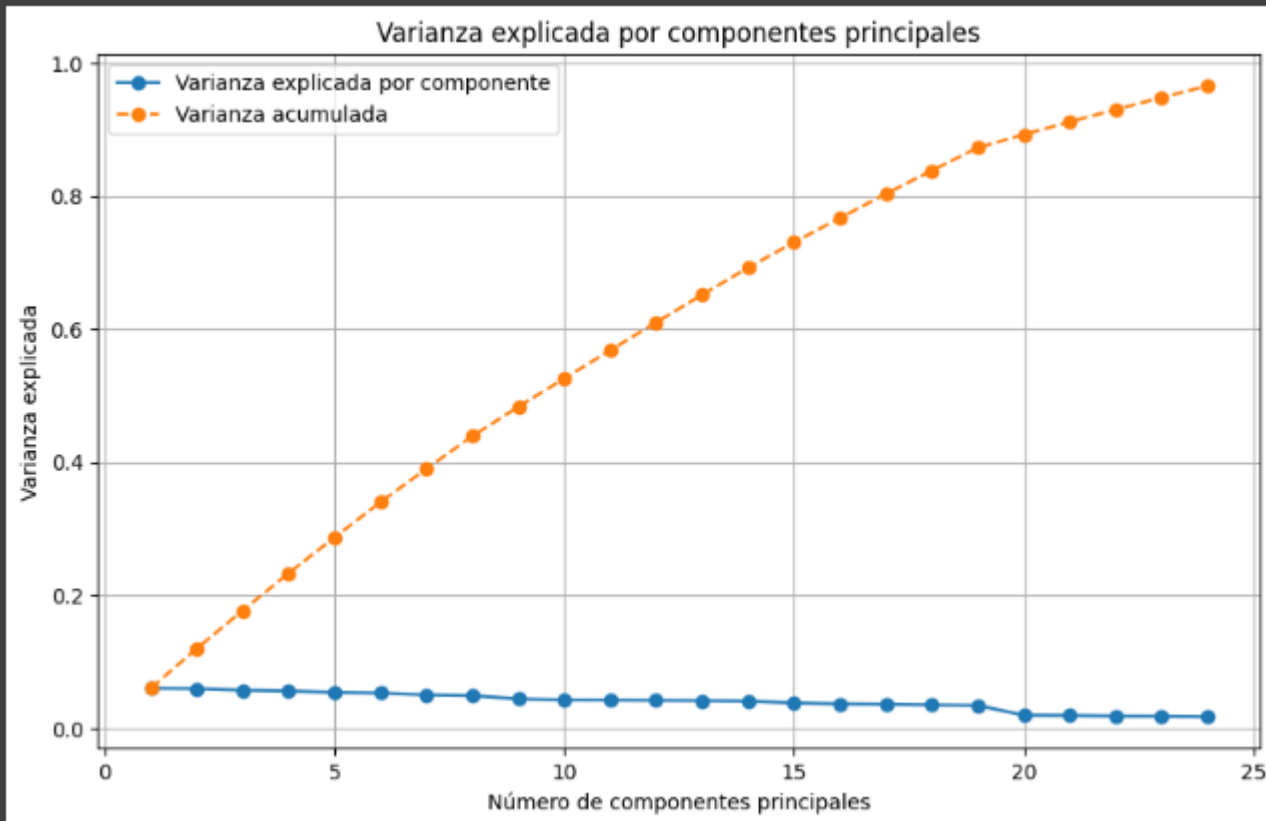


# ANALISIS BIVARIADO



Se puede observar que no existe una correlación LINEAL apreciable entre las columnas que componen a la base, habiendo eliminado las que no representan características de valor para el análisis

# MODELO MACHINE LEARNING



Podemos decir que con un número de 25 componentes se alcanza a comprender el 95% de los datos.

Al retener 25 componentes principales, se está conservando la mayoría de la información importante de los datos originales, lo que permite reducir significativamente la dimensionalidad de los datos mientras se mantiene una cantidad sustancial de la información original.

Esto puede ser útil para reducir el tiempo de procesamiento y mejorar el rendimiento del modelo de machine learning.

# MODELO MACHINE LEARNING

---

## Modelo Linear Regression

Mejores parámetros: {'fit\_intercept': True}

Precisión del modelo (MSE): 0.008710553309397445

Precisión del modelo (R2): 0.9609745144212132

## Modelo Ridge Regression

Mejores parámetros: {'alpha': 1, 'fit\_intercept': True, 'solver': 'sparse\_cg'}

Precisión del modelo (MSE): 0.008711221473415996

Precisión del modelo (R2): 0.9609715813770761

## Modelo Random Forest

Mejores parámetros: {'max\_depth': None, 'min\_samples\_leaf': 1, 'min\_samples\_split': 2, 'n\_estimators': 200}

Precisión del modelo (MSE): 0.012350968749999996

Precisión del modelo (R2): 0.9444169798372708

# Cross Validation de modelos

---

Modelo Linear Regression: Precisión del modelo (MSE): 0.0089 +/- 0.0004; Precisión del modelo (R2): 0.9600 +/- 0.0019

-----

Modelo Ridge Regression: Precisión del modelo (MSE): 0.0089 +/- 0.0004; Precisión del modelo (R2): 0.9600 +/- 0.0019

-----

Modelo Random Forest: Precisión del modelo (MSE): 0.0160 +/- 0.0042; Precisión del modelo (R2): 0.9299 +/- 0.0173

-----

Podemos observar que de los modelos utilizados, dos de ellos, el Ridge Regression y Linear Regression, tienen los mismos desempeños. Si bien el modelo de Random Forest presenta un buen desempeño, se encuentra opacado en comparación a los otros dos modelos utilizados.

Por lo cual se debe optar por uno de los de regresión. Para ello se toma como parámetro, el modelo más robusto, sería el caso de la regresión ridge. Esta se comporta mejor que la regresión lineal cuando hay multicolinealidad entre las características predictoras.

# Modelo Final

---

Mejores parámetros para Ridge Regression: {'alpha': 1, 'fit\_intercept': True, 'solver': 'sparse\_cg'}

Precisión del modelo (MSE) para Ridge Regression: 0.008711221473415996

Precisión del modelo (R2) para Ridge Regression: 0.9609715813770761

Precisión del modelo en datos de prueba: 1.000

Informe de clasificación:

	precision	recall	f1-score	support
0	1.00	1.00	1.00	139
1	1.00	1.00	1.00	129
micro avg	1.00	1.00	1.00	268
macro avg	1.00	1.00	1.00	268
weighted avg	1.00	1.00	1.00	268
samples avg	0.67	0.67	0.67	268



# Modelo Final

---

De los resultados obtenidos, podemos sacar las siguientes afirmaciones:

La precision del modelo alcanzo el 100% en los datos de prueba. Lo que significa que fue capaz de predecir correctamente si un cliente tendría una experiencia mala o muy buena en todos los casos evaluados.

Columna "Precision": La proporción de instancias clasificadas para la experiencia de los clientes, siendo mala o muy buena, la precisión es del 100% para ambas clases.

Columna "recall"(o sensibilidad): es la proporción de instancias positivas que fueron correctamente clasificadas por el modelo. Aquí también se obtuvo un recall del 100%, lo que significa que el modelo identificó correctamente todas las instancias positivas.

Columna F1-score: es la media armónica entre precision y recall. Podemos ver que es del 100% para ambas clases, lo que indica un buen equilibrio entre precision y recall.

En resumen, estos resultados sugieren que el modelo de predicción es altamente efectivo para clasificar si un cliente tendrá una experiencia mala o muy buena, al menos en el conjunto de datos de prueba utilizado

# Conclusiones

---

Dado que los resultados del modelo muestran una precisión del 100% en la predicción de si un cliente tendrá una buena experiencia en el mercado automotriz, podemos concluir lo siguiente:

Apoyo a la hipótesis alternativa (H1): La precisión del 100% del modelo sugiere que existe una diferencia significativa en la probabilidad de que un nuevo cliente tenga una buena experiencia en el mercado automotriz. Esto respalda la idea de que el modelo de predicción tiene un impacto significativo en la capacidad de predecir la experiencia de los nuevos clientes.

Rechazo de la hipótesis nula (H0): Dado que el modelo logra una precisión del 100%, podemos rechazar la hipótesis nula de que no hay diferencia en la probabilidad de que un nuevo cliente tenga una buena experiencia en el mercado automotriz. Los resultados indican claramente que el modelo sí tiene un impacto significativo en la capacidad de predicción.

En resumen, los resultados del modelo respaldan la idea de que el modelo de predicción utilizado tiene un alto nivel de precisión en la predicción de la experiencia de los nuevos clientes en el mercado automotriz, lo que sugiere que el modelo es efectivo y tiene un impacto significativo en la capacidad de predicción.