

Informe del Proyecto

Predicción de Precios de Aguacates

Este informe detalla el desarrollo de un modelo predictivo para estimar el precio promedio de los aguacates utilizando la base de datos de Avocado Prices.

El objetivo principal fue analizar los datos, preparar los datos, construir un modelo predictivo y evaluar su rendimiento, con el fin de ofrecer ideas valiosas para productores y comerciantes.

1. Descripción del problema

El precio promedio de los aguacates puede variar significativamente debido a factores como la variedad, la ubicación geográfica y el tipo de venta (convencional u orgánico). La pregunta central del proyecto fue: *¿Cómo se puede predecir el precio promedio de los aguacates utilizando características históricas?*

2. Análisis Exploratorio de Datos (EDA)

Comprensión de la Estructura de los Datos

Se realizó una revisión inicial del conjunto de datos, identificando variables clave, su tipo y la presencia de valores faltantes. Las principales variables incluían:

Fecha: Momento de la venta.

Tipo: Convencional u orgánico.

Variedad: Diferentes tipos de aguacates.

Precio: Precio promedio por unidad.

Identificación de Valores Faltantes y Distribuciones

Se identificaron algunos valores faltantes en variables categóricas. Además, se analizaron las distribuciones de los precios mediante histogramas y gráficos de barras, observando que el precio promedio seguía una distribución aproximadamente normal.

Visualizaciones

Se generaron gráficos para visualizar la relación entre la cantidad, tipo de aguacate, precio y volumen total. Esto ayudó a identificar patrones y posibles correlaciones.

3. Preparación de los Datos

Limpieza de Datos

Manejo de Valores Faltantes

Conversión de Fechas: Se transformó la columna de fecha a un formato numérico, calculando los días desde una fecha base.

Normalización y Codificación

Codificación de Variables Categóricas: Se utilizaron variables Dummy para convertir las categorías en formatos numéricos, facilitando el uso en el modelo.

División de Datos

Los datos se dividieron en conjuntos de entrenamiento y prueba (80/20) para permitir una evaluación adecuada del modelo.

4. Modelado Predictivo

Elección del Modelo

Se eligió un modelo de regresión lineal debido a su simplicidad y eficacia para problemas de predicción de variables continuas. Este modelo permite interpretar fácilmente la influencia de cada variable independiente en el precio promedio.

Entrenamiento del Modelo

Se utilizó Scikit-learn para construir y entrenar el modelo con el conjunto de datos de entrenamiento. Se ajustó el modelo a las variables independientes y al precio promedio.

5. Evaluación del Modelo

Métricas de Evaluación:

Se evaluó el rendimiento del modelo utilizando:

Error Cuadrático Medio (MSE): Para medir la precisión de las predicciones.

Coefficiente de Determinación (R^2): Para evaluar la proporción de la variabilidad en el precio que es explicada por el modelo.

Resultados del modelo

El modelo mostró un rendimiento aceptable, con un MSE razonable y un R^2 que indicaba que el modelo era capaz de capturar una parte significativa de la variabilidad en los precios de los aguacates.

6. Conclusiones y Recomendaciones

La pregunta inicial se centraba en predecir el precio promedio de los aguacates utilizando un conjunto de datos que incluye diversas variables, como la variedad del aguacate, la ubicación geográfica y el tipo de venta (convencional u orgánico).

A través del análisis y la modelización, se desarrolló un modelo de regresión lineal que logró capturar una parte significativa de la variabilidad en los precios de los aguacates. Las métricas evaluadas, como el Error Cuadrático Medio (MSE) y el coeficiente de determinación (R^2), indican que el modelo es capaz de realizar predicciones razonablemente precisas en relación con los precios históricos, indicando que el modelo es capaz de capturar una parte significativa de la variabilidad en los precios de los aguacates.

Conclusión: Con base en los resultados, se concluye que el modelo es efectivo para predecir el precio promedio de los aguacates, aunque hay margen para mejorar su precisión. Factores como la estacionalidad y otras variables económicas podrían tener un impacto significativo en el precio y deben ser considerados para optimizar el modelo.

En resumen, el modelo proporciona una herramienta útil para entender y predecir los precios de los aguacates, lo que puede ser valioso para productores, comerciantes y consumidores.

Recomendaciones:

Explorar Modelos Alternativos: Se sugiere experimentar con otros algoritmos de aprendizaje automático, como árboles de decisión, bosques aleatorios o redes neuronales, que podrían capturar relaciones no lineales más complejas en los datos.

Incluir Más Variables: Considerar la inclusión de variables adicionales que podrían influir en los precios, tales como factores estacionales, condiciones climáticas, cambios en la oferta y demanda, o datos económicos relevantes. Esto podría mejorar la capacidad predictiva del modelo.

Validación Cruzada: Implementar técnicas de validación cruzada para evaluar la robustez del modelo y prevenir el sobreajuste.

7. Despliegue del Modelo (Opcional)

Estrategias de Despliegue: Para llevar el modelo a un entorno productivo, se recomienda utilizar frameworks como Flask o FastAPI para desarrollar una API que permita a los usuarios ingresar datos y recibir predicciones en tiempo real. Esto facilitaría la integración del modelo en aplicaciones web o móviles.

Plataformas de Nube: Considerar el uso de plataformas en la nube como Heroku o AWS para alojar la aplicación. Esto proporcionaría escalabilidad y accesibilidad, permitiendo que múltiples usuarios accedan al modelo simultáneamente.

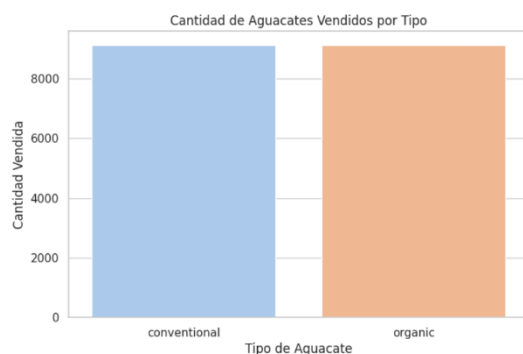
Interfaz de Usuario: Desarrollar una interfaz de usuario amigable que simplifique la interacción con el modelo y mejore la experiencia del usuario, permitiendo a los no expertos ingresar datos y obtener predicciones de manera intuitiva.

Monitoreo y Mantenimiento: Implementar un sistema de monitoreo para evaluar el rendimiento del modelo en producción, así como un plan de mantenimiento regular para actualizar el modelo con nuevos datos y mejorar su precisión a lo largo del tiempo.

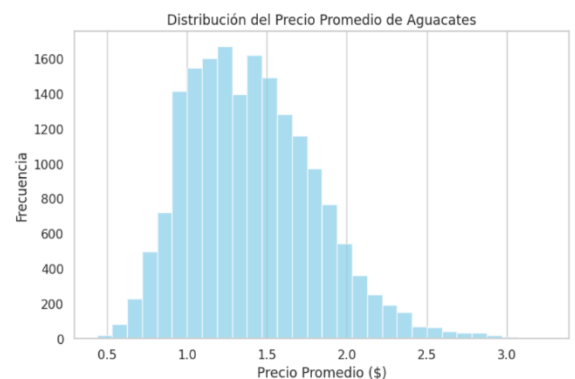
8. COLAB/PYTHON

[Proyecto Final modulo PYTHON - Colab \(google.com\)](#)

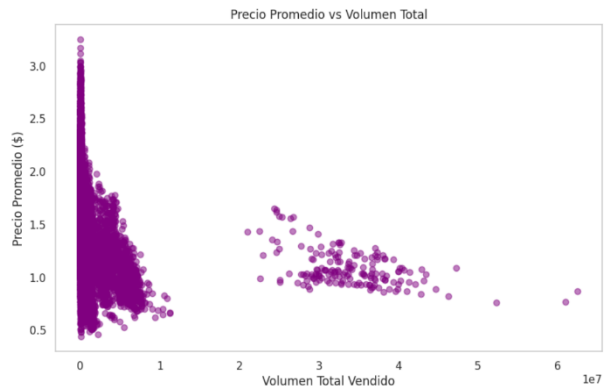
Gráficos principales del proyecto para visualización de los datos



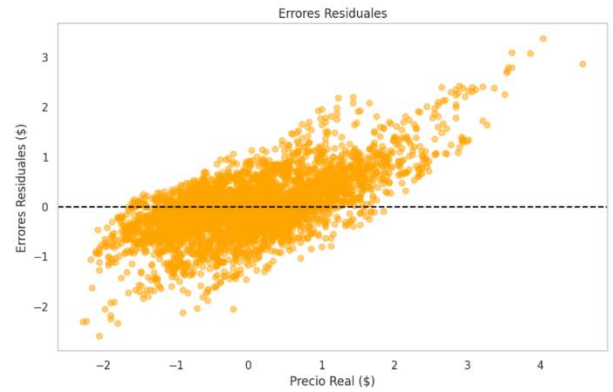
Gráfica 1. Cantidad de aguacates vendidos por tipo de aguacate



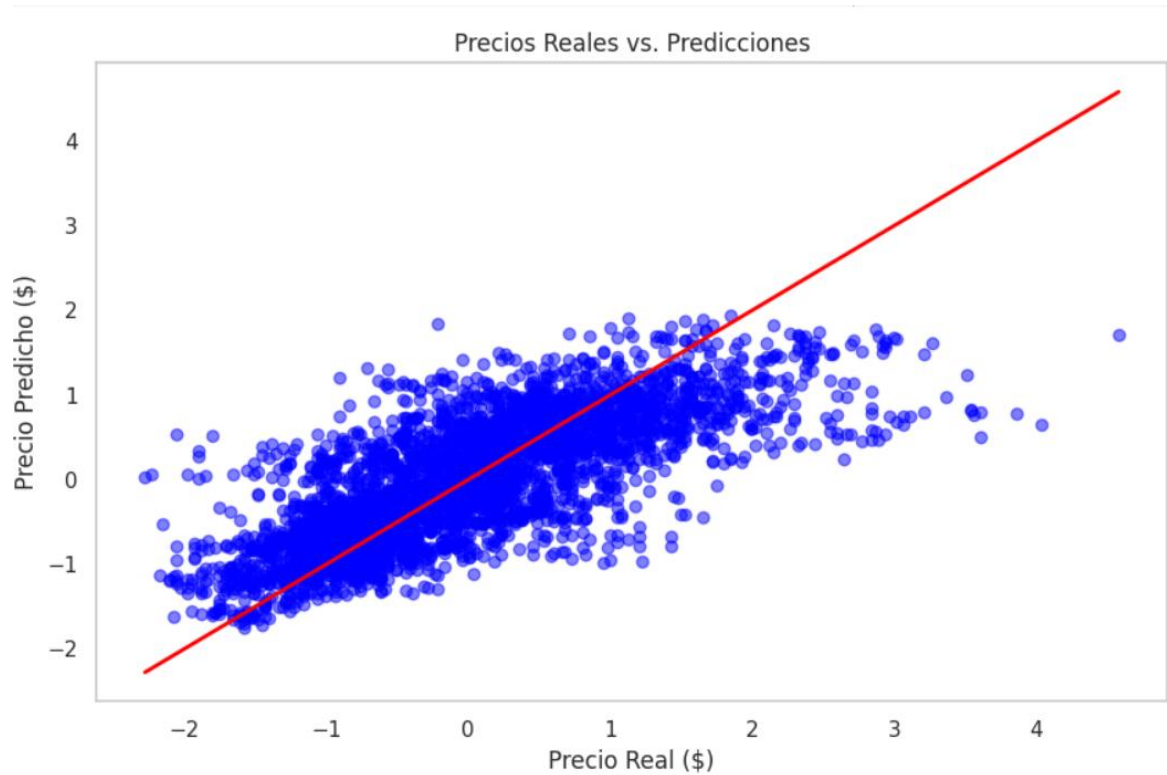
Gráfica 2. Distribución del precio promedio de aguacates



Gráfica 3. Precio promedio por volumen total



Gráfica 4. Errores residuales



Gráfica 5. Precios reales vs. Predicciones