

ANÁLISIS EXHAUSTIVO DEL DATASET “PET_STORE_RECORDS_2020.CSV”

TIPO DE PROYECTO:

Tecnológico y de Investigación

ESPACIO CURRICULAR/MÓDULO:

- Estadística y exploración de datos II
- Ciencia de Datos-II

EJES | UNIDADES CONCEPTUALES:

- Ciencia de datos II:** Numpy, Pandas, Introducción al Machine Learning, Correlación, Regresión lineal, lineal múltiple y logística.
- Estadística II:** Estimación de parámetros, contraste de hipótesis, Test de ANOVA, correlación, Regresión lineal, lineal múltiple y logística.

PROBLEMÁTICA | NECESIDAD | CASO:

El caso de estudio desarrollado a lo largo de las evidencias de aprendizaje y este proyecto consta en, a partir de un dataset, aplicar las técnicas estadísticas y los algoritmos estudiados para la obtención de información a través de los datos, así como el uso de los mismos para la inferencia de escenarios.

OBJETIVO GENERAL:

Analizar y modelar el comportamiento de las ventas de un petshop mediante la aplicación de herramientas estadísticas y técnicas de ciencia de datos, con el propósito de identificar los factores que influyen en las ventas, validar hipótesis sobre las relaciones entre variables y desarrollar modelos predictivos que aporten valor al proceso de toma de decisiones comerciales.

OBJETIVOS ESPECÍFICOS:

- Explorar y depurar el dataset de ventas
- Realizar un análisis descriptivo
- Evaluar la normalidad y homocedasticidad
- Aplicar análisis de varianza (ANOVA)
- Analizar la correlación entre variables cuantitativas
- Desarrollar y ajustar modelos de regresión lineal y múltiple
- Interpretar los resultados obtenidos
- Integrar los conocimientos teóricos de estadística y ciencia de datos

FUNDAMENTACIÓN | HIPÓTESIS:

El proyecto tiene como objetivo aplicar de manera integrada los conocimientos de Estadística II y Ciencia de Datos II en un caso real, utilizando un dataset de ventas de una tienda de mascotas. A partir de variables como precios, categorías y cantidades vendidas, se analizan patrones de consumo y comportamientos del mercado desde una perspectiva cuantitativa. Este trabajo busca transformar los datos en información útil para la toma de decisiones comerciales, optimizando recursos y mejorando la comprensión del negocio. Además, demuestra la importancia del análisis de datos como herramienta clave para generar conocimiento aplicable en contextos reales y fortalecer la formación profesional en ciencia de datos.

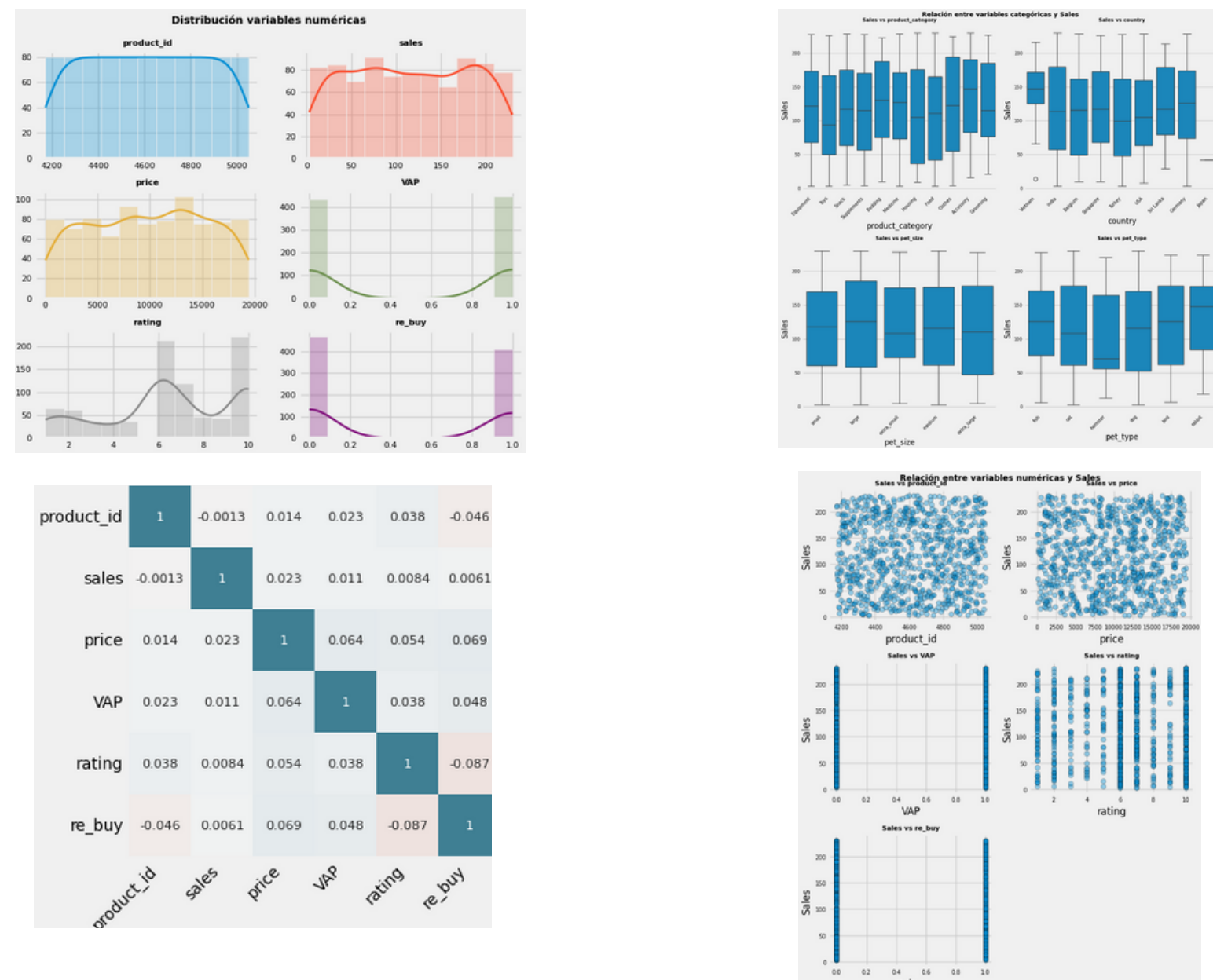
CONCLUSIONES | RESULTADOS ESPERADOS:

El proyecto permitió consolidar los aprendizajes del cuatrimestre mediante una experiencia práctica que integró análisis, interpretación y toma de decisiones basadas en datos reales. Más allá de los resultados obtenidos, fue una oportunidad para comprender cómo los datos pueden transformarse en información valiosa para explicar fenómenos y apoyar la gestión de un negocio, fortaleciendo la capacidad analítica y el pensamiento crítico. A lo largo del proceso se desarrollaron habilidades clave como la organización, la planificación, la resolución de problemas y el trabajo colaborativo. La experiencia del ABP no solo permitió aplicar los contenidos de la carrera en un contexto real, sino que también potenció competencias transversales como la comunicación de resultados y la toma de decisiones fundamentadas en evidencia, convirtiéndose en una instancia de aprendizaje y crecimiento profesional significativo.

ACCIONES | RECURSOS | TIEMPO:

Objetivo Específico	Acciones	Recursos	Tiempos Estimados
Explorar y depurar el dataset	<ul style="list-style-type: none">Importar el dataset desde CSV.Identificar valores nulos, duplicados o inconsistentes.Normalizar formatos de datos y eliminar registros atípicos no representativos.	Python (Pandas, Numpy), Google Colab, Dataset Kaggle.	Semana 1
Realizar un análisis descriptivo	<ul style="list-style-type: none">Calcular estadísticas descriptivas (media, mediana, desviación estándar).Generar visualizaciones (boxplots, histogramas, gráficos de barras).Describir patrones generales y distribución de variables.	Pandas, Matplotlib, Seaborn.	Semana 2
Evaluar la normalidad y homocedasticidad	<ul style="list-style-type: none">Aplicar tests de Shapiro-Wilk y Levene.Utilizar Q-Q plots para contrastar visualmente la normalidad.Determinar la persistencia de pruebas paramétricas.	Scipy, Statsmodels, Matplotlib.	Semana 2
Aplicar análisis de varianza (ANOVA)	<ul style="list-style-type: none">Realizar ANOVA de una vía para comparar montos de venta por categoría.Aplicar ANOVA de dos vías considerando tipo de mascota y categoría.Interpretar los valores p y diferencias.	Statsmodels, Scipy.	Semana 3
Analizar la correlación entre variables cuantitativas	<ul style="list-style-type: none">Calcular coeficientes de correlación (Pearson o Spearman según distribución).Graficar matrices de correlación para visualizar relaciones entre variables.	Pandas, Seaborn (heatmap).	Semana 3
Desarrollar y ajustar modelos de regresión lineal y múltiple	<ul style="list-style-type: none">Seleccionar variables predictoras relevantes.Ajustar modelos lineales y evaluar R² y significancia de los coeficientes.Validar supuestos de los modelos (residuos, colinealidad).	Statsmodels, Scikit-learn, Python.	Semana 4
Interpretar los resultados obtenidos	<ul style="list-style-type: none">Analizar los hallazgos estadísticos en relación con el contexto comercial.Redactar conclusiones sobre los factores que más influyen en las ventas.Traducir los resultados técnicos a un lenguaje gerencial.	Informe final, gráficos de apoyo, conocimientos teóricos.	Semana 5
Integrar los conocimientos teóricos de estadística y ciencia de datos	<ul style="list-style-type: none">Sistematizar los resultados en un documento final.Relacionar los métodos utilizados con los contenidos de ambas materias.Preservar la síntesis del proceso en una exposición oral.	Google Colab, PowerPoint/Canva, documentación del equipo.	Semana 6

PRODUCTO FINAL



REPOSITORIO: [HTTPS://GITHUB.COM/GRUPO-11-CDIA/CIENCIA-DE-DATOS](https://github.com/grupo-11-cdia/ciencia-de-datos)

INTEGRANTES: Cáceres Giménez, Cesia Fiorella
Di Campli, Gastón

Lorenzati, Valentino
Menón, Nicolás
Terreno, Alejo

DOCENTE: NAHUEL PRATTA, MARCOS UGARTE

TECNICATURA: TECNICATURA SUPERIOR EN CIENCIAS DE DATOS E INTELIGENCIA ARTIFICIAL

COHORTE: 2024