

TECNICATURA SUPERIOR EN

Ciencia de Datos e Inteligencia Artificial

Módulo Análisis de datos

Proyecto Final

1. NOMBRE DEL PROYECTO: Análisis descriptivo y visualización de ventas de PepsiCo en Argentina.

Estudio basado en datos recolectados de distribuidores a nivel nacional durante un período determinado.

Integrantes:

- Eglimar Ramirez
- Francisco Oviedo
- Jonathan Guillen
- Jonathan Manuel Palomeque
- Lucas Ledesma
- Maia Majzum
- Valentina Pich

2. TIPO DE PROYECTO: Tecnológico.
3. ESPACIO CURRICULAR: Análisis de datos (Procesamiento y Estadísticas).
4. EJES TEMÁTICOS/ RED DE CONCEPTOS:

1. Procesamiento de Datos

● Descripción:

Análisis y transformación de datos crudos utilizando bibliotecas especializadas como Pandas. El trabajo incluye la importación, limpieza, normalización y estructuración de conjuntos de datos reales para facilitar el análisis estadístico.

● Competencias a fortalecer:

- Capacidad para manipular grandes volúmenes de datos con eficiencia.
 - Manejo de entornos como Jupyter Notebook y Google Colab.
 - Aplicación de buenas prácticas en la documentación y organización de los datos.
-

2. Estadística Descriptiva y Exploración de Datos

● Descripción:

Aplicación de técnicas de estadística descriptiva para obtener conclusiones significativas sobre los datos. Se trabaja con medidas de tendencia central, dispersión, tablas de frecuencia y gráficos estadísticos para interpretar y comunicar resultados.

● Competencias a fortalecer:

- Interpretación de variables cuantitativas y cualitativas.
 - Aplicación de análisis descriptivos e inferenciales.
 - Evaluación de la calidad de los datos y análisis de valores faltantes o atípicos.
-

3. Programación y Automatización del Análisis

- **Descripción:**

Implementación de scripts en Python para automatizar el análisis, filtrado y visualización de datos. Se utilizan librerías como Pandas, NumPy y Matplotlib para desarrollar soluciones replicables y escalables.

- **Competencias a fortalecer:**

- Estructuración de código limpio y modular.
 - Integración de análisis estadístico con programación.
 - Capacidad para reutilizar y documentar código de análisis.
-

4. Visualización y Comunicación de Resultados

- **Descripción:**

Elaboración de gráficos y reportes visuales que permitan comunicar resultados de forma clara, precisa y comprensible. Uso de gráficos de barras, histogramas, diagramas de caja, entre otros.

- **Competencias a fortalecer:**

- Selección adecuada de visualizaciones según tipo de variable.
 - Uso de Matplotlib, Seaborn o herramientas similares.
 - Redacción de conclusiones coherentes con los datos representados.
-

5. PROBLEMÁTICAS/ NECESIDADES:

En el contexto actual, las empresas multinacionales como PepsiCo enfrentan el desafío de comprender con precisión el rendimiento de sus productos en distintas regiones del país, especialmente a través de su red de distribuidores. La falta de una visión integrada y depurada de los datos de ventas dificulta la toma de decisiones estratégicas respecto a stock, distribución, campañas comerciales o desarrollo de nuevos productos. Desde el rol de analistas de datos, surge la necesidad de limpiar, estructurar y analizar estos datos para detectar patrones de consumo, identificar productos destacados y comparar resultados entre zonas geográficas.

6. FUNDAMENTACIÓN:

Este proyecto representa una problemática real del sector comercial de consumo masivo: la necesidad de transformar grandes volúmenes de datos en información útil.

El potencial del análisis de este conjunto de datos reside en su capacidad para relevar el comportamiento de ventas a nivel nacional, reconociendo diferencias regionales y productivas.

Es altamente relevante para nuestro perfil profesional porque nos permite aplicar competencias de análisis descriptivo, programación, procesamiento y visualización de datos, tal como exige el campo de la Ciencia de Datos en entornos reales. Además, tiene impacto en la comunidad al facilitar mejores decisiones logísticas y comerciales que pueden influir en el acceso a productos, la eficiencia en el consumo de recursos y la reducción de costos.

7. VISIÓN DEL PROYECTO:

Construir un análisis integral y visualmente comprensible de las ventas de productos PepsiCo a lo largo de las diferentes regiones de Argentina, a partir de un dataset sin depurar, utilizando herramientas de procesamiento de datos y estadística descriptiva. El proyecto aspira a consolidar un enfoque metodológico profesional para abordar

problemas reales con datos reales, reforzando tanto el criterio técnico como el sentido crítico para interpretar información.

-OBJETIVOS:

- **Objetivo general:**

Analizar y visualizar los datos de ventas de productos PepsiCo distribuidos en Argentina, a partir de un dataset sin depurar, aplicando herramientas de procesamiento de datos y estadística descriptiva para detectar patrones relevantes y comunicar resultados útiles para la toma de decisiones.

- **Objetivos específicos:**

- 1- Limpiar y estructurar el conjunto de datos para su posterior análisis.
- 2- Analizar las variables numéricas y categóricas y detectar valores nulos o mal cargados.
- 3- Calcular y representar medidas estadísticas descriptivas de las ventas por producto, región y rubro.
- 4- Detectar productos más vendidos y su distribución por área geográfica.
- 5- Desarrollar visualizaciones efectivas que permitan comunicar los hallazgos de forma clara.

- **Metas:**

-Asegurar la integridad y consistencia de los datos.

-Completar el proyecto en el plazo estipulado.

-Analizar datos de ventas del último mes de 20 distribuidores de diferentes zonas geográficas del país.

8. SELECCIÓN DE ACCIONES

Objetivos específicos	Acciones
Limpiar y estructurar el conjunto de datos original.	<ul style="list-style-type: none"> - Detectar y corregir valores nulos o erróneos. - Estandarizar nombres de campos. - Organizar las columnas para facilitar análisis.
Calcular medidas estadísticas descriptivas.	<ul style="list-style-type: none"> - Analizar media, mediana y desvío estándar de Cantidad por producto y región - Tabular totales por Rubro y Línea
Detectar productos más vendidos por región	<ul style="list-style-type: none"> - Agrupar datos por Area y Descripción - Ordenar por Cantidad descendente para obtener rankings
Desarrollar visualizaciones	<ul style="list-style-type: none"> - Crear gráficos de barras, tortas o mapas con Matplotlib / Seaborn - Incorporar insights con textos explicativos en el notebook

9. PRODUCTO FINAL

1. Análisis del contexto

En el contexto actual, las empresas multinacionales como PepsiCo enfrentan el desafío de interpretar grandes volúmenes de datos provenientes de sus redes de distribución. La falta de una estructura unificada para recolectar, depurar y analizar esta información dificulta la toma de decisiones estratégicas en aspectos clave como la gestión de inventario, las campañas comerciales o el lanzamiento de nuevos productos. Este proyecto surge de la necesidad de transformar un dataset crudo de ventas en una fuente útil de información para el análisis comercial.

2. Problemas identificados

- **Datos desorganizados:** La información original contenía errores de carga, valores faltantes, nombres mal escritos y estructuras inconsistentes.
- **Falta de visualización clara:** No existían reportes visuales que permitieran interpretar el comportamiento de los productos en las distintas regiones del país.
- **Dificultad para identificar patrones:** Sin un análisis estadístico adecuado, era difícil distinguir los productos más vendidos, su distribución geográfica o su desempeño por línea y rubro.

3. Solución propuesta

Se propuso desarrollar un cuaderno de trabajo en **Jupyter Notebook**, basado en Python, que permita importar el dataset crudo, limpiarlo, procesarlo y generar análisis estadísticos y visuales. Este enfoque permite detectar patrones de consumo, identificar productos destacados y analizar los resultados por zona geográfica.

El objetivo principal fue ofrecer una herramienta accesible, visual y replicable, capaz de acompañar la toma de decisiones de la empresa desde una perspectiva basada en datos.

4. Herramientas tecnológicas utilizadas

- **Lenguaje de programación:** Python
- **Entorno de desarrollo:** Jupyter Notebook
- **Librerías:**
 - Pandas para limpieza y análisis de datos
 - NumPy para operaciones estadísticas
 - Matplotlib y Seaborn para la generación de gráficos

5. Desarrollo

a. Carga y depuración de datos

Se importó el archivo Excel provisto, se exploró su estructura y se realizaron operaciones para corregir errores frecuentes: columnas mal nombradas, valores nulos, registros duplicados, y formatos inconsistentes.

b. Análisis estadístico descriptivo

Se calcularon medidas como la media, mediana y desviación estándar de las distintas variables y Análisis de simetría y asimetría según su distribución.

c. Visualización de datos.

Se generaron gráficos de barras horizontales por Rubro, gráficos de torta por región y gráficos de dispersión que ilustran el comportamiento de ventas según distintas variables categóricas. Esto permitió identificar claramente qué productos se destacan, qué zonas concentran las mayores ventas y qué líneas presentan mayor regularidad.

d. Interpretación y conclusiones

A partir de los análisis, se extrajeron observaciones claves que fueron acompañadas de recomendaciones estratégicas. Se detectó la concentración de ventas ciertos productos en regiones específicas, y la existencia de líneas con gran volumen de ventas que podrían potenciarse aún más.

El informe documental que acompaña al código incluye:

- Descripción metodológica.
- Justificación de decisiones de limpieza y visualización.
- Conclusiones sobre el comportamiento de ventas y recomendaciones.
- Repositorio en GitHub con la documentación del proceso y el trabajo final.

7. Articulación curricular

Desde el espacio **Procesamiento de Datos**, se aplicaron los siguientes contenidos:

- Uso de estructuras *Series* y *DataFrames*.
- Transformación de datos con **pandas**.
- Limpieza avanzada, combinación y análisis exploratorio.

Desde el espacio **Estadística y Exploración de Datos**, se trabajaron:

- Medidas descriptivas (media, mediana, moda, varianza, rango).
- Gráficos estadísticos (torta, barras, histogramas).
- Clasificación de variables y escalas de medición.
- Análisis de simetría, curtosis y forma de la distribución.

8. Vinculación comunitaria (enfoque comercial/marketing)

Para cerrar el ciclo ABP, se propone como producto socialmente útil la comunicación visual de los hallazgos a través de:

- Infografías para redes sociales, que representen:
 - Tendencias de consumo.
 - Rubros más vendidos por zona.
 - Cambios estacionales en ventas.
- **Folletos digitales para pymes y emprendedores**, orientados a:
 - Entender cómo leer sus datos.
 - Optimizar decisiones comerciales a partir de patrones detectados.

9. Conclusión

El proyecto logró cumplir con los objetivos planteados, entregando un cuaderno de análisis de datos funcional, claro y adaptado a una problemática real de la empresa. A través del uso de herramientas tecnológicas y conocimientos de estadística descriptiva, se transformó un conjunto de datos sin depurar en información útil y visualmente comprensible.

Este trabajo permitió no solo fortalecer habilidades técnicas en programación y análisis de datos, sino también adquirir una mirada crítica sobre el proceso de toma de decisiones basado en evidencia. El producto final es una base sólida para futuras extensiones, como la predicción de ventas, segmentación de clientes o automatización de reportes mensuales.

10. BIBLIOGRAFIA

Bibliografía general del módulo:

- Berenson, M. L., Levine, D. M., & Krehbiel, T. C. (2001). *Estadística para Administración* (2.ª ed.). Pearson Educación.
- Cantatore de Frank, N. (1980). *Manual de Estadística Aplicada* (Tomo II). Editorial Hemisferio Sur.
- Cignacco, G., Craveri, A., & Terán, T. (2008). *Estadística descriptiva y probabilidad con aplicaciones en las ciencias veterinarias*. Editorial UNR.
- Cignacco, G., Craveri, A., & Terán, T. (2009). *Inferencia estadística con aplicaciones*. Editorial UNR.
- Cochran, W., & Cox, G. (1978). *Diseños experimentales*. Editorial Trillas, México
- Hunter, J. D. (2007). Matplotlib: A 2D graphics environment. *Computing in Science & Engineering*, 9(3), 90–95. <https://doi.org/10.1109/MCSE.2007.55>
- McKinney, W. (2022). *Python for Data Analysis: Data Wrangling with Pandas, NumPy, and Jupyter* (3rd ed.). O'Reilly Media.
- The Matplotlib Development Team. (s/f). *Matplotlib Documentation*. <https://matplotlib.org/>
- The Pandas Development Team. (s/f). *Pandas Documentation*. <https://pandas.pydata.org/>
- Triola, M. F. (2009). *Estadística* (10.ª ed.). Pearson Educación.
- VanderPlas, J. (2017). *Python Data Science Handbook*. O'Reilly Media.