

Laboratorio 1

Sesión #1 Componente Práctico

Título del Laboratorio: Aplicación de los conceptos del análisis de los datos.

Duración: 2 horas

Objetivos del Laboratorio:

1. *Aplicar lo conceptos vistos en clase con cuestionarios y casos reales.*
2. *Identificar en el dataset los tipos de datos.*
3. *Realizar con un dataset la exploración, análisis y descripción de los datos.*

Objetivos del Laboratorio:

1. *Aplicar lo conceptos vistos en clase con cuestionarios y casos reales.*
2. *Identificar en el dataset los tipos de datos.*
3. *Realizar con un dataset la exploración, análisis y descripción de los datos.*

Materiales Necesarios:

1. Computador.
2. Internet.

Estructura del Laboratorio:

Laboratorio 1

1. Revisión de Conceptos – Cuestionario

- **Pregunta:** ¿Qué es el análisis de datos?
Respuesta: Es el proceso de exportar datos, analizarlos, limpiarlos y validarlos para tomar las mejores decisiones
- **Pregunta:** ¿Cómo beneficia el análisis de datos a la industria manufacturera?
Respuesta: Los beneficios son múltiples desde la predicción de cuando un equipo o maquina podría fallar, la optimización de los consumos de energía o del trabajo, cadena de suministros y mejorar la toma de decisiones.
- **Pregunta:** ¿Qué es un dataset?
Respuesta: Conjunto de datos organizados en una tabla
- **Pregunta:** ¿Qué implica la limpieza de datos?
Respuesta: identificar y corregir o eliminar errores, inconsistencias y datos irrelevantes en un conjunto de datos, con el fin de mejorar su calidad y confiabilidad para el análisis.
- **Pregunta:** ¿Cuál es la primera etapa del ciclo de vida de los datos?
Respuesta: creación y recopilación
- **Pregunta:** ¿Qué se hace en la etapa de procesamiento de datos?
Respuesta: Transformar los datos crudos en datos procesables y confiables para el análisis, eliminando duplicados, corrigiendo errores y asegurando la consistencia.

- **Pregunta:** ¿Qué tecnologías caracterizan la era del Big Data?

Respuesta: Hadoop y Spark, NoSQL Databases, Data Warehouses y Lakes, Machine Learning e IA, Cloud Computing, Data Visualization.

2. Escenarios reales aplicando el Análisis de Datos

1. **Escenario:** Pregunta: Una empresa de retail quiere entender por qué algunas tiendas tienen mejores ventas que otras. ¿Qué proceso podrían utilizar?

Respuesta: Recopilación de Datos, Análisis de Datos, Análisis de Competencia

2. **Escenario:** Pregunta: Un hospital necesita mejorar la precisión en los diagnósticos. ¿Cómo puede ayudar el análisis de datos?

Respuesta: Análisis de Datos, machine learning

3. **Escenario:** Pregunta: Una tienda en línea desea personalizar las recomendaciones de productos para sus clientes. ¿Qué concepto de análisis de datos utilizarían?

Respuesta: El análisis de datos basado en machine learning y algoritmos de recomendación, que analizan el historial de compras, preferencias y comportamientos de cada cliente para sugerir productos de interés.

4. **Escenario:** Pregunta: Un investigador está recopilando datos sobre la satisfacción del cliente. ¿Qué tipos de datos debería considerar?

Respuesta: **datos cualitativos** (opiniones y comentarios) y **datos cuantitativos** (calificaciones y métricas de satisfacción) para obtener una visión completa.

5. **Escenario:** Pregunta: Una encuesta tiene respuestas incompletas. ¿Qué paso en la limpieza de datos debería realizarse?

Respuesta: Se debe realizar la **imputación o eliminación de valores nulos** durante la limpieza de datos, dependiendo de la cantidad de respuestas incompletas y su relevancia en el análisis.

6. **Escenario:** Pregunta: Una empresa de energía renovable busca optimizar la producción de energía solar. ¿Qué tipo de análisis de datos deberían realizar?

Respuesta: Deberían realizar un **análisis predictivo de datos**, utilizando modelos de machine learning para anticipar patrones de producción y optimizar la eficiencia energética en base a condiciones climáticas y patrones de demanda.

7. **Escenario:** Pregunta: Un investigador está recopilando datos sobre el uso de una nueva aplicación móvil. ¿Qué tipos de datos debería considerar?

Respuesta: **datos de uso** (tiempo de uso, frecuencia, funcionalidades usadas) y **datos demográficos** de los usuarios (edad, ubicación)

8. **Escenario:** Pregunta: Un analista de datos descubre errores en un conjunto de datos de encuestas. ¿Qué paso en la limpieza de datos debería realizarse?

Respuesta: **corregir los errores mediante limpieza y validación de datos**, verificando la exactitud de los valores y eliminando duplicados o entradas anómalas.

9. Escenario: Pregunta: Una empresa tiene datos antiguos que ya no necesita. ¿Qué debería hacer con estos datos?

Respuesta: Proceso de eliminación de datos

3. Casos de la vida real en la industria aplicando el Análisis de Datos

El grupo de personas deberá realizar el análisis del caso, cuál sería la acción para seguir, qué resultado se tiene y qué reflexión queda.

Se debe demostrar, cómo el análisis de datos puede aplicarse de manera efectiva en diversas industrias para optimizar operaciones, mejorar la toma de decisiones y proporcionar un valor significativo tanto a las empresas como a los clientes.

Caso 1: Optimización de la cadena de suministro en Retail

Una cadena de supermercados internacional utilizó análisis de datos para optimizar su cadena de suministro. Recopilaron datos de ventas en tiempo real, niveles de inventario y patrones de demanda estacional para cada tienda.

- **Acción:** Esto permitió a la empresa identificar tendencias y ajustar los niveles de inventario en función de la demanda local y estacional, evitando el exceso o la falta de productos. Implementaron un modelo de predicción que anticipaba necesidades y adaptaba las órdenes de reposición para que el suministro fuera óptimo en cada región.
- **Resultado:** La empresa logró reducir el desperdicio de productos perecederos y los costos asociados a la gestión de inventario, al tiempo que mejoró la disponibilidad de productos en las tiendas. Los estantes estaban abastecidos de acuerdo a la demanda real, lo cual aumentó la satisfacción del cliente y disminuyó las pérdidas por productos no vendidos.
- **Reflexión:** El caso demuestra cómo el análisis de datos ayuda a la toma de decisiones estratégicas, al anticipar necesidades de abastecimiento y adaptar operaciones para responder mejor a la demanda del mercado.

Caso 2: Personalización de la experiencia del cliente en E-commerce

Una plataforma de comercio electrónico quería aumentar las ventas y la retención de clientes mediante la personalización de la experiencia de compra.

- **Acción:** usó análisis de datos y machine learning para segmentar clientes e implementar recomendaciones personalizadas basadas en el historial de compras y preferencias de los usuarios.
- **Resultado:** Aumento en las ventas por cliente y una mayor retención gracias a una experiencia de compra más atractiva y relevante.
- **Reflexión:** La personalización impulsa la lealtad y satisfacción del cliente, mostrando cómo el análisis de datos puede optimizar el marketing y el compromiso de los usuarios.

Laboratorio 2

1. Con el siguiente Dataset (Laboratorio_sesion 1_ datos_pandemia) realizar la identificación de tipos de datos, describirlos.

	A	B	C	D	E	F	G
	PacienteID	Edad	Genero	Ubicacion	FechaDiagnostico	EstadoSalud	Resultado
1	1	45	Masculino	CiudadA	2023-01-01 00:00:00	Grave	Fallecido
2	2	48	Femenino	CiudadB	2023-01-02 00:00:00	Leve	Fallecido
3	3	65	Masculino	CiudadA	2023-01-03 00:00:00	Leve	Recuperado
4	4	68	Masculino	CiudadA	2023-01-04 00:00:00	Moderado	Fallecido
5	5	68	Femenino	CiudadB	2023-01-05 00:00:00	Leve	Fallecido
6	6	10	Masculino	CiudadD	2023-01-06 00:00:00	Grave	Recuperado
7	7	84	Masculino	CiudadA	2023-01-07 00:00:00	Grave	Recuperado
8	8	22	Masculino	CiudadC	2023-01-08 00:00:00	Grave	Recuperado
9	9	37	Masculino	CiudadA	2023-01-09 00:00:00	Moderado	Recuperado
10	10	88	Masculino	CiudadA	2023-01-10 00:00:00	Leve	Fallecido
11	11	71	Femenino	CiudadB	2023-01-11 00:00:00	Leve	Fallecido
12	12	89	Femenino	CiudadB	2023-01-12 00:00:00	Leve	Recuperado
13	13	89	Femenino	CiudadB	2023-01-13 00:00:00	Leve	Fallecido
14	14	13	Femenino	CiudadB	2023-01-14 00:00:00	Leve	Fallecido
15	15	59	Femenino	CiudadA	2023-01-15 00:00:00	Leve	Fallecido
16	16	66	Femenino	CiudadC	2023-01-16 00:00:00	Leve	Fallecido
17	17	40	Femenino	CiudadD	2023-01-17 00:00:00	Moderado	Fallecido

R//

Paciente ID:

- Tipo de dato: **Cuantitativo**

Edad:

- Tipo de dato: **Cuantitativo**

Género:

- Tipo de dato: **Cualitativo**

Ubicación:

- Tipo de dato: **Cualitativo**

Fecha Diagnostico:

- Tipo de dato: **Cuantitativo**

Estado Salud:

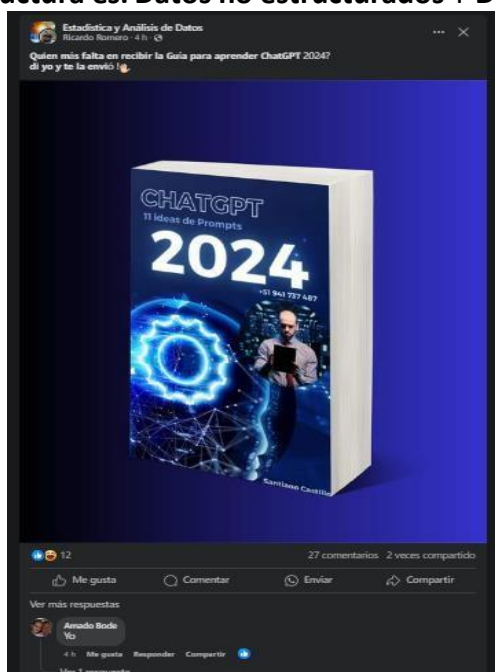
- Tipo de dato: **Cualitativo**

Resultado:

- Tipo de dato: **Cualitativo**

2. Defina la estructura de los datos y archivos: estructurado, no estructurados, semiestructurados.

La estructura es: Datos no estructurados Y Datos semiestructurados



- La estructura es: Datos no estructurados



Piedra, papel o tijeras | Corto Animado (Reflexiones)

YouTube | H De Hanniel | 404,8K visualizaciones | 10 de ene. de 2018

- La estructura es: datos semiestructurados

```

json
{
  "cliente": {
    "id": 12345,
    "nombre": "Juan Pérez",
    "correo_electronico": "juan.perez@example.com",
    "direccion": {
      "calle": "Calle Falsa 123",
      "ciudad": "Ciudad A",
      "codigo_postal": "12345"
    },
    "historial_compras": [
      {
        "id_compra": 67890,
        "fecha": "2023-08-14",
        "monto_total": 150.75,
        "articulos": [
          {
            "id_articulo": 111,
            "nombre_articulo": "Camiseta",
            "cantidad": 2,
            "precio_unitario": 25.00
          },
          {
            "id_articulo": 222,
            "nombre_articulo": "Pantalón",
            "cantidad": 1,
            "precio_unitario": 100.75
          }
        ]
      }
    ]
  }
}

```

- La estructura es: datos no estructurados

```
From: juan.perez@example.com
To: maria.gomez@example.com
Date: Mon, 14 Aug 2024 10:15:00 -0500
Subject: Reunión el viernes

Hola María,

Quisiera confirmar nuestra reunión para este viernes a las 10:00 AM.

Saludos,
Juan
```

ID_Pedido	ID_Cliente	Fecha_Pedido	Monto_Total
1001	1	2024-06-01	250.50
1002	2	2024-06-05	150.75
1003	1	2024-07-15	100.00

- La estructura es: Datos estructurados

1. **Eliminar Filas con Valores Inválidos o Incompletos**
2. **Corregir Errores Tipográficos y Inconsistencias**
3. **Formatear Fechas**
4. **Separar Información Incorrectamente Agrupada**
- 5.