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. 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.

Materiales Necesarios:

1. Computador.

2. Internet.

**Estructura del Laboratorio:**

**Revisión de Conceptos – Cuestionario**

**- Pregunta: ¿Qué es el análisis de datos?**

**Respuesta:** El análisis de datos es el proceso sistemático de inspeccionar, limpiar, transformar y modelar datos con el objetivo de descubrir información útil, sacar conclusiones y apoyar la toma de decisiones

- **Pregunta**: **¿Cómo beneficia el análisis de datos a la industria manufacturera?**

**Respuesta**: El análisis de datos beneficia a la industria manufacturera al transformar operaciones reactivas en procesos proactivos, optimizados y rentables.

- **Pregunta**: **¿Qué es un dataset?**

**Respuesta**: Un dataset es un conjunto estructurado de datos que generalmente se organiza en forma de tabla, donde cada fila representa una observación o registro (por ejemplo, un cliente, una venta, un producto), y cada columna representa una variable o característica (como edad, precio, cantidad, etc.).

- **Pregunta**: **¿Qué implica la limpieza de datos?**

**Respuesta**: La limpieza de datos es el proceso de identificar y corregir errores, inconsistencias o valores irrelevantes en un conjunto de datos para garantizar que sea preciso, completo y confiable antes de su análisis.

- **Pregunta**: **¿Cuál es la primera etapa del ciclo de vida de los datos?**

**Respuesta**: Recolección de datos

- **Pregunta**: **¿Qué se hace en la etapa de procesamiento de datos?**

**Respuesta**: El procesamiento de datos es el momento donde transformamos los datos crudos (a menudo desordenados o sin formato) en un formato útil y estructurado para análisis posteriores. Es la fase donde los datos cobran sentido y están listos para ser "leídos" por modelos, algoritmos o personas.

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

**Respuesta**: La era del Big Data se distingue por el uso de tecnologías avanzadas que permiten manejar, almacenar y procesar volúmenes masivos de datos a gran velocidad y con una variedad de formas. Estas tecnologías son clave para transformar datos no estructurados y estructurados en información útil para la toma de decisiones y la innovación.

**SECCIÓN 2.**

**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 relevantes (Ventas por tienda, Datos demográficos, Inventarios y stock, Factores externos), Limpieza y preparación de los datos, Análisis exploratorio de datos (EDA) [Visualización de patrones, Correlación, Segmentación de tiendas], Análisis de las métricas clave, KPIs de rendimiento, Factores de éxito, Modelado predictivo, Benchmarking y análisis competitivo, Toma de decisiones y acciones

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

**Respuesta**: Personalización de tratamientos y diagnóstico basado en datos

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

**Respuesta**: análisis predictivo y más específicamente, sistemas de recomendación.

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

**Respuesta**:

* Datos cualitativos {Comentarios y opiniones, Entrevistas y grupos focales}
* Datos cuantitativos {Encuestas de satisfacción}
* Datos ordinales {Escalas de satisfacción}
* Datos demográficos {Características del cliente}
* Datos de comportamiento {Interacciones previas con la marca, Historial de contacto con soporte}

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

**Respuesta**: Eliminar registros incompletos, Imputación de valores faltantes {Promedio/mediana/moda, Imputación basada en modelos}, Codificación de valores faltantes, Rellenar con valores predeterminados, Revisar el proceso de recolección de datos.

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

**Respuesta**: Análisis predictivo, Análisis de series temporales, Análisis de optimización

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

**Respuesta**:

Datos cuantitativos:

* Tiempo de uso: ¿Cuánto tiempo pasan los usuarios en la aplicación? Esto incluye el promedio de sesión y la frecuencia de uso.
* Interacciones por sesión: El número de acciones realizadas en la app, como clics, desplazamientos, o interacciones con diferentes características.
* Tasa de retención: Porcentaje de usuarios que regresan a la app después de un período determinado, lo que indica el interés y el valor percibido por los usuarios.
* Tasa de conversión: Cuántos usuarios completan una acción específica (como una compra, registro o descarga) dentro de la aplicación.
* Errores y fallos: Número de fallos o errores que los usuarios encuentran mientras usan la app, que podrían afectar la experiencia y la adopción.

Datos cualitativos:

* Comentarios y opiniones de los usuarios: Los usuarios pueden proporcionar retroalimentación sobre su experiencia a través de reseñas o encuestas dentro de la aplicación.
* Satisfacción del usuario: Encuestas o entrevistas que ayudan a comprender el nivel de satisfacción de los usuarios con la funcionalidad, diseño, y utilidad de la aplicación.
* Preferencias del usuario: Análisis de qué características de la app son más populares o útiles, basándose en comentarios o patrones de uso.

Datos demográficos:

* Edad, ubicación y género: Estos datos ayudan a comprender quién está utilizando la aplicación y permiten personalizar la experiencia o segmentar campañas de marketing.
* Tipo de dispositivo: Información sobre si los usuarios están utilizando dispositivos iOS o Android, y si la app tiene un rendimiento diferente en cada plataforma.

Datos de interacción y navegación:

* Flujo de navegación: ¿Qué caminos siguen los usuarios dentro de la app? Esto ayuda a identificar los puntos donde los usuarios abandonan la app o se encuentran con dificultades.
* Tasa de abandono: Porcentaje de usuarios que abandonan la app después de realizar ciertas acciones, lo que puede indicar posibles problemas en la interfaz de usuario o en la experiencia general.

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**: El analista de datos debería realizar una depuración y validación de los datos, lo cual incluye identificar y corregir valores atípicos, inconsistencias, datos faltantes o mal ingresados. Esto puede implicar pasos como:

* Eliminar duplicados que distorsionen los resultados.
* Corregir errores tipográficos o de formato (por ejemplo, "fememnino" en lugar de "femenino").
* Imputar valores faltantes de forma adecuada (media, mediana o técnicas más avanzadas).
* Estandarizar categorías para asegurar uniformidad.
* Verificar la integridad lógica entre campos (por ejemplo, edades que coincidan con la cohorte esperada).

Todo esto se realiza con el fin de asegurar que el análisis posterior se base en datos confiables, porque, como dice el dicho en nuestro mundo: "Garbage in, garbage out".

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

**Respuesta**: Archivado de datos siempre y cuando hay algún normativa legal que obligue a la empresa a mantenerlos se debe de tener guardado en un lugar aislado de los datos que están en productivo, Eliminarnos

**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:** La cadena de supermercados implementó un sistema de análisis de datos en tiempo real que recopiló información sobre las ventas, los niveles de inventario y los patrones de demanda estacional de cada tienda. Usaron algoritmos de análisis predictivo para prever las fluctuaciones de la demanda y ajustaron sus pedidos y reabastecimientos de inventario de manera más eficiente, minimizando las roturas de stock y optimizando los niveles de inventario. Además, integraron esta información con sus proveedores para mejorar la gestión de la cadena de suministro.

**Resultado:** El análisis de datos permitió a la cadena de supermercados lograr una mejora significativa en la eficiencia de la cadena de suministro. Las tiendas experimentaron menos quiebres de stock, se redujo el sobrestock de productos no vendidos y se optimizó el almacenamiento. Además, se redujo el costo de logística y se mejoró la satisfacción del cliente al asegurar que los productos más demandados estuvieran siempre disponibles en las estanterías. La capacidad para prever la demanda permitió una planificación más precisa y un reabastecimiento más efectivo, lo que redujo significativamente los costos operativos.

**Reflexión:** Este caso demuestra cómo el análisis de datos puede transformar una parte fundamental del negocio como la gestión de la cadena de suministro, proporcionando insights accionables que optimizan tanto los costos como el servicio al cliente. El uso de datos en tiempo real, combinado con análisis predictivos, permite a las empresas no solo reaccionar ante los problemas, sino anticiparse a ellos. En el caso de esta cadena de supermercados, la integración de tecnologías de análisis de datos resultó en ahorros significativos, mayor eficiencia operativa y un mejor servicio al cliente, algo que podría ser replicable para empresas de retail en diferentes sectores.

**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:** La plataforma de comercio electrónico implementó un sistema de recomendación personalizada utilizando análisis de datos para ofrecer productos adaptados a las preferencias y comportamientos de compra de cada cliente. Recopilaron datos sobre las acciones previas de los usuarios, como productos comprados, productos vistos, y tiempo de permanencia en distintas categorías. Usaron algoritmos de machine learning para segmentar a los clientes según sus intereses y comportamientos, y ofrecieron recomendaciones de productos en tiempo real, además de promociones y descuentos personalizados.

**Resultado:** Como resultado, la plataforma experimentó un aumento significativo en las ventas, ya que las recomendaciones más relevantes lograron captar la atención de los clientes y facilitar la compra impulsiva. La tasa de conversión aumentó al lograr que los productos recomendados estuvieran alineados con las preferencias del usuario, mejorando la experiencia de compra. Además, la retención de clientes mejoró, ya que los usuarios encontraron la plataforma más atractiva y fácil de usar. El sistema también permitió a la plataforma conocer mejor a sus clientes, lo que facilitó la creación de ofertas específicas que generaron lealtad a largo plazo.

**Reflexión:** Este caso destaca cómo la personalización basada en datos puede transformar la experiencia del cliente en el comercio electrónico, no solo aumentando las ventas inmediatas sino también construyendo relaciones a largo plazo. Los datos no solo permiten ofrecer productos relevantes, sino que también brindan una ventaja competitiva en un mercado saturado. Al centrarse en las necesidades y preferencias de los clientes, la empresa no solo incrementó sus ingresos, sino que también cultivó una experiencia de usuario más atractiva, lo que se traduce en una mayor lealtad y un engagement más profundo. Este tipo de personalización es ahora clave en la estrategia de cualquier empresa que quiera sobresalir en el mundo digital.

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

|  |  |
| --- | --- |
| La estructura es: |  |
|  | No estructurado |
|  | No estructurado |
|  | Semiestructurado tipo JSON |
|  | Semi estructurado |
|  | Estructurado, Tabla SQL |
|  | Estructurado CSV |

Con el siguiente Dataset realiza la exploración de los datos para la respectiva limpieza, y visualización, describir y analizar los datos.

descripción de los datos



Reporte de Limpieza de Datos

Dataset: Laboratorio \_sesion 1\_exploraciondatos.xlsx

Cantidad de registros analizados: 20

Pasos de limpieza ejecutados:

* Eliminación de filas vacías con valores na
* Se eliminaron las dos últimas filas del dataset, ya que presentaban valores faltantes (na) en todos sus campos.
* Se eliminaron las columnas Comentarios y Observaciones por no aportar valor analítico al objetivo del análisis.
* Se eliminaron registros duplicados basados en la columna Cédula, garantizando la unicidad de cada paciente/usuario.
* Corrección de nombres
* Se normalizaron los valores en la columna de nombres para corregir errores tipográficos, inconsistencias y garantizar uniformidad.
* Tratamiento de valores faltantes
* En las columnas Dirección, Celular y Correo Electrónico, los campos vacíos fueron estandarizados como null para un manejo adecuado en bases de datos o análisis posteriores.
* Corrección de valores categóricos
* En la columna País, se corrigieron nombres mal escritos o inconsistentes (ej.: "Colmbia" → "Colombia").
* En la columna Servicio, se unificaron los valores a solo dos categorías válidas: Comercio y Hogar. Los valores antiguos o mal digitados fueron estandarizados.
* Normalización de valores numéricos
* En la columna Valor Factura, las celdas vacías fueron imputadas con el valor 0, asumiendo ausencia de facturación.
* Limpieza y estandarización de fechas
* En la columna Fecha de Instalación, se corrigieron formatos incorrectos y se reemplazaron los valores faltantes por null.

|  |
| --- |
| **Cantidad de registros** |
| 14 |
| **Suma de valor factura** | **Promedio de valor factura** |
| $ 1.582.994,00 | $ 113.071,00 |
|  |  |

Cantidad de servicios instalado por país

|  |  |  |  |
| --- | --- | --- | --- |
| **Paises** | **Comercio** | **Hogar** | **Total general** |
| Peru | 1 | 1 | 2 |
| Japon | 1 | 1 | 2 |
| Colombia |  | 2 | 2 |
| Chile | 2 |  | 2 |
| Nicaragua | 1 |  | 1 |
| Venezuela | 1 |  | 1 |
| España |  | 1 | 1 |
| italia |  | 1 | 1 |
| Ecuador | 1 |  | 1 |
| Honduras |  | 1 | 1 |
| **Total general** | **7** | **7** | **14** |

Países que más facturan

|  |  |  |  |
| --- | --- | --- | --- |
| **Paises** | **Comercio** | **Hogar** | **Total general** |
| Venezuela | $ 680.000,00 |  | $ 680.000,00 |
| Japon | $ 450.000,00 | $ - | $ 450.000,00 |
| Honduras |  | $ 89.000,00 | $ 89.000,00 |
| Colombia |  | $ 79.876,00 | $ 79.876,00 |
| España |  | $ 69.702,00 | $ 69.702,00 |
| Chile | $ 68.456,00 |  | $ 68.456,00 |
| Ecuador | $ 54.000,00 |  | $ 54.000,00 |
| Peru | $ 23.000,00 | $ 23.000,00 | $ 46.000,00 |
| Nicaragua | $ 45.960,00 |  | $ 45.960,00 |
| italia |  | $ - | $ - |
| **Total general** | **$ 1.321.416,00** | **$ 261.578,00** | **$ 1.582.994,00** |