

## Laboratorio 26

### Sesión #26 Carga y Exploración de Datos

**Título del Laboratorio:** Carga y Exploración de Datos de las encuesta de Empresas en Bogotá

**Duración:** 2 horas

**Objetivos del Laboratorio:**

1. *Análisis es determinar cómo diversos factores (tipo de empresa, respuestas a preguntas clave y factores ponderados) influyen en el desempeño y las decisiones estratégicas de estas empresas.*
2. *El análisis busca proporcionar insights que revelen oportunidades de mejora para las organizaciones.*

**Materiales Necesarios:**

1. Computador con conexión a internet.
2. Dataset confiable de <https://www.datos.gov.co/>

**Estructura del Laboratorio:**

Cargar el archivo "Mediana\_Empresa.csv" en la herramienta de tu elección (Excel, Python, Power BI) y realizar una exploración inicial.

1. Realizar el cargue del Dataset y explorar la información.
2. Formulación de las preguntas iniciales para guiar el análisis, responderlas.
  - a) ¿Cuál es el objetivo del análisis de datos?
    - ❖ El objetivo del análisis de datos es transformar grandes volúmenes de datos en información comprensible y valiosa que apoye la toma de decisiones. Permite identificar patrones, tendencias y relaciones dentro de los datos para predecir comportamientos futuros. También facilita la mejora de procesos operativos y la optimización de recursos, maximizando la eficiencia en diversas áreas de la organización. A través de la inteligencia de datos, las empresas pueden responder de manera más rápida y precisa a los cambios del mercado. Además, el análisis de datos ayuda a obtener una visión más clara de las necesidades y preferencias de los clientes. En resumen, su objetivo es proporcionar insights que ayuden a tomar decisiones más informadas y estratégicas.

b) ¿Qué problemas específicos se quieren resolver?

- ❖ El análisis de datos busca resolver problemas como la incertidumbre en la toma de decisiones, proporcionando hechos y cifras claras. También ayuda a identificar ineficiencias en procesos operativos, como el uso subóptimo de recursos. En el área comercial, facilita la segmentación de clientes, ayudando a personalizar productos y servicios según sus necesidades. También puede prevenir o detectar fraudes, analizando patrones anómalos en datos financieros o transacciones. El análisis de datos mejora la previsión de demanda y la gestión de inventarios, evitando tanto el exceso como la escasez de productos. Finalmente, permite tomar decisiones informadas sobre la calidad y el rendimiento de productos o servicios.

c) ¿Qué decisiones estratégicas se derivarán del análisis?

- ❖ El análisis de datos permite tomar decisiones estratégicas como el lanzamiento de nuevos productos o servicios, basados en las preferencias y comportamientos de los clientes. Ayuda a ajustar las estrategias de marketing, enfocándose en los segmentos más rentables o personalizados. También optimiza la fijación de precios, adaptando ofertas según la demanda y la elasticidad de los precios. Facilita la expansión a nuevos mercados, identificando áreas de crecimiento y posibles oportunidades. Además, el análisis de datos permite mejorar la eficiencia operativa, reduciendo costos y mejorando procesos. Finalmente, ayuda a gestionar riesgos y a asegurar el cumplimiento de normativas, especialmente en sectores regulados.

```
#Mostrar ejemplos nulos
print("Ejemplos de valores nulos: ")
print(df[df.isnull().any(axis=1)].head())
```

Ejemplos de valores nulos:

	ORIGEN	PRODUCTO	GRADO DE ALCOHOL	\
105	I	Tequila Oro (Joven) Mixto Jalieza	35.32	
159	I	V. Ros. Ig Lisboa Mil Caminho	13.13	
176	I	V. Verde Branco Casal Garcia	9.40	
177	I	V. Verde Rosado Casal Garcia	9.40	
227	I	W. Balthus 12 Años	40.00	

  

	REGISTRO SANITARIO	VIGENCIA DE REGISTRO	SANITARIO	\
105	INVIMA 2018L-0009429		24/07/2028	
159	INVIMA 2015L-0007963		21/12/2025	
176	INVIMA 2017L-0008557		17/02/2027	
177	INVIMA 2017L-0008557		17/02/2027	
227	INVIMA 2016L-0008371		3/10/2026	

  

	RESOLUCIÓN DE REGISTRO	SANITARIO	PRODUCTOR	\
105		18026542.0	NaN	
159		15047858.0	NaN	
176		17000164.0	NaN	
177		17000164.0	NaN	
227		19020764.0	NaN	

  

	NOMBRE EMPRESA	DISTRIBUIDORA	NIT
105	RODRIGUEZ Q. ALEXANDER/3&M	DIST BRISSART	1.002786e+08
159	RODRIGUEZ Q. ALEXANDER/3&M	DIST BRISSART	1.002786e+08
176	RODRIGUEZ Q. ALEXANDER/3&M	DIST BRISSART	1.002786e+08
177	RODRIGUEZ Q. ALEXANDER/3&M	DIST BRISSART	1.002786e+08
227	DELGADO MARIN LUISA	FERNANDA	1.088255e+10

▶ #Describir columnas de texto o cadenas de caracteres

```
print("Descripcion de columnas de texto: ")
for column in df.select_dtypes(include=['object']).columns:
    print(f"{column}: {df[column].str.len().describe()}")
```

↗ Descripcion de columnas de texto:

ORIGEN:	count	15886.0
mean		1.0
std		0.0
min		1.0
25%		1.0
50%		1.0
75%		1.0
max		1.0

Name: ORIGEN, dtype: float64

PRODUCTO:	count	15886.00000
mean		29.47759
std		6.62087
min		6.00000
25%		25.00000
50%		30.00000
75%		35.00000
max		50.00000

Name: PRODUCTO, dtype: float64

REGISTRO SANITARIO:	count	15797.00000
mean		17.778312
std		3.764612
min		4.00000
25%		15.00000
50%		20.00000
75%		20.00000

▶ # Valores unicos que podrian ser inconsistencias

```
print("Inconsistencias potenciales: ")
for column in df.select_dtypes(include=['object']).columns:
    print(f"{column}: {df[column].unique()[:10]}")
```

↗ Inconsistencias potenciales:

ORIGEN: ['N' 'I']

PRODUCTO: ['Ap. Brissart Sabor A Cafe' 'Ap. Crema Con Sabor A Chocolate M Harv'  
'Ap. Crema Sab Whisky Harvey Mackys 14'  
'Ap. Crema Sab. A Ron Harvey Mackys' 'Ap. Crema Sabor A Whisky Brissart'  
'Ap. No Vin. Anisado Eduardo Iii' 'Ap. No Vin. Sabor Whisky Wardos'  
'Ap. Sab. Ron Casa La Viña (2.5)' 'Ap. Sabor Cereza Jumbo'  
'Beb Emb Sab A Fresa Casa La Viña']

REGISTRO SANITARIO: ['INVIMA 2019L-0010049' 'INVIMA 2016L-0008145' 'INVIMA 2017L-0008812'  
'INVIMA 2014L-0007076' 'INVIMA 2019L-0009865' 'INVIMA L-002435 R1'  
'INVIMA 2017L-0009117' 'INVIMA 2011L-0005528' 'INVIMA 2017L-0008925'  
'RSA-003197-2017']

VIGENCIA DE REGISTRO SANITARIO: ['19/06/2029' '14/04/2026' '17/07/2027' '12/05/2024' '8/03/2029'  
'27/10/2019' '7/12/2027' '7/10/2021' '18/09/2027' '10/05/2022']

PRODUCTOR: ['LICORES BRISSART S.A.S' 'CANDIOTA DE VINOS Y LICORES S.A.'  
'CASA LA VIÑA LTDA' 'CASA LA VIÑA LTDA.' 'PITASH S.A.S.'  
'IMPORTACIONES GRUCAL SAS' 'MONTAÑA CORREA NICOLAS FERNANDO' 'JONRON SAS'  
'INVERSIONES LA CARETA SAS' 'IMPORTACIONES GRUCAL S.A.S.']

NOMBRE EMPRESA DISTRIBUIDORA: ['RODRIGUEZ Q. ALEXANDER/J&M DIST BRISSART'  
'VELASQUEZ ARANGO JAIME A/ESTANQ. POPULAR' 'GOMEZ CARDONA DIANA CAROLINA'  
'DELGADO MARIN LUISA FERNANDA' 'HENAO CEBALLOS LAURA CAROLINA'  
'INCORPORADO - LIASCOS DIAZ LORENA ANDRE' 'BETANCUR VICTOR MANUEL'  
'VILLEGAS PEÑA PEDRO LUIS-BOD. DEL NORTE'  
'GARZON VALENCIA ANDRES(RANCHO Y LICORES)' 'OÑATE CARVAJAL JAIRO RAFAEL']

### 3. Conclusiones (10 minutos):

## Resultados del análisis del dataset de licores

### 1. Información General:

- El dataset contiene 15,977 filas y 9 columnas, con datos sobre el origen, producto, grado de alcohol, registro sanitario, productor, y distribuidor de diversos licores.

### 2. Valores Nulos:

- Se encontraron 134 valores nulos en "REGISTRO SANITARIO", 6,306 en "RESOLUCIÓN DE REGISTRO SANITARIO", y 168 en "PRODUCTOR". Estos valores faltantes podrían afectar la precisión del análisis si no se manejan adecuadamente

### 3. Filas Duplicadas:

- Se detectaron 2,574 filas duplicadas que pueden distorsionar los resultados y deben eliminarse

### 4. Resumen de Columnas Categóricas:

- **ORIGEN:** 10,417 productos son de origen importado (I), mientras que 5,560 son de origen nacional (N).
- **PRODUCTO:** Existen 7,940 productos únicos, con algunos de los más frecuentes siendo "W. Black & White 40°" y "Gin. Gordons Special Dry".
- **REGISTRO SANITARIO:** Hay 4,148 registros sanitarios distintos.
- **VIGENCIA DE REGISTRO SANITARIO:** Existen 1,945 fechas diferentes de vigencia, siendo las más comunes las de 2024 y 2025.
- **PRODUCTOR:** Los principales productores son "GLOBAL WINE & SPIRITIS LTDA" y "DISTRIBUIDORA DE VINOS Y LICORES S.A.S."
- **NOMBRE EMPRESA DISTRIBUIDORA:** Hay 219 distribuidoras distintas, donde destacan "DIST. DE VINOS Y LICORES S.A." y "COMERCIAL DE LICORES S.A.S."