

Laboratorio 27

Sesión #27 Limpieza, estadística descriptiva

Título del Laboratorio: Limpieza y estadística descriptiva de los datos en los productos de licor registrados en Risaralda para la distribución.

Duración: 2 horas

Objetivos del Laboratorio:

1. *Llevar a cabo un proceso completo de limpieza de datos y aplicar técnicas de estadística descriptiva al dataset de productos de licor, con el fin de mejorar la calidad de los datos y proporcionar insights significativos que puedan ser utilizados para la toma de decisiones estratégicas.*
2. *Esto incluye identificar y resolver problemas de datos como valores nulos, duplicados e inconsistencias, y resumir las principales características del dataset mediante medidas estadísticas que faciliten la comprensión y análisis de la información*

Materiales Necesarios:

1. *Computador con conexión a internet.*
2. *Herramientas de limpieza de los datos y aplicación de estadística descriptiva.*

Estructura del Laboratorio:

1. Entregar un informe escrito que detalle el paso a paso, debe incluir las primeras observaciones sobre la estructura de los datos, los hallazgos, conclusiones y recomendaciones, se debe entregar en PDF o Word.
2. Responder las preguntas planteadas.
 - a) **¿Qué información se debe analizar en el Dataset para garantizar el cumplimiento de las normativas sanitarias y optimizar la distribución de los productos?** Se debe analizar información relacionada con los estándares de calidad y las regulaciones sanitarias, como fechas de vencimiento, condiciones de almacenamiento, registros de inspección y datos de trazabilidad. Además, para optimizar la distribución de los productos, es importante analizar datos sobre demanda, tiempos de entrega, capacidad de almacenamiento y rutas de distribución.
 - b) **¿Cómo pueden los valores nulos y las inconsistencias en el texto afectar la precisión y calidad del análisis de los datos?** Los valores nulos y las inconsistencias en el texto pueden afectar significativamente la precisión y calidad del análisis de datos, ya que pueden introducir sesgos, errores y omisiones en los resultados. Esto puede llevar a conclusiones incorrectas y decisiones basadas en datos incompletos o imprecisos.

- c) **¿Qué estrategias se pueden implementar para asegurar que el análisis de datos sea preciso y útil para la toma de decisiones estratégicas?** Para asegurar un análisis de datos preciso y útil, se pueden implementar estrategias como la limpieza y preparación exhaustiva de los datos, el uso de técnicas avanzadas de modelado y validación, y la implementación de controles de calidad y auditorías regulares. Además, es importante contar con herramientas y tecnologías adecuadas, así como capacitar al personal en las mejores prácticas de análisis de datos.

```
df['NOMBRE EMPRESA DISTRIBUIDORA'].fillna('Desconocido', inplace=True)
```

```
[ ] df = df.dropna()

nulos_despues = df.isnull().sum()

print(nulos_despues)
```

```
ORIGEN      0
PRODUCTO    0
GRADO DE ALCOHOL  0
REGISTRO SANITARIO  0
VIGENCIA DE REGISTRO SANITARIO  0
RESOLUCIÓN DE REGISTRO SANITARIO  0
PRODUCTOR    0
NOMBRE EMPRESA DISTRIBUIDORA  0
NIT          0
dtype: int64
```

```
print("Datos despues de nuestra limpieza: ")

print(df.head())
```


```
Datos despues de nuestra limpieza:
ORIGEN      PRODUCTO  GRADO DE ALCOHOL  \
0      N      Ap. Brissart Sabor A Cafe      10.0
1      N  Ap. Crema Con Sabor A Chocolate M Harv      14.0
2      N  Ap. Crema Sab Whisky Harvey Mackys      14.0
4      N      Ap. Crema Sab. A Ron Harvey Mackys      14.0
6      N      Ap. Crema Sabor A Whisky Brissart      14.0

REGISTRO SANITARIO  VIGENCIA DE REGISTRO SANITARIO  \
0  INVIMA 2019L-0010049      19/06/2029
1  INVIMA 2016L-0008145      14/04/2026
2  INVIMA 2017L-0008812      17/07/2027
4  INVIMA 2014L-0007076      12/05/2024
6  INVIMA 2019L-0009865      8/03/2029

RESOLUCIÓN DE REGISTRO SANITARIO      PRODUCTOR  \
0      19022399.0      LICORES BRISSART S.A.S
1      16009721.0  CANDIOTA DE VINOS Y LICORES S.A.
2      17025467.0  CANDIOTA DE VINOS Y LICORES S.A.
4      14009708.0  CANDIOTA DE VINOS Y LICORES S.A.
6      19005235.0      LICORES BRISSART S.A.S


NOMBRE EMPRESA DISTRIBUIDORA      NIT
0  RODRIGUEZ Q. ALEXANDER/J&M DIST BRISSART  100278640
1  RODRIGUEZ Q. ALEXANDER/J&M DIST BRISSART  100278640
2  RODRIGUEZ Q. ALEXANDER/J&M DIST BRISSART  100278640
4  RODRIGUEZ Q. ALEXANDER/J&M DIST BRISSART  100278640
6  RODRIGUEZ Q. ALEXANDER/J&M DIST BRISSART  100278640
```


Conclusiones: Se debe analizar información relacionada con estándares de calidad y regulaciones sanitarias, como fechas de vencimiento y condiciones de almacenamiento, además de datos sobre demanda y tiempos de entrega para optimizar la distribución. Los valores nulos y las inconsistencias en el texto pueden introducir sesgos y errores en el análisis de datos, afectando su precisión y calidad. Para un análisis de datos preciso y útil, se deben implementar estrategias como la limpieza exhaustiva de datos, técnicas avanzadas de modelado y controles de calidad regulares.

 laboratorio 27

Crear laboratorio 27

Hace 1 minuto

 README



Añadir un README

Ayude a las personas interesadas en este repositorio a comprender su proyecto agregando un README.

Añadir un README