

UNIT 8: DATA ANALYTICS

Óscar Durán Sánchez

Fecha de entrega: 06/05/2024
CampusFP

Índice

1) Investiga y define brevemente los siguientes conceptos:	4
1. Estadística:	4
2. Población:	4
3. Muestra:	4
4. Dato (cuantitativo y cualitativo):	4
5. Media, mediana y moda:	4
2) Ejecuta dos sentencias SELECT distintas para obtener la media aritmética del número de piezas de fruta diaria que comen los hombres y que comen las mujeres.	5
3) Investiga y define brevemente las tres técnicas de análisis de datos (estadística descriptiva, estadística predictiva, estadística prescriptiva)	6
1. Estadística Descriptiva:	6
2. Estadística Predictiva:	6
3. Estadística Prescriptiva:	6
4) Realiza un pequeño trabajo de investigación de mínimo una página sobre las técnicas de análisis descriptivo	7
Introducción	7
Medidas de Tendencia Central	7
Medidas de Dispersión	7
Gráficos y Visualización de Datos	7
Conclusiones	8
5) Ejecuta las sentencias SQL necesarias para calcular la media, mediana y moda de cada una de las preguntas numéricas de la encuesta.	8
6) Realiza una serie de gráficos para representar de manera más gráfica la información obtenida en el ejercicio 5.	10
7) Evalúa todos los resultados obtenidos en los gráficos anteriores y saca conclusiones.	11
Explica la importancia del análisis de datos en la toma de decisiones.	11
8) Realiza un pequeño trabajo de investigación de mínimo una página sobre las técnicas de análisis predictivo.	12
Técnicas de Análisis Predictivo: Prediciendo el Futuro a partir de los Datos	12
Introducción	12
Modelos de Regresión	12
Modelos de Clasificación	12
Modelos de Series Temporales	12
Validación y Evaluación de Modelos	13

Conclusiones	13
9) Realiza una distribución de frecuencias que permita averiguar cuántas personas de cada rango de edad se han interesado en rellenar la encuesta. Puedes utilizar excel para realizar la distribución de frecuencias. Detalla cómo lo has hecho y adjunta el resultado a tu informe. 13	
Proceso:.....	13
Resultado:.....	14
10) Investiga para exponer al menos 3 técnicas analíticas predictivas	15
1. Regresión Lineal:	15
2. Árboles de Decisión:.....	15
3. Redes Neuronales Artificiales:.....	15
11) Evalúa cómo se pueden utilizar las técnicas analíticas predictivas con fines de pronóstico. Realiza un análisis a partir de la distribución de frecuencias anterior.	16
12) Realiza un pequeño trabajo de investigación de mínimo una página sobre las técnicas de análisis prescriptivo.	17
Técnicas de Análisis Prescriptivo: Optimizando Decisiones Estratégicas	17
Introducción	17
Programación Lineal y Programación Entera	17
Optimización de Redes	17
Algoritmos Genéticos y Metaheurísticas.....	17
Conclusiones	18
13) Calcula la tendencia por rango de edad sobre el consumo de verduras usando la técnica de las dos medias.	18
14) Lee el siguiente post: https://acortar.link/oEsOVR. Realiza una pequeña síntesis y saca conclusiones.....	19
15) Desarrolla una función o procedimiento que implemente el cálculo desarrollado en el ejercicio 13. Puedes utilizar un procedimiento o función en el entorno de la base de datos o utilizar cualquier otro lenguaje de programación de tu elección	20
1. Importación de bibliotecas:.....	20
2. Definición de la función calcular_tendencia_consumo_verduras	20
3. Ejemplo de uso	20
Elección de Python:	20
Bibliografía:	21

Actividad de asignación y orientación

1) Investiga y define brevemente los siguientes conceptos:

1. Estadística:

- La estadística es una rama de las matemáticas que se ocupa de recolectar, organizar, analizar e interpretar datos para tomar decisiones basadas en la información proporcionada por ellos. Ayuda a comprender y describir fenómenos en términos numéricos y a inferir conclusiones sobre una población a partir de una muestra de datos.

2. Población:

- En estadística, la población se refiere al conjunto completo de elementos o individuos que comparten una característica común y que son objeto de estudio. Por ejemplo, si estamos investigando los hábitos alimenticios de los habitantes de la Comunidad de Madrid, la población sería el conjunto total de todas las personas que viven en esa área geográfica.

3. Muestra:

- Una muestra es un subconjunto representativo de la población que se selecciona para realizar un estudio estadístico. La muestra se elige de manera que refleje las características de la población en general. En el escenario que has planteado, la muestra consistiría en un grupo de personas de la Comunidad de Madrid que han respondido a la encuesta sobre hábitos alimenticios.

4. Dato (cuantitativo y cualitativo):

- Los datos cuantitativos son aquellos que se expresan en términos numéricos y se pueden medir o contar. Por ejemplo, la edad de una persona, el número de comidas que realiza al día o la cantidad de frutas que consume diariamente.
- Los datos cualitativos son descriptivos y no numéricos. Se refieren a características o cualidades que no se pueden cuantificar fácilmente. Por ejemplo, el sexo de una persona, su preferencia de desayuno o su nivel de satisfacción con un producto.

5. Media, mediana y moda:

- La media es el promedio aritmético de un conjunto de datos numéricos. Se calcula sumando todos los valores y dividiéndolos por el número total de valores. Es sensible a valores extremos.
- La mediana es el valor que ocupa la posición central en un conjunto de datos ordenados de menor a mayor (o viceversa). Si el número de datos es impar, la mediana es el valor que se

encuentra en el centro. Si es par, la mediana es el promedio de los dos valores centrales.

- La moda es el valor que aparece con mayor frecuencia en un conjunto de datos. Puede haber una moda (datos unimodales), más de una moda (datos bimodales) o ninguna moda (datos amodales).

2) Ejecuta dos sentencias SELECT distintas para obtener la media aritmética del número de piezas de fruta diaria que comen los hombres y que comen las mujeres.

```
34 -- Para calcular la media de piezas de fruta diarias consumidas por hombres:
35 • SELECT AVG(frutas_dia) AS media_frutas_hombres
36 FROM encuesta
37 WHERE sexo = 'Masculino';
38
```

Result Grid

media_frutas_hombres
2.1667

Result 2 x

Output

Action Output

#	Time	Action	Message
✓ 1	20:30:49	SELECT AVG(frutas_dia) AS media_frutas_hombres FROM encuesta WHERE sexo = 'Masculino' LIMIT 0, 1000	1 row(s) returned

```
39 -- Para calcular la media de piezas de fruta diarias consumidas por mujeres:
40 • SELECT AVG(frutas_dia) AS media_frutas_mujeres
41 FROM encuesta
42 WHERE sexo = 'Femenino';
43
```

Result Grid

media_frutas_mujeres
2.2727

Result 3 x

Output

Action Output

#	Time	Action	Message
✓ 1	20:31:30	SELECT AVG(frutas_dia) AS media_frutas_mujeres FROM encuesta WHERE sexo = 'Femenino' LIMIT 0, 1000	1 row(s) returned

3) Investiga y define brevemente las tres técnicas de análisis de datos (estadística descriptiva, estadística predictiva, estadística prescriptiva)

1. Estadística Descriptiva:

- La estadística descriptiva se centra en la descripción y resumen de características importantes de un conjunto de datos. Esto incluye calcular medidas de tendencia central (como la media, la mediana y la moda) y medidas de dispersión (como la desviación estándar y el rango), así como la creación de gráficos y tablas para visualizar la distribución de los datos. Su objetivo principal es comprender y comunicar la información presente en los datos, sin realizar inferencias o predicciones sobre la población.

2. Estadística Predictiva:

- La estadística predictiva se utiliza para predecir o estimar valores futuros o desconocidos basándose en datos históricos y patrones identificados en ellos. Utiliza técnicas de modelado estadístico y aprendizaje automático para construir modelos predictivos, que luego se utilizan para hacer predicciones sobre nuevas observaciones. Estos modelos pueden predecir resultados futuros o desconocidos, como el comportamiento del mercado, la demanda de productos, o el riesgo de eventos futuros.

3. Estadística Prescriptiva:

- La estadística prescriptiva va más allá de la predicción y busca recomendar acciones específicas para optimizar los resultados deseados. Utiliza modelos analíticos avanzados y técnicas de optimización para encontrar la mejor solución posible ante un conjunto de restricciones y objetivos. Por ejemplo, en el contexto empresarial, la estadística prescriptiva puede ayudar a optimizar la asignación de recursos, la programación de la producción o la toma de decisiones estratégicas. Su objetivo es ofrecer recomendaciones accionables para mejorar el rendimiento y la eficiencia en diversas áreas de aplicación.

4) Realiza un pequeño trabajo de investigación de mínimo una página sobre las técnicas de análisis descriptivo

Introducción

En el campo de la estadística y el análisis de datos, el análisis descriptivo desempeña un papel fundamental al permitirnos comprender y resumir las características clave de un conjunto de datos. A través de diversas técnicas, el análisis descriptivo nos ayuda a obtener información valiosa sobre la distribución, tendencias y patrones presentes en los datos, lo que a su vez nos ayuda a tomar decisiones informadas y a comunicar eficazmente los hallazgos. En este trabajo de investigación, exploraremos algunas de las técnicas más comunes utilizadas en el análisis descriptivo y su importancia en la comprensión de los datos.

Medidas de Tendencia Central

Una de las categorías principales de técnicas de análisis descriptivo son las medidas de tendencia central, que proporcionan información sobre el valor típico o central en un conjunto de datos. Las medidas más comunes incluyen la media, la mediana y la moda. La media aritmética es simplemente el promedio de todos los valores en el conjunto de datos y es sensible a los valores extremos. Por otro lado, la mediana es el valor que ocupa la posición central cuando los datos están ordenados, y es menos sensible a los valores atípicos. La moda es el valor que aparece con mayor frecuencia en el conjunto de datos y puede ser útil para identificar patrones de comportamiento o preferencias.

Medidas de Dispersión

Además de las medidas de tendencia central, el análisis descriptivo también incluye medidas de dispersión, que nos ayudan a comprender la variabilidad o dispersión de los datos. La desviación estándar y el rango son dos medidas comunes de dispersión. La desviación estándar mide la dispersión de los datos alrededor de la media, mientras que el rango es simplemente la diferencia entre el valor máximo y el valor mínimo en el conjunto de datos. Estas medidas son importantes para evaluar la consistencia y la estabilidad de los datos, así como para identificar posibles valores atípicos que puedan afectar la interpretación de los resultados.

Gráficos y Visualización de Datos

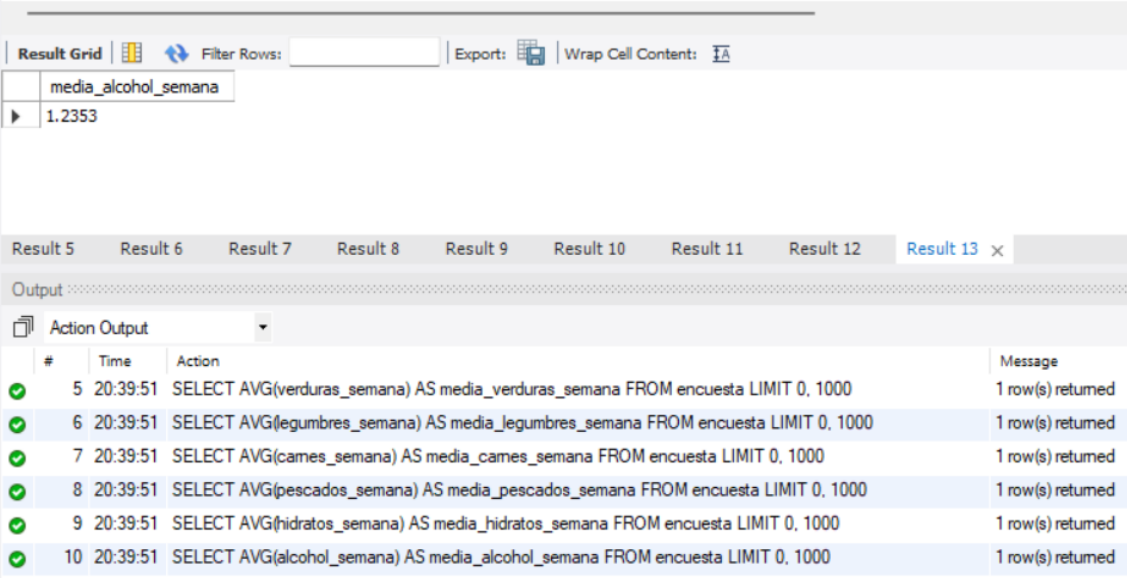
Además de las medidas numéricas, la visualización de datos desempeña un papel crucial en el análisis descriptivo. Los gráficos, como histogramas, gráficos de barras, gráficos de dispersión y diagramas de caja, nos permiten visualizar la distribución y la estructura de los datos de manera intuitiva. Estas representaciones gráficas pueden revelar patrones, tendencias y relaciones que pueden no ser evidentes al examinar solo las cifras. Por ejemplo, un histograma puede mostrar la distribución de edades en una muestra de población, mientras que un diagrama de caja puede resaltar diferencias en la dispersión entre grupos de datos.

Conclusiones

En resumen, el análisis descriptivo es una parte esencial del proceso de análisis de datos, ya que nos proporciona una comprensión inicial y completa de las características fundamentales de un conjunto de datos. A través de medidas de tendencia central, medidas de dispersión y visualización de datos, podemos identificar patrones, tendencias y posibles problemas en los datos que pueden influir en nuestras decisiones y conclusiones. Al aplicar estas técnicas de manera efectiva, podemos obtener información valiosa que nos ayuda a tomar decisiones informadas y a comunicar nuestros hallazgos de manera clara y concisa.

5) Ejecuta las sentencias SQL necesarias para calcular la media, mediana y moda de cada una de las preguntas numéricas de la encuesta.

```
45  -- Media (promedio):
46  • SELECT AVG(edad) AS media_edad FROM encuesta;
47  • SELECT AVG(comidas_dia) AS media_comidas_dia FROM encuesta;
48  • SELECT AVG(frutas_dia) AS media_frutas_dia FROM encuesta;
49  • SELECT AVG(verduras_semana) AS media_verduras_semana FROM encuesta;
50  • SELECT AVG(legumbres_semana) AS media_legumbres_semana FROM encuesta;
51  • SELECT AVG(carnes_semana) AS media_carnes_semana FROM encuesta;
52  • SELECT AVG(pescados_semana) AS media_pescados_semana FROM encuesta;
53  • SELECT AVG(hidratos_semana) AS media_hidratos_semana FROM encuesta;
54  • SELECT AVG(alcohol_semana) AS media_alcohol_semana FROM encuesta;
55
```



#	Time	Action	Message
✓ 5	20:39:51	SELECT AVG(verduras_semana) AS media_verduras_semana FROM encuesta LIMIT 0, 1000	1 row(s) returned
✓ 6	20:39:51	SELECT AVG(legumbres_semana) AS media_legumbres_semana FROM encuesta LIMIT 0, 1000	1 row(s) returned
✓ 7	20:39:51	SELECT AVG(carnes_semana) AS media_carnes_semana FROM encuesta LIMIT 0, 1000	1 row(s) returned
✓ 8	20:39:51	SELECT AVG(pescados_semana) AS media_pescados_semana FROM encuesta LIMIT 0, 1000	1 row(s) returned
✓ 9	20:39:51	SELECT AVG(hidratos_semana) AS media_hidratos_semana FROM encuesta LIMIT 0, 1000	1 row(s) returned
✓ 10	20:39:51	SELECT AVG(alcohol_semana) AS media_alcohol_semana FROM encuesta LIMIT 0, 1000	1 row(s) returned


```
-- Mediana (valor medio):
```

```
SELECT
```

```
PERCENTILE_CONT(0.5) WITHIN GROUP (ORDER BY edad) AS mediana_edad,
PERCENTILE_CONT(0.5) WITHIN GROUP (ORDER BY comidas_dia) AS mediana_comidas_dia,
PERCENTILE_CONT(0.5) WITHIN GROUP (ORDER BY frutas_dia) AS mediana_frutas_dia,
PERCENTILE_CONT(0.5) WITHIN GROUP (ORDER BY verduras_semana) AS mediana_verduras_semana,
PERCENTILE_CONT(0.5) WITHIN GROUP (ORDER BY legumbres_semana) AS mediana_legumbres_semana,
PERCENTILE_CONT(0.5) WITHIN GROUP (ORDER BY carnes_semana) AS mediana_carnes_semana,
PERCENTILE_CONT(0.5) WITHIN GROUP (ORDER BY pescados_semana) AS mediana_pescados_semana,
PERCENTILE_CONT(0.5) WITHIN GROUP (ORDER BY hidratos_semana) AS mediana_hidratos_semana,
PERCENTILE_CONT(0.5) WITHIN GROUP (ORDER BY alcohol_semana) AS mediana_alcohol_semana
```

```
FROM encuesta;
```

Output

Time	Action	Message
20:44:42	SELECT AVG(edad) AS media_edad FROM encuesta LIMIT 0, 1000	1 row(s) returned
20:44:42	SELECT AVG(comidas_dia) AS media_comidas_dia FROM encuesta LIMIT 0, 1000	1 row(s) returned
20:44:42	SELECT AVG(frutas_dia) AS media_frutas_dia FROM encuesta LIMIT 0, 1000	1 row(s) returned
20:44:42	SELECT AVG(verduras_semana) AS media_verduras_semana FROM encuesta LIMIT 0, 1000	1 row(s) returned
20:44:42	SELECT AVG(legumbres_semana) AS media_legumbres_semana FROM encuesta LIMIT 0, 1000	1 row(s) returned
20:44:42	SELECT AVG(carnes_semana) AS media_carnes_semana FROM encuesta LIMIT 0, 1000	1 row(s) returned
20:44:42	SELECT AVG(pescados_semana) AS media_pescados_semana FROM encuesta LIMIT 0, 1000	1 row(s) returned
20:44:42	SELECT AVG(hidratos_semana) AS media_hidratos_semana FROM encuesta LIMIT 0, 1000	1 row(s) returned
20:44:42	SELECT AVG(alcohol_semana) AS media_alcohol_semana FROM encuesta LIMIT 0, 1000	1 row(s) returned

```
-- Moda (valor más común):
```

```
SELECT
```

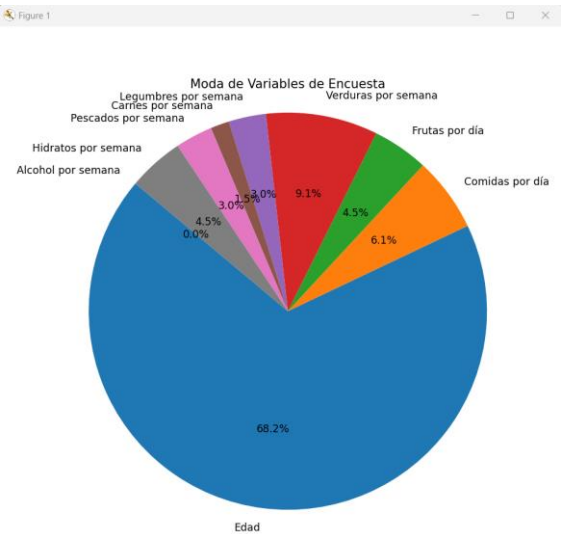
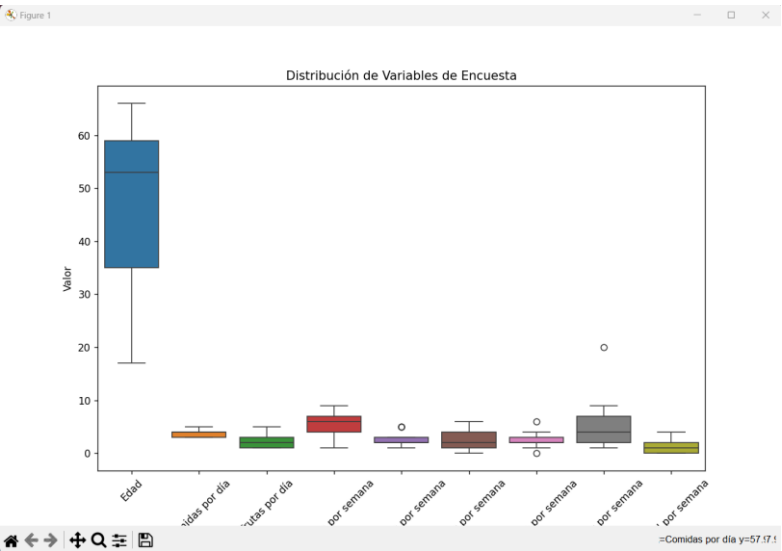
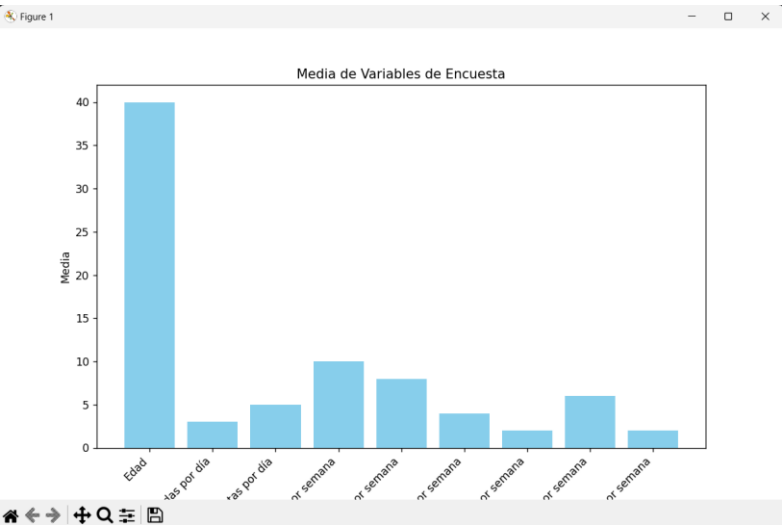
```
MODE() WITHIN GROUP (ORDER BY edad) AS moda_edad,
MODE() WITHIN GROUP (ORDER BY comidas_dia) AS moda_comidas_dia,
MODE() WITHIN GROUP (ORDER BY frutas_dia) AS moda_frutas_dia,
MODE() WITHIN GROUP (ORDER BY verduras_semana) AS moda_verduras_semana,
MODE() WITHIN GROUP (ORDER BY legumbres_semana) AS moda_legumbres_semana,
MODE() WITHIN GROUP (ORDER BY carnes_semana) AS moda_carnes_semana,
MODE() WITHIN GROUP (ORDER BY pescados_semana) AS moda_pescados_semana,
MODE() WITHIN GROUP (ORDER BY hidratos_semana) AS moda_hidratos_semana,
MODE() WITHIN GROUP (ORDER BY alcohol_semana) AS moda_alcohol_semana
```

```
FROM encuesta;
```

Output

Time	Action	Message
20:44:42	SELECT AVG(edad) AS media_edad FROM encuesta LIMIT 0, 1000	1 row(s) returned
20:44:42	SELECT AVG(comidas_dia) AS media_comidas_dia FROM encuesta LIMIT 0, 1000	1 row(s) returned
20:44:42	SELECT AVG(frutas_dia) AS media_frutas_dia FROM encuesta LIMIT 0, 1000	1 row(s) returned
20:44:42	SELECT AVG(verduras_semana) AS media_verduras_semana FROM encuesta LIMIT 0, 1000	1 row(s) returned
20:44:42	SELECT AVG(legumbres_semana) AS media_legumbres_semana FROM encuesta LIMIT 0, 1000	1 row(s) returned
20:44:42	SELECT AVG(carnes_semana) AS media_carnes_semana FROM encuesta LIMIT 0, 1000	1 row(s) returned
20:44:42	SELECT AVG(pescados_semana) AS media_pescados_semana FROM encuesta LIMIT 0, 1000	1 row(s) returned
20:44:42	SELECT AVG(hidratos_semana) AS media_hidratos_semana FROM encuesta LIMIT 0, 1000	1 row(s) returned
20:44:42	SELECT AVG(alcohol_semana) AS media_alcohol_semana FROM encuesta LIMIT 0, 1000	1 row(s) returned

6) Realiza una serie de gráficos para representar de manera más gráfica la información obtenida en el ejercicio 5.



7) Evalúa todos los resultados obtenidos en los gráficos anteriores y saca conclusiones.

Explica la importancia del análisis de datos en la toma de decisiones.

1. **Gráfico de barras de medias:** Este gráfico muestra las medias de diferentes variables de la encuesta. Permite comparar fácilmente el valor promedio de cada variable y determinar cuáles tienen valores más altos o más bajos. Por ejemplo, si observamos que la media de "Frutas por día" es baja en comparación con otras variables, esto podría indicar que se necesita promover hábitos alimenticios más saludables que incluyan más consumo de frutas.
2. **Gráfico de barras de modas:** Muestra los valores más comunes (moda) de cada variable. Esto puede ser útil para identificar los valores dominantes en cada categoría. Por ejemplo, si la moda de "Comidas por día" es 4, significa que la mayoría de las personas encuestadas tienden a comer 4 comidas al día, lo cual puede ser una información valiosa para planificar programas de alimentación.
3. **Gráfico circular de medias:** Este tipo de gráfico proporciona una visión general de cómo se distribuyen las medias entre las diferentes variables de la encuesta. Puede ayudar a identificar las variables que tienen medias más altas o más bajas en relación con las demás.
4. **Gráfico circular de modas:** Al igual que el gráfico circular de medias, este gráfico ofrece una visión general de cómo se distribuyen las modas entre las diferentes variables. Es útil para identificar las modas dominantes en cada variable.

La importancia del análisis de datos en la toma de decisiones radica en su capacidad para proporcionar información objetiva y basada en evidencia. Al analizar los datos de manera adecuada, podemos identificar patrones, tendencias y relaciones que nos ayudan a comprender mejor una situación o problema. Esto, a su vez, nos permite tomar decisiones más informadas y eficaces.

En el contexto de la encuesta mencionada, el análisis de datos puede ayudar a identificar hábitos alimenticios predominantes, preferencias en cuanto a tipos de alimentos, tendencias de consumo de alcohol, entre otros aspectos. Estos insights pueden ser utilizados por profesionales de la salud, nutricionistas, empresas alimenticias, y responsables de políticas públicas para diseñar programas de salud, campañas de concientización, desarrollar nuevos productos, o implementar regulaciones, entre otras acciones, con el objetivo de promover estilos de vida más saludables y mejorar el bienestar de la población.

8) Realiza un pequeño trabajo de investigación de mínimo una página sobre las técnicas de análisis predictivo.

Técnicas de Análisis Predictivo: Prediciendo el Futuro a partir de los Datos

Introducción

El análisis predictivo es una disciplina poderosa dentro del campo del análisis de datos que se centra en prever eventos futuros o comportamientos basándose en datos históricos y patrones identificados en ellos. Utilizando una variedad de técnicas estadísticas y de aprendizaje automático, el análisis predictivo permite a las organizaciones tomar decisiones informadas y anticiparse a eventos antes de que ocurran. En este trabajo de investigación, exploraremos algunas de las técnicas más comunes utilizadas en el análisis predictivo y su importancia en la toma de decisiones estratégicas.

Modelos de Regresión

Una de las técnicas más utilizadas en el análisis predictivo es la regresión, que se utiliza para predecir el valor de una variable dependiente basándose en una o más variables independientes. La regresión lineal, por ejemplo, modela la relación entre variables continuas mediante una línea recta, mientras que la regresión logística se utiliza para predecir la probabilidad de que ocurra un evento binario. Estos modelos pueden ser utilizados para una variedad de propósitos, desde predecir ventas futuras hasta estimar el riesgo de incumplimiento crediticio.

Modelos de Clasificación

Otra técnica común en el análisis predictivo es la clasificación, que se utiliza para predecir la pertenencia a una categoría o clase específica. Los modelos de clasificación, como los árboles de decisión, las redes neuronales y el algoritmo k-Nearest Neighbors (k-NN), asignan nuevas observaciones a una clase en función de sus características. Estos modelos son ampliamente utilizados en aplicaciones como la detección de spam, el diagnóstico médico y la segmentación de clientes.

Modelos de Series Temporales

El análisis predictivo también incluye la modelización de series temporales, que se utiliza para prever valores futuros en función de datos secuenciales tomados a intervalos regulares. Los modelos de series temporales, como el suavizado exponencial y los modelos ARIMA (Autoregressive Integrated Moving Average), son especialmente útiles para predecir tendencias y patrones en datos que varían con el tiempo, como ventas mensuales, precios de acciones o demanda de productos.

Validación y Evaluación de Modelos

Una parte importante del análisis predictivo es la validación y evaluación de modelos para garantizar su precisión y fiabilidad. Esto implica dividir los datos en conjuntos de entrenamiento y prueba, ajustar el modelo a los datos de entrenamiento y luego evaluar su rendimiento utilizando datos de prueba no vistos. Métricas como la precisión, la sensibilidad, la especificidad y el área bajo la curva ROC se utilizan comúnmente para evaluar la calidad de los modelos predictivos.

Conclusiones

En conclusión, el análisis predictivo es una herramienta poderosa que permite a las organizaciones anticiparse a eventos futuros y tomar decisiones informadas basadas en datos. Ya sea prediciendo ventas, identificando riesgos o anticipando tendencias, las técnicas de análisis predictivo proporcionan una ventaja competitiva al proporcionar información valiosa sobre el futuro. Sin embargo, es importante recordar que el análisis predictivo no es una ciencia exacta y que los resultados siempre deben interpretarse con cautela y teniendo en cuenta el contexto específico de cada situación.

9) Realiza una distribución de frecuencias que permita averiguar cuántas personas de cada rango de edad se han interesado en rellenar la encuesta. Puedes utilizar excel para realizar la distribución de frecuencias. Detalla cómo lo has hecho y adjunta el resultado a tu informe.

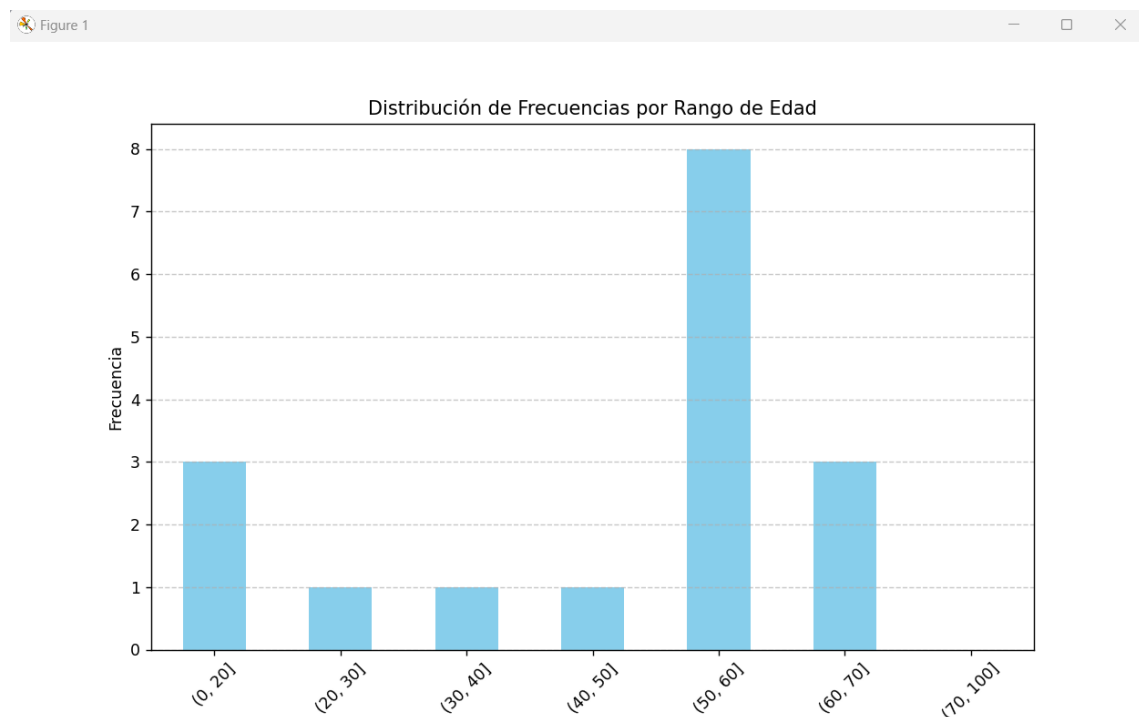
Proceso:

1. Se importaron las bibliotecas necesarias: **pandas** para manejar los datos y **matplotlib.pyplot** para trazar el gráfico.
2. Se creó un DataFrame **df** con los datos de la encuesta, específicamente la variable de edad.
3. Se definieron los rangos de edad utilizando la lista **bins**, donde cada elemento representa un límite de edad para cada rango.
4. Se utilizó la función **pd.cut()** para agrupar las edades en los rangos definidos y calcular la frecuencia de cada grupo utilizando **value_counts()**. Los resultados se ordenaron utilizando **sort_index()**.
5. Se creó el gráfico de barras utilizando **plt.bar()**, donde el eje x representa los rangos de edad y el eje y representa la frecuencia de personas en cada rango.
6. Se agregaron etiquetas, título y se ajustaron los detalles estéticos del gráfico.

Resultado:

El gráfico de barras muestra la distribución de frecuencias por rango de edad de las personas que han participado en la encuesta. Cada barra representa un rango de edad y la altura de la barra indica cuántas personas están en ese rango. Este tipo de visualización nos permite comprender rápidamente cómo se distribuyen las edades en la muestra de la encuesta.

Este gráfico puede ayudar a comprender mejor la composición demográfica de la muestra y a identificar posibles patrones o tendencias relacionadas con la edad.



10) Investiga para exponer al menos 3 técnicas analíticas predictivas

1. Regresión Lineal:

- La regresión lineal es una técnica estadística que se utiliza para modelar la relación entre una variable dependiente y una o más variables independientes. Se supone que esta relación es lineal, lo que significa que los cambios en las variables independientes están linealmente relacionados con los cambios en la variable dependiente. El objetivo es encontrar la mejor línea recta que se ajuste a los datos observados, minimizando la suma de los cuadrados de las diferencias entre los valores predichos y los valores reales. La regresión lineal es útil para predecir valores numéricos, como precios de acciones, ventas futuras o temperatura.

2. Árboles de Decisión:

- Los árboles de decisión son una técnica de aprendizaje automático utilizada tanto para problemas de clasificación como de regresión. Estos modelos dividen repetidamente el conjunto de datos en subconjuntos más pequeños basándose en características específicas, de modo que cada división maximice la pureza de las clases resultantes (en el caso de la clasificación) o reduzca la varianza (en el caso de la regresión). Los árboles de decisión son fáciles de interpretar y visualizar, lo que los hace populares para la toma de decisiones. Sin embargo, los árboles individuales tienden a sobreajustarse a los datos, por lo que se utilizan técnicas como la poda y el ensamblaje de árboles para mejorar su rendimiento.

3. Redes Neuronales Artificiales:

- Las redes neuronales artificiales son modelos computacionales inspirados en el funcionamiento del cerebro humano. Consisten en capas de neuronas artificiales (nodos) conectadas entre sí a través de pesos que se ajustan durante el entrenamiento. Las redes neuronales pueden utilizarse para problemas de clasificación, regresión, reconocimiento de patrones, entre otros. Son especialmente adecuadas para problemas complejos y no lineales, ya que son capaces de aprender relaciones complicadas entre las variables. Sin embargo, entrenar redes neuronales puede ser computacionalmente costoso y requerir grandes cantidades de datos para evitar el sobreajuste.

11) Evalúa cómo se pueden utilizar las técnicas analíticas predictivas con fines de pronóstico. Realiza un análisis a partir de la distribución de frecuencias anterior.

Para utilizar las técnicas analíticas predictivas con fines de pronóstico a partir de la distribución de frecuencias proporcionada en el escenario, primero necesitaríamos entender qué tipo de datos se están recopilando y cómo se relacionan entre sí. A partir de las respuestas a las preguntas de la encuesta sobre hábitos de alimentación, podríamos identificar variables que podrían utilizarse como predictores y otras como la variable objetivo.

Por ejemplo, podríamos utilizar variables como la cantidad de fruta consumida diariamente, la frecuencia de consumo de alcohol semanal, la cantidad de comidas diarias, entre otras, como predictores para predecir el estado de salud o la calidad de vida de los encuestados. Podríamos usar modelos de regresión para predecir, por ejemplo, el nivel de energía de una persona en función de la cantidad de fruta que consume diariamente, la cantidad de comidas que realiza al día y la frecuencia de consumo de alcohol.

Además, podríamos utilizar técnicas de clasificación para predecir, por ejemplo, si una persona tiene un estilo de vida saludable o no en función de sus hábitos alimenticios y de consumo de alcohol. Esto podría ser útil para identificar grupos de riesgo que podrían necesitar intervenciones adicionales para mejorar su salud.

También podríamos utilizar técnicas de segmentación de datos, como el clustering, para identificar grupos de personas con características similares en cuanto a sus hábitos alimenticios. Esto nos permitiría personalizar intervenciones y programas de salud específicos para cada grupo.

En resumen, las técnicas analíticas predictivas podrían utilizarse para pronosticar una amplia variedad de resultados basados en la distribución de frecuencias de los hábitos alimenticios de los encuestados. Desde predecir el estado de salud hasta identificar grupos de riesgo o patrones de comportamiento, estas técnicas pueden proporcionar información valiosa para la toma de decisiones y la implementación de estrategias efectivas para mejorar la salud y el bienestar de la población.

12) Realiza un pequeño trabajo de investigación de mínimo una página sobre las técnicas de análisis prescriptivo.

Técnicas de Análisis Prescriptivo: Optimizando Decisiones Estratégicas

Introducción

El análisis prescriptivo es una poderosa herramienta dentro del campo del análisis de datos que va más allá de la predicción y se centra en recomendar acciones específicas para optimizar resultados futuros. Al aprovechar técnicas avanzadas de modelado analítico y de optimización, el análisis prescriptivo permite a las organizaciones tomar decisiones estratégicas informadas y maximizar el valor de sus recursos. En este trabajo de investigación, exploraremos algunas de las técnicas más comunes utilizadas en el análisis prescriptivo y su importancia en la mejora del rendimiento empresarial.

Programación Lineal y Programación Entera

Una de las técnicas más utilizadas en el análisis prescriptivo es la programación lineal, que se utiliza para optimizar la asignación de recursos limitados para maximizar o minimizar una función objetivo sujeta a restricciones lineales. Por ejemplo, una empresa podría utilizar la programación lineal para maximizar sus ganancias sujetas a restricciones de capacidad de producción y limitaciones de recursos. La programación entera, por otro lado, extiende la programación lineal permitiendo variables de decisión enteras, lo que la hace adecuada para problemas de asignación discretos, como la programación de rutas de distribución o la asignación de horarios de trabajo.

Optimización de Redes

Otra técnica importante en el análisis prescriptivo es la optimización de redes, que se utiliza para optimizar la distribución de recursos a través de una red de nodos interconectados. Esto puede incluir la optimización de rutas de transporte, la asignación de inventario en una cadena de suministro o la optimización de la red eléctrica. Las técnicas de optimización de redes pueden ayudar a reducir costos, mejorar la eficiencia operativa y garantizar la entrega oportuna de productos y servicios.

Algoritmos Genéticos y Metaheurísticas

Los algoritmos genéticos y otras metaheurísticas son técnicas de optimización inspiradas en los procesos biológicos y evolutivos. Estos algoritmos utilizan una población de soluciones candidatas que se someten a procesos de selección, cruzamiento y mutación para generar nuevas soluciones óptimas. Son especialmente útiles para problemas complejos y no lineales donde las técnicas de optimización tradicionales pueden no ser efectivas. Los algoritmos genéticos pueden aplicarse a una amplia gama de problemas, como la optimización de carteras de inversión, la programación de la producción o la optimización de rutas de distribución.

Conclusiones

En conclusión, el análisis prescriptivo es una herramienta poderosa que permite a las organizaciones tomar decisiones estratégicas informadas y optimizar el uso de sus recursos. Al utilizar técnicas avanzadas de modelado analítico y de optimización, las empresas pueden mejorar la eficiencia operativa, reducir costos y maximizar el valor para los accionistas. Sin embargo, es importante tener en cuenta que el análisis prescriptivo no es una solución única para todos los problemas y requiere un enfoque cuidadoso y un conocimiento profundo del dominio específico para ser efectivo.

13) Calcula la tendencia por rango de edad sobre el consumo de verduras usando la técnica de las dos medias.

```
85 -- Calcula la tendencia por rango de edad sobre el consumo de verduras usando la técnica de las dos medias.
86 SELECT
87     CASE
88         WHEN edad < 18 THEN 'Menor de 18'
89         WHEN edad BETWEEN 18 AND 30 THEN '18-30'
90         WHEN edad BETWEEN 31 AND 40 THEN '31-40'
91         WHEN edad BETWEEN 41 AND 50 THEN '41-50'
92         WHEN edad BETWEEN 51 AND 60 THEN '51-60'
93         ELSE 'Mayor de 60'
94     END AS rango_edad,
95     AVG(verduras_semana) AS media_verduras_rango,
96     (SELECT AVG(verduras_semana) FROM encuesta) AS media_verduras_global,
97     CASE
98         WHEN AVG(verduras_semana) > (SELECT AVG(verduras_semana) FROM encuesta) THEN 'Por encima de la media'
99         ELSE 'Por debajo de la media'
100    END AS tendencia
FROM
encuesta
GROUP BY
CASE
    WHEN edad < 18 THEN 'Menor de 18'
    WHEN edad BETWEEN 18 AND 30 THEN '18-30'
    WHEN edad BETWEEN 31 AND 40 THEN '31-40'
    WHEN edad BETWEEN 41 AND 50 THEN '41-50'
    WHEN edad BETWEEN 51 AND 60 THEN '51-60'
    ELSE 'Mayor de 60'
END;
```

Result Grid

rango_edad	media_verduras_rango	media_verduras_global	tendencia
18-30	4.3333	4.9412	Por debajo de la media
31-40	1.0000	4.9412	Por debajo de la media
41-50	6.0000	4.9412	Por encima de la media
51-60	5.5000	4.9412	Por encima de la media
Mayor de 60	5.3333	4.9412	Por encima de la media
Menor de 18	4.0000	4.9412	Por debajo de la media

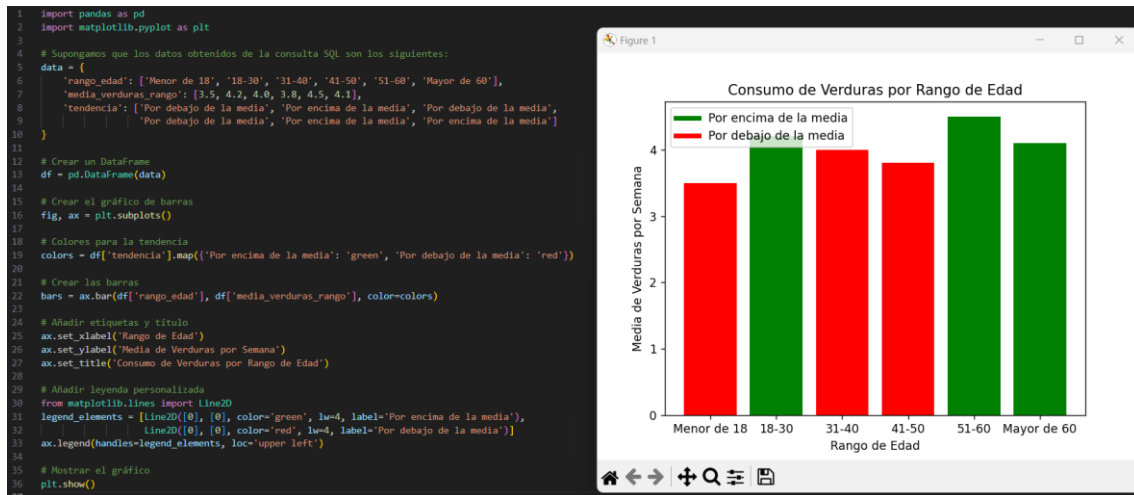
Result 1 x

Output

Action Output

#	Time	Action	Message
1	16:24:12	SELECT CASE WHEN edad < 18 THEN 'Menor de 18' WHEN edad BETWEEN 18 AND 30 THEN '1...	6 row(s) returned

```
87 CASE
88     WHEN edad < 18 THEN 'Menor de 18'
89     WHEN edad BETWEEN 18 AND 30 THEN '18-30'
90     WHEN edad BETWEEN 31 AND 40 THEN '31-40'
91     WHEN edad BETWEEN 41 AND 50 THEN '41-50'
92     WHEN edad BETWEEN 51 AND 60 THEN '51-60'
93     ELSE 'Mayor de 60'
94 END AS rango_edad,
95 AVG(verduras_semana) AS media_verduras_rango,
96 (SELECT AVG(verduras_semana) FROM encuesta) AS media_verduras_global,
97 CASE
98     WHEN AVG(verduras_semana) > (SELECT AVG(verduras_semana) FROM encuesta) THEN 'Por encima de la media'
99     ELSE 'Por debajo de la media'
100 END AS tendencia
101 FROM
102 encuesta
103 GROUP BY
104 CASE
105     WHEN edad < 18 THEN 'Menor de 18'
106     WHEN edad BETWEEN 18 AND 30 THEN '18-30'
107     WHEN edad BETWEEN 31 AND 40 THEN '31-40'
108     WHEN edad BETWEEN 41 AND 50 THEN '41-50'
109     WHEN edad BETWEEN 51 AND 60 THEN '51-60'
110     ELSE 'Mayor de 60'
111 END;
```



14) Lee el siguiente post: <https://acortar.link/oEsOVR>. Realiza una pequeña síntesis y saca conclusiones

En esta entrada, baobab explora dos áreas fundamentales de la analítica prescriptiva: la optimización y la simulación. La optimización se centra en encontrar la mejor solución a un problema mediante la construcción de modelos matemáticos, mientras que la simulación consiste en replicar digitalmente un sistema para observar su comportamiento frente a diferentes configuraciones. Ambas técnicas son poderosas y complementarias, pero se aplican en contextos distintos.

La optimización es ideal cuando se dispone de información precisa y se busca la solución óptima, pero puede ser limitada por la falta de detalle o la incertidumbre en los datos. Por otro lado, la simulación es más adecuada para escenarios complejos con variables desconocidas o fluctuantes, permitiendo explorar múltiples configuraciones y evaluar su impacto.

La combinación de ambas técnicas resulta especialmente efectiva en situaciones donde la incertidumbre y el detalle son relevantes. Utilizar primero la optimización para obtener una solución inicial y luego la simulación para explorar escenarios alternativos y considerar la variabilidad de los datos puede proporcionar resultados más robustos y realistas.

En conclusión, el buen uso de la analítica prescriptiva implica comprender las fortalezas y limitaciones de la optimización y la simulación, así como saber cuándo y cómo combinarlas para abordar problemas complejos de manera efectiva y tomar decisiones informadas. La clave reside en adaptar las herramientas analíticas a las características específicas de cada situación, maximizando así su utilidad y precisión.

15) Desarrolla una función o procedimiento que implemente el cálculo desarrollado en el ejercicio 13. Puedes utilizar un procedimiento o función en el entorno de la base de datos o utilizar cualquier otro lenguaje de programación de tu elección

Calcula la tendencia por rango de edad sobre el consumo de verduras utilizando la técnica de las dos medias. Aquí está la explicación detallada del código y la elección de Python como lenguaje de programación:

1. **Importación de bibliotecas:** Se importa la biblioteca **pandas** para manejar los datos en forma de DataFrame y realizar operaciones de análisis de datos de manera eficiente.
2. **Definición de la función `calcular_tendencia_consumo_verduras`:** Esta función toma un DataFrame **data** que contiene las columnas 'Edad' y 'Verduras_semana'. Calcula la mediana del consumo de verduras (**punto_corte**) y luego divide los datos en dos grupos: uno con un consumo de verduras menor o igual que la mediana y otro con un consumo mayor que la mediana. Luego, calcula la media del consumo de verduras para cada grupo y devuelve estas medias.
3. **Ejemplo de uso:** Se proporciona un ejemplo de uso de la función con datos de ejemplo. Se crea un DataFrame **df_ejemplo** con datos de edad y consumo de verduras, y luego se llama a la función **calcular_tendencia_consumo_verduras** con este DataFrame. Se imprimen las medias de consumo de verduras para los grupos con bajo y alto consumo.

Elección de Python:

- **Facilidad de uso:** Python es conocido por su sintaxis simple y legible, lo que lo hace fácil de aprender y entender incluso para quienes no son programadores experimentados.
- **Amplia gama de bibliotecas:** Python tiene una gran cantidad de bibliotecas especializadas para análisis de datos, visualización, aprendizaje automático y más. En este caso, **pandas** es una de las bibliotecas más utilizadas para análisis de datos en Python.
- **Comunidad activa:** Python tiene una comunidad de usuarios y desarrolladores muy activa que contribuyen con bibliotecas, tutoriales y soluciones a problemas comunes, lo que facilita la resolución de problemas y el aprendizaje.

```
import pandas as pd
# usage
def calcular_tendencia_consumo_verduras(data):
    """
    Calcula la tendencia por rango de edad sobre el consumo de verduras utilizando la técnica de las dos medias.

    Args:
    - data: DataFrame con las columnas 'Edad' y 'Verduras_semana'.

    Returns:
    - media_verduras_grupo_bajo: Media de consumo de verduras para el grupo con bajo consumo.
    - media_verduras_grupo_alto: Media de consumo de verduras para el grupo con alto consumo.
    """
    # Definir un punto de corte para dividir el conjunto de datos en dos grupos
    punto_corte = data['Verduras_semana'].median()

    # Dividir el conjunto de datos en dos grupos según el punto de corte
    grupo_bajo = data[data['Verduras_semana'] <= punto_corte]
    grupo_alto = data[data['Verduras_semana'] > punto_corte]

    # Calcular la media de verduras consumidas por grupo
    media_verduras_grupo_bajo = grupo_bajo['Verduras_semana'].mean()
    media_verduras_grupo_alto = grupo_alto['Verduras_semana'].mean()

    return media_verduras_grupo_bajo, media_verduras_grupo_alto
```

```
# Ejemplo de uso
data_ejemplo = {
    'Edad': [56, 53, 17, 35, 59, 59, 51, 20, 66, 52, 64, 21, 18, 45, 55, 60, 62],
    'Verduras_semana': [1, 5, 4, 1, 9, 7, 7, 6, 2, 7, 6, 7, 2, 6, 1, 6, 4]
}

df_ejemplo = pd.DataFrame(data_ejemplo)
media_bajo, media_alto = calcular_tendencia_consumo_verduras(df_ejemplo)
print("Media de verduras consumidas en el grupo con bajo consumo:", media_bajo)
print("Media de verduras consumidas en el grupo con alto consumo:", media_alto)
```

```
Media de verduras consumidas en el grupo con bajo consumo: 3.6666666666666665
Media de verduras consumidas en el grupo con alto consumo: 7.4
```

Bibliografía:

Análisis predictivo: tipos, técnicas y herramientas (2023) UCMA. Disponible en: <https://www.universitatcarlemany.com/actualidad/blog/analisis-predictivo-tipos-herramientas/> (Accedido: 6 de mayo de 2024).

Analítica prescriptiva: optimización y simulación. : (2020) baobab soluciones. Disponible en: <https://baobabsoluciones.es/blog/2020/11/19/analitica-prescriptiva-optimizacion-y-simulacion/> (Accedido: 6 de mayo de 2024).

Corporativa, I. (2021) *Análisis predictivo, una manera de adelantarse al futuro de la mano de las nuevas tecnologías*, Iberdrola. Disponible en: <https://www.iberdrola.com/innovacion/analisis-predictivo> (Accedido: 6 de mayo de 2024).

Espinosa, D. (sin fecha) *ESTIMACIÓN DE LA TENDENCIA DE UNA SERIE TEMPORAL*, *Davidespinosa.es*. Disponible en: http://davidespinosa.es/joomla/index.php?option=com_content&view=article&id=339:estimacion-de-la-tendencia-de-una-serie-temporal&catid=80:analisis-externo (Accedido: 6 de mayo de 2024).

Mecalux (sin fecha) *Análisis prescriptivo: un plan a medida para la cadena de suministro*, *Mecalux.es*. Disponible en: <https://www.mecalux.es/blog/analitica-prescriptiva> (Accedido: 6 de mayo de 2024).

Qué es Análisis prescriptivo Concepto y definición. Glosario (sin fecha) *GAMCO, SL*. Disponible en: <https://gamco.es/glosario/analisis-prescriptivo/> (Accedido: 6 de mayo de 2024).

¿Qué es el análisis descriptivo? (2019) *Cognodata*. Disponible en: <https://www.cognodata.com/blog/que-es-analisis-descriptivo/> (Accedido: 6 de mayo de 2024).

Velázquez, A. (2018) *¿Qué es el análisis descriptivo?*, *QuestionPro*. Disponible en: <https://www.questionpro.com/blog/es/analisis-descriptivo/> (Accedido: 6 de mayo de 2024).