

Inteligencia de Negocios

05 Septiembre 2023

Tarea U2T01: Tarea de exploración y análisis de un conjunto de datos proporcionado

Docente: MsC. Arlen Jeannette López

Tarea 1: Exploración y Análisis de Datos del Titanic con Python

En esta tarea, se trabajará con el conjunto de datos "Titanic", que contiene información sobre los pasajeros del famoso barco Titanic.

Objetivo:

- Utilizar Python y bibliotecas de análisis de datos para explorar y analizar el conjunto de datos del Titanic.
- Aplicar estadísticas descriptivas y visualización para obtener información relevante del conjunto de datos.

Instrucciones:

1. Preparación del Entorno:

- Asegúrate de tener Python y las bibliotecas `pandas`, `numpy`, `matplotlib` y `seaborn` instaladas en tu entorno de desarrollo.

2. Cargar el Conjunto de Datos:

- Descarga el conjunto de datos del Titanic desde el siguiente enlace:
<https://web.stanford.edu/class/archive/cs/cs109/cs109.1166/stuff/titanic.csv>

3. Utiliza la biblioteca `pandas` para cargar el conjunto de datos en un DataFrame. Asegúrate de manejar los valores faltantes adecuadamente.

4. Exploración Inicial:

- Muestra las primeras filas del conjunto de datos para familiarizarte con su estructura.
- Calcula estadísticas descriptivas básicas para las características numéricas, como la edad y la tarifa del pasaje.

5. Visualización de Datos:

- Crea visualizaciones para explorar los datos del Titanic. Puedes usar histogramas, gráficos de barras, gráficos de dispersión y gráficos de caja para mostrar la distribución y las relaciones entre las características.
- Presta atención a la edad, el género, la clase de pasaje y la supervivencia de los pasajeros.

6. Análisis de Supervivencia:

- Analiza la tasa de supervivencia en función de diferentes características como género, edad y clase de pasaje.
- ¿Existen diferencias significativas en la supervivencia entre grupos de pasajeros?

7. Identificación de Patrones:

- Identifica patrones o relaciones notables en los datos del Titanic.
- ¿Qué características parecen estar relacionadas con una mayor tasa de supervivencia?

8. Informe de Resultados:

- Prepara un informe que resuma tus hallazgos. Incluye visualizaciones y estadísticas destacadas.
- Describe cualquier patrón o relación que hayas identificado.
- Presenta tus conclusiones y observaciones en un formato claro y organizado.

Entrega:

- Los estudiantes deben entregar un cuaderno Jupyter (archivo `.ipynb`) que contenga todo el código y los informes escritos.
- Los informes deben ser claros y bien estructurados.

Tarea 2: Análisis Exploratorio de Datos Financieros en Python

Objetivo:

- Utilizar Python y bibliotecas de análisis de datos para realizar un análisis exploratorio de datos financieros.
- Aplicar estadísticas descriptivas y visualización para obtener información relevante del conjunto de datos.

Instrucciones:

1. Preparación del Entorno:

- Asegúrate de tener Python y las bibliotecas ``pandas``, ``numpy``, ``matplotlib`` y ``seaborn`` instaladas en tu entorno de desarrollo.

2. Cargar el Conjunto de Datos:

- Descarga un conjunto de datos financiero que contenga información sobre diferentes acciones o activos financieros. Puedes buscar conjuntos de datos financieros en fuentes como Yahoo Finance o usar APIs para obtener datos en tiempo real. <https://www.kaggle.com/datasets/atharvaarya25/financials>
- Utiliza la biblioteca ``pandas`` para cargar el conjunto de datos en un DataFrame. Asegúrate de nombrar las columnas adecuadamente, incluyendo el símbolo de la acción, la fecha y el precio de cierre.

3. Exploración Inicial:

- Muestra las primeras filas del conjunto de datos para familiarizarte con su estructura.
- Calcula estadísticas descriptivas básicas para el precio de cierre, como la media, la mediana, la desviación estándar, el mínimo y el máximo.

4. Visualización de Datos:

- Crea visualizaciones para explorar los datos financieros. Puedes usar gráficos de líneas para mostrar la evolución del precio de cierre a lo largo del tiempo.
- Presta atención a la tendencia general de los precios y busca eventos importantes que puedan haber afectado a las acciones.

5. Análisis de Rendimiento:

- Calcula el rendimiento diario de las acciones utilizando los precios de cierre. El rendimiento diario se calcula como el cambio porcentual en el precio de cierre con respecto al día anterior.
- Visualiza el rendimiento diario utilizando un gráfico de líneas.

6. Identificación de Patrones:

- Identifica patrones o tendencias notables en los datos financieros. ¿Hay períodos de volatilidad? ¿Existen eventos que hayan afectado significativamente el precio de las acciones?

7. Informe de Resultados:

- Prepara un informe que resuma tus hallazgos. Incluye visualizaciones y estadísticas destacadas.
- Describe cualquier patrón o relación que hayas identificado.
- Presenta tus conclusiones y observaciones en un formato claro y organizado.

Entrega:

- Debes entregar un cuaderno Jupyter (archivo `.ipynb`) que contenga todo el código y los informes escritos.
- Los informes deben ser claros y bien estructurados.

Tarea 3: Exploración y Análisis de Datos del Mercado de Bienes Raíces con Python

Objetivo:

- Utilizar Python y bibliotecas de análisis de datos para explorar y analizar un conjunto de datos del mercado de bienes raíces.
- Aplicar estadísticas descriptivas y visualización para obtener información relevante del conjunto de datos.

Instrucciones:

1. Preparación del Entorno:

- Asegúrate de tener Python y las bibliotecas `pandas`, `numpy`, `matplotlib` y `seaborn` instaladas en tu entorno de desarrollo.

2. Cargar el Conjunto de Datos:

- Descarga el conjunto de datos del mercado de bienes raíces desde el siguiente enlace: <https://www.kaggle.com/datasets/ahmedshahriarsakib/usa-real-estate-dataset>
- Utiliza la biblioteca `pandas` para cargar el conjunto de datos en un DataFrame. Asegúrate de nombrar las columnas adecuadamente.

3. Exploración Inicial:

- Muestra las primeras filas del conjunto de datos para familiarizarte con su estructura.
- Calcula estadísticas descriptivas básicas para las características numéricas, como el precio de venta, el número de habitaciones y el área del terreno.

4. Visualización de Datos:

- Crea visualizaciones para explorar los datos del mercado de bienes raíces. Puedes usar histogramas, gráficos de dispersión, gráficos de barras y gráficos de caja para mostrar la distribución y las relaciones entre las características.
- Presta atención a la distribución de precios, la relación entre el número de habitaciones y el precio, y otras características relevantes.

5. Análisis de Precios:

- Realiza un análisis más detallado de los precios de venta. Calcula la media, la mediana y la desviación estándar de los precios.
- Visualiza la distribución de precios mediante un histograma y un gráfico de caja.

6. Identificación de Patrones:

- Identifica patrones o tendencias notables en los datos del mercado de bienes raíces.
- ¿Existen características específicas que estén relacionadas con los precios más altos?

7. Informe de Resultados:

- Prepara un informe que resuma tus hallazgos. Incluye visualizaciones y estadísticas destacadas.
- Describe cualquier patrón o relación que hayas identificado.
- Presenta tus conclusiones y observaciones en un formato claro y organizado.

Entrega:

- Los estudiantes deben entregar un cuaderno Jupyter (archivo `.ipynb`) que contenga todo el código y los informes escritos.
- Los informes deben ser claros y bien estructurados.

Tarea 4: Análisis de Datos de Salud Pública en Python

Objetivo:

- Utilizar Python y bibliotecas de análisis de datos para realizar un análisis exploratorio de datos relacionados con la salud pública.
- Aplicar estadísticas descriptivas y visualización para obtener información relevante del conjunto de datos.

Instrucciones:

1. Preparación del Entorno:

- Asegúrate de tener Python y las bibliotecas ``pandas``, ``numpy``, ``matplotlib`` y ``seaborn`` instaladas en tu entorno de desarrollo.

2. Cargar el Conjunto de Datos:

- Descarga un conjunto de datos de salud pública que contenga información relevante, como tasas de enfermedades, factores de riesgo, o información demográfica de una población .
<https://www.kaggle.com/datasets/thedevastator/public-health-indicators-in-chicago>
- Utiliza la biblioteca ``pandas`` para cargar el conjunto de datos en un DataFrame. Asegúrate de nombrar las columnas adecuadamente.

3. Exploración Inicial:

- Muestra las primeras filas del conjunto de datos para familiarizarte con su estructura.
- Calcula estadísticas descriptivas básicas para las variables numéricas, como la media, la mediana, la desviación estándar, el mínimo y el máximo.

4. Visualización de Datos:

- Crea visualizaciones para explorar los datos de salud pública. Puedes usar gráficos de barras, histogramas, gráficos de dispersión y gráficos de líneas según la naturaleza de los datos.
- Presta atención a las tendencias en las tasas de enfermedades o a las relaciones entre factores de riesgo y la salud de la población.

5. Análisis de Correlación:

- Calcula la matriz de correlación entre las variables numéricas. Utiliza un mapa de calor (heatmap) para visualizar la matriz de correlación.
- Identifica cualquier correlación fuerte o inversa entre variables.

6. Identificación de Patrones:

- Identifica patrones o tendencias notables en los datos de salud pública. ¿Existen factores de riesgo que estén fuertemente relacionados con ciertas enfermedades?

7. Informe de Resultados:

- Prepara un informe que resuma tus hallazgos. Incluye visualizaciones y estadísticas destacadas.
- Describe cualquier patrón o relación que hayas identificado.
- Presenta tus conclusiones y observaciones en un formato claro y organizado.

Entrega:

- Debes entregar un cuaderno Jupyter (archivo `.ipynb`) que contenga todo el código y los informes escritos.
- Los informes deben ser claros y bien estructurados.

Tarea 5: Análisis de Datos Ambientales en Python

Objetivo:

- Utilizar Python y bibliotecas de análisis de datos para realizar un análisis exploratorio de datos relacionados con el medio ambiente.
- Aplicar estadísticas descriptivas y visualización para obtener información relevante del conjunto de datos.

Instrucciones:

1. Preparación del Entorno:

- Asegúrate de tener Python y las bibliotecas pandas, numpy, matplotlib y seaborn instaladas en tu entorno de desarrollo.

2. Cargar el Conjunto de Datos:

- Descarga un conjunto de datos ambientales que contenga información sobre la calidad del aire, niveles de contaminantes, o cualquier otro aspecto relacionado con el medio ambiente que te interese.
<https://www.kaggle.com/datasets/moazzimalibhatti/co2-emission-by-countries-year-wise-17502022>
- Utiliza la biblioteca pandas para cargar el conjunto de datos en un DataFrame. Asegúrate de nombrar las columnas adecuadamente.

3. Exploración Inicial:

- Muestra las primeras filas del conjunto de datos para familiarizarte con su estructura.
- Calcula estadísticas descriptivas básicas para las variables numéricas, como la media, la mediana, la desviación estándar, el mínimo y el máximo.

4. Visualización de Datos:

- Crea visualizaciones para explorar los datos ambientales. Puedes usar gráficos de líneas, gráficos de barras, histogramas y mapas de calor según la naturaleza de los datos.

- Presta atención a las tendencias en la calidad del aire o a las relaciones entre diferentes contaminantes.

5. Análisis de Tendencias Temporales:

- Si el conjunto de datos incluye información temporal, realiza un análisis de tendencias a lo largo del tiempo. Puedes usar gráficos de líneas para visualizar cómo han cambiado las condiciones ambientales con el tiempo.

6. Identificación de Patrones:

- Identifica patrones o tendencias notables en los datos ambientales. ¿Hay estaciones del año en las que la calidad del aire sea peor? ¿Existen correlaciones entre diferentes contaminantes?
- Informe de Resultados:
- Prepara un informe que resuma tus hallazgos. Incluye visualizaciones y estadísticas destacadas.
- Describe cualquier patrón o relación que hayas identificado.
- Presenta tus conclusiones y observaciones en un formato claro y organizado.

7. Entrega:

- Debes entregar un cuaderno Jupyter (archivo .ipynb) que contenga todo el código y los informes escritos.
- Los informes deben ser claros y bien estructurados.

Tarea 5: Análisis de Datos de Comercio Electrónico en Python

Objetivo:

- Utilizar Python y bibliotecas de análisis de datos para realizar un análisis exploratorio de datos relacionados con el comercio electrónico.
- Aplicar estadísticas descriptivas y visualización para obtener información relevante del conjunto de datos.

Instrucciones:

1. Preparación del Entorno:

- Asegúrate de tener Python y las bibliotecas pandas, numpy, matplotlib y seaborn instaladas en tu entorno de desarrollo.

2. Cargar el Conjunto de Datos:

- Descarga un conjunto de datos de comercio electrónico que contenga información sobre transacciones, productos vendidos o comportamiento de los clientes en línea. <https://www.kaggle.com/datasets/zusmani/pakistans-largest-ecommerce-dataset>
- Utiliza la biblioteca pandas para cargar el conjunto de datos en un DataFrame. Asegúrate de nombrar las columnas adecuadamente.

3. Exploración Inicial:

- Muestra las primeras filas del conjunto de datos para familiarizarte con su estructura.
- Calcula estadísticas descriptivas básicas para las variables numéricas, como la media, la mediana, la desviación estándar, el mínimo y el máximo.

4. Visualización de Datos:

- Crea visualizaciones para explorar los datos de comercio electrónico. Puedes usar gráficos de barras, histogramas, gráficos de dispersión y gráficos de líneas según la naturaleza de los datos.
- Presta atención a las tendencias en las ventas, la frecuencia de compras y el comportamiento de los clientes.

4. Segmentación de Clientes:

- Si el conjunto de datos incluye información de clientes, realiza una segmentación de clientes utilizando técnicas de clustering como K-Means.
- Visualiza los grupos de clientes y describe sus características distintivas.

5. Identificación de Patrones:

- Identifica patrones o tendencias notables en los datos de comercio electrónico. ¿Hay productos que se venden más que otros? ¿Existen segmentos de clientes que compran con más frecuencia?

6. Informe de Resultados:

- Prepara un informe que resuma tus hallazgos. Incluye visualizaciones y estadísticas destacadas.
- Describe cualquier patrón o relación que hayas identificado.
- Presenta tus conclusiones y observaciones en un formato claro y organizado.

Entrega:

- Debes entregar un cuaderno Jupyter (archivo .ipynb) que contenga todo el código y los informes escritos.
- Los informes deben ser claros y bien estructurados.

Estructura del informe:

Creating a structured paper for data analytics involves following a logical flow to effectively communicate your analysis and findings. Here's a suggested structure for a data analytics paper:

1. Title Page:

- Title of the paper
- Author(s) name(s)
- Affiliation(s)
- Contact information (email)

2. Abstract:

- A brief summary of the paper (150-250 words) that highlights the problem, methodology, key findings, and implications.

3. Introduction:

- Provide an introduction to the problem or question you're addressing.
- State the purpose and objectives of your analysis.
- Highlight the significance of the problem and why it's relevant.

4. Literature Review:

- Discuss relevant existing research, models, or theories related to your analysis.
- Show how your work builds upon or differs from previous research.

5. Data Collection and Preprocessing:

- Describe the data sources and data collection methods.
- Explain how the data was cleaned, transformed, and prepared for analysis.

6. Methodology:

- Explain the analytical methods, algorithms, or models you used.
- Discuss the rationale behind your choice of methods.
- Include any assumptions made during the analysis.

7. Results:

- Present your findings in a clear and organized manner.
- Use tables, charts, graphs, and visualizations to support your results.
- Discuss any patterns, trends, or insights you have discovered.

8. Discussion:

- Interpret the results in the context of your research objectives.
- Discuss the implications of your findings.
- Address any limitations or uncertainties in your analysis.

9. Conclusion:

- Summarize the key findings and their implications.
- Restate the research objectives and how they were met.
- Suggest areas for further research, if applicable.

10. References:

- List all the sources (books, articles, datasets, etc.) cited in your paper.
- Follow a specific citation style (e.g., APA, MLA, Chicago) consistently.

11. Appendices (if needed):

- Include supplementary information, code snippets, or additional data that supports your analysis but isn't essential for the main text.

12. Acknowledgments (optional):

- Thank individuals, organizations, or funding sources that contributed to your research.

Remember to maintain a clear and concise writing style throughout your paper, use proper headings and subheadings, and provide citations for any external sources or data used in your analysis. Tailor the structure to the specific requirements of your target audience or journal if you are submitting your paper for publication.

Examples: <https://www.analyticsinsight.net/top-10-must-read-data-science-research-papers-in-2022/>

Criterios de Evaluación	Puntos Asignados
Comprende y Manipula los Datos (10 puntos)	
Carga de Datos	3 puntos
Exploración Inicial	4 puntos
Manipulación y Preparación de Datos	3 puntos
Análisis y Descubrimiento de Patrones (10 puntos)	
Análisis de Datos	4 puntos
Interpretación de Resultados	4 puntos
Identificación de Patrones Relevantes	2 puntos
Visualización y Presentación de Datos (10 puntos)	
Calidad y Variedad de Visualizaciones	4 puntos
Presentación y Claridad	4 puntos
Uso Creativo de Gráficos	2 puntos
Comunicación y Calidad General (5 puntos)	
Comunicación Efectiva	2 puntos
Calidad General del Trabajo	3 puntos

