

Trabajo Final Exploración de Datos Multivariados

Primera Defensa: Semana del 22/04.

Indicaciones

La entrega y presentación es obligatoria para la regularidad o promoción de la materia y la fecha es inamovible.

El trabajo final constará de tres archivos: el primero es un JUPYTER con todo el análisis realizado y las conclusiones pertinentes, donde el código debe poder ejecutarse. El segundo archivo consiste en un documento Word donde se entregará el informe final seleccionando la información que surgió como importante del análisis realizado. La entrega se realiza por mail el día antes de la defensa. El tercer archivo es una presentación que se corresponde con la exposición de los resultados y puede enviarse el mismo día de la exposición conjunto con el código.

La defensa es una presentación de los hallazgos obtenidos, de una duración aproximada de 25 minutos por grupo, ante la profesora y compañeros.

Algunas cuestiones a tener en cuenta:

- Un cliente que llega con una base de datos muchas veces no sabe sobre la calidad de sus datos y la información que contiene su dataset, por lo cual siempre es imprescindible describir la base de datos recibida.
- Siempre hay alguna pregunta o prioridad a tratar de contestar, no obstante si bien el análisis debe poner el foco ahí, siempre es posible extraer información adicional que puede resultar de utilidad para el cliente, el análisis debe ser exhaustivo.
- No todas las variables son útiles, hay información que tal vez no tiene sentido considerar, y que el analista debe saber reconocer. Por ejemplo, el nombre de una persona en un análisis de venta.
- Por otra parte, siempre es posible generar nuevas variables que tal vez resulten de utilidad para el análisis, por ejemplo si se conoce la dirección de una persona, se puede generar una variable con la localidad o el barrio, y esto puede servir para segmentar los clientes de algún determinado rubro.
- La calidad de un informe no se mide por su extensión, puede haber un informe de cinco páginas claro y conciso que contenga toda la información relevante, como puede haber un informe de 50 páginas que redunde sobre lo mismo o muestre información irrelevante para el problema.
- El informe no debe contener ni todas las pruebas que se realizaron, ni todas las tablas, ni todos los gráficos, debe seleccionar o confeccionar solo aquellos que sirvan para comprender los datos y que sean de valor para el cliente. Es decir, si encontró que hay una correlación entre la altura y el peso, pero lo que está estudiando es el rendimiento académico de un estudiante, dicha información resultará irrelevante.

Respecto al informe

Elaborar un informe sobre un dataset es una tarea crucial para comunicar hallazgos y conclusiones derivadas del análisis de datos. Algunas recomendaciones y consideraciones importantes que un analista de datos debería tener en cuenta al elaborar dicho informe:

- **Comprender el contexto y los objetivos:** Antes de comenzar el análisis, es crucial comprender el contexto en el que se recopilaron los datos y los objetivos del análisis.
- **Limpiar y preprocesar los datos:** Antes de realizar cualquier análisis, es fundamental realizar una limpieza exhaustiva de los datos para eliminar valores atípicos, datos faltantes y errores. Además, es posible que sea necesario preprocesar los datos mediante técnicas como normalización, estandarización, codificación de variables categóricas, etc.

Es importante mencionar en el informe una descripción del estado en el que se encontró el dataset y mencionar las tareas que se debieron llevar a cabo para poder manipularlo. Luego, en conclusiones se pueden extraer recomendaciones para la toma futura de datos.

- **Exploración de datos:** Realiza un análisis exploratorio de los datos para comprender su estructura, distribución, relaciones entre variables y posibles patrones. Utiliza visualizaciones como gráficos, tablas y diagramas para comunicar estos hallazgos de manera efectiva.
- **Selección de técnicas analíticas adecuadas:** Elige las técnicas analíticas más apropiadas en función de los objetivos del análisis y la naturaleza de los datos. Esto puede incluir análisis estadísticos descriptivos, estadística inferencial, modelado predictivo, segmentación de datos, análisis de series temporales, etc.
- **Interpretación de resultados:** Asegúrate de interpretar los resultados de manera clara y concisa, evitando jerga técnica innecesaria. Explica las implicaciones prácticas de tus hallazgos y cómo se relacionan con los objetivos del análisis.
- **Presentación visual efectiva:** Utiliza visualizaciones efectivas y claras para comunicar tus hallazgos de manera visualmente atractiva. Asegúrate de que tus gráficos y tablas sean fáciles de entender y estén correctamente etiquetados.
- **Narrativa coherente:** Organiza tu informe de manera lógica y estructurada, siguiendo una narrativa coherente que guíe al lector a través de tus análisis, hallazgos y conclusiones.
- **Consideraciones éticas y de privacidad:** Si estás trabajando con datos sensibles, asegúrate de cumplir con todas las regulaciones de privacidad y ética de datos. Anonimiza los datos según sea necesario y protege la privacidad de los individuos involucrados.

Estructura del informe

Un informe completo sobre un dataset debería contener varias secciones clave para garantizar que los hallazgos y conclusiones se presenten de manera clara y comprensible. Aquí hay una estructura sugerida para un informe de análisis de datos:

- **Portada:**
 - Título del informe.
 - Nombre del analista de datos.
 - Fecha de creación del informe.
- **Resumen**
 - Breve descripción del problema o pregunta de investigación.
 - Objetivos del análisis.
 - Principales hallazgos y conclusiones.
 - Recomendaciones clave.
- **Introducción:**
 - Contexto del análisis.
 - Objetivos del informe.
 - Descripción del dataset y su relevancia para el análisis.
- **Metodología:**
 - Descripción de las técnicas y herramientas utilizadas en el análisis.
 - Detalles sobre la limpieza, preprocesamiento y exploración de datos.
 - Explicación de los modelos o métodos analíticos aplicados.
- **Resultados:**
 - Presentación de los principales hallazgos derivados del análisis.
 - Utilización de tablas, gráficos y visualizaciones para respaldar los resultados.
 - Interpretación de los hallazgos y su relevancia para los objetivos del análisis.

- **Discusión:**

- Análisis más detallado de los resultados.
- Exploración de patrones, tendencias o relaciones importantes encontradas en los datos.
- Comparación con investigaciones anteriores o referencia a la literatura relevante.

- **Conclusiones:**

- Recapitulación de los principales hallazgos.
- Respuesta a los objetivos del análisis.
- Implicaciones prácticas de los resultados.
- Limitaciones del estudio y áreas para futuras investigaciones.

- **Recomendaciones:**

- Sugerencias para acciones futuras basadas en los hallazgos del análisis.
- Estrategias para mejorar el rendimiento, tomar decisiones informadas o abordar desafíos identificados.

- **Referencias:**

- Cualquier fuente de datos, literatura o herramientas utilizadas en el análisis.

Criterios de Evaluación:

- **Entrega y Cumplimiento de Plazos:**

- Se evaluará la entrega puntual y completa de los archivos requeridos según las indicaciones establecidas.
- La fecha de entrega es inamovible y la presentación es obligatoria para la regularidad o promoción de la materia.

- **Calidad Técnica y Ejecución del Código:**

- Se verificará que el código proporcionado sea ejecutable sin errores y esté bien documentado.
- Se evaluará la calidad técnica del análisis realizado, incluyendo la limpieza, preprocesamiento y exploración de los datos.

- **Coherencia y Claridad en el Informe:**

- Se evaluará si el informe sigue las pautas establecidas, incluyendo la estructura sugerida y la coherencia del texto.
- Se considerará la capacidad del informe para comunicar de manera clara y concisa los hallazgos y conclusiones derivados del análisis.

- **Análisis de Datos:**

- Se examinará la profundidad y la variedad de los análisis realizados, así como la aplicación adecuada de los contenidos de la materia.
- Se evaluará la capacidad del equipo para seleccionar y aplicar técnicas analíticas apropiadas para abordar los objetivos del análisis.

- **Comunicación Oral de Resultados:**

- Se evaluará la capacidad de ambos integrantes del equipo para comunicar de manera efectiva las pruebas realizadas y los resultados obtenidos durante la defensa del trabajo final.
- Se considerará la demostración de un claro conocimiento del dataset y de los contenidos de la asignatura durante la presentación oral.

- **Presentación Visual de Resultados:**

- Se tendrá en cuenta el impacto visual de la presentación, incluyendo la selección y elaboración de gráficos, tablas y visualizaciones.
- Se evaluará la efectividad de la presentación en la comunicación de los hallazgos y conclusiones derivados del análisis de datos.

Espero que disfruten la ardua tarea de poner en juego lo aprendido.