Caso Práctico

1. Descripción

El objetivo del caso práctico es crear una visualización siguiendo las etapas y los principios de diseño vistos en la asignatura. El caso práctico se desarrollará sobre una fuente de datos que el estudiante consideré de su interés. Deberá plantear unos objetivos globales del análisis del dataset y responder mediante el diseño e implementación de una visualización orientada al objetivo.

En los siguientes apartados se detallan las características del trabajo, los criterios de evaluación y algunos ejemplos de propuestas de trabajos realizados en cursos anteriores.

Para realizar la visualización se utilizará python y librerías de visualización, preferiblemente Altair, pero si se necesita generar algún tipo de visualización que no ofrezca Altair, se pueden utilizar otras librerías.

2. Criterios de Evaluación

En la evaluación del caso práctico se tendrán en cuenta los siguientes puntos:

- **Complejidad de las fuentes de datos**: Se considerará un conjunto de datos complejo a aquellos conjuntos de datos derivados del Big Data (con gran volumen, variedad de datos y cierta velocidad de producción de datos) o bien conjuntos de datos cuya naturaleza sea desestructurada (textos, videos, audios,...), se encuentren en distintos repositorios, etc.

Por ejemplo, el dataset debería tener las siguientes características:

- Número de tablas/clases: > 5 tablas
- Número de instancias o casos: > 1000 instancias
- Tipos de características: como en cualquier sistema, deberá haber características propias de los objetos y relaciones entre objetos.
    - Número de atributos propios de cada instancia (>5 atributos): permitirán filtrar, agrupar, ...

- Deberá haber características espaciales y temporales para que sea posible distintos tipos de visualizaciones.
- Número de relaciones entre clases (>2): permitirán analizar interacciones entre objetos.

- **Funcionamiento de la aplicación**: la aplicación debe ser autocontenida y ejecutarse sin errores. Se valorará positivamente que se haya publicado en internet, la inclusión de elementos interactivos en la visualización, etc.
- **Claridad de la memoria**: la memoria está bien organizada, el objetivo de la visualización está claro, los pasos realizados en el análisis son correctos y están bien explicados, etc.
- **Uso correcto de los gráficos**: los gráficos utilizados están en consonancia con los objetivos o subobjetivos marcados, el gráfico presenta las variables adecuadas y con el nivel de detalle y precisión necesario, el diseño elegido es correcto y los distintos elementos están proporcionados, etc.
- **Interpretación de los gráficos**: en la memoria, los gráficos utilizados se han interpretado correctamente y las conclusiones obtenidas son correctas.
- **Integración de la visualización:** los gráficos que responden a subobjetivos parciales están perfectamente integrados dentro de la visualización final, se han aplicado correctamente los principios de composición y diseño para hacer una visualización estéticamente correcta y orientada al objetivo.

En caso de que se detecte alguna forma de plagio en alguna parte o la totalidad del caso práctico, el/la estudiante será suspendido por todo el curso.

3. Entregas

Habrá dos convocatorias para la entrega del caso práctico: la convocatoria ordinaria, en la que se realiza el trabajo durante el cuatrimestre en el que se imparte la asignatura, y la convocatoria extraordinaria de septiembre. Las fechas de las entregas se publicarán en el curso docente con suficiente antelación.

Se realizarán varias entregas relacionadas con el caso práctico: una entrega inicial con la propuesta del caso práctico, que es obligatoria y debe recibir el visto bueno del equipo docente para continuar, una entrega intermedia que es optativa, y una entrega final. La única entrega que se evaluará será la entrega final, pero las entregas inicial e intermedia son importantes ya que servirán para que el equipo docente proporcione realimentación al alumno durante el desarrollo del caso práctico. A continuación se explican en detalle las diferentes entregas.

• Entrega 1: Propuesta del caso práctico.

En esta primera entrega, el/la estudiante deberá definir el objetivo de la visualización, las fuentes de datos que va a emplear, así como una breve planificación de los subobjetivos y la manera de abordarlos.

Esta entrega es obligatoria, y el/la estudiante debe contar con el visto bueno del equipo docente para continuar con el caso práctico.

Habrá dos posibilidades de entrega: una al principio del cuatrimestre en el que se cursa la asignatura y otra antes de comenzar el periodo estival para poder entregar el trabajo final en septiembre. Ambas fechas se publicarán convenientemente en el curso virtual.

El formato de esta entrega será un documento PDF que se debe subir a la plataforma docente en el lugar habilitado (se accede a través del icono Tareas). Para facilitar la tarea de evaluación, el nombre del fichero seguirá este patrón: E1-<apellidos-nombre-estudiante>.pdf

• Entrega 2: Entrega intermedia.

Esta entrega se realizará a principios de mayo y servirá para detectar deficiencias en la memoria (esta entrega no estará disponible para la convocatoria de septiembre). La memoria entregada deberá contener, al menos, los siguientes apartados, y su tamaño estará limitado:

- Portada (Título, autor y resumen) (1 pg.)
- Planteamiento del problema y objetivos de la visualización (1-2 pgs)
- Preparación de los datos (preprocesado) (máx. 3 pgs)
- Procesado y análisis (estudios y visualizaciones parciales) (máx. 8 pgs)
- Visualización (integración de los resultados parciales) (máx. 4 pgs)
- Discusión, Conclusiones y posibles mejoras. (máx. 2 pgs)

Al igual que la entrega 1, el formato de esta entrega será un documento PDF que se debe subir a la plataforma virtual docente en el lugar habilitado (acceso desde el icono *Tareas*). Esta entrega no es obligatoria pero el/la estudiante debe ser consciente de que es una oportunidad para recibir realimentación por parte del equipo docente. Obviamente, la realimentación recibida dependerá del grado de elaboración de la memoria entregada. Para facilitar la tarea de evaluación, el nombre del fichero seguirá este patrón: E2-<apellidos-nombre-estudiante>.pdf

Es importante tener en cuenta que esta memoria servirá para evaluar la asignatura de "Visualización de Datos", por tanto, es importante no solo el resultado final, la visualización, sino justificar las decisiones de diseño tomadas. Por ejemplo, no vale con seleccionar una paleta de colores para uniformizar la visualización, sino que es necesario justificar por qué se ha elegido esa paleta, explicitando la relación de los colores elegidos con el objetivo de la visualización.

• Entrega 3: Entrega final.

Esta es la única entrega evaluable y deberá incluir un fichero comprimido (E3-<apellidos-nombre-estudiante>.zip) con los siguientes elementos:

- E3-<apellidos-nombre-estudiante>.pdf: La memoria del caso práctico, que incluirá la explicación del trabajo realizado, la implementación y la visualización resultante (ver apartados en el punto anterior relativo a la entrega intermedia).
- Todos los ficheros, tanto de código como de datos (dataset), necesarios para que el equipo docente pueda ejecutar la visualización. Si, además, se ha subido a internet y se puede ejecutar en un explorador, se asegura de que funciona correctamente.
- Vídeo-presentación de la visualización (solo en el caso de presentación offline).

La presentación del trabajo realizado se ha considerado como una manera de obtener el 10% (1 punto) dedicado a "Otras actividades evaluables" en la evaluación de la asignatura. Se proponen dos opciones: presentación online y offline. La presentación online se realizará en directo y consistirá en 10 minutos de exposición por parte del estudiante, más un turno de preguntas por parte de los evaluadores para aclarar o profundizar en algún punto. La presentación offline consistirá en un video de 10 minutos máximo en formato mp4.

Dado que el tiempo está muy ajustado, la presentación se debe centrar en lo importante, aunque se dejen cosas sin contar. Ya sabemos que el trabajo realizado es mayor que el presentado, pero esto se valora indirectamente en el resultado final. En ambos casos recomendamos realizar varias pruebas, grabarlas y analizarlas para refinar la presentación final.

Dado que la presentación online se considera más exigente que la offline, la valoración online permitirá obtener hasta 1 punto, mientras que la presentación offline solo permitirá obtener hasta 0.7 puntos.

4. Ejemplos de propuestas

Lo ideal a la hora de realizar el caso práctico sería seguir la secuencia natural de tareas:

1) proponerse un objetivo,
2) analizar qué datos se necesitan,
3) buscar los datos a partir de distintas fuentes e integrarlos,
4) realizar el estudio exploratorio y sacar resultados parciales
5) integrarlos en la visualización final.

Sin embargo, esto podría llevar demasiado tiempo y no podríamos asegurar que encontrarais finalmente los datos necesarios para cumplir con el objetivo. En su lugar, os proponemos que, partiendo también de un objetivo inicial que os interese, busquéis los datasets relacionados pero que, en caso de no encontrar todos los datos que necesitáis, retoquéis ligeramente los objetivos para adaptarlos a los datos existentes.

Fijado el objetivo, será necesario encontrar qué relaciones entre variables, métricas y/o descriptores (KPI) mejor se relacionan con el objetivo global desde algún punto de vista o perspectiva (estos serán los subobjetivos en los que se descompone el objetivo general). Para cada uno de estos subobjetivos, se realizarán análisis parciales y, finalmente, una visualización final donde se integren estas distintas perspectivas para proporcionar una respuesta al objetivo global.

A modo de ejemplo, supongamos que tenemos los datos de ventas de las tiendas de una compañía X implantada a nivel mundial. Primero, tenemos que fijar un objetivo:

- Analizar el estado de la compañía.
- Analizar la evolución de la compañía.
- Analizar la manera de aumentar las ventas.
- Analizar las tiendas / departamentos más / menos rentables.
- ...

El objetivo deberá ir acompañado de un contexto y un público receptor del trabajo (no es lo mismo realizar una visualización para el CEO que para publicarlo en un periódico). En cualquier caso, este objetivo global podrá descomponerse en distintas preguntas más concretas (cada pregunta dará lugar a una visualización parcial):

- ¿Quién/dónde/cuándo se obtiene mayor rendimiento? (Por tipo de tienda, por departamento, por vendedor, por producto, por zona, por fecha, ...)
- ¿Los días festivos se vende más? ¿Merece la pena tener las tiendas abiertas? ¿Qué hay de las fiestas escolares?

- ¿Convendría tener más refuerzo en las tiendas? ¿cuándo?
- ¿Cómo afectan a las tiendas de la compañía X la existencia de tiendas cercanas de la competencia (compañías Y1, Y2, …)?
- ¿Qué campañas funcionan mejor (para vender más, para atraer más clientes, etc.)?
- ¿Cuál es la evolución de la empresa? ¿Se ve un crecimiento en los últimos años? ¿Sería buena idea abrir más tiendas?
- …

Finalmente, estas visualizaciones parciales se deberán integrar en una visualización final que responda al objetivo global de la visualización.

También a modo de ejemplo, a continuación os listamos algunos ejemplos de trabajos que se han realizado en cursos anteriores:

- Análisis de atentados a nivel mundial (1997-2017)
  Resumen: "Los atentados suceden durante todo el año en gran cantidad de países, la mayoría en países del tercer mundo y por consecuencia no nos llegan noticias de estos atentados. En este trabajo se realiza una visualización de los atentados en el mundo entre 1997 y 2017 donde se analizan diferentes tipos de ataques, los tipos de armas empleadas, y objetivos de dichos ataques."

- Efecto de la pandemia en el alquiler de apartamentos de uso turístico.
- Análisis del olivar en Andalucía en el siglo XXI.
- Análisis en tiempo real de los precios de los Carburantes en gasolineras de España.
- Análisis de la relación entre inmigración/emigración y empleo en Barcelona (por barrios).
- Influencia del tráfico en la calidad del aire en Madrid (por barrios)
- Análisis del impacto de las noticias más importantes de cada día en el mercado de valores en EE. UU.
- Evolución meteorológica en los Países Bajos desde 1901 hasta 2020
- Características de los éxitos musicales de la década del 2000.
- Análisis del precio de los jugadores de futbol a nivel mundial.
- Análisis de perfiles de jugadores del juego "League of Legends".
- …

El equipo docente recomienda las siguientes fuentes de datos abiertas, además de aquellas que se sugieran a lo largo del curso:

- Datos en abierto de la UNED: http://gesmant.uned.es/otom_opendata/

- Datos en abierto del Gobierno de España: https://datos.gob.es/es/catalogo
- Datasets de la plataforma para competición de Machine Learning Kaggle: https://www.kaggle.com/datasets
- Datos en abierto del ayuntamiento de Madrid: https://datos.madrid.es/portal/site/egob/
- Portal de datos en abierto de la Unión Europea: https://data.europa.eu/euodp/es/data/
- Portal de datos en abierto IEEE: https://ieee-dataport.org/
- Generador de conjuntos de datos aleatorios: https://www.mockaroo.com/