Organización de Datos 75.06. Primer Cuatrimestre de 2018. Examen parcial, primera oportunidad: Criterio

1- Es importante filtrar los datos que son necesarios antes de comenzar a trabajar, si no lo hacen se descuenta un min de 5ptos. Hay descuentos de 3 ptos si realizan operaciones de mas, o ineficientes (por ejemplo realizar un takeordered cuando necesitan solo obtener mínimo).
Si los formatos para realizar el join no se corresponde a (K, V) descuento de 5 puntos.

2- Una opción para la resolución es generar una nueva columna en el Data Frame que podamos usar para saber en qué casos estuvo involucrado Batman. Para ello se puede realizar un join filtrando aquellos casos que tuvieron respuesta a la batiseñal (respuesta en 1), más allá de que hay distintas formas de hacerlo. Se descuentan puntos si la solución no es eficiente, o si se intentan realizar a mano operaciones que ya están resueltas por funciones de pandas..

3- V/F: Respuestas sin justificación valen cero.

a- Falso, nos quedaremos con los valores singulares más significativos (mayores) los cuales además de permitirnos acumular la mayor cantidad de energía de la matriz, serán los que describen la dimensión intrínseca de los datos.

b- Falso, es solamente recomendable realizar una reducción de dimensiones si por motivos específicos (propios del algoritmo de Machine Learning, de la cantidad de datos o de los medios en los que podemos ejecutar el algoritmo, por ejemplo hardware) no podemos ejecutarlo con toda la información. Siempre será recomendable como etapa inicial usar toda la información disponible. Esto se basa en el Teorema Fundamental de la Dimensionalidad.

c- Falso, T-SNE tiene como objetivo que dos puntos que estaban cercanos en el espacio original permanezcan cercanos en el espacio reducido, pero nada podemos inferir de los puntos que originalmente se encontraban alejados.

d- Verdadero, ya que con ISOMAP aplicamos MDS a la matriz de distancia de todos los puntos contra todos.

4- Se evalúa la resolución general, si no aplican correctamente el esquema planteado, o si no evalúan los candidatos de a pares se realizan descuentos en base al tipo de error cometido.

5-a) Si aplican mal las fórmulas cero. Si no aplican corrección de Laplace -3. Errores de cuentas no descuentan salvo que sea evidente que el resultado obtenido está mal, en ese caso descuenta por error conceptual en base al tipo de error cometido.
Si dicen que los documentos recuperados son todos, descuento de 3 puntos.
Si solo dicen que se recupera D6, descuento de 3 puntos.
Si calculan mal en b) precisión, recall o F1 -2 por cada uno.

6- Debe comenzar dándole más probabilidad a los bytes bajos. No debería tener en cuenta contextos, ya que BS los rompe. Soluciones que comiencen con todos los caracteres equiprobables valen 0 puntos, ya que no es lo que se puede suponer luego de MTF+BS. Es importante que se describa una solución, mencionar métodos sin explicar cómo se podrían utilizar no se consideran una solución válida.

7- Existen distintos enfoques que se pueden encarar, pero a partir del enunciado se puede ver claramente que el mismo puede considerarse una visualización que nos permita mostrar una serie temporal en la que se pueden mostrar:

· Variación de la tasa de resolución de crímenes en una cierta categoría con la participación de Batman.
· Relación entre crímenes sin resolver y crímenes resueltos a lo largo del tiempo desde la aparición de Batman. (que podrían por ejemplo mostrarse para una categoría específica que se quiera indicar como la de crimen organizado).
Esto puede lograrse con line plots de y axes, bar plots o stacked bar plots dependiendo de que se quiera mostrar.
La propuesta debe comunicar efectivamente lo que se desee mostrar. Descuento de puntaje si la visualización no tiene título, ni ejes rotulados.