

Laboratorio 7.

Análisis de redes sociales

INSTRUCCIONES:

Puede realizar uno de los 3 problemas que se le presentan. Debe extraer los datos de una red social del tema que escoja trabajar. Debe hacer un análisis exploratorio para entender mejor los datos. Debe descubrir patrones interesantes. Recuerde que la investigación debe ser reproducible por lo que debe guardar el código que ha utilizado para resolver los ejercicios y/o cada uno de los pasos llevados a cabo si utiliza una herramienta visual. Este laboratorio debe realizarse en grupos de 3. Inscribese en uno de los grupos que hay en canvas para la actividad.

PROBLEMAS A RESOLVER

PROBLEMA 1.

Extraiga los datos relacionados con el covid en Guatemala, puede utilizar cualquier red social, a la que se tenga acceso, y pueden extraerse datos de varias. Por ejemplo si extrae datos de twitter podría utilizar el hashtag #COVID19gt o #CoronavirusGT. Explore los datos que extraiga y descubra conocimiento, tendencias y elementos interesantes.

PROBLEMA 2.

Extraiga los datos relacionados con el tráfico en la ciudad de Guatemala, puede utilizar cualquier red social, a la que se tenga acceso, y pueden extraerse datos de varias. Por ejemplo si extrae datos de twitter podría utilizar los hashtag #TraficoGT, #TransitoGT o @amilcarmontejo . Explore los datos que extraiga y descubra conocimiento, tendencias y elementos interesantes. ¿Será que surgió efecto para disminuir el tráfico la estrategia de circulación de placas pares o impares?

PROBLEMA 3.

Extraiga los datos de las redes sociales de una empresa, una campaña de marketing de una empresa o el lanzamiento de un producto. Agregue la descripción de la empresa, campaña o producto y analice las interacciones de los clientes con ellos, determine el éxito de la campaña, producto o empresa en las interacciones de los clientes.

PROBLEMA 4.

Extraiga los datos relacionados con la popularidad del gobierno en curso en Guatemala, puede utilizar cualquier red social, a la que se tenga acceso, y pueden extraerse datos de varias. Explore los datos que extraiga y descubra conocimiento, tendencias y elementos interesantes. ¿Será que el actuar del presidente durante la pandemia hizo que su popularidad aumentara o disminuyera? ¿Encontró algún influencer?

EJERCICIOS

1. Descargue los archivos de datos
2. Cargue los archivos de datos a R o a Python o cualquier herramienta visual que utilice.
3. Limpie y preprocese los datos. Describa de forma detallada las actividades de preprocesamiento que llevó a cabo.
 - 3.1. Se pueden hacer tareas como:
 - Convertir el texto a mayúsculas o a minúsculas
 - Quitar los caracteres especiales que aparecen como “#”, “@” o los apóstrofes.
 - Quitar las url
 - Revisar si hay emoticones y quitarlos
 - Quitar los signos de puntuación
 - Quitar los artículos, preposiciones y conjunciones (stopwords)
 - Quitar números si considera que interferirán en las predicciones.
4. Haga un análisis exploratorio de los datos para entenderlos mejor, documente todos los análisis. Escriba una serie de insights que se puedan seguir investigando.
5. Llegue a conclusiones sobre lo que descubrió.

EVALUACIÓN

(15 puntos) Extracción de información:

- Está documentado el proceso de extracción de los datos de las redes sociales. Se extraen suficientes datos para descubrir información.

(25 puntos) Análisis exploratorio:

- Se elaboró un análisis exploratorio en el que se explican los cruces de variables, hay gráficos explicativos y análisis que permiten comprender el conjunto de datos.

(20 puntos) Limpieza y preprocesamiento de los datos:

- Se documentan las tareas de limpieza, incluyendo los paquetes/módulos que se usaron.

(30 puntos) Descubrimiento de información.

- Se describe de forma comprensible y amena los descubrimientos encontrados. Se calculan algunas métricas. Se detalla la información que es interesante.

(10 puntos) Conclusiones.

- Se llega a conclusiones sobre los descubrimientos hechos utilizando los análisis.

MATERIAL A ENTREGAR

- Archivo .pdf con el informe que contenga, los resultados de los análisis y las explicaciones.
- Link de Google drive donde trabajó el grupo.
- Script de R (.r o .rmd) o de Python que utilizó.
- Link del repositorio usado para versionar el código.

FECHAS DE ENTREGA

- **AVANCE:** Descarga de los datos de las redes sociales seleccionadas. Descripción de los datos (tamaño, origen y tipo de corpus), preprocesamiento y sus explicaciones. Planificación de las tareas para el grupo: jueves 3 de septiembre de 2020 a las 19:00.
 - **DOCUMENTO FINAL COMPLETO:** domingo 6 de septiembre de 2020 a las 23:59
- NOTA:** Solo se calificará el Documento Final si está entregado el avance con todo lo que se pide.