

Laboratorio 8. Visualización

INSTRUCCIONES:

Debe extraer los datos de la importación de vehículos de la página web de la SAT (<https://portal.sat.gob.gt/portal/alza-e-importacion-vehiculos/#1510763502681-dff4b62b-fd76>). Haga un análisis exploratorio para entender mejor los datos. Debe descubrir patrones interesantes, específicamente sobre las motos. Cruce la información con el conjunto de datos de accidentes de tránsito que provee el Instituto Nacional de Estadística (INE) (<https://www.ine.gob.gt/ine/estadisticas/bases-de-datos/accidentes-de-transito/>). Recuerde que la investigación debe ser reproducible por lo que debe guardar el código que ha utilizado para resolver el ejercicio y/o cada uno de los pasos llevados a cabo si utiliza una herramienta visual. Este laboratorio debe realizarse en grupos de 3. Inscribise en uno de los grupos que hay en canvas para la actividad.

PROBLEMA A RESOLVER

Debe mostrar patrones y resultados importantes sobre importación, exportación de motos y el impacto que causan en la sociedad guatemalteca. Presentar el resultado de aplicación de modelos mediante una visualización estática en forma de infografía comprensible para todo público.

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.
4. Haga un análisis exploratorio de los datos para entenderlos mejor, documente todos los análisis.
5. Elabore una visualización estática en forma de infografía para dar información a una audiencia no especializada en análisis estadístico. En ella debe presentar de una forma comprensible los resultados más relevantes, importantes y/o interesantes que encontró en la exploración que hizo de los datos de importaciones de vehículos de la Superintendencia de Administración Tributaria y del conjunto de datos de Accidentes de tránsito. Tenga en cuenta aspectos como combinación de colores, tipos de gráficos e información que está presentando, para construir su visualización.
6. Elabore en un documento en el que explique la razón de la selección de los gráficos incluidos. Agregue la paleta de colores usada y explique por qué la seleccionó. Adicione también el orden que utilizó en su infografía.

EVALUACIÓN

(30 puntos) Limpieza y preprocesamiento de los datos. Análisis exploratorio y Proyecciones:

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

- Modelos y resultados. Se elaboran modelos de predicción sencillos que permitan hacer alguna proyección a futuro sobre el comportamiento de las motos.

(40 puntos) Infografía

- Se entiende bien lo que debe transmitir la infografía, el orden en el que están posicionados los distintos elementos es intuitivo y comprensible.
- La paleta de colores seleccionada funciona bien, no cuesta trabajo leer los textos ni interpretar los gráficos.
- La infografía tiene suficientes gráficos para dar una idea de los elementos explorados en los datos, los gráficos son adecuados para el tipo de dato que representan.
- Los gráficos transmiten la información de forma entendible, tiene el nivel de detalles adecuado y no tiene errores de escala, o de otro tipo. ¡No mienten!

(30 puntos) Documento de explicación.

- El documento de la explicación es coherente con la infografía presentada, y se plasman en él las decisiones tomadas para construir la visualización.
- Las decisiones del diseño se basan en la teoría estudiada, la paleta de colores fue seleccionada de acuerdo con la naturaleza y significado de los resultados que se desean comunicar con la visualización.
- Están explicadas las decisiones del uso de los gráficos seleccionados y se basan en el tipo de dato que representan.
- Se explica el orden seleccionado para mostrar los resultados.

MATERIAL A ENTREGAR

- Archivo .pdf con el informe que contenga, los resultados de los análisis y las explicaciones.
- Archivo pdf con la explicación de las decisiones de diseño que tomaron para construir la visualización
- Imagen con la infografía
- 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 y filtrado de los datos. Descripción de los datos (tamaño, origen y tipo de corpus), preprocesamiento y sus explicaciones. Análisis Exploratorio de Conjunto de datos de Importación de Vehículos. Selección de paleta de colores y justificación Planificación de las tareas para el grupo: miércoles 30 de octubre 13:55.
 - **DOCUMENTO FINAL COMPLETO:** sábado 2 de noviembre de 2019 a las 23:59
- NOTA:** Solo se calificará el Documento Final si está entregado el avance con todo lo que se pide.