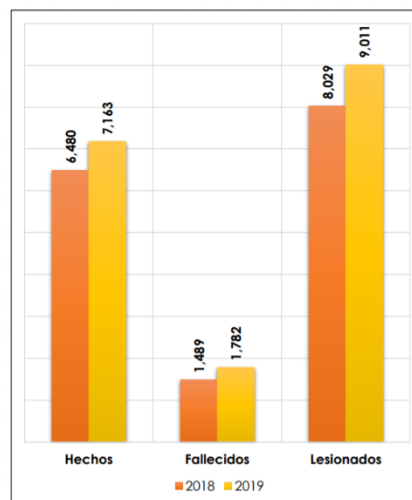


Guía de Análisis Exploratorio. Proyecto

INTRODUCCIÓN:

Para hacer una investigación formal es necesario que esta se base en una situación problemática y por consiguiente un problema que justifique la investigación. Revisar la teoría que rodea la problemática y los antecedentes de investigaciones similares. La investigación puede ser aplicada a diversos temas incluyendo finanzas, economía y negocios. Uno de los grandes problemas que aqueja a la sociedad guatemalteca es el tráfico y por ende los accidentes de tránsito. Según el boletín Número 12-2019 de las estadísticas del departamento de tránsito de la Policía Nacional Civil (http://transito.gob.gt/wp-content/uploads/2020/01/Boletin-No_-12-2019-15.pdf)

En 2019 se incrementaron los hechos de tránsito en un 11%, los fallecidos por estos hechos en un 20% y los lesionados en un 12% con respecto al año anterior.



Fuente: Boletín estadístico Número 12-2019. Policía Nacional de tránsito.

El objetivo de este proyecto es intentar descubrir, de qué forma está impactando, si es que lo hace, el parque vehicular existente, en el tema del tráfico y hechos viales, así como sus consecuencias.

Para esto puede utilizar los datos de importaciones de vehículos que provee el portal de la SAT (<https://portal.sat.gob.gt/portal/alza-e-importacion-vehiculos/>) así como los datos de hechos de tránsito, fallecidos y lesionados que tiene el Instituto Nacional de Estadística en su página (<https://www.ine.gob.gt/ine/estadisticas/bases-de-datos/accidentes-de-transito/>). Si cree que los datos del censo de 2018 puedan ayudar puede encontrarlos en el siguiente vínculo (<https://www.censopoblacion.gt/descarga>). Si tiene acceso a otros conjuntos de datos que creen que puedan servirle para atacar el problema, es libre de utilizarlos, respetando siempre las condiciones de quien los publica.

Dado que es un tema de gran envergadura su grupo puede decidir enfocarse en un límite de tiempo, en un tipo de vehículo, etc. Esto es posible siempre que la explicación del acotamiento del problema esté basada en datos.

ACTIVIDADES

1. Describa la situación problemática que lo lleva a acotar el problema.
2. Enuncie un problema científico y unos objetivos preliminares.
3. Describa los datos que tiene para responder el problema planteado. Esto incluye el estado en que encontró el o los conjuntos de datos y las operaciones de limpieza que le realizó, en caso de que hayan sido necesarias.
4. Explore los datos.
 - a. Comience describiendo cuantas variables y observaciones tiene disponibles, el tipo de cada una de las variables.
 - b. Haga un resumen de las variables numéricas e investigue si siguen una distribución normal y tablas de frecuencia para las variables categóricas, escriba lo que vaya encontrando.
 - c. Cruce las variables que considere que son las más importantes para hallar los elementos clave que lo pueden llevar a comprender lo que está causando el problema encontrado.
 - d. Haga gráficos exploratorios que le de ideas del estado de los datos.
 - e. Haga un agrupamiento (clustering) e interprete los resultados.
5. Escriba unas conclusiones con los hallazgos encontrados durante el análisis exploratorio

EVALUACIÓN

- **(10 puntos) Situación Problemática:** Describe la situación problemática que da lugar al problema.
- **(10 puntos). Problema científico:** Se enuncia el problema científico que se desprende de la situación planteada. Se comprende bien cuál es el problema.
- **(10 puntos). Objetivos:** Se plantean los objetivos a cumplir para darle solución al problema planteado. Se enuncia al menos un objetivo general y 2 específicos. Los objetivos deben ser medibles y alcanzables durante la investigación.
- **(20 puntos). Descripción de los datos:** Se describen los datos, tanto las variables y observaciones como las operaciones de limpieza que se le hicieron si fueron necesarias.
- **(30 puntos). Análisis Exploratorio:**
 - Estudia las variables cuantitativas mediante técnicas de estadística descriptiva
 - Hace gráficos exploratorios como histogramas, diagramas de cajas y bigotes, gráficos de dispersión, que ayudan a explicar los datos.
 - Analiza las correlaciones entre las variables, trata de explicar los outliers (puntos atípicos) y toma decisiones acertadas ante la presencia de valores faltantes.
 - Estudia las variables categóricas.
 - Elabora gráficos de barra, tablas de frecuencia y de proporciones
 - Explica muy bien todos los procedimientos y los hallazgos que va haciendo.

- Determina el mejor número de clusters a utilizar.
- Hace el agrupamiento con cualquiera de los algoritmos estudiados.
- Verifica la calidad del agrupamiento usando el método de la silueta.
- Interpreta los grupos, usando para eso las variables numéricas y categóricas dentro de cada grupo.
- **(20 puntos). Hallazgos y conclusiones:**
 - Hace un resumen de los hallazgos en el análisis exploratorio
 - Le pone un nombre a los grupos que reflejen sus características principales
 - Llega a conclusiones sobre los siguientes pasos a seguir.

MATERIAL A ENTREGAR

- Archivo .pdf con el informe de análisis exploratorio que debería tener:
- Script de R (.r o .rmd) o de Python que utilizó para responder las preguntas con el código utilizado o archivo de flujo de trabajo de KNime.