

Guía para de Proyecto 2. Análisis Exploratorio

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. Desarrolle el tema que seleccionó desde el punto de vista de la investigación científica. Este es el primer paso para elaborar el artículo que presentarán en forma de poster en la feria científica que se desarrollará el próximo 9 de noviembre.

ACTIVIDADES

1. Describa la situación problemática que lo lleva a seleccionar el tema del proyecto a trabajar.
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 dataset 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 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 hayar 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 e interprete los datos
 - f. Valore si el posible aplicar la técnica de análisis de componentes principales, si es así aplíquela e interprete los componentes.
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.
 - Valora la posibilidad de hacer un análisis de componentes principales y si es factible aplica la técnica
 - Determina el mejor número de clusters a utilizar usando el método de Ward
 - 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.