

Guía para de Proyecto 2.

Análisis Exploratorio

INTRODUCCIÓN

Lancasco es la primera compañía de la Industria Químico-Farmacéutica de Centroamérica, fundada en 1927. Se divide en dos unidades de negocios con la producción y comercialización de productos farmacéuticos de prescripción médica, venta libre (vitaminas, productos naturales y suplementos alimenticios) y genéricos, así como la producción y comercialización de productos del cuidado de la piel, salud y belleza que se comercializan por medio de venta directa. En este último renglón se encuentra Escentia. Esta empresa ha proporcionado sus datos para que se realicen análisis y se creen modelos de proyección de ventas. Tienen presencia en 4 países centroamericanos por lo que el trabajo se dividirá en los siguientes módulos.

- Análisis de Ventas Directas en Guatemala
- Análisis de Ventas Directas en Honduras
- Análisis de Ventas Directas en El Salvador
- Análisis de Ventas Directas en Nicaragua
- Análisis de Ventas Directas en Centroamérica
- Análisis de las Ventas por sector en Centroamérica.

INSTRUCCIONES

Grupos:

- Se formarán grupos de 6 estudiantes para analizar cada uno de los puntos mencionados anteriormente.
- En cada grupo debe estar inscrito un estudiante de Administración en Empresas que aportará, además del trabajo con los datos la visión empresarial de las soluciones.
- Temas de cada Grupo:
 - o **Grupo 1:** Análisis de Ventas Directas en Guatemala
 - o **Grupo 2:** Análisis de Ventas Directas en Honduras
 - o **Grupo 3:** Análisis de Ventas Directas en El Salvador
 - o **Grupo 4:** Análisis de Ventas Directas en Nicaragua
 - o **Grupo 5:** Análisis de Ventas Directas en Centroamérica
 - o **Grupo 6:** Análisis de las Ventas por sector en Centroamérica.

Archivos:

- Deben contar con un repositorio privado en el cual versionen los archivos y demostrar trabajo de TODOS los miembros del grupo a lo largo del tiempo que van a dedicar al proyecto.
- Deben contar con un archivo en línea que permita comprobar la participación de todos en la generación del informe.
- Tenga en cuenta que firmó un acuerdo de confidencialidad por lo que los enlaces no deben ser públicos.

Datos:

- Dado que debe firmar un acuerdo de confidencialidad los datos serán entregados en el momento en que todos los integrantes del grupo hayan firmado el documento.

ACTIVIDADES

1. Describa la situación problemática que está viviendo la empresa.
2. Enuncie un problema que quiere resolver Escentia 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 disponible, 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 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 datos.
 - f. Haga reglas de asociación y discuta los hallazgos más interesantes.
 - g. 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.
- Hace reglas de asociación y explica las reglas más interesantes.
- 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.
- Link de Google Drive donde trabajó el grupo
- Link del repositorio usado para versionar el código.
- Presentación de Power Point a usar para presentar resultados a los directivos de Scintia.

FECHAS DE ENTREGA

- **AVANCES:** Análisis Exploratorio de cada una de las variables de los conjuntos de datos. miércoles **28 de agosto 13:55.**
- **DOCUMENTO FINAL COMPLETO:** martes **3 de septiembre** de 2019 **23:59**
- **PRESENTACIÓN:** Presentación a los directivos de Scintia: miércoles, **4 de septiembre** de 2019. **11:30.**
- **NOTA:** Solo se calificará el Documento Final si está entregado el avance con todo lo que se pide.