**Objetivo:**

Utilizar el análisis exploratorio e inferencial de los datos que permita la adaptación en modelos de aprendizaje estadístico.

**Contexto:**

**Paso 1: Contexto y descarga de archivo**

Empresa de viajes compartidos que desean implementar una estrategia de precios dinámica para optimizar las tarifas en función de las condiciones del mercado en tiempo real. Actualmente, la empresa solo utiliza la duración del viaje para decidir las tarifas. La empresa pretende aprovechar técnicas basadas en estadística para analizar datos históricos.

Se requiere realizar un análisis estadístico de tipo explicativo e inferencial que permitan conocer los patrones que influyen en el costo histórico.

**(Solución en PYTHON O R)**

**Interpretar cada resultado**

**inividual**

**Análisis exploratorio e inferencial.**

**Paso 2: Desarrollo de informe de actividades**

**Análisis Unidimensional.**

1. Realizar una tabla de diccionario de datos que tenga tres columnas: nombre de la variable, descripción de la variable y clasificación en numérica o categórica u otra.
2. Realizar las tablas de distribuciones de frecuencias y gráficas de cada una de las **variables categóricas**, colocando para cada una, su correspondiente interpretación descriptiva.
3. Calcular las medidas de resumen de número de viajeros, calificación, expectativa de duración del viaje y costo histórico del viaje realizar su correspondiente interpretación descriptiva.
4. Analizar la distribución de las variables del punto 3, mediante un Boxplot o Violín plot, realizar su correspondiente interpretación descriptiva.

**Análisis Bidimensional e Inferencial.**

1. Analizar la correlación dos a dos entre número de viajeros, calificación, expectativa de duración del viaje contra el costo histórico, siguiendo las indicaciones:

* Realizar los gráficos scatterplot correspondientes (número de viajeros vs costo histórico, calificación vs costo histórico y expectativa de duración del viaje vs costo histórico), describa las posibles correlaciones de acuerdo con las visualizaciones.
* Calcular la correlación de Pearson dos a dos (número de viajeros vs costo histórico, calificación vs costo histórico y expectativa de duración del viaje vs costo histórico) y realizar prueba de correlación para determinar si la correlación entre las variables es fuerte, media o débil e interprete cada uno de los resultados.

6) Analizar las asociaciones estado de fidelidad del cliente, reservación y tipo de vehículo, siguiendo las indicaciones:

* Realiza una tabla de contingencia para cada una de las asociaciones (deben salir tres tablas), coloque observaciones relevantes frente a las posibles relaciones que encuentre entre categorías.
* Realice pruebas Chi – Cuadrado para cada una de las tablas de contingencia, puede hacerlo en paralelo e interprete los resultados, determinando que la existencia o no de asociación entre las variables.

1. Analizar la relación entre número de viajeros, calificación y expectativa de duración del viaje, siguiendo las indicaciones:

* Diseñe gráficos de boxplot o violín plot para analizar si se detectan diferencias en las distribuciones de viajeros, calificación y expectativa de duración del viaje según la fidelidad del cliente. (Debe obtener tres gráficas).
* Realice una prueba ANOVA de una sola vía, que compruebe la igualdad o diferencia de medias, interprete la significancia estadística y si encuentra diferencias significativas con algún grupo, implementa una prueba Post Hoc, que permita determinar si la fidelidad del cliente influye en la cantidad de viajeros, su calificación y expectativa del viaje. (Debe obtener tres pruebas)

**Paso 3.**

**Presentación del informe de actividades**

Realizar una presentación en Genially, PowerPoint u otra herramienta similar, en donde muestre en cada diapositiva los ítems solicitados en el **paso 2**, en diez diapositivas, así:

1. Primera: portada, título, autor, fecha.
2. Debe diseñar y presentar un objetivo general que sea coherente con la problemática propuesta en el paso 1 y los análisis estadísticos hecho en el paso 2 y en la misma diapositiva redactar tres objetivos específicos del análisis estadístico realizado en el paso 2, donde el primer objetivo se relacione con el análisis unidimensional, el segundo objetivo con el análisis bidimensional y el tercer objetivo con el análisis inferencial. Para el diseño de los objetivos (general y específico) es obligatorio seguir las indicaciones de al guia.
3. Tercera a la novena: presentar cada una de las actividades del paso 2, resumidas y organizadas. Cada actividad se debe colocar en una diapositiva, en este caso serían siete diapositivas. III. Décima: presentar mínimo seis conclusiones con los resultados de las actividades propuestas, dos para cada análisis estadístico realizado (unidimensional, bidimensional e inferencial), que sean coherentes y estén respaldadas por los resultados estadísticos más relevantes obtenidos en el desarrollo de las actividades propuestas

Requelas diapositivas: solo 10 ni mas ni menos.

Grabar sustentación de 8 minutos y subir a youtuber publixco

**Paso 4. Aportes en el foro**

vaya realizando durante los pasos 2 y 3, teniendo en cuenta los siguientes parámetros:

* Subir archivos Word o pdfs, no archivos de Python o r al foro.
* Subir con la plantilla dada los aportes.

**Paso 5:** **Compilación trabajo final – Informe de actividades**

Entregar informe con plantilla **Anexo 1 – Plantilla entrega Tarea**

* **Portada.**
* **● Objetivo General.**
* **● Objetivos Específicos.**
* **● Desarrollo del informe de actividades (Paso 2): colocar los códigos utilizados en Python o R para generar el desarrollo de cada una de las actividades propuestas.**
* **● Conclusiones. (Paso 3)**
* **● Enlace de la sustentación del informe (Paso 3)**
* **● Referencias Bibliográficas en normas APA.**