



UANL

UNIVERSIDAD AUTÓNOMA DE NUEVO LEÓN

FCFM



FACULTAD DE CIENCIAS FÍSICO MATEMÁTICAS

TIPO DE EVALUACIÓN: EXAMEN DIAGNÓSTICO.

MATERIA/UNIDAD DE APRENDIZAJE: Minería de datos.

I. Lee el caso de ejemplo y responde las preguntas correspondientes:

La Coca Cola Company tiene 366,000 accionistas y efectúa una encuesta por medio de la selección aleatoria de 30 accionistas de cada uno de los 50 estados de Estados Unidos. Se registra el número de acciones de cada accionista de la muestra.

- a) ¿Los valores obtenidos son discretos o continuos?
Los valores son discretos
- b) Identifique el nivel de medición (nominal, ordinal, de intervalo, de razón) de los datos muestrales.
De razón
- c) ¿Qué tipo de muestreo (aleatorio, sistemático, de conveniencia, estratificado, por racimos) se utiliza?
Estratificado
- d) Si se calcula el número promedio (la media) de acciones. ¿El resultado es un estadístico o un parámetro?
Estadístico, ya que se calcula sobre la muestra.
- e) Si usted fuera ejecutivo en jefe de la Coca Cola Company, ¿qué característica del conjunto de datos consideraría que es extremadamente importante?
La proporción de acciones que tiene cada accionista
- f) ¿Qué es lo que está incorrecto al evaluar la opinión del accionista enviado un cuestionario por correo que los accionistas podrían llenar y regresar por el mismo medio?
Puede prestarse a que no contesten con la realidad, ya que al no ser anónimo podrían estar forzados a contestar de cierta manera.

II. Describe con tus palabras para que se utiliza el análisis uni-variante de datos:

Para analizar por separado alguna variable de interés, esto es sin tomar en cuenta el impacto de las demás variables.

III. Describe qué es la media, la mediana y la desviación estándar:

La media es el promedio de una variable, se calcula con la suma de todos los datos entre el número de datos.

La mediana es el valor que se encuentra en la posición de exactamente la mitad.

La desviación estándar es la variación que hay entre los datos, ayuda para medir la dispersión.

IV. Con tus palabras explica qué es minería de datos y en donde puede ser aplicada:

Es una herramienta que nos permite analizar bases de datos, para poder trabajar con ellos, ya sea para hallar relación entre los datos, realizar una predicción al futuro de los datos, entre otras cosas. Puede ser aplicada en prácticamente cualquier lugar, mientras halla un historial correcto, por ejemplo, en bancos.

V. Que lenguajes de programación dominas o en qué porcentaje lo manejas:

Visual Basic – 95% R-50% Python-50%



UANL

FCFM



UNIVERSIDAD AUTÓNOMA DE NUEVO LEÓN

FACULTAD DE CIENCIAS FÍSICO MATEMÁTICAS

VI. Sabes manejar notebooks digitales como Jupyter, Spyder, Google Collab, etc. (Si la respuesta es otro, anota cuál):

Si, Google Collab y Spyder

VII. ¿Trabajas o has trabajado alguna vez en alguna empresa que maneje técnicas de ciencia de datos?

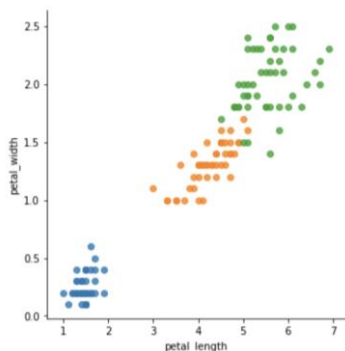
No

VIII. Has utilizado alguna herramienta de aprendizaje de máquina. Si la respuesta es afirmativa, anota cuál:

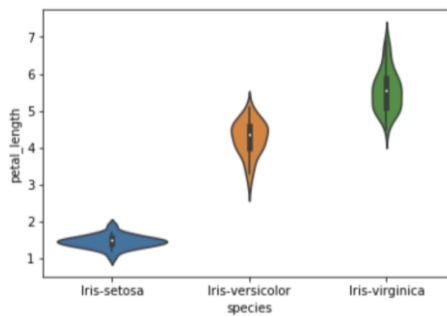
No

IX. Relaciona los conceptos con sus gráficas correspondientes:

- Se utilizan para evaluar visualmente cómo se distribuyen los puntos de datos con respecto a su frecuencia.
- Forma estandarizada de mostrar la distribución de datos basada en un resumen de cinco números ("mínimo", primer cuartil (Q1), mediana, tercer cuartil (Q3) y "máximo"). Puede informarle si sus datos son simétricos, qué tan estrechamente están agrupados y si sus datos están sesgados.
- Método para visualizar la distribución de datos numéricos de diferentes variables. Es similar al diagrama de caja, pero con un diagrama rotado en cada lado, que brinda más información sobre la estimación de densidad en el eje y.
- Visualización de datos bidimensionales que utiliza puntos para representar los valores obtenidos para dos variables diferentes: una trazada a lo largo del eje "xy" otra trazada a lo largo del eje y.
- Produce una matriz de relaciones entre cada variable en sus datos para un examen instantáneo de nuestros datos. También puede ser un gran punto de partida para determinar los tipos de análisis de regresión a utilizar.



D)



C)

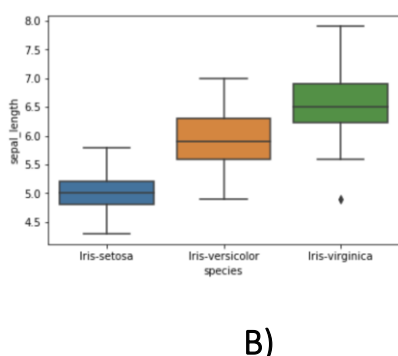
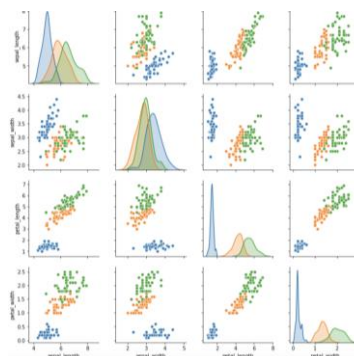
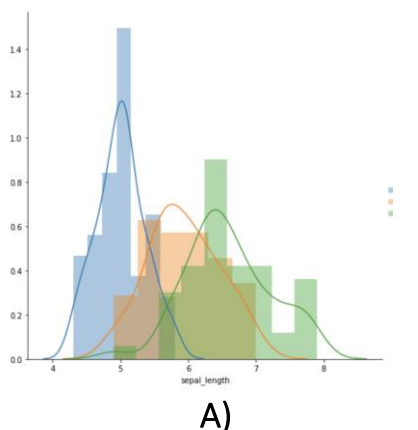


UANL

UNIVERSIDAD AUTÓNOMA DE NUEVO LEÓN

FCFM

FACULTAD DE CIENCIAS FÍSICO MATEMÁTICAS



- X. Tienes problemas de compartir pantalla, o hablar en micrófono en el Teams: (Menciona cuales son los problemas en caso de tenerlos)
No tengo problemas
- XI. Tienes problemas de abrir o utilizar Nexus: (Menciona cuales si tienes alguno)
El único detalle es que no me aparece el grupo en Nexus, pero si estuviera no tendría ningún problema.
- XII. Que sistema operativo tienes:
Windows 10
- XIII. De las siguientes herramientas menciona si tienes conocimiento de alguna y en que porcentaje.
- Github. 0%
 - Jupyter Notebook 0%
 - Google Collab 40%
 - Python 3 50%
 - Terminal de tu computadora. 0%