



UANL

UNIVERSIDAD AUTÓNOMA DE NUEVO LEÓN

FCFM

FACULTAD DE CIENCIAS FISICO MATEMÁTICAS



TIPO DE EVALUACIÓN: EXAMEN DIAGNÓSTICO.

MATERIA/UNIDAD DE APRENDIZAJE: Minería de datos.

Nombre: Ricardo González Berumen.

Matrícula: 1941497

Grupo: 001

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?
 - **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?
 - **Aleatorio.**
- d) Si se calcula el número promedio (la media) de acciones. ¿El resultado es un estadístico o un parámetro?
 - **Parámetro estadístico.**
- e) Si usted fuera ejecutivo en jefe de la Coca Cola Company, ¿qué característica del conjunto de datos consideraría que es extremadamente importante?
 - **El estado con mayor número de acciones.**
- 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?
 - **Que el correo puede llegar a no ser contestado por el accionista y pudiese prestarse a errores por la urgencia de contestar.**

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

- **Se basa en describir los datos observados de una variable y encontrar los patrones que existen en ellos.**

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

- **Media:** Valor que se obtiene al dividir la suma de un conglomerado de números entre la cantidad de ellos.
- **Mediana:** Valor que se encuentra a la mitad de los otros valores, es decir, que, al ordenar los números de menor a mayor, éste se encuentra justamente en medio entre los que están por arriba.



UANL

UNIVERSIDAD AUTÓNOMA DE NUEVO LEÓN

FCFM

FACULTAD DE CIENCIAS FÍSICO MATEMÁTICAS



- Desviación estándar: Medida de la dispersión de una distribución de frecuencias respecto de su media. Equivale a la raíz cuadrada de la varianza.

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

- Es un paso particular en el proceso de la interpretación de datos, consistiendo en la aplicación de algoritmos específicos para extraer patrones (modelos) de los datos.
- Puede ser de gran ayuda en el sector empresarial, informático y de Desarrollo.

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

- VBA – 80% a 90%
- C y C++ - 50% a 60%.
- R – 30%
- Python – 20%

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

- Sólo Google Collab.

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

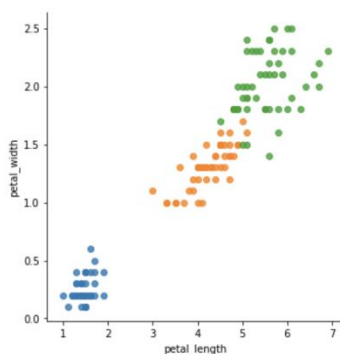
- No.

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:

- a) Se utilizan para evaluar visualmente cómo se distribuyen los puntos de datos con respecto a su frecuencia.



- b) 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.

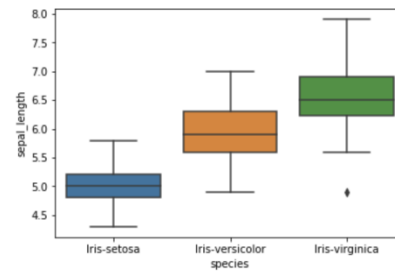


UANL

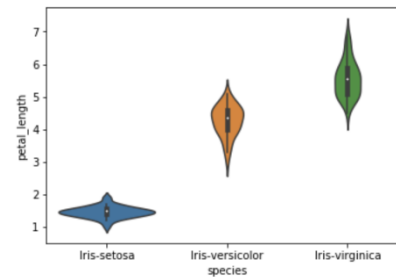
UNIVERSIDAD AUTÓNOMA DE NUEVO LEÓN

FCFM

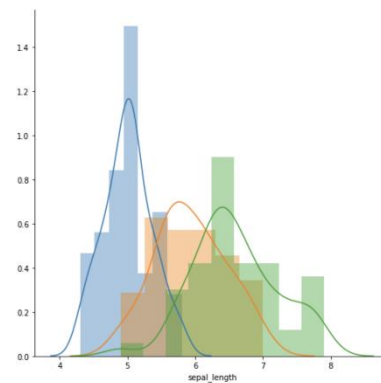
FACULTAD DE CIENCIAS FÍSICO MATEMÁTICAS



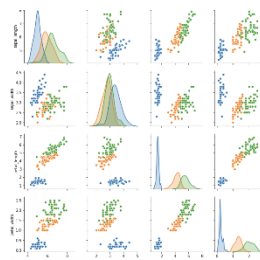
- c) Método para visualizar la distribución de datos numéricos de diferentes variables. Es similar al diagrama de caja pero con un diagrama rotado en cada lado, que brinda más información sobre la estimación de densidad en el eje y .



- d) 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 .



- e) 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.





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.
- XI. Tienes problemas de abrir o utilizar Nexus: (Menciona cuales si tienes alguno)
- Sí, no me aparece el grupo.
- XII. Qué sistema operativo tienes:
- Windows 10.
- XIII. De las siguientes herramientas menciona si tienes conocimiento de alguna y en qué porcentaje.
- Github: Sí, lo que nos mostró en clase.
 - Jupyter Notebook: No.
 - Google Collab: Sí, 60%
 - Python 3: Sí, 20%
 - Terminal de tu computadora. Sí, 20%