



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? **R = Discreta**
- b) Identifique el nivel de medición (nominal, ordinal, de intervalo, de razón) de los datos muestrales.
R = Nominal
- c) ¿Qué tipo de muestreo (aleatorio, sistemático, de conveniencia, estratificado, por racimos) se utiliza? **R = Estratificado**
- d) Si se calcula el número promedio (la media) de acciones. ¿El resultado es un estadístico o un parámetro? **R = Estadístico**
- e) Si usted fuera ejecutivo en jefe de la Coca Cola Company, ¿qué característica del conjunto de datos consideraría que es extremadamente importante? **R = La cantidad de acciones por estado por cada 30 accionistas**
- 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? **R = Que el correo se puede llegar a perder si este es una misma respuesta del correo y no un correo nuevo**

II. Describe con tus palabras para que se utiliza el análisis uni-variante de datos: Se utiliza para resumir información y encontrar patrones

III. Describe qué es la media, la mediana y la desviación estándar:
La media es el promedio de ciertos datos, la mediana es el punto central de la información acomodada de orden ascendente y la desviación estándar es que tanto se mueve la información de la media

IV. Con tus palabras explica qué es minería de datos y en donde puede ser aplicada:
Usar la información para entregar la información que se busca, es como en un curso que realice alguna vez, tienes que encontrar la manera de usar la información que se te presenta para contestar las preguntas que se realizan. Puede ser aplicado en prácticamente todo lo que necesite análisis de mucha información, análisis financiero, de riesgos, de comportamiento, censos, etc.

V. Que lenguajes de programación dominas o en que porcentaje lo manejas: VBA 50%, R 25%, C 20%



UANL

UNIVERSIDAD AUTÓNOMA DE NUEVO LEÓN

FCFM

FACULTAD DE CIENCIAS FISICO MATEMÁTICAS



VI. Sabes manejar notebooks digitales como Jupyter, Spyder, Google Collab, etc. (Si la respuesta es otro, anota cuál): No

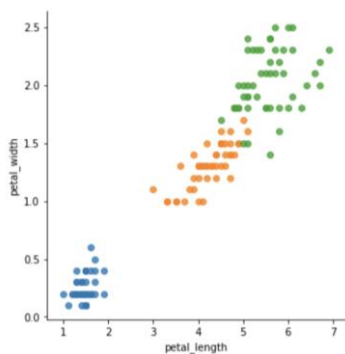
VII. ¿Trabajas o has trabajado alguna vez en alguna empresa que maneje técnicas de ciencia de datos?

Si, actualmente trabajo en riesgos paramétricos, pero la información no es lo suficientemente grande para utilizar algún lenguaje de programación, por ahora solo es Excel.

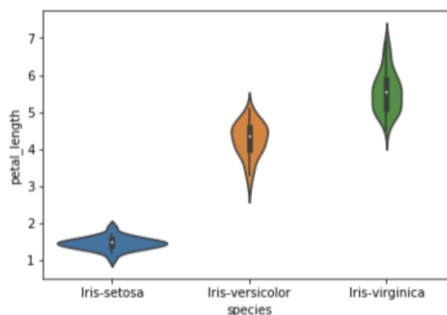
VIII. Has utilizado alguna herramienta de aprendizaje de máquina. Si la respuesta es afirmativa, anota cuál: No

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.
- 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.



a



c

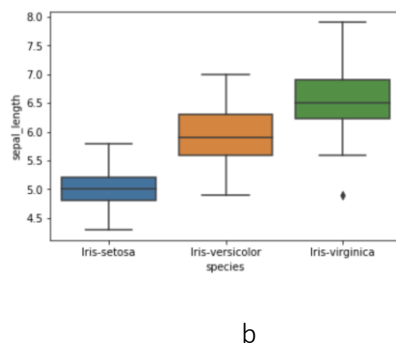
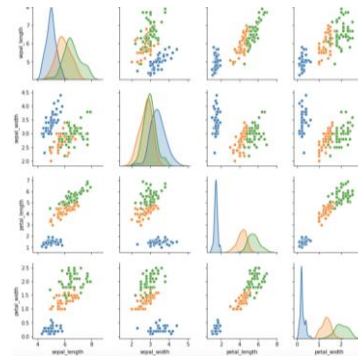
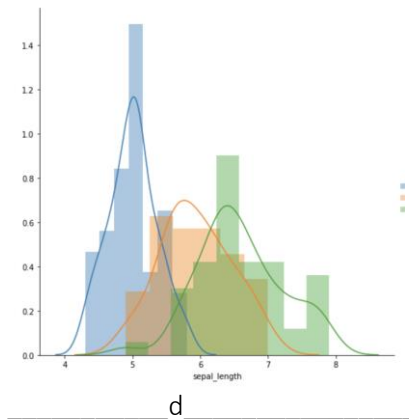


UANL

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
- XI. Tienes problemas de abrir o utilizar Nexus: (Menciona cuales si tienes alguno) No
- XII. Que sistema operativo tienes: Windows 10
- XIII. De las siguientes herramientas menciona si tienes conocimiento de alguna y en que porcentaje.
 - Github. 30%
 - Jupyter Notebook
 - Google Collab
 - Python
 - Terminal de tu computadora.