



UANL

UNIVERSIDAD AUTÓNOMA DE NUEVO LEÓN

FCFM



FACULTAD DE CIENCIAS FÍSICO MATEMÁTICAS

TIPO DE EVALUACIÓN: EXAMEN DIAGNÓSTICO.

Nombre: Diana Villarreal Garza 1877498

Grupo: 001

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?

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?

Muestreo estratificado.

- d) Si se calcula el número promedio (la media) de acciones. ¿El resultado es un estadístico o un parámetro?

Es un estadístico.

- e) Si usted fuera ejecutivo en jefe de la Coca Cola Company, ¿qué característica del conjunto de datos consideraría que es extremadamente importante?

La desviación estándar para conocer qué tan desviados están los datos entre ellos y saber que tanta variabilidad hay en el número de acciones de cada accionista de la muestra, al igual el promedio para conocer cuál es el número promedio de acciones por 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?

Lo incorrecto sería el riesgo por parte del encuestador a no recibir ninguna respuesta, por lo que serían datos vacíos dentro de su análisis.

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

El análisis univariante se usa cuando el objeto de estudio es una sola variable de los datos. Esto se hace para enfocarse en una sola característica de los datos.

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



UANL

UNIVERSIDAD AUTÓNOMA DE NUEVO LEÓN

FCFM



FACULTAD DE CIENCIAS FÍSICO MATEMÁTICAS

La media es una medida que indica el promedio de un conjunto de datos, la mediana representa el valor central cuando los datos son ordenados de manera ascendente o descendente, y la desviación estándar representa la variabilidad de los datos y que tan dispersos están entre ellos.

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

La minería de datos es una ciencia que estudia el manejo, el análisis, y la interpretación de los conjuntos de datos. Asimismo, con ella se pueden generar patrones o tendencias que contribuyen a la comprensión del comportamiento de los datos. Puede ser aplicada en el área de marketing para conocer cuales productos son vendidos de manera conjunta y así recomendarlos.

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

Lenguaje R (30%)

Lenguaje Visual Basic (60%)

Lenguaje C (30%)

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

He trabajado en ocasiones únicamente con 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.
- 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.

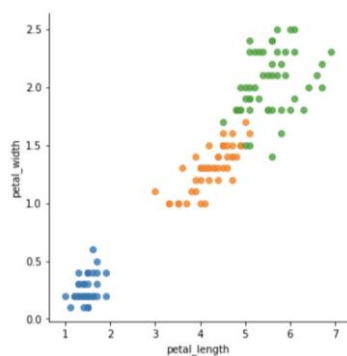


UANL

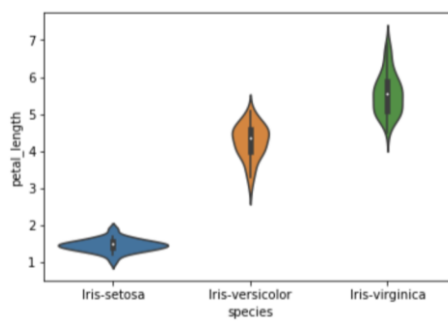
UNIVERSIDAD AUTÓNOMA DE NUEVO LEÓN

FCFM

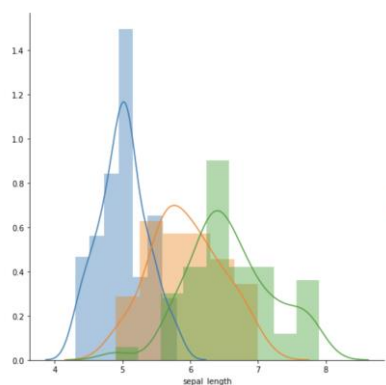
FACULTAD DE CIENCIAS FÍSICO MATEMÁTICAS



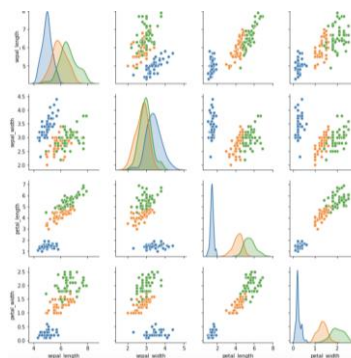
D



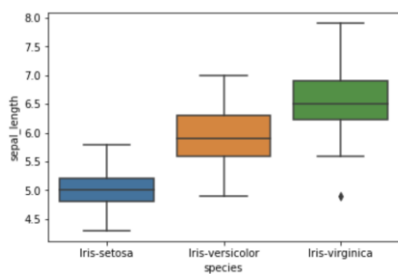
C



A



E



B

- X. Tienes problemas de compartir pantalla, o hablar en micrófono en el Teams: (Menciona cuales son los problemas en caso de tenerlos)

No



UANL

UNIVERSIDAD AUTÓNOMA DE NUEVO LEÓN

FCFM



FACULTAD DE CIENCIAS FÍSICO MATEMÁTICAS

XI. Tienes problemas de abrir o utilizar Nexus: (Menciona cuales si tienes alguno)

No tengo problema al abrir o utilizar Nexus, pero la unidad de aprendizaje “Minería de Datos” no está habilitada.

XII. Que sistema operativo tienes:

Windows 10

XIII. De las siguientes herramientas menciona si tienes conocimiento de alguna y en qué porcentaje.

- Github.
-Tengo poco conocimiento (5%).
- Jupyter Notebook
-No tengo conocimiento (0%).
- Google Collab
-Tengo poco conocimiento (10%).
- Python 3
-Tengo poco conocimiento (10%).
- Terminal de tu computadora.
-No tengo conocimiento (0%).