



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.

- ¿Los valores obtenidos son discretos o continuos? **Discretos**
- Identifique el nivel de medición (nominal, ordinal, de intervalo, de razón) de los datos muestrales.
Nivel de medición de razón debido a que el 0 representa ausencia de acciones
- ¿Qué tipo de muestreo (aleatorio, sistemático, de conveniencia, estratificado, por racimos) se utiliza? **Muestreo estratificado debido a que se seleccionan aleatoriamente accionistas de cada estado lo cual puede entenderse como una selección de elementos de cada estrato.**
- Si se calcula el número promedio (la media) de acciones. ¿El resultado es un estadístico o un parámetro? **Es un estadístico si se basa en la muestra debido a que el parámetro se refiere al valor único poblacional.**
- Si usted fuera ejecutivo en jefe de la Coca Cola Company, ¿qué característica del conjunto de datos consideraría que es extremadamente importante? **Que el número promedio de acciones de los inversionistas esté por encima de un determinado valor.**
- ¿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? **La pérdida de respuestas que se puede presentar donde al ser un cuestionario por correo puede que parte de los accionistas no responda.**

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

Es la distinción de la composición de un conjunto de datos donde solamente se estudia una variable con el fin de ver el comportamiento de estos.

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

La media es una medida de tendencia central que representa el promedio de un conjunto de datos y se calcula como la suma del conjunto de valores dividida entre el número total de valores.

La mediana es una medida de tendencia central que representa el valor central dentro de un conjunto de datos ordenados.

La desviación estándar es una medida de dispersión que representa qué tan dispersos se encuentran los valores de un conjunto de datos con respecto a la media.



UANL

UNIVERSIDAD AUTÓNOMA DE NUEVO LEÓN

FCFM



FACULTAD DE CIENCIAS FÍSICO MATEMÁTICAS

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

La minería de datos es un proceso donde se estudia el comportamiento un conjunto de datos con el fin de encontrar patrones o características. Se puede aplicar en muchas áreas como en la financiera, industrial, de medicina, entre otras para la elaboración de predicciones o pronósticos de distintos datos que puedan aportar a la toma de decisiones.

V. Que lenguajes de programación dominas o en qué porcentaje lo manejas: Lenguaje C (90%), Visual Basic (85%), R (70%) y Python (60%).

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

VII. ¿Trabajas o has trabajado alguna vez en alguna empresa que maneje técnicas de ciencia de datos? No he trabajado en alguna, pero me interesaría que pronto pueda hacerlo.

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.



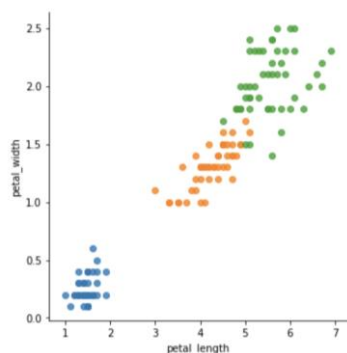
UANL

UNIVERSIDAD AUTÓNOMA DE NUEVO LEÓN

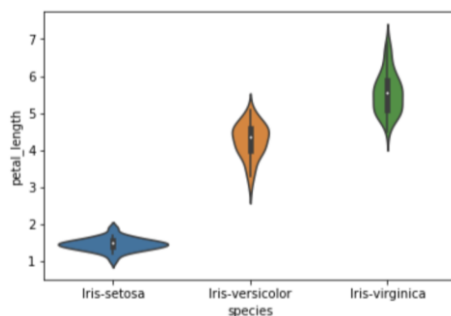
FCFM



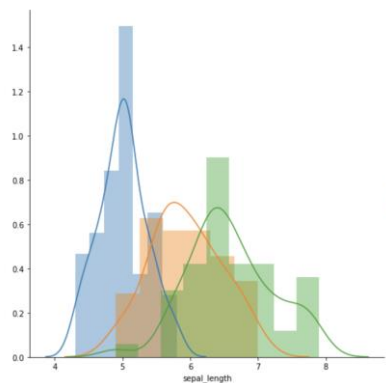
FACULTAD DE CIENCIAS FÍSICO MATEMÁTICAS



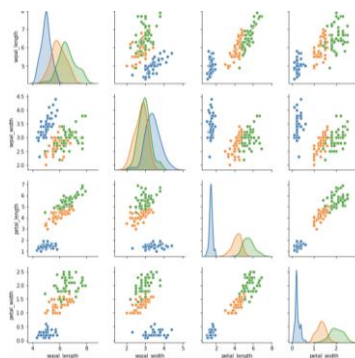
d) Gráfico de dispersión



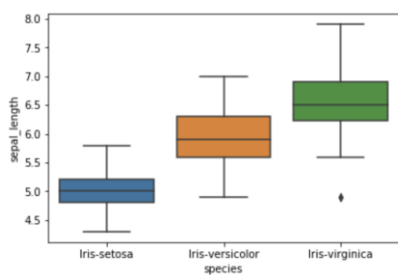
c)



a) Histograma



e)



b) Box plot o diagrama de caja

- 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) **Para abrir Nexus no, pero para utilizarlo un impedimento es que no me aparece el grupo de nuestra clase.**



UANL

UNIVERSIDAD AUTÓNOMA DE NUEVO LEÓN

FCFM



FACULTAD DE CIENCIAS FÍSICO MATEMÁTICAS

- XII. Qué sistema operativo tienes: **Windows (Versión de 64 bits, procesador basado en x64)**
- XIII. De las siguientes herramientas menciona si tienes conocimiento de alguna y en qué porcentaje.
- Github. (15%)
 - Jupyter Notebook (0%)
 - Google Collab (50%)
 - Python 3 (60%)
 - Terminal de tu computadora. (20%)

Datos del alumno:

Arturo Isaac Sánchez Tovar 1867031