



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?

Son valores discretos

b) Identifique el nivel de medición (nominal, ordinal, de intervalo, de razón) de los datos muestrales.

Medición 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?

Estados de donde hay mas 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?

Al evaluar la opinión de tantos encuestados los resultados tendrían mucha variación, es mejor un cuestionario de opciones. Por otro lado al ser por correo es mas difícil recordar que lo envíen y puede que no se reciba cantidad de respuestas esperada.

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

Se utiliza para analizar una sola variable para tomarla independiente de lo que otras variables puedan impactar en ella

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

Media: podría ser descrito como el valor esperado o como el promedio de los datos

Mediana: es el valor central del conjunto de datos al tenerlos ordenados

Desviación estándar: representa la dispersión que tienen los datos, es decir que tan dispersos están de 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:

Son las técnicas y uso de tecnologías que se utiliza para analizar, manipular y trabajar bases de datos extensas, se aplica para estudios de muestreo, predicciones, crear nuevas bases y mas necesidades que tenga una empresa con su información o la de sus clientes.

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

Visual Basic: lo domino muy bien y lo aplico en mi trabajo, 85-90%

R: Se lo básico del lenguaje, 30-35%

SQL: Se lo básico igual, 15%

Alteryx: No se si cuente como programación pero lo manejo en mi trabajo y he creado mis propios flujos y estoy trabajando en la nivelación, 50-55%

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

No, no había escuchado de ellas

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

Si, estoy como practicante de revenue management en 7-eleven y manejo bases de datos en R y Alteryx

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.



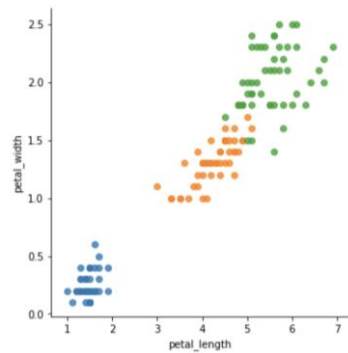
UANL

UNIVERSIDAD AUTÓNOMA DE NUEVO LEÓN

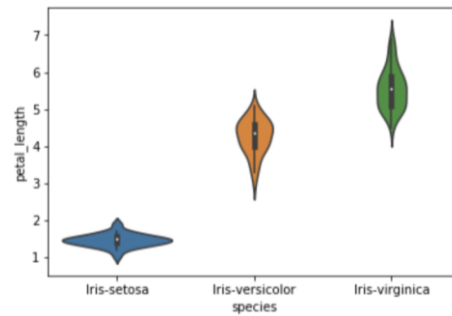
FCFM



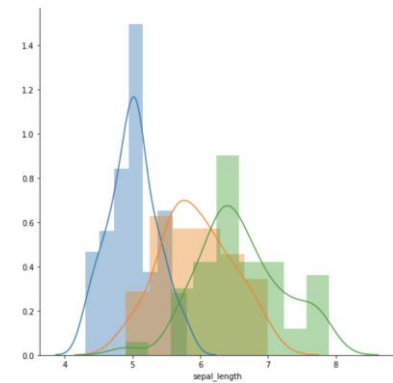
FACULTAD DE CIENCIAS FÍSICO MATEMÁTICAS



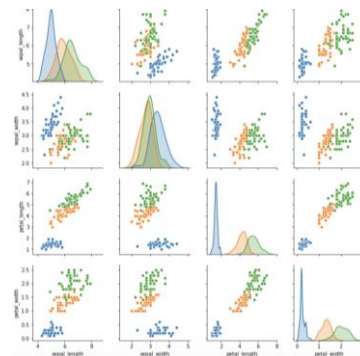
a



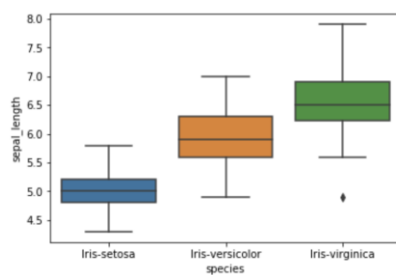
c



d



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 cuento con ningún problema



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, no hay 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: 0%
- Python 3: 0%
- Terminal de tu computadora: 0%