



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.- Discretos.
- b) Identifique el nivel de medición (nominal, ordinal, de intervalo, de razón) de los datos muestrales.
R.- Razón.
- c) ¿Qué tipo de muestreo (aleatorio, sistemático, de conveniencia, estratificado, por racimos) se utiliza?
R.- Sistemático.
- d) Si se calcula el número promedio (la media) de acciones. ¿El resultado es un estadístico o un parámetro?
R.- Parámetro.
- e) Si usted fuera ejecutivo en jefe de la Coca Cola Company, ¿qué característica del conjunto de datos consideraría que es extremadamente importante?
R.- Número de acciones por persona.
- 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 la información del cuestionario puede ser interpretada de formas distintas por el que lo redactó y el que lo contestó, además de que no se le da la importancia y respeto al no hablarlo en persona.

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

R.- Para comprender cada variable de forma individual.

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

R.- Las tres son parte del análisis descriptivo de variables cuantitativas, **la media** es el promedio de todos los datos analizados, **la mediana** es un valor que se obtiene al acomodar los valores de mayor a menor y se selecciona el del centro en caso de tener una cantidad impar de valores, en el caso de ser par se promedian los dos del centro, por último, la **desviación estándar** es la raíz de la varianza y es como fluctúan los datos.



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:

R.- La verdad no tengo conocimiento de la materia, pero por el nombre me imagino la obtención y guardado de datos que pueden ser de importancia para un analista. Un ejemplo podría ser en las elecciones, el análisis de votos por casilla y cosas por estilo, en una empresa sus compradores, salidas y entradas de dinero entre otros datos, algo así como una base de datos.

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

La verdad no soy muy buena en ninguno pero manejo un poco de R tal vez un 30%, C++ en lo mas básico únicamente lo aprendido en clase de programación.

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

No, pero podría buscar el funcionamiento de estas.

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

Actualmente solo estudio.

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.

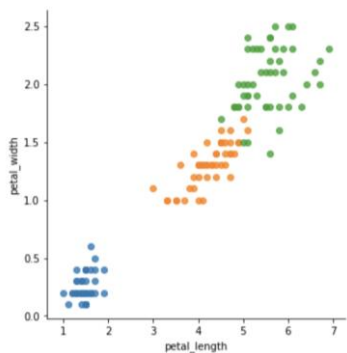


UANL

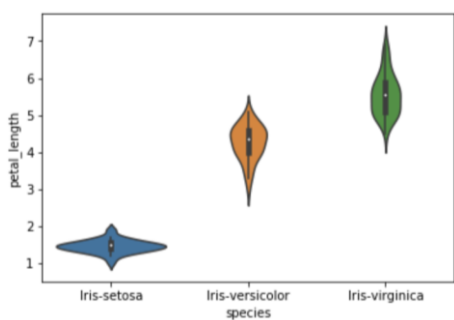
UNIVERSIDAD AUTÓNOMA DE NUEVO LEÓN

FCFM

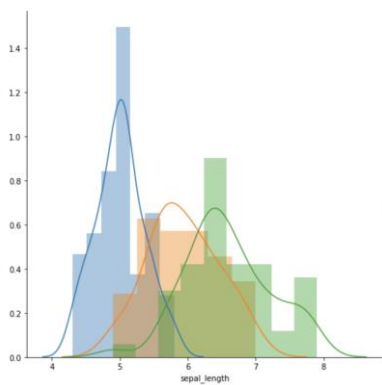
FACULTAD DE CIENCIAS FISICO MATEMÁTICAS



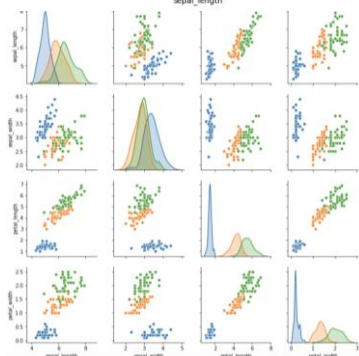
d)



c)



a)



e)

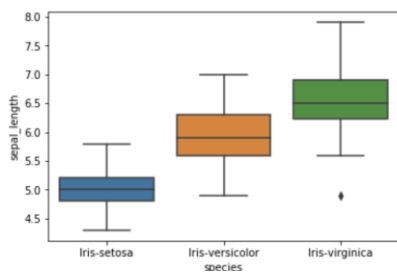


UANL

UNIVERSIDAD AUTÓNOMA DE NUEVO LEÓN

FCFM

FACULTAD DE CIENCIAS FÍSICO MATEMÁTICAS



b)

- X. Tienes problemas de compartir pantalla, o hablar en micrófono en el Teams: (Menciona cuales son los problemas en caso de tenerlos)
Mi internet es algo lento y no puedo utilizarlas todas a la vez, en especial si el clima se nubla o llueve se vuelve mas lento.
- 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. NO
 - Jupyter Notebook NO
 - Google Collab NO
 - Python 3 NO
 - Terminal de tu computadora. NO