



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?
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?
aleatorio
- d) Si se calcula el número promedio (la media) de acciones. ¿El resultado es un estadístico o un parámetro?
estadístico
- e) Si usted fuera ejecutivo en jefe de la Coca Cola Company, ¿qué característica del conjunto de datos consideraría que es extremadamente importante?
El porcentaje, de acciones, perteneciente al accionista elegido.
- 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?
Muy posiblemente contaría con sesgo, ya que velarían por su propio interés. Cabe agregar que al ser en línea no se puede interpretar la reacción del encuestado a la encuesta.

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

Para notar cambios más específicos. Al tener el análisis en una sola variable, se podría considerar al resto de las variables, en dado caso de que haya más, como constantes; lo cual nos dejaría ver los efectos que ésta causaría al cambiarla.



UANL

FCFM



UNIVERSIDAD AUTÓNOMA DE NUEVO LEÓN

FACULTAD DE CIENCIAS FÍSICO MATEMÁTICAS

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

- La media es el valor promedio de la muestra. En cierta forma se podría ver como lo que la muestra, como un individuo, obtuvo de resultado.
- La mediana es el valor medio, ordenando todos los datos de menor a mayor o viceversa.
- La desviación estándar indica, en unidades lineales, que tanta dispersión existe entre los valores y la media.

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

Es el proceso de encontrar patrones en grandes conjuntos de datos; muy comúnmente utilizada en redes sociales. En éstas se utiliza para mostrar ciertas preferencias que podríamos tener, Netflix lo utiliza para mostrarnos series o películas que nos podrían gustar, Spotify en la música o podcasts, la gran mayoría en publicidad, etc.

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

R 15-20%

C 30%

Visual basic 60%

Julia 5-10%

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

He trabajado muy poco con Pluto.

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.



UANL

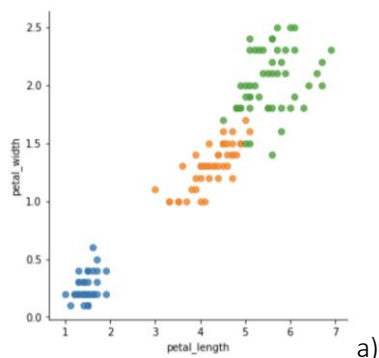
UNIVERSIDAD AUTÓNOMA DE NUEVO LEÓN

FCFM

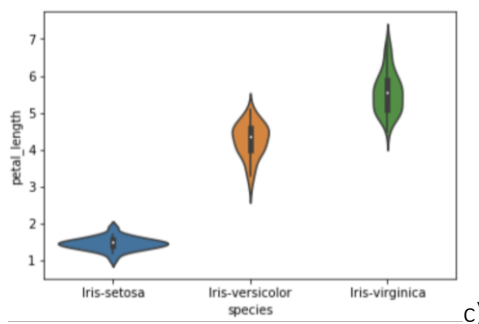
FACULTAD DE CIENCIAS FISICO MATEMÁTICAS



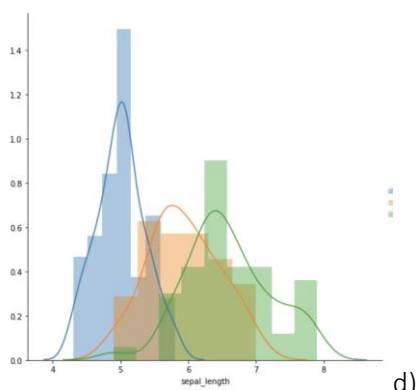
- e) Produce una matriz de relaciones entre cada variable en sus datos para un examen instantáneo de nuestros datos. También puede ser un gran punto de partida para determinar los tipos de análisis de regresión a utilizar.



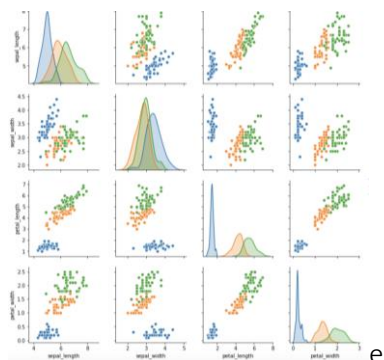
a)



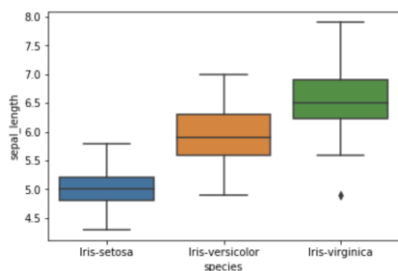
c)



d)



e)



b)



UANL

FCFM



UNIVERSIDAD AUTÓNOMA DE NUEVO LEÓN

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, solamente que no me aparece la materia.
- XII. Qué sistema operativo tienes:
Windows 10
- XIII. De las siguientes herramientas menciona si tienes conocimiento de alguna y en qué porcentaje.
- Github. 3%
 - Jupyter Notebook 3%
 - Google Collab 0%
 - Python 3%
 - Terminal de tu computadora. 3%