



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 intervalo**

- 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?

**La cantidad 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?

**Se podría sesgar el muestreo ya que no se puede confirmar que la persona que está llenando dicha encuesta sea el accionista.**

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

**Dado que se trata de un análisis de datos por medio de una sola variable, se podría utilizar en análisis que contenga múltiples variables y se desee examinar cada una de ellas por separado y de esta forma poder estudiarlas más a fondo.**

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

**La media se trata del promedio de datos, es decir la suma de todos ellos dividido entre la cantidad total de datos.**

**La mediana se trata del dato que se encuentra en el centro de un conjunto de datos ordenados.**



# UANL

UNIVERSIDAD AUTÓNOMA DE NUEVO LEÓN

# FCFM



FACULTAD DE CIENCIAS FÍSICO MATEMÁTICAS

La desviación estándar es el dato que nos dice que tan dispersos están nuestros datos de la media.

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

La minería de datos es un análisis profundo de un conjunto de datos de los cuales se puede extraer información importante para a partir de ella establecer patrones, tendencias, proyecciones, etc.

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

C++: Básico

Visual Basic: Avanzado

R: Básico

Python: Intermedio

SQL: Básico

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

Si, Jupyter

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

Si, actualmente trabajo en Johnson Controls en OTC Reporting.

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.

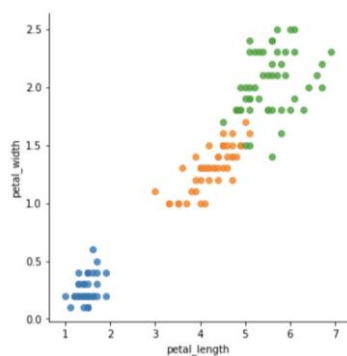


# UANL

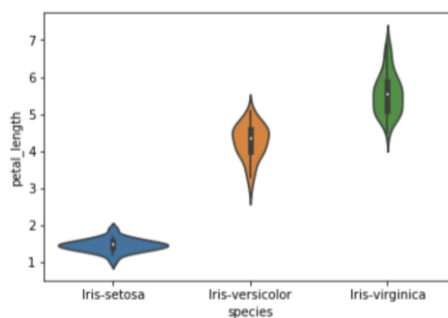
UNIVERSIDAD AUTÓNOMA DE NUEVO LEÓN

# FCFM

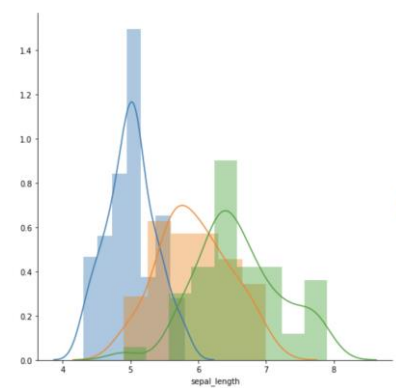
FACULTAD DE CIENCIAS FÍSICO MATEMÁTICAS



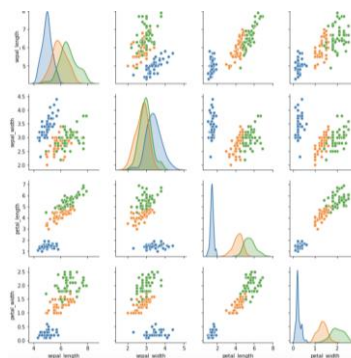
d)



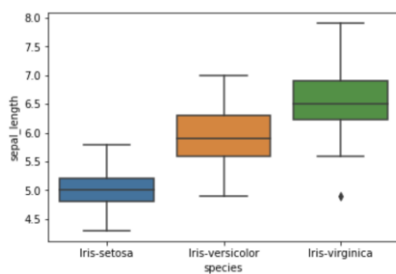
c)



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**

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 - **Intermedio**
- Google Collab - **No**
- Python 3 - **Intermedio**
- Terminal de tu computadora. - **Basico**