**TIPO DE EVALUACIÓN: EXAMEN DIAGNÓSTICO.**

MATERIA/UNIDAD DE APRENDIZAJE: Minería de datos.

1. **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.

1. ¿Los valores obtenidos son discretos o continuos?

Son discretos ya que se registrarán el número de acciones de cada accionista, por lo tanto, no se pueden tener 1.5 acciones.

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

De razón

1. ¿Qué tipo de muestreo (aleatorio, sistemático, de conveniencia, estratificado, por racimos) se utiliza?

Creo que puede ser un muestreo aleatorio estratificado debido a que se pueden dividir los elementos de la población en grupos, ya sea por estado, etc. Pero también pienso que podría ser aleatorio debido a que a N=30 para los 50 estados.

1. Si se calcula el número promedio (la media) de acciones. ¿El resultado es un estadístico o un parámetro?

Es un estadístico, ya que proviene de la muestra.

1. Si usted fuera ejecutivo en jefe de la Coca Cola Company, ¿qué característica del conjunto de datos consideraría que es extremadamente importante?

El valor que obtendré de cada accionista, y de ahí poder determinar estadísticos como la media de la muestra.

1. ¿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?

Que siempre habrá un margen de error, es decir se podría perder información quizás porque el accionista no comprendió al cien o no contesto todo, entonces ahí se perdería información.

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

Puede ser utilizado para realizar pronósticos a futuro sobre esos datos.

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

Bueno la media es el valor promedio de los datos, o de la muestra en este caso, es el valor esperado de la muestra.

La mediana es el valor que se encuentra en el centro de los datos, es decir a la mitad de mi muestra.

Y la desviación estándar indica que tan dispersos están los datos respecto a la media.

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

La minería de datos en lo personal creo que es explorar los datos, recopilados de una gran base de datos y puede ser aplicada yo creo que, para todo, en el sector medicinal, finanzas, biológico, tecnológico ya que de todo se puede obtener información de los datos.

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

Manejo C y C++, en un porcentaje quizás un 70%, y no se si se considere como lenguaje de programación, pero quizás Visual Basic en un 50%.

1. **Sabes manejar notebooks digitales como Jupyter, Spyder o algún otro (Si la respuesta es otro, anota cuál):**

No, no manejo ninguno.

1. **Has utilizado alguna herramienta de aprendizaje de máquina. Si la respuesta es afirmativa, anota cuál:**

No

1. **Relaciona los conceptos con sus gráficas correspondientes:**
2. Se utilizan para evaluar visualmente cómo se distribuyen los puntos de datos con respecto a su frecuencia.
3. 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.
4. 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.*
5. 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.
6. 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.

|  |  |
| --- | --- |
| \_\_\_\_\_\_\_\_\_\_\_\_\_d)\_\_\_\_\_\_\_\_\_\_\_\_\_ | \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_c)\_\_\_\_\_\_\_\_\_\_\_\_\_ |
| \_\_\_\_\_\_\_\_\_\_\_\_\_\_a)\_\_\_\_\_\_\_\_\_\_\_\_\_ | \_\_\_\_\_\_\_\_\_\_\_\_e)\_\_\_\_\_\_\_\_\_\_\_\_\_ |
| \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_b)\_\_\_\_\_\_\_\_\_\_\_\_\_ |  |

1. Tienes problemas de compartir pantalla, o hablar en micrófono en el Teams: (Menciona cuales son los problemas en caso de tenerlos)

No tengo problema alguno.

1. Que sistema operativo tienes:

Sistema operativo de 64 bits

1. De las siguientes herramientas menciona si tienes conocimiento de alguna y en que porcentaje.

* Github. (0%)
* Jupyter Notebook (0%)
* Python 3 (0%)
* Terminal de tu computadora. (0%)