**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 valores discretos.**

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

**Es una variable de escala ordinal.**

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

**Un muestreo aleatorio.**

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

**Es un estadístico.**

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

**Consideraría que mi número total de acciones tiene que ser un número entero y que las acciones registradas al final representan menos del 1% de sus accionistas.**

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 esos 30 accionistas no representen realmente el comportamiento de la población, por lo tanto, que no se cuente con una muestra adecuada y se caiga en un muestreo sesgado.**

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

**A grandes rasgos, creo que este análisis se utiliza para estudiar únicamente una cualidad o característica de ciertos datos.**

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

**La media es el promedio de una serie de valores que se obtiene con la suma de esos valores entre la cantidad de valores de la serie, la mediana es el valor central de una serie de datos o valores ordenados y la desviación estándar es una medida que representa la dispersión de un conjunto de datos.**

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

**La minería de datos es un proceso que estudia la información de los datos y su comportamiento, se puede aplicar en casi todas las situaciones de la vida cotidiana, pero actualmente podemos verla mayormente reflejada en las aplicaciones móviles y redes sociales que se manejan por medio de información, ya sea de productos, servicios, personas, etc.**

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

**No domino ni un lenguaje al 100%, el único lenguaje con el que he trabajado es con C aproximadamente un 40%**

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

**Nunca he manejado alguna notebook**

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 problemas de compartir pantalla ni con el micrófono, pero he tenido problemas al prender la cámara ya que me traba la computadora y me saca de Teams.**

1. Que sistema operativo tienes:

**Windows 8.**

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

* Github.
* Jupyter Notebook
* Python 3
* Terminal de tu computadora.

**De ninguna.**