**GRUPO 5**

**Integrantes:**

* **Rosales Alarcon Armando**
* **Huaman Atuncar Cristina**
* **Yancul Millones Ohami**

Pregunta 1:

De las siguientes preguntas, clasifica cada una como descriptiva, exploratoria, inferencia, predictiva o causal, y razona brevemente (una frase) el porqué:

1. Dado un registro de vehículos que circulan por una autopista, disponemos de su marca y modelo, país de matriculación, y tipo de vehículo (por número de ruedas). Con tal de ajustar precios de los peajes, ¿Cuántos vehículos tenemos por tipo? ¿Cuál es el tipo más frecuente? ¿De qué países tenemos más vehículos?

|  |
| --- |
| Las preguntas son de tipo Descriptivas.    Debido a que las preguntas hacer referencia a las caracteristicas propias de los vehiculos. |

2. Dado un registro de visualizaciones de un servicio de video-on-demand, donde disponemos de los datos del usuario, de la película seleccionada, fecha de visualización y categoría de la película, queremos saber ¿Hay alguna preferencia en cuanto a género literario según los usuarios y su rango de edad?

Las preguntas son de tipo explorativa

Debido que buscan encontrar patrones en los datos existentes. En este caso se podria utilizar para visualizar los datos para ver si hay diferencias en las preferencias de genero literario según el rango de edad.

3. Dado un registro de peticiones a un sitio web, vemos que las peticiones que provienen de una red de telefonía concreta acostumbran a ser incorrectas y provocarnos errores de servicio. ¿Podemos determinar si en el futuro, los próximos mensajes de esa red seguirán dando problemas? ¿Hemos notado el mismo efecto en otras redes de telefonía?

|  |
| --- |
| Las preguntas son de tipo inferenciales.  Se busca usar en base a la información que se tiene si este escenario puede repetirse en otros redes. |

4. Dado los registros de usuarios de un servicio de compras por internet, los usuarios pueden agruparse por preferencias de productos comprados. Queremos saber si ¿Es posible que, dado un usuario al azar y según su historial, pueda ser directamente asignado a un o diversos grupos?

|  |
| --- |
| La pregunta es de tipo Predictivas.    Debido que busca predecir la posibilidad de asignar a un usuario, basándose en su historial de compras, a uno o varios grupos de consumo. |

Pregunta 2: Considera el siguiente escenario: Sabemos que un usuario de nuestra red empresarial ha estado usando esta para fines no relacionados con el trabajo, como por ejemplo tener un servicio web no autorizado abierto a la red (otros usuarios tienen servicios web activados y autorizados). No queremos tener que rastrear los puertos de cada PC, y sabemos que la actividad puede haber cesado. Pero podemos acceder a los registros de conexiones TCP de cada máquina de cada trabajador (hacia donde abre conexión un PC concreto). Sabemos que nuestros clientes se conectan desde lugares remotos de forma legítima, como parte de nuestro negocio, y que un trabajador puede haber habilitado temporalmente servicios de prueba. Nuestro objetivo es reducir lo posible la lista de posibles culpables, con tal de explicarles que por favor no expongan nuestros sistemas sin permiso de los operadores o la dirección.

Se tienen que considerar los datos de los logs de todas las PCS para obtener las conexiones entrantes y salientes, en el campo de entrantes se verificaria la red VPN o el segmento autorizado, dado que con ello se tendria un universo mas acotado y con proposito de obtener atributos de los usuarios (data science), como por ejemplo: edad, area, tipos de consumo, entre otros para poder hace un clustering.

Adicional a esta fuente de logs considerar tambien el log de DNS y FW.

Luego se podrá agrupar el rango de posibles usuarios con acceso indebido, con lo cual se debe analizar la correlaciòn de variable como tipo de eje X y Y para ver QUÉ variable definen el comportamiento agrupado (cluster).

Finalmente, analizar un modelo de propersión para obtener qué tipo de segmento es más probable que tenga un mal comportamiento y responder en caso sea de alto nivel.

Como resultado se debe generar un informe con la información de los posibles usuarios que estarían accediendo a servicios NO AUTORIZADOS.