**PRUEBA TECNICA INGENIERO Y ADMINISTRADOR DE DATOS**

Para el desarrollo de esta prueba puede hacer uso de lenguajes R y/o Python únicamente.

Debe anexar el notebook de trabajo técnico y adicionalmente realizar una presentación de máximo 8 slides que resuma el análisis y los resultados obtenidos.

Estas pruebas podrían tomarle un buen tiempo así que le estamos dando **el fin de semana para realizarlas, tenga en cuenta que tiene un plazo máximo de entrega hasta el lunes 8:00 am** y deberá enviar los resultados a donde le indicara talento humano Porvenir y a los siguientes correos electrónicos [flgordillo@porvenir.com.co](mailto:flgordillo@porvenir.com.co), [gordillorodasfrancisco@gmail.com](mailto:gordillorodasfrancisco@gmail.com) . Por favor enviar en un solo archivo comprimido con su nombre y el cargo al que se está postulando.

El archivo “segmentación.xlsx” contiene la información de las características físicas y de comportamiento transaccional de las terminales (datafonos). **En el diccionario debe guiarse con la que se denomina AMBAS y BASE NO SUPERVISADO**:

1. Prepare los datos y mencione las características importantes de cada columna, sea lo más específico posible.
2. Realice un EDA (Exploratory Data Analysis) completo, explicando en cada línea de código el objetivo de cada paso. Utilice lo que considere necesario, gráficos, imágenes o tablas, también puede descargar librerías para utilizar lenguaje SQL dentro del Python o R si desea hacerlo.
3. A partir del análisis realizado exponga las conclusiones y recomendaciones importantes con vista de negocio. La base no es totalmente explicita, sin embargo, se tendrá en cuenta su análisis.
4. A partir del análisis realizado exponga las conclusiones y recomendaciones importantes, mencionando que inconvenientes ha encontrado en la calidad de los datos y como los debería resolver, utilice conceptos de gobierno de información y metodologías de conceptualización de dominios y atributos, así como sus implementaciones de reglas de calidad.

DICCIONARIO

|  |  |  |
| --- | --- | --- |
| **DONDE ESTÁ** | **NOMBRE** | **DESCRIPCIÓN** |
| AMBAS BASES | Mcodes | Código de identificación única de un establecimiento comercial |
| AMBAS BASES | Terminal | Código de identificación del punto o los puntos de un comercio, estas pueden ser físicas o virtuales, las físicas tendrán atributos asociados a un datáfono, las virtuales no contarán con esa información |
| AMBAS BASES | estado\_terminal | Especifica si el punto de acceso está Activo o Bloqueado |
| AMBAS BASES | producto\_red | Atributo que indica para qué tipo de servicio/ funcionalidad fue creada la terminal |
| AMBAS BASES | Tecnologiapp | Tecnología soportada por un dispositivo físico (datáfono) asociado a una terminal |
| AMBAS BASES | Modelopp | Modelo del datáfono asociado a una terminal registrado en inventarios |
| AMBAS BASES | Tecnologiaaf | Tecnología leída en el monitoreo remoto (arquitectura flexible) a las terminales físicas que están bajo ese esquema |
| AMBAS BASES | Modeloaf | Modelo leído en el monitoreo remoto (arquitectura flexible) a las terminales físicas que están bajo ese esquema |
| AMBAS BASES | Browser | Browser por el que corre la aplicación de Arquitectura flexible en un datáfono |
| AMBAS BASES | Nombretipoter | Tipo de terminal configurada lógicamente - indica para qué servicio se creó la terminal |
| AMBAS BASES | grupo\_version | Versión de arquitectura flexible cargada en el datáfono agrupada por el número y el parche, los que no estén en el modelo de arquitectura flexible se registran como trad |
| BASE NO SUPERVISADO | fecha\_version | Fecha en la que se cargó la versión de arquitectura flexible en la terminal |
| AMBAS BASES | Mcc | Segmento económico, cada letra correspondería a un segmento (ejemplo restaurantes, vestuario, aerolíneas) |
| BASE NO SUPERVISADO | segmento\_actual | Agrupación de segmentos económicos (MCC) según categorización ( línea blanca, viajes) |
| AMBAS BASES | grupo\_version\_ant | Versión de arquitectura flexible que tenía el datáfono antes de la última descarga de versión actual agrupada por el número y el parche, los que no estén en el modelo de arquitectura flexible se registran como tradicional. |
| AMBAS BASES | Ciudadfin | Ciudad |
| BASE NO SUPERVISADO | Seccionalfin | Agrupación de ciudades |
| BASE NO SUPERVISADO | fecha\_creacion\_terminal |  |
| AMBAS BASES | trx\_dia\_cero | transacciones del día de observación (ver fecha full) |
| AMBAS BASES | trx\_dia\_uno | transacciones un día antes del día de observación (ver fecha full) |
| AMBAS BASES | trx\_dia\_dos | transacciones dos días antes del día de observación (ver fecha full) |
| AMBAS BASES | trx\_dia\_tres | transacciones tres días antes del día de observación (ver fecha full) |
| AMBAS BASES | trx\_dia\_cuatro | Transacciones cuatro días antes del día de observación (ver fecha full) |
| AMBAS BASES | trx\_dia\_cinco | transacciones cinco días antes el día de observación (ver fecha full) |
| AMBAS BASES | trx\_dia\_seis | transacciones seis días antes del día de observación (ver fecha full) |
| AMBAS BASES | trx\_mes\_tres | Transacciones realizadas 2 meses antes del mes en que se realiza la observación (ver fecha full) ej. Si fecha full es 15/12/2018, se traerán las transacciones registradas en el mes 10 para la terminal |
| AMBAS BASES | trx\_mes\_dos | Transacciones realizadas 1 mes antes del mes en que se realiza la observación (ver fecha full) ej. Si fecha full es 15/12/2018, se traerán las transacciones registradas en el mes 11 para la terminal |
| AMBAS BASES | trx\_mes\_uno | Transacciones realizadas en el mes en que se realiza la observación (ver fecha full) ej. Si fecha full es 15/12/2018, se traerán las transacciones registradas en el mes 12 para la terminal |
| AMBAS BASES | prom\_mes\_tres | promedio diario de transacciones en el mes de observación menos dos |
| AMBAS BASES | prom\_mes\_dos | promedio diario de transacciones en el mes de observación menos uno |
| AMBAS BASES | prom\_mes\_uno | promedio diario de transacciones en el mes de observación |
| AMBAS BASES | total\_terminales | Cuantas terminales están asociadas al código único en la fecha de observación |
| AMBAS BASES | cantidad\_falla\_m3 | Número de fallas reportadas en el mes de observación menos dos |
| AMBAS BASES | cantidad\_falla\_m2 | Número de fallas reportadas en el mes de observación menos uno |
| AMBAS BASES | cantidad\_falla\_mes1 | Número de fallas reportadas en el mes de observación |
| BASE NO SUPERVISADO | fecha\_compra | Cuando se compró el datáfono que tiene asociada la terminal |
| BASE NO SUPERVISADO | inicio\_actividades\_pos | Cuando se instaló por primera vez el datáfono que tiene asociada la terminal (no necesariamente en la terminal en la que se encuentra actualmente) |
| BASE NO SUPERVISADO | fecha\_Full | Fecha de observación del parque de terminales |
| BASE SUPERVISADO | *antiguedad\_creacion\_terminal* | número de días transcurridos desde la creación del terminal hasta el día de observación |
| BASE SUPERVISADO | *antiguedad\_version* | número de días transcurridos desde la instalación de la versión en la terminal hasta el día de observación |
| BASE SUPERVISADO | *antiguedad\_compra\_pos* | número de días transcurridos desde la compra del datáfono hasta el día de observación |
| BASE SUPERVISADO | *antiguedad\_puesta\_pos* | número de días transcurridos desde la primera vez que inició actividades el datáfono hasta el día de observación |
| BASE SUPERVISADO | *clase\_falla* | no aplica- omitir campo |
| BASE SUPERVISADO | *Fallo* | etiqueta de clasificación de fallo para la terminal observada |
| BASE SUPERVISADO | *fecha\_fallo* | fecha en la que se presentó el fallo |
| BASE SUPERVISADO | *fecha\_observacion* | Fecha de observación del parque de terminales |