**PRUEBA TECNICA ANALISTA**

**TEORIA:**

1. ¿Qué es Apache Spark?
2. Necesito crear una tabla en una base de datos sql, dicha tabla almacenará información resultante de un proceso de pagos a una aseguradora y tiene un promedio de 20 a 30 mil registros por día. Mencione una buena práctica que podría implementar para la ingesta de dicha información.
3. Nombre 2 motores de base de datos que conozca para almacenamiento de información estructurada.
4. ¿Qué es sql?
5. Tengo un conjunto de datos estructurados, los cuales almacenan una lista de puntos físicos distribuidos por el país. Cada punto físico tiene el departamento y la ciudad en la que está ubicado. Que función de sql usaría para determinar la cantidad de puntos físicos que tengo por cada uno de los departamentos.
6. Tengo dos tablas que almacenan datos estructurados, existe una columna id la cual me permite relacionar dichas tablas. Mencione una o varias funciones de sql que me permitan relacionarlas.
7. Tengo una base en Excel con 2 millones de registros. Necesito aplicar unas reglas de perfilamiento de clientes sobre dicha base. Mencione a alto nivel de que manera llevaría a cabo dicha marcación. (lenguajes, aplicativos, infraestructura, ciclos, condicionales etc).
8. Tengo un repositorio en GIT y necesito hacer una actualización sobre uno de los archivos del repositorio. Mencione la secuencia de comandos GIT con los que se realizaría dicha actualización.
9. Necesito leer y guardar un archivo de Excel con Python, que librería utilizaría para resolver dicha necesidad.
10. Mencione mínimo 3 casos de uso en los que considera que podemos utilizar inteligencia artificial para optimizar procesos en la operación de Bancolombia.
11. Necesito buscar un archivo que tenga la palabra ‘data’ en una carpeta del ordenador con Python, mencione una librería que podría utilizar para resolver dicha necesidad.
12. Necesito obtener la fecha actual y definir si es día hábil o día festivo en Python, que librería utilizaría para resolver dicha necesidad.
13. Necesito implementar un Front en Python, que librería utilizaría para resolver dicha necesidad.

**PRACTICA:**

1. Con la base “data.xlsx” tenemos las siguientes necesidades a resolver por medio de SQL:

* Necesitamos saber la cantidad de clientes nacidos entre el 1970 y el 1990.
* Necesitamos saber la cantidad de clientes pertenecientes al grupo de clientes A y B con capacidad de endeudamiento mayor a 10’000.000.
* Necesitamos identificar el promedio de clientes que tienen correos de tipo Outlook o Hotmail y que viven en Brooklyn o New York.

1. Con la base “data.xlsx” desarrolle una solución en Python que sea capaz de:

* Leer el archivo en un dataframe.
* Cargarlo a un motor de base de datos sql (de su preferencia), en un servidor local host.
* Ejecutar los sql desarrollados en el punto 1.
* Retornar la información a Python en dataframes con los resultados.
* Enviar un correo automático a ‘elkduque@bancolombia.com.co’ y ‘halopez@bancolombia.com.co’ con un informe de los resultados obtenidos en el perfilamiento de dicha base.
* Guardar los reportes en una carpeta de su ordenador en extensión .xlsx.