Prueba técnica - Data Scientist

**Prueba Técnica para el Puesto de Científico de Datos Junior**

**Introducción:**

Como Científico de Datos Junior, desempeñarás un papel clave en el apoyo al equipo de Atención al Cliente. Tu tarea principal será crear un sistema de clasificación de textos que categorice los correos electrónicos de los clientes. Esto ayudará a agilizar el manejo de consultas y preocupaciones de los clientes. Además, después de la clasificación, analizarás los datos categorizados para proporcionar información que podría mejorar la eficiencia del departamento.

**Parte Uno: Clasificación de Textos**

**Objetivo:** Desarrollar un sistema que clasifique los correos electrónicos de los clientes en categorías significativas para ayudar al equipo de Atención al Cliente a gestionar y responder de manera más efectiva.

**Pasos:**

* **Definir Categorías:** Define las categorías apropiadas basándote en los datos proporcionados. Asegúrate de que sean equilibradas entre ser demasiado generales y demasiado específicas.
* **Preparación de Datos:** Trabajarás con tres tablas:
  + **Clientes:** Contiene información sobre los clientes.
  + **Correos Electrónicos:** Compilación de correos electrónicos de clientes.
  + **Impagos:** Un subconjunto de clientes que no deseamos analizar debido a su estado de cuenta moroso. Asegúrate de que estos correos sean filtrados en tu análisis.
* **Desarrollar Pipeline:** Libertad para usar la herramienta que se considere oportuna.

**Parte Dos: Análisis e Información**

**Objetivo:** Analizar los datos categorizados para identificar patrones, obtener información y proponer mejoras para el departamento de Atención al Cliente, utilizando como base los emails proporcionados en la base de datos.

**Pasos:**

* **Exploración de Datos:** Profundiza en los correos categorizados para entender la distribución, como:
  + Volumen de correos por categoría.
  + Distribución temporal de correos (por ejemplo, ¿hay problemas que surgen en momentos específicos del año/mes/día?).
* **Proporcionar Recomendaciones:** Basado en tu análisis:
  + Sugiere posibles estrategias para una mejor asignación de recursos dentro del departamento.
  + Identifica posibles áreas de formación o capacitación para el equipo de Atención al Cliente.
  + Propón medidas preventivas que podrían reducir el volumen de ciertas categorías de problemas.

**Entorno de Desarrollo**:

Te proporcionaremos una carpeta que contiene un archivo **docker-compose.yml** que se encarga de levantar dos servicios: una base de datos MySQL y una aplicación de Python que implementarás.

**Instrucciones**:

* **Configuración del Entorno**:
  + Una vez hayas descargado la carpeta proporcionada, ejecuta el comando **docker-compose up** para levantar ambos servicios (la base de datos y la aplicación).
  + Eres libre de modificar el Dockerfile del servicio **app** para añadir las herramientas y utilidades que necesites.
* **Desarrollo del Sistema de Clasificación**:
  + En la carpeta **./app**, encontrarás un archivo llamado **main.py**. Tu tarea será:
  + Crear una API que, al recibir una fila de la tabla **emails** como entrada, devuelva la predicción de la categoría correspondiente. (en el README tienes más detalles)
  + Integrar el código adicional y cualquier otro recurso necesario para que la pipeline de clasificación funcione correctamente.
* **Análisis e Información**:
  + Además de la implementación anterior, debes desarrollar un notebook en el que realices el análisis descrito en la "Parte Dos" del ejercicio. En este notebook, explora los correos categorizados, identifica patrones y presenta tus recomendaciones.

**Entregables**:

* La carpeta proporcionada con las modificaciones realizadas en **main.py** y cualquier otro archivo necesario.
* Un notebook con el análisis y conclusiones extraídas de los datos.

Ten en cuenta que, dada la reducida cantidad de datos que te proporcionamos, no esperamos un análisis muy exhaustivo de los datos.