**CIENTIFICO DE DATOS**

**Keraunos**

**Proceso de Selección**

**Nombre Completo:**

**Numero Identificación:**

**Observaciones:**

* Use este documento para incluir la mayor cantidad de la información textual y gráfica que responda a las preguntas.
* Es posible que requiera utilizar herramientas externas. Incluya los enlaces de los artefactos externos, o los archivos adjuntos correspondientes dentro del correo. Se recomienda que pueda respaldar los distintos hallazgos con su debido soporte.
* Los siguientes casos, corresponden a la mezcla de información cierta y ficticia, construidos para fines de medir ciertos conocimientos y habilidades requeridos para el cargo en mención.
* No hay una extensión mínima o máxima.
* Guarde este documento en formato PDF y envíelo.
* Envíe todos los documentos y soportes en un único correo, con el asunto: “Nombre del aspirante\_CientificoDatos”, al siguiente correos: mrestrepo@keraunos.co . En el correo, explique brevemente el contenido de los archivos adicionales, si llega a tenerlos.

**Prueba científico de Datos.**

1. De todo el scope de ciencia de datos, cual considera que es su fuerte? En que temas considera que tiene debilidades técnicas y/o de negocio?
2. Se le compartió un Excel, el cual esta compuesto por dos hojas. En la hoja DataFramePrueba se encuentra el set de datos y en la hoja diccionario se encuentra la información asociada a cada variable.
   1. Datos: Se comparte un set de datos donde cada registro corresponde a un cliente de un banco que tiene tarjeta de crédito. En este caso, el banco tiene una preocupación por las ventas con tarjetas de crédito en un ecommerce en el extranjero llamado “Temu”, en el cual observa que hay un deterioro de la cartera mayor al promedio. El banco ha generado una muestra de clientes en un periodo de tiempo que han usado su tarjeta de crédito en compras en “Temu”, y ha evaluado si dicho cliente se encuentra o no en mora.

El banco esta dispuesto a negar el uso de la tarjeta de crédito en Temu, a aquellos clientes que muestren que no van a cumplir con su obligación de pago. Dentro del dataset se encuentran clientes que ya han tenido uso de la tarjeta de crédito en otros establecimientos, y también clientes cuyo primer uso fue en Temu (esto se puede identificar en la variable NumeroCreditosGPrevius, donde aquellos clientes que tengan cero, es porque su primer uso de la tarjeta de crédito fue en Temu).

* 1. Reto: El reto del candidato consiste en usar los datos para entender las dinámicas de incumplimiento de pago de las tarjetas de crédito al ser usadas en Temu, y lograr predecir si el cliente va a pagar o no. Se espera que demuestre habilidades en el manejo, limpieza y análisis de datos, ingeniería de características, así como en la aplicación de algoritmos de Machine Learning y demás técnicas de ciencia de datos que el candidato considere.

Es clave que el candidato exhiba competencias para entender el problema y las variables que lo componen. Más allá de buscar la mejor solución de manera inmediata, se valorará el razonamiento crítico del candidato para escoger el(los) modelo(s), analizar el problema, proponer mejoras continuas del modelo, identificando falencias en los aspectos técnicos, prácticos o teóricos que puedan afectar los resultados.

* 1. Entregable: El candidato debe entregar el repositorio que explique el problema, insights obtenidos de los datos, las metodologías aplicadas, el porqué de la selección de modelos o algoritmos aplicados, la relevancia de las variables escogidas y el análisis de los resultados obtenidos. Además, deberá identificar y proponer estrategias para mejorar el modelo, discutiendo las limitaciones encontradas y cómo éstas pudieron influir en los resultados. La presentación deberá ser clara y comprensible, capaz de traducir términos técnicos a un contexto de negocio, permitiendo entender cómo los insights obtenidos pueden ser aplicados para optimizar las decisiones comerciales. Se espera que el candidato muestre la habilidad de reflexionar sobre su trabajo, manifestando una comprensión profunda de la ciencia de datos no solo como una disciplina técnica, sino también como una herramienta para la toma de decisiones estratégicas.
  2. El repositorio entregado debe tener mínimo 10 commits realizados y al menos 2 PR que permita evaluar las habilidades del candidato en el versionamiento y el manejo de las incorporaciones de nuevas funcionalidades. Los PR en github pueden ser aprobados por el mismo candidato y hacer la incorporación con merge o squash como más le parezca.