



Instructions to candidates

Welcome to the selection process for ProColombia. This test is designed to simulate real working conditions at our organization. It gives you a chance to evaluate if this role aligns with your career aspirations for the coming years.

Within this document, you'll find a case study that you need to solve to advance to the next stage of our selection process. Your analytical abilities, programming, writing and summarization skills, information utilization, and time management will be assessed. We encourage strategic responses as only the top 5 candidates will progress to the subsequent stage.

For this assessment, you are free to consult any online sources. Although the estimated time for completing all questions is 8 hours, the final deadline is Tuesday, 19th December 2023, before 7:00 A.M.

When responding to each question, it's mandatory to include the code used in your solutions. Please upload this code to your GitHub account and share the public link with us. Submissions should be made via email and must include all the materials utilized in your research (e.g., datasets, pivot tables, PDFs, web pages, etc.). Clarity and organization of your submissions are crucial.

Limits on words, slides, or pages: To ensure fairness, there's a strict limit on the amount of content you can submit. Exceeding the specified limits for words, slides, or pages will result in a 2% penalty per case on your final score.

Late submissions: Please note, submissions after 7:00 A.M on 19th December 2023 will not be accepted.

Plagiarism policy: Plagiarism, or uncredited borrowing of content, is considered a serious violation and will lead to disqualification. By submitting your answers, you affirm that your work is original, free from plagiarism, and has not received external assistance.

Evaluation Criteria:

How effectively have you answered the question?

The quality of your arguments and evidence.

Your understanding of the relevant aspects of the question.

Depth and critical analysis demonstrated.

Clarity of your answer.

Knowledge of key programming language concepts.

An outstanding submission will be characterized by exceptional originality, insightful analysis, comprehensive synthesis, and excellent argumentation and writing skills. Conversely, a lower score may result from weakly structured arguments, insufficient evidence, factual inaccuracies, theoretical inconsistencies, or inadequate source documentation.

Thank you for considering a career with ProColombia.

Prueba técnica profesional

Coordinación de Analítica

Introducción

La Coordinación de Analítica se dedica a la utilización de datos para generar análisis detallados, con el objetivo de fomentar el comercio internacional y el crecimiento económico. Nuestra labor consiste en proporcionar y administrar conocimiento especializado en economía, junto con herramientas analíticas avanzadas. Estas herramientas son fundamentales para apoyar los procesos de internacionalización, como las exportaciones, la inversión y el turismo.

En el equipo de Analítica, nos enfocamos en crear valor de negocio y en desarrollar ventajas competitivas tanto para la entidad como para los empresarios y el país. Esto lo logramos mediante la gestión eficaz de los datos como un activo estratégico clave. Ofrecemos información precisa y análisis relevantes que son esenciales para respaldar la toma de decisiones informadas y para identificar oportunidades que promuevan el comercio, la inversión y el turismo a nivel nacional.

Dada la importancia de nuestra misión, es fundamental que las respuestas proporcionadas estén alineadas con la visión y el enfoque de nuestro equipo. La prueba presentada es un caso práctico que brinda la oportunidad de experimentar de forma directa el trabajo en el equipo de Analítica.

Para resolver el caso asignado, sugerimos seguir las seis fases clave del proceso *Cross Industry Standard Process for Data Mining* (CRISP-DM) para análisis predictivo de datos: 1) Entendimiento del negocio: identificar claramente el problema de negocio que se quiere resolver y diseñar una solución de análisis de datos para esto. 2) Entendimiento de los datos: comprender las distintas fuentes de datos disponibles y los tipos de datos contenidos en ellas. 3) Preparación de los datos: convertir las diversas fuentes de datos en una tabla base analítica bien estructurada para generar modelos que solucionen el problema de negocio o la necesidad. 4) Modelado: construir distintos modelos predictivos con algoritmos de aprendizaje automático y seleccionar el mejor. 5) Evaluación: probar rigurosamente los modelos para asegurar que sean precisos, no tengan sobre-ajuste o bajo-ajuste. 6) Implementación: integrar con éxito el modelo en los procesos organizacionales a través de un producto que sea para el consumo del usuario final. Su habilidad para navegar estas etapas determinará la eficacia de la solución propuesta.

Este caso le preparará para la prueba presencial que hará en caso de que su puntuación final se encuentre en el top 5 de los postulantes al cargo.



I. CASO PARA RESOLVER

La vicepresidenta de exportaciones le pidió al equipo de Analítica que haga un análisis del tejido empresarial colombiano y de los pronósticos de las exportaciones para el año 2024.

Respecto al tejido empresarial debe identificar las empresas que tienen un alto potencial de atención en ProColombia, es decir aquellas que pueden tener potencial de convertirse en exportadoras, esto le permitirá al equipo de la Vicepresidencia de Exportaciones (VPE) contar con un listado corto de empresas que tengan alta probabilidad de cerrar negocios en el exterior. Ese listado será compartido con la VPE como insumo para que las diferentes gerencias y asesores las contacten e inicien procesos de asesoramiento.

Para llevar a cabo este análisis, dispondrán de diversas bases de datos: Base de Exportaciones, Directorio de Empresas del DANE, Listado de Super Sociedades, Listado del Registro Único Empresarial (RUES) y un diccionario (Diccionario.xlsx) detallado de cada tabla donde se detalla la interpretación de cada variable. La conexión entre estas bases de datos se realiza a través del NIT de las empresas. Los archivos se guardaron en “.txt”, con separador |, comas como separador de decimal y encoding = 'utf-8'.

Utilizando lo anterior, elabore un modelo que permita identificar las empresas con alto potencial de atención. A la vicepresidencia le debe entregar un listado con las empresas que sugiere contactar (Archivo en Excel). Sumado a lo anterior, la Vicepresidencia requiere *insights* sobre el análisis, es decir, información reveladora y conocimiento que le permita tomar decisiones. Es libre de hacer el análisis de los resultados según considere más adecuado teniendo en cuenta el enfoque del equipo de Analítica y la necesidad de la VPE.

La segunda tarea asignada por la VPE consiste en analizar las exportaciones no minero-energéticas (NME), proyectando su comportamiento para el cierre de 2023 y 2024: si aumentarán, se mantendrán estables o disminuirán. Este análisis requiere el uso de la Base de Exportaciones (expo_nme.xlsx) y la aplicación de un modelo predictivo que le permita identificar de la manera más objetiva posible el comportamiento que tendrán las exportaciones NME. Usen datos hasta diciembre de 2022 para entrenamiento y de enero a septiembre de 2023 para validación, midiendo el error cuadrático medio para evaluar la calidad del pronóstico, y luego realicen proyecciones hasta diciembre de 2024.

Desde Analítica, los productos siempre se presentan formalmente a los clientes, por lo tanto, para la vicepresidencia, debe hacer una presentación de máximo 7 diapositivas explicando lo que se hizo para llegar al listado de empresas y el pronóstico de las exportaciones. Se sugiere que la presentación incluya qué hizo en sus modelos propuestos, los resultados y los *insights* más importantes al analizar las bases de datos. Los candidatos seleccionados para la siguiente fase (top 5) tendrán 10 minutos para presentar ante los clientes.

Es esencial que todos los recursos utilizados, incluyendo los scripts de programación (en Python, R o Stata), estén bien documentados y comentados según estándares internacionales. Cada script y análisis debe ser auto explicativo y fácil de seguir. Esto implica explicar claramente el propósito de cada función, sus parámetros de entrada y salida. La totalidad de los documentos utilizados deben alojarse en un repositorio en GitHub con acceso público. Para cada tarea asignada debe compartir el enlace correspondiente y debe ser lo más claro posible para facilitar su entendimiento y claridad. Recuerde ser empático con el cliente para que el producto esté en versión final. Asegurarse de que la estructura del repositorio en GitHub sea intuitiva, con un ReadMe claro que guíe al usuario a través del contenido. Muchos éxitos.