



## Prueba Técnica para: Especialista Ciencia de Datos

A continuación, encontrará los puntos que debe de desarrollar y las especificaciones requeridas.

1. Realizar **Web scraping** sobre el programa académico derecho en 3 universidades de la ciudad de Medellín, para lo cual deberá presentar **a).** código en el lenguaje de programación Phyton o R y **b).** los resultados de la búsqueda, teniendo en cuenta las siguientes variables: nombre del programa, duración, costo, modalidad y acreditación en alta calidad.
2. Con los resultados obtenidos en el punto anterior, realizar **a).** un **tablero de visualización**, ya sea en Power BI o Tableau, y **b).** en un texto Word, donde analice e interprete los datos, en máximo dos párrafos.
3. Presente la ruta a seguir para implementar un modelo predictivo de deserción del proyecto de Sapiencia Talento Especializado, incluyendo las variables para alimentar el modelo.
4. En la actualidad se cuenta con servicios en la nube y de modelación, como Vertex AI, Big Query, Machine Learning, API Tensor Flow. ¿En su experiencia laboral ha utilizado alguno de estos?, en caso de ser afirmativo, describa en tres viñetas cuál servicio ha utilizado, en qué proyecto (s) lo ha implementado y la ventaja de utilizar ese servicio en la nube.
5. El archivo “Bd\_modelo\_predictivo\_desercion.xlsx” contiene la información de las características socioeconómicas de los beneficiarios de los programas de Sapiencia, además, se encuentra la variable denominada “desertor” que indica si el beneficiario desertó de alguno de los programas de la entidad. Teniendo en cuenta lo anterior, resuelva: **a)** Código en R o Python en el cual lleve **a cabo dos técnicas de machine learning** que permitan predecir si un beneficiario de los programas de Sapiencia desertará en el semestre en curso. **b)** Responda las siguientes preguntas: ¿Cuáles son **las variables más relevantes** para que un beneficiario deserte y cómo las identifica? ¿Qué variables adicionales propondría usted incorporar en este análisis y así **mejorar la capacidad de predicción** de sus modelos?

Tener en cuenta: **nombrar los archivos con su primer nombre y primer apellido**, así

- Punto 1a: WS\_primernombre\_primerapellido
- Punto 1b: RB\_primernombre\_primerapellido
- Punto 2a: TV\_primernombre\_primerapellido
- Punto 2b: AD\_primernombre\_primerapellido
- Punto 3: R\_primernombre\_primerapellido
- Punto 4: ML\_primernombre\_primerapellido
- Punto 5a: MP\_primernombre\_primerapellido
- Punto 5b: Mr\_primernombre\_primerapellido