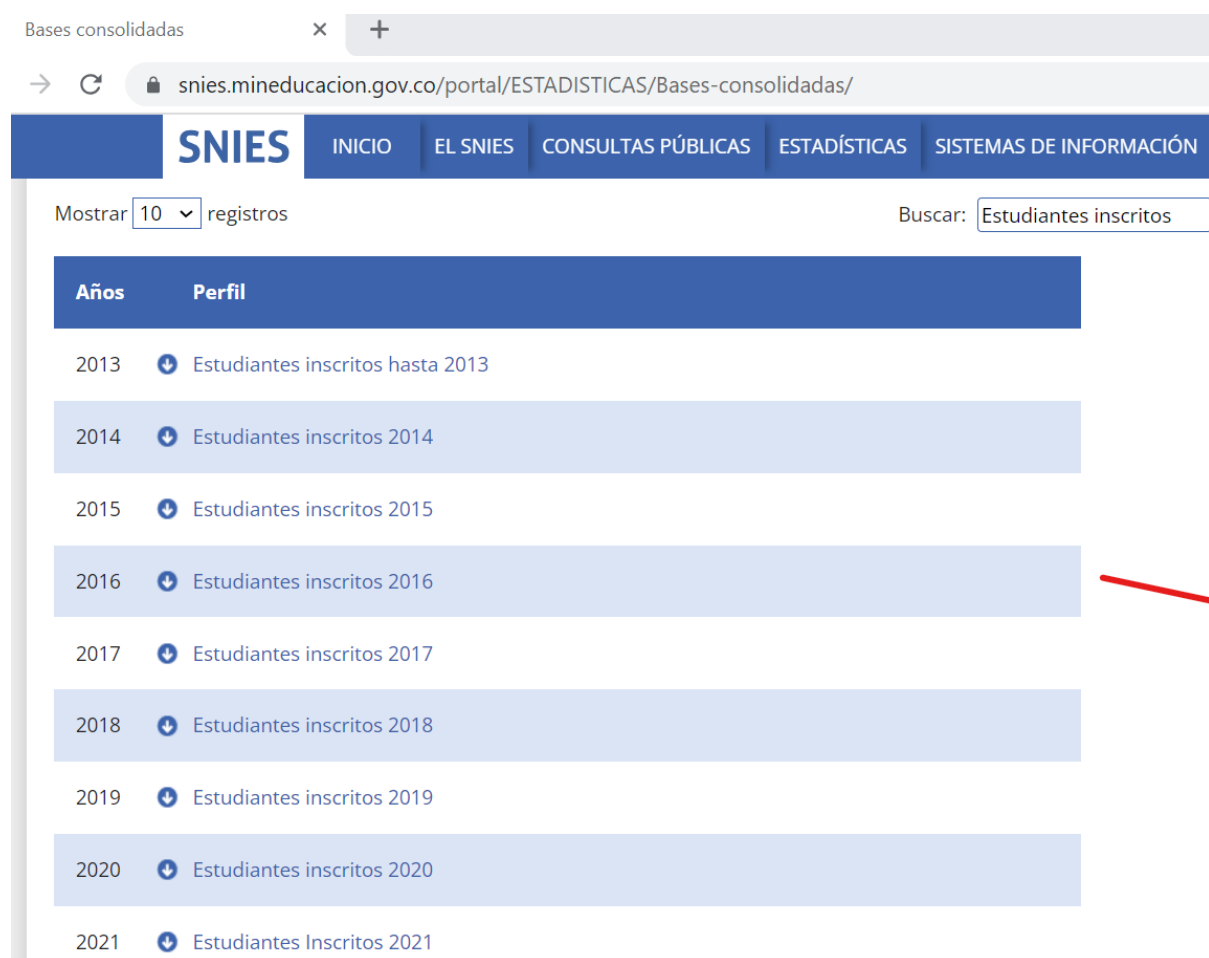


# Prueba Análisis de datos

Queremos agradecer por haberse postulado para hacer parte del equipo de desarrollo del proyecto que esta permitiendo visualizar la gestión de la Universidad Nacional de Colombia Sede Bogotá.

Estimado/a candidato/a, por favor resuelva los siguientes puntos y entréguelos a más tardar el próximo jueves 8 de junio a las 11:59 a.m. (Si lo entrega antes lo tendremos en cuenta).

1. Limpieza de datos y ETL (40%): usando los datos del SNIES <https://snies.mineducacion.gov.co/portal/ESTADISTICAS/Bases-consolidadas/> , mas específicamente de Estudiantes inscritos



Bases consolidadas x +

→ ↻ [snies.mineducacion.gov.co/portal/ESTADISTICAS/Bases-consolidadas/](https://snies.mineducacion.gov.co/portal/ESTADISTICAS/Bases-consolidadas/)

**SNIES** INICIO EL SNIES CONSULTAS PÚBLICAS ESTADÍSTICAS SISTEMAS DE INFORMACIÓN

Mostrar 10 registros Buscar: Estudiantes inscritos

Años	Perfil
2013	Estudiantes inscritos hasta 2013
2014	Estudiantes inscritos 2014
2015	Estudiantes inscritos 2015
2016	Estudiantes inscritos 2016
2017	Estudiantes inscritos 2017
2018	Estudiantes inscritos 2018
2019	Estudiantes inscritos 2019
2020	Estudiantes inscritos 2020
2021	Estudiantes Inscritos 2021

Realice los siguientes procedimientos:

- Eliminar valores nulos o faltantes.
- Estandarizar el formato de los datos (dejan en letras capitales, mayúsculas corregir ortografía etc.).

- Eliminar registros duplicados.
- Reducir redundancia en los datos (ejemplos Educación técnica, Educación para técnicos, agrupar en Educación técnica)
- Realizar transformaciones o cálculos necesarios para preparar los datos para su análisis deje todo integrado en un único archivo.

Para tener este punto como valido deberá compartir el archivo resultante y el código empleado, puede usar las librerías que considere necesarias.

2. Generación de visualizaciones en Python (10%): con el dataset de administrativos del SNIES cree los siguientes gráficos con cualquier variable y sus combinaciones que considere atractivas, deberá usar la librería plotly, debe usar al menos 1 gráficos por 4 tipos diferentes de gráficos, para un mínimo de 4 gráficos, y exportar estos en formato png o html con su respectivo código
3. Visualización de datos en Tableau (40%): Crear un tablero usando el dataset entregado con anterioridad por la Universidad Nacional.

Genere un archivo en Tableau con al menos 4 gráficos de diferentes variables (ej. Inscritos, universidades, programas, sexos, nivel de formación) en la que se muestre que está pasando con los inscritos a la educación superior a través de una historia. Se le solicita incluir filtros clasificados por año, departamento y entidad educativa. A continuación, agrupar en tres grandes grupos:

1. Universidades NO SUE corresponde a las universidades privadas
2. Universidades del SUE corresponde a universidades estatales o públicas,
3. Universidad Nacional como tercer grupo.

En el aspecto visual se espera que el tablero sea interactivo, atractivo y estético.

4. Documento de análisis (10%): En un documento de 3 páginas hacer un breve análisis de sus conclusiones, hallazgos en los datos procesados y datos interesantes a destacar.

Deberá nombrar enviar un correo a [prig\\_bog@unal.edu.co](mailto:prig_bog@unal.edu.co) indicando:

- Nombre del postulante
- Teléfono del postulante
- Correo del postulante
- El link de la carpeta compartida en GoogleDrive con el nombre "Pba\_Datos\_Nombre\_Apellidos", reemplazando su nombre y apellidos, que contenga

- Enlace al tablero en tableau public
- Enlace descargable al código de Python en Github
- Documento de análisis

De antemano agradecemos por su participación y por el tiempo dedicado en la elaboración de esta prueba.