



PROYECTO

Cristian Alarcón

Sarah González

Daniel Muñoz

INTRODUCCIÓN

El Servicio Nacional de Aprendizaje (SENA) cuenta con una amplia red de Centros de Aprendizaje distribuidos a lo largo del territorio colombiano. La correcta georreferenciación de estos centros es fundamental para la planeación estratégica, la toma de decisiones institucionales y el análisis territorial de la oferta educativa. El presente proyecto busca analizar, depurar y preparar un dataset de georreferenciación de los Centros de Aprendizaje del SENA, con el fin de garantizar su calidad y permitir posteriores análisis avanzados como segmentación geográfica, clustering con K-meas.

OBJETIVOS GENERAL

Analizar preparar y ocupar un modelo para un conjunto de datos de georreferenciación de los Centros de Aprendizaje del SENA mediante la metodología ASUM-DM, garantizando la calidad de los datos y sentando las bases para modelos analíticos posteriores.

OBJETIVOS ESPECÍFICOS

- Comprender la estructura y características del *dataset*.
- Identificar patrones, anomalías y posibles inconsistencias en los datos geográficos.
- Realizar un proceso de limpieza y transformación de los datos.
- Preparar un *dataset* confiable para etapas de modelado y optimización.
- Utilizar un Modelo de IA o analítica de datos.

METODOLOGÍA

El proyecto se desarrolla siguiendo la metodología ASUM-DM,

abordando las fases de:

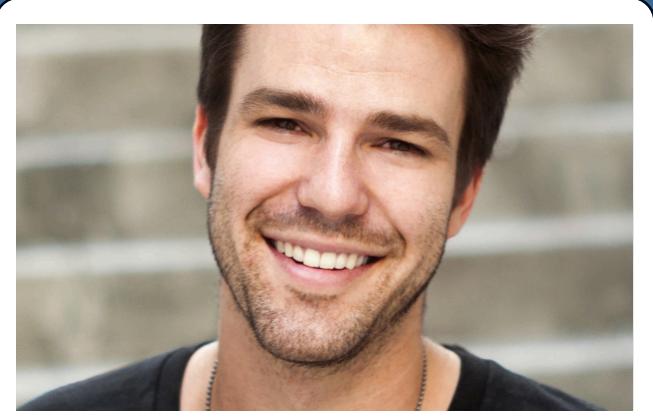
- Comprensión del Negocio
- Comprensión de los Datos
- Preparación de los Datos

EQUIPO

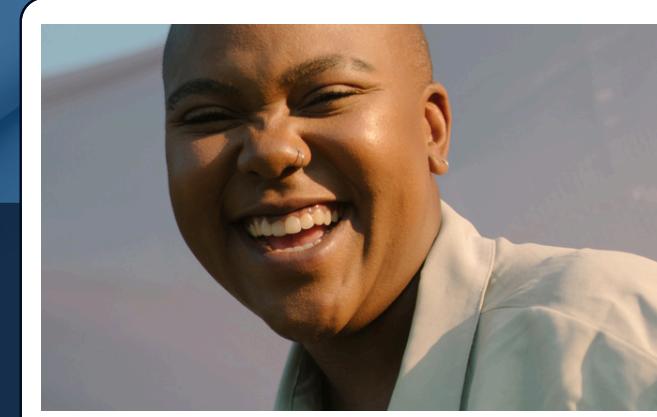
ROLES



- SARAH: DISEÑO DE LA COMUNICACIÓN FINAL, DISEÑO DEL DASHBOARD Y CREACIÓN DEL INFORME.



DANIEL: MODELADO INICIAL Y OPTIMIZACIÓN DEL MODELO.



CRISTIÁN: ANÁLISIS TÉCNICO, EDA, LIMPIEZA DE DATOS

COMPRENSIÓN DE LOS DATOS (EDA)

<https://colab.research.google.com/drive/1KZOC9pxLRJVZGqIJjSVMgnnE5gFsGCGc#scrollTo=sPAWbeN5bTFY>

PREPARACIÓN DE LOS DATOS

<https://colab.research.google.com/drive/1KZOC9pxLRJVZGqIJjSVMgnnE5gFsGCGc#scrollTo=sPAWbeN5bTFY>

MODELADO: CLUSTERING CON K-MEANS

https://github.com/Cristianalarconn/Proyecto_Desaf-o-Integral-de-An-lisis-Modelado-y-Optimizaci-n-de-Datos/blob/main/notebooks/.ipynb_checkpoints/kmeans_clustering-checkpoint.ipynb



CONCLUSIONES GENERALES

- El dataset presenta una buena calidad inicial y permitió un proceso de análisis y limpieza eficiente.
- El uso de técnicas de EDA facilitó la identificación de anomalías y patrones espaciales.
- El modelo K-Means con 4 clusters, StandardScaler y tres variables demostró ser la mejor configuración.
- La variable distancia a Bogotá aporta información relevante al proceso de segmentación.
- Los resultados obtenidos permiten comprender mejor la distribución geográfica de los Centros de Aprendizaje del SENA y sirven como base para análisis estratégicos futuros.

GRACIAS

Andrea Flores

