**Entregable 1: Definición del Proyecto y Adquisición de Datos**

1. **Descripción del proyecto**

El proyecto, titulado “Análisis de rendimiento educativo en España y factores que influyen en el mismo”, consiste en poner en práctica los conocimientos adquiridos durante el Bootcamp de Data Science para realiza un análisis a nivel, nacional, autonómico, provincial y municipal de desempeño educativo, y los factores económicos y demográficos que explican dicho desempeño.

Este proyecto se llevará a cabo de la siguiente manera, se extraerán datos educativos, económicos y demográficos de una serie de fuentes que se detallan más adelante. Se hará un proceso ETL en Python usando la librería Pandas con los mismos y se cargarán en una base de datos de SQL. Se realizará un EDA y modelos de clasificación en clusters de localidades que compartan características semejantes en los aspectos a analizar. Como tecnología no explicada en el aula y que se investiga se usará Power Bi para presentar resultados.

Para dar un contexto temporal unitario al análisis, los datos se extraen del año 2022, eligiendo este año por ser el más cercano a la actualidad del que se disponen de datos completos.

Como medidores del desempeño educativo se usan las puntuaciones del informe PISA de 2022, la tasa de repetidores por curso, comunidad autónoma y provincia y el porcentaje de aprobados en la EBAU.

Los datos educativos, demográficos y económicos que se usan para el proyecto se obtienen mediante web scrapping, y son los siguientes:

* **Ministerio de educación y ciencia:** 
  + Nota EBAU por comunidad autónoma
* **Ministerio de Ciencia, Innovación y Universidades**
  + Universidades por provincia
* **INE:** 
  + Censo nacional, autonómico, provincial y municipal
  + Nivel de estudios completados: Población de 15 y más años por sexo y nivel de estudios (agrupado) (Municipios entre 50 y 500 habitantes)
  + Nivel de estudios completados: Población de 15 y más años por sexo, lugar de nacimiento (España/extranjero) y nivel de estudios (detalle) (Municipios de 500 habitantes o más). Estos datos también se obtienen para Comunidades Autónomas.
  + Estudios en curso: Población de 15 y más años por sexo, nacionalidad (española/extranjera), nivel más alto de estudios en curso (agrupado) y relación entre lugar de residencia y lugar de estudios (Capitales de provincia y principales ciudades)
  + Personas entre 18 y 69 años de edad según máximo nivel educativo alcanzado por sexo e ingresos mensuales netos del hogar
  + Gasto educativo por persona por comunidad autónoma
  + Tasa repetidores por curso, comunidad autónoma, provincia
  + Tasa de paro por provincia, comunidad y nacional
  + Renta media por comunidad, provincia y municipio.
* **Wikipedia (El informe PISA lo publica la O.E.C.D y está disponible en su web, pero en la Wikipedia se encuentra la tabla que resume las puntuaciones que nos interesan, por lo que se escrapea desde ahí)**
  + Puntuación en informe PISA de Matemáticas, ciencias y comprensión lectora, a nivel nacional y autonómico

1. **Contexto de negocio**

Desde el punto de vista de la utilidad empresarial o de negocio, este proyecto se contextualiza en el marco de las administraciones públicas, para mejorar la situación de la educación en España. Concretamente, el Gobierno de la Nación nos pide un análisis sobre el estado de la educación en España sobre el estado de la educación a nivel nacional, autonómico, provincial y municipal. El objetivo es encontrar los factores demográficos y educativos que causan desigualdad en el desempeño educativo. El objetivo es encontrar medidas que favorezcan la igualdad de oportunidades en el acceso y desempeño en la educación pública.

Pretendemos responder a preguntas como:

“¿Influye el nivel de ingresos del hogar en el nivel educativo máximo alcanzado?”

“¿Qué relación hay entre gasto medio educativo por alumno y el desempeño educativo?”

¿Hay diferencias significativas en el nivel máximo educativo alcanzado respecto al sexo y el lugar de nacimiento?

En las conclusiones del análisis trataremos de dar algún insight que sirva para favorecer la igualdad de oportunidades en el derecho a la educación pública. Por ejemplo, encontrar donde hace falta mayor inversión, o que grupos de población necesitan mayor integración e igualdad educativa.