

## PRÁCTICA 1

### DISEÑOS DE PROCESO ETL EN DATA SCIENCE

#### Proceso de extracción: Análisis de Notas del Máster en Data Science

Este documento describe el proceso ETL (Extract, Transform, Load) aplicado al archivo 'notas\_master\_data\_science.csv'. El objetivo del proceso es extraer los datos de los estudiantes, calcular el promedio de notas, clasificar a los estudiantes en aprobados y reprobados, y presentar los resultados mediante diagramas de barras y gráficos tipo rosco (pie chart).

##### 1. Extracción

En esta fase se realiza la lectura del archivo CSV 'notas\_master\_data\_science.csv', el cual contiene 50 registros de estudiantes y 5 columnas correspondientes a las materias del máster en Data Science: Machine Learning, Big Data Analytics, Deep Learning, Data Visualization y Statistics & Probability.

##### 2. Transformación

Durante esta etapa se calculará el promedio de cada estudiante considerando las cinco materias. Posteriormente, se clasificará a los estudiantes como 'Aprobados' si su media es igual o superior a 60, y 'Reprobados' en caso contrario.

##### 3. Carga y Visualización

Los resultados se presentarán mediante gráficos de barras que muestran la distribución de notas por materia y un diagrama tipo rosco que ilustra el porcentaje de aprobados y reprobados. Estas visualizaciones permiten analizar rápidamente el desempeño general de la cohorte.

##### 4. Tareas Detalladas

1. Cargar el archivo CSV utilizando la librería pandas.
2. Calcular el promedio de cada estudiante.
3. Clasificar a los estudiantes como Aprobados ( $\geq 60$ ) o Reprobados ( $< 60$ ).
4. Generar un gráfico de barras mostrando la media por materia.
5. Generar un gráfico de rosco mostrando el porcentaje de aprobados y reprobados.
6. Guardar las gráficas y el resumen estadístico en un reporte final.