## **Análisis Académico - Tarea 1: La Primera Inspección**

### **Pregunta 1: Separador de Columnas**

Se identifican dos separadores distintos:

* **Coma (,)**: Es el separador dominante, utilizado por **seis** archivos: Alumnos.csv, Calificaciones.csv, Cursos.csv, Grupos.csv, Horas.csv, Modulos.csv, Lineas.csv, Objetivos.csv y Procesos.csv.
* **Punto y Coma (;)**: Es utilizado por un único archivo del set de indicadores: Indicadores\_Finales.csv.

### **Pregunta 2: Encabezados (Nombres de Columnas)**

En todos los ficheros la primera fila contiene los nombres de las columnas.

* **Claridad**: Los encabezados son generalmente claros y descriptivos, aunque utilizan una mezcla de idiomas (español, valenciano) y formatos. Por ejemplo, en Alumnos.csv encontramos fecha\_nac y fecha\_matricula, ambos utilizando snake\_case y siendo concisos.
* **Excepción**: En Indicadores\_Finales.csv, los encabezados como Valor-A, Valor-T1, Cod\_SQ son menos explícitos a primera vista, pero el contexto sugiere que son valores de evaluación (Anual, Trimestres) y códigos de calidad.

### **Pregunta 3: Valores Extraños o Faltantes (Inspección Visual)**

La inspección visual revela varias áreas críticas que necesitarán limpieza de datos:

* **Celdas Vacías y Nulos Explícitos**: Se observan numerosos valores faltantes:
  + En Alumnos.csv y Grupos.csv, se aprecian celdas completamente vacías en columnas como DNI, municipio\_nac\_ext, y aula, que deberán codificarse como NaN (Not a Number) en Pandas.
  + En Calificaciones.csv, la columna observacion contiene el texto literal "No hay datos guardados", que actúa como un valor nulo y deberá ser reemplazado.
  + En Cursos.csv, la columna padre (que define jerarquía) tiene celdas vacías ( ).
* **Consistencia en Fechas**:
  + En Alumnos.csv, la columna fecha\_nac utiliza el formato DD/MM/AAAA (ej: 28/05/1982).
  + En Cursos.csv, Grupos.csv y Modulos.csv, la columna fecha\_exportacion utiliza el formato AAAA-MM-DD HH:MM:SS.
  + Ambos formatos son consistentes dentro de sus respectivos ficheros, pero deberán unificarse si se combinan.

### **Pregunta 4: Formatos de Datos Inconsistentes**

El principal problema de formato se encuentra en la representación de los números y las fechas:

* **Separador Decimal**: La columna de notas en Calificaciones.csv y las columnas de valor en Indicadores\_Finales.csv (ej: Valor-A) probablemente usan la coma (,) como separador decimal (ej: 7,5). Esto es inconsistente con el estándar informático anglosajón, por lo que las comas deberán ser reemplazadas por puntos (.) para que las columnas puedan ser tratadas como datos numéricos (*float*).

## **Pregunta 5: Claves y Relaciones en Datos\_2022.zip (Académico)**

Este conjunto se centra en el estudiante, sus calificaciones, y la estructura de la oferta educativa.

| Fichero | Clave Primaria (PK) | Función y Relaciones (FK) |
| --- | --- | --- |
| Alumnos.csv | NIA (Número de Identificación del Alumno) | Es la tabla principal de la dimensión Alumno. Su NIA se utiliza como clave foránea en Calificaciones.csv (columna alumno) para vincular a cada nota con el estudiante que la obtuvo. |
| Cursos.csv | codigo | Define la estructura curricular. Su codigo se usa como FK para: 1) Grupos.csv (para saber a qué curso pertenece un grupo) y 2) Cursos.csv (columna padre) para crear una jerarquía (ej. para enlazar un Grado Medio a su Familia Profesional). |
| Modulos.csv | codigo | Define las asignaturas/unidades didácticas. Su codigo se utiliza como FK en: 1) Calificaciones.csv (columna contenido) para identificar la asignatura calificada y 2) Horas.csv (columna CODIGO) para determinar la carga lectiva del módulo. |
| Grupos.csv | codigo | Identifica grupos específicos de estudiantes. Su codigo se relaciona con la tabla de matriculación (que no está incluida, pero que a su vez se vincularía con Alumnos.csv). Su principal FK es hacia Cursos.csv. |
| Calificaciones.csv | (No tiene una PK simple) | Es una tabla de Hechos (transaccional). Su clave es compuesta (implícita) por: alumno (FK a Alumnos), curso (FK a Cursos), contenido (FK a Modulos) y evaluacion. |
| Horas.csv | CODIGO (compuesta por CODIGO y CICLO) | Registra la carga lectiva. Su CODIGO se relaciona con el codigo de Modulos.csv. |

La cadena de relaciones más importante comienza con el NIA del alumno, que se vincula a Calificaciones.csv. Dentro de Calificaciones.csv, la nota se une al codigo del módulo (en Modulos.csv) y al codigo del curso (en Cursos.csv).

## 

## 

## 

## **Claves y Relaciones en Indicadores.zip (Gestión Estratégica)**

Este conjunto modela la planificación y la medición de la calidad, siguiendo una jerarquía clara:

| Fichero | Clave Primaria (PK) | Función y Relaciones (FK) |
| --- | --- | --- |
| Lineas.csv | Linea | Es la capa superior de la estrategia. Su Linea se utiliza como clave foránea en Objetivos.csv (columna Linea) para agrupar los objetivos bajo esa línea estratégica. |
| Objetivos.csv | Objetivo\_PAA | Define los objetivos específicos. Su Objetivo\_PAA se relaciona como FK en Procesos.csv (probablemente una columna llamada Objetivo o similar) para vincular los objetivos a los procesos que los ejecutan. |
| Procesos.csv | Proceso | Identifica las acciones operacionales. Su Proceso se relaciona como FK en Indicadores\_Finales.csv (a través de la columna Cod\_PAA o Identificador) para medir el rendimiento de esa actividad. |
| **Indicadores\_Finales.csv** | **Identificador** (compuesta con Curso) | Es la tabla de Hechos (Métrica). Contiene los valores alcanzados (Valor-A). Sus claves foráneas apuntan a los otros tres archivos a través de las columnas Cod\_PAA o Identificador, permitiendo desagregar el resultado de la medición hasta el objetivo o línea estratégica. |

La jerarquía de gestión se construye de arriba hacia abajo: Una Linea incluye varios Objetivos\_PAA, que a su vez son gestionados por distintos Procesos. Los Indicadores\_Finales.csv miden el éxito de todo este árbol, utilizando Cod\_PAA o el Identificador para vincular la cifra con el punto exacto de la estrategia al que corresponde.