**PRÁCTICA PROFESIONALIZANTE 2**

**TRABAJO PRÁCTICO #4**: EDA DataSet Alumnos / Carreras / Materias: Descripción y Pronóstico

**Escenario o Situación Profesional:**

Deben suponer que yo soy una persona interesada en Comprar el Instituto, y como tal, los he contratado para que puedan “Describirme” el funcionamiento del mismo, y la tendencia que tiene en cuanto a matrícula y deserción cada carrera. Esto debería permitirme evaluar los recursos y estrategias publicitarias utilizadas hasta el momento, determinar la viabilidad o posibilidades de continuidad de las carreras, etc…

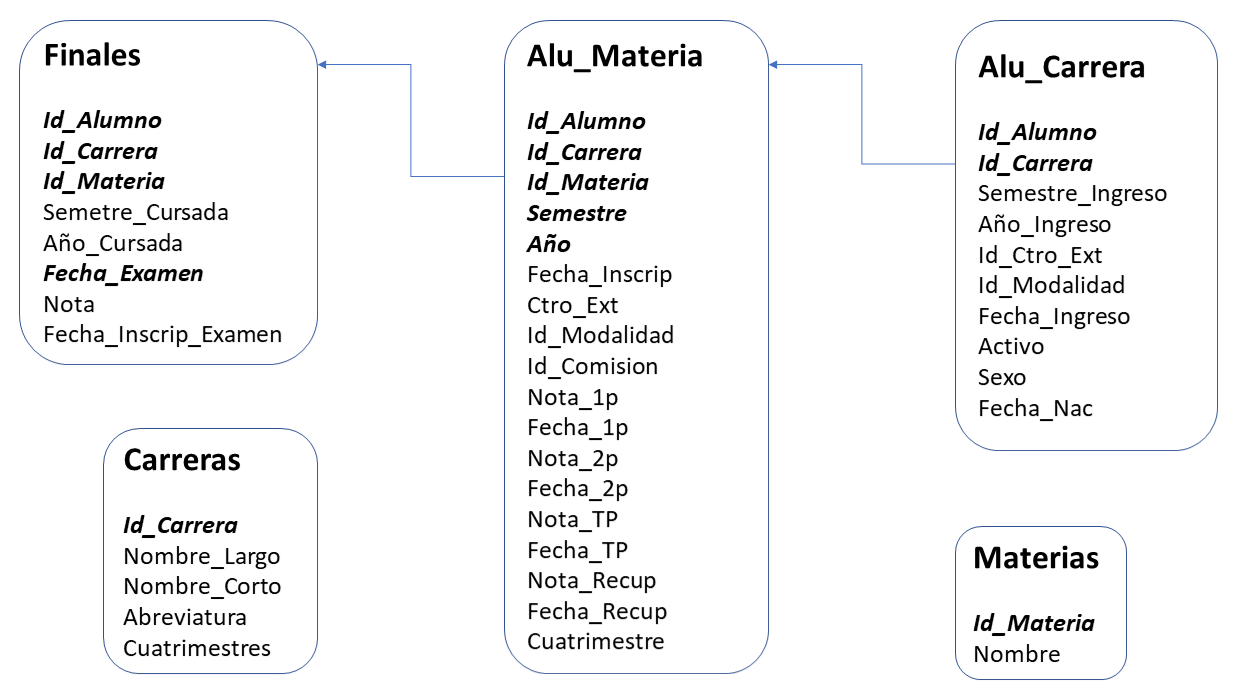
No deben darme consejos de marketing ni publicidad, ni tampoco económicos ni financieros. Para eso cuento con mi equipo de especialistas: pero estos especialistas no pueden hacer su trabajo, asesorándome, si no cuentan con información altamente descriptiva! Finalmente, la posibilidad de pronosticar (si fuera posible) la matrícula para 2011.

Para ello cuentan con los datos de 10 años! Desde el 2001 hasta 2010. En ellos podrán observar para cada alumno o exalumno qué carrera cursó, hasta dónde llegó en la misma (es decir, cuál fue el máximo cuatrimestre cursado por ese alumno en su carrera).

**Sobre el DataSet:**

Datos de matrícula (inscripción a la carrera) y estado de avance de los alumnos del Instituto en sus respectivas carreras, incluyendo el grado de avance de cada uno, materias cursadas y rendidas.

El DataSet se provee como conjunto de archivos CSV con la siguiente estructura:



Los nombres de las tablas y los campos (columnas) son suficientemente autodescriptivos. No obstante, siguen a continuación algunas referencias:

* La tabla **Materias** es paramétrica: simplemente contiene el nombre de cada materia referida en las restantes tablas, con su **Id\_Materia** asociado.
* Con la tabla **Carreras** sucede lo mismo, su Id\_**Carrera** asociada a tres versiones de su nombre de carrera o abreviatura. Esta tabla contiene además un campo importante: **Cuatrimestres**. Este campo indica cuántos cuatrimestres tiene el plan de estudio de la carrera. Es decir, si para la carrera X encontramos que tiene 4 o 6 cuatrimestres, los registros de cursado de cada materia por cada alumno (tabla Alu\_Materia), en el campo en que refieren el Cuatrimestre al que pertenece la materia cursada, no podrán tener valores mayores a 4 o 6 respectivamente.
* **Id\_Alumno** : Dato “encriptado” que representa el DNI de cada alumno. Un alumno aparecerá en las tablas tantas veces como carreras haya cursado, o materias, o exámenes haya rendido.
* **Sexo**: M = Masculino ; F = Femenino
* **Fecha\_Nac**: Ojo, cuidado con la interpretación del formato en pandas, Excel, o las herramientas que usen.
* **Año\_Ingreso y Semestre\_Ingreso**: Lo dicho. El año y semestre (1 ó 2) en que ingresó a la carrera en cuestión.
* **Fecha\_Ingreso**: Este dato puede parecer redundante atento a que tenemos en las dos columnas previas el año y semestre de ingreso, pero de hecho es una precisión adicional y un dato importante a tener en cuenta ya que pueden producirse ingresos prematuros o tardíos a un determinado semestre. Dicho de otro modo, una persona puede, el 15 de Septiempre, inscribirse (tardíamente) para cursar en el 2do semestre de ese año, o por el contrario, estar inscribiéndose (tempranamente) para cursar en el 1er semestre del año próximo. Quizá haya alguna incidencia de qué tan pronto o tan tarde se inscribe un alumno en una carrera, respecto de la probabilidad de que el mismo termine siendo un desertor.
* **Alu\_Materia->Cuatrimestre**: Cuatrimestre (dentro del plan de estudio) en el que está situada la materia cursada.
* **Notas** (calificaciones): En la tabla de cursado (Alu\_Materia) y también en la de finales tenemos notas. Estas son valores enteros comprendidos entre 0 (cero) y 10 (diez). También encontrará que muchos registros tienen como nota un 11 (once): cuidado!! Esto significa “Ausente” y evidentemente dicho valor no debe computarse en ningún cálculo de promedio.
* Atributos implícitos que pueden inferirse a partir de los datos datos

Pueden extraerse muchos atributos (datos nuevos) a partir de los existentes, como por ejemplo los siguientes:

* + **Edad del alumno al inscribirse en la carrera**

Es una simple resta... Este atributo es interesante de analizar por carrera, por semestre, etc

* + **Condición de Recursante en una materia**

El alumno es recursante si tiene un registro previo de la misma materia en la misma carrera: evidentemente ha quedado libre y por eso debió recursarla.

* + **Condición de finalización en el cursado de una materia**

Al terminar el cursado de una materia el alumno puede encontrarse en una de tres condiciones posibles:

* + - **Promocionado**: aprobó los dos Exámenes Parciales y la nota de Trabajo Práctico con 7 (siete) o más. Evidentemente no rindió recuperatorio.
    - **Regular**: puede haber desaprobado uno o ambos parciales, o haber estado ausente, y haber aprobado el recuperatorio. Independientemente de ello, tiene que tener aprobado la nota de Trabajo Práctico. En todos los casos, la nota mínima para aprobar es 4 (cuatro).

Otra forma de estar regular es que no haya necesitado rendir el recuperatorio, ya que aprobó en tiempo y forma los dos parciales y el trabajo práctico, pero en al menos una de esas tres notas obtuvo una calificación menor a 7 (siete).

* + - **Libre**: si el alumno no aprobó o faltó a uno o ambos parciales, y tampoco aprobó o faltó al recuperatorio está Libre. También pasa automáticamente a la categoría de Libre si tiene reprobado o ausente la nota de Trabajo Práctico, ya que esa nota no se puede levantar con el recuperatorio.
  + **Condición de Desertor / Alumno regular / Egresado**

A partir del máximo cuatrimestre que el alumno llegó a cursar (puede ser un valor comprendido entre 1 y 6), y en función de las fechas o periodos en que cursó, podemos considerar si el alumno ha desertado, si está cursando de manera regular, o si es un egresado en su respectiva carrera.

A los efectos prácticos, consideraremos que quienes llegaron a 6to cuatrimestre no han sido desertores, y los consideraremos “**Egresados**” si, además, tienen rendidas y aprobadas todas las materias cursadas.

Por otra parte, consideraremos a los alumnos que llegaron al cuatrimestre que les correspondería según su año de ingreso como alumnos “**Regulares**”: por ejemplo, un alumno que se inscribió en el primer semestre de 2009 y tiene materias cursadas en 2009/1, 2009/2, 2010/1 y 2010/2, evidentemente no es un desertor ya que “está cursando”. Obviamente no llegó al último semestre de su carrera porque el tiempo y su ritmo de cursado no se lo ha permitido, pero está cursando regularmente.

Dicho en otras palabras, cuando un alumno no alcanzó el último semestre de su carrera y ha dejado de cursar, lo consideraremos “**Desertor**”.

**Enfoque de los análisis de inscripciones**

En el Instituto todas las materias son cuatrimestrales, es decir que el ciclo de análisis es cada semestre: resulta conveniente comparar el desempeño de las inscripciones en el 1er semestre de un año respecto del 1er semestre del anterior, o del siguiente, pero no debe compararse el 1er semestre con el 2do, ya que los ingresos en Agosto son siempre mucho menores que en Marzo.

**Objetivo:**

* Analizar los datos provistos procurando entender la lógica de la información suministrada.
* Generar informes (especialmente gráficos) que permitan “Describir” los datos, sus características, tendencias y diferentes niveles de agrupamiento en la información. Pueden hacerse cortes por sexo, por edad (o año de nacimiento por ejemplo) – incluso considerar banas etarias-, además de lo obvio que sería por carrera y semestre-año de análisis.
* En los informes considerar especialmente los atributos artificiales que haya podido generar tal como condición de regularidad de los alumnos en cada materia, condición del alumno frente a la carrera, etc…
* Pronosticar la matrícula (inscripción de ingresantes) y deserción esperable para el periodo siguiente a los datos suministrado: 1er y 2do semestre de 2011.

**Metodología:**

Haremos esto como en la realidad sucede: cada quien hará por su cuenta y criterio la tarea indicada, y me presentará la información a la que arriben. Obviamente, por motivos académicos, deberán entregarme también el código fuente con el que lo sustentan. Pueden hacer un análisis previo con alguna otra herramienta como Excel, manejo de datos con algún gestor de base de datos, etc: también deberá estar documentado.

La información y avance serán discutidos con cada uno en reunión de trabajo, y podré darle una segunda o tercera vuelta de rosca pidiéndoles detalles más precisos, u otros informes, según lo que cada uno me presente.

**Presentación Final y Calificación**

El trabajo desarrollado y presentado será considerado para la calificación de esta materia; y el mismo trabajo - con las modificaciones y agregados que se indiquen en su calificación - será presentado en defensa oral como examen final de esta asignatura.