

# Estructuras de Datos

## Universidad Nacional de Tres de Febrero

### Guía de ejercicios 4

#### Archivos CSV

Los archivos CSV nos permite guardar información estructurada en un archivo de texto, separando los campos con un delimitador que generalmente es ";" (punto y coma) o "," (coma) En esta guía de ejercicios vamos a trabajar con los datasets de ciencia y tecnología de Argentina, publicados en:

<https://datasets.datos.mincyt.gob.ar/dataset> (<https://datasets.datos.mincyt.gob.ar/dataset>)

Los datasets publicados son:

1. **Empresas de ejecución de proyectos de ciencia, tecnología e innovación:** Empresas de ejecución de proyectos financiados por la Agencia Nacional de Promoción Científica y Tecnológica (Argentina, 2001-2016). Registro Nacional de Proyectos de Investigación, Desarrollo e Innovación
2. **Proyectos de ciencia, tecnología e innovación:** Registro Nacional de Proyectos de Investigación, Desarrollo e Innovación (Argentina, 2001-2016)
3. **Recursos humanos en ciencia y tecnología:** Registro Unificado y Normalizado a nivel nacional de los Datos Curriculares del personal científico y tecnológico que se desempeña en las instituciones argentinas (Registro CVar) (Argentina, 2016)
4. **Equipamiento científico-tecnológico:** Equipamiento científico-tecnológico registrado en Instituciones del Sistema Nacional de Ciencia y Tecnología Argentino adheridas al Programa de Grandes Instrumentos y Bases de Datos (Argentina, 2016). Registro Nacional de Instituciones de Ciencia, Tecnología e Innovación
5. **Instituciones del Sistema Nacional de Ciencia y Tecnología Argentino:** Instituciones del Sistema Nacional de Ciencia y Tecnología Argentino adheridas al Registro CVar (Argentina, 2016)

A continuación se presentan una serie de consultas que para resolverlas vamos a utilizar Python, programando las funciones que sean necesarias para lo cual vamos a utilizar las versiones csv de los datasets. Estos datasets utilizan "," (coma) como delimitador.

Para leer correctamente los caracteres especiales (vocales con tildes, ñ, etc.) se recomienda utilizar la codificación utf-8. Por ejemplo para abrir el primer dataset solo para lectura y con codificación utf-8 se puede usar la siguiente instrucción:

```
ifile=open("1 - empresas-ejecutoras-proyectos.csv",encoding='utf-8', mode="r")
```

#### Ejercicio 1: ¿Cuál es la empresa que más proyectos ejecutó?

Dataset 1

In [14]:

```
import csv  
  
ifile=open("1 - empresas-ejecutoras-proyectos.csv",encoding='utf-8', mode="r")
```

**Ejercicio 2: ¿En cuál provincia se ejecutaron más proyectos de empresas? y ¿Cuál provincia tiene más empresas que ejecutan proyectos?**

Dataset 1

In [15]:

```
import csv  
  
ifile=open("1 - empresas-ejecutoras-proyectos.csv",encoding='utf-8', mode="r")
```

**Ejercicio 3: ¿Cuál es la región de la Argentina que recibió más dinero para financiar proyectos? y ¿Cuál es el tipo de organización que menos aporte realizó?**

Dataset 2 El aporte que realiza una organización es el monto\_total - monto\_financiado

In [16]:

```
import csv
```

**Ejercicio 4: Escribir un archivo en formato csv, usando como delimitador ";" (punto y coma) con los siguientes campos: provincia, gran\_area\_de\_formación, cantidad\_RRHH.**

Dataset 3

In [ ]:

**Ejercicio 5: Escribir un archivo en formato csv, usando como delimitador ";" con los siguientes campos: provincia, tipo\_personal, cantidad\_RRHH.**

Dataset 3

In [ ]:

**Ejercicio 6: Escribir un archivo CSV con los siguientes campos provincia, cantidad de empresas, monto total financiado, totalizado por provincia (cada provincia aparece una única vez).**

Datasets 1 y 2

In [ ]:

**Ejercicio 7: Mostrar por pantallas todas las provincias que cuentan con personal en área de formación "ciencias médicas y de la salud" y no cuentan con "Resonador Magnético"**

Datasets 3 y 4

In [ ]:

**Ejercicio 8: Mostrar por pantallas todas las provincias que cuentan con personal en área de formación "ingenierías y tecnologías" y con "Computadora de Alto Desempeño" o "Láseres" como equipamiento y la cantidad de proyectos ejecutados por "Consortio público" o "Consortio público-privado"**

Datasets 2, 3 y 4

In [ ]: