

Trabajando con datos externos con SciPy:

Ejercicios propuestos

Cargar datos con pandas

Sobre el ejemplo de clase ("churn"):

- Seleccionar tres campos numéricos del dataset, diferentes de las vistas en clase.
- Comprobar si tienen correlaciones entre ellos, consultando una medida de correlación primero, y después tratando de hacer una regresión lineal.
- Observar la distribución de los atributos numéricos. Alguno parece una distribución conocida?
- Comprobar si el dataset está balanceado, es decir, la proporción (en porcentaje) de files en las que Churn? Es Yes respecto a las que es No.
- Examinar el campo "State". ¿Cuántos valores diferentes tiene? Contar el número de veces que Churn?=Yes. para cada estado, para observar si tiene alguna influencia.

Trabajo con XML

Sobre el ejemplo de clase:

- Comprobar los tipos de ayuda diferentes que existen. Nota: está en el elemento `AssistanceType`
- Dibujar los 5 años que USA ha dado más cantidad de ayuda internacional y los 5 que menos, en unidades monetarias constantes.

Sobre un nuevo dataset:

- Descargar la información de aparcamientos del gobierno de UK:
<http://data.gov.uk/dataset/car-parks>
- Obtener la estadística de parkings por ubicación. Nota: está en el element Location: `<Location>Aberystwyth</Location>` pero solamente cuando está directamente bajo el elemento `<CarPark>`.
- Obtener la capacidad en plazas por location.

Scrapping

Tema libre, tomar una página Web con contenido tabular que combine cadenas y números y tratar de moverlo a un DataFrame.

Trabajando con APIs

Sobre el ejemplo de AngellList:

- Utilizando bucles y teniendo en cuenta que en las propias respuestas del API nos indica datos de paginación, capturar todos los datos de AngelList.
- Analizar los salarios que ofrecen las startups de AngelList.
- Buscar las startups que tienen más followers y dibujar una distribución del número de followers. Puede utilizarse la llamada del API: **GET /startups?filter=raising**

SQL en Python

- Crear una base de datos para modelar la app Wallapop e insertar algunos datos de ejemplo. Hacer consultas con SQL que impliquen a más de una tabla.
- Cargar datos con SQL en un DataFrame de pandas.