

# Programación para *Data Science*

## Unidad 8: Visualización de datos en Python - Ejercicios y preguntas

### Ejercicio 0

Cargad el conjunto de datos `pulitzer-circulation-data.csv` en un *dataframe* de pandas. La fuente original de este conjunto de datos es el repositorio de datos de [FiveThirtyEight](https://github.com/fivethirtyeight/data) (<https://github.com/fivethirtyeight/data>).

```
In [1]: # Respuesta
```

### Ejercicio 1

Generad un diagrama de dispersión que permita visualizar las muestras del conjunto de datos de pulitzer según las variables 'Daily Circulation 2004' y 'Daily Circulation 2013'. Incluid una recta con el ajuste lineal entre ambas variables. Por lo que respecta a los detalles de visualización, limitad la visualización de ambos ejes al intervalo (0, 1000000), usad el estilo `whitegrid` de seaborn, y dibujad los puntos y la línea en color verde. **(2.5 puntos)**

Pista: Podéis usar la función `jointplot` que hemos visto en el notebook de explicación. Pensad en qué tipo de gráfica, [de entre las que ofrece jointplot](http://seaborn.pydata.org/generated/seaborn.jointplot.html) (<http://seaborn.pydata.org/generated/seaborn.jointplot.html>), se ajusta a los requerimientos del enunciado.

```
In [ ]: # Respuesta
```

### Ejercicio 2

En la universidad se está preparando un viaje de fin de semestre para ir a esquiar a Suiza. El viaje se realizará por grupos reducidos, lo que agilizará la coordinación.

Dadas las características del viaje, cada grupo deberá disponer de algunos recursos para realizarlo: como mínimo se necesitará una furgoneta y un conductor. Solo los grupos que tengan estos recursos podrán realizar el viaje.

1. ¿Cuántos grupos de estudiantes podrán realizar el viaje?
2. ¿Cuáles son los estudiantes que podrán ir de viaje?

Para tomar la decisión, tenéis acceso a los siguientes datos sobre los estudiantes de la universidad:

- Un grafo con las relaciones entre los estudiantes (dos estudiantes se conocen si hay una arista entre ellos en el grafo),
- Dos listas que indican los estudiantes que tienen furgoneta y cuáles tienen carné de conducir, respectivamente

Para poder formar parte de un grupo de viaje, un estudiante debe conocer como mínimo a otro estudiante del grupo. De este modo puede ponerse en contacto con el grupo y coordinar el viaje. Los estudiantes tienden a formar grupos tan grandes como les sea posible.

Nota 1: podéis asumir que la capacidad de una furgoneta es infinita ;)

Nota 2: para resolver el ejercicio, **debéis generar una visualización del grafo que os permita responder a las dos preguntas planteadas**. Después, **recordad responder a ambas preguntas!**

**(2.5 puntos)**

```
In [2]: # Datos disponibles para el ejercicio

import networkx as nx

# Grafo G, que representa las relaciones entre los estudiantes
G = nx.Graph()

G.add_edges_from([(1,2), (1,3), (3, 5), (16, 18), (3, 4), (15, 16), (1, 4), \
                  (7, 8), (8, 9), (9, 7), (9, 10), (7, 10), (4, 5), (2, 5), (10, 11), \
                  (11, 7), (13, 8), (14, 8), (16, 17), (17, 18), (18, 19), (17, 19), \
                  (0, 1), (0, 3), (12, 6)])

# Lista de estudiantes con furgoneta
tienen_furgoneta = [17, 18, 19, 5, 12]

# Lista de estudiantes conductores
conductores = [15, 16, 5, 6, 7, 8, 9, 10]
```

```
In [4]: # Respuesta
```

### Ejercicio 3

Crea un mapa con la librería `geoplotlib` que muestre las 7 maravillas del mundo moderno. El mapa debe permitir ver la localización de cada maravilla y su nombre. **(2.5 puntos)**

Nota 1: Podéis guardar las localizaciones en un fichero `.csv` como el del ejemplo de las paradas de autobuses del notebook de explicación o bien [cargar directamente los datos](https://github.com/andrea-cuttone/geoplotlib/wiki/User-Guide#loading-data) (<https://github.com/andrea-cuttone/geoplotlib/wiki/User-Guide#loading-data>) a partir de un diccionario o un *dataframe*.

Nota 2: Podéis obtener las coordenadas geográficas de un lugar con el servicio [LatLong.net](http://www.latlong.net/) (<http://www.latlong.net/>) o usando [Google Maps](https://www.google.es/maps) (<https://www.google.es/maps>) (en este último caso, una vez hayáis buscado el lugar, podéis extraer las coordenadas de la url).

```
In [5]: # Respuesta
```

### Ejercicio 4

Generad una gráfica que muestre el número de votos y el número de escaños obtenidos de los 10 partidos más votados en las elecciones generales en España de 2016 para el congreso. Podéis generar el tipo de gráfica que consideréis más oportuno, siempre y cuando permita visualizar en **una única gráfica** ambos resultados (votos y escaños), de manera que sea fácil comparar ambos resultados. Recordad incluir toda la información necesaria para poder interpretar luego la gráfica adecuadamente.

Nota: para realizar el ejercicio, en primer lugar necesitaréis buscar y descargar los datos de interés. Después, habrá que preprocesarlos, para finalmente generar la visualización que se pide en el ejercicio.

**(2.5 puntos)**

```
In [2]: # Respuesta
```