

## Actividad U1. Análisis de valoraciones (50%)

El objetivo de esta actividad es que seáis capaces de desarrollar un flujo de acciones robusto que permitan obtener una serie de indicadores a partir de información textual.

Para ello utilizaremos el catálogo de valoraciones sobre hoteles que podéis encontrar en:

<http://kavita-ganesan.com/entity-ranking-data/>

### **Dataset Description**

- Full reviews of hotels in 10 different cities (Dubai, Beijing, London, New York City, New Delhi, San Francisco, Shanghai, Montreal, Las Vegas, Chicago)
- There are about 80-700 hotels in each city
- Extracted fields include date, review title and the full review
- Total number of reviews: ~259,000

### **Format**

There should be 10 different folders representing the 10 cities mentioned earlier. Each file (within these 10 folders) would contain all reviews related to a particular hotel. The filename represents the name of the hotel. Within each file, you would see a set of reviews in the following format:

```
Date1<tab>Review title1<tab>Full review 1
Date2<tab>Review title2<tab>Full review 2
```

Each line in the file represents a separate review entry. Tabs are used to separate the different fields.

### **Analizaréis los siguientes apartados:**

Para el conjunto de hoteles de cada ciudad:

1. ¿Qué partes de una **habitación** son las más mencionadas? P.ej. baño, cama, jacuzzi, ...
2. ¿Qué servicios pueden detectarse por cada hotel? Por ejemplo, gimnasio, spa, piscina, restaurante, ...
3. ¿Qué lugares y qué porcentaje de revisiones sobre cada ciudad mencionan otras ciudades, regiones o puntos de interés turístico? Por ejemplo, con vuestra información podríamos llegar a responder a la pregunta: ¿en Las Vegas se menciona el Gran Cañón del Colorado?

### **La entrega ha de contener la siguiente información:**

- Scripts de Python o url a vuestro repositorio.
- Pequeño informe en pdf sobre el proceso seguido, y los resultados obtenidos.

**Importante:** Como ya sois conscientes, el alcanzar una alta eficiencia, eficacia y robustez en vuestra propuesta conlleva un planteamiento minucioso de una gran cantidad de casos. No pretendo que tengáis unas cuotas perfectas, pero me gustaría que cuando os decantéis por un conjunto de procedimientos intentéis explicar los puntos fuertes y débiles. En otras palabras, deberéis identificar o describir el alcance de vuestra propuesta