



Guía de ejercicios de la segunda clase

Estos ejercicios son para trabajar analizando la base de datos sobre árboles en parques y plazas en la Ciudad de Buenos Aires y con la de árboles en veredas.

Primero deben descargar el archivo de árboles en parques, en formato csv desde el siguiente link: [Arbolado en espacios verdes](#).

Ejercicios

1. Definir una función `leer_parque(nombre_archivo, parque)` que abra el archivo indicado y devuelva una lista de diccionarios con la información del parque especificado. La lista debe tener un diccionario por cada árbol del parque elegido. Dicho diccionario debe tener los datos correspondientes a un árbol (recordar que cada fila del csv corresponde a un árbol).

Probar la función en el parque 'GENERAL PAZ' y debería dar una lista con 690 árboles.

2. Escribir una función `especies(lista_arboles)` que tome una lista de árboles como la generada en el ejercicio anterior y devuelva el conjunto de especies (la columna 'nombre_com' del archivo) que figuran en la lista.
3. Escribir una función `contar_ejemplares(lista_arboles)` que, dada una lista como la generada con `leer_parque(...)`, devuelva un diccionario en el que las especies sean las claves y tengan como valores asociados la cantidad de ejemplares en esa especie en la lista dada.
Debería verse que en el parque General Paz hay 20 Jacarandás, en el Parque Los Andes hay 3 Tilos y en Parque Centenario hay 1 Laurel.
4. Escribir una función `obtener_alturas(lista_arboles, especie)` que, dada una lista como la generada con `leer_parque(...)` y una especie de árbol (un valor de la columna 'nombre_com' del archivo), devuelva una lista con las alturas (columna 'altura_tot') de los ejemplares de esa especie en la lista.

Observación: Conviene devolver las alturas como números (de punto flotante) y no como cadenas de caracteres. Sugerimos hacer esto modificando `leer_parque(...)` o modificando el tipo del valor antes de utilizarlo.



Usar la función para calcular la altura promedio y altura máxima de los 'Jacarandá' en los tres parques mencionados. Debería obtenerse esto:

| Medida | General Paz | Los Andes | Centenario |
|--------|-------------|-----------|------------|
| max | 16.0 | 25.0 | 18.0 |
| prom | 10.2 | 10.54 | 8.96 |

5. Escribir una función `obtener_inclinaciones(lista_arboles, especie)` que, dada una lista como la generada con `leer_parque(...)` y una especie de árbol, devuelva una lista con las inclinaciones (columna 'inclinacio') de los ejemplares de esa especie.
6. Combinando la función `especies()` con `obtener_inclinaciones()` escribir una función `especimen_mas_inclinado(lista_arboles)` que, dada una lista de árboles devuelva la especie que tiene el ejemplar más inclinado y su inclinación.

Correrlo para los tres parques mencionados anteriormente. Debería obtenerse, por ejemplo, que en el Parque Centenario hay un Falso Guayabo inclinado 80 grados.

7. Volver a combinar las funciones anteriores para escribir la función `especie_promedio_mas_inclinada(lista_arboles)` que, dada una lista de árboles devuelva la especie que en promedio tiene la mayor inclinación y el promedio calculado.

Resultados. Debería obtenerse, por ejemplo, que los Álamos Plateados del Parque Los Andes tiene un promedio de inclinación de 25 grados.

Vamos a trabajar ahora también con el archivo [de árboles en veredas](#). Queremos estudiar si hay diferencias entre los ejemplares de una misma especie según si crecen en un un parque o en la vereda. Para eso tendremos que juntar datos de dos bases de datos diferentes.

Explorar el dataset nuevo de árboles en veredas.

Armamos un DataFrame `data_arboles_veredas` que tenga solamente las siguientes columnas: 'nombre_cientifico', 'ancho_acera', 'diametro_altura_pecho', 'altura_arbol'

Sugerimos trabajar al menos con las siguientes especies seleccionadas:

```
especies_seleccionadas = ['Tilia x moltkei', 'Jacaranda mimosifolia', 'Tipuana tipu']
```



Advertencia: El GCBA usa distintos nombres para especie, altura y diámetro según el dataset, por ejemplo 'altura_tot' en uno y 'altura_arbol' en otro. Los nombres científicos varían de un dataset al otro. Por ejemplo 'Tipuana Tipu' se transforma en 'Tipuana tipu'.

Proponemos los siguientes pasos para comparar los diámetros a la altura del pecho de las *tipas* en ambos tipos de entornos.

8. Para cada dataset, armar otro seleccionando solamente las filas correspondientes a las tipas (llamalos `df_tipas_parques` y `df_tipas_veredas`, respectivamente) y las columnas correspondientes al diámetro a la altura del pecho y alturas. Usar como copias (usando `.copy()`) para poder trabajar en estos nuevos dataframes sin modificar los dataframes grandes originales. Renombrar las columnas necesarias para que se llamen igual en ambos dataframes.
9. Agregar a cada dataframe (`df_tipas_parques` y `df_tipas_veredas`) una columna llamada 'ambiente' que en un caso valga siempre 'parque' y en el otro caso 'vereda'.
10. Concatenar los dataframes.
11. Explorar y analizar sobre la cuestión planteada:

¿Hay diferencias entre los ejemplares de una misma especie según si crecen en un un parque o en la vereda?