

En esta práctica usted ejercitará el procesamiento de archivos usando las funciones propias de python, las más avanzadas de **numpy**, y **pandas** para tratar bases de datos. Los archivos que requiere para la guía lo encontrará en el directorio **Data/Guia** de la unidad.

1. Use las funciones de python (sin utilizar numpy) para leer el archivo "visitantes\_cine\_2016.dat". Defina correctamente la ruta al archivo con la variable **str\_file** y use las tres funciones descritas abajo.

```
# -*- coding: UTF-8 -*-
```

```
str_file = '[Reemplazar con ruta al archivo visitantes_cine_2016.dat]'
```

```
f_in = open(str_file, 'r')
```

```
%ejemplo de read
```

```
data = f.read
```

```
%ejemplo de readline
```

```
data = f.readline
```

```
%ejemplo de readlines
```

```
data = f.readlines
```

No olvide abrir y cerrar el archivo después de procesarlo con las funciones **open** y **close**. Compare los distintos tipos y contenidos de la variable **data** como resultado de cada función.

2. Lea el primer renglón del mismo archivo y asigne el contenido a dos variables, una con nombre **mes** que contenga el mes y la otra **vis** con el número de visitantes.
3. Procese todo el archivo con un bucle (loop) y guarde el contenido en dos listas con el mismo nombre de las variables del ejercicio anterior.
4. Grafique con matplotlib el número de visitantes en función del mes. Calcule el número de visitantes anuales y el promedio de visitantes mensuales.
5. Escriba en un nuevo archivo el número de visitantes anuales y el promedio de visitantes mensuales en dos líneas diferentes.
6. Use la función **genfromtxt** o **loadtxt** del módulo **numpy** para procesar el archivo **msd.xvg** que se encuentra en el directorio **Data/Guia**. Este archivo contiene la desviación cuadrática media de un partícula (CO<sub>2</sub> en agua) en  $nm^2$  en función del tiempo de simulación en *ps*. Grafique estas dos variables y realice una regresión lineal para calcular el coeficiente de difusión que equivale a un sexto de la pendiente de la regresión. (El valor obtenido por otro programa lo encuentra como un comentario al inicio del archivo **msd.xvg**)

Finalmente, vamos a trabajar con una base de datos del banco central. Bajo el siguiente link <http://si3.bcentral.cl/Siete/secure/cuadros/arboles.aspx> podrá encontrar el valor del dolar observado en el año 2016.

7. Gaurde los valores del dolar con formato excel
8. Abra el archivo descargado con Microsoft excel y gaurde el archivo en formato csv (en caso de no tener Microsoft excel puede usar el archivo que se encuentra bajo el directorio Data/Guia.)
9. Procese el archivo con la función `genfromtxt` para obenter el valor promedio del dolar observado en el año 2016. (Para eliminar los valores ausentes de los días feriados puede utilizar la función `isnan()` de numpy.)
10. **Bonus Track (opcional):** Busque en internet bajo el link <http://www.datos.gob.cl> una base de datos con las precipitaciones en Chile del primer semestre del año 2015. Descargue el archivo con formato `.csv`. Procese el archivo con el módulo `pandas` y analice los datos para responder a la siguientes preguntas: a. Cuál estación registro la mayor precipitación diaria y en que fecha ocurrió esto? b. Cuál fue la precipitación acumulada en la estación de Pichoy, Valdivia durante el primer semestre de este año? (Para esta última pregunta se recomienda usar la función `loc` en `pandas`)