



ISIS-1221 INTRODUCCIÓN A LA PROGRAMACIÓN

Proyecto de Nivel 4

Boletín Estadístico Uniandes

Objetivo general

El objetivo general de este proyecto es que usted practique todos los conceptos estudiados en el nivel 4 del curso. Recuerde que este proyecto debe realizarse de forma **completamente individual**.

Objetivos específicos

1. Ejercitar la implementación de algoritmos para construir y recorrer matrices.
2. Fomentar la habilidad de descomponer un problema en subproblemas y de implementar funciones que los resuelven.
3. Acercarse a las librerías de Pandas y Matplotlib.

Descripción de la aplicación

Anualmente la Universidad de los Andes publica un informe estadístico bajo el nombre “Boletín Estadístico Uniandes”. Este boletín, como su nombre lo indica, contiene una serie de reportes con información estadística de estudiantes, profesores, facultades, centros de apoyo, etc. El boletín del 2019 se encuentra disponible para consulta pública, en el siguiente enlace:

<https://planeacion.uniandes.edu.co/pdi/boletin-estadistico/boletin-estadistico>

Para el desarrollo del proyecto de nivel 4 hemos extraído información de este boletín con el propósito de permitirle hacer algunas consultas de interés y conocer así mejor su Universidad.

La información que hemos recopilado es la siguiente:

- Datos generales de las facultades: cantidad de profesores, cantidad de estudiantes, promedio de créditos inscritos por los estudiantes en un semestre, promedio general acumulado de los estudiantes, número de títulos otorgados y cantidad de grupos de investigación. Estos datos se encuentran en el archivo adjunto “estadisticas_facultades.csv”.
- Atención de puestos estudiantes entre facultades. Esto es, el número de puestos estudiante de una Facultad, atendidos por otra Facultad. Esta información se encuentra en el archivo adjunto “matriz_puestos.csv”.
- Cantidad de estudiantes que cursan doble programa y programas en los cuales están inscritos. Esta información se encuentra en el archivo adjunto “matriz_dobles.csv”.

Trabajando con este conjunto de datos usted puede convertirse en un verdadero conocedor de la Universidad.

iii Le deseamos la mejor de las suertes en esta exploración universitaria!!!

Parte 1: Cargar y entender los datos

La primera opción de la aplicación debe permitir que se carguen los datos de los tres archivos csv a tres matrices diferentes. En el esqueleto del módulo adjunto a este enunciado, se encuentran implementadas las funciones que cargan 2 de las 3 matrices a utilizar. Usted solo debe desarrollar la función que carga la tercera matriz. **La función que implemente esta opción debe recibir como parámetro el nombre del archivo y debe retornar una matriz (lista de listas).** A continuación, hay una explicación más detallada del contenido de las 3 matrices con las que se va a trabajar en este proyecto.

Matriz de estadísticas de facultades:

La matriz de estadísticas contiene datos generales de cada facultad de la universidad. La primera columna (columna 0) tiene los nombres de cada una de las facultades de la universidad, mientras que la primera fila (fila 0), tiene los nombres (o títulos) de los datos contenidos en cada columna. Estos son, en orden: **Profesores hombres, Profesores mujeres, Estudiantes hombres, Estudiantes mujeres, Promedio de créditos, PGA (promedio general acumulado) promedio, Títulos de pregrado, Títulos de maestría y Grupos de investigación.** Cada posición en la matriz representa el valor de dicho atributo para dicha facultad. Por ejemplo, Economía está en la fila 8, la casilla en la posición (8,5) representa el promedio de créditos inscritos por los estudiantes de la Facultad de Economía. Note también que cada fila completa de la matriz contiene el total de la información de una misma facultad. Al momento de cargarse, la matriz tiene 12 filas y 10 columnas.

Matriz de puestos estudiante:

Un “puesto estudiante” es un cupo en un curso dictado por una facultad y tomado por un estudiante. Cada facultad ofrece (o atiende) un número determinado de puestos que son tomados por diferentes facultades según las necesidades académicas de los estudiantes. En esta matriz, la fila 0 y la columna 0 contienen los nombres de las diferentes facultades que tiene la universidad. Cada posición en esta matriz representa el número de cupos atendidos por la facultad que se encuentra en la fila y que son tomados por estudiantes de la facultad que se encuentra en la columna. Por ejemplo, Ciencias se encuentra en la fila 4 e Ingeniería se encuentra en la fila 10, el número contenido en la casilla de la posición (4,10) indica el número de puestos de estudiantes que son atendidos por la Facultad de Ciencias y que son tomados por estudiantes de la Facultad de Ingeniería. Esta matriz es una matriz de 12 filas x 13 columnas. Esto se debe a que estudiantes de Estudios Dirigidos toman cursos en otras facultades, pero Estudios Dirigidos no tiene cursos propios para ofrecer a otras facultades. La columna de Estudios Dirigidos es la columna 13, de modo que usted puede estar seguro de que las posiciones de la diagonal ((1,1), (2,2), etc.) representan en todos los casos los puestos de estudiantes ocupados por la misma facultad que los ofrece.

Matriz de dobles programas:

A diferencia de las dos matrices anteriores, esta matriz está expresada en programas académicos en lugar de facultades. En la columna 0 y en la fila 0 se encuentran los nombres de casi todos los programas académicos ofrecidos por la universidad, con la excepción de las licenciaturas que están condensadas en una sola entrada llamada “Licenciatura”. Cada posición en esta matriz representa el número de estudiantes que están haciendo dicho doble programa. Por ejemplo, Psicología se encuentra en la fila 21, y Medicina en la fila 35, el número en la posición (21,35) indica el número de estudiantes que estudian Psicología (como programa principal) y están haciendo doble programa con Medicina. A su vez, la posición (35,21) representa aquellos que están matriculados en Medicina como programa principal y su doble programa es Psicología. Esta matriz es cuadrada.

Parte 2: Desarrollar las funciones

Luego de haber completado la función para cargar la matriz de puestos estudiante y de entender bien los formatos en los cuales están representadas las tres matrices, usted deberá completar las siguientes funciones, que responden a consultas sobre los datos.

NOTA IMPORTANTE: Usted puede implementar más funciones de ser necesario, así como reutilizar funciones dentro de otras funciones. Recuerde que una buena práctica de programación es la de dividir y conquistar. Esto es, dividir un problema en subproblemas más pequeños, desarrollar funciones para resolver estos subproblemas y componer funciones para resolver el problema principal.

Función 1 – Puestos estudiante ofrecidos (o atendidos) por una Facultad

Implemente una función que reciba como parámetro la matriz de puestos estudiante y una facultad de interés y retorne un valor entero con la cantidad de puestos ofrecidos por dicha facultad. Por ejemplo, la Facultad de Economía atiende un total de 4107 puestos estudiante.

Función 2 – Puestos estudiante ocupados por una Facultad

Implemente una función que reciba como parámetro la matriz de puestos estudiante y una facultad de interés y retorne un valor entero con la cantidad de puestos ocupados por estudiantes de dicha facultad. Por ejemplo, la Facultad de Educación ocupa un total de 185 puestos estudiante.

Función 3 – Facultad más servicial

Implemente una función que reciba como parámetro la matriz de puestos estudiante, y retorne una tupla con la información de la facultad más servicial. La facultad más servicial es aquella que atiende un mayor número de puestos estudiante de otras facultades, comparativamente con sus propios puestos estudiante atendidos. Por ejemplo: la facultad de medicina atiende un total de 4444 puestos estudiante. De estos puestos, 4318 son de estudiantes de medicina y tan solo 126 son de otras facultades. Por consiguiente, la relación puestos de otras facultades/puestos propios es: $126/4444 = 0.028$. Esto quiere decir que solo el 2.83% de los puestos estudiante de la Facultad de Medicina son tomados por estudiantes de otras facultades. Es fácil deducir que Medicina no es la facultad más servicial, según esta definición. La primera posición de la tupla que retorna esta función debe ser el nombre de la facultad más servicial y la segunda posición de la tupla, el porcentaje de puestos estudiante de las demás facultades que son atendidos por dicha facultad (redondeado a dos decimales).

Función 4 – Existe facultad generosa

Implemente una función que reciba como parámetro la matriz de puestos estudiante, el nombre de una facultad de interés y un porcentaje (valor real de 0 a 100) de puestos estudiante, e indique si en la Universidad existe alguna facultad que atienda un porcentaje de puestos estudiante de la facultad de interés, que sea igual o superior al ingresado por parámetro. La función debe retornar una tupla cuyo primer valor sea el nombre de la facultad encontrada, y el segundo valor sea el porcentaje de puestos estudiante cubiertos por dicha facultad (redondeado a 2 decimales). En caso de que ninguna facultad atienda más de este porcentaje de puestos estudiante, la función debe retornar la tupla (“No existe facultad generosa”, 0). En caso de que varias facultades sobrepasen este porcentaje, la función debe retornar la primera que encuentre.

Por ejemplo: supongamos que la facultad de interés es Medicina y el porcentaje recibido por parámetro es 5. Sabemos que la Facultad de Medicina ocupa un total de 6350 puestos estudiante. El 5 por ciento de este valor es 317. Esta función debe entonces buscar en la matriz de puestos, una facultad que atienda al menos 317 puestos estudiante de medicina. La Facultad de Ciencias cumple con esta condición, ya que atiende 1179 puestos estudiante de Medicina, lo cual equivale al 18.57% del total de puestos estudiante de Medicina. La Facultad de Ciencias Sociales cumple igualmente con esta condición, ya que atiende 477 puestos estudiante de Medicina, lo cual equivale al 7.51% del total de puestos estudiante de Medicina. Esta función debe entonces retornar alguna de estas dos facultades, la que encuentre primero, junto con el porcentaje de atención correspondiente.

Función 5 – Porcentaje de autocubrimiento

Implemente una función que reciba como parámetro la matriz de puestos estudiante y la matriz de estadísticas de facultades y calcule el porcentaje de autocubrimiento de todas las facultades. El porcentaje de autocubrimiento de una facultad X es definido de la siguiente manera:

$$\text{autocubrimiento de } X = \frac{\text{puestos estudiante ocupados por } X}{\text{puestos estudiante ofrecidos por } X}$$

Después de calcular esta información para cada facultad de la universidad, usted debe agregar una nueva columna a la matriz de información estadística de facultades, en la cual guardará el valor calculado para cada una de las facultades (redondeados a 2 decimales). Puede poner el título que desee a esta nueva columna. Si ya existía la columna no importa, de todas formas se agrega. Esta función retorna la matriz de estadísticas modificada.

Función 6 – Doble programa más popular

Implemente una función que reciba como parámetro la matriz de dobles programas, e indique cuál es el doble programa más popular entre los estudiantes. Esta función debe retornar una tupla, cuya primera posición es una cadena de caracteres de la forma “Programa1 – Programa2” y la segunda posición es el número de estudiantes que cursan este doble programa. Para este requerimiento, es equivalente estar inscrito en el Programa1 y tener un doble programa con el programa2, que estar inscrito en el Programa2 y hacer doble con el Programa1. Por ejemplo: hay 15 estudiantes inscritos en Ingeniería Mecánica y que están cursando Diseño como segundo programa. Por otra parte, hay 1 estudiante de Diseño, cursando Ingeniería Mecánica como segundo programa. Por consiguiente el doble programa “Ing. Mecanica – Disenio” cuenta con 16 estudiantes.

Función 7 – Dobles programas con un programa de interés

Implemente una función que reciba como parámetro la matriz de dobles programas y el nombre de un programa de interés, y retorne un diccionario, cuyas llaves sean los nombres de todos los programas diferentes al programa de interés y los valores de cada llave, sean el número de personas que realizan dicho doble programa (siempre y cuando este número sea mayor a cero). Para este caso es equivalente estar inscrito en el Programa1 y hacer un doble programa con el programa2, que estar inscrito en el Programa2 y hacer doble programa con el Programa1. Por ejemplo: si el programa de interés es Geociencias, esta función debe retornar un diccionario con las siguientes parejas llave-valor:

```
{'Administracion': 1, 'Musica': 1, 'Biologia': 5, 'Fisica': 12, 'Quimica': 1, 'Antropologia': 1, 'Ciencia Politica': 1, 'Historia': 1, 'Lenguas y Cultura': 1, 'Derecho': 1, 'Gobierno y A. Publicos': 2, 'Ing. Ambiental': 13, 'Ing. Civil': 36, 'Ing. Electronica': 1, 'Ing. Industrial': 2, 'Ing. Mecanica': 5, 'Ing. Quimica': 2}
```

Función 8 – Estadísticas del PGA

Implemente una función que reciba como parámetro la matriz de estadísticas, y retorne una lista con 3 tuplas, que corresponden respectivamente a la facultad con mayor PGA promedio, a la facultad con menor PGA promedio y a la facultad con el PGA promedio igual a la mediana de los PGA de todas las facultades. Cada tupla tiene 2 posiciones, en la primera posición debe ir el nombre de la facultad y en la segunda posición su PGA promedio.

Función 9 – Existe facultad con un determinado porcentaje de estudiantes de un género dado

Implemente una función que reciba como parámetro la matriz de estadísticas, una cadena que represente un género (usted puede decidir cómo deben ser estas cadenas cuando las pida al usuario, ya sea “m”/“f”, “mujer”/“hombre”, “masculino”/“femenino”, etc.) y un porcentaje (valor real entre 0 y 100), e indique si existe alguna facultad, que tenga un porcentaje de estudiantes del género dado por parámetro superior al porcentaje dado.

La función debe retornar una tupla de 3 posiciones, donde la primera posición sea True o False, según se haya encontrado o no una facultad con dicha característica, la segunda posición el nombre de la facultad, o la cadena vacía en caso de no encontrar ninguna, y la tercera posición el porcentaje que supera al ingresado por parámetro, o 0 en caso de que no se haya encontrado ninguna facultad. En caso de que varias facultades sobrepasen este porcentaje, la función debe retornar la primera que encuentre.

Ejemplo:

Supongamos que los parámetros de entrada son: “femenino” (para el género) y 60 (para el porcentaje). En este caso estamos preguntando al programa si existe al menos una facultad que tenga más del 60% de estudiantes mujeres. La función debe entonces calcular, para cada facultad, la cantidad total de estudiantes (teniendo en cuenta tanto hombres como mujeres) y el porcentaje de estudiantes mujeres con respecto a este total. Posteriormente, debe buscar si existe alguna facultad, cuyo porcentaje de estudiantes mujeres sea superior al 60%. La Facultad de Arquitectura y Diseño, así como la Facultad de Ciencias Sociales cumplen con esta condición. La primera con un 64.04% de estudiantes mujeres y la segunda con un 65.68%. Esta función debe entonces retornar alguna de estas dos facultades, la que encuentre primero, junto con True y el porcentaje de mujeres correspondiente.

Actividad 0: Preparación del ambiente de trabajo

1. Cree una carpeta para trabajar, poniéndole su nombre o login.
2. Descargue de BrightSpace el archivo con el “esqueleto” del proyecto (n4-boletin-esqueleto.zip) y descomprímalo en su carpeta de trabajo. El esqueleto consiste en un conjunto de archivos que usted va a usar o a modificar.
3. Abra Spyder y cambie la carpeta de trabajo para que sea la carpeta donde descargó el archivo con los datos.

Actividad 1: Construir el módulo de funciones

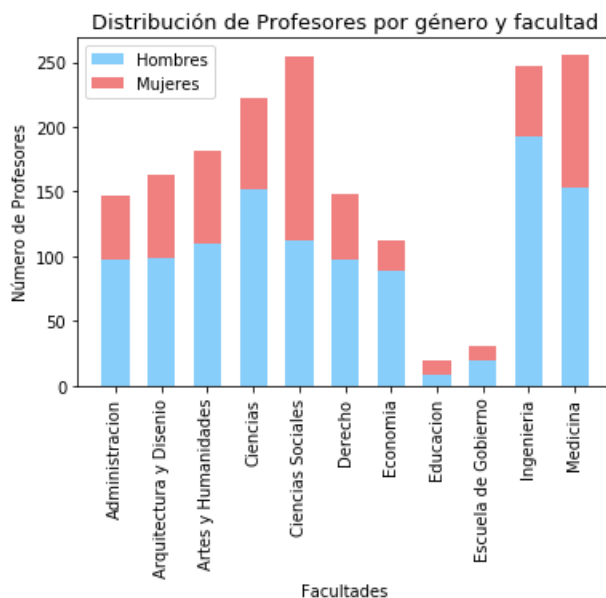
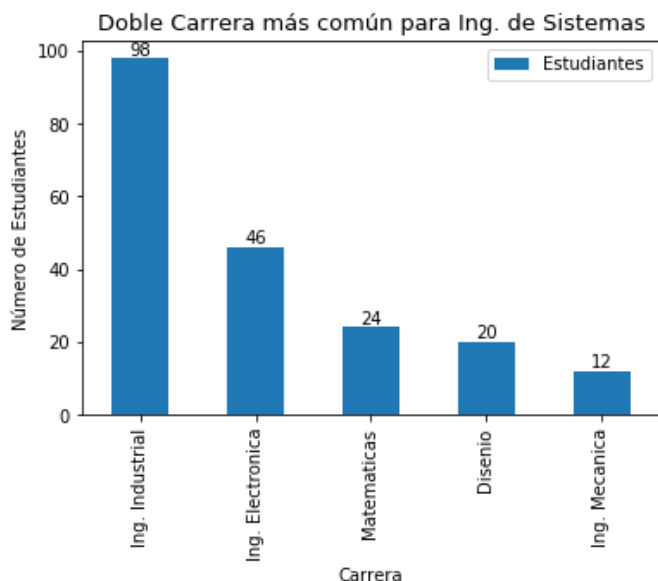
Usando Spyder, complete el archivo de nombre “boletin.py”. Este archivo ya cuenta con funciones para cargar 2 de las 3 matrices que se deben trabajar. En este archivo usted va a construir el módulo en el que va a implementar las funciones que responden a los requerimientos de la aplicación. **Defina, documente e implemente** las funciones en este archivo. Usted puede crear cuántas funciones considere necesarias dentro de su módulo. Mínimo debe haber una función por cada una de las acciones que debe realizar el programa.

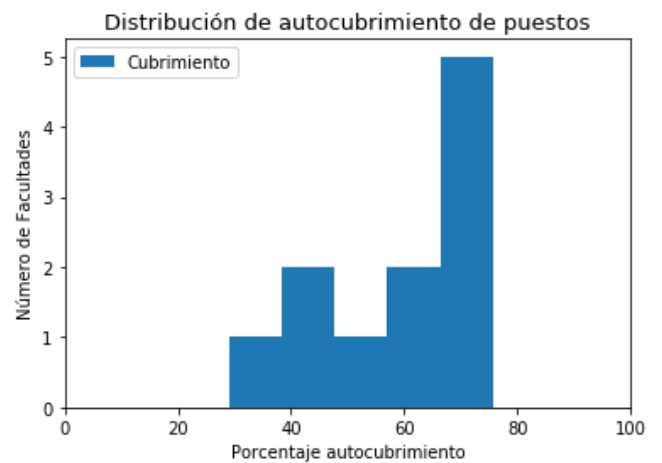
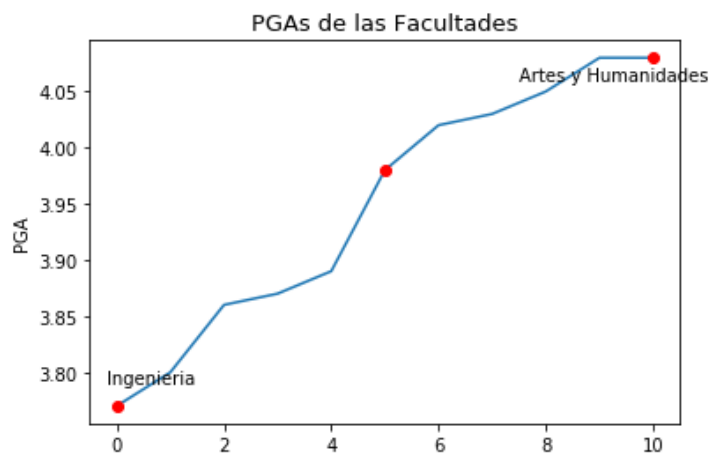
Actividad 2: Completar la interfaz de usuario basada en consola

1. En esta actividad usted tiene que completar la interfaz basada en consola para que el usuario interactúe con la aplicación. Para construir esta interfaz usted debe completar el archivo “consola_boletin.py”, el cual ya tiene una parte implementada que le facilitará su trabajo. Usted debe modificar los elementos marcados con la etiqueta TODO.
2. Pruebe la interfaz por consola ejecutando el archivo “consola_boletin.py”. Verifique que las funcionalidades de su aplicación se comporten de acuerdo con lo esperado. Puede probar el correcto funcionamiento de su programa cargando la información que se encuentra en los archivos “matriz_dobles.csv”, “estadisticas_facultades.csv” y “matriz_puestos.csv”. Utilizando estas 3 matrices, haga uso de las funciones que ya desarrolló y verifique que funcionen sin problema alguno.

Actividad 3: Uso de Pandas y Matplotlib

Adicional a las pruebas que usted debe realizar usando su consola, hemos puesto a su disposición un pequeño programa desarrollado en Python y llamado “graficas_boletin.py”, el cual hemos incluido en el esqueleto del proyecto. Este programa le permitirá probar 3 funciones de su proyecto (y otras más) para ver gráficamente los resultados haciendo uso de las librerías Pandas y Matplotlib. Para probar sus propias funciones, busque los “TODO” en este programa y rempácelos por los llamados correspondientes a sus funciones. Cada función en este programa viene con una breve documentación que explica lo que hace. Al usar la información de las matrices de ejemplo, las gráficas que usted podrá generar quedarán almacenadas en la **carpeta** llamada “graficas” y son las siguientes:





Entrega

1. Comprima los dos archivos: boletin.py y consola_boletin.py en un solo archivo .zip. El archivo comprimido debe llamarse **"N4-PROY-login.zip"**, donde login es su nombre de usuario de Uniandes.
2. Entregue el archivo comprimido a través de BrightSpace en la actividad designada como **"Proyecto N4"**.