

Laboratorio 1

Desarrollo de una herramienta analítica usando recursos nativos de Python

Para el desarrollo de esta actividad debes utilizar únicamente recursos nativos de Python (puedes usar librerías como `math` y `random`). Está prohibido el uso de librerías externas, a excepción de `matplotlib`. Te recomendamos leer por completo el enunciado del laboratorio antes de comenzar, de forma que tengas claro el propósito global de la actividad y puedas desarrollar tu solución apuntando a él desde el inicio.

En algunas de las misiones encontrarás: una **versión base**, la cual indica el requerimiento mínimo de la misión; una **versión completa**, la cual indica mejoras deseables o complementos a la **versión base**. Un desarrollo aceptable del laboratorio debe satisfacer la **versión base** y un desarrollo correcto del laboratorio debe satisfacer la **versión completa**. Cumplir con el desarrollo de la **versión completa** desde un inicio hará más sencillas las misiones siguientes.

Al desarrollar este laboratorio pondrás a prueba tus habilidades para:

1. Identificar y abordar preguntas de negocio y de *analytics*.
2. Leer datos desde archivos y almacenarlos en estructuras de Python.
3. Crear y llamar funciones.
4. Implementar condicionales diferenciando `if`, `elif` y `else`.
5. Implementar ciclos diferenciando el propósito de `for` y `while`.

Contexto: exploración del potencial para asignar becas Unialpes

La Universidad de los Alpes se ha caracterizado por su compromiso con la generación de impacto en la sociedad a través de diferentes programas de becas. Programas como Becas para los Mejores, entre otros, han sido revolucionarios al permitir que estudiantes de bajos recursos puedan acceder a una de las universidades más prestigiosas del país.

Actualmente, la universidad quiere evaluar el potencial impacto que tendrían algunos programas de acción afirmativa para asignación de becas, con base en información demográfica que ha obtenido a través de encuestas a personas interesadas en becas. Se

cuenta con información respecto a la edad, género, estado civil, estrato socioeconómico, región de proveniencia, desempeño académico y nivel de escolaridad de las personas encuestadas. Los encargados de diseñar el programa saben que utilizar criterios basados en este tipo de variables es un tema sensible, pero es necesario si se quieren focalizar las oportunidades en poblaciones menos representadas.

El objetivo de este laboratorio es crear una herramienta computacional que permita a un tomador de decisiones explorar qué efecto podría tener el uso de diferentes reglas para seleccionar a los beneficiarios de las becas. Puntualmente, se espera que, para distintas reglas de selección (ej., pertenecer a cierta región, rango de edad, etc.), la herramienta ofrezca una caracterización de los potenciales beneficiarios de las becas (a partir de estadísticas descriptivas y visualizaciones), de forma que exista un sustento para elegir la regla de selección adecuada.

Fase 1: obtener e inspeccionar archivos

Los datos de encuestas fueron obtenidos como parte de un trabajo de campo y, por limitaciones tecnológicas, solo se cuenta con la información en archivos de texto plano (.txt). Cada archivo corresponde a uno de los atributos antes mencionados y están estructurados de forma que todos relacionan a las personas encuestadas en el mismo orden (i.e., en cada archivo la primera línea corresponde a un dato de la primera persona encuestada, la segunda línea corresponde a un dato de la segunda persona encuestada, etc.).

Los archivos con los que cuentas son:

- "edad.txt": un archivo con la edad de cada persona encuestada;
- "escolaridad.txt": un archivo con el nivel de escolaridad de cada persona encuestada;
- "estado_civil.txt": un archivo con el estado civil de cada persona encuestada;
- "estrato.txt": un archivo con el estrato socioeconómico de cada persona encuestada;
- "genero.txt": un archivo con el género de cada persona encuestada;
- "promedio.txt": un archivo con el promedio académico de cada persona encuestada;
- "region.txt": un archivo con la región de proveniencia de cada persona encuestada.

Misión 1

Versión base: utiliza métodos de lectura de archivos para guardar la información de cada archivo en una estructura de datos de Python, de manera que se respete el orden y la categoría de los datos.

Versión completa: escribe una función que sea capaz de ejecutar la **versión base** y llámala dentro de un ciclo para automatizar la lectura de los archivos.

Misión 2

Versión base: calcula estadísticas descriptivas básicas (por lo menos, media/mediana y varianza/desviación) para cada variable relevante en el contexto del problema y repórtalas concluyendo acerca de las principales características de los datos.

Versión completa: complementa la **versión base** con visualizaciones informativas sobre los datos (ej., histogramas, gráficos de dispersión, boxplots, etc.), relevantes para el problema.

Fase 2: abordar la pregunta de negocio

El comité encargado de diseñar el programa de becas ha tenido dificultades para llegar a acuerdos. Algunos piensan que considerar las características sociodemográficas (ej., género, región, etc.) es parte del espíritu del programa, mientras que otros piensan que el criterio de selección debería centrarse estrictamente en lo económico y académico. Dado esto, el comité ha sugerido evaluar dos propuestas, suponiendo un número NN de becas disponibles.

- La primera propuesta sugiere que se deben tomar las diferentes regiones y géneros y, para cada una de las combinaciones de estos dos criterios, asignar una cantidad de becas proporcional al número de personas de esa subpoblación con respecto al total de personas encuestadas. Dentro de cada subpoblación, las becas deben aprobarse a las personas con mejor promedio académico.
- La segunda propuesta sugiere que se deben tomar en orden ascendente los estratos y, en cada uno de ellos, asignar una beca a cada una de las personas que se encuentren dentro del 2% con mejor promedio académico en esa subpoblación. En caso de quedar becas disponibles luego de haber cubierto todos los estratos, se repite el mismo esquema hasta asignar todas las becas disponibles, excluyendo a las personas a quienes se les haya asignado una beca en alguna ronda anterior.

Misión 3

Define la pregunta de negocio y la pregunta de *analytics* para esta situación.

Misión 4

Versión base: desarrolla un algoritmo que seleccione a los potenciales beneficiarios de las becas, según la primera propuesta.

Versión completa: escribe una función que sea capaz de ejecutar la **versión base** para distintos números NN de becas disponibles.

Misión 5

Versión base: desarrolla un algoritmo que seleccione a los potenciales beneficiarios de las becas, según la segunda propuesta.

Versión completa: escribe una función que sea capaz de ejecutar la **versión base** para distintos números NN de becas disponibles.

Misión 6

Versión base: con base en tu desarrollo de las misiones 2, 4 y 5, implementa un procedimiento que le permita al comité comparar las dos propuestas, exponiendo métricas y visualizaciones de las principales características demográficas de los potenciales beneficiarios de las becas que se obtendrían con cada propuesta.

Versión completa: escribe una función que sea capaz de ejecutar la **versión base** de manera automática para ambas propuestas.

Fase 3: desarrollar una solución analítica

Reconociendo la complejidad de este problema y viendo el potencial del análisis de datos para enriquecer la discusión, el comité requiere una herramienta analítica que permita definir y evaluar nuevas propuestas de asignación de becas. Se debe poder:

1. Ingresar el número de becas disponibles (ej., `N=0`).
2. Ingresar un rango de edad para los candidatos factibles (ej., `min_edad=99`, `max_edad=100`).
3. Elegir cualesquiera de las siguientes tres características: género, estrato, región (ej., `caracteristicas=["genero", "estrato"]`).
4. Ingresar interactivamente (durante la ejecución de la herramienta) el porcentaje del total de las becas que se asignará a las personas con mejor promedio académico dentro de cada una de las combinaciones de las características elegidas. Por ejemplo, si se eligen las características "género" y "estrato", deberá asignarse un porcentaje a cada combinación (ej., asignar 3% de las becas al género "masculino" del estrato 1, asignar 5% de las becas al género "femenino" del estrato 1, etc.).

La herramienta debe permitir al comité configurar una nueva propuesta y seleccionar al grupo de potenciales beneficiarios de las becas de acuerdo con la propuesta configurada. Además, la herramienta debe exponer métricas y visualizaciones de las principales características demográficas de los potenciales beneficiarios de las becas.

Misión 7

Desarrolla un algoritmo que reciba los parámetros para configurar la nueva propuesta y seleccione a los potenciales beneficiarios de las becas.

Misión 8

Con base en tu desarrollo de la misión 7, implementa un algoritmo que: le permita al comité configurar una nueva propuesta; exponga métricas y visualizaciones de las principales características demográficas del grupo de potenciales beneficiarios para la propuesta configurada.

Misión 9

Utiliza tu herramienta para configurar propuestas a tu criterio. Con sustento en las métricas y visualizaciones que obtengas, elige una de las propuestas que configuraste. Compara y concluye acerca de las virtudes y falencias de tu propuesta con respecto a las de las dos propuestas del comité.

Referencias

- Discriminación Positiva
https://es.wikipedia.org/wiki/Discriminaci%C3%B3n_positiva
- Análisis de formulación del programa "Ser Pilo Paga"
<https://repository.urosario.edu.co/handle/10336/12526>
- La acción afirmativa en la Universidad de los Andes: El caso del programa "Oportunidades para talentos nacionales"
http://www.scielo.org.co/scielo.php?script=sci_arttext&pid=S0123-885X2004000300005
- Acción afirmativa y equidad: un análisis desde la propuesta de Thomas Nagel
http://www.scielo.org.mx/scielo.php?script=sci_arttext&pid=S2007-24062016000200049

Créditos

Autores: Camilo Hernando Gómez Castro, Alejandro Mantilla Redondo, Diego Alejandro Cely Gómez.

Fecha última actualización: 11/07/2022