

Contenido

Aplicantes Universitarios Dataset	2
New York University.....	2
Descripción	2
Imagen identificativa.....	2
Contexto	2
Contenido	2
Agradecimientos.....	3
Inspiración	3
Licencia.....	4
Código fuente y dataset	4
Recursos.....	4

Aplicantes Universitarios Dataset

New York University

Descripción

El conjunto de datos generado como parte de esta actividad práctica reúne diferentes características de prospectos que han aplicado a la universidad de Nueva York entre los años 2012 y 2018. Algunas de las variables que se recogen en el conjunto de datos son el género, estado, tipo de escuela, GPA, SAT, si es atleta.

Imagen identificativa



Contexto

El conjunto de datos contiene información calificaciones y decisiones de admisión de estudiantes que han aplicado a la universidad de Nueva York, se puede identificar a los estudiantes admitidos, rechazados o en lista de espera desde durante los años 2012 a 2018.

Contenido

Para cada accidente, el cual se corresponde con un registro en el conjunto de datos, se recogen las siguientes características:

- **Profile**, identificador del aplicante
- **Gender**, genero del aplicante
- **ACT**, promedio del resultado de los exámenes del aplicante

- **Status**, estado de la aplicación
- **EA/ED**, indica si la aplicación es exclusiva a la universidad
- **Legacy**, indica si algún familiar asistió anteriormente
- **Athlete**, indica si es atleta o no
- **Type**, tipo de secundaria
- **State**, estado de la secundaria
- **UW**, promedio del estudiante sin tomar en cuenta dificultad de las materias
- **W**, promedio del estudiante tomando en cuenta dificultad de las materias
- **M**, resultado de la prueba SAT en matemáticas
- **CR**, posición en su clase
- **W**, resultado de la prueba SAT en escritura

El sitio de College data <https://www.collegedata.com> es miembro de la Asociación nacional para el asesoramiento de admisión a la universidad en los Estados Unidos (NACAC por sus siglas en ingles).

NACAC se dedica a brindarles a los profesionales de consejería y admisión herramientas que necesitan para prestar un mejor servicio a los estudiantes y mejorar el proceso de transición a la universidad.

NACAC obtiene incorpora información del Sistema de Datos de Educación Postsecundaria Integrado (IPEDS por sus siglas en ingles) del Departamento de Educación de EE. UU.

Agradecimientos

Los datos han sido recolectados desde la página [Collegedata](https://www.collegedata.com). Para ello, se ha hecho uso del lenguaje de programación Python y de técnicas de *Web Scraping* para extraer la información alojada en las páginas HTML. Para el preprocesado de datos se utilizo R.

Inspiración

El presente conjunto de datos podría utilizarse para realizar estadísticas y estudios del ingreso a las universidades en los Estados Unidos. Así como en el campo de la *minería de datos*, a la hora de elaborar modelos de clasificación. Así, se podría querer elaborar un modelo que permita clasificar quien va a ser seleccionado o no para la universidad en la que aplico.

Licencia

La licencia escogida para la publicación de este conjunto de datos ha sido **CC BY-SA 4.0 License**. Los motivos que han llevado a la elección de esta licencia tienen que ver con la idoneidad de las cláusulas que esta presenta en relación con el trabajo realizado:

- *Se debe proveer el nombre del creador del conjunto de datos generado, indicando los cambios que se han realizado.* De esta manera, se reconoce el trabajo ajeno y en qué medida se han realizado aportaciones en relación con el trabajo original.
- *Se permite un uso comercial.* Esto haría que incrementen las probabilidades de que una empresa utilice los datos generados y realicen trabajos de calidad que reporten cierto reconocimiento al autor original.
- *Las contribuciones realizadas a posteriori sobre el trabajo publicado bajo esta licencia deberán distribuirse bajo la misma.* Esto hace que el trabajo del autor original continúe distribuyéndose bajo los términos que él mismo planteó.

Código fuente y dataset

Tanto el código fuente escrito para la extracción de datos como el dataset generado pueden ser accedidos a través de [este enlace](#).

Recursos

1. Lawson, R. (2015). Web Scraping with Python. Packt Publishing Ltd. Chapter 2. Scraping the Data
2. Simon Munzert, Christian Rubba, Peter Meißner, Dominic Nyhuis. (2015). Automated Data Collection with R: A Practical Guide to Web Scraping and Text Mining. John Wiley & Sons.