



Informe Prueba Auxiliar Analista de Datos Uniautónoma del Cauca.

Nombre: John Eduard Delgado Gómez

CC: 1061820435

Celular: 3148068955

1. Introducción.

El siguiente documento presenta los resultados obtenidos del análisis al conjunto de datos de estudiantes de Uniautónoma. Se plantea una limpieza de datos, seguido de un análisis exploratorio de Datos (EDA), seguido de una creación de Dashboard que ayude al seguimiento de indicadores referentes a los estudiantes de Uniautónoma. Por último, se recopila los resultados en el presente informe.

2. Fuente de datos.

El dataset estudiantes.csv está conformado inicialmente por 8 columnas:

- **ID_Estudiante:** Identificador del estudiante.
- **Facultad:** Nombre de la facultad a la que pertenece el estudiante. Las posibles facultades fueron Economía, Artes, Salud, Ciencias Sociales, Humanidades y Ingeniería.
- **Facultad:** Nombre del programa al que pertenece el estudiante. Se cuenta con 17 programas diferentes.
- **Año_Ingreso:** Representa el año en el que ingreso el estudiante a la universidad.
- **Estado:** Estado actual del estudiante representado en las categorías de Activo, Egresado y Retirado.
- **Promedio:** Promedio obtenido por el estudiante para esta prueba se dio un límite entre 0 – 10.
- **Asistencia:** Porcentaje de asistencia del estudiante. Tiene como límite 0 – 100.
- **Indicadores_Adicionales:** Cadena de texto donde se especifica si un estudiante tiene beca y además si realiza actividades extracurriculares.

3. Preprocesamiento

Se llevaron a cabo las siguientes actividades para la limpieza y transformación de los datos:

- Eliminación de espacios al inicio y final de cada palabra por precaución.
- Corrección de valores en la variable Año de ingreso cuando se presentaban dos fechas simultáneamente (Ej: "2017 - 2020"), seleccionando el año menor.
- Corrección de errores tipográficos en la variable Estado.
- División de la columna Indicadores Adicionales en dos nuevas columnas: Beca y Actividad extracurricular, para mejorar la interpretación de los datos.
- Manejo de datos faltantes.

En cuanto al tratamiento de valores ausentes:

- Para Promedio y Asistencia, se eliminaron directamente las filas con datos faltantes.
- Para Año de ingreso, se completaron los valores faltantes con 0.

- Para Estado, se reemplazaron los datos ausentes por "No registrado".
- Para Beca y Actividad extracurricular, se imputaron los valores faltantes utilizando la moda.
- Para Facultad, se determinó el valor faltante a partir del programa académico asignado al estudiante.

Tras aplicar este proceso, se logró conservar el **80.35%** de los datos originales.

4. Principales hallazgos:

- La facultad de ciencias sociales tiene la mayor cantidad de estudiantes.
- La facultad con el menor número de estudiantes es la facultad de ingeniería.
- El programa de Contaduría tiene mayor cantidad de estudiantes, seguido de Sociología.
- El año 2014 fue el peor año respecto a cantidad de estudiante. Sin embargo, se observan 78 estudiantes de los cuales no se tiene registro, una suposición inicial puede ser que parte de estos 78 estudiantes podrían estar distribuidos en los años 2021 y 2014 para alcanzar un valor más cercano al promedio.

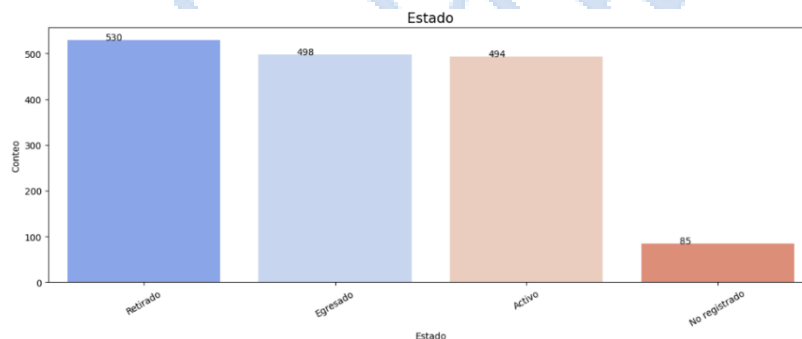


Figura 1. Número de estudiantes por estado.

- No se cuenta con 85 registros, pero algo preocupante es ver que la cantidad de estudiantes Retirados sea mayor respecto a Egresados o Activos. Aunque los datos faltantes se distribuyeran en egresado o activo, el número de estudiantes no debería ser tan alto para retirados. Una recomendación es realizar un posterior estudio para conocer a que se debe el posible retiro y actuar para reducir estos números.
- No se encuentran indicios de alguna posible correlación entre el promedio académico y la asistencia al tomar datos de toda la facultad, por tanto, no se entra en detalle a calcular correlación o p-value.

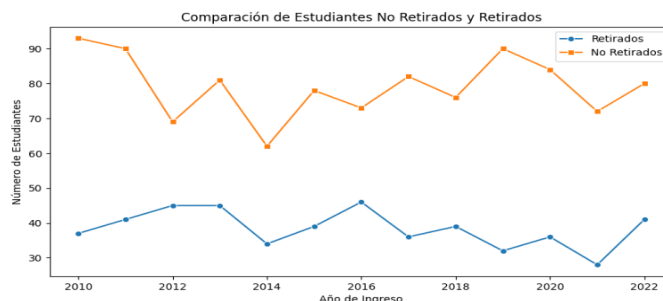
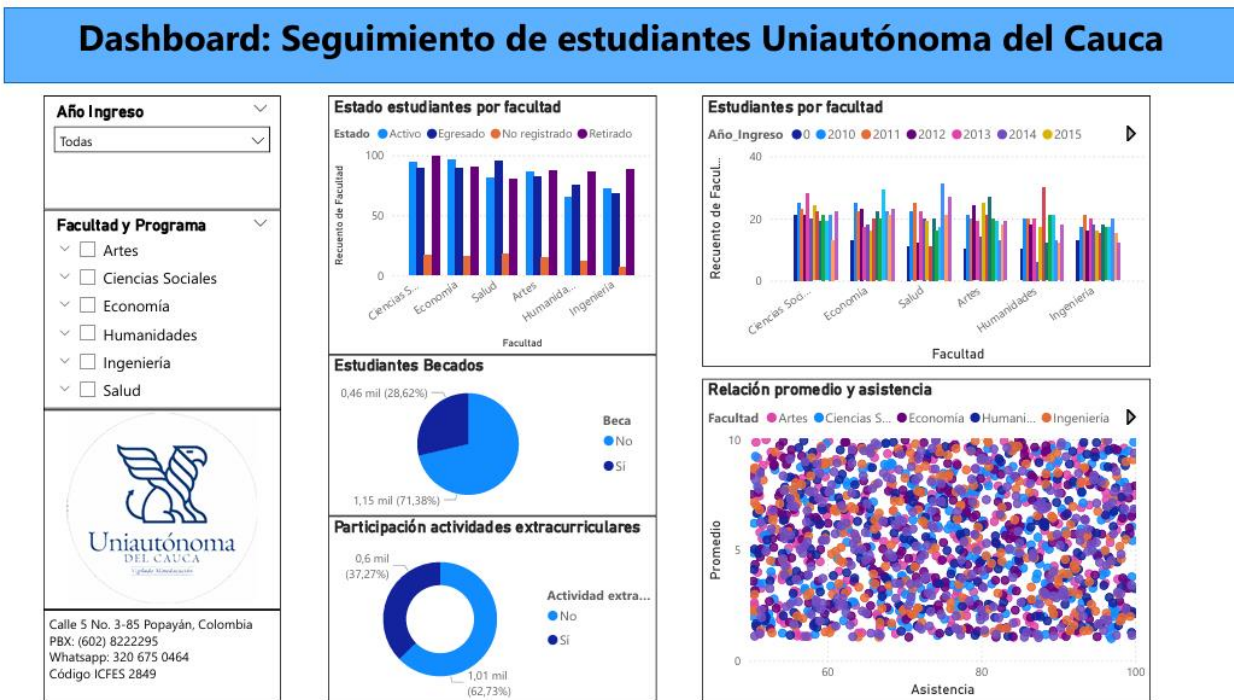


Figura 1. Número de estudiantes no retirados y retirados.

- Si consideramos a los estudiantes activos y egresados como un solo conjunto, se encuentra que de cada año alrededor de un 20 - 25% de los estudiantes se retiran.



5. Dashboard: Seguimiento de estudiantes Uniautónoma del Cauca.



Se presenta un dashboard interactivo conformado por 5 graficas principales que ayudaran a conocer el numero de estudiantes por facultad, el estado, el índice de becados, participación en actividades extracurriculares y un grafico de dispersión para la visualización de alguna posible relación entre el promedio y la asistencia.

El dashboard se complementa con 2 segmentadores de datos uno para seleccionar el año de ingreso del estudiante y otro para seleccionar las facultades o los programas de interés.

6. Conclusiones, limitantes y Recomendaciones.

Se trabajo con un conjunto de datos de estudiantes de 2000 filas conservando posterior a la limpieza alrededor del 80% de los datos. Se extrajo insights de importancia complementándose con un dashboard interactivo para realizar el seguimiento de los actuales y futuros estudiantes. Para este análisis se tiene como limitantes la cantidad de datos por programa que limitan encontrar una posible relación entre el promedio y la asistencia. Como recomendación final, se propone incentivar las actividades extracurriculares y reducir la tasa de retiros en especial en el programa de Diseño.