

Práctica Final:

Parte I - Selección del conjunto de datos



Visualización de Datos

Daniel Felipe Gómez Aristizabal
dfgomez@uoc.edu

Índice

Introducción	2
1. Justificación breve	2
2. Relevancia del conjunto de datos	3
3. Complejidad del conjunto de datos	3
4. Originalidad del análisis	9
5. Preguntas de investigación	10
Preguntas de investigación principales:	10
6. Diccionario de variables	11
Ingresos mensuales de estudiantes por fuente	11
Uso de ahorros para estudios o manutención	12
Grado de dificultad financiera percibida	12
Bibliografía	14

Introducción

Hace cinco años, me encontraba terminando bachillerato y me enfrentaba a la gran pregunta: ¿Qué hago ahora? ¿Qué rumbo académico y profesional debo tomar?. Una decisión compleja que no solo determinaría mi futuro sino que también comprometía el aspecto económico. Ir a la universidad me obligaba a irme de casa de mis padres y asumir una serie de gastos económicos que no sabía si serían demasiado elevados para mi familia. Temía comenzar una carrera universitaria que finalmente no pudiera completar por motivos financieros.

Sin embargo, gracias a las subvenciones y becas que se ofrecen para respaldar el estudio, así como, al apoyo incondicional de mis padres, hoy puedo decir con orgullo que soy Graduado en Ingeniería Informática.

Esta experiencia personal despertó en mí el interés por comprender y analizar cómo las condiciones económicas, sociales y emocionales afectan las trayectorias académicas de los estudiantes en Europa. Sentando así las bases motivacionales para seleccionar la temática del proyecto que se desarrolla a continuación.

Tras investigar en diversas fuentes, he encontrado un proyecto llamado EUROSTUDENT. Se trata de una iniciativa a nivel europeo que busca recopilar datos sobre diversos factores que afectan a los estudiantes de educación superior. Dado que su objetivo es facilitar comparaciones entre países, el proyecto cuenta con una sólida red de socios nacionales en cada país participante, lo que permite detectar fortalezas y debilidades en las distintas situaciones académicas.

EUROSTUDENT ha publicado diversos estudios basados en los datos que recopila, además de ofrecer acceso a su base de datos con los resultados de las encuestas realizadas. Por tanto, tanto las bases de datos como los estudios ya presentados pueden resultar muy útiles para el desarrollo del proyecto de la asignatura, ya que permiten comparar y contrastar la información y las conclusiones que obtengamos.

1. Justificación breve

Como comentaba previamente, he seleccionado el proyecto EUROSTUDENT como fuente de datos, debido al interés, tanto personal como académico, que me genera comprender las diferencias y similitudes que existen en las condiciones de vida y de estudio de los estudiantes universitarios europeos. Mi intención es determinar cómo el background de los alumnos puede condicionar su experiencia universitaria y si existen determinados patrones de algunos países que influyan en la educación.

Con el objetivo de comprender mejor el entorno estudiantil, se analizarán factores que inciden directamente en el rendimiento académico, el bienestar personal y la igualdad de

oportunidades. Para ello, nos beneficiaremos de la gran variedad de variables que presenta, que permiten estudiar aspectos relevantes como los costes de vida, las fuentes de ingresos, la movilidad académica y el acceso a ayudas. Esta información proporciona una visión rica y detallada de la realidad estudiantil en diferentes países.

2. Relevancia del conjunto de datos

La educación superior es una preocupación internacional actual, especialmente por su relación con la igualdad de oportunidades, la equidad de género y el bienestar estudiantil. Comprender las condiciones de vida y estudio del alumnado europeo permite identificar desigualdades y patrones que influyen en su experiencia académica, aportando claves para desarrollar políticas más óptimas e inclusivas.

En este contexto, veo que los datos recogidos en EUROSTUDENT se presentan como una fuente de información privilegiada, al recoger datos desglosados por variables como género, nivel socioeconómico, tipo de institución y país.

Estos datos permiten abordar preguntas relevantes sobre desigualdad de género en el acceso a recursos, diferencias en las cargas laborales o los efectos de las condiciones de vida en el abandono académico. Así, el análisis de este conjunto de datos se alinea directamente con los debates contemporáneos en materia de educación y equidad social.

3. Complejidad del conjunto de datos

Por fortuna, contamos con una gran variedad de datos para llevar a cabo nuestro estudio. Desde la página web donde se pueden descargar, disponemos de un sistema de filtros iniciales que nos permite, por ejemplo, seleccionar temáticas, grupos, países a incluir y la ronda del estudio (actualmente disponibles: rondas 5, 6, 7 y 8). Estos filtros nos permiten obtener tablas muy simples que contienen únicamente las variables de interés.

Por ejemplo, a continuación se muestra un caso en el que se selecciona la temática “Age Groups”, en la ronda ocho, con todos los grupos incluidos y una selección de países determinados. Como se puede observar, la información proporcionada es clara y sencilla de interpretar.

Country	All students											
	< 22 years			22 - 24 years			25 - 29 years			30 years and older		
	Value	Unit	Count	Value	Unit	Count	Value	Unit	Count	Value	Unit	Count
AT	20.9	%	9,050	28.8	%	12,466	29.1	%	12,594	21.3	%	9,205
AZ	77.0	%	1,929	15.2	%	381	4.9	%	122	2.9	%	73
CH (D)	17.1	%	3,913	37.2	%	8,525	32.1	%	7,344	13.6	%	3,121
CZ	35.7	%	5,287	36.8	%	5,447	16.1	%	2,376	11.4	%	1,689
DE	23.9	%	1,527	29.5	%	1,883	28.1	%	1,796	18.5	%	1,185
DK (D)	13.8	%	1,945	39.4	%	5,545	33.6	%	4,722	13.1	%	1,847
EE	27.9	%	1,029	27.1	%	996	16.0	%	590	29.0	%	1,067
ES	45.5	%	4,126	23.6	%	2,139	11.2	%	1,015	19.6	%	1,779

Por fortuna, la plataforma también ofrece la posibilidad de descargar los datos sin agrupar. En concreto, se ponen a disposición 17 conjuntos de datos separados por temática. En general, todos presentan una estructura muy similar, que analizaremos más adelante, donde se muestran los resultados de los encuestados organizados en diversas agrupaciones.

Lo que varía entre un dataset y otro es, principalmente, el conjunto de preguntas formuladas. De este modo, podremos alternar entre los distintos conjuntos de datos según las variables que deseemos analizar o incluir en nuestro estudio.

Dataset	Resumen
E8 topic A Students characteristics.xlsx	Contiene información sobre las características personales del estudiantado, como edad, género, nacionalidad y situación familiar.
E8 topic B Socioeconomic background.xlsx	Presenta datos sobre el contexto socioeconómico de los estudiantes, incluyendo nivel educativo de los padres, ingresos familiares y situación económica general.
E8 topic C Transition and a	Incluye información relacionada con el acceso a la

ccess.xlsx	educación superior y la transición desde la educación secundaria, como los motivos de elección de estudios y barreras percibidas.
E8 topic D Types and modes of study.xlsx	Describe los tipos de estudios realizados y las modalidades de cursado, como tiempo completo o parcial, presencial, virtual o híbrido.
E8 topic E Housing situation.xlsx	Contiene datos sobre la situación habitacional del estudiantado, como tipo de vivienda, convivencia y distancia respecto al centro de estudios.
E8 topic F Students expenses.xlsx	Proporciona información sobre los gastos habituales de los estudiantes, incluyendo alojamiento, alimentación, transporte y materiales de estudio.
E8 topic G Students resources.xlsx	Recoge datos sobre los recursos económicos y ayudas con los que cuentan los estudiantes, como becas, ayudas familiares o empleo.
E8 topic H Employment.xlsx	Muestra información sobre la situación laboral de los estudiantes, incluyendo tipo de empleo, horas trabajadas y conciliación con los estudios.
E8 topic I Time budget.xlsx	Detalla cómo distribuyen su tiempo los estudiantes entre estudio, trabajo, ocio y otras actividades cotidianas.
E8 topic J International student mobility.xlsx	Abarca datos sobre la movilidad internacional, como estancias en el extranjero, programas de intercambio y motivaciones.
E8 topic K Assessment of studies and performance.xlsx	Incluye percepciones del estudiantado sobre la calidad de sus estudios y su propio rendimiento académico.

E8_topic_L_Counselling_Services.xlsx	Ofrece información sobre el uso y valoración de los servicios de orientación y apoyo psicológico o académico.
E8_topic_M_Health_wellbeing.xlsx	Presenta datos sobre la salud física y mental del estudiantado, incluyendo hábitos, niveles de estrés y bienestar general.
E8_topic_N_Effects_of_the_COVID-19_Pandemic.xlsx	Analiza los efectos de la pandemia de COVID-19 en la vida académica y personal de los estudiantes.
E8_topic_O_Equipment_at_home.xlsx	Contiene información sobre el equipamiento disponible en el hogar, como acceso a internet, dispositivos electrónicos y condiciones para estudiar.
E8_topic_P_Digitalisation.xlsx	Aborda el uso de tecnologías digitales en el proceso de aprendizaje, incluyendo herramientas, competencias digitales y metodologías en línea.
E8_topic_Q_Discrimination_safety_feeling.xlsx	Recoge datos sobre experiencias de discriminación, acoso o percepción de seguridad dentro y fuera del ámbito educativo.
Importante: Cada uno de los nombres contiene un enlace a la descarga de los datasets.	

A continuación, y con el fin de comprender mejor cómo funciona el dataset, se presenta un breve recorrido sobre su estructura. Como ejemplo, tomaremos el dataset G_Student_resources, el cual contiene información sobre las fuentes de ingresos de los estudiantes encuestados.

Al abrir el archivo en formato Excel, nos encontramos con una tabla de contenidos desde la cual es posible visualizar las preguntas incluidas en la encuesta. Además, esta tabla permite navegar fácilmente hacia las pestañas correspondientes donde se encuentran registradas las respuestas.

+ TOC income_all_total income_all_total_N income_all_sources_5 income_all_sources_5_N income_all_sources_macrocalc																
	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P
1																
2																
3																
4																
5																
6																
7																
8																
9																
10																
11																
12																
13																
14																
15																
16																
17																
18																
19																
20																
21																
22																
23																
24																
25																
26																
27																
28																
29																
30																
31																
32																
33																
34																
35																
36																
37																
38																
39																
40																
41																
42																
43																
44																
45																
46																
47																
48																
49																
50																
51																
52																
53																
54																
55																
56																
57																
58																
59																
60																
61																
62																
63																
64																
65																
66																
67																
68																
69																
70																
71																
72																
73																
74																
75																
76																
77																
78																
79																
80																
81																
82																
83																
84																
85																
86																
87																
88																
89																
90																
91																
92																
93																
94																
95																
96																
97																
98																
99																
100																

Por ejemplo, si navegamos por la pregunta “Composition of students’ total monthly income”, veremos que contamos con columnas que indican el grupo al que pertenecen los resultados de esa fila, como por ejemplo “All students”. Luego, en la columna indicator item, se especifica qué valor concreto representa esa respuesta. En este caso, podríamos encontrar, por ejemplo, las contribuciones familiares, el apoyo estudiantil no nacional, las ganancias personales, las ayudas nacionales o los ingresos de otras fuentes.

A continuación, observamos que hay una columna por país, en la cual se presentan, en este caso, las medias de los ingresos correspondientes a cada una de las secciones mencionadas anteriormente. Lo interesante de esto es que, según el tipo de pregunta, los valores pueden expresarse en porcentajes o en valores absolutos, y en general en distintos formatos que resultan útiles para el análisis.

Como se puede apreciar en la siguiente imagen, en cada pestaña se presentan posteriormente los mismos resultados, pero agrupados según otras variables, como la edad, el sexo, el nivel educativo previo, el tipo de educación, entre otros.

income Composition of students' total monthly income																
Description:		Monthly income including transfers in kind / Arithm. mean														
	focusgroup	focusgroup item	indicator item	unit	AT	AZ	CL	DE	DK	EE	ES	FI	FR	GE		

Evidentemente, el tipo de preguntas y las posibles respuestas varían según el tema. En este caso, observamos que se consultaban los ingresos provenientes de diversas fuentes, por lo que, de cierta forma, los encuestados podían asignar un valor a cada una de ellas. Sin embargo, también encontramos otros ejemplos en los que se utiliza una escala del 1 al 5 para determinar el grado de acuerdo o desacuerdo frente a determinadas afirmaciones. Por ejemplo, ante la pregunta “It was always clear I would study one day”, los resultados se presentan tanto en porcentaje como en valores absolutos.

Uno de los retos que identifico a primera vista es que estos datos no son del todo amigables para ser procesados mediante un lenguaje de programación, que sería mi opción preferida. Por lo tanto, idealmente extraeré los datos que me interesen de cada estudio y los almacenaré en un archivo CSV separado, donde sea más sencillo cargarlos en un dataframe y manipularlos adecuadamente.

Otro aspecto al que debo prestar atención es que no todos los conjuntos de datos comparten la misma estructura, por lo que tendré que ajustar la carga de cada uno de manera personalizada, dependiendo de los que decida utilizar.

Finalmente, me gustaría destacar que se trata de una muestra que incluye a miles de personas provenientes de diversas regiones de Europa, con una participación de 9.059 personas en el caso de España. En total, se entrevistaron a 295.150 personas.

Country	All students											
	Female			Male			Other			Prefer not to assign myself to the categories offered		
	Value	Unit	Count	Value	Unit	Count	Value	Unit	Count	Value	Unit	Count
AT (D)	54.7	%	23,696	43.7	%	18,942	0.9	%	407	0.6	%	270
AZ	50.9	%	1,274	49.1	%	1,229	0.0	%	0	0.1	%	2
CH (D)	53.0	%	12,145	47.0	%	10,758	n. a.	%	n. a.	n. a.	%	n. a.
CZ (D)	56.3	%	8,332	42.6	%	6,303	0.0	%	0	1.1	%	163
DE (D)	49.7	%	3,176	49.4	%	3,153	0.6	%	37	0.4	%	24
DK (D)	57.6	%	8,099	42.4	%	5,961	0.0	%	0	0.0	%	0
EE	60.4	%	2,224	38.4	%	1,413	0.0	%	0	1.2	%	45
ES	55.6	%	5,042	41.8	%	3,797	0.6	%	55	2.0	%	178
FI	57.5	%	3,883	41.5	%	2,803	1.0	%	65	0.0	%	0
FR	55.4	%	24,608	43.7	%	19,432	0.9	%	412	0.0	%	0

4. Originalidad del análisis

Desde mi perspectiva, el proyecto resulta especialmente interesante, ya que aborda un tema tan cotidiano y relevante como es la educación, sobre el cual no siempre se reflexiona con la profundidad que merece. Considero que con el enfoque se le espera dar a este estudio es de gran valor y puede ofrecer conclusiones muy llamativas, que inviten a reflexionar incluso sobre cómo mejorar y potenciar la educación superior.

Además, considero que comparar los resultados a nivel europeo puede ser de gran valor para dimensionar cuál es la situación española en relación con otros países, y evaluar con mayor precisión la situación de los estudiantes universitarios y el funcionamiento de distintos aspectos del sistema educativo superior.

Sumado a estos aspectos, considero que se podrá aportar aún más contexto y relevancia si en determinados análisis también se tiene en cuenta el PIB per cápita y el gasto público en educación, para ello nos apoyaremos en fuentes como Eurostat del que obtendremos datos como el producto interno bruto de los diferentes países de la Unión Europea.

5. Preguntas de investigación

Antes que nada, quiero aclarar que las preguntas que presentaré a continuación, aunque han sido las que más me han llamado la atención tras analizar y reflexionar sobre los datos, no son rígidas. Considero que pueden variar o requerir la incorporación de nuevas preguntas complementarias, ya que, conforme avance en el trabajo con los datos, surgirán nuevas incógnitas que podrían tener mayor relevancia. No obstante, he procurado plasmar mis dudas o enfoque en las siguientes preguntas.

Preguntas de investigación principales:

1. ¿Cuál es el costo medio de vida de los estudiantes en cada país?
2. ¿Qué fuentes de ingreso son más comunes entre los estudiantes?
3. ¿Qué porcentaje de estudiantes trabaja durante sus estudios? ¿Existe un impacto sobre su rendimiento académico?
4. ¿Cuántas horas dedican los estudiantes al estudio y al trabajo?
5. ¿Qué tan satisfechos están los estudiantes con la calidad de la enseñanza recibida?

6. ¿Qué porcentaje de estudiantes expresa intención de abandonar los estudios?
7. ¿Cuál es el nivel de felicidad de los estudiantes según su país de residencia?
8. ¿Qué tipos de discriminación son más reportados entre los estudiantes?
9. ¿Qué relación existe entre el nivel de ingresos familiares y la posibilidad de participar en movilidad internacional?
10. ¿Cuál es el estado de salud general de los estudiantes?

6. Diccionario de variables

Como se comentaba en el apartado 3, acerca de la complejidad del conjunto de datos, trabajaremos con varios de los 17 datasets disponibles. Aunque tengo claro cuáles serán imprescindibles, aún no estoy seguro de cuáles no serán necesarios. Entre cada dataset se observan ligeras diferencias, principalmente en cuanto a las preguntas, pero en general, la estructura suele corresponder con la que se presentó en dicho apartado.

Teniendo esto en cuenta, debemos considerar que los conjuntos con los que estamos trabajando presentan diversas variables complejas, tratadas como datasets independientes, lo cual impide analizarlas de forma columnar tradicional, ya que son complejos en sí mismos. No obstante, presentaré algunas de las variables que analizaremos a lo largo de la investigación, extraídas directamente de los datasets.

En concreto tomaremos algunas de las secciones del dataset G_Student_resources.

Ingresos mensuales de estudiantes por fuente		
Variable	Descripción	Tipo de dato
self_earned_income	Ingresos medios obtenidos por el estudiante mediante empleo	Numérico
family_partner_contribution	Aporte económico de la familia o pareja (en dinero o especie)	Numérico

national_public_student_support	Ayuda pública nacional recibida	Numérico
non_country_sources_support	Apoyo financiero en efectivo desde fuentes extranjeras	Numérico
income_other_sources	Otros ingresos (herencias, premios, regalos, etc.)	Numérico

Uso de ahorros para estudios o manutención		
Variable	Descripción	Tipo de dato
savings_from_jobs	Porcentaje de estudiantes que usaron ahorros de trabajos anteriores	Numérico
savings_from_inheritance_or_gifts	Porcentaje que usó ahorros provenientes de herencias, regalos u otras fuentes	Numérico
no_savings_used	Porcentaje que no utilizó ahorros para vivir o estudiar	Numérico

Grado de dificultad financiera percibida		
Variable	Descripción	Tipo de dato
financial_difficulties_very_serious	Porcentaje que experimenta dificultades financieras muy serias	Numérico
financial_difficulties_2	Porcentaje con nivel de dificultad 2 (moderado)	Numérico
financial_difficulties_3	Porcentaje con nivel de dificultad 3	Numérico
financial_difficulties_4	Porcentaje con nivel de dificultad 4 (bajo impacto)	Numérico
financial_difficulties_none	Porcentaje sin dificultades financieras	Numérico

Bibliografía

EUROSTUDENT (s. f.) *EUROSTUDENT – Comparación de las condiciones de vida de los estudiantes en Europa*. Disponible en: <https://www.eurostudent.eu/> (Accedido: 2 de mayo de 2025).

EUROSTUDENT (s. f.) *EUROSTUDENT Database*. Disponible en: <https://database.eurostudent.eu/drm/> (Accedido: 2 de mayo de 2025).

Eurostat (s. f.) *PIB y componentes principales (nombres actuales)*. Disponible en: https://ec.europa.eu/eurostat/databrowser/view/nama_10_gdp/default/table?lang=en&category=na10.nama10.nama_10_ma (Accedido: 2 de mayo de 2025).