



UNIVERSITAS
Miguel Hernández

Automatización de informes estadísticos con Google Colab y Python

—

Una aplicación para las encuestas Erasmus+

Trabajo de Fin de Grado

Grado en Estadística Empresarial

Curso 2024-2025

Autor: Mario Goñi Arrivi

Tutora: Maria Asunción Martínez Mayoral

Índice de contenidos

[1. Resumen](#)

[Abstract](#)

[2. Palabras clave](#)

[3. Contexto](#)

[4. Objetivos](#)

[5. Información disponible](#)

[6. Desarrollo](#)

[6.1 Acceso y Procesado](#)

[Acceso y recopilación de las bases de datos](#)

[Corrección de errores tipográficos y duplicados](#)

[Unificación de nomenclaturas y categorías](#)

[Integración de bases de datos](#)

[6.2 Análisis estadístico](#)

[6.3 Elaboración de mapas](#)

[6.4 Generación del informe Word](#)

[6.5 Interfaz de usuario](#)

[Selección de intervalo temporal](#)

[Selección de titulaciones](#)

[Selección según duración de estancia](#)

[6.6 Software y hardware](#)

[7. Validación](#)

[8. Conclusiones](#)

[Referencias](#)

[Anexo](#)

1. Resumen

Este Trabajo de Fin de Grado aborda la problemática derivada del tratamiento anual y manual de los datos recogidos en las encuestas Erasmus+ gestionadas por la Oficina de Relaciones Internacionales (ORI) de la Universidad Miguel Hernández. La realización de informes anuales para describir las prácticas ERASMUS realizadas por los distintos colectivos de la universidad, y su nivel de satisfacción, es una tarea susceptible de ser automatizada para reducir los tiempos de análisis y agilizar su elaboración cada año. Estas encuestas, realizadas entre los años 2014 y 2020, presentan además, una notable heterogeneidad en cuanto a formato, estructura y nomenclatura, lo que complica en cierto modo la automatización, no dejando por ello de ser viable, por la utilidad que va a generar al Servicio de cara a la toma de decisiones estratégicas cada curso.

Como solución, se ha desarrollado una herramienta automatizada en entorno Google Colab con Python que permite procesar los datos, estandarizarlos, aplicar los análisis estadísticos descriptivos e inferenciales pertinentes en función de los objetivos y el tipo de información disponible, y generar informes completos en formato Word, con gráficos, mapas, tablas y comentarios interpretativos. Esta herramienta facilita la personalización de los informes según grado, año o país de destino, y automatiza un proceso que antes requería una considerable carga de trabajo manual cada año.

Los resultados de su aplicación, muestran que el sistema funciona correctamente, ahorrando tiempo, reduciendo errores y mejorando notablemente la presentación de la información respecto a lo que se venía haciendo hasta el momento. Se concluye que esta solución permite optimizar la gestión de los datos Erasmus y representa una mejora significativa en los flujos de trabajo de la ORI. Abre las puertas a la automatización de la generación de cualquier tipo de informe en cualquier tipo de servicio, incorporando por supuesto otro tipo de funcionalidades según las necesidades específicas de cada servicio y cada informe.

Abstract

This Final Degree Project addresses the issues arising from the annual and manual processing of data collected from the Erasmus+ surveys managed by the International Relations Office (ORI) of the Miguel Hernández University. The preparation of annual reports to describe the Erasmus practices carried out by different university groups, and their level of satisfaction, is a task that could be automated to reduce analysis time and streamline its completion each year. These surveys, conducted between 2014 and 2020, also present considerable heterogeneity in terms of format, structure, and nomenclature, which complicates automation to some extent. However, automation remains a viable option due to the usefulness it brings to the Service for strategic decision-making each academic year.

As a solution, an automated tool has been developed in the Google Colab environment using Python. This tool processes and standardizes the data, applies the appropriate descriptive and inferential statistical analyses depending on the objectives and type of information available, and generates complete Word reports including charts, maps, tables, and interpretative comments. The tool enables customization of the reports by degree, year, or destination country, automating a process that previously required a considerable amount of manual work each year.

The results of its implementation, as reported by the ORI, show that the system functions correctly, saving time, reducing errors, and significantly improving the presentation of information compared to previous practices. It is concluded that this solution optimizes the management of Erasmus data and represents a significant improvement in the ORI's workflows. It also opens the door to the automation of report generation in any type of service, potentially incorporating additional functionalities tailored to the specific needs of each service and report.

2. Palabras clave

Automatización de informes estadísticos, Google Colab, Python, Visualización de datos.

3. Contexto

La oficina de Relaciones Internacionales de la Universidad Miguel Hernández recibe todos los cursos académicos una base de datos con las respuestas a la encuesta Erasmus+, dirigida tanto a los alumnos como al personal de la Universidad que hayan realizado movilidad Erasmus en el curso correspondiente. Esta encuesta es una herramienta fundamental para recopilar información sobre la experiencia y la satisfacción de los participantes en el programa Erasmus+, que promueve la movilidad y cooperación entre universidades europeas y de otros países asociados.

Las bases de datos con la información recopilada en las encuestas están en formato de hojas de cálculo de Excel y contienen las respuestas a una amplia variedad de preguntas incluidas en la encuesta. Estas preguntas cubren aspectos tan diversos como la percepción de la beca, la duración de la estancia, el nivel de idioma alcanzado, el coste de vida en el país de destino, y la valoración general de la experiencia Erasmus. Es importante señalar que, aunque el modelo general de la encuesta se mantiene constante para todos los años, la estructura presenta ligeras modificaciones anuales. Además, la formulación de preguntas similares puede variar en función del año, lo que genera diferencias en la manera en que se registran y organizan los datos.

Debido a la falta de homogeneidad en las encuestas, la tarea de combinar todas las bases de datos en un único conjunto coherente se vuelve lenta y compleja. La integración manual de estos datos requiere un esfuerzo considerable para ajustar las diferencias en estructura y nomenclatura, así como para depurar la información redundante o inconsistente. Este proceso manual no solo consume tiempo, sino que también incrementa la probabilidad de cometer errores que pueden afectar la fiabilidad de los resultados finales.

Dado el interés de la oficina de Relaciones Internacionales por conocer en profundidad la opinión de los estudiantes que han realizado la movilidad Erasmus y observar la evolución de estas opiniones a lo largo de los años, es necesario, cada vez que llegan los informes, trabajar con las bases de datos, realizar tareas de depuración y manipular los datos para extraer conclusiones relevantes. De hecho, la elaboración de los informes anuales, globales y de

evolución, es fundamental para tomar decisiones informadas que puedan mejorar la experiencia de futuros estudiantes y optimizar los recursos destinados al programa.

Mediante una oferta de prácticas realizada por la ORI, se me brindó la oportunidad de trabajar directamente sobre esta cuestión de la automatización de informes con las bases de datos de la encuesta Erasmus entre los años de 2014 a 2021.

Esta experiencia me permitió conocer de primera mano los retos asociados a la gestión de grandes volúmenes de información heterogénea y a la necesidad de desarrollar soluciones eficientes para su análisis.

La propuesta inicial consistía en elaborar informes detallados, segmentados por titulaciones, facultades y también de forma general para todos los grados. El objetivo era obtener conclusiones específicas que pudieran revelar posibles diferencias en las movilidades dependiendo de factores como el país de destino, la percepción de la beca, la duración de la estancia, el nivel adquirido de idioma, el coste de vida y la experiencia global de los participantes. Estos informes tenían un gran valor para el seguimiento y mejora continua del programa Erasmus en la universidad.

La principal dificultad residía en la necesidad de generar múltiples informes de forma individualizada, teniendo en cuenta variables como el año, el grado académico o el tiempo de duración de la movilidad (por ejemplo, un semestre o más de un semestre). Esto implicaba adaptar el código y los procesos de análisis para cada caso particular, aumentando considerablemente el tiempo requerido y la posibilidad de cometer errores durante la elaboración de los informes. Esta complejidad limitaba la rapidez y eficiencia con la que se podía disponer de resultados fiables.

Para abordar estos desafíos, se planteó desarrollar, a través de esta práctica y, mediante el uso de Cuadernos Jupyter en Google Colab con Python, un proyecto de Automatización de Informes.

La intención era diseñar un interfaz único y flexible que permitiera al usuario, en este caso la ORI, seleccionar en tiempo real los datos que fueran de su interés y obtener un informe específico adaptado a esas necesidades. Esta solución buscaba simplificar el proceso, reducir el margen de error y mejorar la accesibilidad de la información para la toma de decisiones.

El uso de herramientas como Jupyter y Google Colab no solo facilita la ejecución de código en entornos colaborativos y accesibles desde cualquier lugar, sino que también permite una integración directa con diferentes formatos de datos y una fácil visualización de resultados. Esto hace que la propuesta de automatización no solo sea efectiva desde el punto de vista técnico, sino también práctica y adaptable a futuros requerimientos. Además, al estar integrado en la Universidad Miguel Hernández la Google Suite, la accesibilidad a los cuadernos Colab es inmediata para todo el personal, haciendo aún más extensivo su uso.

En resumen, el contexto de este trabajo se centra en la necesidad real de gestionar y analizar grandes volúmenes de datos heterogéneos relacionados con la movilidad Erasmus, y en la oportunidad de mejorar este proceso a través de la automatización, optimizando así el tiempo y recursos de la ORI, y facilitando la obtención de conclusiones útiles para la mejora continua del programa Erasmus.

4. Objetivos

El objetivo principal de este proyecto es el diseño y desarrollo de una herramienta avanzada de automatización de informes estadísticos para las encuestas Erasmus que gestiona anualmente la Oficina de Relaciones Internacionales (ORI) de la Universidad Miguel Hernández de Elche.

Esta herramienta no solo se limitará a analizar los datos actualmente disponibles, sino que se diseñará con la perspectiva de poder adaptarse y evolucionar conforme se incorporen nuevas bases de datos en el futuro.

Objetivos Específicos

Se plantean como objetivos específicos del presente trabajo:

1. Desarrollar un sistema de automatización, bajo entorno Google Colab con Python, para la generación de informes estadísticos anuales en formato Word, bajo parámetros estándar de usabilidad, y procesados en la nube.
2. Automatizar el acceso a los resultados de las encuestas en hojas de cálculo de Google Drive, así como la limpieza, depuración y homogeneización de las preguntas y respuestas, para integrar toda la información, ya procesada, en una base de datos única y fácilmente migrable y accesible.
3. Resolver el análisis estadístico, descriptivo e inferencial, para dar respuesta a las necesidades e intereses informativos planteados por la ORI sobre las encuestas Erasmus. Integrar gráficos visualmente atractivos e informativos, aportando así mismo la interpretación y comentarios pertinentes para facilitar su interpretación y publicación a distintos colectivos.
4. En particular, crear e integrar mapas de distribución y satisfacción de usuarios Erasmus UMH en los diversos países europeos de destino.

5. Información disponible

Se han utilizado datos procedentes de las encuestas Erasmus de satisfacción realizadas por el estudiantado y personal de la UMH que formó parte del programa de movilidad de la Universidad entre los años 2014 y 2021, ambos incluidos. Estas encuestas son planteadas y lanzadas cada año por el programa Erasmus y son comunes a todas las universidades participantes.

Los datos fueron proporcionados en forma de archivos Excel, organizados en hojas de cálculo separadas por año académico. Cabe señalar que el fichero correspondiente al año 2020 incluía también las respuestas del año 2021. En el caso del estudiantado, cada archivo contenía aproximadamente 170 observaciones (una por participante), mientras que los archivos correspondientes al personal recogían una media de 20 observaciones por año.

Las respuestas de los encuestados que pertenecen a grados de **ESIC-UMH** no se incluyen en este estudio, por disponer de un sistema de garantía de calidad independiente.

Las encuestas cumplimentadas constan de distintos apartados que valoran la experiencia de movilidad desde múltiples perspectivas. Los ítems vinculados a preguntas de la encuesta que han sido objeto de estudio en este trabajo son los siguientes:

1. Datos Personales y de la Movilidad

id : Identificador único de la movilidad por persona

tipo : Variable categórica que identifica si la movilidad la realizó un estudiante o personal UMH

anyo: Variable numérica respecto al año de la movilidad (desde 2014 a 2021)

ini : Fecha de comienzo de la movilidad

fin : Fecha de finalización de la movilidad

pais : País de destino

tit : Titulación de la persona que realiza la movilidad

universidad: Universidad de destino (Datos disponibles a partir de 2016)

2. Motivaciones para realizar la movilidad

motivos : Variable categórica que representa los motivos para realizar la movilidad, en la que los encuestados podían seleccionar varias opciones simultáneamente

Posibles respuestas de la variable motivos:

- Conocer otro país
- Desarrollar aptitudes sociales
- Realizar prácticas en una lengua extranjera
- Crear redes personales y profesionales
- Experimentar diferentes contenidos educativos
- Aumentar las perspectivas laborales en mi país
- Realizar las prácticas en una lengua extranjera
- Vivir en el extranjero y conocer gente

3. Satisfacción general con la experiencia

Categóricas:

deseo_rep : Variable categórica respecto al deseo de repetir la movilidad.

Niveles de respuesta: Sí, No, Sin decidir.

satis : Variable categórica que representa la satisfacción general con la movilidad.

Niveles de respuesta: Muy insatisfecho/a, Bastante insatisfecho/a, Ni satisfecho/a ni insatisfecho/a, Bastante satisfecho/a, Muy satisfecho/a.

4. Calidad de los estudios en la universidad de destino

Categóricas con niveles de respuesta: Muy insatisfecho/a, Bastante insatisfecho/a, Ni satisfecho/a ni insatisfecho/a, Bastante satisfecho/a, Muy satisfecho/a.

calidad1_centrod : Calidad de las clases y cursos

calidad2_centrod : Calidad de los métodos de enseñanza

5. Factores de elección del centro de destino

Categóricas con niveles de respuesta: Importante, Poco importante, Muy importante, Nada importante, Sin opinión, Más importante.

selec_centro1: Selección del destino en base a reputación

selec_centro2 : Selección del destino en base a instalaciones y servicios

selec_centro3: Selección del destino en base a idioma

selec_centro4: Selección del destino en base al país

selec_centro5: Selección del destino en base a ciudad y cultura

selec_centro6: Selección del destino en base a la vida social

6. Conocimiento de lenguas extranjeras y apoyo lingüístico

Categorías:

clases_idioma: Principal idioma utilizado en las clases

Niveles de respuesta: Inglés, Italiano, Alemán, Portugués, Finlandés,
Lituanio, Francés, Holandés, Rumano, Eslovaco, Irlandés,
Estonio, Griego, Polaco.

mejora_idioma: Mejora del idioma tras la estancia.

Niveles de respuesta: Sí, No, 'No, ya tenía cierta fluidez'.

7. Desarrollo personal a través de la movilidad

Categorías con niveles de respuesta:

Bastante de acuerdo, Totalmente de acuerdo, Ni de acuerdo ni en desacuerdo,
Totalmente en desacuerdo, Bastante en desacuerdo.

mejora_personal: Mejora en la confianza y habilidades personales tras la movilidad

adapta_personal: Mejora en la adaptación ante nuevas situaciones tras la movilidad

resol_personal: Mejora en la capacidad resolutive tras la movilidad

8. Preparativos prácticos y organizativos

Categorías con niveles de respuesta: Muy insatisfecho/a, Bastante insatisfecho/a, Ni satisfecho/a ni insatisfecho/a, Bastante satisfecho/a, Muy satisfecho/a.

proceso_selec: Transparencia del proceso de selección para la movilidad

apoyoadm_envio: Apoyo administrativo en la organización de envío

tutor_envio: Apoyo de los tutores en la organización de envío

apoyoorg_envio: Apoyo de organizaciones/iniciativas estudiantiles en la organización de envío

apoyopers_envio: Ayuda con problemas personales y necesidades en la organización de envío

visa_envio: Ayuda para resolver problemas con el visado en la organización de envío

seguro_envio: Ayuda con cuestiones del seguro en la organización de envío

apoyoadm_dest: Apoyo administrativo en la organización de acogida

tutor_dest: Apoyo de los tutores en la organización de acogida

apoyoorg_dest: Apoyo de organizaciones/iniciativas estudiantiles en la organización de acogida'

apoyopers_dest: Ayuda con problemas personales y necesidades en la organización de acogida

visa_dest: Ayuda para resolver problemas con el visado en las organizaciones de acogida

seguro_dest: Ayuda con cuestiones del seguro en las organizaciones de destino

Nota:

En este apartado, la variable **proceso_selec** está disponible para todos los años y todos los colectivos.

El resto de variables solo están disponibles para el colectivo de estudiantes, y en el caso de las variables **apoyopers_envio** y **apoyopers_dest**, están disponibles a partir de 2016.

9. Satisfacción con las infraestructuras y servicios de la universidad de destino

Categorías con niveles de respuesta: Muy insatisfecho/a, Bastante insatisfecho/a, Ni satisfecho/a ni insatisfecho/a, Bastante satisfecho/a, Muy satisfecho/a.

calidad_alojamiento: Satisfacción con el alojamiento

restauracion_centrod: Calidad de las instalaciones y servicios

aulas_centrod: Calidad de las aulas

10. Coste medio de la movilidad

'gasto_medio': Variable numérica (en euros) que representa el gasto medio mensual de la persona que realiza la movilidad.

6. Metodología y desarrollo

En este apartado se describe de manera detallada la metodología seguida para alcanzar los objetivos propuestos en este trabajo. La metodología planteada busca no solo optimizar la gestión y procesamiento de datos, sino también garantizar la calidad, coherencia y profesionalidad en la presentación final de los informes.

Para ello, se ha diseñado un proceso estructurado que incluye desde el procesado de información (limpieza y homogeneización de los datos en las diferentes hojas de cálculo), fundamentales para asegurar la validez de los análisis posteriores, hasta la aplicación de técnicas básicas de análisis estadístico, elaboración de mapas, y la generación automatizada de documentos formateados. Se han integrado pues, destrezas en programación y estadísticas, con el fin de ofrecer una solución integral usable y fácilmente interpretable a los usuarios finales

Cada una de las fases que componen esta metodología ha sido cuidadosamente planificada para responder a las necesidades específicas planteadas desde un primer momento por la ORI y que derivaron en los objetivos propuestos en este proyecto. Se ha resuelto aprovechando herramientas de programación en Python, entornos de trabajo colaborativos y librerías especializadas. De esta forma, se garantiza la escalabilidad y flexibilidad de la herramienta, permitiendo adaptaciones futuras y facilitando la actualización periódica de los informes sin necesidad de intervenciones manuales extensas.

Describimos pues a continuación, el proceso seguido para la elaboración del proyecto, estructurado en las siguientes fases: 1) acceso a los datos y procesado de la información, 2) integración de las bases de datos, 3) análisis estadísticos, 4) elaboración de mapas, 5) automatización del informe, 6) interfaz de usuario, y finalizamos describiendo el software y hardware utilizado para el desarrollo del proyecto,

6.1 Acceso a los datos e integración de encuestas

En esta fase del proyecto, se abordaron una serie de problemas derivados de la gestión manual y anual de las bases de datos proporcionadas. Las respuestas a las encuestas disponibles, recogidas en ficheros Excel independientes por años, presentaban una amplia variabilidad tanto en la estructura como en el contenido, debido a las variaciones implementadas en las encuestas en años consecutivos, lo que dificultaba significativamente su tratamiento automatizado. Para dar respuesta a esta situación, se llevó a cabo un procedimiento sistematizado de chequeo, homogeneización y depuración de datos, para poder llevar a cabo finalmente la integración de toda la información en un único fichero de datos, incorporando como una variable más el año del que provenía. Este proceso se detalla a continuación.

Chequeo y unificación de preguntas de interés

El primer paso consistió en la localización, recopilación y verificación de todas las encuestas correspondientes a los distintos años del periodo analizado, en formato pdf. Este proceso requirió la confrontación de las encuestas con las hojas de cálculo proporcionadas por el servicio, para identificar diferencias entre los años disponibles e identificar las preguntas de interés. Se chequeó a continuación el formato de presentación de dichas preguntas de interés y las variaciones existentes entre los diferentes ficheros/años, para abordar las tareas de homogeneización pertinentes.

Uno de los principales retos fue la falta de uniformidad en la forma en que se formulaban las preguntas a lo largo de los años. Las mismas variables de interés podían estar asociadas a preguntas redactadas de manera distinta según el año o el colectivo encuestado (estudiantado o personal de la UMH), lo que impedía una asociación directa entre encabezados y variables clave para el análisis.

La revisión manual realizada permitió crear un sistema de correspondencias que vinculaba las diferentes formulaciones de una misma variable de interés. ,

El resultado del proceso de revisión se presenta en la Tabla 1, donde se muestran las variables de interés y las diferentes formulaciones encontradas.

Tabla 1. Variables disponibles y nomenclaturas alternativas encontradas en las diferentes encuestas realizadas entre los años 2014 y 2020.

| Variables | Nomenclaturas |
|--------------------|--|
| id | 'Identificador de la movilidad' , 'Mobility ID' |
| tipo | 'Tipo de movilidad' , 'Mobility Type' |
| ini | 'Fecha de inicio' , 'Start Date' |
| fin | 'Fecha de finalización' , 'End Date' |
| pais | 'País de acogida' , 'Receiving Country' |
| tit | 'TITULACIÓN' , 'Titulación' , 'TITULACION' |
| universidad | 'Nombre legal de la organización de acogida' |
| motivos | '¿Cuáles fueron los principales factores que le motivaron para estudiar en el extranjero?' '¿Cuáles fueron los principales motivos para realizar prácticas en el extranjero?' '¿Cuáles fueron los principales factores que le motivaron para realizar prácticas en el extranjero?' |
| deseo_rep | '¿Considera la posibilidad de realizar otra movilidad Erasmus+ en el futuro?' |
| satis | 'En general, ¿cómo está de satisfecho/a con su experiencia de movilidad Erasmus+?' |

| | |
|-------------------------|--|
| calidad1_centrod | <p>'¿Estuvo satisfecho/a con la calidad del aprendizaje y de la docencia que recibió en su institución de acogida?: Calidad de los cursos'</p> <p>'Calidad de los cursos'</p> |
| calidad2_centrod | <p>'¿Estuvo satisfecho/a con la calidad del aprendizaje y de la docencia que recibió en su institución de acogida?: Calidad de los métodos de enseñanza'</p> <p>'Calidad de los métodos de enseñanza'</p> |
| selec_centro1 | <p>'¿Qué criterios tuvo en cuenta para seleccionar su institución de acogida?: Reputación/rankings'</p> <p>'¿Qué criterios tuvo en cuenta para seleccionar su empresa/organización de acogida?: Reputación de la empresa',</p> <p>'¿Qué criterios tuvo en cuenta para seleccionar su institución de acogida?: Reputación/ránkings'</p> <p>'Reputación de la empresa',</p> <p>'Reputación/ránkings'</p> |
| selec_centro2 | <p>'¿Qué criterios tuvo en cuenta para seleccionar su institución de acogida?: Instalaciones y servicios de apoyo para estudiantes internacionales'</p> <p>'¿Qué criterios tuvo en cuenta para seleccionar su empresa/organización de acogida?: Instalaciones'</p> <p>'Instalaciones'</p> <p>'Instalaciones y servicios de apoyo para estudiantes internacionales'</p> |

| | |
|----------------------|--|
| selec_centro3 | <p>'¿Qué criterios tuvo en cuenta para seleccionar su institución de acogida?: Idioma'</p> <p>'Idioma'</p> <p>'Lengua'</p> <p>'¿Qué criterios tuvo en cuenta para seleccionar su empresa/organización de acogida?: Idioma'</p> |
| selec_centro4 | <p>'¿Qué criterios tuvo en cuenta para seleccionar su institución de acogida?: País'</p> <p>'País'</p> <p>'¿Qué criterios tuvo en cuenta para seleccionar su empresa/organización de acogida?: País'</p> |
| selec_centro5 | <p>'¿Qué criterios tuvo en cuenta para seleccionar su institución de acogida?: Ciudad y cultura'</p> <p>'¿Qué criterios tuvo en cuenta para seleccionar su empresa/organización de acogida?: Ciudad y cultura'</p> <p>'Ciudad y cultura'</p> |
| | <p>'¿Qué criterios tuvo en cuenta para seleccionar su institución de acogida?: Vida social'</p> <p>'¿Qué criterios tuvo en cuenta para seleccionar su empresa/organización de acogida?: Vida social'</p> <p>'Vida social'</p> |

| | |
|------------------------|--|
| selec_centro6 | |
| clases_idioma | <p>'¿Cuál fue el principal idioma utilizado por sus profesores?'</p> <p>'¿Cuál fue el principal idioma utilizado en su actividad de movilidad?'</p> <p>'¿Cuál fue el principal idioma utilizado en su actividad de prácticas?'</p> |
| mejora_idioma | <p>'¿Considera que ha mejorado sus competencias en esa lengua durante su estancia en el extranjero?'</p> <p>'¿Considera que ha mejorado sus competencias en ese idioma durante su estancia en el extranjero?'</p> |
| mejora_personal | <p>'Después de haber tomado parte en esta actividad de movilidad...: tengo más confianza en mi persona y en mis habilidades'</p> <p>'tengo más confianza en mi persona y en mis habilidades'</p> |
| adapta_personal | <p>'Después de haber tomado parte en esta actividad de movilidad...: me adapto y puedo actuar mejor ante situaciones nuevas'</p> <p>'me adapto y puedo actuar mejor ante situaciones nuevas'</p> |

| | |
|-----------------------|--|
| resol_personal | <p>'Debido a mi estancia en el extranjero con Erasmus+, he aprendido mejor a ...: encontrar soluciones en contextos difíciles o problemáticos (capacidad resolutive)'</p> <p>'encontrar soluciones en contextos difíciles o problemáticos (capacidad resolutive)'</p> |
| proceso_selec | <p>'¿El proceso de selección en su institución de origen fue justo y transparente?'</p> <p>'¿El proceso de selección en su institución de envío fue justo y transparente?'</p> |
| apoyoadm_envio | <p>'Estuvo satisfecho/a con el apoyo ofrecido por su institución de envío?: Apoyo administrativo'</p> <p>'Apoyo administrativo envio'</p> |
| tutor_envio | <p>'¿Estuvo satisfecho/a con el apoyo ofrecido por su institución de envío?: Tutorización académica'</p> <p>'Tutorización académica envio'</p> |
| apoyoorg_envio | <p>'¿Estuvo satisfecho/a con el apoyo ofrecido por su institución de envío?: Apoyo de organizaciones/iniciativas estudiantiles'</p> <p>'Apoyo de organizaciones/iniciativas de estudiantes envio'</p> <p>'¿Estuvo satisfecho/a con el apoyo ofrecido por su institución de envío?: Apoyo de organizaciones/iniciativas de estudiantes'</p> |

| | |
|------------------------|--|
| | |
| apoyopers_envio | 'Ayuda con necesidades y problemas personales envío' |
| visa_envio | '¿Estuvo satisfecho/a con la asistencia ofrecida para solventar cuestiones de visado?: Por su institución de envío' 'Por su institución de envío visa' 'Por su empresa/organización de acogida visa' |
| seguro_envio | '¿Estuvo satisfecho/a con la asistencia ofrecida para solventar cuestiones del seguro?: Por su institución de envío' 'Por su institución de envío' |
| apoyoadm_dest | '¿Estuvo satisfecho/a con el apoyo ofrecido en su institución de acogida?: Apoyo administrativo' 'Apoyo administrativo' |
| tutor_dest | '¿Estuvo satisfecho/a con el apoyo ofrecido en su institución de acogida?: Tutorización académica' 'Tutorización académica' |
| apoyoorg_dest | '¿Estuvo satisfecho/a con el apoyo ofrecido en su institución de acogida?: Apoyo de organizaciones/iniciativas estudiantiles' 'Apoyo de organizaciones/iniciativas de estudiantes' |
| apoyopers_dest | 'Ayuda con necesidades y problemas personales' |

| | |
|-----------------------------|---|
| visa_dest | <p>'¿Estuvo satisfecho/a con la asistencia ofrecida para solventar cuestiones de visado?: Por su institución de acogida'</p> <p>'Por su institución de acogida visa'</p> <p>'¿Estuvo satisfecho/a con la asistencia ofrecida para solventar cuestiones de visado?: Por su empresa/organización de acogida'</p> <p>'Por su empresa/organización de acogida visa'</p> |
| seguro_dest | <p>'Por su empresa/organización de acogida'</p> <p>'¿Estuvo satisfecho/a con la asistencia ofrecida para solventar cuestiones del seguro?: Por su empresa/organización de acogida'</p> <p>'¿Estuvo satisfecho/a con la asistencia ofrecida para solventar cuestiones del seguro?: Por su institución de acogida'</p> <p>'Por su institución de acogida'</p> |
| calidad_alojamiento | <p>'¿Estuvo satisfecho/a con su alojamiento?'</p> |
| restauracion_centrod | <p>'Evalúe los servicios o instalaciones para estudiantes en la institución de acogida: Cafetería, servicios de restauración'</p> <p>'Cafetería, servicios de restauración'</p> |

| | |
|----------------------|---|
| | 'Cafetería / servicios de restauración' |
| aulas_centrod | 'Evalúe los servicios o instalaciones para estudiantes en la institución de acogida: Aulas' 'Aulas' |
| gasto_medio | '¿Cuál fue el gasto medio (alojamiento y manutención) por mes durante su estancia en el extranjero?' '¿Cuál fue el gasto medio (alojamiento y manutención) por mes, en euros, durante su estancia en el extranjero?' '¿Cuál fue el gasto medio (alojamiento y manutención) por mes, en euros, durante su estancia en el extranjero?' |

A partir de la tabla de correspondencias para las formulaciones alternativas de las diversas preguntas de interés, se implementaron funciones en Python que reconocieran los encabezados mediante patrones de texto (por ejemplo, palabras clave o estructuras comunes), permitiendo agrupar automáticamente todas las versiones de una misma pregunta bajo una única variable homogénea, sin necesidad de intervención manual año tras año.

Corrección de errores tipográficos y duplicados

Durante la revisión de los datos, se detectaron errores tipográficos recurrentes tanto en los encabezados como en las respuestas de las encuestas. También se identificaron registros duplicados que distorsionaban el análisis. Para resolver estos problemas, se aplicaron técnicas de procesamiento de texto y filtros automáticos que permitieron estandarizar las entradas, corregir errores comunes y eliminar duplicados o registros redundantes, garantizando la integridad de la información.

Unificación de nomenclaturas y categorías

Otro aspecto clave a corregir fue la heterogeneidad en las categorías de respuesta en algunas preguntas que, aunque conceptualmente eran similares, se formulaban de manera distinta. Por

ejemplo, una misma titulación podía aparecer con nombres diferentes a lo largo de los años. Se construyeron diccionarios de equivalencias para normalizar estas denominaciones y asegurar la coherencia de las categorías en todas las bases de datos.

Tabla 2. Errores tipográficos en las respuestas respecto a la titulación

| Titulación | Nombres |
|--|--|
| Grado en Administración y Dirección de Empresas | 'Grado en Administración y Dirección de Empresas', 'Grado en Administración y Dirección de Empresas', 'Grado en Administración y Dirección de Empresa' |
| Grado en Fisioterapia | 'Grado en Fisioterapia', 'Grado en Fisoterapia' |
| Grado en Farmacia | 'Grado en Farmacia', 'Grado de Farmacia' |
| Grado en Psicología | 'Grado en Psicología', 'Grado en Psicologia' |
| Licenciado en Enología | 'Licenciado en Enología', 'Licenciado en Enologia' |
| Grado en Ingeniería de Tecnologías de Comunicación | 'Grado en Ingeniería de Tecnologías de Comunicación', 'Grado en Ingeniería de Tecnologias de comunicacion' |

Integración de bases de datos

Una vez finalizado el proceso de limpieza y homogeneización de las bases de datos, almacenadas todas ellas en hojas de cálculo de Google en Google Drive, estas se cargan de forma individualizada en el espacio de trabajo de Google Colab, mediante la conexión directa a Google Drive.

A través de esta conexión, el sistema accede a los archivos y detecta automáticamente los encabezados presentes en cada hoja de cálculo. Para cada uno de ellos, se aplica un proceso de asignación de variables basado en un diccionario de correspondencias previamente definido, que vincula distintos posibles nombres de encabezado para cada variable. Se crea además, para cada fichero, una variable que permite identificar el año de realización de cada encuesta (año), así como el colectivo.

Posteriormente, una vez cargada toda la información en una única tabla de datos que identifica perfectamente el año en que se hizo la encuesta y el colectivo, junto con todas las variables de interés, se guarda en un archivo en formato ".pkl", que optimiza su posterior acceso y tratamiento en Python. Este formato, además de acelerar los tiempos de carga, permite conservar los tipos de datos y estructuras complejas de las tabla (dataframes), lo que resulta especialmente útil cuando se trabaja con grandes volúmenes de información y se requiere de sucesivos accesos a los datos ya depurados para su tratamiento analítico.

Programación y software

El cuaderno de procesado comienza pues, estableciendo una conexión con Google Drive a través de "google.colab.drive", lo que permite acceder directamente a las hojas de cálculo almacenadas en esta plataforma. A continuación, se procede a la lectura de varios archivos Excel que contienen las respuestas anuales de las encuestas Erasmus.

Durante el proceso, se automatiza la detección de los encabezados de las hojas de cálculo. Además, se aplica una correspondencia predefinida entre los nombres de las columnas y las variables estándar, lo que facilita la unificación de los encabezados. Al final, se consolidan todas las bases de datos en un único DataFrame estructurado. Este archivo consolidado se guarda en formato .pkl, lo que preserva la estructura del DataFrame y sus tipos de datos, mejorando la eficiencia al trabajar con el archivo en futuros cuadernos.

Para realizar este proceso, se utilizan varias librerías:

- **pandas**, que se emplea para leer y manipular los datos tabulares
- **os** para gestión de ficheros
- **re**, para trabajar con los nombres de los mismos mediante expresiones regulares.
- **openpyxl**, que se utiliza como motor para leer los archivos con extensión .xlsx.
- **pickle**, que permite guardar el DataFrame resultante en un archivo .pkl.

6.2 Análisis estadístico

El análisis estadístico, se basa fundamentalmente en el procesado y descripción de variables de recuento, variables cualitativas y escalas de valoración, todas ellas comunes en el ámbito de las encuestas sociales.

Se estructura esta sección en el procesado de los datos, el análisis descriptivo y los contrastes de hipótesis resueltos para responder las cuestiones de interés planteadas.

Procesado de los datos

Gran parte de las variables recogidas en el cuestionario están estructuradas como escalas de tipo Likert, utilizando cinco niveles de respuesta: 'Muy satisfecho/a', 'Bastante satisfecho/a', 'Ni satisfecho/a ni insatisfecho/a', 'Bastante insatisfecho/a' y 'Muy insatisfecho/a'. Estas escalas permiten captar matices en la percepción de los encuestados sobre distintos aspectos de su experiencia Erasmus+, desde la calidad académica hasta el acompañamiento institucional. Sin embargo, para facilitar el análisis y la interpretación de los datos, en algunos casos se ha optado por simplificar estas escalas de cinco niveles a tres, como comentamos en la sección sobre procesado de datos.

Por ejemplo, en la variable *satis3*, que recoge la satisfacción global con la experiencia, se ha reducido la escala a tres niveles mediante la agrupación de las respuestas. Las opciones 'Muy satisfecho/a' y 'Bastante satisfecho/a' se combinaron en una categoría de satisfacción general, mientras que 'Bastante insatisfecho/a' y 'Muy insatisfecho/a' se agruparon bajo una categoría de insatisfacción. Por su parte, la opción intermedia 'Ni satisfecho/a ni insatisfecho/a' se mantuvo como una categoría neutral. Esta simplificación tiene como objetivo facilitar tanto el análisis descriptivo como el uso posterior de modelos estadísticos o visualizaciones más claras.

Además de las variables ordinales tipo Likert, el cuestionario también incluye variables categóricas de respuesta múltiple. Un ejemplo relevante es la variable relacionada con los motivos para participar en el programa Erasmus+, en la que los encuestados podían seleccionar varias opciones simultáneamente. Esto genera un tipo de dato compuesto que requiere un tratamiento específico para su análisis. Para ello, fue necesario descomponer cada respuesta en sus componentes individuales y contabilizar la frecuencia con la que aparecía cada motivo, independientemente de si formaba parte de una selección de dos, tres o más respuestas por parte de un mismo participante. De este modo, se pudo construir una matriz de motivos a nivel individual, que permite analizar su distribución y detectar patrones en las motivaciones del alumnado.

Otro aspecto importante en la limpieza y procesamiento de los datos se refiere a aquellas variables en las que la respuesta era abierta, es decir, el encuestado debía escribir manualmente su respuesta. Un caso típico es el campo correspondiente a la titulación del estudiante.

Dado que no existía una lista cerrada o desplegable de opciones, las respuestas recogidas presentaban una gran heterogeneidad en su formato, incluyendo errores tipográficos, abreviaciones no estandarizadas, diferencias de capitalización, omisiones o formas inconsistentes de nombrar un mismo grado. Esto dificultaba la categorización automática de los datos y requería un trabajo manual de revisión y mapeo.

En concreto, se creó un diccionario que relaciona las respuestas erróneas, incompletas o alternativas con la titulación oficial correspondiente, permitiendo así estandarizar este campo y agrupar correctamente a los estudiantes según su programa académico. Este proceso fue esencial para asegurar la fiabilidad de los análisis comparativos por titulación, ya que sin una normalización previa, los resultados habrían estado fragmentados o distorsionados por entradas inconsistentes.

En conjunto, todo este trabajo de transformación, agrupación y limpieza de datos ha sido necesario para garantizar la coherencia del conjunto, asegurar una correcta interpretación de los resultados y facilitar posteriores análisis más complejos. La calidad del tratamiento previo de las variables, tanto en su forma como en su contenido, es clave para que las conclusiones obtenidas sean sólidas, representativas y útiles para la toma de decisiones o la mejora de futuras ediciones del programa.

Análisis descriptivo

La función principal del análisis descriptivo es generar los gráficos, tablas e imágenes necesarias que respondan a las preguntas de interés planteadas por la ORI para obtener un informe sobre la actuación Erasmus. Las representaciones gráficas no solo facilitan la comprensión de la información, sino que también permiten una interpretación visual que complementa el análisis cuantitativo, haciendo que los resultados sean accesibles para audiencias tanto técnicas como no especializadas. Además, su generación automatizada garantiza la consistencia y uniformidad en la presentación de los informes, contribuyendo a una comunicación profesional y eficiente de los hallazgos.

Para obtener una visión clara y detallada de las opiniones, percepciones y valoraciones de los encuestados, facilitando tanto el análisis de tendencias (a lo largo de los años), como la comparación entre diferentes grupos o períodos.

Los gráficos considerados incluyen una variedad de visualizaciones, tales como:

- diagramas de barras y gráficos de sectores (pastel), para variables de tipo categórico, con la opción de apilamiento cuando queremos mostrar la asociación entre diversas variables categóricas;
- histogramas que permiten ilustrar de manera clara la distribución de datos de tipo numérico
- gráficos boxplots para para visualizar la distribución de los datos, identificar la mediana, la dispersión y detectar posibles valores atípicos.

Contrastes de hipótesis

Además del análisis estadístico descriptivo, se incorporaron pruebas de hipótesis con el fin de profundizar en el estudio de las relaciones entre determinadas variables y la satisfacción general de los estudiantes con su experiencia Erasmus. Para ello, se formularon y resolvieron algunos casos de contraste de hipótesis utilizando la prueba de chi-cuadrado, una técnica estadística adecuada para examinar la asociación entre variables cualitativas. Estas pruebas permiten detectar asociaciones estadísticamente significativas entre determinadas condiciones y los niveles de satisfacción declarados por los estudiantes. Así, complementan el análisis descriptivo al ofrecer evidencia sobre cómo factores como el apoyo recibido o la percepción económica pueden influir en la valoración final de la experiencia Erasmus.

Software

Gráficas Descriptivas

Para la creación de los gráficos incluidos en el análisis estadístico, se han empleado dos bibliotecas ampliamente utilizadas en Python: Seaborn y Matplotlib. La combinación de ambas ofrece una solución eficiente y flexible para representar visualmente datos tanto cualitativos como cuantitativos. Mientras que Matplotlib actúa como base para la generación de gráficos detallados y altamente personalizables, Seaborn simplifica la construcción de visualizaciones estadísticas más complejas, añadiendo automáticamente elementos estéticos y funcionales.

Gracias a estas herramientas, ha sido posible plasmar los resultados de manera clara, coherente y visualmente atractiva, facilitando su interpretación dentro del informe final.

Tablas

Para la creación de las tablas en el análisis de los datos, se ha empleado principalmente Pandas. A continuación, se ha utilizado Tabulate para formatear y mejorar la presentación de las tablas generadas. Esta biblioteca permite crear tablas bien estructuradas y visualmente atractivas, lo que facilita la lectura y comprensión de los resultados. Tabulate proporciona un estilo de cuadrícula que organiza claramente los datos, permitiendo que los resultados sean fácilmente interpretados en el informe final.

Se ha utilizado principalmente el cuaderno:

🔗 240. Visualización de datos multivariantes con Seaborn.ipynb , desarrollado por IA4LEGOS UMH (2023).

Resultados: análisis descriptivo

En este apartado se presentan algunos de los gráficos generados , representativos del análisis descriptivo llevado a cabo. A continuación, vamos a ver algunos de los gráficos creados principalmente con las librerías *seaborn* y *matplotlib*.

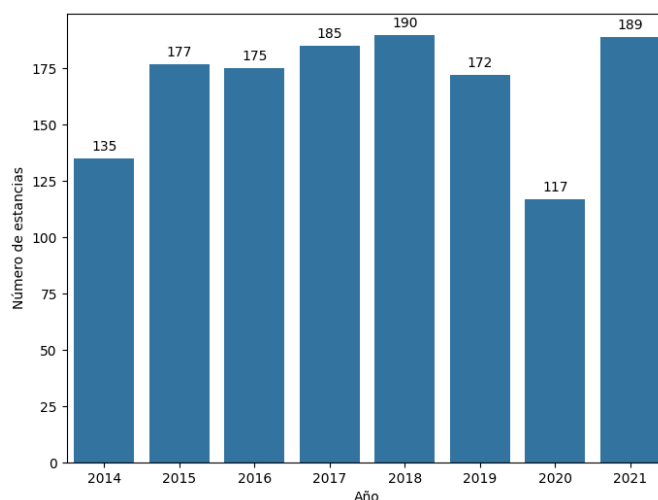


Figura 1. Número de movilidades por año

Esta figura muestra el número de movilidades por año en los años disponibles, incluyendo a todos los grados participantes en el programa Erasmus. Es la figura con la que comienzan todos los informes generados, proporcionando una visión general de la evolución de las movilidades a lo largo del tiempo.

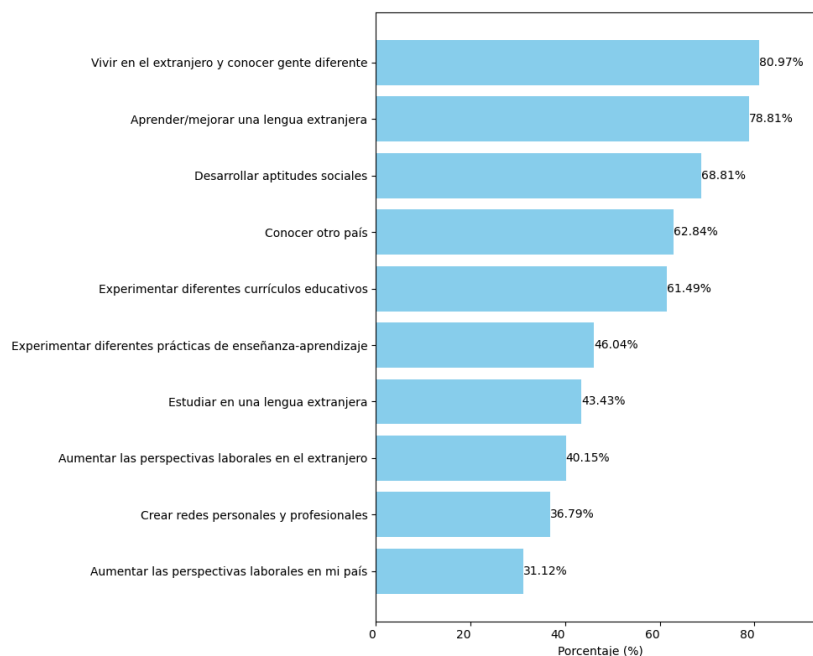


Figura 2. Motivos para la realización de la movilidad

En la figura 2, se opta por reflejar de mayor a menor las respuestas más votadas, de forma que se vea antes la respuesta más común para los encuestados.

Esta disposición facilita una lectura rápida y clara de las prioridades percibidas por los participantes, permitiendo identificar de inmediato cuáles son los principales motivos que impulsan la movilidad educativa internacional.

Además, cabe destacar que este gráfico de barras muestra las respuestas en porcentaje, y el orden descendente visualmente destaca la diferencia de peso entre unas razones y otras, ayudando a interpretar no sólo qué factores son más importantes, sino también la proporción relativa de su relevancia en comparación con el resto.

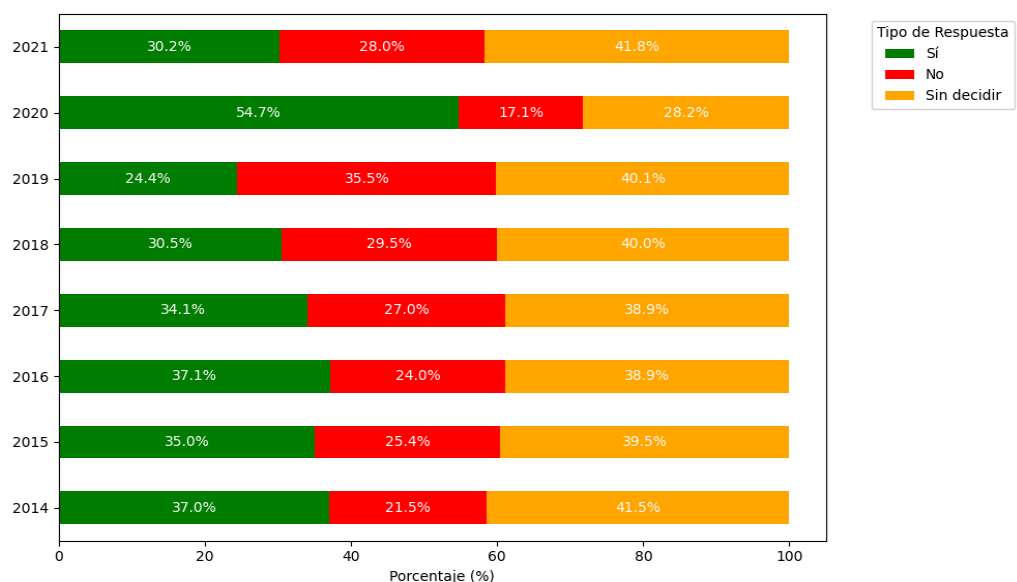


Figura 3. Repetir la experiencia en caso de ser posible

En la figura 3 se presenta un gráfico de barras apiladas por año, según los años que se hayan seleccionado, en este caso todos los años disponibles en el estudio, con el objetivo de reflejar la evolución del deseo de poder repetir la experiencia.

Se ha utilizado una paleta de colores tipo semáforo —verde para “Sí”, rojo para “No” y amarillo para “Sin decidir”— que favorece una lectura rápida y clara de los datos. Esta representación permite informarse de todos los años en un solo gráfico y, además, observar de forma comparativa cómo ha variado la disposición de los participantes a repetir la experiencia a lo largo del tiempo.

También se incluye un gráfico de boxplot, respecto al gasto medio, se puede observar en la siguiente figura:

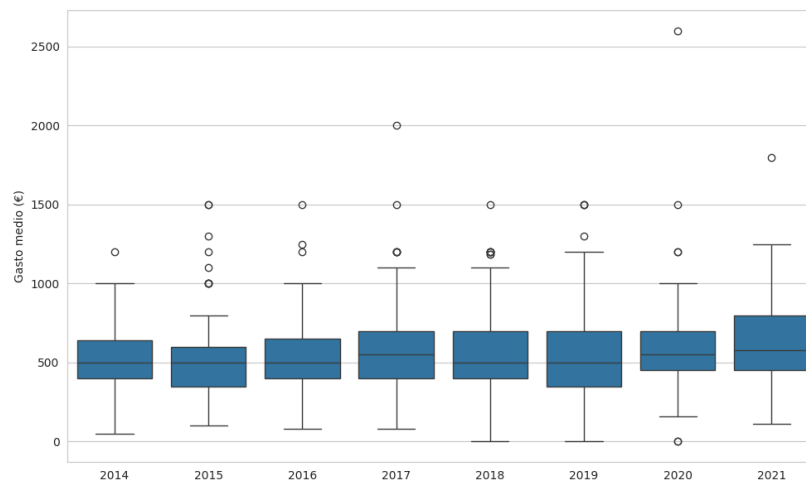


Figura 4. Diagrama de Boxplot del Gasto Medio

La figura 4 muestra un gráfico de boxplots que representa la distribución del gasto medio por año, permitiendo comparar visualmente cómo ha variado este aspecto a lo largo del tiempo. Cada boxplot resume la dispersión de los datos anuales, mostrando la mediana, el rango intercuartílico y los valores atípicos, si los hay.

Esta representación resulta útil para identificar años con mayor o menor variabilidad en el gasto, detectar tendencias generales, y observar posibles cambios en el comportamiento económico de los participantes. Además, facilita una comparación directa entre años, tanto en términos de gasto central como de la amplitud de los datos

A continuación, se presenta otro tipo de contenido descriptivo incluido en los informes: tablas informativas organizadas por países. En estas tablas se muestra un ranking de universidades basado en el número de movildades realizadas hacia cada una de ellas, así como los distintos grados desde los cuales se han llevado a cabo dichas movildades. Esta información permite identificar los destinos más frecuentes y los perfiles académicos de los estudiantes que participan en los programas de movilidad.

Tabla 3. Ranking de Universidades en Eslovaquia

| Ranking | Universidad | Número de Visitantes | Grados |
|---------|--|----------------------|---|
| 1 | UNIVERZITA KOMENSKEHO V BRATISLAVE | 8 | Grado en Ciencias Ambientales, Grado en Ciencias Ambientales. Plan 2016 |
| 2 | SLOVENSKA ZDRAVOTNICKA UNIVERZITA V BRATISLAVE | 2 | Grado en Fisioterapia |
| 3 | SLOVENSKA POLNOHOSPODARSKA UNIVERZITA V NITRE | 1 | Grado en Ciencia y Tecnología de los Alimentos |

Como se puede observar en la tabla presentada, este formato resulta muy útil para visualizar de forma descriptiva las universidades visitadas en los distintos países. Permite identificar fácilmente cuál es la universidad más visitada en cada país y qué grados han realizado movildades hacia ella.

En este caso concreto, se muestra únicamente la tabla correspondiente a Eslovaquia a modo de ejemplo, pero los informes completos incluyen tablas similares para todos los países de los que se disponga de información, en función de los datos seleccionados por el usuario.

Resultados: contrastes de hipótesis

En esta sección se exponen algunos de los casos de contraste de hipótesis que se han llevado a cabo para aportar una dimensión inferencial al estudio. Estos contrastes permiten explorar y confirmar la existencia de relaciones estadísticamente significativas entre determinadas variables de interés, tales como factores que pueden influir en la satisfacción general de los estudiantes con su experiencia Erasmus. La inclusión de estas pruebas fortalece la rigurosidad del análisis y proporciona evidencia sólida sobre las asociaciones detectadas en los datos.

A continuación, se muestran algunos gráficos realizados para investigar la relación entre variables categóricas (y vinculados a los contrastes de hipótesis realizados con la prueba de chi-cuadrado, para analizar posibles relaciones significativas).

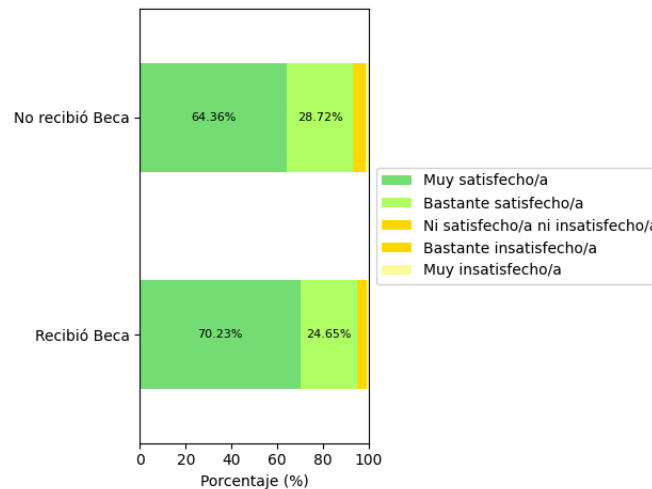


Figura 5. Efecto de percibir beca para la satisfacción con la experiencia.

Esta figura muestra los porcentajes de satisfacción con la experiencia según si el estudiante percibió o no una beca.

A nivel interno de código, se realiza un contraste chi-cuadrado, en el que se almacena una respuesta en función del análisis, para luego mostrar en el informe.

En este caso, la respuesta que muestra el informe sería esta:

“Observamos que entre los que no recibieron beca el 93,08% están satisfechos o muy satisfechos con la movilidad, y este porcentaje es del 94.88% para los que recibieron beca. Al testar con el test chi-cuadrado diferencias estadísticas, concluimos que no se evidencian ($p\text{-valor}=0.38$), es decir, no podemos afirmar que realmente el hecho de percibir beca repercuta más o menos sobre la satisfacción del estudiantado con la movilidad.”

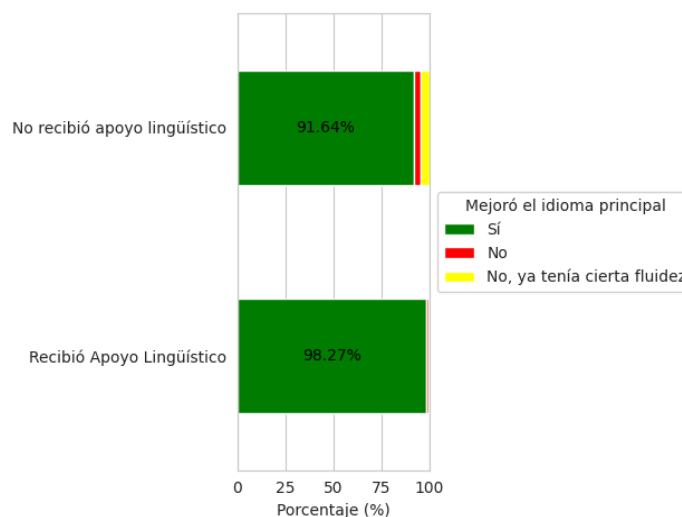


Figura 6. Efecto de recibir apoyo lingüístico en la mejora del idioma

En esta figura, se analiza el efecto de recibir apoyo lingüístico para mejorar el idioma. En este caso el p-valor tiene como resultado <0.001 , es decir, el apoyo lingüístico supuso ventajas para conseguir mejoras con el idioma.

En los informes finales se incluyen contrastes de hipótesis adicionales que permiten profundizar en el análisis de la experiencia estudiantil.

Por ejemplo, se examina la relación entre el nivel de satisfacción con la experiencia académica y la percepción de apoyo lingüístico, así como la mejora en el dominio del idioma en función de si el estudiante ha realizado una movilidad de un semestre o de una duración superior. Estos análisis complementarios tienen como objetivo ofrecer una visión más completa y detallada sobre los factores que influyen en la experiencia de los estudiantes durante su estancia académica.

Estos resultados mostrados en relación con los contrastes de hipótesis de las figuras 8, 9 y 10, se han obtenido utilizando datos agregados de todos los años y de todos los grados.

Es importante tener en cuenta que, en función de los años o grados que se seleccionen, los resultados podrían variar, lo que afectaría tanto al p-valor como a su posible interpretación.

En algunos casos, la limitada cantidad de datos disponibles, por ejemplo, al filtrar por un grado específico dentro de un año concreto, puede dificultar o incluso impedir la realización del

contraste de hipótesis mediante la prueba de chi-cuadrado. Esto se debe a que esta prueba requiere un tamaño de muestra mínimo para que sus resultados sean estadísticamente válidos.

Cuando no se cumplen estas condiciones, el contraste puede arrojar resultados erróneos o directamente no puede calcularse. Por ello, en ciertos informes donde no es posible aplicar el contraste de manera adecuada, se opta por mostrar únicamente las figuras o gráficos descriptivos, omitiendo cualquier comentario relacionado con la significación estadística o la interpretación de los resultados.

6.3 Elaboración de mapas

Para enriquecer el informe, y dado el carácter geográfico de la encuesta, se incluyen mapas que permiten visualizar la distribución de los usuarios Erasmus de la UMH en los diversos países europeos de destino, integrando en ellos su nivel de satisfacción.

Este enfoque no solo enriquece la presentación visual del informe, sino que también facilita la comprensión contextualizada de los datos, al vincular los resultados estadísticos con su dimensión territorial.

En la Figura 7 se representa un mapa con una escala de color que va del amarillo al rojo, en función del número de movilidades registradas en cada país. Junto a la abreviatura de cada país se indica entre paréntesis el número total de movilidades, siendo Italia el país con mayor participación, con un total de 498 movilidades.

La Figura 8, por su parte, muestra un mapa con una escala de color que va del amarillo al verde, en función de la satisfacción media expresada por los estudiantes en relación con su experiencia Erasmus en cada país.

Junto a la abreviatura de cada país aparece la puntuación media obtenida, en una escala del 1 al 5, donde 5 representa el mayor nivel de satisfacción.

Para calcular esta media, las respuestas cualitativas se transforman en valores numéricos mediante una escala de valoración: se asigna un valor de 1 a 'Muy insatisfecho/a', 2 a 'Bastante insatisfecho/a', 3 a 'Ni satisfecho/a ni insatisfecho/a', 4 a 'Bastante satisfecho/a' y 5 a 'Muy satisfecho/a'.

Esta codificación permite cuantificar las percepciones subjetivas de los estudiantes y representar gráficamente los niveles medios de satisfacción por país, facilitando así una comparación clara y visualmente accesible entre los distintos destinos europeos.

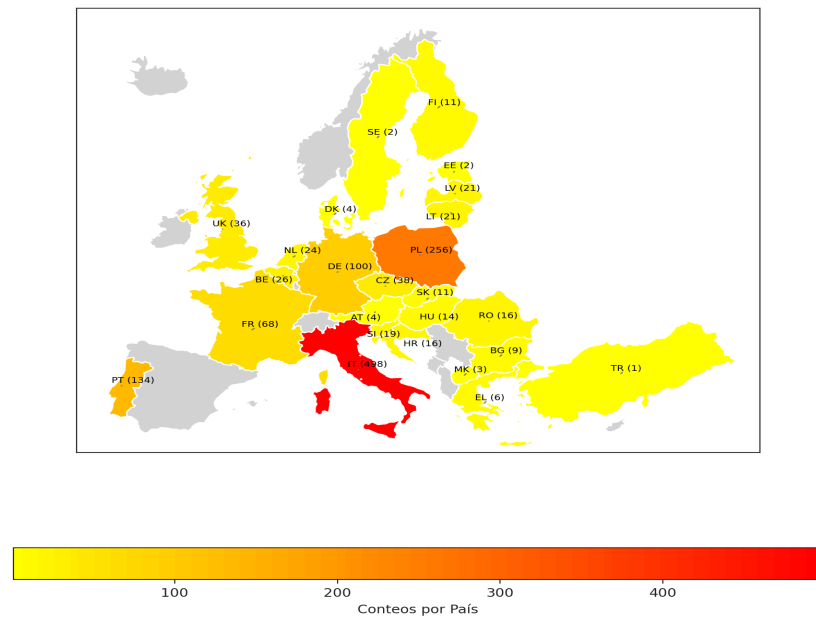


Figura 7. Mapa de densidad de movilizaciones

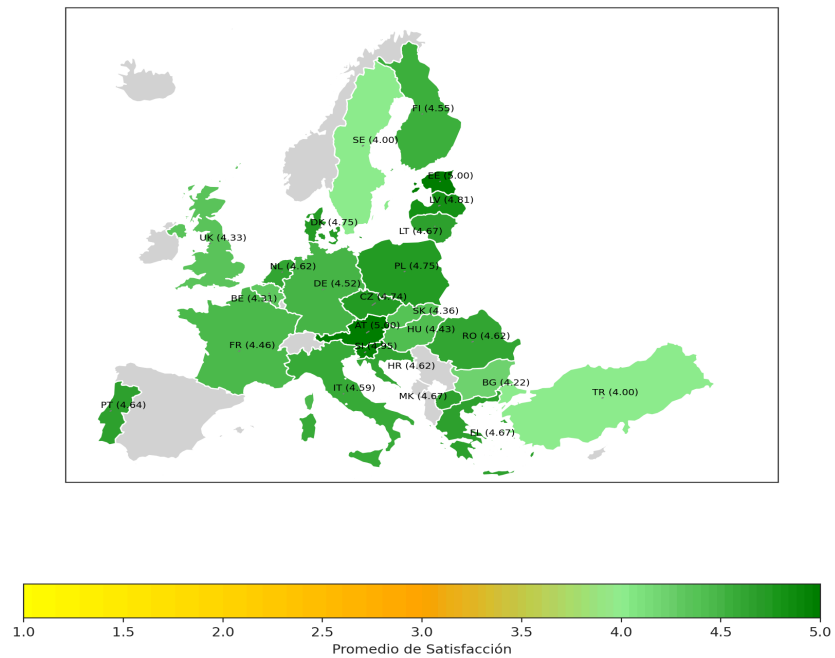


Figura 8. Mapa de Satisfacción con la Movilidad

Programación y software

Para la elaboración de los mapas geográficos incluidos en los informes, el proceso comenzó con la clonación de un [repositorio de GitHub](#), que proporciona los recursos necesarios para el tratamiento de datos espaciales. A continuación, se procedió a la extracción de archivos ZIP que contienen datos NUTS (Nomenclatura de Unidades Territoriales Estadísticas) de la Unión Europea, fundamentales para representar correctamente las distintas regiones del espacio europeo.

Estos archivos, que incluyen tanto shapefiles como archivos .dbf, fueron manipulados mediante el uso de librerías especializadas como geopandas y simpledbf, que permiten integrar datos estadísticos con información geoespacial. El resultado son mapas que muestran, por un lado, la densidad de movilidades Erasmus por país y, por otro, los niveles medios de satisfacción de los estudiantes, ambos expresados mediante escalas de color fácilmente interpretables.

Para la generación de los mapas, se utilizó el cuaderno de Google Colab, [Mapas eurostat.ipynb](#), desarrollado por Morales (2021).

6.4 Generación del informe Word

La generación automatizada de informes en formato Word constituye una parte fundamental del flujo de trabajo, ya que permite recopilar y presentar los resultados del análisis de manera clara, coherente y profesional. Esta etapa no se limita únicamente a la inclusión de gráficos o tablas, sino que organiza el contenido de forma estructurada, facilitando su comprensión y lectura. La incorporación de elementos visuales junto con los comentarios interpretativos correspondientes permite contextualizar la información y comunicar los hallazgos de forma efectiva, mejorando así la utilidad del informe tanto para uso interno como para su presentación a terceros.

Además, el informe final incluye funciones que mejoran notablemente la experiencia del usuario, como la creación automática de un índice con hipervínculos que facilita la navegación entre secciones, y la numeración de páginas, que aporta orden y profesionalismo al documento. Este enfoque automatizado no solo optimiza el tiempo dedicado a la redacción y maquetación, sino que también asegura la consistencia formal en cada entrega, reduciendo la posibilidad de errores y adaptándose con facilidad a diferentes volúmenes de información o requerimientos específicos. En conjunto, esta herramienta representa una solución eficaz y escalable para la generación de documentación técnica o académica.

Insertar imágenes

La función principal del módulo de generación de informes es organizar y presentar de forma clara y estructurada las figuras generadas durante el análisis.

Cada vez que se crea una figura, como las mostradas en el apartado 6.1, esta se guarda automáticamente en los archivos del cuaderno mediante la librería `matplotlib`, utilizando la instrucción `plt.savefig()`. Esta función permite al programador asignar un nombre personalizado al archivo y definir su formato de salida. En este caso, todas las figuras se almacenan en formato `.png`, asegurando una buena calidad gráfica y compatibilidad con su posterior inclusión en el documento Word final.

Por ejemplo, si se desea guardar la figura 1, bastaría con incluir la instrucción `plt.savefig('figura1.png')` al final de la celda en la que se genera dicha figura. La imagen quedará almacenada automáticamente en el panel de archivos del cuaderno de Google Colab.

Además de la imagen de la figura, también se suele guardar una variable que contiene texto comentado, para luego plasmar en el informe junto a la figura.

Para insertar las figuras guardadas en el informe final, hay que utilizar el comando `doc.add_picture`, dentro del cual podemos elegir el tamaño y localización de la figura.

Índice Navegable

Una vez generadas todas las figuras, gráficos y resultados comentados en los apartados anteriores, se integran en el documento Word final. No obstante, con el objetivo de mejorar la presentación y usabilidad del informe, se incorporan también ciertos elementos adicionales que contribuyen a una lectura más cómoda, organizada y accesible. Uno de los más relevantes es la creación de un índice inicial con hipervínculos internos, que permite al usuario navegar directamente entre los distintos apartados del documento con un solo clic. Esta funcionalidad resulta especialmente útil en informes extensos, donde desplazarse manualmente por varias páginas puede resultar poco eficiente o engorroso.

La implementación de este índice navegable se llevó a cabo mediante la combinación de distintos procedimientos ofrecidos por la librería `python-docx`, complementados con la manipulación directa de la estructura XML del documento, lo cual permite ir más allá de las funcionalidades básicas incluidas en su interfaz de alto nivel. En concreto, se definen marcadores invisibles dentro del texto, que actúan como puntos de referencia o anclas. Estos marcadores se colocan justo antes de cada sección relevante del informe, identificados mediante un nombre único. Posteriormente, en la parte inicial del documento, se insertan enlaces que apuntan directamente a esos marcadores, funcionando como accesos rápidos a cada parte del texto.

Este sistema de hipervínculos internos se construye a partir de dos operaciones clave, que deben ejecutarse de forma consecutiva y cuidadosamente coordinada para que el índice funcione correctamente. En primer lugar, es necesario insertar marcadores en el documento, los cuales funcionan como puntos de referencia invisibles dentro del texto. Cada marcador se asocia a un bloque de contenido importante, por ejemplo, el comienzo de una sección, subsección o figura destacada, y se define mediante una estructura XML específica del formato Word (Office Open XML). Técnicamente, esto se logra introduciendo elementos `w:bookmarkStart` y `w:bookmarkEnd`, los cuales enmarcan el contenido que se desea etiquetar,

utilizando un nombre identificador único que luego puede ser referenciado desde otras partes del documento.

Una vez creados estos marcadores, la segunda operación consiste en generar los hipervínculos que permiten al lector saltar directamente a esos puntos desde el índice inicial. Para ello, se construye un enlace dentro del párrafo correspondiente del índice, también empleando etiquetas XML, concretamente el elemento `w:hyperlink`, al cual se le asigna como destino el nombre del marcador previamente definido. Dentro del hipervínculo se introduce el texto visible que actuará como enlace, que puede corresponder al título de la sección o a cualquier otra descripción relevante.

Además de la funcionalidad técnica, se cuida también la presentación visual de estos enlaces. Se les aplica un formato específico que los distingue como elementos interactivos, adoptando el estilo convencional de los hipervínculos en documentos digitales: texto azul subrayado. Este detalle no solo mejora la estética del informe, sino que también facilita al lector la identificación inmediata de los elementos clicables, haciendo más intuitiva la navegación.

Gracias a este sistema combinado de marcadores y enlaces internos, el documento final no solo contiene los resultados del análisis de forma estructurada, sino que se convierte en una herramienta interactiva que mejora sustancialmente la experiencia del usuario. El lector puede desplazarse rápidamente a cualquier apartado del informe sin necesidad de buscar manualmente o recorrer varias páginas, lo que resulta especialmente útil en documentos extensos o de carácter técnico.

Cabe destacar que esta funcionalidad no está directamente soportada por la interfaz de alto nivel de la librería `python-docx`, por lo que su implementación requiere intervenir directamente en la estructura XML subyacente del documento. Esto implica una comprensión más profunda del formato de archivos DOCX y una manipulación controlada de sus nodos internos, lo que se ha conseguido mediante funciones personalizadas desarrolladas específicamente para este propósito. Este enfoque demuestra cómo, con el uso adecuado de herramientas avanzadas, es posible extender las capacidades de una librería y dotar a un proceso automatizado de un nivel de detalle y profesionalidad que, en muchos casos, iguala o incluso supera el de una edición manual.

Numeración de páginas

Para complementar la presentación del informe automatizado, resulta fundamental ofrecer una estructura clara y profesional también en aspectos formales como la numeración de las páginas. Esta función se encarga de añadir de manera automática los números de página en el pie de cada sección del documento, contribuyendo así a mejorar la organización y facilitar la consulta rápida del contenido.

La implementación se basa en recorrer todas las secciones del documento, accediendo al pie de página correspondiente en cada una de ellas. En este espacio, se crea o reutiliza un párrafo destinado a contener el número de página. Para insertar el número, se construye una secuencia específica de elementos XML que representan un campo dinámico de Word encargado de mostrar la numeración automática. Este campo, compuesto por etiquetas que indican el inicio, la instrucción y el final del contenido, asegura que el número se actualice correctamente según la página en la que se encuentre el lector.

Además, para mantener una presentación homogénea y estética, el párrafo que contiene el número de página se centra horizontalmente en el pie de página. De este modo, la función no sólo automatiza un elemento fundamental en cualquier informe formal, sino que también contribuye a mejorar la experiencia visual y funcional del documento final.

La función comienza recorriendo todas las secciones del documento mediante un bucle que itera sobre `doc.sections`. Esto es importante porque cada sección en un documento Word puede tener un pie de página diferente, por lo que es necesario aplicar la numeración en todas para mantener la coherencia en el informe.

Para cada sección, se accede al pie de página correspondiente a través del atributo `footer`. A continuación, se verifica si dicho pie de página ya contiene algún párrafo consultando `footer.paragraphs`. Si no existe ningún párrafo, se crea uno nuevo utilizando la función `add_paragraph()`. Este párrafo será el contenedor donde se insertará la numeración de página.

El siguiente paso es añadir un fragmento de texto o "corrida" dentro del párrafo, que se genera con el método `add_run()`. Esta corrida es necesaria para incluir los elementos dinámicos que Word interpreta como campo de número de página.

Dentro de este proceso, se añaden tres elementos XML que construyen el campo dinámico. Primero, se inserta un elemento `w:fldChar` con el atributo `w:fldCharType` configurado como `begin`, que indica el inicio del campo. Después, se añade un elemento `w:instrText` que contiene la instrucción `PAGE`, la cual indica que debe mostrarse el número de página actual. Finalmente,

se incorpora otro elemento `w:fldChar` con el atributo `w:fldCharType` en `end`, que marca el fin del campo. Estos elementos se añaden directamente a la estructura XML del documento mediante la función `append()` aplicada al contenido interno del `run`, lo que permite que Word reconozca y actualice automáticamente la numeración.

Por último, se ajusta la alineación del párrafo que contiene el número de página utilizando la propiedad `alignment` y asignándole el valor `WD_PARAGRAPH_ALIGNMENT.CENTER`. Esto centra el número en el pie de página, mejorando la presentación visual del documento.

Gracias a esta función, la numeración de páginas se añade de manera automática y uniforme en todo el documento, contribuyendo a la profesionalidad y facilidad de uso del informe final.

La automatización que proporciona esta implementación es especialmente relevante en entornos donde se requiere la generación periódica o masiva de documentos con formatos estándar. Al automatizar tanto la estructura como la numeración, se elimina la necesidad de ajustes manuales, reduciendo errores y optimizando el tiempo de producción. Esta escalabilidad convierte al sistema en una herramienta eficiente para la gestión documental en contextos profesionales o académicos.


En conclusión, el desarrollo del software para la generación automática y estructurada de documentos Word refleja una integración avanzada de técnicas que van desde la manipulación de la estructura XML interna hasta la aplicación cuidadosa de estilos y formatos para asegurar la coherencia visual.

Esta solución no sólo automatiza procesos tediosos y propensos a error, sino que también garantiza la creación de informes profesionales y homogéneos, adaptados a las necesidades específicas de cada usuario. La implementación destaca por su robustez y flexibilidad, posicionándose como una herramienta eficaz en la optimización de flujos de trabajo relacionados con la documentación formal.

Software

Para la integración de los gráficos y mapas en el informe final en formato Word, se utiliza la librería *python-docx*, que permite insertar estos elementos directamente desde el entorno de trabajo en Python, manteniendo el formato y la estructura definidos. Esta herramienta permite un control preciso sobre la disposición del contenido en el documento, incluyendo títulos, tablas, imágenes y estilos. Además, se han incorporado funciones específicas para la creación

automática de índices y enlaces internos, lo que mejora significativamente la navegación dentro del informe y facilita el acceso a secciones concretas.

Para la generación de los informes Word, se utilizó el cuaderno de Google Colab  9. Automatización de ficheros word.ipynb , desarrollado por Borrás (2024).

6.5 Interfaz de usuario

Para facilitar el uso de la herramienta desarrollada, se ha diseñado una interfaz gráfica básica pero funcional, implementada en el entorno Google Colab. Esta interfaz permite a cualquier usuario sin conocimientos técnicos seleccionar fácilmente los parámetros necesarios para generar informes personalizados. La elección de Google Colab responde a su accesibilidad, facilidad de uso y compatibilidad con Python, lo que permite una interacción sencilla mediante widgets de selección desplegable (dropdowns) sin necesidad de modificar el código fuente.

Al ejecutar la herramienta, el usuario se encuentra con un conjunto de opciones interactivas que le permiten configurar el informe según sus necesidades. Estas opciones están organizadas en un panel claro e intuitivo, que guía al usuario paso a paso en la personalización de los datos a analizar. Los elementos principales de esta interfaz son los siguientes:

Selección de intervalo temporal

- Año 1: El primer campo desplegable permite seleccionar el año de inicio del análisis. El usuario puede elegir un año concreto (por ejemplo, 2014) o bien optar por la opción "Todos los años", que agrupa la información desde el inicio del periodo disponible (2014) hasta el más reciente (2021).
- Año 2: Este segundo campo se actualiza automáticamente en función de la selección de "Año 1", para facilitar un rango cronológico coherente. Por ejemplo, si se selecciona 2016 como año inicial, el valor predeterminado para "Año 2" se ajusta automáticamente a 2017. Esto evita errores comunes de superposición temporal o selección inconsistente de periodos.

Selección de titulaciones

Otro elemento interactivo es el desplegable de Titulaciones, donde el usuario puede elegir si desea generar el informe con datos de:

- Todas las titulaciones disponibles.
- Una titulación específica (por ejemplo, "Grado en Comunicación Audiovisual y Periodismo" o "Doble Grado en DADE").

- O bien varias titulaciones, permitiendo así hacer informes de por ejemplo, todos los grados que pertenezcan a una facultad.

Este nivel de filtrado permite enfocar el análisis en colectivos específicos y comparar titulaciones concretas a lo largo del tiempo.

Selección según duración de estancia

Finalmente, se ofrece un filtro adicional relacionado con la duración de la estancia Erasmus:

- Todos: Incluye al estudiantado con cualquier duración de movilidad.
- Solo un semestre: Filtra únicamente a quienes realizaron estancias semestrales.
- Más de un semestre: Considera únicamente aquellos casos de estancia superior a un semestre académico.

Esta opción permite adaptar el informe al tipo de experiencia Erasmus vivida, lo que resulta relevante para ciertos análisis de satisfacción o rendimiento.

Esta capa de personalización representa una mejora significativa frente a los procesos manuales anteriores, facilitando la toma de decisiones basada en datos y aumentando la eficiencia del trabajo de la ORI.

Figura 9. Interfaz de selección de parámetros para el informe.

Para facilitar la interacción con el sistema y permitir una personalización sencilla de los informes, se ha diseñado una interfaz dinámica utilizando widgets de Jupyter en Google Colab.

Esta interfaz responde de forma automática a las selecciones del usuario, ajustando las opciones disponibles según corresponda. Por ejemplo, al seleccionar un rango de años, el segundo campo se adapta o se desactiva si no es necesario. Lo mismo ocurre con los demás filtros: cualquier cambio activa una función que actualiza los datos filtrados en tiempo real, sin necesidad de intervención manual.

De este modo, se consigue una experiencia fluida e intuitiva, minimizando errores y permitiendo generar informes comparativos con rapidez y flexibilidad.

6.6 Software y hardware

Con el fin de recopilar en un apartado único toda la información ya incorporada sobre el software utilizado para los diferentes procedimientos abordados en este trabajo, mostramos a continuación un breve resumen de las librerías utilizadas para responder a los diferentes objetivos resueltos.

Carga y manipulación de datos

pandas: Lectura y manipulación de datos tabulares (Excel, DataFrames, análisis).

os, re: Navegación por carpetas y manejo de nombres de archivos con expresiones regulares.

openpyxl: Motor para leer archivos .xlsx.

pickle: Guardado de objetos (como DataFrames) en formato .pkl para un acceso más rápido.

Análisis estadístico

matplotlib: Generación de gráficos detallados y personalizables.

seaborn: Visualizaciones estadísticas estilizadas sobre matplotlib.

tabulate: Formateo de tablas para hacerlas más legibles y estéticamente agradables.

Elaboración de mapas

geopandas: Manipulación y visualización de datos geográficos (shapefiles).

simplifiedbf: Lectura de archivos .dbf asociados a shapefiles geoespaciales.

Generación del informe final

python-docx: Creación automática de documentos Word, incluyendo texto, imágenes y tablas.

Interfaz de Usuario

IPython.display, ipywidgets: Creación de menús desplegables, selectores múltiples y elementos visuales interactivos dentro del entorno del cuaderno..

7. Validación

El sistema desarrollado ha sido probado de manera exhaustiva utilizando las bases de datos correspondientes a las encuestas Erasmus+ recopiladas por la Oficina de Relaciones Internacionales (ORI) de la Universidad Miguel Hernández entre los años 2014 y 2020, ambos inclusive.

Durante el proceso de validación, se consideraron diversos escenarios de uso, relativos a la selección de diferentes grados académicos, rangos de años, filtros por país de destino y combinaciones de variables relevantes. En todos los casos, el sistema fue capaz de generar informes completos en formato Word y adaptar dinámicamente el contenido generado según los parámetros introducidos por el usuario. Este comportamiento confirma que se han alcanzado los objetivos definidos en este trabajo, tanto en lo relativo a la automatización del flujo de trabajo como en la estandarización del análisis y presentación de los resultados.

El enfoque modular y reutilizable de la herramienta, desarrollado en entornos de código abierto y accesibles como Google Colab y Python, permite que esta solución pueda mantenerse y adaptarse fácilmente en el futuro conforme se incorporen nuevas encuestas o cambien las necesidades de análisis de la ORI. Además, su diseño intuitivo, basado en cuadernos interactivos, facilita su uso por parte de personal no experto en programación, ampliando su potencial de impacto institucional.

A continuación, se presentan los principales resultados obtenidos durante la validación de la herramienta, organizados según sus funcionalidades clave y su comportamiento ante diferentes situaciones prácticas de uso.

Generación automática de informes personalizados

Una de las funcionalidades clave del sistema es la capacidad de generar informes estadísticos completamente personalizados, en formato Word, a partir de parámetros definidos por el usuario. En las pruebas realizadas, el sistema fue capaz de:

- Detectar correctamente los grados y años disponibles en las bases de datos
- Permitir la selección simultánea de múltiples grados y años.
- Generar documentos que integran automáticamente gráficos, mapas, tablas y textos explicativos.

- Mantener una estructura coherente y profesional, con numeración de páginas, índice interactivo e hipervínculos internos.

Esto demuestra que el sistema es escalable, flexible y apto para usos reales por parte del personal de la Oficina de Relaciones Internacionales.

Validación del acceso y procesado

Una de las mayores dificultades era la heterogeneidad en los nombres de columnas y estructuras de datos entre años. La herramienta resolvió con éxito este problema gracias a un sistema de correspondencia que mapea automáticamente los distintos nombres al formato estándar interno.

En todas las pruebas, los datos fueron procesados de forma coherente, permitiendo trabajar con un único DataFrame consolidado sin necesidad de intervención manual, lo cual reduce significativamente el tiempo de trabajo y los errores humanos

Visualizaciones y análisis estadístico automatizado

Los informes generados incluyen:

- Gráficos descriptivos relevantes y personalizables.
- Contrastes de hipótesis (chi-cuadrado) aplicados automáticamente.
- Comentarios interpretativos generados de forma contextual según los resultados.
- Mapas geográficos que enriquecen visualmente la presentación de resultados.

Estas visualizaciones se adaptan correctamente al subconjunto de datos seleccionado por el usuario, lo que valida la capacidad de la herramienta para ajustarse dinámicamente a diferentes contextos analíticos.

Optimización del flujo de trabajo

Otro resultado destacado ha sido la mejora en el flujo de trabajo respecto a la situación inicial. La herramienta permite:

- Reducir tiempos de generación de informes de varias horas a minutos.
- Eliminar procesos manuales de copia, limpieza y estructuración de datos.
- Centralizar todo el flujo de trabajo en dos Cuadernos Jupyter consecutivos.

Este nuevo enfoque representa un salto cualitativo en la manera en que la información de las encuestas Erasmus puede ser analizada, presentada y reutilizada.

8. Conclusiones

El presente Trabajo de Fin de Grado ha tenido como finalidad principal el diseño, implementación y validación de una herramienta automatizada para el tratamiento y análisis de los datos procedentes de las encuestas Erasmus+ de la Universidad Miguel Hernández, en el marco de las actividades gestionadas por la Oficina de Relaciones Internacionales (ORI). Esta herramienta responde a una necesidad concreta: optimizar un proceso que hasta ahora se realizaba de forma manual, consumiendo gran cantidad de tiempo y recursos humanos, y con una elevada probabilidad de errores debido a la heterogeneidad de los datos y la variabilidad de las encuestas a lo largo de los años.

A lo largo del trabajo se han planteado varios objetivos específicos, todos ellos orientados a construir una solución eficaz, funcional y escalable. En este sentido, puede afirmarse que todos los objetivos propuestos se han alcanzado satisfactoriamente:

- Automatización de informes personalizados: La herramienta permite generar informes estadísticos en formato Word de manera completamente automática, integrando tablas, gráficos, mapas y textos explicativos. Esta automatización ahorra tiempo, reduce el margen de error humano y garantiza una presentación homogénea, profesional y accesible.
- Homogeneización y limpieza de datos: Se ha creado un sistema de correspondencias que permite unificar variables con diferentes denominaciones en los distintos años, resolviendo así una de las mayores dificultades del proyecto: la variabilidad estructural entre encuestas. Esta normalización permite trabajar con un único conjunto de datos consolidado, lo que simplifica enormemente el análisis posterior.
- Aplicación de análisis estadístico descriptivo e inferencial: La herramienta incorpora diferentes técnicas estadísticas, desde gráficos exploratorios hasta contrastes de hipótesis utilizando la prueba de chi-cuadrado, lo que permite detectar relaciones significativas entre variables. Estos análisis no solo son generados automáticamente, sino que también incluyen comentarios interpretativos generados desde el propio código.
- Visualización de resultados y mapas interactivos: Se ha conseguido representar los resultados de forma visual y comprensible, mediante gráficos de barras, sectores, boxplots y mapas de distribución. Esta capacidad de visualización aporta un valor añadido importante, especialmente para el análisis institucional y la presentación pública de resultados.
- Facilidad de uso y adaptabilidad futura: Gracias a su desarrollo en Google Colab y al uso de librerías ampliamente documentadas de Python, la herramienta es fácilmente

mantenible y utilizable por personal con conocimientos técnicos básicos. Su estructura modular permite incorporar nuevas funcionalidades sin necesidad de rediseñar el sistema desde cero.

A nivel global, el trabajo ha cumplido de forma exitosa su propósito principal: dotar a la ORI de una solución capaz de generar informes estadísticos completos, fiables y personalizables a partir de los datos de las encuestas Erasmus+, sin necesidad de realizar procesos manuales tediosos y propensos a errores.

Además, el trabajo tiene una importante dimensión de transferencia de conocimiento: no se trata solo de un proyecto académico o de investigación, sino de una aplicación real y útil que puede integrarse directamente en los procedimientos internos de una institución pública. Esto le otorga una dimensión práctica que va más allá del ámbito universitario, demostrando que el uso de tecnologías accesibles como Python, Colab y Word puede suponer un cambio sustancial en la forma de gestionar y aprovechar la información en entornos administrativos.

Desde una perspectiva más reflexiva, este TFG también ha supuesto una oportunidad para integrar y aplicar de forma práctica conocimientos adquiridos en varias áreas: análisis de datos, estadística aplicada, programación en Python, automatización de procesos, visualización de resultados y documentación técnica. El hecho de haber trabajado con datos reales, con errores, vacíos y diferencias entre años, ha contribuido a desarrollar una mirada crítica hacia el tratamiento de la información y la necesidad de aplicar soluciones adaptativas y robustas en proyectos de análisis social.

Una de las mayores aportaciones del trabajo es haber demostrado que es posible automatizar completamente un proceso complejo de análisis y presentación de resultados, sin necesidad de recurrir a herramientas propietarias ni sistemas cerrados. Todo el desarrollo se ha realizado con herramientas de código abierto, accesibles, documentadas y sostenibles, lo que refuerza la validez del proyecto en contextos donde los recursos pueden ser limitados.

A partir de los resultados obtenidos y de la experiencia de desarrollo, se pueden identificar diversas líneas de mejora y profundización que podrían abordarse en futuros trabajos, principalmente desde el punto de vista técnico y metodológico.

En primer lugar, sería altamente enriquecedor incorporar análisis estadísticos más avanzados, como regresión logística, análisis multivariante o técnicas de agrupamiento (*clustering*). Estas

metodologías permitirían detectar con mayor precisión patrones y perfiles entre los estudiantes, así como identificar factores clave que influyen en su nivel de satisfacción con la movilidad.

Asimismo, una evolución natural del proyecto consistiría en su integración con herramientas de análisis visual interactivo, mediante la conexión con plataformas como Power BI o Tableau. Esto permitiría complementar los informes en formato Word con *dashboards* dinámicos y actualizables, mejorando la experiencia del usuario y facilitando la exploración de los datos en tiempo real.

En definitiva, este trabajo ha demostrado que es posible dar respuesta a una necesidad concreta y operativa mediante el uso de herramientas tecnológicas accesibles y metodologías de trabajo bien estructuradas. A través del desarrollo de esta herramienta, se ha logrado optimizar un proceso complejo, que hasta ahora requería una inversión considerable de tiempo y esfuerzo, estandarizar el tratamiento de datos provenientes de múltiples fuentes heterogéneas y mejorar sustancialmente la presentación y el análisis de la información recogida en las encuestas Erasmus+ de la Universidad Miguel Hernández.

El sistema desarrollado permite transformar datos desorganizados y dispersos en informes claros, visuales y coherentes, facilitando así la toma de decisiones informadas por parte del personal de la Oficina de Relaciones Internacionales. La automatización implementada no solo agiliza el flujo de trabajo, sino que también aumenta la fiabilidad del análisis, al eliminar errores derivados de la intervención manual y asegurar una metodología repetible y consistente.


Más allá del resultado técnico, este proyecto ha supuesto una experiencia de aprendizaje transversal, que ha permitido aplicar de forma integrada conocimientos de programación, análisis estadístico, gestión de datos, documentación técnica y comprensión institucional. Ha puesto de manifiesto la relevancia de vincular competencias técnicas con una lectura crítica del entorno donde se aplican, adaptando las soluciones no solo a los requerimientos técnicos, sino también a las dinámicas reales del funcionamiento universitario.

Con las posibles mejoras ya esbozadas y la capacidad de adaptación del sistema, la herramienta desarrollada tiene el potencial de consolidarse como un recurso útil y sostenible dentro de la Universidad Miguel Hernández, contribuyendo a una gestión más eficiente, ordenada y basada en datos de los programas de movilidad internacional. Este trabajo, por tanto, no solo ha cumplido los objetivos propuestos, sino que ha abierto nuevas oportunidades de evolución y refinamiento de cara a futuros desarrollos.


Referencias

Conforme a la última edición de las [normas APA \(7ª Edición\)](#) en 2025).

J. Morales, 2021. Mapas Eurostat, disponible como Cuaderno Colab en:

 Mapas eurostat.ipynb . Accedido el 16 de junio de 2025.

IA4LEGOS, 2023. Visualización de datos multivariantes con Seaborn, disponible como cuaderno Colab en:

 240. Visualización de datos multivariantes con Seaborn.ipynb .

Accedido el 16 de junio de 2025.

F. Borrás, 2024. Automatización de ficheros word, disponible como Cuaderno Colab en:


 9. Automatización de ficheros word.ipynb . Accedido el 16 de junio de 2025.

Anexo


Cuaderno de Lectura de Datos:

 [Lectura de Datos .ipynb](#)

Cuaderno de Generación de Informes de Estudiantes:

 [Informes Encuestas Estudiantes.ipynb](#)

Cuaderno de Generación de Informes de Personal en Prácticas:

 [Informes Encuestas Personal en Prácticas.ipynb](#)