



Universitat Oberta
de Catalunya

uoc.edu



Universitat Oberta
de Catalunya

Universitat Oberta de Catalunya (UOC)
Máster en Ciencia de Datos

TRABAJO FIN DE MÁSTER

Calidad Asistencial Sanitaria Europea

Alicia Perdices Guerra

Máster en Ciencia de Datos

Área Medicina (TFM-Med), Salud Pública y Medical Devices.

Nombre Consultor/a Susana Pérez Álvarez

Nombre Profesor/a responsable de la asignatura: Àngels Rius Gavidia.

Fecha de entrega PEC3 (entrega parcial): **13 de abril 2021**



Esta obra está sujeta a una licencia de Reconocimiento-NoComercial-

SinObraDerivada [3.0 España de Creative Commons](https://creativecommons.org/licenses/by-nc-nd/3.0/es/)

FICHA DEL TRABAJO FINAL

Título del trabajo:	<i>Calidad Asistencial Sanitaria Europea</i>
Nombre del autor:	<i>Alicia Perdices Guerra</i>
Nombre del consultor/a:	<i>Susana Pérez Álvarez</i>
Nombre del PRA:	<i>Àngels Rius Gavidia</i>
Fecha de entrega (mm/aaaa):	04/2021
Titulación:	<i>Máster Universitario en Ciencia de Datos</i>
Área del Trabajo Final:	<i>Área Medicina (TFM-Med)</i>
Idioma del trabajo:	<i>Castellano</i>
Palabras clave	<i>Calidad asistencial europea, Estadística de Gasto, Recursos y Actividades, Predicción y patrones por países.</i>
Resumen del Trabajo (máximo 250 palabras): <i>Con la finalidad, contexto de aplicación, metodología, resultados i conclusiones del trabajo.</i>	
<p>El Trabajo fin de máster tiene como objetivo <u>determinar la situación asistencial sanitaria europea</u> a través del análisis estadístico de los gastos generados por diferentes tipos de hospitales y países sobre los diversos cuidados necesarios por enfermedades, de los recursos utilizados como personal médico, enfermería e investigadores, así como de las actividades y/o procedimientos relacionados con altas hospitalarias en diferentes ratios y tasa de ocupación de camas, tipos de operaciones quirúrgicas y tratamientos, tecnología médica, servicios preventivos, y cuidados domiciliarios.</p> <p>Para ello se ha utilizado los datos disponibles en el repositorio de Eurostat. Las tablas contendrían información sobre el cuidado sanitario relacionado con el gasto, recursos y actividades hospitalarias.</p> <p>Después de realizar tareas de limpieza y combinación de tablas, se aplican métodos estadísticos descriptivos que reflejan la realidad de la calidad asistencial en los distintos países europeos. A continuación, se realizan predicciones atendiendo a distintos factores y se obtienen patrones entre los distintos países estudiados.</p> <p>Tras la investigación se persigue obtener la información necesaria para mejorar la calidad asistencial en hospitales gestionando el gasto y los recursos adecuadamente, optimizando los presupuestos para disminuir las estancias hospitalarias con una tasa de ocupación equilibrada, mejorar los servicios preventivos y cuidados domiciliarios, reduciendo de este modo el número de muertes y mejorando la calidad de vida de los pacientes.</p>	

Abstract (in English, 250 words or less):

The objective of the present study that I carry out is to determine the European healthcare status through the statistical analysis of the **expenses** applied by different types of hospitals and countries to the various necessary care for diseases, of the **resources** used as medical personnel, nurses and researchers , as well as **activities** related to procedures such as hospital discharges in different ratios and bed occupancy rates, types of surgeries and treatments, medical technology, preventive services, causes of death and home care.

For this, data has been made available in the Eurostat repository. The tables would contain information on Health care related to hospital expenses, resources and activities.

After cleaning and combining the tables, descriptive statistical methods are applied that reflect the reality of healthcare quality in different European countries. Then, predictions are made based on different factors and patterns are obtained between the different studied countries.

After the investigation, the aim is to obtain the necessary information to improve the quality of care in hospitals, managing spending and resources appropriately, optimizing budgets to reduce hospital stays with a balanced occupancy rate, improve preventive services and home care, thus reducing the number of deaths and improving the quality of patient's life.

Índice

1. Introducción	1
1.1 Contexto y justificación del Trabajo	1
1.2 Objetivos del Trabajo	1,2
1.3 Enfoque y método seguido	2,3
1.4 Planificación del Trabajo	3,4
1.5 Breve resumen de productos obtenidos	5
1.6 Breve descripción de los otros capítulos de la memoria	5
2. Estado del Arte	6,11
3. Diseño e Implementación	
3.1 Introducción	12,13
3.2 Proceso de obtención y limpieza de datos	13,25
3.3 Análisis exploratorio de datos	25,34
3.4 Modelización de los datos	35,36
4. Conclusiones	-
5. Glosario	-
6. Bibliografía	-
7. Anexos	-

Lista de figuras y tablas

- [Figura 1.](https://agileisnow.com/2021/02/21/por-que-usar-el-tablero-kanban/) Casafranca, Johana. **La agilidad es ahora.** Recuperado de: <https://agileisnow.com/2021/02/21/por-que-usar-el-tablero-kanban/>
 - [Figura 2.](https://www.ticportal.es/glosario-tic/scrum-implementacion-proyectos) Tic. Portal. **Scrum: una manera de ejecutar los principios de la metodología Agile.** Recuperado de: <https://www.ticportal.es/glosario-tic/scrum-implementacion-proyectos>
 - [Figura 3.](#) Diagrama de Gantt.
 - [Figura 4.](https://ec.europa.eu/eurostat/statistics-explained/index.php/Health_statistics_at_regional_level#Health_status_and_health_care) Estadísticos de Salud por Regiones Europeas Eurostat Statistics Explained. **Health statistics at regional level.2021.** Recuperado de: https://ec.europa.eu/eurostat/statistics-explained/index.php/Health_statistics_at_regional_level#Health_status_and_health_care.
 - [Figura 5.](https://twitter.com/eu_health/status/933684222614753280) Cinco formas de hacer que la Sanidad sea Efectiva, Accesible y Resistente. **EU_HEALTH. State of Health in the EU 2017. 5 ways to make healthcare effective, accessible and resilient.** 2017. Recuperado de: https://twitter.com/eu_health/status/933684222614753280
 - [Figura 6.](#) Estructura de los datos.
 - [Figura 7.](#) Diagrama de flujo de la Extracción, Transformación y Carga de Datos (ETL).
 - [Figura 8.](#) Análisis Exploratorio de Datos.
 - [Figura 9.](#) EDA Regresión Logística.
 - [Figura 10.](#) EDA Contraste de Muestras.
 - [Figura 11.](#) Modelización de los Datos.
-
- [Tabla 1.](#) Descripción de las Fases del ETL.
 - [Tabla 2](#) EDA por Países y Años.
 - [Tabla 3](#) EDA Regresión logística.
 - [Tabla 4.](#) EDA Contraste de Muestras
 - [Tabla 5.](#) Modelización de Datos.

1. Introducción

1.1 Contexto y justificación del Trabajo

Lo que se pretende estudiar con este trabajo es la situación real **de la asistencia sanitaria en Europa y cómo se responde a las necesidades cotidianas en los distintos países**. Es de vital importancia poder darse cuenta si los pacientes son atendidos correctamente, si se están utilizando los recursos necesarios, o si se dispone de ellos, y si su gestión se está realizando correctamente.

En la situación actual que estamos viviendo con la pandemia de la COVID-19, se está poniendo de manifiesto las carencias de los sistemas sanitarios mundiales, evidenciando en muchos casos, no solo falta de recursos, sino una gestión deficiente de éstos.

Por este motivo, es primordial tener un plan de actuación en relación a las distintas situaciones sanitarias que pueden vivir los distintos países, saber qué recursos son los más solicitados, y cómo gestionarlos para evitar un colapso hospitalario y que todos los pacientes, puedan ser tratados en tiempo y forma.

En lo que respecta a la **motivación** para elegir esta línea de investigación, tengo que decir, que mi experiencia vivida durante muchos años en hospitales mientras me curaba de una leucemia y recibía diálisis debido a que la quimioterapia dañó mis riñones, hace que sienta cierta implicación y necesidad de compartir lo aprendido y seguir aprendiendo para poder conseguir que todos los pacientes no solo sean atendidos, sino que se haga a tiempo y en las mejores condiciones posibles. Que todo paciente sienta que está en las mejores manos, que se sienta seguro y confiado, y que se haga por él todo lo posible. Esto solo lo podremos conseguir si se tienen recursos y se gestionan bien.

1.2 Objetivos del Trabajo

- El objetivo principal es **descubrir la situación real de los hospitales** en relación con los **gastos presupuestados, recursos disponibles y actividades** y hacer una **comparativa por países**, poniendo el foco **en aquellos con más éxito**, es decir, aquellos con más recursos y con menos mortalidad prevenible o tratable, pudiendo encontrar las claves que llevarán a modificar los aspectos necesarios al resto de países. Además de establecer **predicciones** sobre el futuro de esta calidad asistencial que marcarán las pautas a seguir.

- Analizar **qué países han aplicado medidas de prevención sanitarias** a su población para la detección de enfermedades como el cáncer de mama y de colon, por ejemplo.
- Obtener **patrones por países** que ayuden a explicar el **éxito o fracaso conseguido** para modificar la gestión realizada en aquellos con peores resultados.
- Realizar **predicciones** de la calidad asistencial para poder rectificar en tiempo y forma todos los aspectos negativos que encuentre.
- Realizar **varias visualizaciones e infografía** sobre todo lo investigado para facilitar la comprensión del estudio.

1.3 Enfoque y método seguido

El método de trabajo está basado en **metodologías ágiles** que son un conjunto de prácticas innovadoras para el desarrollo de proyectos que prioriza la **agilidad y la flexibilidad** en la evolución de éstos. El tipo de metodología ágil seguido es **un tablero Kanban**.

Un tablero Kanban, es una herramienta visual que ofrece una visión general del estado actual del trabajo, y así, poder optimizar y mejorar todas las tareas planificadas para lograr los objetivos buscados.

En este tipo de tableros, se asignan **tareas o elementos colocados en columnas**. Estas columnas son una secuencia de los pasos a seguir de principio a fin en el proyecto o trabajo. Los elementos se pueden organizar por colores para facilitar conseguir los objetivos marcados.

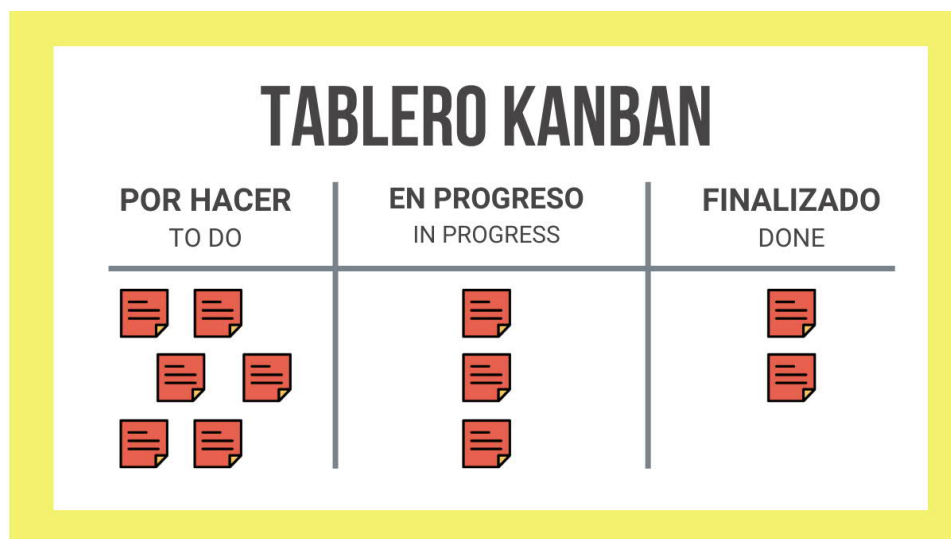


Fig. 1 Ejemplo de Tablero Kanban.

Existen otros tipos de metodologías ágiles como **SCRUM**. Ésta, se caracteriza por planificar tareas de un proyecto complejo en distintas fases con el fin de obtener resultados relativamente pronto ofreciendo un trabajo de calidad.

En SCRUM, los proyectos se ejecutan en ciclos temporales cortos y de duración fija en los que cada iteración tiene que proporcionar un resultado completo para poder ser entregado. Ésta es la razón por la que no he seguido esta metodología, ya que el proyecto desarrollado en el TFM no lo permite de ese modo.

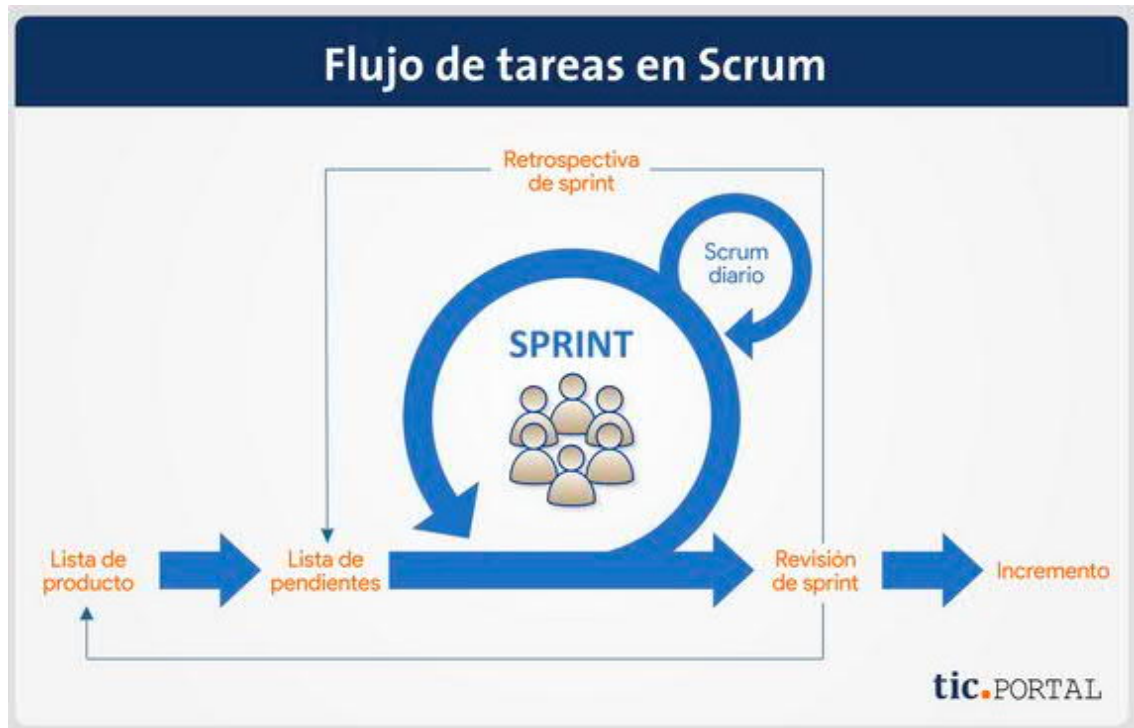


Fig.2 Flujo de Tareas Scrum

1.4 Planificación del Trabajo

La planificación total abarcaría 20 semanas dividida en varios bloques claves:

- **Definición y Planificación del trabajo Final de Máster** en la que se concretan el contexto, justificación y motivación personal. Por otro lado, se marcan los objetivos, el método y la planificación del trabajo en una línea temporal. Se ha empleado 2 semanas.
- **Estado del Arte.** En el que se plasma la investigación llevada a cabo en el pasado intentando resolver la problemática del trabajo que se está realizando. Se ha empleado 3 semanas
- **Diseño e implementación del trabajo.** En el que se diseña, desarrolla e implementa el producto. Esta fase del proyecto está dividida en varias tareas:
 - **Obtención y limpieza de datos (ETL, Extract Transform and Load).** Que abarcan 2 semanas cada una, aunque se han

simultaneado en los últimos días dedicados a estas tareas. Se estudian valores extremos, inconsistencias y valoración de posibles imputaciones.

- **Análisis de los Datos:**
 - **Análisis Descriptivo y Exploratorio de datos (EDA, Exploratory Data Analysis).** Es una de las fases a la que más tiempo se ha dedicado, 3 semanas.
 - Una vez terminado el análisis, se pasó a las siguientes fases de creación de **Modelos de Regresión** y **patrones** mediante **Machine Learning**.
- Por último, se sacaron las **conclusiones** de la investigación y el análisis
- **Redacción y presentación de la Memoria.** Que se completaron en 4 semanas y que incluye la explicación de la necesidad del proyecto, los objetivos, el estado del arte o análisis de mercado, el diseño e implementación del proyecto y sus conclusiones.

La división de tareas quedaría plasmada en el siguiente diagrama temporal:

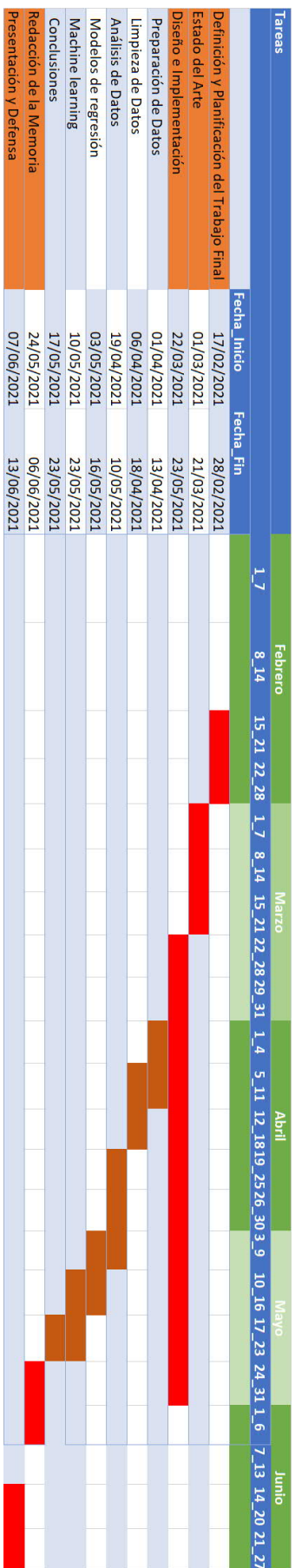


Fig. 3 Diagrama de Gantt

1.5 Breve sumario de productos obtenidos

Por detallar

1.6 Breve descripción de los otros capítulos de la memoria

Por detallar

2. Estado del Arte

La calidad asistencial sanitaria ha sido objeto de estudio en múltiples ocasiones, realizado desde ópticas distintas.

Entre los estudios a los que se hace referencia encontramos el artículo **La calidad Asistencial: concepto y medida** de María del Mar Villegas e Isabel María Rosa Díaz [\[1\]](#).

Se ha intentado definir su concepto con claridad y sobre todo descubrir cuál es la medida óptima que acredite si la calidad es buena o mala, pero **nunca existe un único criterio** que nos ayude a medir la calidad de la atención al paciente.

En general, la calidad asistencial responde a la interacción de un conjunto de dimensiones que son:

- ✓ **Científico-Técnicas** en relación con el correcto diagnóstico, tratamiento y destreza médicas.
- ✓ **Interpersonales.** Relacionadas con:
 - La confianza y comunicación entre médico y paciente.
 - Empatía, teniendo especial cuidado en el tacto a la hora de tratar con el paciente.
 - Trato también a familiares.
- ✓ **Ambientales**, entorno, relacionadas con el confort, servicios prestados, intimidad en los ingresos...
- ✓ **Sociales.** Poniendo el foco en toda la infraestructura necesaria para que familias socialmente desfavorecidas reciban los tratamientos que necesitan con las relaciones interpersonales óptimas.
- ✓ **Económicas.**

La dimensión científico-técnica es algo que se da por hecho, ser atendido y recibir el tratamiento adecuado para la resolución del problema médico.

Sin embargo, en los últimos años, **las expectativas y las necesidades de los pacientes están jugando un papel muy importante** a la hora de valorar la calidad de la asistencia sanitaria.

El punto de vista del paciente está siendo considerado como un factor esencial hasta tal punto que está produciendo cambios en los sistemas sanitarios regionales o nacionales. Por ejemplo, el plan de Calidad de la Conserjería de salud de la junta de Andalucía explica que *“El ciudadano es el centro del sistema sanitario público, por lo tanto, la satisfacción de sus necesidades, demandas y expectativas se convierten en objetivos fundamentales en toda la política sanitaria andaluza”*.

Aunque hay que destacar, que **todavía las opiniones de los pacientes no provocan cambios realmente importantes** en los aspectos organizativos de nuestro sistema de salud.

Del mismo modo, en la investigación realizada por Miguel López-Coronado, Isabel de la Torre, Jesús Herreros y Javier Cabo se evidencia una mejora de la calidad asistencial mediante la **telemedicina y la teleasistencia** debido en muchos casos a la falta no solo de personal especializado sino además de su disponibilidad [2].

Por otro lado, en el artículo “La aplicación del modelo europeo de gestión de calidad Total al Sector Sanitario” [3], se explica cómo la **EFQM (The European Foundation for Quality Management)** ha elaborado un modelo que es entendido como un conjunto de procesos de mejora en los que todo el personal sanitario está involucrado con el objetivo de obtener la calidad asistencial deseada. Actualmente existen dos formas distintas de enfocarlo:

- ✓ **Monitorización** de dimensiones como la mortalidad, cirugía inapropiada, tratamientos inadecuados etc....
- ✓ Y lo que se llama **Individualización**, que pone el foco en las necesidades de los pacientes y/o profesionales. Aunque sigue siendo **más frecuente el enfoque en la calidad científico-técnico**, las expectativas de los pacientes y profesionales se abren camino rápida y eficazmente.

Es decir, para la gestión de la calidad total de la EFQM se ha elaborado un modelo que se basa en la siguiente premisa: *“la satisfacción del cliente, de los empleados y el impacto positivo en la sociedad se consiguen mediante el liderazgo en la política y estrategia, gestión de personal, el uso eficiente de los recursos y una adecuada definición de los procesos”*. Esto es, constituye una herramienta que posibilita a cualquier organización ya sea sanitaria o no, orientar su gestión hacia la **satisfacción del cliente**.

Y para finalizar con una aplicación específica, el estudio sobre la **evaluación y seguimiento de la calidad asistencial, ofrecida a los intoxicados en los Servicios de Urgencias** [4], revelaron que, con la disponibilidad de un protocolo específico de tratamiento de las intoxicaciones, de material adecuado para llevar a cabo estos procesos como sondas especiales, stock de antídotos suficientes y realización de análisis específicos, se obtuvieron resultados satisfactorios. Por lo que concluimos que la **existencia de protocolos y recursos, redundan en una mejora de la calidad científico- técnicas** y en este estudio en concreto en la **calidad de satisfacción del paciente medida por el número de reclamaciones** que se hicieron que no superó la cifra general de las reclamaciones del Servicio de Urgencias.

En los diferentes artículos mencionados hasta ahora, se comprueba que la calidad asistencial sanitaria no se puede medir o valorar únicamente por un único criterio, aunque si existe un foco común en los procesos de mejora que se llevan a cabo, la **satisfacción del paciente**.

Hasta ahora se ha analizado qué variables se tienen en cuenta para valorar la calidad asistencial sanitaria ofrecida a los pacientes y cómo la satisfacción de éstos es el foco de los procesos de mejora que se llevan a cabo.

Para medir la satisfacción de los pacientes se tienen en cuenta 5 puntos básicos, recogidos en el artículo **Health Statistics at Regional Level** [5]:

- Estado físico.
- Estado Emocional.
- Situación Social.
- Síntomas
- Medidas biométricas.

El estudio arroja los siguientes datos:

- Casi el 70% de la población adulta europea percibe su salud como buena o muy buena, encontrando diferencias por edades y por regiones (zonas rurales o urbanas), siendo la población adulta más joven que vive en las ciudades la que se considera en mejor estado de salud.
- La población de las zonas rurales no accede de forma satisfactoria a los servicios sanitarios bien porque son caros, por la imposibilidad de desplazarse a los hospitales por largas distancias o por falta de medios de transporte, o debido a las largas lista de espera para la realización de ciertas pruebas médicas que disuaden a los pacientes de realizárselas.
- Como media, en 2017 había 1 médico por cada 266 habitantes.

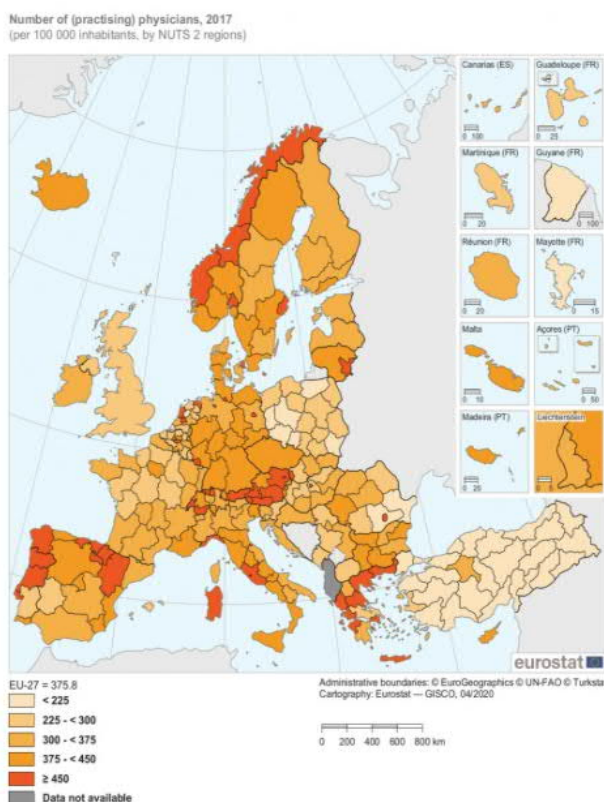
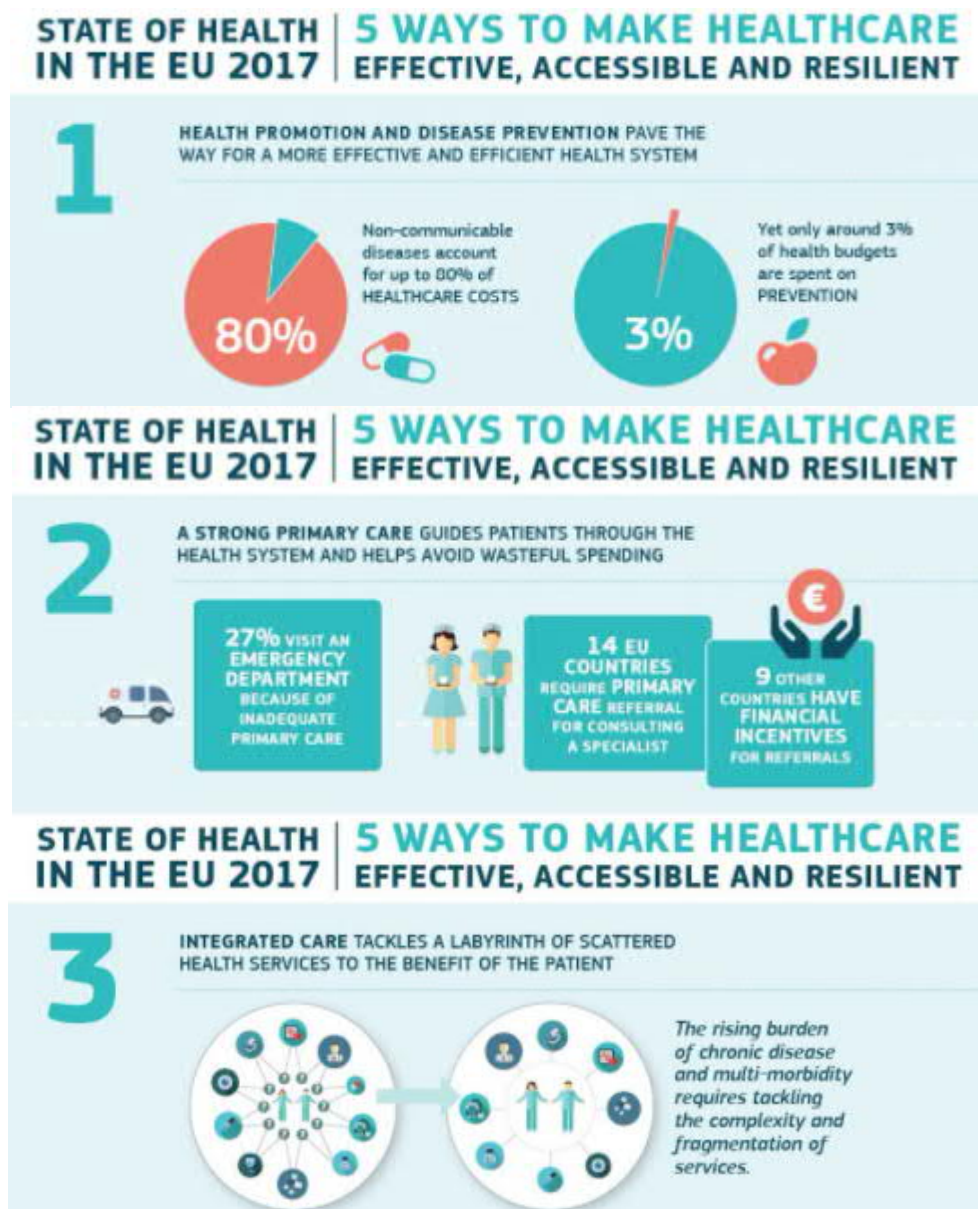


Fig. 4. Estadísticos de Salud por Regiones Europeas

De entre los procesos de mejora, cabe mencionar las recomendaciones de la **Comisión Europea para la Salud y Seguridad Alimentaria (SANTE, EU Commission's DG Health & Food Safety)** que encontramos resumidas en los siguientes gráficos [\[6\]](#):

- Promover la Prevención.
- Fortalecer la Asistencia Médica primaria en los Centros de Salud.
- Fragmentación de los Servicios Sanitarios en beneficio del paciente.
- Previsión y Planificación proactiva.
- Base de datos centralizada de los pacientes.



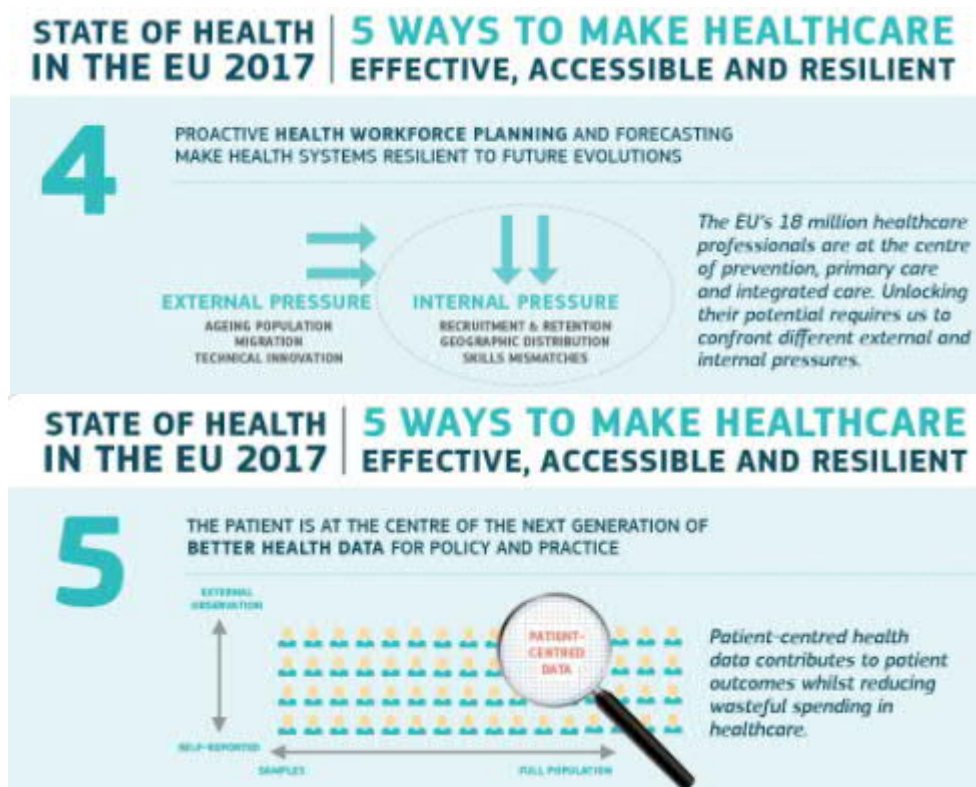


Fig. 5 Cinco formas de hacer que la Sanidad sea Efectiva, Accesible y Resistente.

Y para finalizar, a modo de resumen, se describen las instituciones que integran los Sistemas Sanitarios de varios Países Europeos y se comparan con el establecido en USA. **Howard SJ y Davison P** lo resumen en el artículo: **Comparing Health protection systems across six European nations and the USA** [7].

- **Finlandia.** El Sistema de Salud se integra en el Instituto Nacional de Salud y Bienestar (**THL**) financiado en parte por empresas privadas y farmacéuticas, con un modelo que integra la investigación y la práctica.
- **Francia.** Como consecuencia de las reformas en el sistema de protección de la Salud, se establece una nueva organización llamada **l'institute** que engloba diferentes instituciones. Además, existen 26 Agencias Regionales de Salud.
- **Alemania.** El instituto **Robert Koch (RKI)** como responsable del control y la prevención de enfermedades, la **Oficina Federal de Estadísticas (SB)** con funciones epidemiológicas y de vigilancia, el **instituto Paul Ehrlich (PEI)** con responsabilidad en los programas de vacunación integran el sistema de protección para la salud alemán.
- **Hungría.** La responsabilidad nacional de protección de la salud recae en el director médico, que supervisa ocho institutos nacionales de salud y salud pública dentro de una organización conocida como **ÁNTSZ**. Cabe destacar que en Hungría no hay un programa obligatorio de vacunación.
- **Países Bajos.** Las funciones para la protección de la salud a nivel regional son realizadas por los **servicios municipales**. Sin embargo, a

nivel nacional, es el **Centro para el Control de Enfermedades Infecciosas (Cib)** el encargado de estas funciones.

- **USA.** La protección de la salud se proporciona a nivel de nacional en los Centros para la Prevención y el Control de Enfermedades (CDC) que se organizan en Oficinas (Ej. Oficina de Enfermedades Infecciosas) y subdividen en Centros (Ej. Centro nacional de Enfermedades infecciosas Emergentes y Zoonóticas).

3. Diseño e implementación del trabajo.

3.1 Introducción

La implementación del trabajo cumple las siguientes especificaciones de **diseño**:

Se deberán cumplir los objetivos marcados, es decir principalmente **descubrir la situación real de los hospitales** en relación con los **gastos presupuestados, recursos disponibles y actividades** y hacer una **comparativa por países**.

Para ello, se obtiene la información necesaria para mejorar la calidad asistencial en hospitales, gestionando los recursos adecuadamente y de este modo mejorar la calidad de vida de los pacientes.

De entre los programas y lenguajes utilizados tanto en el proceso de extracción, transformación y carga de datos (ETL), así como el Análisis Exploratorio de los Datos (EDA) y la aplicación de Modelos, destacamos:

- **RStudio** en lenguaje **R**.
- **Anaconda** en lenguaje **Python**.
- Por otro lado, se ha utilizado el repositorio **GitHub** para almacenar, los datos, y los ficheros con el código en R tanto del proceso ETL como EDA y Modelización.
- **Tableau** para la realización de visualizaciones.

La estructura de los datos es la siguiente:

ESTRUCTURA DE LOS DATOS					
1. Bloque Gasto Sanitario (4 archivos)	<ul style="list-style-type: none">• Financiación del Gasto Sanitario.• Gasto Sanitario por Función (cuidados curativos, rehabilitación...).• Gasto Sanitario por proveedor (destinatario del gasto sanitario).• Ingresos en el Sector Sanitario por Países y Organizaciones.				
	2. Bloque Recursos Sanitarios <ul style="list-style-type: none">•(Staff: 5 archivos,•Facilities: 4 archivos)	<table><tr><td>Staff (Personal Hospitalario)</td><td>Facilities (Recursos Auxiliares, que facilitan las actividades hospitalarias)</td></tr><tr><td><ul style="list-style-type: none">• Médicos por especialidad</td><td><ul style="list-style-type: none">• N.º de camas por propiedad.• Recursos</td></tr></table>	Staff (Personal Hospitalario)	Facilities (Recursos Auxiliares, que facilitan las actividades hospitalarias)	<ul style="list-style-type: none">• Médicos por especialidad
Staff (Personal Hospitalario)	Facilities (Recursos Auxiliares, que facilitan las actividades hospitalarias)				
<ul style="list-style-type: none">• Médicos por especialidad	<ul style="list-style-type: none">• N.º de camas por propiedad.• Recursos				

	<ul style="list-style-type: none"> • Médicos por Países. • Personal de Enfermería y Cuidados. • Personal Sanitario Hospitalario. • Personal Sanitario Hospitalario exceptuando Enfermería. 	<p>Técnicos Hospitalarios.</p> <ul style="list-style-type: none"> • Tecnología Médica. • Tipos de Camas Hospitalarias.
3. Bloque de Actividades (4 archivos)	<ul style="list-style-type: none"> • Altas, Media de Estancia Hospitalaria. • Ocupación de camas. • Pacientes Trasplantados y en Diálisis. • Ratio de Altas en el Hospital de Día por Diagnóstico. 	
4. Bloque de Prevención (3 archivos)	<ul style="list-style-type: none"> • Detección de Cáncer de Mama y Cérvix Uterino. • Detección de Enfermedades Cardiovasculares y Diabetes. • Vacunación a la población mayor de 65 años. 	
5. Bloque de Causas de Mortalidad (2 archivos)	<ul style="list-style-type: none"> • Mortalidad Tratable y Prevenible • Muertes por Enfermedades Infecciosas. 	
6. Cuidados Domiciliarios (1 archivo)	<ul style="list-style-type: none"> • Cuidados Domiciliarios. 	
7. Estado de Salud (1 archivo)	<ul style="list-style-type: none"> • Estado de Salud. 	

Fig. 6 Estructura de los Datos

3.2 Proceso de obtención y Limpieza de Datos (ETL, Extract, Transform, Load).

El diagrama de flujo del proceso ETL es: (fases y tareas de cada bloque):

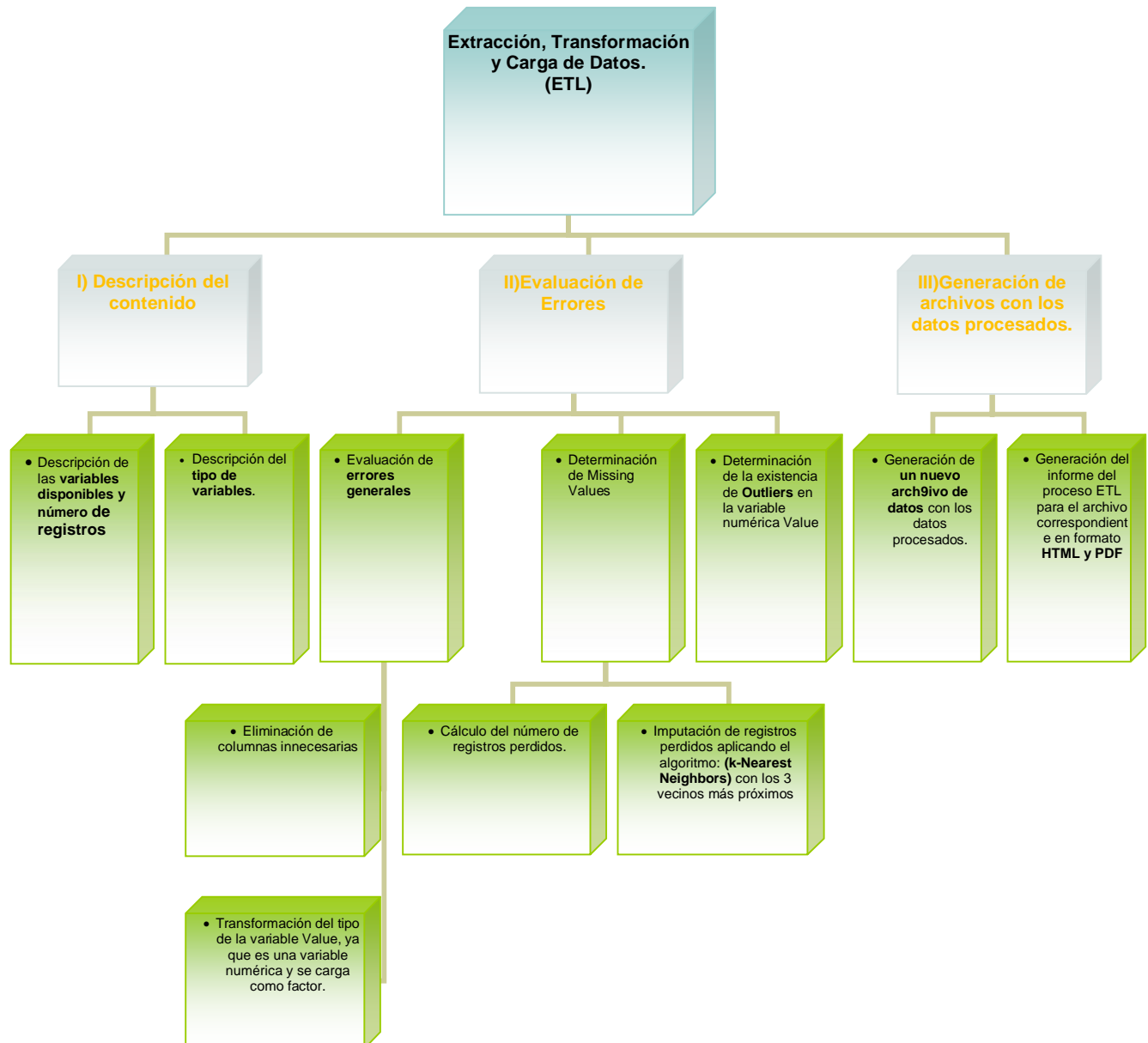


Fig. 7 Diagrama de flujo de la Extracción, Transformación y Carga de Datos (ETL)

La descripción de las fases de forma genérica:

DESCRIPCIÓN DE LAS FASES DEL ETL		
Descripción del contenido	Evaluación de errores	Generación de archivos con los datos procesados.
<ul style="list-style-type: none"> • Descripción de las variables disponibles y del número de registros. 	<ul style="list-style-type: none"> • Evaluación de errores generales. Para todos los archivos se seguirán las mismas acciones. <ul style="list-style-type: none"> ▪ Se eliminan las columnas innecesarias como Flag.and.Footnotes. ▪ Se modifica el tipo de dato de la variable "Value" 	<ul style="list-style-type: none"> • Se genera el archivo resultante del proceso de ETL
<ul style="list-style-type: none"> • Descripción del tipo de variables. Estructura común de los archivos: <ul style="list-style-type: none"> ▪ TIME: variable cuantitativa. Indica el año en el que se ha realizado la medida, en este caso el valor de la variable "Value". Se carga bien como número entero. ▪ GEO: variable cualitativa. Indica el país o región en el que se ha realizado la medida. Se carga bien como factor. ▪ UNIT: variable cualitativa. Indica la medida de la variable valor. Se carga bien como factor. ▪ Value: Variable cuantitativa. Indica el número del recurso en cuestión. Es una variable numérica, pero se carga como factor. Se hace transformación. ▪ Fal.and.footnotes: Notas sobre etiquetas. Eliminamos esta columna. 	<ul style="list-style-type: none"> • Determinación de Datos faltantes o Missing Values <ul style="list-style-type: none"> ▪ Se comprueba para cada archivo los registros que faltan. ▪ Se ha utilizado el algoritmo KNN por la gran versatilidad que presenta para estimar todo tipo de patrones de valores perdidos ya sean continuos o discretos. Funciona del siguiente modo: Dado un conjunto de datos incompleto, se seleccionan los K casos más cercanos del conjunto de datos conocidos que estén más cercanos al valor desconocido de tal forma que minimicemos una métrica de distancia. 	<ul style="list-style-type: none"> • Se genera el informe con formato HTLM y PDF del código usado.

<ul style="list-style-type: none"> •Además, se describen las variables específicas de cada archivo en cuestión. 	<ul style="list-style-type: none"> •Comprobación de la existencia de outliers para cada archivo. 	
--	--	--

[Tabla.1](#) Descripción de las fases del ETL

Descripción de cada bloque de datos en relación con las fases del proceso ETL: (los puntos genéricos están descritos en la [tabla1](#):

1. BLOQUE GASTO SANITARIO

1) Financiación del Gasto Sanitario.

(i) Descripción del contenido.

- **ICHA11_HF**: variable cualitativa. Indica el organismo de financiación, ya sea gubernamental o por seguros privados etc.

(ii) Evaluación de Errores

1. **Evaluación de Errores Generales.**
2. **Determinación de Datos faltantes o Missing Values.**
Se observan **513** valores perdidos de 2000 disponibles.
3. Se comprueba que existen **outliers** en los valores de la variable "**Value**".

(iii) Generación de archivos con la información corregida.

1. Se genera el archivo resultado del proceso de los datos: **GastoSanitario_Financiacion_clean.csv**

2) Gasto Sanitario en relación con la Función

(i) Descripción del contenido:

- **ICHA11_HC**: variable cualitativa. Indica cómo se aplica el gasto sanitario por función.

(ii) Evaluación de errores

1. **Evaluación de Errores Generales.**
2. **Determinación de Datos faltantes o Missing Values.**
Se observan **595** valores perdidos de 2000 disponibles.
3. Se comprueba que existen **outliers** en los valores de la variable "**Value**".

(iii) Generación de archivos con la información corregida.

1. Se genera el archivo resultado del proceso de los datos: **GastoSanitario_Funcion_clean.csv**.

3) Gasto Sanitario en relación con la Entidad sobre la que se aplica el gasto sanitario.

(i) Descripción del contenido:

- **ICHA11_HP**: variable cualitativa. Indica la entidad sobre la que se aplica el gasto sanitario.

(ii) Evaluación de Errores.

1. Evaluación de Errores Generales.
2. Determinación de Datos faltantes o Missing Values.
 - Se observan **694** valores perdidos de 2000.
3. Se comprueba que existen **outliers** en los valores de la variable “Value”.

(iii) Generación de archivos con la información corregida.

1. Se genera el archivo resultado del proceso de los datos: **GastoSanitario_Proveedor_clean.csv**.

4) Ingresos Sanitarios por países.

(i) Descripción del contenido:

- **ICHA11_FS**: variable cualitativa. Indica los ingresos por países.

(ii) Evaluación de Errores.

1. Evaluación de Errores Generales
2. Determinación de Datos faltantes o Missing Values.
 - Se observan **62** valores perdidos de 220.
3. Se comprueba que existen **outliers** en los valores de la variable “Value”.

(iii) Generación de archivos con la información corregida.

- Se genera el archivo resultado del proceso de los datos:
IngresosSanitarios_Financiacion_clean.csv.

2. BLOQUE RECURSOS SANITARIOS

STAFF

1) Médicos por Edad y Países.

(i) Descripción del contenido

- **AGE**: Variable cualitativa. Indica el rango de edad del número de médicos disponibles por países. Variable factor.
- **SEX**: Variable cualitativa. Solo un valor “Total”. Variable total.

(ii) Evaluación de Errores.

1. Evaluación de Errores Generales.

2. **Determinación de Datos faltantes o Missing Values.**
 - Se observan **419** valores perdidos de 1900.
3. Comprobamos, que la variable “Value” **tiene valores extremos.**

(iii) Generación de archivos con la información corregida.

- Se genera el archivo resultado del proceso de los datos: **Medicos_Por_Paises_clean.csv.**

2) Médicos por Especialidad.

(iv) Descripción del contenido

- **MED_SPEC:** Variable cualitativa. Indica la especialidad médica. Variable factor.

(v) Evaluación de Errores.

1. **Evaluación de Errores Generales.**
2. **Determinación de Datos faltantes o Missing Values.**
 - Observamos **390** valores perdidos de 1800.
3. Comprobamos, que la variable “Value” **tiene valores extremos.**

(vi) Generación de archivos con la información corregida.

- Se genera el fichero limpio: **Medicos_x_especialidad_clean.csv.**

3) Personal Sanitario Hospitalario.

(i) Descripción del contenido

- **ISCO08:** Variable cualitativa. Indica el tipo de personal al que se refiere la variable “Value”, esto es, enfermeros/as, personal de cuidados, matronas, Médicos, Matronas, o Personal hospitalario.
- **WSTATUS:** Variable cualitativa. Indica el estatus laboral del personal sanitario hospitalario.

(ii) Evaluación de Errores.

1. **Evaluación de Errores Generales.**
2. **Determinación de Datos faltantes o Missing Values.**
 - Se observan **883** valores perdidos de 1800.
3. Se comprueba, que la variable “Value” **tiene valores extremos.**

(iii) Generación de archivos con la información corregida.

- Se genera el archivo resultado del proceso de los datos: **PersonalSanitarioHospitalario_clean.csv.**

4) Personal de Enfermería y Cuidados.

(i) Descripción del contenido

- **WSTATUS:** Variable cualitativa. Indica el estatus laboral del personal de Enfermería y Cuidados.

- **ISCO08:** Variable cualitativa. Indica el tipo de personal al que se refiere la variable “Value”, esto es, enfermeros/as, personal de cuidados, matronas...

(ii) Evaluación de Errores.

1. **Evaluación de Errores Generales.**
2. **Determinación de Datos faltantes o Missing Values.**
 - Se observan **9347 valores perdidos de 17100.**
3. Se comprueba, que la variable “Value” **tiene valores extremos.**

(iii) Generación de archivos con la información corregida.

- Se genera el archivo resultado del proceso de los datos:
Personal_Enfermeria_Cuidados_clean.csv.

5) Personal Sanitario Exceptuando Enfermería.

(i) Descripción del contenido

- **WSTATUS:** Variable cualitativa. Indica el estatus laboral del personal sanitario (Exceptuando Enfermería).
- **ISCO08:** Variable cualitativa. Indica el tipo de personal al que se refiere la variable “Value”, esto es, Médicos, Dentistas, Fisioterapeutas o Farmacéuticos.

(ii) Evaluación de Errores.

1. **Evaluación de Errores Generales.**
2. **Determinación de Datos faltantes o Missing Values.**
 - Se observan **847 valores perdidos de 1600.**
3. Se comprueba, que la variable “Value” **tiene valores extremos.**

(iii) Generación de archivos con la información corregida.

- Se genera el archivo resultado del proceso de los datos:
PersonalSanitario_No_enfermeria_clean.csv.

FACILITIES

1) Número/Ratio de Camas por Propiedad y Países.

(i) Descripción del contenido

- **OWNER:** Variable cualitativa. Indica la propiedad de las camas hospitalarias.

(ii) Evaluación de Errores.

1. **Evaluación de Errores Generales.**
2. **Determinación de Datos faltantes o Missing Values.**
 - Se observan **786 valores perdidos de 2790.**
3. Se comprueba, que la variable “Value” **tiene valores extremos.**

(iii) Generación de archivos con la información corregida.

- Se genera el archivo resultado del proceso de los datos: **Camas_Propiedad_clean.csv**.

2) Recursos Técnicos Hospitalarios.

(i) Descripción del contenido

- **FACILITY**: Variable cualitativa. Indica el tipo de recurso hospitalario.

(ii) Evaluación de Errores.

1. Evaluación de Errores Generales.
2. Determinación de Datos faltantes o Missing Values.
 - Se observan **1030 valores perdidos de 1860**.
3. Se comprueba, que la variable “Value” tiene valores extremos.

(iii) Generación de archivos con la información corregida.

- Se genera el archivo resultado del proceso de los datos:
RecursosTecnicos_hospitalarios_clean.csv.

3) Tecnología Médica.

(i) Descripción del contenido

- **FACILITY**: Variable cualitativa. Indica el tipo de Tecnología Médica.
- **ICHA_HP**: Variable cualitativa. Indica dónde se usa el recurso tecnológico: Hospitales, Ambulatorios, etc....

(ii) Evaluación de Errores.

1. Evaluación de Errores Generales.
2. Determinación de Datos faltantes o Missing Values.
 - Se observan **7005 valores perdidos de 15750**.
3. Se comprueba, que la variable “Value” tiene valores extremos.

(iii) Generación de archivos con la información corregida.

- Se genera el fichero con los datos procesados:
TecnologiaMedica_clean.csv.

4) Tipos de camas hospitalarias.

(i) Descripción del contenido

- **FACILITY**: Variable cualitativa. Indica el tipo de cama hospitalaria.

(ii) Evaluación de Errores.

1. Evaluación de Errores Generales.
2. Determinación de Datos faltantes o Missing Values.
 - Se observan **1604 valores perdidos de 5850**.
3. Se comprueba, que la variable “Value” tiene valores extremos.

(iii) Generación de archivos con la información corregida.

- Se genera el fichero con los datos procesados:
TiposCamasHospitalarias_clean.csv.

3. ACTIVIDADES

1) Altas, Estancias y Camas Hospitalarias en Pacientes Ingresados y por Servicios Curativos.

(i) Descripción del contenido

- **OWNER:** Variable cualitativa. Indica la propiedad de las camas hospitalarias.

(ii) Evaluación de Errores.

1. Evaluación de Errores Generales.
2. Determinación de Datos faltantes o Missing Values.
 - Se observan 786 valores perdidos de 2790.
3. Se comprueba, que la variable "Value" **tiene valores extremos.**

(iii) Generación de archivos con la información corregida.

- Se genera el fichero con los datos procesados: **Camas_Propiedad_clean.csv.**

2) Recursos Técnicos Hospitalarios.

(i) Descripción del contenido

- **FACILITY:** Variable cualitativa. Indica el tipo de recurso hospitalario.

(ii) Evaluación de Errores.

1. Evaluación de Errores Generales.
2. Determinación de Datos faltantes o Missing Values.
 - Se observan **1030 valores perdidos de 1860.**
3. Se comprueba, que la variable "Value" **tiene valores extremos.**

(iii) Generación de archivos con la información corregida.

- Se genera el fichero con los datos procesados: **RecursosTecnicos_hospitalarios_clean.csv.**

3) Tecnología Médica.

(i) Descripción del contenido

- **FACILITY:** Variable cualitativa. Indica el tipo de Tecnología Médica.
- **ICHA_HP:** Variable cualitativa. Indica dónde se usa el recurso tecnológico: Hospitales, Ambulatorios, etc....

(ii) Evaluación de Errores.

1. Evaluación de Errores Generales.
2. Determinación de Datos faltantes o Missing Values.
 - Se observan **7005 valores perdidos de 15750.**
3. Comprobamos, que la variable "Value" **tiene valores extremos.**

(iii) Generación de archivos con la información corregida.

- Se genera el fichero con los datos procesados: **TecnologiaMedica_clean.csv.**

4) Tipos de camas hospitalarias.

- **FACILITY:** Variable cualitativa. Indica el tipo de cama hospitalaria.

(iv) Evaluación de Errores.

- 1. Evaluación de Errores Generales.**
- 2. Determinación de Datos faltantes o Missing Values.**
 - Se observan **1604 valores perdidos de 5850.**
- 3. Comprobamos, que la variable “Value” tiene valores extremos.**

(v) Generación de archivos con la información corregida.

- Se genera el fichero con los datos procesados: **TiposCamasHospitalarias_clean.csv.**

4. PREVENCIÓN

1) Detección del Cáncer de Mama y Cérvix Uterino.

(i) Descripción del contenido

- **SOURCE:** Variable cualitativa. Indica la fuente del estudio.
- **ICD10:** Indica el tipo de cáncer foco de estudio: De mama o Cérvix Uterino.

(ii) Evaluación de Errores.

- 1. Evaluación de Errores Generales.**
- 2. Determinación de Datos faltantes o Missing Values.**
 - Se observan 915 valores perdidos de 1440.
- 3. Se comprueba, que la variable “Value” tiene valores extremos.**

(iii) Generación de archivos con la información corregida.

- Se genera el fichero con los datos procesados: **Deteccion_Cancer_Mama_Cervix_clean.csv.**

2) Detección de Enfermedades Cardiovasculares y Diabetes.

(i) Descripción del contenido

- **HLTHCARE:** Variable cualitativa. Indica el tipo de medida.
- **ISCED11:** Variable cualitativa. Estándar en estadísticas de educación en el que se hacen las mediciones.
- **DURATION:** Variable cualitativa. Indica el tiempo durante el cual se hace la medición.
- **SEX:** Variable cualitativa. En cómputo total. No hay distinción entre sexos.
- **AGE:** Variable cualitativa. En cómputo total. No hay distinción entre edades.

(ii) Evaluación de Errores.

1. **Evaluación de Errores Generales.**
2. **Determinación de Datos faltantes o Missing Values.**
 - No se observan valores perdidos.
3. Se comprueba, que la variable “Value” **no tiene valores extremos.**

(iii) Generación de archivos con la información corregida.

- Se genera el fichero con los datos procesados:
Deteccion_enfCardio_Diabetes_clean.csv.

3) Vacunación a la población mayor de 65 años.

(i) Descripción del contenido

- Variables comunes descritas al inicio.

(ii) Evaluación de Errores.

1. **Evaluación de Errores Generales.**
2. **Determinación de Datos faltantes o Missing Values.**
 - Se observan **119 valores perdidos de 380.**
3. Se comprueba, que la variable “Value” **no tiene valores extremos.**

(iii) Generación de archivos con la información corregida.

- Se genera el fichero con los datos procesados:
Vacunacion_+65_clean.csv.

5. CAUSAS DE MUERTE

1) Mortalidad Tratable y Prevenible.

(i) Descripción del contenido

- **MORTALIT:** Variable cualitativa. Indica el tipo de mortalidad: Tratable o Prevenible.
- **ICD10:** Variable Cualitativa. En la clasificación de enfermedades, se indica el total de ellas.
- **SEX:** Variable cualitativa. Indica el sexo de la población estudiada.

(ii) Evaluación de Errores.

1. **Evaluación de Errores Generales.**
2. **Determinación de Datos faltantes o Missing Values.**
 - Se observan 54 valores perdidos de 4536.
3. Se comprueba, que la variable “Value” **tiene valores extremos.**

(iii) Generación de archivos con la información corregida.

- Se genera el fichero con los datos procesados:
Mortalidad_Tratable_Prevenible_clean.csv.

2) Mortalidad por Enfermedades infecciosas.

(i) Descripción del contenido

- **ICD10:** Variable Cualitativa. Indica la clasificación de enfermedades infecciosas que se han estudiado en la población.
- **SEX:** Variable cualitativa. Indica el sexo de la población estudiada. Total (No hay diferencia entre sexos).

(ii) Evaluación de Errores.

1. **Evaluación de Errores Generales.**
2. **Determinación de Datos faltantes o Missing Values.**
 - Se observan 156 valores perdidos de 1332.
3. Se comprueba, que la variable “Value” **tiene valores extremos.**

(iii) Generación de archivos con la información corregida.

- Se genera el fichero con los datos procesados: **Deteccion_enfCardio_Diabetes_clean.csv.**

6. BLOQUE CUIDADOS DOMICILIARIOS

1) Cuidados Domiciliarios.

(i) Descripción del contenido.

- **ISCED11:** variable cualitativa. Indica estándar en estadísticas de educación en el que se hacen las mediciones.
- **SEX:** Variable cualitativa. Indica el sexo de la población estudiada.
- **AGE.** Variable cualitativa. En cómputo total. No hay distinción entre edades.

(ii) Evaluación de Errores

1. **Evaluación de Errores Generales.**
2. **Determinación de Datos faltantes o Missing Values.**
 - Se observan 12 valores perdidos de 396.
3. Se comprueba que existen **outliers** en los valores de la variable “Value”.

(iii) Generación de archivos con la información corregida.

1. Se genera el fichero con los datos procesados: **CuidadosDomiciliarios_clean.csv**

7. BLOQUE ESTADO DE SALUD

1) Estado de salud por Sexo y por Países.

(i) Descripción del contenido.

- **INDIC_HE:** variable cualitativa. Explica el valor de la variable “Value”, esto es, años de vida sana en valores absolutos desde el nacimiento.
- **SEX:** Variable cualitativa. Indica el sexo de la población estudiada.

(ii) Evaluación de Errores

1. **Evaluación de Errores Generales.**
2. **Determinación de Datos faltantes o Missing Values.**
 - Se observan 18 valores perdidos de 990.
3. Se comprueba que existen **outliers** en los valores de la variable **“Value”**.

(iii) Generación de archivos con la información corregida.

1. Se genera el fichero con los datos procesados:
EstadoDeSalud_Sexo_clean.csv

3.3 Análisis Exploratorio de datos (EDA, Exploratory Data Analysis).

I. Estudio de los distintos aspectos de la calidad sanitaria por países.

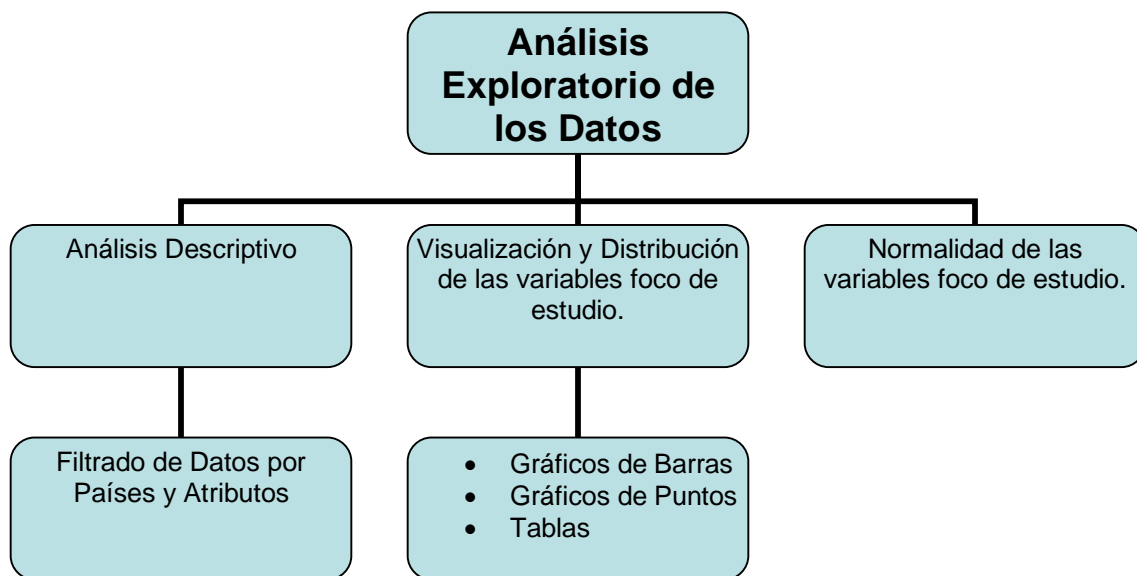


Fig. 8 Análisis Exploratorio de Datos.

ANÁLISIS EXPLORATORIO POR PAÍSES		
Análisis Descriptivo	Visualización y Distribución de Value	Estudio de la Normalidad de la Variable “Value”
<ul style="list-style-type: none"> • Descripción de las variables disponibles y del número de registros. 	<ul style="list-style-type: none"> • Generación de Gráficos de barras en relación con los años (Variable TIME). Para todos los archivos se seguirán las mismas acciones. <ul style="list-style-type: none"> ▪ Se utiliza la librería ggplot2 para generarlos. 	<p>Para cada archivo se realizan las siguientes acciones:</p> <ul style="list-style-type: none"> • Descripción de forma visual la distribución de la variable en cuestión utilizando gráficas de Densidad y Curvas Q-Q.

<ul style="list-style-type: none"> • Filtrado de la información, seleccionando los países foco de estudio 	<ul style="list-style-type: none"> • Generación de Gráficos de puntos en relación con los Países (Variable Pais). <ul style="list-style-type: none"> ▪ Se utiliza la librería ggplot2 para generarlos. 	<ul style="list-style-type: none"> • Se realizan pruebas de hipótesis siendo: <ul style="list-style-type: none"> ○ H0: La muestra proviene de población normal ○ H1: la muestra no proviene de una población normal. • Se confirma normalidad aplicando el test Shapiro-Wilk.
<ul style="list-style-type: none"> • Filtrado particular de cada archivo en relación con la información foco de estudio. 	<ul style="list-style-type: none"> • Obtención de tablas con los 5 países con mayor recuento en el aspecto considerado en cuestión, como gasto sanitario por años de forma individual. • Obtención de tablas con los 5 países con mayor recuento en el aspecto considerado en cuestión, como gasto sanitario por rango de años (cómputo Global de la información disponible) 	

Tabla 2 EDA por Países y Años

I. Análisis Descriptivo.

- Filtrado de la información, **seleccionando los países foco de estudio** (Se eliminan los campos que hacen referencia a los países de la Unión Europea de forma global en un período de tiempo determinado)
 - European Union - 27 countries (from 2020)
 - European Union - 28 countries (2013-2020)
 - European Union - 27 countries (2007-2013)
 - European Union - 15 countries (1995-2004)
 - Euro area - 19 countries (from 2015)
 - Euro area - 18 countries (2014)
 - Euro area - 12 countries (2001-2006)
- Filtrado particular de cada archivo en relación con la información foco de estudio. (Diferentes factores en las variables a considerar).

Por ejemplo, en gasto por proveedor, se selecciona la información relativa al gasto destinado a Hospitales, Hospitales Mentales y Hospitales Especiales.

II. Visualización y Distribución de “Value”

- **Value:** Variable que en cada archivo nos indica el valor del aspecto considerado como por ejemplo gasto sanitario, etc.
- Generación de GRÁFICOS DE BARRAS en relación con los años (variable **TIME**).
- Generación de GRÁFICOS DE PUNTOS en relación con los Países Europeos.
- Obtención de tablas con los **5 países** con mayor recuento en el aspecto considerado en cuestión, como gasto sanitario **por años de forma individual**.
- Obtención de tablas con los **5 países** con mayor recuento en el aspecto considerado en cuestión, como gasto sanitario **por rango de años (cómputo Global de la información disponible)**

III. Estudio de la Normalidad de “Value”.

- Para cada archivo se realizan las siguientes acciones:
- Descripción de forma visual la distribución de la variable en cuestión utilizando gráficas de **Densidad y Curvas Q-Q**.
- Se realizan pruebas de hipótesis siendo:
 - **HO (Hipótesis Nula):** La muestra proviene de población normal
 - **H1 (Hipótesis Alternativa):** la muestra no proviene de una población normal.
 - Se confirma normalidad aplicando el test **Shapiro-Wilk**.

IV. Código y Gráficos.

- Gráficos y código disponibles en: Enlace a GitHub. [Enlace a GitHub](#).

II. Análisis Exploratorio de Datos en Europa: Regresión Logística.

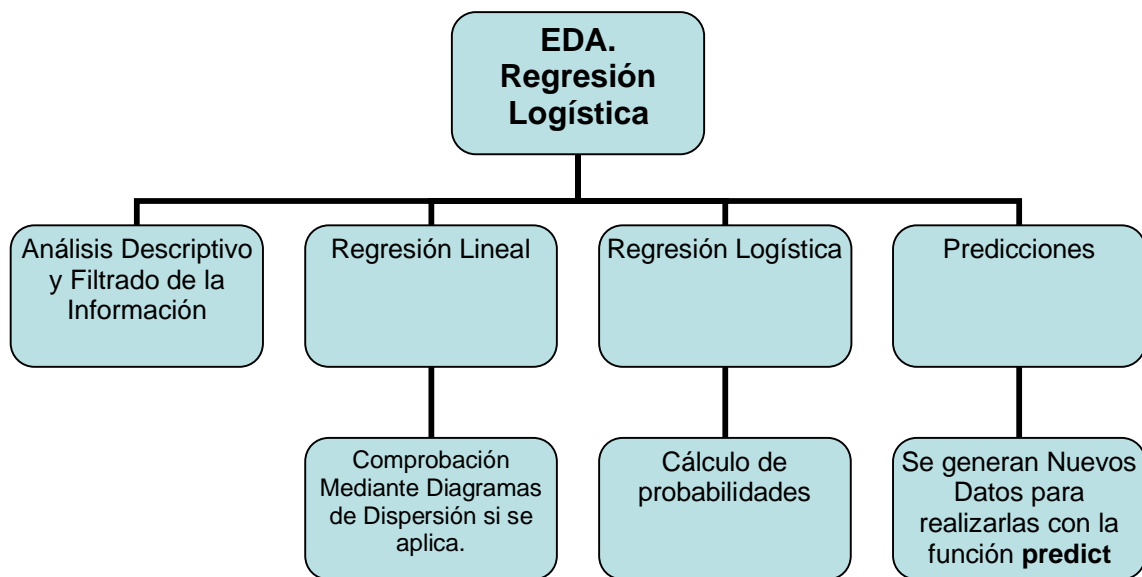


Fig. 9 EDA Regresión Logística

REGRESIÓN LOGÍSTICA			
Análisis Descriptivo. Filtrado de la información.	Regresión Lineal	Regresión Logística	Predicciones
<ul style="list-style-type: none"> Filtrado de la información, seleccionando los países foco de estudio (Reino Unido, Alemania, Italia, Francia y España) 	<ul style="list-style-type: none"> Generación de un modelo de regresión lineal que explique la variable “Mortalidad” en función de: <ul style="list-style-type: none"> Médicos de Atención Primaria. Unidades de Resonancia Magnética Escáneres Tomográficos, Actividades de Diagnóstico de Cáncer de Mama Diagnóstico de Cáncer de Cérvix Uterino Vacunación. 	<ul style="list-style-type: none"> Se discretiza las variables Mortalidad Prevenible y Mortalidad Tratable a 0 si no supera la media poblacional de la mortalidad y 1 en caso contrario. 	<ul style="list-style-type: none"> Se crean datos nuevos sobre los que realizar predicciones utilizando los modelos generados. <ul style="list-style-type: none"> Disminuyendo Recursos Aumentando Recursos Se utiliza la función predict y se calculan probabilidades.

<ul style="list-style-type: none"> • Filtrado particular de cada archivo en relación con los años de recopilación de información. (2011-2017) 	<ul style="list-style-type: none"> • En primer lugar, se comprueba si la generación del modelo lineal tiene sentido, es decir, si existe una relación lineal entre las variables a estudiar. <ul style="list-style-type: none"> ▪ La comprobación se realiza mediante diagramas de dispersión. 	<ul style="list-style-type: none"> • Se estiman modelos de regresión logística para predecir la probabilidad de tener una Mortalidad Tratable o Prevenible que supere la media en función de variables predictoras como: <ul style="list-style-type: none"> ○ Médicos de Atención Primaria. ○ Unidades de Resonancia Magnética ○ Escáneres Tomográficos 	
--	---	---	--

Tabla 3. EDA Regresión Logística

I. Análisis Descriptivo. Filtrado de la información.

- Filtrado de la información, seleccionando los países foco de estudio (**Reino Unido, Alemania, Italia, Francia y España**)
Se ha escogido estos países ya que son los 5 que aparecen con más frecuencias en las tablas resumen.

II. Regresión Lineal

- El objetivo es Generación de un modelo de regresión lineal que explique la variable “Mortalidad” en función de:
 - Médicos de Atención Primaria.
 - Unidades de Resonancia Magnética
 - Escáneres Tomográficos,
 - Actividades de Diagnóstico de Cancer de Mama
 - Diagnóstico de Cáncer de Cérvix Uterino
 - Vacunación
- En primer lugar, se comprueba si la generación del modelo lineal tiene sentido, es decir, si existe una relación lineal entre las variables a estudiar.
 - La comprobación se realiza mediante **diagramas de dispersión**.
 - Se comprueba que no existe una relación lineal entre variables por lo que se concluye que la

estimación de modelos de regresión lineales no tiene sentido.

De haberse podido realizar la estimación, se hubiera generado los modelos para después:

- Comprobar si el valor de **R-squared** (el coeficiente de determinación ajustado) indica si existe una fuerte dependencia lineal entre las variables estudiadas.
Cuanto mayor sea el coeficiente de determinación, mayor dependencia entre variables.
- Realizar **predicciones** de los modelos generados sobre nuevos datos para obtener información valiosa, como por ejemplo la variabilidad de la mortalidad en función de incrementos de gastos o personal hospitalario. Posteriormente se hubieran calculado **intervalos de confianza** sobre las predicciones realizadas.

III. Regresión Logística.

- Se discretiza las variables **Mortalidad Prevenible y Mortalidad Tratable** a 0 si no supera la media poblacional de la mortalidad y 1 en caso contrario.
- Se **estiman modelos de regresión logística** para predecir la **probabilidad** de tener una Mortalidad Tratable o Prevenible que supere la media en función de variables predictoras como:
 - Médicos de Atención Primaria.
 - Unidades de Resonancia Magnética
 - Escáneres Tomográficos

IV. Predicciones.

- Se crean datos nuevos sobre los que realizar predicciones utilizando los modelos generados.
 - Disminuyendo Recursos
 - Aumentando Recursos
- Se utiliza la función **predict** y se calculan **probabilidades**.

V. Código y Gráficos.

- Gráficos y código disponibles en: Enlace a GitHub. [Enlace a GitHub](#).

III. Análisis Exploratorio de Datos en Europa: Contraste de muestras.

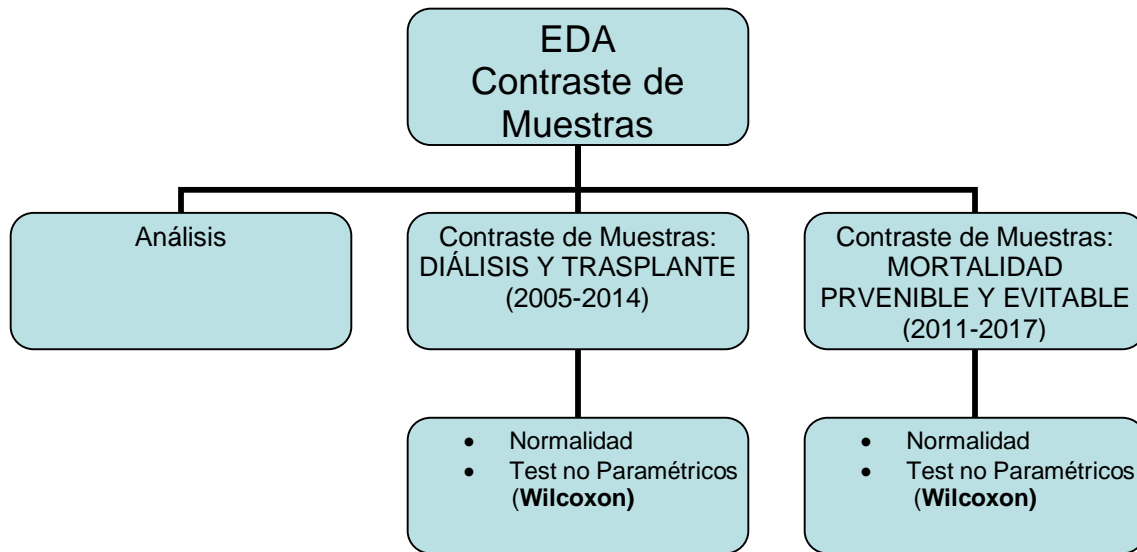


Fig. 10. EDA Contraste de Muestras

CONTRASTE DE MUESTRAS		
Análisis	Contraste de muestras: DIÁLISIS Y TRASPLANTE	Contraste de muestras: MORTALIDAD TRATABLE Y PREVENIBLE
<ul style="list-style-type: none"> • Descripción de las variables disponibles y del número de registros. 	<p>Contraste de Muestras para 2005 y 2014. Evolución del número de Pacientes en Diálisis y Trasplantados.</p> <ul style="list-style-type: none"> • Contraste de hipótesis de muestras pareadas 	<p>Contraste de Muestras para 2011 y 2017. Evolución del índice de Mortalidad Prevenible y Tratable.</p> <ul style="list-style-type: none"> • Contraste de hipótesis de muestras pareadas
<ul style="list-style-type: none"> • Filtrado de la información, seleccionando los años objeto de estudio <ul style="list-style-type: none"> ▪ 2005 y 2014 para Diálisis y Trasplante. ▪ 2011 y 2014 para Mortalidad Prevenible y Tratable. 	<ul style="list-style-type: none"> • Asunción de Normalidad mediante: <ul style="list-style-type: none"> ▪ Gráficas Q-Q ▪ Test de Shapiro. 	<ul style="list-style-type: none"> • Asunción de Normalidad mediante: <ul style="list-style-type: none"> ▪ Gráficas Q-Q ▪ Test de Shapiro.

<ul style="list-style-type: none"> •Se agrupa toda la información. 	<ul style="list-style-type: none"> •Se decide si se emplean test paramétricos o no paramétricos para hacer el contraste dependiendo del perfil de normalidad obtenido anteriormente. De entre los test no paramétricos tenemos: <ul style="list-style-type: none"> ▪ Test de suma de rangos (Mann-Whitney) ▪ Test de rangos y signos de Wilcoxon (muestras pareadas) 	<ul style="list-style-type: none"> •Se decide si se emplean test paramétricos o no paramétricos para hacer el contraste dependiendo del perfil de normalidad obtenido anteriormente. De entre los test no paramétricos tenemos: <ul style="list-style-type: none"> ▪ Test de suma de rangos (Mann-Whitney) ▪ Test de rangos y signos de Wilcoxon (muestras pareadas)
---	---	--

Tabla 4. EDA Contraste de Muestras

I. Análisis Descriptivo. Filtrado de la información.

- Descripción de las variables disponibles y del número de registros.
- Filtrado de la información, seleccionando los años objeto de estudio
 - 2005 y 2014 para Diálisis y Trasplante.
 - 2011 y 2014 para Mortalidad Prevenible y Tratable.

II. Contraste de Muestras. Diálisis y Trasplante.

- Contraste de Muestras para **2005 y 2014. Evolución del número de Pacientes en Diálisis y Trasplantados.** Contraste de hipótesis de **muestras pareadas**
- Asunción de Normalidad mediante:
 - **Gráficas Q-Q**
 - **Test de Shapiro.**
- Se decide si se emplean test paramétricos o no paramétricos para hacer el contraste dependiendo del perfil de normalidad obtenido anteriormente. De entre los test no paramétricos tenemos:
 - Test de suma de rangos (**Mann-Whitney**)
 - Test de rangos y signos de **Wilcoxon** (muestras pareadas).

Se utilizan test no paramétricos para realizar el contraste ya que no puede asumirse normalidad en ninguna muestra (Estudio de cada variable en el EDA y en cada apartado de la Regresión)

III. Contraste de Muestras. Mortalidad Tratable y Prevenible.

- Contraste de Muestras para **2011 y 2017**. Evolución de la **Mortalidad Tratable y Prevenible**. Contraste de hipótesis de **muestras pareadas**
- Asunción de Normalidad mediante:
 - **Gráficas Q-Q**
 - **Test de Shapiro.**
- Se decide si se emplean test paramétricos o no paramétricos para hacer el contraste dependiendo del perfil de normalidad obtenido anteriormente. De entre los test no paramétricos tenemos:
 - Test de suma de rangos (**Mann-Whitney**)
 - Test de rangos y signos de **Wilcoxon** (muestras pareadas).

Se utilizan test no paramétricos para realizar el contraste ya que no puede asumirse normalidad en ninguna muestra (Estudio de cada variable en el EDA y en cada apartado de la Regresión)

IV. Código y Gráficos.

- Gráficos y código disponibles en: Enlace a GitHub. [Enlace a GitHub](#)

3.4 Modelización de los datos

- 1 Modelización de las variables, con modelos no supervisados como Kmeans. (En Python).

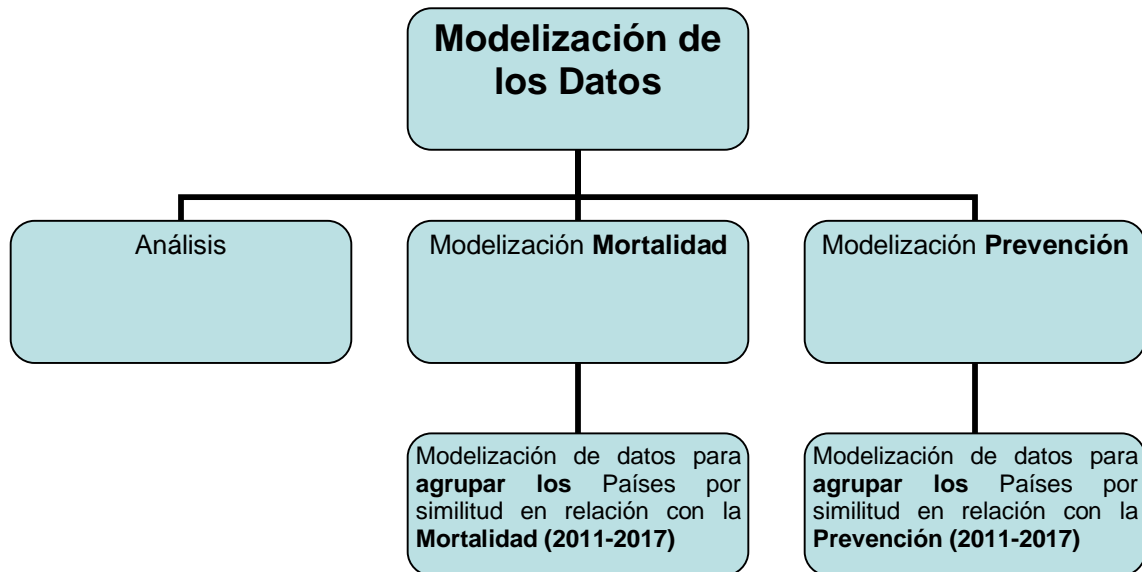


Fig. 11 Modelización de los Datos

MODELIZACIÓN DE DATOS		
Análisis	Modelización: MORTALIDAD	Modelización: PREVENCIÓN
<ul style="list-style-type: none"> • Visualización de las variables y sus atributos: <ul style="list-style-type: none"> ▪ Histogramas ▪ Diagramas de cajas. 	Modelización de datos para agrupar los Países por similitud en relación con la Mortalidad (2011-2017)	Modelización de datos para agrupar los Países por similitud en relación con la Prevención (2011-2017)
<ul style="list-style-type: none"> • Normalización de las variables para que tengan el mismo peso. 		

<ul style="list-style-type: none"> • Estimación del número de Clústeres a detectar por Kmeans, se utiliza “la regla del codo” 		
--	--	--

Tabla 5. Modelización de Datos

1 Análisis de la información.

- **Visualización** de las variables y sus atributos:
 - **Histogramas**
 - **Diagramas de cajas.**
- **Normalización** de las variables para que tengan el **mismo peso**.
- Estimación del número de Clústeres a detectar por Kmeans, se utiliza “**la regla del codo**”.

2 Modelización. Mortalidad.

- Modelización de datos para **agrupar los Países** por **similitud** en relación con la **Mortalidad** (2011-2017)

3 Modelización. Prevención.

- Modelización de datos para **agrupar los Países** por **similitud** en relación con la **Prevención** (2011-2017)

V. Código y Gráficos.

- Gráficos y código disponibles en: Enlace a GitHub. [Enlace a GitHub](#)

4. Conclusiones

Por Determinar. Se añaden:

- Conclusiones
- Gráficos

5. Glosario

Por Determinar

6. Bibliografía

- [1] Villegas Perriñán María del Mar, Rosa Díaz Isabel María. **La calidad Asistencial: concepto y medida**. 2003. Recuperado de:
[La calidad asistencial: concepto y medida | Villegas Perriñán | Dirección y Organización \(revistadyo.es\)](http://www.revistadyo.es)
- [2] López-Coronado Miguel, De la Torre Isabel, Herreros Gonzales Jesús, Cabo Salvador Javier. **Mejora de la calidad asistencial mediante la telemedicina y la teleasistencia**. Ediciones Díaz de Santos.2014. Capítulo 24.
- [3] Mira José Joaquín, Lorenzo Susana, Rodríguez-Marín Jesús, Aranaz Jesús, Sitges Esher. **La aplicación del modelo europeo de gestión de la Calidad Total al Sector Sanitario: Ventajas y Limitaciones**.1998. Recuperado de:
<http://www.calidadasistencial.es/images/gestion/biblioteca/39.pdf>
- [4] Nogué-Xarau S, Amigó-Tadín M, Sánchez-Sánchez M y Salmerón Bargo JM. **Evaluación y seguimiento de la calidad asistencial ofrecida a los intoxicados en un Servicio de Urgencias**.2007. Recuperado de:
<https://www.redalyc.org/pdf/919/91924105.pdf>
- [5] Eurostat Statistics Explained. **Health statistics at regional level.2021**. Recuperado de:
https://ec.europa.eu/eurostat/statistics-explained/index.php/Health_statistics_at_regional_level#Health_status_and_health_care.
- [6] EU_HEALTH. **State of Health in the EU 2017. 5 ways to make healthcare effective, accessible and resilient**. 2017. Recuperado de:
https://twitter.com/eu_health/status/933684222614753280
- [7] Howard, Simon John. **Comparing health protection systems across six European nations and the USA**. Recuperado de:
https://www.researchgate.net/publication/286611133_Comparing_health_protection_systems_across_six_European_nations_and_the_USA

6. Anexos

Los datos utilizados para el análisis se encuentran disponibles en el siguiente enlace:

<https://ec.europa.eu/eurostat/web/health/data/database>

En el propio repositorio de GitHub, se ha creado una carpeta con los datos procesados para el análisis.

Toda la documentación generada durante la fase de diseño e implementación: ficheros pdf y Rmd generados en el estudio estadístico con RStudio en lenguaje R y con Anaconda en lenguaje Python se encuentra alojada en el repositorio GitHub:

[aperdices-g/TFM---Calidad-Asistencial-Sanitaria-Europea \(github.com\)](#)

Esquema del repositorio: TFM Calidad Asistencial Sanitaria Europea.

1. **Datos Disponibles.** Conjunto de datos utilizados para el análisis.
2. **Limpieza de Datos (ETL)**
 - a. En relación con Actividades
 - b. Causas de Mortalidad
 - c. Cuidados Domiciliarios
 - d. Estados de Salud por Sexos
 - e. Gasto Sanitario
 - f. Prevención
 - g. Recursos Sanitarios
3. **Análisis Exploratorio de Datos (EDA)**
 - a. Análisis Exploratorio por Países.
 - b. Análisis de Regresión Logística.
 - c. Contraste de Muestras.
4. **Modelización de datos**
5. **Memoria**