

Fundamentos de Ciencia de Datos

PEC3: Proyectos y proyectos de datos. Ciclos y metodologías.

Itziar Ricondo (iricondoi@uoc.edu)

Criterios de evaluación generales de la PEC.

- Demostrar que se han asimilado correctamente los contenidos teóricos y que se aplican con precisión.
- Razonar todas las respuestas de forma analítica y sintética.
- Estructurar y maquetar correctamente el documento.
- Esquematizar las respuestas mediante el uso de gráficos, infografías y diagramas explicativos.
- Utilizar con precisión las citas y las referencias bibliográficas.
- Ajustar la extensión del documento a un máximo de 20 páginas.

Contexto.

Continuamos con el caso tratado en las anteriores prácticas.

Resumen: El 2019 el grupo HM Hospitales inició un plan de transformación digital que supuso un cambio en la cultura de la organización. En 2020 arranca la iniciativa *Covid data save lives*, poniendo a disposición de la comunidad científica el repositorio de los historiales clínicos de los pacientes atendidos durante la pandemia.

De nuevo, responde a las siguientes preguntas tomando como punto de partida la bibliografía del Bloque 3 y la búsqueda de información adicional que realices.

Pregunta 1 (15% puntuación)

En base a las indicaciones de la metodología **PMBOK** referentes a la figura del director de proyecto:

- ¿A qué miembro de la organización le asignarías el rol de director de proyecto *Covid data save lives?*. Describe sus responsabilidades y competencias.

Asignaría el rol de **director de proyecto** Covid data saves lives a **José María Castellano**, director científico de la Fundación de Investigación HM Hospitales [1]. La iniciativa tiene como objetivo poner a disposición de la comunidad científica internacional datos clínicos de la evolución de los pacientes covid para ayudar a conocer mejor los efectos del virus. Es por ello que creo que el director científico de la Fundación de Investigación puede conocer bien (de manera ágil, puesto que el tiempo apremiaba) las necesidades de los *stakeholders* principales, la comunidad científica.

Existen otros roles relacionados con la gestión de proyectos, no de manera individual, que creo que es importante considerar. En la bibliografía [2], [3] el Director General el Dr Juan Abarca aparece en varias entrevistas. Tal y como se preguntaba también en la PEC1, considero que el perfil de Abarca encaja en el de **patrocinador del proyecto** general de transformación digital. Él es el primer convencido en el poder de los datos: “Creemos que el dato, gestionado a través de distintas herramientas como la inteligencia artificial puede ser el gran diferencial de la sanidad del siglo XXI”.

En otras entrevistas aparece Alberto Estirado [4], director de Sistemas de Información y Transformación Digital de HM Hospitales. Estirado tiene la función de coordinar e implementar el Plan Estratégico de Transformación Digital. Creo que un director de portafolio sería un director de varios proyectos orientados a cierta área de la estrategia de transformación. El rol de Estirado creo que es superior al director de portafolio.

El director de proyecto es la persona asignada como **responsable de gestionar** el proyecto, cumplir con los objetivos y guiar el proyecto hacia el éxito. Las responsabilidades y competencias del director del proyecto van más allá de las capacidades técnicas y engloban competencias personales orientadas a la relación.

- La dirección técnica de proyectos. El director de proyectos debe conocer y dominar aquellos **conocimientos, métodos, técnicas y habilidades** propios de la dirección de proyectos y específicos de esta materia. Si bien en la gestión de proyectos se está llevando a cabo un cambio de modelo, de experto técnico a facilitador, es necesario tener unos conocimientos técnicos en la materia, para entender las necesidades del

proyecto y conocer también el lenguaje común entre los miembros del equipo, para poder aportar, ser eficaz y eficiente. En este caso, el proyecto *Covid data save lives* busca compartir información valiosa de los pacientes Covid para que la comunidad científica pueda utilizar estos datos en sus algoritmos para aprender sobre la enfermedad. Es por eso que he elegido a **José María Castellano**, director científico de la Fundación de Investigación HM Hospitales, ya que tiene conocimiento y puede saber por su experiencia qué tipo de datos se requieren (mucho más que los datos demográficos mostrados en otros datasets) y en qué formato para poder aplicar algoritmos de aprendizaje

- **Gestión estratégica y de negocios.** El director de proyectos debe conocer el entorno competitivo y de negocio, para poder identificar los proyectos más interesantes y alineados con la estrategia. En este caso Castellanos [1] aporta conocimiento del sector/actividad de la investigación clínica, conoce qué tipo y formato de datos son interesante para la aplicación.
- **Liderazgo.** El director de proyectos debe guiar, motivar y dirigir un equipo. Cada vez se da más importancia a estas habilidades personales, sin desmerecer las responsabilidades y competencias técnicas y de negocio expuestas anteriormente. Estas habilidades se pueden clasificar en:
 - Comunicación
 - Liderazgo
 - Dirección (incluye la resolución de conflictos)
 - Habilidades cognitivas

A nivel más detallado, el director de proyecto es responsables, en el marco de un proyecto, de:

- Identificar los **requerimientos** del proyecto y los riesgos asociados.
- Desarrollar el **plan** de proyecto y los subplanes asociados.
- Gestionar y **guiar** el proyecto para cumplir con sus objetivos.
- Identificar y gestionar las necesidades, preocupaciones y expectativas de los **interesados** del proyecto
- Gestionar la **influencia** de los **interesados**
- Gestionar y dirigir el **equipo** de proyecto y los **recursos**.
- **Mantener** el proyecto según lo planificado y gestionar las restricciones: alcance, tiempo, coste, calidad y recursos.
- Monitorizar, comunicar y reportar las **métricas** del proyecto y su estado a los interesados.

- Entender y aplicar el conocimiento, las habilidades, las herramientas y las técnicas a las actividades del proyecto para cumplir con los objetivos. Esta responsabilidad está orientada a la experiencia, con objeto de utilizar tanto las habilidades interpersonales como las mejores lecciones aprendidas en beneficio de proyecto.
- Asume el rol de director. ¿Qué entregables planificarías?, ¿qué recursos marcarías como indispensables y por qué?

Entregables:

- **E1. Informe de requisitos del proyecto:**
 - **Tipo:** Informe
 - **Descripción:** Entregable donde se describen los requisitos de datos:
 - **Contenido:** todas las tablas y variables a compartir con el exterior
 - **Gobernanza:** los requisitos de gobernanza de datos, incidiendo en la privacidad, requisitos de acceso y uso de los mismos.
- **E2. Informe de desarrollo para la obtención del dataset Covid data save lives**
 - **Tipo:** Informe
 - **Descripción:** Informe donde se describe el trabajo de desarrollo realizado en las bases de datos, operaciones de transformación y consulta de datos para obtener las tablas de datos necesarias-
- **E3. Conjunto de datos**
 - **Tipo:** Conjunto de datos
 - **Descripción:** Conjunto de datos que recibirá el solicitante del dataset Covid data save lives
- **E4. Informe de cierre**
 - **Tipo:** Informe

Los recursos asociados al proyecto consideran tanto los recursos materiales y personales. En el proyecto Covid Data Save Lives no he estimado la compra de recursos materiales, ya que se cuenta con toda la infraestructura hardware y software. En cuanto a los recursos personales, en el proyecto se requerirán personas de los siguientes departamentos:

- Fundación de Investigación HM Hospitales: Se necesita una persona especializada en machine learning para medicina predictiva, no tanto por su conocimiento en técnicas de

análisis, sino para que establezca los requisitos de datos interesantes de todos los datos que se registran de los pacientes en HM Hospitales.

- **Sistemas de Información:** Se necesitan recursos del departamento de sistemas de información, en concreto con conocimientos en gestión de bases de datos para la obtención de la información necesaria sobre los pacientes.
- **Comité de Data Science:** Persona del comité de Data Science de HM Hospitales para la elaboración del proceso de aprobación de solicitud de acceso a los datos *Covid data save lives*, que está muy ligado al establecimiento de condiciones que deben cumplir los futuros usuarios de los datos.

Pregunta 2 (25% puntuación)

Siguiendo la metodología del **PMBOK**, responde a los siguientes apartados en base al proyecto *Covid Data Save Lives*:

- **Define formalmente el proyecto:**

El proyecto *Covid data save lives* es un trabajo llevado a cabo por HM Hospitales que tiene por objetivo desarrollar un dataset con información (datos clínicos, analíticas, tratamientos) de pacientes de Covid que han pasado por HM Hospitales y ponerlo a disposición de la comunidad científica para posterior uso en investigación.

- **Explica las fases y principales características del ciclo de vida del proyecto.**

El ciclo de vida de un proyecto es el conjunto de fases (secuenciales o solapadas) por las que pasa un proyecto desde su inicio hasta el cierre. Las fases seguidas en este proyecto siguiendo la metodología del PMBOK serían:

- **Inicio del proyecto:** El proyecto recibe la aprobación y financiación del proyecto. En este caso, en plena pandemia, HM Hospitales decide llevar a cabo este proyecto [5], en un momento donde contaban con 3000 historiales de pacientes de los hospitales de Madrid, Barcelona y Galicia. Voy a asumir que el inicio del proyecto ha sido el 1 de marzo de 2020. Esta fecha coincide también con la nominación del Dr. José María Castellano, nombrado director científico de la Fundación de Investigación HM

Hospitals (y que en este trabajo se ha definido como director del proyecto *Covid data save lives*).

- Planificación del proyecto: En esta fase se definen y revisan los objetivos, el alcance, y las actividades a realizar. Se planifican las actividades a realizar para conseguir el alcance y los objetivos, así como su cronograma en el tiempo y los recursos necesarios para llevarlo a cabo. El resultado es el plan de proyecto, con la estructura de tareas WBS (*Work Breakdown Structure*). Considero que el plan de proyecto se orientó a minimizar el tiempo de ejecución, para poder compartir los datos con la comunidad para fines de investigación cuanto antes.
- Ejecución del proyecto: En esta fase el equipo de proyecto realiza las actividades planificadas. Entre las actividades/tareas de Covid data save lives podríamos identificar:
 - 1) el diseño del conjunto de datos a compartir, conjunto de tablas y variables sobre los pacientes que compondrán el dataset a compartir.
 - 2) transformación, anonimización y consulta de datos para la obtención del conjunto de datos final,
 - 3) procedimiento de publicación y publicación del conjunto de datos.
- Seguimiento y control: En esta fase el director de proyecto asegura que el trabajo se realiza de acuerdo con el plan. En caso contrario, se toman medidas correctoras.
- Cierre: formaliza la aceptación y el cierre del proyecto o de una fase.

- ¿Qué grado de complejidad le asignarías? Argumenta la respuesta

A día de hoy HM Hospitals cuenta con proceso sistematizado de captura de datos de paciente, por lo que mucha de la información del dataset final ya se adquiriría por defecto [6]. Siguiendo el marco de clasificación de Cynefin, considero que la complejidad del proyecto es **simple**. Tanto los requerimientos como la tecnología requerida por la ejecución están lo suficientemente definidos y el proyecto se puede gestionar con una metodología simple.

Como el marco de clasificación se basa en el grado de conocimiento de la tecnología y de los requerimientos, a continuación se explica el grado de conocimiento de ambos términos para el proyecto *Covid data save lives*.

En cuanto al conocimiento de los requerimientos, no hay una especificación única de requerimientos que deba cumplirse para la ejecución del proyecto. Es iniciativa de HM Hospitals poner a disposición los datos. El hecho de que no haya unas especificaciones claras en este caso no es motivo de complejidad. El equipo de proyecto y director de proyecto tienen conocimiento para determinar qué

datos pueden ser interesantes para su posterior análisis. Tal y como afirman en [6], son conscientes del valor de su dataset con respecto a otros conjuntos de datos que se comparten sobre Covid, que se centran en datos demográficos

En cuanto a la tecnología, el hecho de que actualmente ya se recojan los historiales de los pacientes implica que el hardware y software requerido ya está disponible. Por lo tanto, la incertidumbre tecnológica es muy baja.

- ¿Quiénes serían los *stakeholders* o interesados en el proyecto?

El PMBOK define los *stakeholders* como la persona, grupo u organización que puede afectar o verse afectado por el proyecto. Es una definición amplia, donde pueden diferenciarse los integrantes del equipo del proyecto, otros interesados internos a la organización y los interesados externos a la organización.

- Miembros del equipo: Entre los miembros del equipo destacaría las personas del departamento de investigación (machine learning para investigación médica)
- Interesados internos de la organización: Otros integrantes de la organización que no participan directamente en el proyecto, como podría ser el promotor, el propietario de los recursos, otros directores de programas y portafolio. Interesados internos del proyecto son el Dr Abarca, patrocinador de la iniciativa de digitalización, y Alberto Estirado [1], director de Sistemas de Información y Transformación Digital.
- Interesados externos de la organización: Organismos de investigación médica o, como se cita en [6], otros ‘players’ públicos y privados del sector sanitario nacionales y extranjeros. Los proveedores y otros partners/alianzas estarían también dentro de esta categoría, como es el caso del Massachussets Institute of Technology (MIT) [7]

Pregunta 3 (30% puntuación)

Detalla las actividades realizadas, los roles involucrados y los productos resultantes para las distintas áreas de conocimiento en el proyecto *Covid Data Save Lives*:

- **Gestión de la integración:** Son las actividades para identificar, definir, combinar, unificar y coordinar los diferentes procesos y actividades de la dirección de proyectos. Según esta definición de PMBOK, la gestión de la integración sería el trabajo principal del director del proyecto, gestionar todos los elementos del proyecto para que se obtengan los objetivos del mismo. En el proyecto *Covid Data Save Lives* el director de proyecto debe gestionar que todas las partes que lo componen lo hagan según la planificación y en el orden adecuado para conseguir los resultados. El director de proyecto es el responsable de desarrollar el acta de constitución del proyecto, dirigir y controlar el trabajo, así como de cerrar las fases y el proyecto. Creo que el proyecto *Covid Data Save Lives* se inicia con el convencimiento de la Alta Dirección y uno de los retos para el director del proyecto es asegurar su ejecución en el menor tiempo posible, en un momento donde la enfermedad estaba en una situación límite en España y a nivel mundial se necesitaban datos para investigar.
- **Gestión del alcance.** Dentro de esta área de conocimiento, las actividades son:
 - Planificar la gestión del alcance: Se planifica el proceso y se crea el plan de gestión del alcance.
 - Recopilar requisitos: En esta tarea se definen y documentan las necesidades de las partes interesadas.
 - Definir el alcance: El alcance define qué se incluye y qué queda fuera del trabajo a realizar. En el proyecto, el alcance del proyecto se establece como el conjunto de datos relacionados con pacientes Covid que sirvan para su análisis basado en técnicas estadísticas y machine learning con objeto de avanzar en el conocimiento de la enfermedad.
 - Crear la EDT/WBS: Se crea el plan de trabajo, las diferentes actividades para conseguir los entregables del proyecto. En el proyecto *Covid Data Save Lives* se crean las tareas del proyecto.
 - Validar el alcance: Los entregables del proyecto se aceptan formalmente, como medida de avance del proyecto hacia la consecución del alcance. Es importante por ello que los entregables sean aceptados por un *stakeholder*, para lo cual previamente debe definirse quién debe aprobar los entregables.

- **Controlar el alcance:** Los cambios de necesidades y requerimientos de las partes interesadas del proyecto deben ser integrados en una definición del alcance y estructura de trabajo, para evitar problemas entre las expectativas de las partes interesadas.
- **Gestión del cronograma.** En este grupo de procesos se realiza la planificación temporal de las actividades del proyecto. Las actividades son: a) Planificar la gestión del cronograma, b) Definir las actividades, c) Secuenciar las actividades, d) Estimar la duración de las actividades y e) Desarrollar el cronograma

Como se ha comentado anteriormente, se ha establecido el inicio de proyecto en 1 de marzo y el 25 de abril de 2020 se publicó el acceso (bajo solicitud) a los mismos. Se ha realizado en este ejercicio una propuesta de cronograma (Tabla 1) .

Tabla 1: Cronograma propuesto del proyecto

Cronograma	Marzo 2020				Abril 2020			
Tareas	S1	S2	S3	S4	S1	S2	S3	S4
Inicio proyecto	●							
Definición de requerimientos		●						
Diseño de conjunto de datos, tablas y variables			●					
Desarrollo: transformación, anonimización y consulta de datos para la obtención del conjunto de datos final							●	
Cierre: Publicación de datos y cierre del proyecto								●

- **Gestión de costes.** Las actividades a realizar son la planificación y estimación de los costes, el presupuesto y el control de los costes. En la Tabla 2 se realiza un presupuesto estimado del proyecto, estimando un coste promedio de 50 €/h.

Tabla 2: Presupuesto propuesto de proyecto

Dedicación y coste personas	Horas	Coste Promedio	Coste Total	Comentarios
Definición de requerimientos	64 h	50 €/h	3.200 €	4 personas, 16 h/persona
Diseño de conjunto de datos, tablas y variables	72 h	50 €/h	3.600 €	3 personas, 24 h/persona
Desarrollo: transformación, anonimización y consulta de datos para la obtención del conjunto de datos final	80 h	50 €/h	4.000 €	1 persona desarrollo
Cierre: Publicación de datos y cierre del proyecto	40 h	50 €/h	2.000 €	8 h información difusión medios, 24 horas proceso aprobación de solicitud de acceso
	256 h		12.800 €	

- **Gestión de la calidad.** Incluye las actividades para asegurar que se cumplen las normas de gestión de proyectos y normativas del sector, así como la calidad medida como satisfacción del cliente. Los elementos principales de gestión de la calidad son la planificación de proyectos, la gestión de los recursos y la gestión de costes. Indicadores típicos de gestión de la calidad serían el porcentaje de proyectos finalizados en el tiempo planificado, la satisfacción del cliente con respecto a los recursos obtenidos.
- **Gestión de los recursos.** Dentro de este área de conocimiento se realiza la identificación, adquisición y gestión de los recursos. Se incluyen tanto los recursos materiales como personales. En el proyecto *Covid Data Save Lives* no he estimado la compra de recursos materiales, ya que se cuenta con toda la infraestructura hardware y software. En cuanto a los recursos personales, en el proyecto se requerirán personas de las áreas de:
 - Fundación de Investigación HM Hospitales: Establecimiento de los requisitos sobre los datos.
 - Sistemas de Información: Gestión de bases de datos para la obtención de la información necesaria sobre los pacientes.
 - Comité de Data Science: Elaboración del proceso de aprobación de solicitud de acceso a los datos Covid data save lives.

- **Gestión de la comunicación.** Incluye la gestión de toda la documentación del proyecto, así como la difusión del mismo al exterior. Por ejemplo, el 15 de abril HM Hospitales hacía el anuncio de que iba a poner el conjunto de datos disponible [5] y el conjunto se puso a disposición pública el 25 de abril. Además, la noticia fue recogida por diferentes medios.
- **Gestión de riesgos.** Incluye los procesos necesarios para identificar, evaluar y mitigar/eliminar los riesgos del proyecto y las potenciales causas que pueden tener un impacto sobre los objetivos. Uno de los riesgos o aspectos críticos del proyecto es la privacidad de los datos y el acceso a los mismos. En el proyecto se adoptó como medida de protección la anonimización de los datos. Además, se ha desarrollado un proceso para el control y para aprobar que una persona/organización pueda recibir el conjunto de datos de *Covid data saves lives*. El investigador que desee utilizar el conjunto de datos, debe escribir un correo a HM Hospitales, para que su solicitud sea evaluada por la Comisión de Data Science y, en su caso, por el Comité de Ética de la Investigación de HM Hospitales o cualquier otro comité de ética de la investigación acreditado.
- **Administración de compras y contratos.** Incluye todos los procesos necesarios para la compra o adquisición de productos, servicios o resultados necesarios y externos al equipo de trabajo, tanto en nuestra relación con el cliente. En el caso del proyecto *Covid data save lives* ya se contaba con todos los equipos necesarios, pero en otros proyectos podría darse la necesidad de tener que comprar por ejemplo un servidor para el almacenamiento de datos, infraestructura para la ciberseguridad, o un equipo de altas características para el desarrollo de algoritmos de aprendizaje automático. Por otro lado, puede darse la circunstancia de que cierto trabajo sea subcontratado, ya sea por carecer del conocimiento necesarios para llevarlo a cabo, o razones de tiempo y coste. En cualquier caso, tanto las compras de materiales, equipos y subcontrataciones deben ser gestionadas durante el proyecto, que requiere también de la participación del departamento de compras (con el que deberá relacionarse el director del proyecto).
- **Gestión de los interesados.** Implica las actividades para relacionarse y tratar con los interesados del proyecto, dependiendo de su rol y posición en el proyecto. En el proyecto *Covid data save lives* es importante la visión del director de proyecto sobre las necesidades de los investigadores expertos en análisis de datos, clientes usuarios de este proyecto.

Pregunta 4 (30% puntuación)

Propón cómo una organización suscrita a la iniciativa *Covid Data Save Lives* coordinaría su participación siguiendo la metodología *Scrum*. Para ello:

En este ejercicio se va a suponer que el departamento de **eHealth Center de la UOC** ha solicitado a HM Hospitales acceso a los datos de *Covid data save lives*.

El conjunto de datos que se ha descargado está compuesto por 6 tablas. Se lista el nombre y el número de filas de las mismas:

- Tabla principal, datos de ingreso y básicos de paciente, 2.226 registros.
- Tabla de medicación, >60.000 registros
- Tabla de constantes vitales, 54.000 registros
- Tabla de laboratorio, 398.884 registros
- Tabla de diagnósticos, >1.600 para ingresados y >1.900 para urgencias.

Utilizando el conjunto de datos recibido por HM Hospitales, se ha definido un proyecto Scrum con los siguientes objetivos:

- 1) Disponer de una tipología de pacientes
- 2) Determinar la efectividad de diferentes tratamientos.
- 3) Identificar cuanto antes si un paciente puede derivar en paciente grave de covid

Para este ejercicio se va a planificar un Sprint orientado al **análisis descriptivo de los datos**, mediante técnicas de Estadística Descriptiva.

- **Detalla una iteración o sprint. ¿Qué entradas y salidas tendría?**

El Sprint es cada una de las partes en las que se divide el proyecto Scrum. Está orientado a producir un entregable potencialmente entregable. El Sprint que se va a trabajar estas dos semanas (en este ejercicio) es el **análisis exploratorio y descriptivo** de los datos de pacientes covid. Las entradas y salidas para este Sprint se describen a continuación.

Entradas: 1) Conjunto de datos original, aportado por HM Hospitales,

Salida: Archivo Jupyter del análisis exploratorio y descriptivo de los datos del conjunto de datos *Covid data save lives*.

Se va a definir un etapa de Sprint de 2 semanas, aunque la metodología Scrum establece una duración máxima de 1 mes [8]. Sin embargo, se considera suficiente un periodo de realización de 2 semanas, como etapa donde se realiza un análisis de conocimiento inicial de los datos.

- ¿Qué roles y artefactos estarían involucrados?

Roles:

- **Propietario de Producto (Product Owner):** Es el encargado de optimizar y gestionar el flujo de valor del producto. Además de mantener el Product Backlog actualizado y priorizado, debe transmitir a los stakeholders el valor del producto, es el altavoz del cliente. Propongo como Product Owner al director del departamento de eHealth de la UOC.
- **Scrum Master:** Su función es gestionar que el proceso Scrum se lleva a cabo correctamente y eliminar impedimentos. En este ejemplo propongo que Itziar Ricondo va a ser Scrum Master. Su Background es la ingeniería de organización industrial, pero tiene experiencia en metodologías de mejora. Tras finalizar el Master se ha especializado en metodologías ágiles. Su labor en este proyecto es actuar de facilitadora frente al equipo de desarrollo y eliminar cualquier impedimento que pueda surgir.
- **Equipo de desarrollo:** El equipo de desarrollo es el encargado de desarrollar el producto. En este caso va a estar compuesto por 4 personas. Estas personas cuentan con un Master de Ciencias de Datos por la UOC y se acaban de incorporar al departamento de investigación de la UOC. Han recibido formación en la metodología por parte de Itziar Ricondo y este va a ser su primer proyecto Scrum. Van a trabajar exclusivamente en este proyecto. De los cuatro integrantes del equipo, dos de ellos son licenciados en matemáticas, uno de ellos es informático y otro es ingeniero.

Artefactos:

Los artefactos principales son:

- **Pila de Producto (Product Backlog):** Es la principal fuente de información sobre el producto en Scrum, una lista que contiene todos los requerimientos necesarios para implementar el producto. El Product Backlog se gestiona por el **Product Owner**, cuya función es priorizar aquellos elementos que tienen más valor en cada etapa. Se propone el Product Backlog de la Tabla 3.

Tabla 3: Backlog de producto propuesto

Product Backlog
El trabajo a realizar se basa en el conjunto de datos Covid data save lives-
El análisis debe ser presentado al usuario mediante página html o código de archivo Jupyter Notebook
El trabajo debe contener un análisis descriptivo de los datos, para entender qué ha pasado
El trabajo debe contener un análisis basado en técnicas de aprendizaje automatizado para la clasificación de pacientes, predicción temprana de la gravedad y eficiencia de tratamientos
El análisis debe contener tablas y gráficos visuales, para ayudar en la comprensión

- **Sprint Backlog:** Se trata de la lista de elementos a trabajar durante esta etapa Sprint. Este artefacto permite entender cuál es la evolución del trabajo durante el Sprint y visualizar en qué grado se han cumplido los objetivos del Sprint.

Tabla 4: Evolución del Sprint Backlog

Sin empezar	En desarrollo	Terminado
	Análisis t. constantes vitales Análisis t. laboratorio Análisis t. diagnósticos	Análisis t. principal Análisis t. medicación

- **Incremento:** Un incremento es el resultado del Sprint, es la suma de todo el trabajo desarrollado durante el Sprint y dispuesto en forma de resultado para el cliente final. El incremento debe estar en condiciones de utilizarse sin importar si el Product Owner decide liberarlo o no. Al final de esta etapa de Sprint, el resultado es un software basado en Jupyter con el análisis descriptivo de las 6 tablas.

En la metodología Scrum hay otros artefactos no nucleares, como son la definición de “hecho” (**definition of done**) y definición de “listo” (**definition of ready**).

- ¿Qué ceremonias se realizarían?

Scrum define 5 ceremonias para el control de los procesos, eventos con una duración determinada.

- **El Sprint (*Sprint*):** El Sprint es el elemento nuclear del Sprint. El Sprint es un contenedor de los demás eventos y actividades de Scrum: Planificación del Sprint (Sprint Planning), Scrums Diarios (Daily Scrums), trabajo de desarrollo, Revisión del Sprint (Sprint Review) y la Retrospectiva del Sprint (Sprint Retrospective).

- **Planificación del Sprint (*Sprint Planning*):** Es una reunión de planificación donde se define el trabajo a realizar en el Sprint. Para Sprints de 1 mes tiene una duración de 8 horas. El Sprint que se ha planteado es de 2 semanas, por lo que reunión de planificación del Sprint será de 4 horas. La entrada a esta reunión está constituida el Product Backlog), el último Incremento de producto, la capacidad proyectada del Equipo de Desarrollo para el Sprint y el rendimiento pasado del Equipo de Desarrollo. El número de elementos de la Pila del Producto seleccionados para el Sprint depende únicamente del Equipo de Desarrollo
- **Daily Scrum (*Scrum Diario*):** Reunión diaria de 15 minutos en la que participa exclusivamente el Equipo de Desarrollo.
- **Revisión del Sprint (*Sprint Review*):** Al final del Sprint se lleva a cabo una Revisión de Sprint para inspeccionar el Incremento y adaptar la Pila del Producto si fuese necesario. Participan el Equipo de Desarrollo, los interesados del proyecto (invitados por el Product Owner) y el Product Owner. En la reunión se discute el Incremento realizado, los problemas que han aparecido y su resolución. Se discute cuáles pueden ser los siguientes desarrollados, como entrada a la planificación del siguiente Sprint.
- **Retrospectiva del Sprint (*Sprint Retrospective*):** El objetivo de la retrospectiva es hacer de reflexión sobre el último Sprint e identificar posibles mejoras para el próximo.

Itziar Ricondo (iricondoi@uoc.edu)
13/06/2021

Referencias

- [1] «José María Castellano, nuevo director científico de la Fundación de Investigación HM Hospitales». <https://www.infosalus.com/actualidad/noticia-jose-maria-castellano-nuevo-director-cientifico-fundacion-investigacion-hm-hospitales-20200224125313.html> (accedido jun. 09, 2021).
- [2] «Juan Abarca, de HM Hospitales, Premio ComputerWorld 2020 al presidente del año», *ComputerWorld*, 2021. <https://www.computerworld.es/entrevistas/juan-abarca-de-hm-hospitales-premio-computerworld-2020-al-presidente-del-ano> (accedido mar. 18, 2021).
- [3] «HM Hospitales y Microsoft avanzan en la ciencia del dato en el ámbito de la salud», *Computing*, mar. 16, 2021. <https://www.computing.es/analytics/noticias/1124401046201/hm-hospitales-y-microsoft-avanzan-ciencia-del-dato-ambito-de-salud.1.html> (accedido abr. 27, 2021).
- [4] «Alberto Estirado, nuevo director de Sistemas de Información y Transformación Digital de HM Hospitales». <https://www.hmhospitales.com/prensa/notas-de-prensa/alberto-estirado-nuevo-director-sistemas-informacion-transformacion-digital> (accedido mar. 20, 2021).
- [5] «Covid Data Save Lives». <https://www.hmhospitales.com/coronavirus/covid-data-save-lives> (accedido mar. 18, 2021).
- [6] H, «HM Hospitales aúna investigación y Big Data para cercar a la Covid-19», *La Razón*, dic. 31, 2020. <https://www.larazon.es/salud/20201231/yzm7vnrlzgrbpvc4f76fbak7q.html> (accedido mar. 18, 2021).
- [7] «HM Hospitales y el Massachusetts Institute of Technology unen sinergias para predecir el comportamiento del Covid-19». <https://www.hmhospitales.com/prensa/notas-de-prensa/hm-hospitales-y-massachusetts-institute-of-technology-unen-sinergias-para-predecir-comportamiento-covid-19> (accedido jun. 13, 2021).
- [8] «Las 5 ceremonias Scrum: claves para la gestión de procesos», *Deloitte Spain*. <https://www2.deloitte.com/es/es/pages/technology/articles/ceremonias-scrum.html> (accedido jun. 13, 2021).