

Introducción a la Ciencia de Datos – PEC1

Declaración de trabajo original (no plagio) de la o el estudiante

Yo, Vinicio Alejandro Naranjo Mosquera, declaro que esta entrega es el resultado de mi propio esfuerzo y creatividad. He generado todo el contenido basándome en mi conocimiento, habilidades y material proporcionado por mis profesores, sin copiar ni utilizar el trabajo de otros sin el debido crédito.

Ejercicio 1 (20% puntuación)

1) Responded:

- a) Teniendo en cuenta las tres disciplinas básicas que se combinan en el rol de científico o científica de datos, ¿qué dos disciplinas se combinan en el Aprendizaje Automático o Machine Learning?

Las dos disciplinas son:

- **Matemáticas y estadística:** Son las herramientas necesarias para desarrollar modelos predictivos y comprender las relaciones y patrones dentro de los datos.
- **Ciencias de la computación y programación:** Convierten teorías matemáticas en soluciones prácticas de aprendizaje automático. Facilitan la manipulación de datos y el desarrollo de modelos. Son esenciales para experimentar y desplegar tecnologías de IA efectivamente.

- b) ¿Qué dos disciplinas se combinan en el desarrollo de software?

- **El dominio de la informática:** Implica entender desde programación hasta sistemas, crucial para crear y mantener software eficaz.
- **Conocimiento del área de negocio:** Entender el área de negocio es clave para desarrollar software que sea técnica y comercialmente valioso, adaptándose a objetivos y desafíos específicos.

- c) ¿Por qué el perfil de Machine Learning a veces se considera **peligroso** en una empresa?

La implementación de Machine Learning requiere de un conocimiento profundo del dominio de negocio y consideraciones éticas para evitar resultados dañinos y asegurar que los proyectos reflejen los objetivos y valores de la empresa.

- d) Explique brevemente uno de los diez puntos en los cuales se resume el trabajo diario en el rol de Data Scientist, **indicando además el número que ocupa en la lista que los recoge en los contenidos de referencia.**

El punto 9: Crear programas y productos que proporcionen conocimiento a la empresa.

El desarrollo de aplicaciones que integran predicciones de Machine Learning es valioso para ampliar soluciones empresariales y facilitar decisiones informadas, transformando datos en insights estratégicos.

- e) Según la propuesta que se hace en el documento del proceso de la ciencia de datos, ¿cuál sería el paso previo al Data Mining ? Explíquelo brevemente.

El mantenimiento del dato: Este paso es esencial para procesar los datos sin que se genere aún un valor claro para la organización. Incluye tareas fundamentales como la integración, limpieza, enriquecimiento de los datos, así como procesos de extracción, transformación y carga del dato (ETL - Extract, Transform, Load).

- f) ¿Cuál es el ámbito de aplicación de la Ciencia de datos de un algoritmo que puede aprender de múltiples perfiles y determinar que un individuo tiene el potencial de cometer un robo? ¿Puede llevar a la detención de una persona por un crimen «hipotético» al estilo de **The Minority Report**(Philip K. Dick, 1956)?

La aplicación de la ciencia de datos para predecir comportamientos delictivos, como detener a alguien basado en el potencial de cometer un crimen, se enfrentaría a serias cuestiones éticas y legales, incluyendo sesgos en los datos, falta de transparencia, y el riesgo de falsos positivos. La detención por crímenes no cometidos, al estilo de "The Minority Report", violaría muchos principios legales como la presunción de inocencia y daría pie a muchas falsas acusaciones.

- g) ¿Cuál de los términos fundamentales de la Ciencia de datos se trata en el siguiente video: <https://www.youtube.com/watch?v=DQWI1kvmwRg>? ¿Podrías explicarlo brevemente?

El sobreajuste, o **overfitting****, es cuando un modelo de aprendizaje automático aprende los datos de entrenamiento tan detalladamente que incluye el ruido, lo que afecta negativamente su capacidad para realizar predicciones precisas sobre nuevos datos no vistos. Es como memorizar las respuestas de un examen sin entender realmente los conceptos, lo que falla al enfrentar preguntas formuladas de manera diferente. Para evitarlo, se utilizan técnicas como la validación cruzada, la regularización, y los métodos de ensamble, buscando que el modelo mantenga un buen equilibrio entre capturar la esencia de los datos y poder generalizar bien a nuevos ejemplos.**

- h) Según el material compartido en el aula, ¿qué rol se define en clave de humor y cómo se define?

El rol definido en clave de humor es el del científico de datos. Se insinúa que, aunque se pueda simplificar su trabajo a la gestión de grandes volúmenes de datos para obtener ventajas competitivas, la realidad es que su papel es mucho más complejo e involucra un profundo análisis y comprensión de los datos para informar decisiones estratégicas en organizaciones.

A continuación, vamos a ver un ejemplo práctico.

“En un proyecto de análisis de datos en salud, cada rol trabaja de manera independiente, demostrando la confianza en la capacidad de cada profesional para llevar a cabo sus tareas eficientemente. La Data Analyst inicia el proceso haciendo visualizaciones sin una revisión exhaustiva de la calidad de los datos, confiando en la eficacia de los sistemas de recopilación de datos de hospitales y laboratorios. El Data Engineer comienza a crear modelos sin realizar una limpieza previa de los datos recopilados de diversas fuentes, lo que le permite concentrarse directamente en la construcción de modelos predictivos para predecir diagnósticos médicos. La Data Scientist, a pesar de las posibles limitaciones en los datos provenientes de registros médicos y pruebas de laboratorio, lleva a cabo un análisis exploratorio sólido, considerando la situación como una oportunidad para identificar patrones complejos y desafiantes en la información de salud. Por último, el Machine Learning Engineer valora la oportunidad de mejorar la calidad de los datos y modelos al regresar al principio del proceso y preparar toda la infraestructura para el almacenamiento y el flujo de los datos, demostrando su capacidad para asegurar la adecuada preparación de los datos antes de su análisis.

Este enfoque, aunque aparentemente carece de coordinación y limpieza inicial de los datos, explota y resalta las habilidades y capacidades individuales de cada miembro del equipo en el análisis de datos en salud. Además, aumenta la velocidad del procesamiento de los datos y entrenamiento de los modelos ahorrando tiempo en pasos que no son importantes ni estrictamente necesarios.”

- 2) Analiza y comenta de manera crítica el ejemplo práctico expuesto teniendo en cuenta el material de los Bloques 1 y 2 de nuestro curso y los siguientes aspectos:
 - a) El orden de los pasos y procesos descritos en el ejemplo.

El orden descrito en el ejemplo práctico parece desafiar la metodología tradicional en la ciencia de datos, donde la limpieza y preparación de datos preceden cualquier análisis o modelado. Este enfoque puede aumentar el riesgo de generar insights incorrectos o modelos predictivos ineficaces debido a la mala calidad de los datos. La fase de análisis debe basarse en datos limpios y preparados adecuadamente para evitar conclusiones erróneas.

b) Las responsabilidades de cada rol.

Cada rol en el equipo de análisis de datos tiene responsabilidades claras que, si se llevan a cabo aisladamente sin una coordinación efectiva, pueden comprometer la eficacia del proyecto. La Data Analyst debe asegurarse de la calidad de los datos antes de la visualización, y el Data Engineer debe enfocarse en la limpieza y preparación de datos antes de modelar. La Data Scientist y el Machine Learning Engineer deben colaborar estrechamente para garantizar que los modelos sean robustos y los datos de alta calidad.

c) La importancia de los distintos pasos mencionados en el ejemplo.

La limpieza y preparación de datos son fundamentales y deben ser los primeros pasos en cualquier proyecto de análisis de datos. Saltarse estos pasos puede llevar a insights erróneos y afectar negativamente la toma de decisiones basada en el análisis de datos. La modelización y visualización de datos deben basarse en datos limpios y de calidad para ser efectivas.

d) La interrelación entre los distintos roles así como “la aparentemente carencia de coordinación y limpieza inicial de los datos” que se señala en el ejemplo práctico.

La efectividad de un equipo de análisis de datos depende de la coordinación y colaboración entre sus miembros. La falta de coordinación y limpieza inicial de datos señalada en el ejemplo práctico puede llevar a un desperdicio de recursos y tiempo, ya que los problemas de calidad de datos eventualmente requerirán atención, retrasando el análisis y la generación de modelos predictivos. La interacción y comunicación constantes entre los roles son esenciales para el éxito del proyecto.

3) ¿Cómo quedaría redactado el ejemplo si tuvieras que añadir algún cambio? Al margen de que quieras explicar o no los cambios, lo que se pide en esta pregunta es la versión del ejemplo práctico con los cambios que consideres necesarios basándote en su análisis crítico.

El proceso inicia con la Data Analyst, quien antes de proceder a la visualización de los datos, lleva a cabo una revisión preliminar de la calidad de los mismos. Esta revisión asegura que los datos sean fiables y representativos, eliminando o corrigiendo anomalías.

El Data Engineer, antes de comenzar con la creación de modelos predictivos, se enfoca en la limpieza y preparación de los datos recopilados de diversas fuentes. Este paso es vital para asegurar que los modelos sean construidos sobre una base sólida y confiable de datos limpios y bien estructurados, lo que a su vez permite una concentración más efectiva en la construcción de modelos predictivos para predecir diagnósticos médicos.

La Data Scientist, consciente de las limitaciones en los datos provenientes de registros médicos y pruebas de laboratorio, lleva a cabo un análisis exploratorio de los datos ya limpios y preparados por el Data Engineer. Este análisis permite identificar patrones complejos y desafiantes en la información de salud, considerando la calidad de los datos como una base para generar insights valiosos.

Finalmente, el Machine Learning Engineer trabaja en estrecha colaboración con el resto del equipo desde el inicio del proyecto. Este enfoque integrado permite identificar oportunidades para mejorar la calidad de los datos y los modelos desde las etapas iniciales. El Machine Learning Engineer prepara toda la infraestructura necesaria para el almacenamiento eficiente y el flujo de los datos, asegurando que estén adecuadamente preparados para su análisis.

Este enfoque coordinado no solo destaca las habilidades y capacidades individuales de cada miembro del equipo en el análisis de datos en salud, sino que también mejora significativamente la calidad del proceso de análisis de datos.

Observación: En la práctica, las funciones descritas anteriormente pueden ser asimiladas por diferentes roles, en función del tamaño de la empresa, del equipo y los roles que lo constituyen, por lo que vuestras respuestas deben considerar solamente el equipo que se ha planteado para el caso del ejemplo.

Criterios de evaluación

- El apartado 1 se valorará con **1 punto** como máximo, **0.125 puntos** por cada rol bien definido.

- El **apartado 2** se valorará con **0.5 puntos** como máximo, **0.125 puntos** por cada aspecto correctamente argumentado: orden lógico del proceso, responsabilidades de cada rol, importancia de los pasos, interrelación y cooperación entre los distintos roles.
- El **apartado 3** se valorará con **0.5 puntos** como máximo, **0.125 puntos** por cada aspecto correctamente tratado en el ejemplo práctico: orden lógico del proceso, responsabilidades de cada rol, importancia de los pasos, interrelación y cooperación entre los distintos roles.
- La extensión máxima del apartado 1 es de **500 palabras**.
- La extensión máxima del apartado 2 es de **600 palabras**.
- La extensión máxima del apartado 3 es de **300 palabras**.
- Se valorará la capacidad de identificar los textos que justifican las respuestas.
- Todas las respuestas tienen que estar desarrolladas y argumentadas.

Ejercicio 2 (30% puntuación)

En este contexto planteado, imaginad que un sistema de salud solicita sus servicios como científicos de datos con el objetivo de mejorar su atención médica. Con este fin, desarrollad los siguientes puntos:

- 1 Describe brevemente qué problema o problemas de los descritos anteriormente os gustaría solucionar. Establece claramente **una pregunta o una hipótesis de trabajo**.

Enfocándonos en el desafío de la Medicina Participativa, que busca incrementar la implicación de los pacientes en su propio cuidado de salud para promover una toma de decisiones informada y mejorar el cumplimiento del tratamiento, reformulamos nuestra pregunta de trabajo a: ¿Cómo podemos utilizar la tecnología y los datos para fomentar una mayor participación de los pacientes en la gestión de su salud, asegurando que tengan la información necesaria para tomar decisiones informadas sobre su tratamiento?

- 2 **Captura de los datos.** Basándote en el caso de estudio proporcionado ¿Qué datos crudos utilizarías? ¿Dónde podrías obtenerlos? No hace falta que os limitéis exclusivamente a las fuentes de datos proporcionadas directamente por el sistema de salud.

Para abordar este desafío, recopilaríamos datos de:

Historiales médicos electrónicos (HME), para obtener un panorama completo de la salud del paciente, incluyendo diagnósticos, tratamientos previos y resultados de pruebas.

Plataformas de feedback de pacientes, como encuestas de satisfacción y foros en línea, para entender las necesidades, preocupaciones y preferencias de los pacientes.

Aplicaciones de salud móviles e inalámbricas, que pueden proporcionar datos en tiempo real sobre la salud y el estilo de vida del paciente, como la actividad física, el sueño y la alimentación.

Bases de datos de educación en salud, para identificar los recursos informativos más efectivos y las mejores prácticas en la comunicación de información médica a pacientes.

Añadir una funcionalidad de suscripciones dentro de la aplicación móvil para que los usuarios puedan compartir sus hábitos alimenticios y estilos de vida que han sido favorables en su vida

- 3 **Almacenamiento y/o Preprocessing.** ¿Cómo y dónde los guardarías? ¿Les harías alguna modificación (cambio de formato, *data cleansing*, anonimización, etc.)?

Al almacenar los datos limpios, normalizados y anonimizados en una blockchain, avanzamos hacia una infraestructura de almacenamiento segura y transparente, cumpliendo con las normativas de privacidad. La blockchain asegura la integridad de los datos de salud mediante un registro inmutable, promoviendo la participación activa del paciente en su cuidado gracias a su naturaleza descentralizada. Este método no solo fortalece la confianza en el sistema de salud al garantizar la protección de la información personal, sino que también facilita el intercambio seguro de datos entre investigadores y profesionales de la salud.

Utilizando la blockchain, los pacientes ejercen un control mayor sobre sus datos, lo que fomenta una mayor implicación en la gestión de su salud. Paralelamente, investigadores pueden acceder a datos anonimizados para estudios, sin comprometer la privacidad. Esta innovación revela patrones clave para la prevención de enfermedades y la promoción de hábitos saludables, contribuyendo significativamente a la medicina participativa y a la generación de conocimientos colectivos sobre la salud global.

- 4 **Análisis y/o visualización.** ¿Qué métricas, indicadores clave o datos calculados en base a datos crudos utilizarías para tu hipótesis? ¿Alguna idea de qué algoritmos u operaciones de las enumeradas en el caso de estudio proporcionado (a nivel muy general) usarías? (Por ejemplo: Una clasificación, regresión o agrupamiento (clustering))? ¿Cómo medirías el éxito del proyecto?

Desarrollaríamos modelos para analizar las interacciones entre los pacientes y el sistema de salud, identificando patrones que indican una alta o baja participación en el cuidado de su salud. Usaríamos técnicas de análisis de sentimientos en feedback de pacientes y machine learning para predecir los factores que influyen en el cumplimiento del tratamiento y la satisfacción del paciente. Se crearían dashboards personalizados para pacientes y proveedores de salud, mostrando progresos, alertas y recomendaciones basadas en datos para mejorar la toma de decisiones y la participación del paciente.

Para medir el éxito del proyecto, nos enfocaríamos en indicadores clave como:

1. Aumento en la participación activa de los pacientes mediante el uso de herramientas digitales.
2. Mejora en el cumplimiento del tratamiento, observando cambios en las tasas de adherencia.
3. Elevación de la satisfacción del paciente, basada en encuestas y análisis de sentimientos.

4. Avances en la alfabetización en salud de los pacientes, evaluando su comprensión sobre condiciones y tratamientos.
5. Impacto positivo en los resultados de salud, como la reducción de incidencias de enfermedades y mejora en indicadores de control de enfermedades crónicas.
6. Feedback positivo de los proveedores de salud sobre la eficacia de la comunicación y gestión de pacientes.

Estos indicadores nos darán una visión integral del impacto del proyecto, permitiendo ajustes continuos para maximizar el éxito.

- 5 **Publicación.** Es importante que vuestros análisis generen conocimiento útil y/o valor en vuestro contexto de trabajo. ¿Qué harías con los resultados obtenidos? ¿Cómo los presentarías? ¿Quién sería el destinatario final de tus resultados?

Los insights generados serían compartidos con los profesionales de la salud y los pacientes a través de plataformas interactivas y aplicaciones móviles, diseñadas para educar, informar y empoderar a los pacientes en la gestión de su salud. Se organizarían talleres y seminarios web para educar a los pacientes sobre cómo interpretar y utilizar esta información. Además, se desarrollarían guías y materiales de apoyo para ayudar a los profesionales de la salud a comunicar efectivamente la información y fomentar una mayor participación de los pacientes en su cuidado.

Criterios de evaluación

- Cada **apartado** se valorará con **0.6 puntos** como máximo.
- Extensión máxima por apartado: **250 palabras**.
- Todas las respuestas tienen que estar **desarrolladas y argumentadas**.
- Las respuestas tienen que mostrar conexión con el temario teórico, de manera que se pueda percibir el aprendizaje desarrollado.
- Siendo un ejercicio de **asunción y simulación de un rol**, los errores que pueda haber, en temas que no sean estrictamente vinculados con el temario teórico estudiado, no serán tenidos en cuenta.

Ejercicio 3 (30% puntuación)

Caso 2: Agricultura de precisión.

La transformación digital en el sector agrícola representa uno de los avances más significativos hacia una producción sostenible y eficiente, marcada especialmente por la aplicación de la Ciencia de Datos e Inteligencia Artificial. Un caso emblemático de esta evolución es el enfoque adoptado por John Deere, empresa pionera en la fabricación de maquinaria agrícola, que ha redefinido el concepto de maquinaria agrícola a través de la integración de tecnologías avanzadas. Este análisis explora cómo John Deere ha utilizado la Ciencia de Datos para transformar sus productos y servicios, sus objetivos detrás de esta integración tecnológica, y el impacto generado tanto para la empresa como para la sociedad.

1. Contexto de la aplicación de la Ciencia de Datos en el caso escogido

Durante las últimas tres décadas, John Deere ha transitado de fabricar maquinaria agrícola puramente mecánica a desarrollar lo que su CTO, Jahmy Hindman, describe como "suites de sensores móviles con capacidad computacional" (The Verge). Esta transición es el núcleo de la estrategia "industria inteligente" de John Deere, que busca ofrecer un sistema completo que incluye el equipo, los datos, el análisis y la automatización necesarios para que los agricultores brinden cuidados individualizados a gran escala. En el Data + AI Summit de este año, representantes de John Deere explicaron cómo su plataforma de datos permite manejar volúmenes crecientes de información generada por su maquinaria, integrar nuevas fuentes de datos, y unificar estos datos para mejorar los resultados de los clientes.

2. Objetivos y expectativas. ¿Para qué usaron la Ciencia de Datos?

El principal objetivo de la aplicación de la Ciencia de Datos en John Deere es habilitar la agricultura de precisión a gran escala, permitiendo a los agricultores optimizar el uso de recursos como agua, nutrientes y pesticidas, y maximizar los rendimientos de sus cultivos. Esto se logra a través de la recopilación y análisis de datos masivos provenientes de sensores en la maquinaria, sistemas de visión, y conectividad inalámbrica, entre otros. Con la ayuda de algoritmos de machine learning, John Deere puede analizar la calidad de los granos en tiempo real y ajustar automáticamente los parámetros operativos de las máquinas para prevenir daños. Además, la integración de fuentes de datos externas, como información meteorológica, permite anticipar necesidades de riego o aplicación de pesticidas, haciendo el proceso agrícola más eficiente y menos dependiente de intervenciones humanas directas.

3. Impacto para la empresa y/o la sociedad

El impacto de la aplicación de la Ciencia de Datos en John Deere se manifiesta en múltiples niveles. Para la empresa, representa un avance hacia la consolidación de un modelo de negocio basado en la oferta de soluciones integradas que van más allá de la maquinaria agrícola, posicionándose como líder en la transformación digital del sector agrícola. Este enfoque no solo ha abierto nuevas vías de ingresos a través de servicios digitales, sino que también ha fortalecido su relación con los clientes al ofrecerles herramientas que mejoran directamente su productividad y éxito.

Para la sociedad, el impacto es aún más profundo. La optimización de recursos que permite la agricultura de precisión tiene el potencial de reducir significativamente el impacto ambiental de la agricultura. Como indica John Deere, la reducción del uso de químicos hasta en un 70% es una realidad alcanzable, lo que disminuye la contaminación del suelo y el agua, y reduce la dependencia de pesticidas y fertilizantes. Además, al incrementar los rendimientos de las cosechas, se contribuye a la seguridad alimentaria en un contexto global de crecimiento poblacional y cambio climático.

Sin embargo, mirando hacia el futuro, la ambición de alcanzar un 0% de uso de pesticidas se vuelve cada vez más concebible. Esto se debe en gran medida a la creciente contribución de conocimientos y datos por parte de la comunidad apasionada por la agricultura ecológica, combinada con los avances tecnológicos en el sector. Estos datos, que abarcan desde técnicas de cultivo alternativas hasta métodos de control biológico de plagas, cuando se analizan y aplican correctamente, pueden desbloquear nuevas estrategias de cultivo que prescinden completamente de los pesticidas químicos.

Para alcanzar este ideal, será crucial la colaboración entre agricultores, científicos, tecnólogos y la sociedad en general, aprovechando la riqueza de conocimientos compartidos en internet y las capacidades de análisis de datos para experimentar, validar y escalar prácticas agrícolas que sean genuinamente regenerativas y libres de químicos. La visión de una agricultura sin pesticidas no solo es un testimonio de lo que la tecnología y la ciencia de datos pueden lograr, sino también un llamado a reimaginar nuestras prácticas agrícolas para el bienestar del planeta y de las futuras generaciones.

Este caso de John Deere ilustra cómo la Ciencia de Datos y la Inteligencia Artificial no son solo herramientas para el avance tecnológico y la eficiencia empresarial, sino también catalizadores de sostenibilidad y bienestar social. La capacidad de gestionar la agricultura a una escala nunca antes vista, cuidando cada planta individualmente, representa un cambio paradigmático en cómo entendemos y practicamos la agricultura, abriendo la puerta a un futuro donde la tecnología y la sostenibilidad van de la mano.

Criterios de evaluación

- Cada apartado se valorará con **1 punto** como máximo.
- Relacionar lo encontrado en el caso seleccionado con el material lectivo.
- Analizar las aplicaciones de Ciencia de datos.
- Contextualizar la ventaja competitiva que aporta la Ciencia de datos.
- La redacción que contenga **menos de 500 palabras o más de 1000 palabras se penalizará con 0.2 puntos**.

Ejercicio 4 (20% puntuación)

- [Barcelona Open Data](#)

Analiza el portal y responde:

1. ¿Qué es Open Data Barcelona?

Open Data Barcelona es un servicio del Ajuntament de Barcelona dedicado a proporcionar datos abiertos.

2. ¿Qué son los datos abiertos? ¿Cuáles son sus principales características?

Los datos abiertos se caracterizan por ser accesibles al público de manera gratuita y están destinados a fomentar la transparencia, la innovación y el desarrollo de aplicaciones que beneficien a la sociedad.

Después de explorar y analizar el portal, responded:

- ¿Cuántos datasets hay en el Catálogo de Open Data BCN? Describid el portal y los contenidos que ofrece, comentando al menos tres de las secciones del menú principal.



Hay 605 datasets:

Buscar conjuntos de datos...

605 conjuntos de datos encontrados

Ordenar por: Fecha publicación (DESC)

El portal Open Data BCN se presenta como un espacio integral para el acceso a los datos abiertos del Ayuntamiento de Barcelona. Ofrece diferentes secciones en su menú principal:

Visualizaciones y Aplicaciones: Área que muestra representaciones gráficas de datos y aplicaciones desarrolladas a partir de estos. Es un recurso para entender mejor los datos a través de su visualización.

Estadísticas: Se proporcionan métricas y análisis estadísticos de los datos disponibles en el portal, lo cual puede ser de interés para investigadores, periodistas y ciudadanos interesados en la información municipal.

Desarrolladores: Aquí se incluye una introducción y recursos clave como APIs para la gestión del catálogo y consulta de recursos CSV, tokens, y límites de acceso a datasets. También hay enlaces al GitHub del Ayuntamiento, secciones de programación, enlaces de interés y otros recursos útiles para desarrolladores que buscan utilizar y aplicar los datos abiertos en sus proyectos.

A continuación en la Clasificación por temas podremos encontrar el tema Territorio. A su vez, en este tema podremos encontrar datasets como el [Inventario de los separadores de carriles bici de la red de infraestructuras de la ciudad de Barcelona](#).

4. ¿Cuál es el grado de apertura del dataset? Comentad por qué motivo se le da ese grado de apertura(según la propia página), si hay en el dataset recursos con un formato asociado a datos jerárquicos y describid brevemente el esquema que se está utilizando, quién creó el esquema de clasificación, etc.



El conjunto de datos del Inventario de los separadores de carriles bici de la red de infraestructuras de la ciudad de Barcelona se ofrece en múltiples formatos, incluidos CSV y JSON. Estos formatos son legibles por máquinas y son indicativos de un nivel de tres estrellas de apertura según el esquema utilizado por el portal Open Data BCN. El esquema de clasificación es creado y mantenido por el Ajuntament de Barcelona, la fuente de estos datos.

El nivel de tres estrellas significa que los datos no sólo están disponibles en línea, sino que también están presentados en un formato estructurado. A pesar de esto, para alcanzar niveles más altos de apertura, como cuatro o cinco estrellas, los datos tendrían que estar disponibles en formatos abiertos no propietarios y estar interconectados con otros datos para proporcionar un contexto más rico, permitiendo su integración en la web de datos.

El hecho de que los datos estén en formatos como CSV y JSON permite que puedan ser procesados y utilizados fácilmente en aplicaciones y sistemas, promoviendo así su uso y reutilización en una variedad de proyectos y servicios. Los datos también están disponibles para su descarga en archivos ZIP, lo cual es conveniente para el manejo de grandes cantidades de información.

5. ¿Qué formatos de datos podemos encontrar en el portal? Comentad al menos tres formatos diferentes y su grado de apertura.

En el portal Open Data BCN, podemos encontrar conjuntos de datos en diversos formatos que reflejan diferentes grados de apertura. Aquí hay tres ejemplos con sus respectivos grados de apertura:

CSV (Comma-Separated Values): Este es un formato de texto plano utilizado para representar datos tabulares. Cada línea del archivo es un registro de datos. Los registros están separados por saltos de línea y las columnas por comas. El formato CSV es legible por máquinas y por humanos hasta cierto punto, y generalmente se considera como un formato de tres estrellas de apertura porque es estructurado y puede ser utilizado por una variedad de herramientas de software, pero no es un formato abierto estándar como RDF o JSON.

JSON (JavaScript Object Notation): JSON es un formato basado en texto que se utiliza para representar datos estructurados basados en la sintaxis de objetos de JavaScript. Es ampliamente utilizado para intercambiar datos en aplicaciones web. JSON tiene un grado de apertura de tres estrellas ya que es legible por máquinas y es más estructurado y jerárquico en comparación con CSV.

El formato XML (eXtensible Markup Language) es un lenguaje de marcado que define un conjunto de reglas para la codificación de documentos de manera que sea legible tanto para máquinas como para humanos. En el contexto de los datos abiertos, el XML es apreciado por su capacidad de representar datos complejos y jerárquicos con estructura anidada. A menudo, el XML es utilizado para intercambiar datos en la web y para almacenar datos con una estructura compleja. En términos de grado de apertura, los archivos XML generalmente reciben un grado de apertura de tres estrellas, ya que son formatos legibles por máquinas y, debido a su estructura jerárquica, pueden representar relaciones más complejas entre los datos que los formatos como CSV.

Criterios de evaluación

- Los dos primeros apartados se valorarán con **0.25 puntos** como máximo respectivamente.
- El resto de apartados (del 3 al 5), se valorarán con **0.5 puntos como máximo** cada uno.
- Extensión máxima por pregunta: **350 palabras**.
- Explorar los portales de búsqueda de datos abiertos y comprender el contenido de estos.
- Comprender el significado de Open Data y su utilidad.
- Comprender el esquema de 5 estrellas.
- Comentar los distintos formatos de los datos.