



MERKLE DATATHON

V EDICIÓN

© 2024 Merkle, Inc. All Rights Reserved. Confidential.

MERKLE



Gijón | Innovación

Gijón | impulsa



Universidad de
Oviedo

GIJÓN **LUCE**
LABORATORIO URBANO

MERKLE

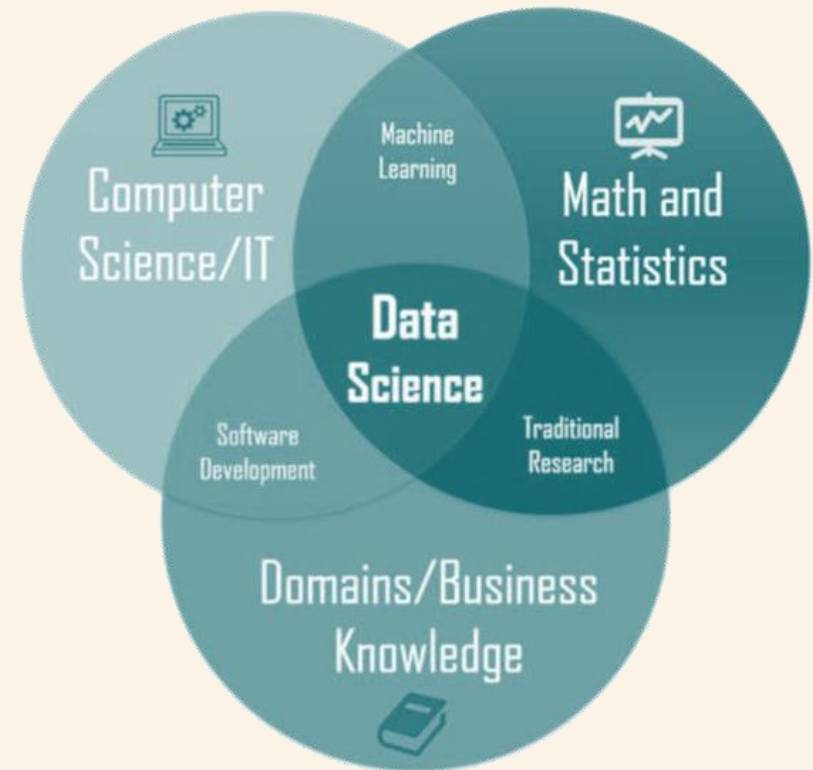
Merkle es una compañía especializada en Customer Experience Management, experta en desarrollar experiencias de cliente únicas y personalizadas a través de diferentes plataformas y dispositivos.



CIENCIA DE DATOS EN MERKLE

Resolución de problemas empresariales mediante:

- Estadística.
- Matemáticas.
- Programación (R, Python, SQL).
- Conocimiento y tratamiento de modelos de datos.
- Visualización de datos.
- Conocimiento de Negocio.
- Mente crítica.



DATATHON

Un “Datathon” es un evento en el que un grupo de científicos de datos y expertos en un área de conocimiento se **reúnen y trabajan en forma multidisciplinaria con el objetivo de abordar preguntas de investigación importantes** para la especialidad a través del análisis de grandes volúmenes de datos.

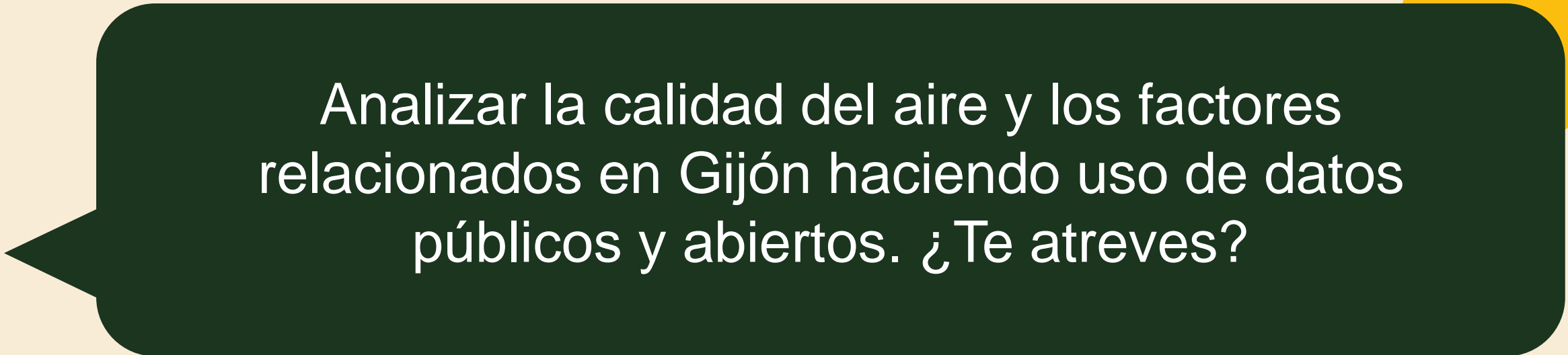


DESAFÍO 2024

El desafío de esta edición surge de la idea propuesta por el Ayuntamiento de Gijón a la Cátedra MERKLE de Ciencia de Datos y Marketing, de cruzar datos de **contaminación y calidad del aire** con datos públicos disponibles de **movilidad urbana** en Gijón y **meteorología**. El objetivo es **visibilizar el valor y calidad de los datos abiertos, y mostrar la información y las oportunidades de los modelos predictivos** a la ciudadanía en este caso, con una clara vocación de servicio público.



El ayuntamiento de Gijón te propone un reto:

A large, dark green speech bubble with a white border, containing the challenge text. A yellow sun is partially visible behind the bubble on the right side.

Analizar la calidad del aire y los factores relacionados en Gijón haciendo uso de datos públicos y abiertos. ¿Te atreves?

Hora	Tarea
9:15 - 9:40	Recibimiento y formación de equipos
09:40 - 10:00	Acto inaugural
10:00 - 10:30	Presentación del reto descriptivo – Parte I (Merkle)
10:30 - 12:30	Trabajo en cuestiones descriptivas
12:30 - 12:45	Presentación del reto predicción & campaña impacto social (medio ambiente) – Partes II & III (Merkle)
12:45 - 14:45	Trabajo en predicción & modelo (PRIMERA PREDICCIÓN)
14:45 - 15:30	COMIDA
15:30 - 17:45	Trabajo en predicción & modelo II (SEGUNDA PREDICCIÓN)
17:45 - 18:00	Envío de PPT y segunda predicción
18:00 - 19:20	Presentaciones (cronometrado entre 7 y 10 minutos)
19:20 - 19:45	Deliberación del jurado - Presentación resultado Merkle
19:45 - 20:00	Anuncio ganadores y entrega de premios



PREMIOS

Todos los grupos seréis evaluados de todos los bloques de los que consta el reto.



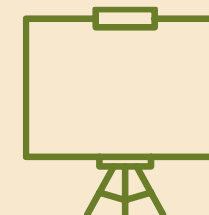
MEJOR PREDICCIÓN
Ganador del ranking



**MEJOR
PLANTEAMIENTO TÉCNICO**
Cuestionario



**MEJOR PROPUESTA
CON IMPACTO SOCIAL**
Pregunta



MEJOR PRESENTACIÓN
Se evalúa el contenido y la forma en que se hace la presentación.



CRITERIOS DE VALORACIÓN

Los criterios de valoración que tendremos en cuenta para cada uno de los premios son los siguientes:

Mejor predicción:

- El menor error en la predicción
- La originalidad y validez de las técnicas utilizadas para construir un modelo predictivo.

Mejor propuesta con impacto social:

- El ajuste de la propuesta a los análisis y conclusiones extraídas de la exploración de los datos.
- La originalidad y de la propuesta y su capacidad de ajuste a la realidad urbana de Gijón.
- Innovación y practicidad de las medidas y de la campaña de sensibilización.

Mejor presentación:

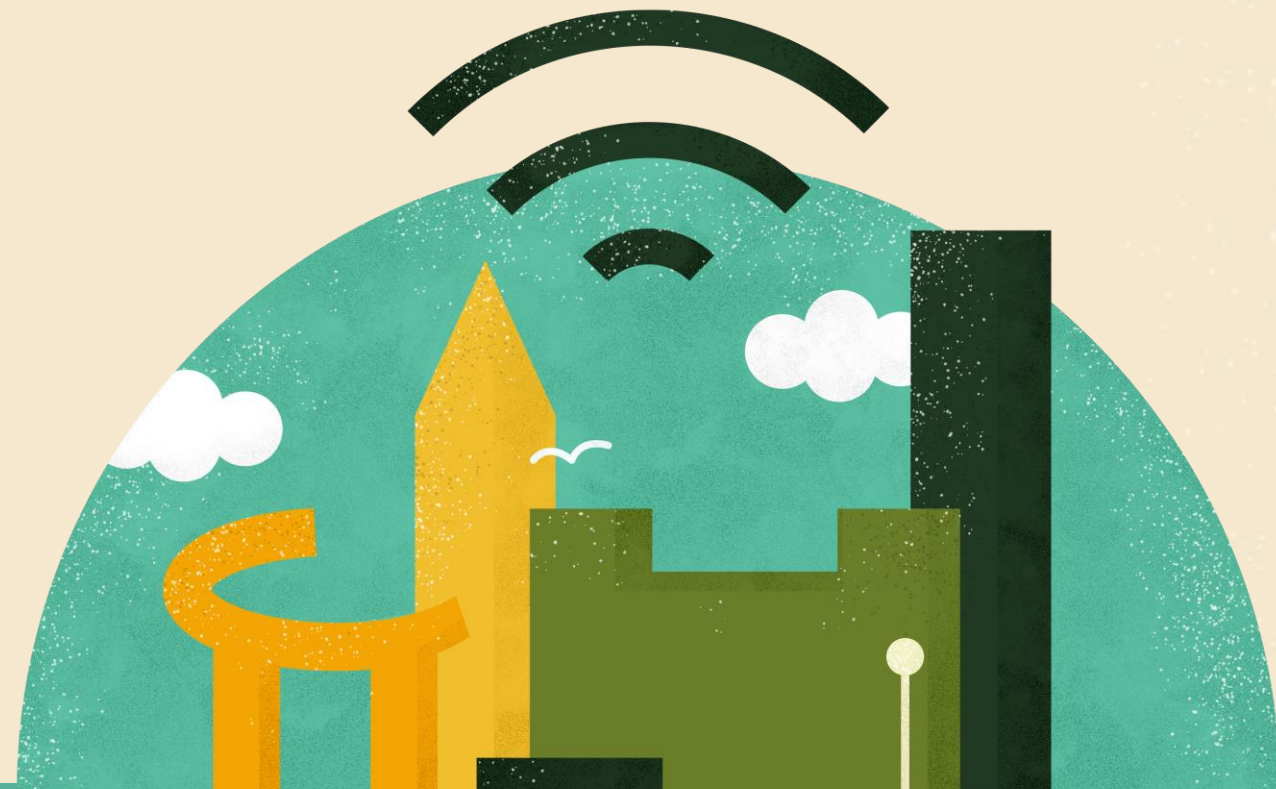
- Para optar a este premio es imprescindible que presenten al menos dos personas del grupo.
- Se valorará la propia presentación (estructura, diseño, claridad en la visualización) y las habilidades de comunicación de los representantes de cada equipo.

Mejor planteamiento técnico:

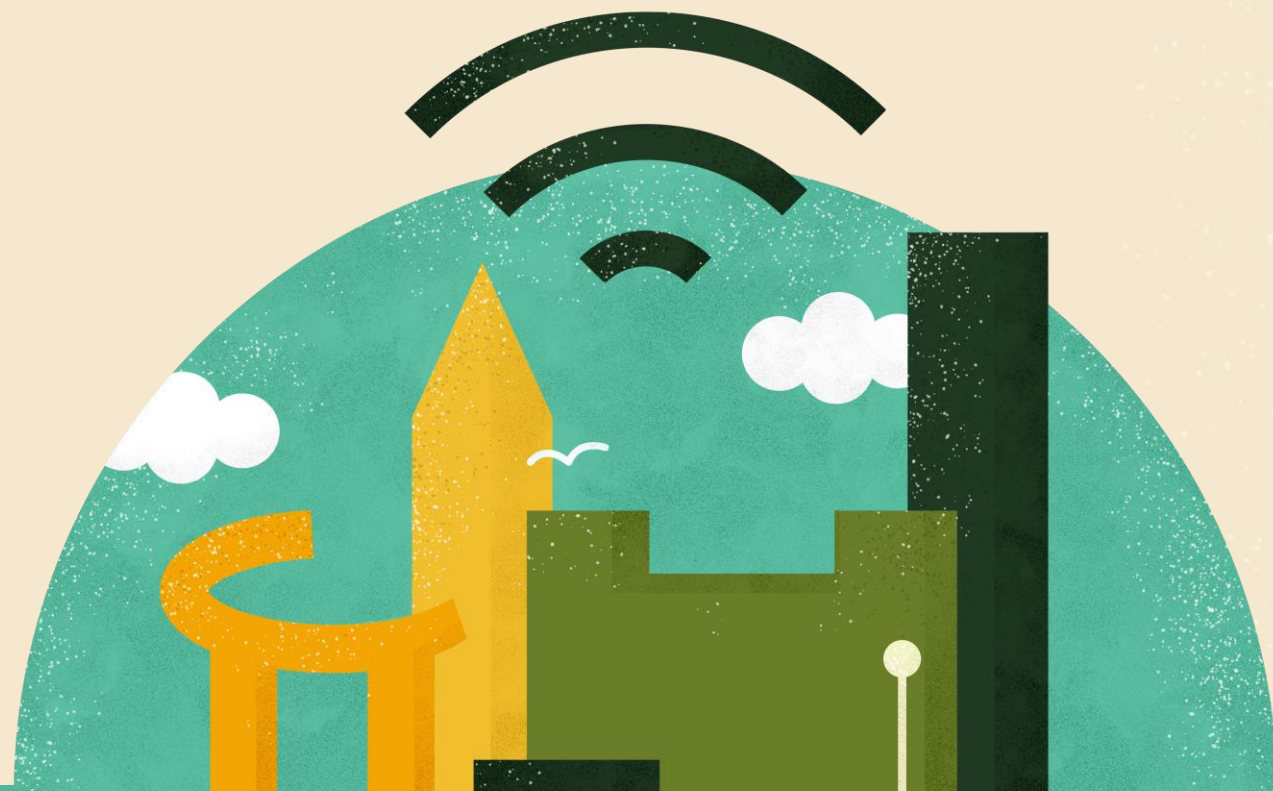
- La capacidad de demostrar si existe o no relación entre las variables, y si existen o no diferencias en los análisis que os pedimos.
- Valoraremos la elección adecuada de las técnicas y gráficos que expliquen relaciones entre variables o las diferencias en los datos.
- Igualmente, se valorará la correcta interpretación de los resultados y las explicaciones razonadas de las conclusiones extraídas.



BASES DE DATOS



DATOS DE CALIDAD DEL AIRE



DATOS DE CALIDAD DEL AIRE

Contaminantes en medición:

- **SO₂:** El dióxido de azufre (SO₂) está regulado en la normativa a causa de su potencial efecto sobre la salud y los ecosistemas.
- **NO₂:** Es un óxido de nitrógeno que representa la principal forma química con efectos adversos sobre la salud. El NO se oxida con facilidad, dando lugar a NO₂.
- **NO:** El otro óxido de nitrógeno medido. Tanto NO como NO₂ representan contaminantes generados por la oxidación del Nitrógeno atmosférico.
- **CO:** El monóxido de carbono es un gas emitido como consecuencia de la combustión que contribuye a la formación de gases de efecto invernadero.
- **O₃:** El ozono tiene un efecto positivo en la estratosfera (10-50 km de la superficie), ya que protege de la radiación ultravioleta. Sin embargo, a cotas inferiores, en la troposfera (la capa de la atmósfera en contacto con la tierra), se convierte en un contaminante que actúa como un agresivo agente oxidante.
- **PM₁₀:** Las partículas son una mezcla heterogénea y compleja de sustancias orgánicas e inorgánicas de tamaño y composición química muy variable. PM₁₀ hace referencia a las partículas de tamaño igual o inferior a 10 micras.
- **PM₂₅:** Igual que las PM₁₀ en este caso nos referimos a partículas de tamaño igual o inferior a 2,5 micras. Las partículas son los contaminantes del aire más importantes en términos de peligrosidad para la salud humana y en mayor medida las PM₂₅ por su reducido tamaño.

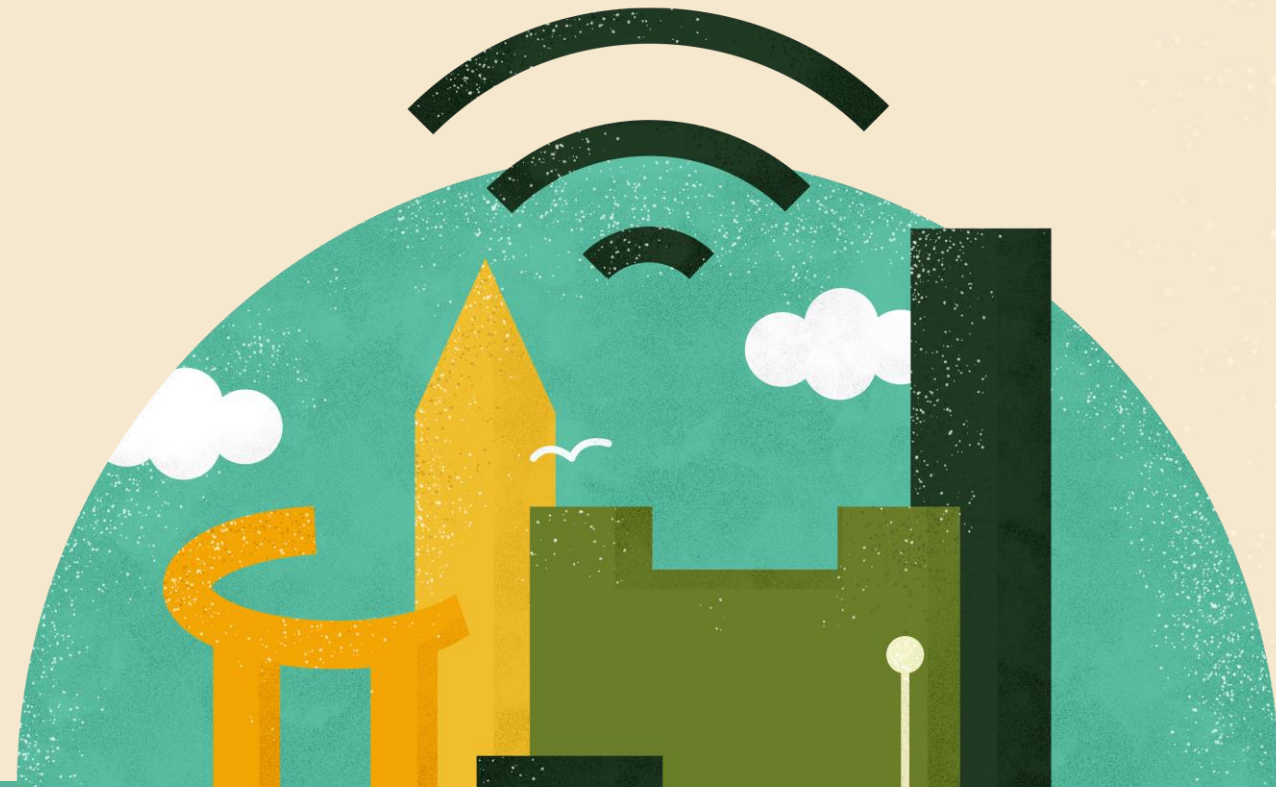


DATOS DE CALIDAD DEL AIRE. Estructura de datos.

Variable	Descripción	
date	Día de medición en formato fecha	
SO2_OesteAvdaArgentina	Concentración de dióxido de azufre de carbono medido en zona oeste de la ciudad (estación Avenida Argentina)	*solo para bloque I (Ej. este vs. Oeste)
SO2_EsteAvdaCastilla	Concentración de dióxido de azufre medido en zona este de la ciudad (estación Avenida Castilla)	*solo para bloque I (Ej. este vs. Oeste)
NO_OesteAvdaArgentina	Concentración de monóxido de nitrógeno medido en zona oeste de la ciudad (estación Avenida Argentina)	*solo para bloque I (Ej. este vs. Oeste)
NO_EsteAvdaCastilla	Concentración de monóxido de nitrógeno medido en zona este de la ciudad (estación Avenida Castilla)	*solo para bloque I (Ej. este vs. Oeste)
NO2_OesteAvdaArgentina	Concentración de dióxido de nitrógeno medido en zona oeste de la ciudad (estación Avenida Argentina)	*solo para bloque I (Ej. este vs. Oeste)
NO2_EsteAvdaCastilla	Concentración de dióxido de nitrógeno medido en zona este de la ciudad (estación Avenida Castilla)	*solo para bloque I (Ej. este vs. Oeste)
CO_OesteAvdaArgentina	Concentración de monóxido de carbono medido en zona oeste de la ciudad (estación Avenida Argentina)	*solo para bloque I (Ej. este vs. Oeste)
CO_EsteAvdaCastilla	Concentración de monóxido de carbono medido en zona este de la ciudad (estación Avenida Castilla)	*solo para bloque I (Ej. este vs. Oeste)
PM10_OesteAvdaArgentina	Concentración de partículas (menos 10 micras) medida en zona oeste de la ciudad (estación Avenida Argentina)	*solo para bloque I (Ej. este vs. Oeste)
PM10_EsteAvdaCastilla	Concentración de partículas (menos 10 micras) medida en zona este de la ciudad (estación Avenida Castilla)	*solo para bloque I (Ej. este vs. Oeste)
O3_OesteAvdaArgentina	Concentración de ozono medido en zona oeste de la ciudad (estación Avenida Argentina)	*solo para bloque I (Ej. este vs. Oeste)
O3_EsteAvdaCastilla	Concentración de ozono medido en zona este de la ciudad (estación Avenida Castilla)	*solo para bloque I (Ej. este vs. Oeste)
SO2_GijonGlobal	Concentración de dióxido de azufre de carbono global para todo Gijón	
NO_GijonGlobal	Concentración de monóxido de nitrógeno global para todo Gijón	
NO2_GijonGlobal	Concentración de dióxido de nitrógeno global para todo Gijón	
CO_GijonGlobal	Concentración de monóxido de carbono global para todo Gijón	
PM10_GijonGlobal	Concentración de partículas (menos 10 micras) global para todo Gijón	
O3_GijonGlobal	Concentración de ozono global para todo Gijón	



DATOS DE MOVILIDAD URBANA



DATOS DE MOVILIDAD EN GIJÓN.

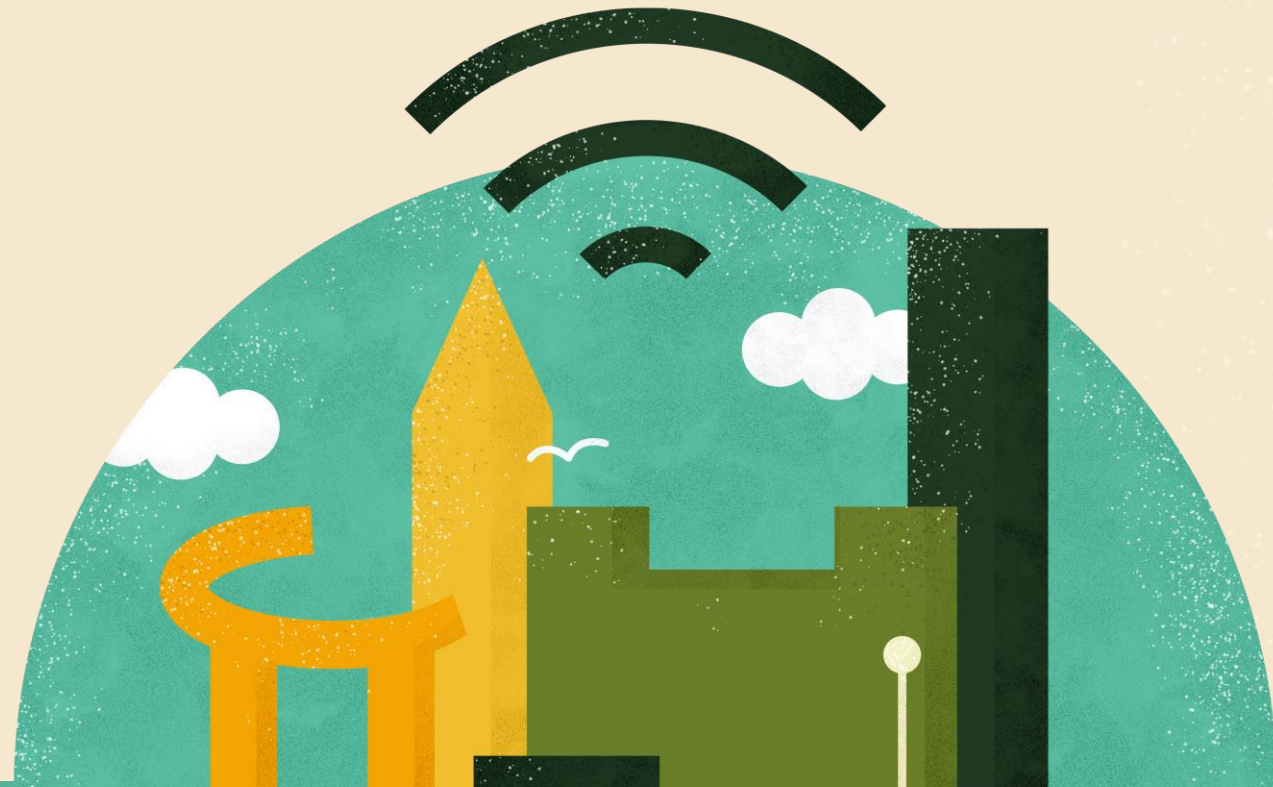
Los datos de movilidad provienen de estaciones de medición de tráfico. Existe en Gijón una importante red de sensores para el conteo de vehículos, agrupados en distintas estaciones de medición. Los datos de movilidad que os proporcionamos son medidas resumen en base diaria de las distintas estaciones de medición.

Estructura de datos.

Variable	Descripción
date	Día de medición en formato fecha
C01_BibioArenaMestas_trafico	Medición de tráfico en la zona de Bibio, Arena, Mestas
C02_CentroPlaya_trafico	Medición de tráfico en la zona centro y playa San Lorenzo
C03_CentroLlano_trafico	Medición de tráfico en la zona centro interior de Gijón
C04_LaviadaPoligono_trafico	Medición de tráfico en la zona de Laviada, Polígono de Pumarín y alrededores
C05_CotoViesques_trafico	Medición de tráfico en la zona de Coto y Viesques
C06_LlanoPumarin_trafico	Medición de tráfico en la zona del Llano y Pumarín
C07_LlanoMontevil_trafico	Medición de tráfico en la zona del Llano Alto y Montevil
C08_NuevoGijonBrana_trafico	Medición de tráfico en la zona de Nuevo Gijón y zona Braña
C09_LaCalzada_trafico	Medición de tráfico en la zona de la Calzada
C10_ContruecesGijonSurRoces_trafico	Medición de tráfico en la zona de Contrueces y Gijón Sur
C11_NatahoyoMoredaTremañes_trafico	Medición de tráfico en la zona de Natahoyo, Moreda y Tremañes
C12_Ceares_trafico	Medición de tráfico en la zona de Ceares



DATOS METEREOLÓGICOS



DATOS METEOROLÓGICOS

Los datos meteorológicos en base diaria se obtienen también de fuentes de datos abiertos. En este caso los datos se obtienen a través de los servicios API de la Agencia Estatal de Meteorología (AEMET).

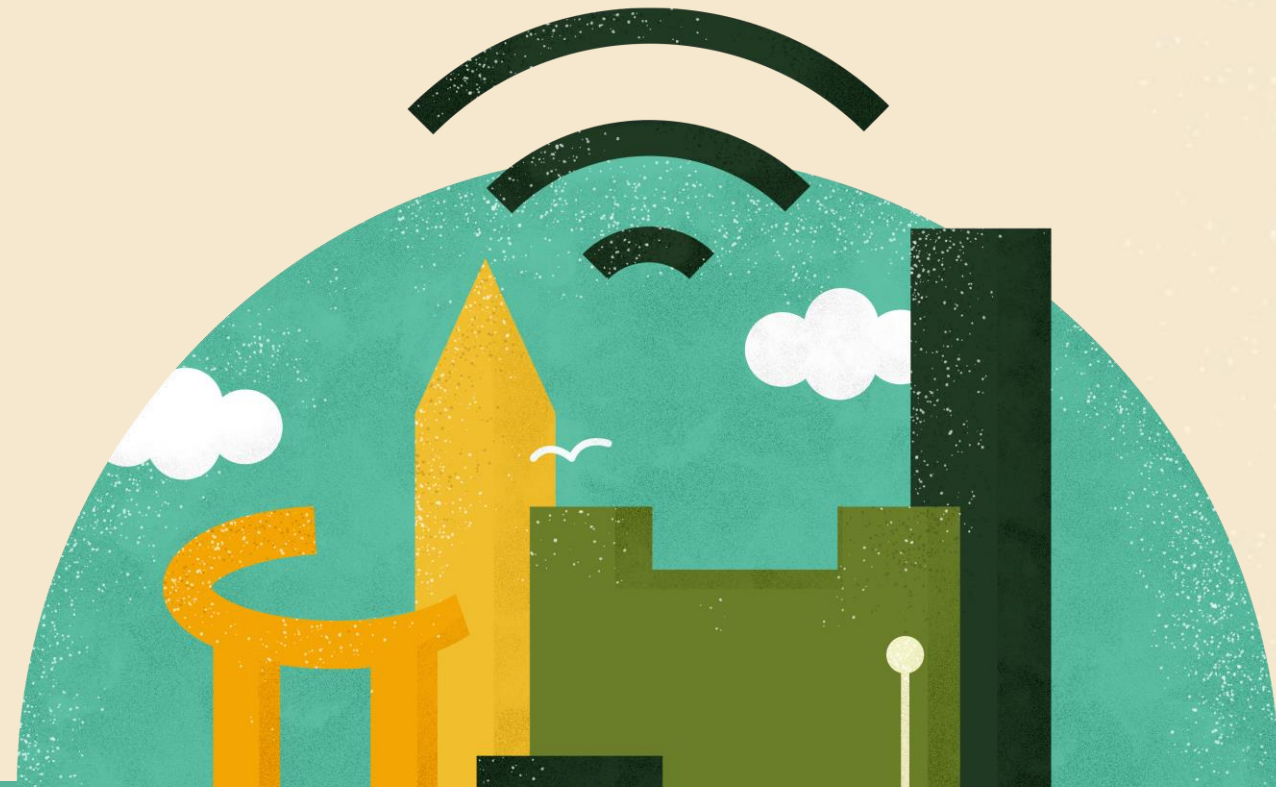
La base de datos se proporciona también en base diaria con información para las siguientes variables.

Estructura de datos.

Variable	Descripción
date	Día de medición en formato fecha
tmed_Gijon	Temperatura media del día de medición en Gijón en °C
prec_Gijon	Precipitación acumulada medida en mm/m2
sol_Gijon	Insolación medidas en horas de sol directo
presMin_Gijon	Presión mínima al nivel de referencia de la estación medida en hectopascal (hPa - coincide con mbar)
presMax_Gijon	Presión máxima al nivel de referencia de la estación medida en hectopascal (hPa - coincide con mbar)
dir_Gijon	Dirección de la racha máxima medido en decenas de grado
velmedia_Gijon	Velocidad media del viento medido en m/s
racha_Gijon	Racha máxima del viento medido en m/s



EL RETO



BLOQUES DEL RETO

El reto consta de **3 partes diferenciadas** en las que tendréis que analizar con distintos objetivos los datos que os proporcionamos

- 1 Preguntas sobre relaciones entre variables y descriptivo de datos.
- 2 Propuesta para la mejora la calidad de vida de los ciudadanos en Gijón.
- 3 Modelo predictivo de la calidad del aire.

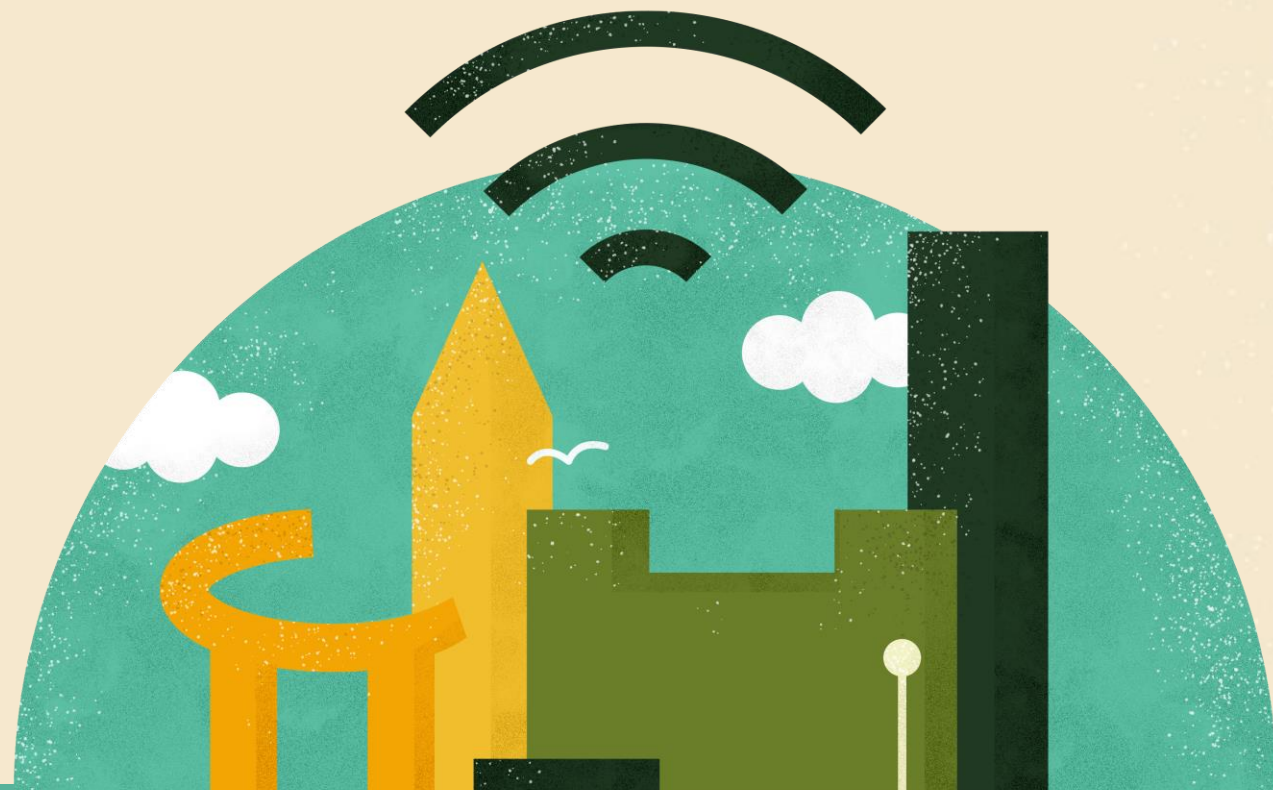
Antes de empezar el primer bloque, tomaos tiempo para familiarizaros con los datos y valorar el tipo de información que vais a manejar y el tipo de técnicas más adecuada para solucionar el reto.

Importante:

Todos los resultados y conclusiones relevantes de vuestro trabajo deberán recogerse en una presentación que será evaluada al final del reto.



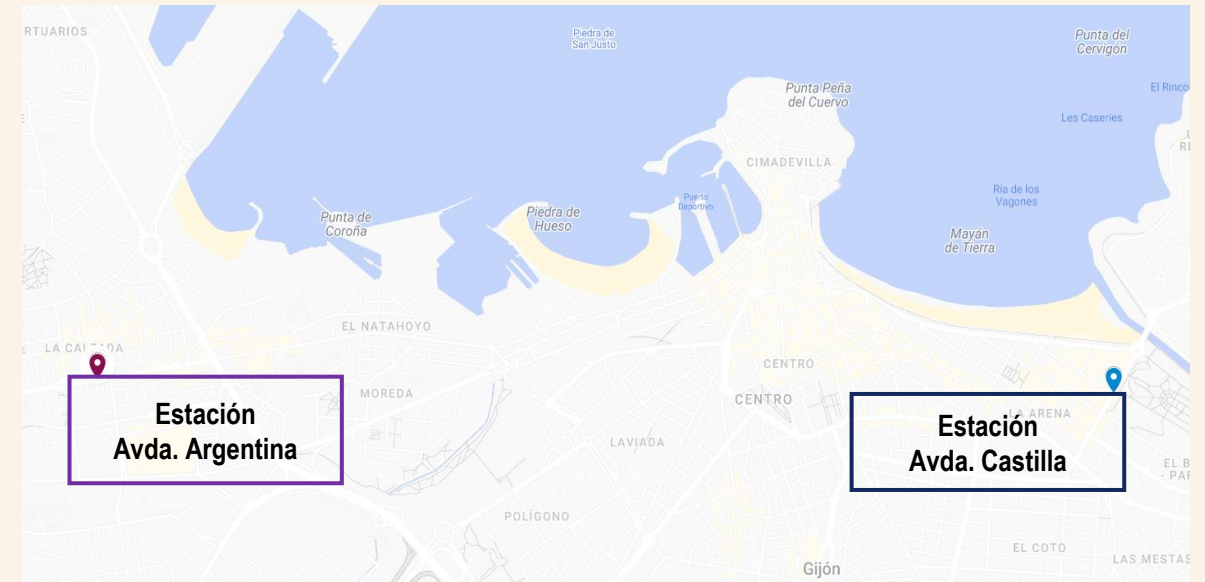
EL RETO: BLOQUE I



BLOQUE I: PREGUNTAS SOBRE RELACIONES ENTRE VARIABLES Y DESCRIPTIVO DE DATOS

Deberéis responder razonadamente a las siguientes preguntas:

1. ¿Qué variables de calidad del aire tienen más relación con la meteorología y con el tráfico o movilidad en la ciudad de Gijón?
2. ¿Existen diferencias espaciales de contaminación en Gijón? Comparando los datos de calidad del aire de la Avda. Castilla (ESTE) y Avda. Argentina (OESTE).
3. ¿Se advierten diferencias en el tráfico y la contaminación entre el periodo Covid y los mismos periodos en años anteriores y posteriores?



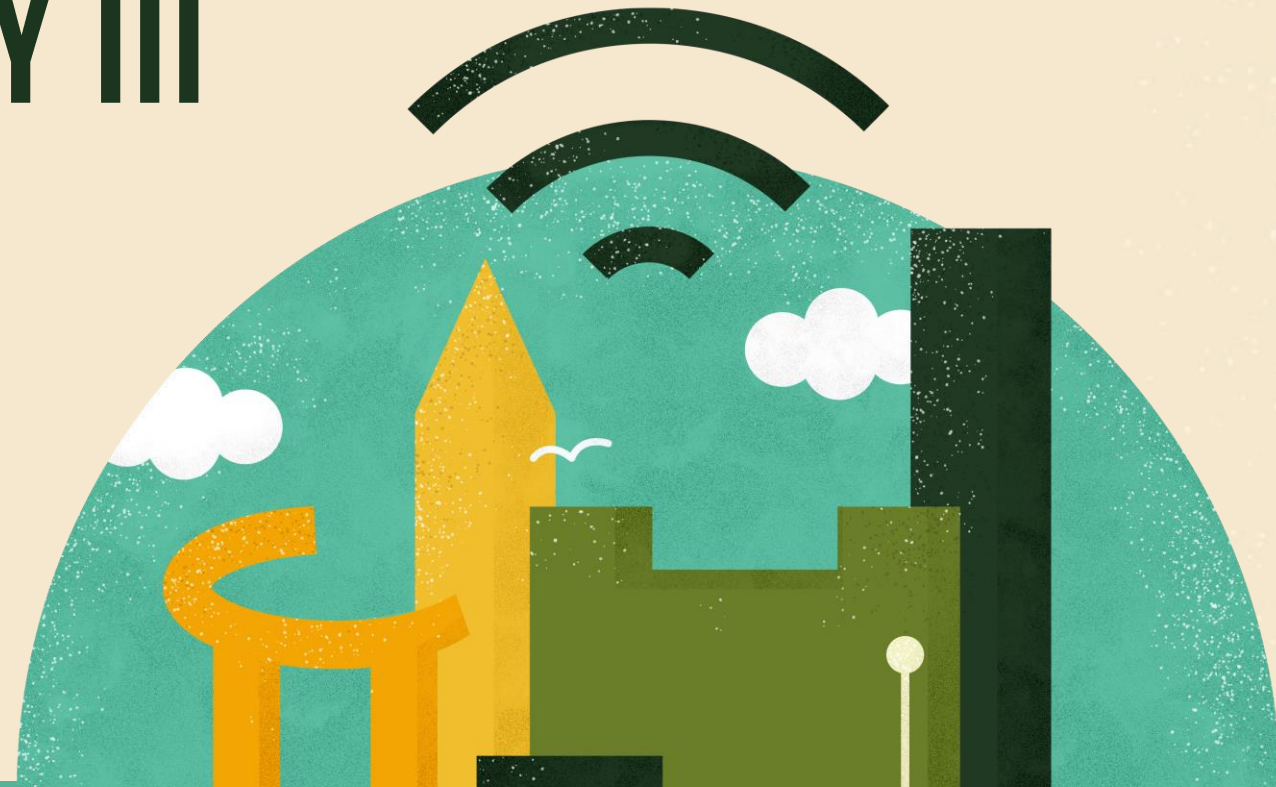
BLOQUE I: PREGUNTAS SOBRE RELACIONES ENTRE VARIABLES Y DESCRIPTIVO DE DATOS

¿Qué vamos a valorar?

1. La capacidad de demostrar si existe o no relación entre las variables y si existen o no diferencias en los análisis que os pedimos.
2. Valoraremos la elección adecuada de las técnicas y gráficos que expliquen relaciones entre variables o las diferencias en los datos.
3. Igualmente, se valorará la correcta interpretación de los resultados y las explicaciones razonadas de las conclusiones extraídas.



EL RETO: BLOQUE II Y III



BLOQUE II: PROPUESTA PARA LA MEJORA LA CALIDAD DE VIDA DE LOS CIUDADANOS EN GIJÓN.

Gracias a las conclusiones obtenidas en los ejercicios anteriores y en base a todos los análisis que hayáis realizado tendréis que diseñar un plan de mejora ambiental para la ciudad de Gijón.

Teniendo en cuenta las variables disponibles, plantead un enfoque integral con diferentes medidas que puedan mejorar la calidad de vida de los ciudadanos, promoviendo prácticas ambientalmente responsables y garantizando un entorno urbano saludable y equitativo para Gijón.

- Se valorarán propuestas innovadoras, prácticas y adaptables a las necesidades cambiantes de la ciudad a lo largo del tiempo.
- Con vuestras propuestas e iniciativas, deberéis **diseñar una campaña de sensibilización** que busque informar e involucrar activamente a la comunidad en la creación de un Gijón más sostenible combinando enfoques educativos, participación ciudadana y colaboraciones de instituciones u organismos locales.

¡IMAGINAD EL FUTURO SOSTENIBLE QUE OS GUSTARÍA PARA GIJÓN!



BLOQUE II: PROPUESTA PARA LA MEJORA LA CALIDAD DE VIDA DE LOS CIUDADANOS EN GIJÓN.

¿Qué vamos a valorar?

1. El ajuste de la propuesta a los análisis y conclusiones extraídas de la exploración de los datos.
2. La originalidad de la propuesta y su capacidad de ajuste a la realidad urbana de Gijón.
3. Innovación y practicidad de las medidas y de la campaña de sensibilización.



BLOQUE III: MODELO PREDICTIVO DE CALIDAD DEL AIRE.



Con el conocimiento de los datos acumulado en los ejercicios anteriores y pudiendo conocer:

1. Los datos de tráfico,
2. Los datos meteorológicos,
3. Las concentraciones de 'SO2_GijonGlobal', 'NO_GijonGlobal', 'NO2_GijonGlobal', 'CO_GijonGlobal', 'PM10_GijonGlobal' y 'O3_GijonGlobal'.

La variable dependiente del modelo (la que vamos a predecir) es una **variable dicotómica basada en puntuaciones ICA** que informa si la **calidad del aire en el día de referencia es buena (0) o mala (1)**. El modelo ha de predecir para un conjunto ciego de 31 días si la calidad del aire en Gijón va a ser buena o aceptable o mala.

¿Qué vamos a valorar?

1. El **menor error** en la predicción
2. La **originalidad y validez** de las técnicas utilizadas para construir un modelo predictivo.



¡Suerte!

