

DEFINE TU ESTRATEGIA

Cómo definir la estrategia de tu proyecto

Estos son los pasos para definir la estrategia:

1. Plantear el problema y sus objetivos
2. Nivel de la pirámide dónde nos enfocamos
3. Proyecto tipo que te encuentras
4. Definir la estructura de tu tabla de datos y los grupos de variables
5. Definir las variables con detalle
6. Define la estructura entradas - salida si aplica

En esta hoja de trabajo vas a definir la idea de tu proyecto pasando por 6 pasos que utilizo en las sesiones estratégica de mis alumnos y clientes.

Se trata de ver el problema de forma global y tratar de llevarlo a una estructura entradas – salida. Si no se puede tratar de identificar otro proyecto tipo que puedas aplicar.

¡A por ello!

1. Plantear el problema y sus objetivos

Contesta a las siguientes preguntas para definir el problema de tu proyecto:

¿Qué problema quieres resolver? Explícalo con palabras sencillas para que te todo el mundo lo pueda entender. Cuando más te alejes del lenguaje técnico mejor.

Quiero conocer la exposición a sustancias tóxicas del águila imperial ibérica a través de la dieta, utilizando un método no invasivo para obtener una muestra significativa, centrándome en el consumo de plástico, plomo y aluminio.

¿Por qué es importante resolver este problema?

La exposición a contaminantes puede provocar problemas de salud importantes, el águila imperial ibérica es una especie en peligro de extinción y protegida en nuestro país.

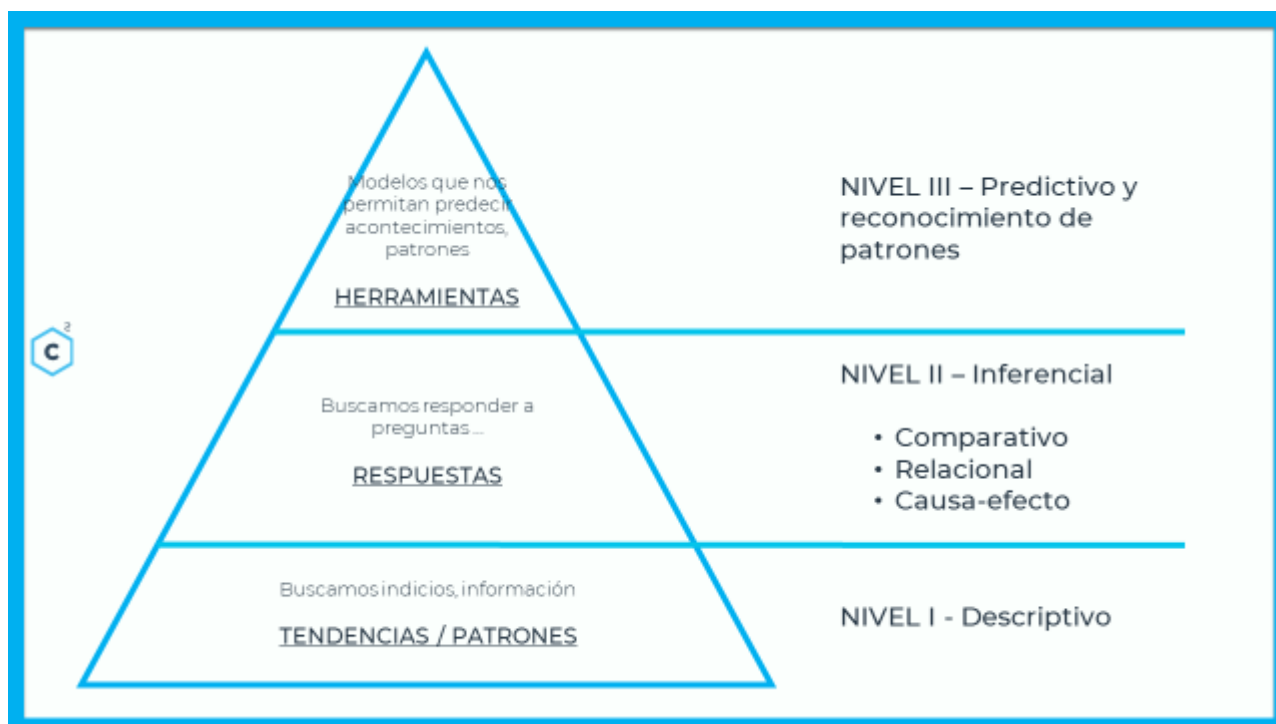
¿Qué se ha trabajado hasta hoy sobre ello? ¿Qué otros proyectos has podido ver que se parezcan al tuyo? Lista temas que has encontrado ya sea dentro de tu empresa o fuera en investigaciones o aplicaciones que hayas podido ver en la red.

- Se ha estudiado el efecto del DDT sobre la especie a través de análisis de la estructura de huevos.
- La influencia de la contaminación de organoclorados y pesticidas sobre la reproducción.
- Se ha estudiado la mortalidad por cebos envenenados.



2. Nivel de la pirámide dónde nos enfocamos

Ya sabes que hay 3 niveles de extracción de información de los datos y quiero que pienses en cada uno de estos tres niveles para enfocar tu proyecto y tus esfuerzos.



Usa la tabla para hacerlo:

ETAPA DE ANÁLISIS DE DATOS	¿QUÉ VAS A HACER EN CADA ETAPA? Escribe lo que se te ocurra sin demasiados filtros.
Descriptivo. El objetivo es lograr explicar los datos con gráficos y estadísticos descriptivos que nos permitan encontrar tendencias y empezar a sacar valor de los datos.	Inicialmente hare una tabla con los datos, en las columnas pondré las variables y en las filas los casos registrados. Obtendré los estadísticos de las diferentes variables descritas y realizare histogramas y boxplot para conocer la distribución de la muestra. Comprobar si existen tendencia a partir de gráficos que representen las principales variables.
Inferencial. Recuerda que aquí puedes comparar, puedes relacionar variables o incluso crear modelos entradas - salida que nos permiten obtener patrones clave de los datos.	En esta parte de la investigación se comprobará si hay o no correlaciones entre las diferentes variables estudiadas. Comprobar si hay correlaciones entre ciertas sustancias toxicas y el consumo de ciertas presas. Comparar la exposición a tóxicos entre territorios y según la dieta. Describir la causa efecto, si la explosión a los contaminantes tiene efecto sobre la reproducción a corto plazo. En esta parte se aplicarán test estadísticos para conocer la significancia de los resultados.



Reconocimiento de patrones. Puedes solucionar problemas de: • Clasificación • Regresión • Clustering - agrupado de datos • Reglas de asociación • Reducir dimensiones • Predicción de series temporales	También es el momento de hacer modelos de entrada y salida para prever que va a pasar en el futuro, o en otros territorios con dietas similares. Es un proyecto de clasificación.
---	--

Marca qué etapa es la más importante para tu proyecto;

- o Descriptivo – si solo quiero ver patrones a nivel gráfico
- o Inferencial – si pretendo contestar hipótesis de investigación y comprender qué factores influyen en una determinada respuesta
- o Reconocimiento de patrones – si pretendo crear una herramienta capaz de predecir, encontrar grupos similares, reducir dimensiones o reglas de asociación.

3. ¿En qué tipo de proyecto te estás enfrentando? Marca, máximo, dos tipos de proyectos.

- o Proyecto 1 – exploratorio/investigativo:
 - o Comparativo
 - o Relacional
 - o Causal
- o Proyecto 2 – clasificador
- o Proyecto 3 – regresión
- o Proyecto 4 – clustering
- o Proyecto 5 – reducción dimensional
- o Proyecto 6 – reglas de asociación
- o Proyecto 7 – series temporales



	<i>PERIODO: 0= No reproductor; 1= reproductor.</i> <i>EGAGÓPILAS: Número de egagrópilas por territorio.</i> <i>PRESAS: Número de presas por egagrópilas.</i> <i>IDT: Índice de diversidad trófica (basado en el índice de Shannon).</i> <i>CC: Porcentaje de recurrencia de conejo de campo.</i> <i>LI: Porcentaje de recurrencia de liebre.</i> <i>PR: Porcentaje de recurrencia de perdiz roja.</i> <i>PD: Porcentaje de recurrencia de paloma doméstica.</i> <i>PT: Porcentaje de recurrencia de paloma torcaz.</i> <i>URR: Porcentaje de recurrencia de urraca.</i> <i>CUL: Porcentaje de recurrencia de culebra sp.</i> <i>PAS: Porcentaje de recurrencia de passeriforme.</i> <i>ESC: Porcentaje de recurrencia de culebra escalera.</i> <i>CO: Porcentaje de recurrencia de conejo de granja.</i> <i>OV: Porcentaje de recurrencia de oveja.</i> <i>SIS: Porcentaje de recurrencia de sisón.</i> <i>ANA: Porcentaje de recurrencia de ánade real.</i> <i>PM: Porcentaje de recurrencia de paloma mensajera.</i> <i>CDZ: Porcentaje de recurrencia de codorniz.</i> <i>ZCH: Porcentaje de recurrencia de zorzal charlo.</i> <i>ACCP: Porcentaje de recurrencia de accipitriforme.</i> <i>COL: Porcentaje de recurrencia de coleóptero.</i> <i>VC: Porcentaje de recurrencia de vaca.</i> <i>GR: Porcentaje de recurrencia de garza real.</i> <i>OTRAS: Porcentaje de recurrencia de otras presas.</i>
¿Qué son las filas u observaciones?	Son los resultados de los análisis de dieta de los territorios estudiados, tamaño de la muestra, ubicación y habitantes por municipio.



¿Qué grupos de variables o factores vas a incluir? Estos factores son los que nos darán información de tu respuesta.	- Variables relacionadas con el territorio. - Variables relacionadas con el tamaño de la muestra. - Variables relacionadas con la dieta. - Variables relacionadas con la exposición a contaminantes. Trata de hacer una agrupación por conceptos, por factores. Dentro de cada grupo factor habrá diferentes variables que den esa característica. - Relativas a la información del territorio. - Relativas al tamaño de la muestra. - Relativas a la dieta consumida. - Relativas a la exposición a los contaminantes.
Lista las variables de entrada que crees que pueden tener un impacto importante	-El número de habitantes por municipio -Todas las variables de entrada relacionadas con la dieta.

Recomendable: dibujar la tabla de datos y pensar en una estructura entradas – salida.

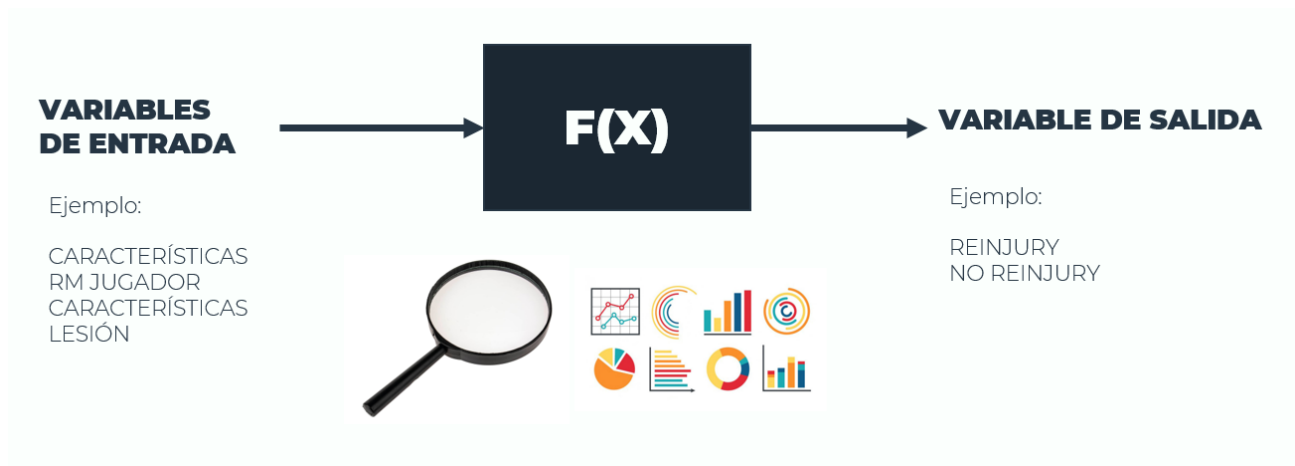
5. Define tus variables con detalle

	Nombre Variable	Código Cómo la escribes en R o Python o SPSS	Grupo. Relativa a... o si es la respuesta	Tipo Variable	Tipo variable en R /Python	Unidades Cualitativa indica los campos Cualitativa indica las unidades	Explicación
1	Territorio	TRR	Relativa al territorio	nominal	texto	Cualitativa	Pareja de águila imperial ibérica muestreada
2	Provincia	PRV	Relativa al territorio	numérica	numérica	código provincia	Código provincia donde está su nido
3	Municipio	MUN	Relativa al territorio	numérica	numérica	código localidad	Código localidad donde está el nido
4	Año	AÑO	Relativa al territorio	numérica	numérica	Años	Año de recogida de la muestra
5	Habitantes	HAB	Relativa al territorio	numérica	numérica	nº de habitantes	Número de Habitantes
6	Periodo	PRD	Relativa al territorio	Nominal binaria	factor	0: No reproductor; 1: Reproductor	Si la muestra ha sido obtenida en el periodo de reproducción o fuera del periodo de reproducción
7	Egagrópilas	EGA	Relativa al tamaño de la muestra	numérica	numérica	cuantitativa discreta	Número de egagrópilas recogidas en cada territorio
8	Presas	PRE	Relativa al tamaño de la muestra	numérica	numérica	cuantitativa discreta	Número de presas por egagrópilas
9	Índice de diversidad trófica	IDT	Relativa a la dieta	numérica	numérica	cuantitativa continua	Índice de diversidad trófica, basado en el Índice de Shannon
10	Conejo de campo	CC	Relativa a la dieta	numérica	numérica	cuantitativa continua	Porcentaje de recurrencia de conejo de campo
11	Liebre	LI	Relativa a la dieta	numérica	numérica	cuantitativa continua	Porcentaje de recurrencia de liebre
12	Perdiz roja	PR	Relativa a la dieta	numérica	numérica	cuantitativa continua	Porcentaje de recurrencia de perdiz roja
13	Paloma doméstica	PD	Relativa a la dieta	numérica	numérica	cuantitativa continua	Porcentaje de recurrencia de paloma doméstica
14	Paloma torcaz	PT	Relativa a la dieta	numérica	numérica	cuantitativa continua	Porcentaje de recurrencia de paloma torcaz
15	Urraca	URR	Relativa a la dieta	numérica	numérica	cuantitativa continua	Porcentaje de recurrencia de urraca
16	Culebra sp	CUL	Relativa a la dieta	numérica	numérica	cuantitativa continua	Porcentaje de recurrencia de culebra sp
17	Paseriforme	PAS	Relativa a la dieta	numérica	numérica	cuantitativa continua	Porcentaje de recurrencia de paseriforme
18	Culebra escalera	ESC	Relativa a la dieta	numérica	numérica	cuantitativa continua	Porcentaje de recurrencia de culebra escalera
19	Conejo de granja	CO	Relativa a la dieta	numérica	numérica	cuantitativa continua	Porcentaje de recurrencia de conejo de granja

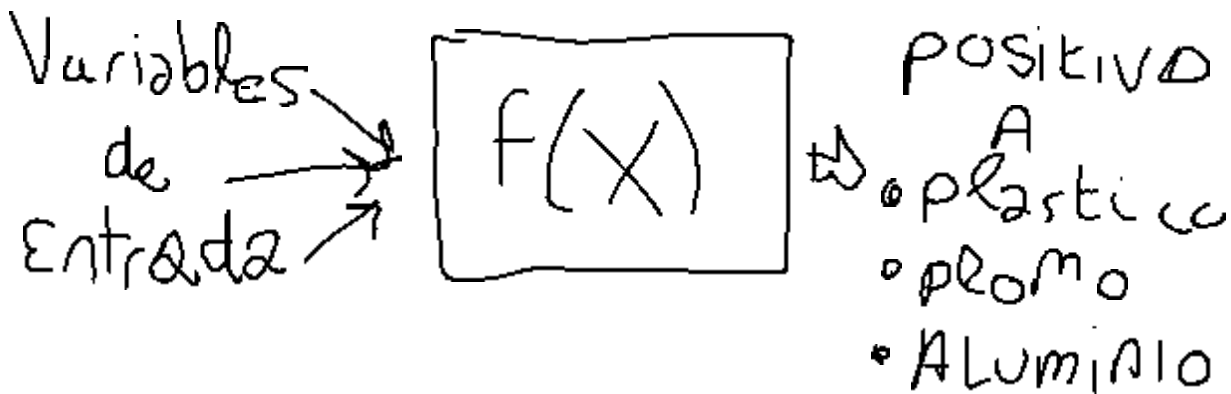


20	Oveja	OV	Relativa a la dieta	numérica	numérica	cuantitativa continua	Porcentaje de recurrencia de oveja
21	Sisón	SIS	Relativa a la dieta	numérica	numérica	cuantitativa continua	Porcentaje de recurrencia de sisón
22	Ánade real	ANA	Relativa a la dieta	numérica	numérica	cuantitativa continua	Porcentaje de recurrencia de ánade real
23	Paloma mensajera	PM	Relativa a la dieta	numérica	numérica	cuantitativa continua	Porcentaje de recurrencia de paloma mensajera
24	Codorniz	CDZ	Relativa a la dieta	numérica	numérica	cuantitativa continua	Porcentaje de recurrencia de codorniz
25	Zorzal charlo	ZCH	Relativa a la dieta	numérica	numérica	cuantitativa continua	Porcentaje de recurrencia de zorzal charlo
26	Accipitriformes	ACCP	Relativa a la dieta	numérica	numérica	cuantitativa continua	Porcentaje de recurrencia de accipitriforme
27	Coleópteros	COL	Relativa a la dieta	numérica	numérica	cuantitativa continua	Porcentaje de recurrencia de coleóptero
28	Vaca	VC	Relativa a la dieta	numérica	numérica	cuantitativa continua	Porcentaje de recurrencia de vaca
29	Garza real	GR	Relativa a la dieta	numérica	numérica	cuantitativa continua	Porcentaje de recurrencia de garza real
30	Otras presas	OTR	Relativa a la dieta	numérica	numérica	cuantitativa continua	Porcentaje de recurrencia de otras presas
31	Exposición a contaminantes	EAC	Variable respuesta	Nominal binaria	factor	0: no expuesto; 1: Expuesto	Exposición a contaminantes
32	Exposición a macro plásticos	EMP	Variable respuesta	numérica	numérica	cuantitativa continua	Exposición a macro plásticos % de egagrópila positivas

6. Dibuja el esquema entradas – salida



Dibuja tu esquema:



Puedes aplicar este guion a tus datos o a datos abiertos de kaggle por ejemplo:

- <https://www.kaggle.com/datasets/fedesoriano/stroke-prediction-dataset>
- <https://www.kaggle.com/datasets/fedesoriano/heart-failure-prediction>
- <https://www.kaggle.com/datasets/arashnic/hr-analytics-job-change-of-data-scientists>
- <https://www.kaggle.com/datasets/benroshan/factors-affecting-campus-placement>
- <https://www.kaggle.com/datasets/ayessa/salary-prediction-classification>

Usa este guion siempre que te enfrentes a un nuevo proyecto y te ayudará a enfocarte en cómo puedes sacar valor de tus datos con un proceso de 6 pasos.

¡Deseando ver tu estrategia ya definida!



Si necesitas apoyo con tus proyectos o formación en Analítica de Datos escribe whatsapp y vemos tu caso



1 fuerte abrazo
Jordi Ollé