

# Índice Calidad del Aire y consumos energéticos






Alba | Katerina | Guilherme | Virginia

3/11/2025

Proyecto Final - Bootcamp Data Analytics - NEOLAND



# Índice Calidad del Aire y consumos energéticos

	Objetivo del Proyecto	1
	Contexto	2
	Alcance	3
	Metodología	4
	Conclusiones	5



## Objetivo del Proyecto

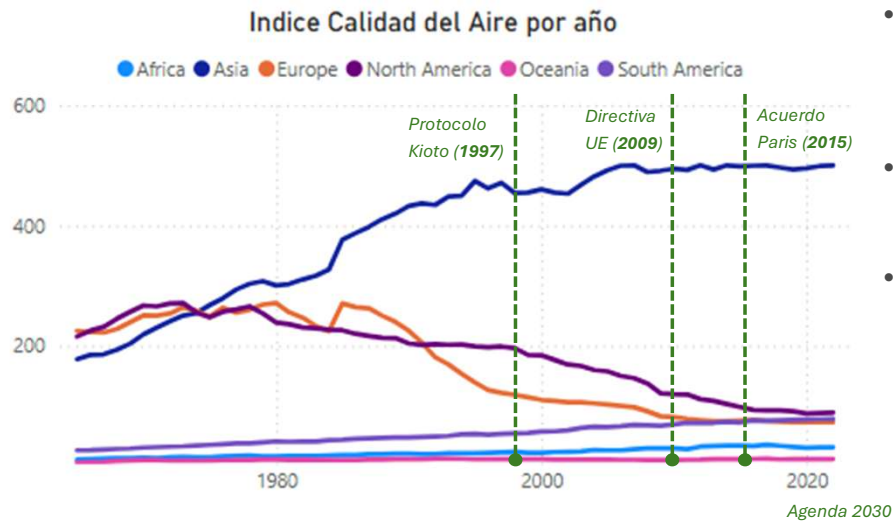
Predecir la calidad del aire en función de la energía renovable.

Identificar los países que tienen mejor y peor calidad del aire

Identificar tendencias del consumo energético

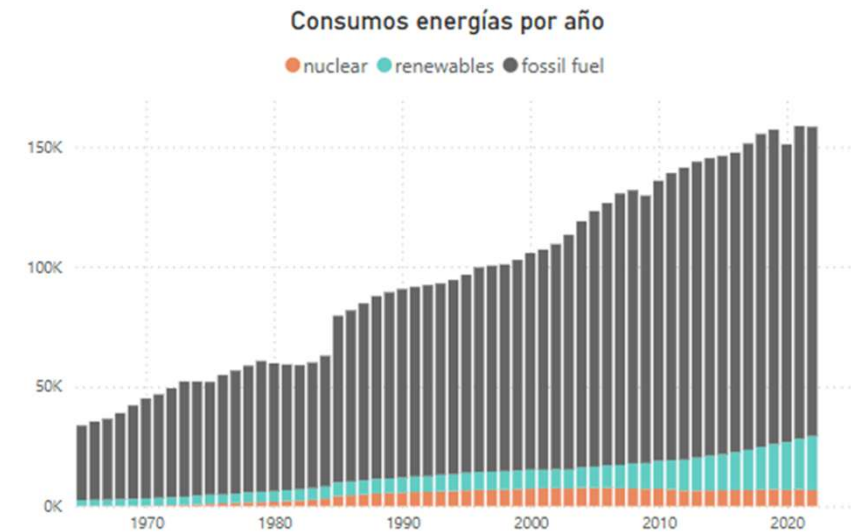


## Contexto



- Las iniciativas globales implementadas para mitigar la calidad del aire han contribuido para aumentar consumos en energías renovables, bajando los índices de contaminación.
- Los niveles de calidad del aire, aunque bajando en Estados Unidos y Europa, siguen por encima de los niveles aceptables.
- Debido a la alta población en China e India y su grande dependencia de energías fósiles, sus índices de contaminación siguen aumentando.

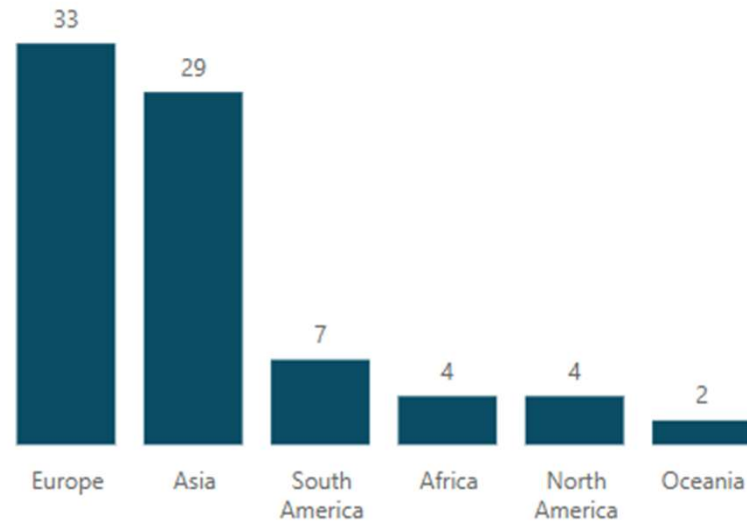
- Desde 1965 hasta 2022, el consumo de energías fósiles, nucleares y renovables ha evolucionado significativamente
- Aumento del consumo de la energía nuclear, pues tiene la ventaja de tener grande eficiencia energética y bajas emisiones





## Alcance

Datos de los consumos energéticos (renovables, nuclear y fósiles) y los diferentes contaminantes desde 1965 hasta 2022 de **79 países**.  
El *dataset* se ha centrado mas en países de Europa, Asia y América.



Además, hemos incluido los datos demográficos y socioeconómicos (población y del PIB) durante el mismo periodo para entender la posible relación con el Índice de Calidad del Aire.



**Herramientas:** *Kaggle* (Fuente de Datasets) | *Python* (Limpieza y Modelo ML) | *PBI* (visualización)

- Definir el periodo de análisis entre los años 1965 y 2022.
- Eliminar las filas donde no constan datos de consumo de energía.
- Eliminar las filas donde no existen datos sobre contaminantes.
- Eliminar las filas donde no hay datos de población.
- Completar los casos en que no exista el PIB, en base al promedio del continente
- Resultado: 35% de los datos iniciales



Indicador que mide la calidad del aire basándose en los diferentes contaminantes atmosféricos: (unidad medida: toneladas)

- Óxidos de Nitrógeno (NOX)
- Dióxido de Azufre (SO<sub>2</sub>)
- Compuestos Orgánicos Volátiles No Metánicos (NMVOC)
- Monóxido de Carbono (CO)
- Carbón Negro (BC)
- Amoniacó (NH<sub>3</sub>)

Se expresa en una escala numérica oficial permitiendo alertar a la población sobre posibles impactos en la salud.

## AIR QUALITY INDEX

	Good	0-50
	Moderate	51-100
	Unhealthy for sensitive groups	101-150
	Unhealthy	151-200
	Very Unhealthy	201-300
	Hazardous	301-500



Del dataset energía obtenemos información relativa a los consumos energéticos, así como la población y el PIB de país.

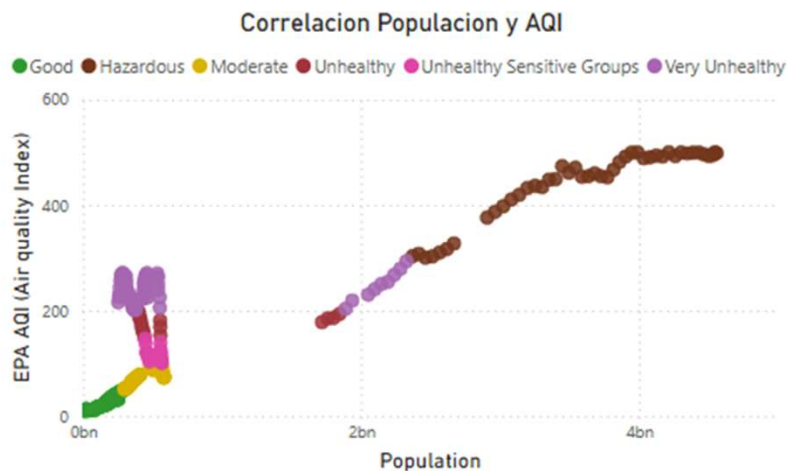
- Energías fósiles (carbón, petróleo, gas)
- Energías renovables (eólica, hídrica, solar)
- Energía Nuclear

Se expresan en terawatts-horas (Kilowatts-horas si per capita).

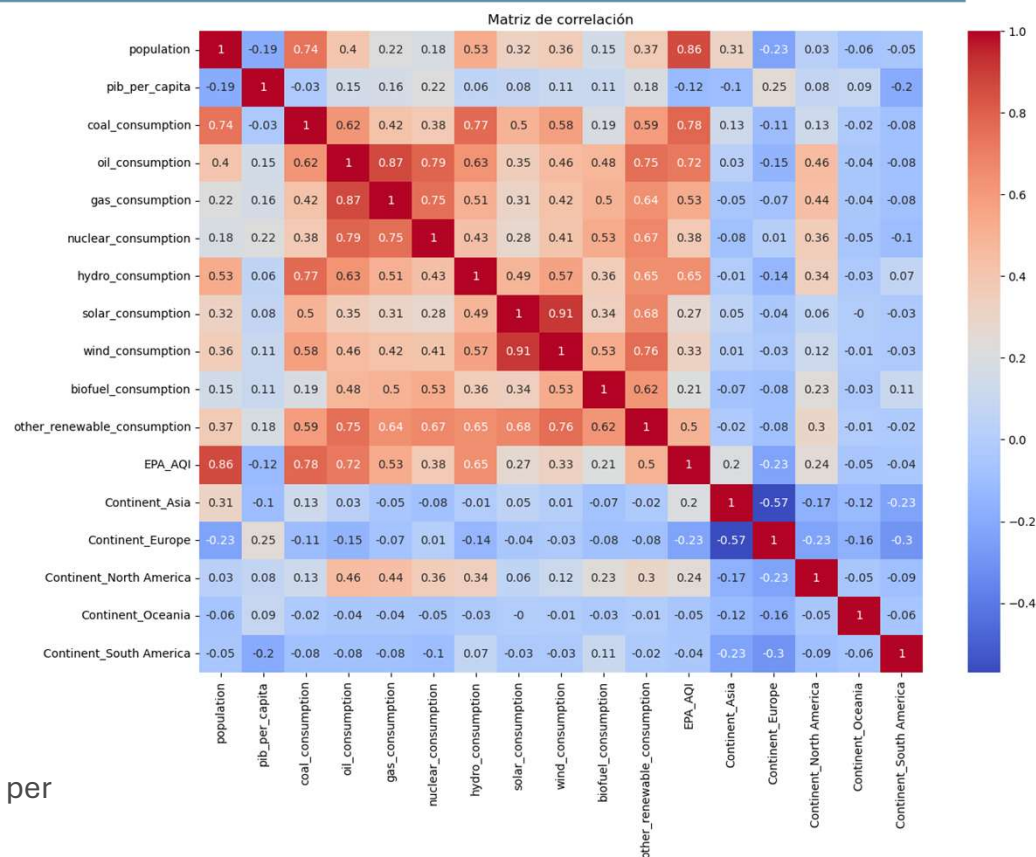




- Alto nivel de correlación entre las variables población y el AQI.
- A mayor población, los consumos de energía crecen y el Índice de Calidad del Aire (AQI) presenta una tendencia al alza.



**SOLUCION:** Ajustar los consumos de energía y emisiones a un nivel per cápita.



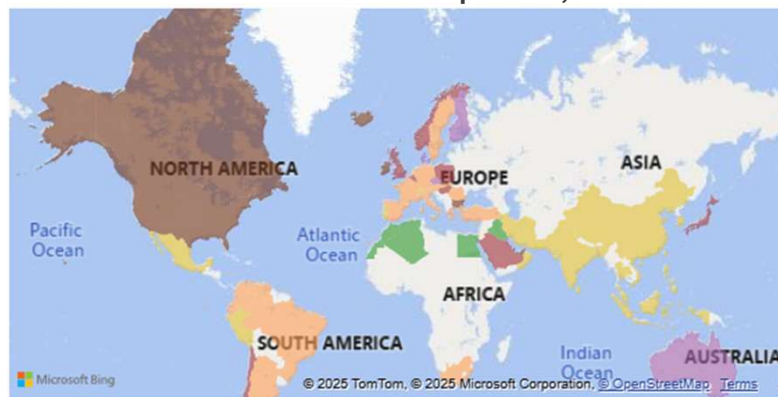


## Metodología | Análisis Exploratorio

Índice Calidad del aire por País, 1965



Índice Calidad del aire por País, 1980



Índice Calidad del aire por País, 2000



Índice Calidad del aire por País, 2022



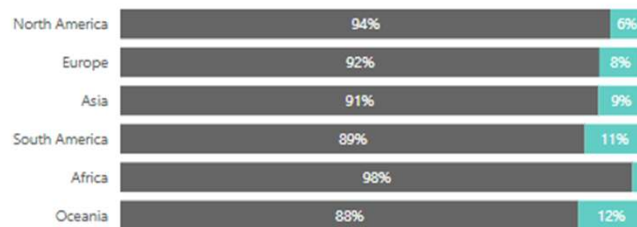
- Al largo de los años los países mas poblados (China, India e EEUU) y igualmente con AQI alto, salen mas beneficiados en términos per capita.
- En cambio, países con menos población (Qatar, Luxemburgo, Nueva Zelandia) resultan mas penalizados



### AÑO 1965

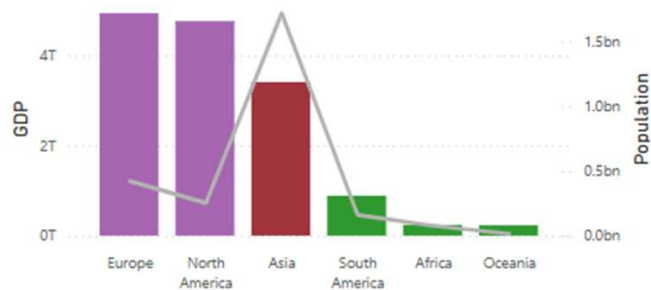
Fuentes Energia por Continente

● fossil ● renewables ● nuclear



Distribucion del AQI y relacion con GDP y Populacion

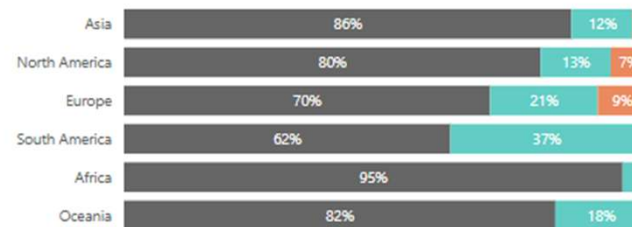
● Good ● Unhealthy ● Very Unhealthy ● Population



### AÑO 2022

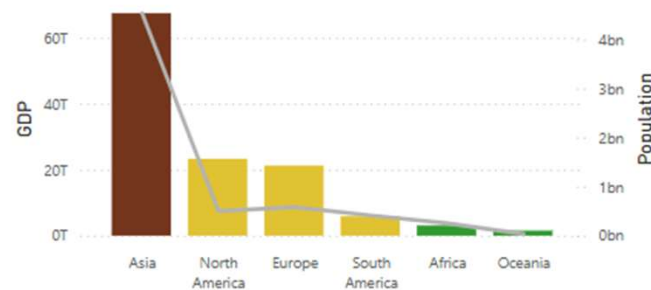
Fuentes Energia por Continente

● fossil ● renewables ● nuclear



Distribucion del AQI y relacion con GDP y Populacion

● Good ● Hazardous ● Moderate ● Population



- Se verifica un incremento del consumo de las energías renovables. Europa y America Sur.
- Aumento del AQI en Asia, específicamente en China y India, debido a la gran población (35% del total)
- En cambio, en Europa y EEUU mejoran la calidad del aire



**OBJETIVO:** Predecir el INDICE DE CONTAMINACION (AQI) en función del CONSUMO ENERGETICO (fósiles, nucleares y renovables) y del PIB de un país a nivel per capita.

**MODELO:** **Random Forest**

**CARACTERISTICAS:**

- Alta precisión;
- Capacidad para manejar tanto problemas de clasificación como de regresión;
- Resistencia al *overfitting* gracias al uso de múltiples árboles de decisión.

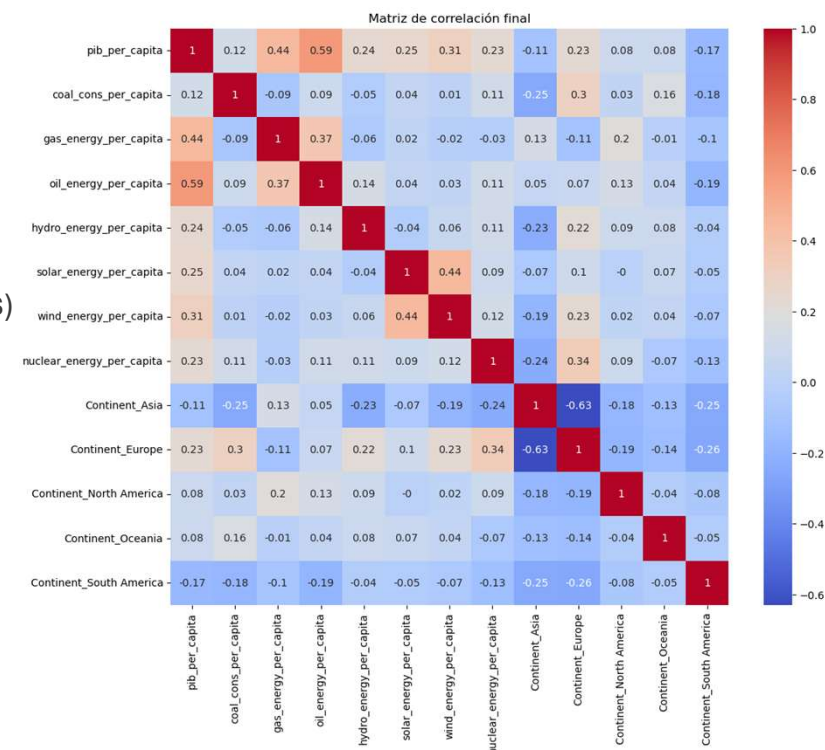


- Selección de columnas para modelo
- Convertir continentes en *dummies*
- Análisis de multicolinealidad: (correlación de variables independientes)

# VIF

- Separación de datos en train y test

	pib_per_capita	coal_cons_per_capita	gas_energy_per_capita	oil_energy_per_capita	hydro_energy_per_capita	solar_energy_per_capita	wind_energy_per_capita	nuclear_energy_per_capita	AQI_Category	Continent_Asia	Continent_Europe	Continent_North America	Continent_Oceania	Continent_South America
0	2880.0	65.753	600.141	1244.239	95.531	0.0	0.0	0.0	Good	False	False	False	False	False
1	2690.0	62.698	611.988	1607.236	83.224	0.0	0.0	0.0	Good	False	False	False	False	False
2	2876.0	46.891	580.614	1468.702	94.003	0.0	0.0	0.0	Good	False	False	False	False	False





- Generación del árbol de clasificación

Profundidad = 16      Accuracy = 79,78%

- Generación del Random Forest

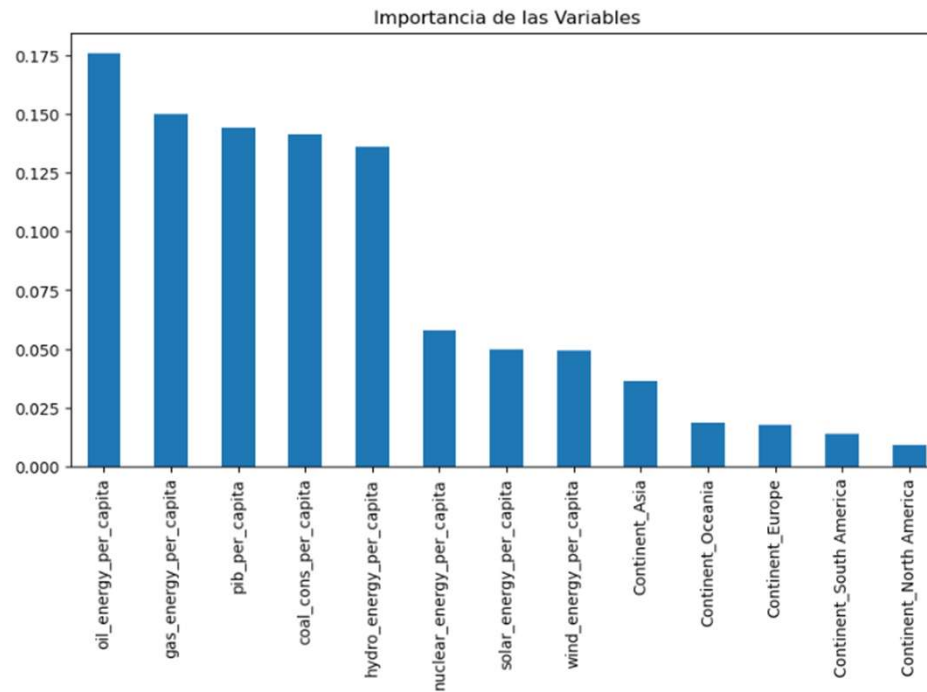
Profundidad = 16      Accuracy = 88,09%

- (Alternativa) Valoración de regresión lineal múltiple

Variables = 14 (todas)    $R^2 = 0,4922$    Error medio = 71,8732



- Método *Feature Importance* (variables que más influyen en el resultado del modelo)







## Metodología | Modelo Machine Learning

### TEST Y CONCLUSIONES (2/2)

PIB	Consumo energético	Coal 0.2	Gas 0.4	Oil 0.4	TOTAL Fossil	Hydro 0.35	Solar 0.25	Wind 0.25	Nuclear 0.15	TOTAL Clean	% Fossil	% Clean	Indice
10,000	3,000	600	1,200	1,200	3,000	-	-	-	-	-	100%	0%	Good
10,000	3,000	300	600	600	1,500	525	375	375	225	1,500	50%	50%	Moderate
10,000	3,000	-	-	-	-	1,050	750	750	450	3,000	0%	100%	Moderate
10,000	50,000	10,000	20,000	20,000	50,000	-	-	-	-	-	100%	0%	Unhealthy
10,000	50,000	5,000	10,000	10,000	25,000	8,750	6,250	6,250	3,750	25,000	50%	50%	Moderate
10,000	50,000	-	-	-	-	17,500	12,500	12,500	7,500	50,000	0%	100%	Moderate
10,000	100,000	20,000	40,000	40,000	100,000	-	-	-	-	-	100%	0%	Very Unhealthy
10,000	100,000	10,000	20,000	20,000	50,000	17,500	12,500	12,500	7,500	50,000	50%	50%	Moderate
10,000	100,000	-	-	-	-	35,000	25,000	25,000	15,000	100,000	0%	100%	Moderate
10,000	200,000	40,000	80,000	80,000	200,000	-	-	-	-	-	100%	0%	Unhealthy
10,000	200,000	20,000	40,000	40,000	100,000	35,000	25,000	25,000	15,000	100,000	50%	50%	Unhealthy
10,000	200,000	-	-	-	-	70,000	50,000	50,000	30,000	200,000	0%	100%	Moderate
50,000	3,000	600	1,200	1,200	3,000	-	-	-	-	-	100%	0%	Moderate
50,000	3,000	300	600	600	1,500	525	375	375	225	1,500	50%	50%	Moderate
50,000	3,000	-	-	-	-	1,050	750	750	450	3,000	0%	100%	Moderate
50,000	50,000	10,000	20,000	20,000	50,000	-	-	-	-	-	100%	0%	Very Unhealthy
50,000	50,000	5,000	10,000	10,000	25,000	8,750	6,250	6,250	3,750	25,000	50%	50%	Moderate
50,000	50,000	-	-	-	-	17,500	12,500	12,500	7,500	50,000	0%	100%	Moderate
50,000	100,000	20,000	40,000	40,000	100,000	-	-	-	-	-	100%	0%	Very Unhealthy
50,000	100,000	10,000	20,000	20,000	50,000	17,500	12,500	12,500	7,500	50,000	50%	50%	Moderate
50,000	100,000	-	-	-	-	35,000	25,000	25,000	15,000	100,000	0%	100%	Moderate
50,000	200,000	40,000	80,000	80,000	200,000	-	-	-	-	-	100%	0%	Unhealthy
50,000	200,000	20,000	40,000	40,000	100,000	35,000	25,000	25,000	15,000	100,000	50%	50%	Unhealthy
50,000	200,000	-	-	-	-	70,000	50,000	50,000	30,000	200,000	0%	100%	Moderate
100,000	3,000	600	1,200	1,200	3,000	-	-	-	-	-	100%	0%	Moderate
100,000	3,000	300	600	600	1,500	525	375	375	225	1,500	50%	50%	Moderate
100,000	3,000	-	-	-	-	1,050	750	750	450	3,000	0%	100%	Moderate
100,000	50,000	10,000	20,000	20,000	50,000	-	-	-	-	-	100%	0%	Very Unhealthy
100,000	50,000	5,000	10,000	10,000	25,000	8,750	6,250	6,250	3,750	25,000	50%	50%	Moderate
100,000	50,000	-	-	-	-	17,500	12,500	12,500	7,500	50,000	0%	100%	Moderate
100,000	100,000	20,000	40,000	40,000	100,000	-	-	-	-	-	100%	0%	Very Unhealthy
100,000	100,000	10,000	20,000	20,000	50,000	17,500	12,500	12,500	7,500	50,000	50%	50%	Moderate
100,000	100,000	-	-	-	-	35,000	25,000	25,000	15,000	100,000	0%	100%	Moderate
100,000	200,000	40,000	80,000	80,000	200,000	-	-	-	-	-	100%	0%	Very Unhealthy
100,000	200,000	20,000	40,000	40,000	100,000	35,000	25,000	25,000	15,000	100,000	50%	50%	Unhealthy
100,000	200,000	-	-	-	-	70,000	50,000	50,000	30,000	200,000	0%	100%	Moderate

- **Índice de contaminación** -> Consumo energético
  - **Consumos bajos** -> índice de contaminación bajo.
  - **Consumos medios/altos** -> índice de contaminación condicionado por el mix empleado
    - Índice de contaminación aceptable si energía renovable > 50%.
  - **Consumos extremos** -> necesidad del uso de energía renovable
    - Índice de contaminación aceptable si energía renovable ~ 100%.
- **PIB alto** -> capacidad para mitigar los efectos de un consumo extremo





- Fuerte correlación entre la población y el consumo energético, el análisis de datos a nivel per cápita ofrece resultados más fiables.
- El índice de calidad del aire de un país está fuertemente condicionado por su consumo energético global, no siendo determinante el mix de energías utilizado.
- El mix de energías utilizado tiene cierta influencia cuando se dan consumos energéticos altos.
- El PIB per cápita de un país influye en su capacidad para mitigar estos efectos.

**En conclusión**, si buscamos mejorar la calidad del aire y reducir la contaminación en las ciudades a nivel global, es fundamental implementar políticas ambientales efectivas y realizar **inversiones en tecnología sostenible conjuntas**.

*Un esfuerzo conjunto será clave para avanzar hacia un entorno menos contaminado y más saludable.*

# Gracias por vuestra atención

Alba | Katerina | Guilherme | Virginia

3/11/2025

Proyecto Final - Bootcamp Data Analytics - NEOLAND

