



DEPARTAMENTO  
DE COMPUTACION

Facultad de Ciencias Exactas y Naturales - UBA

# Laboratorio de datos

## Regresión 2

### Verano 2025

¿Cuánto medirá de adulte?



*Basado en una clase de Mariela Sued*

# ¿Cuánto medirá de adulte?

- + Sin información → ¿Qué podemos decir?

# ¿Cuánto medirá de adulte?

- + Sin información → ¿Qué podemos decir?

[Completemos la Altura](#)



¿Cuánto medirá de adulte?

¿Promediamos?

ESTIMAMOS: 170.85

# Información



Es varón

# ¿Cuánto medirá de adulte?

- + Sin información ✓
- + Es varón →

Completemos  
columna "sexo"



¿Cuánto medirá de adulte?

¿Promediamos entre varones?

ESTIMAMOS: 173.7



# Información





Es varón



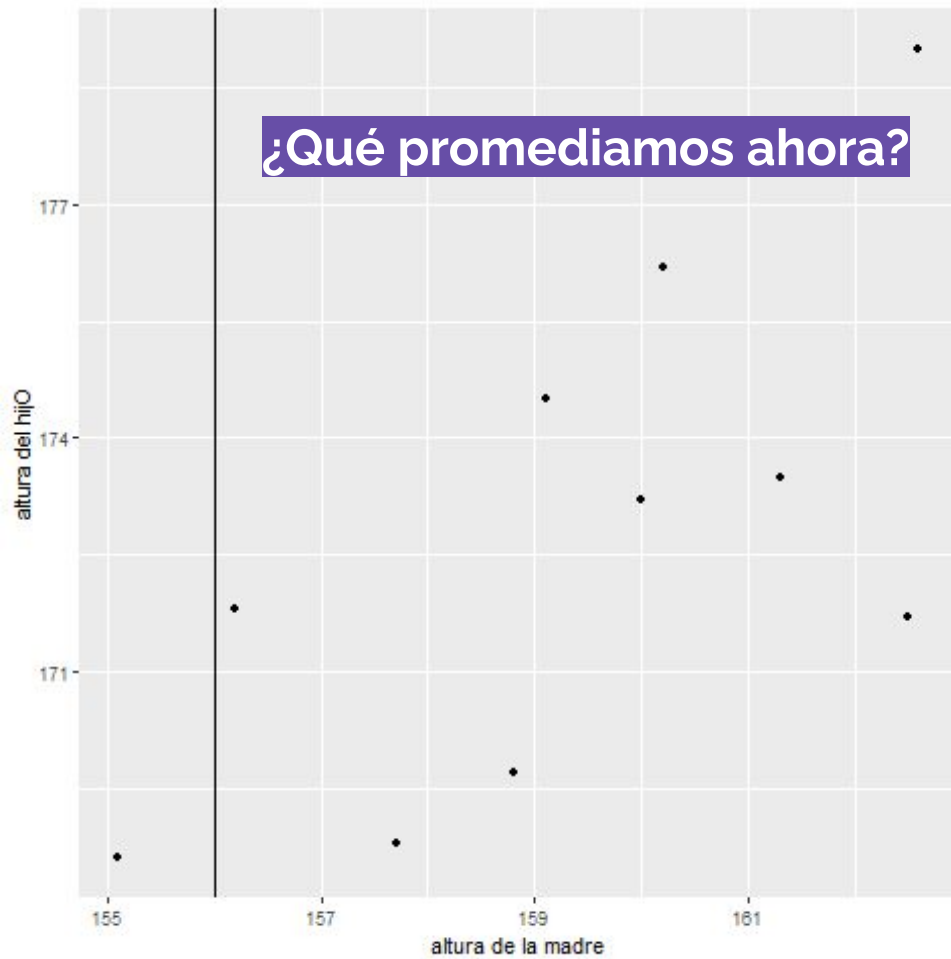
La mamá mide 156

# ¿Cuánto medirá de adulto?

- + Sin información → 
- + Es varón → 
- + La mamá mide 156 →

Completemos columna  
**"altura mamá"**





# Una posibilidad: KNN

Idea: Promediamos los valores de casos parecidos

kNN: k nearest neighbors - k vecinos más cercanos

Ej. Consideramos los 5 valores más **cercanos**\* al valor nuevo (altura madre).  
Promediamos las alturas de esos 5 varones

\*Cercanos: en la o las variables explicativas,  
y con la distancia que consideremos.

Estimamos: 168.9

# K Nearest Neighbors (KNN) - para regresión

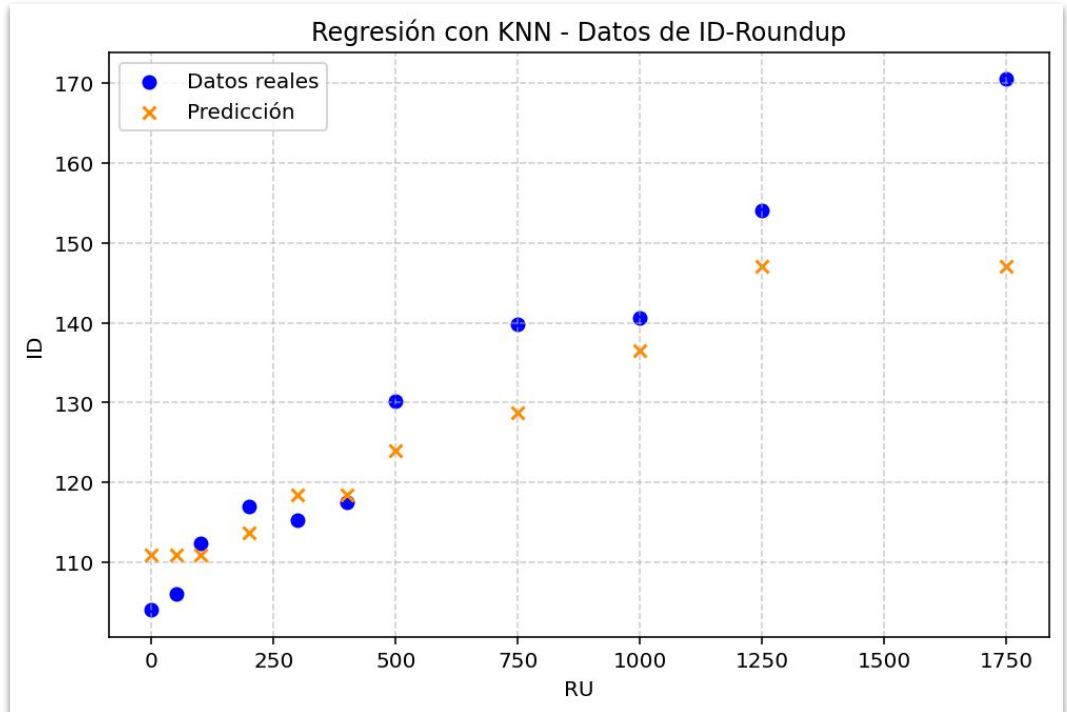
Para determinar el valor de  $y$  para una nueva observación:

1. Buscar los puntos más cercanos, dentro del conjunto de entrenamiento
2. Ver qué valores de  $y$  tienen
3. Promediar

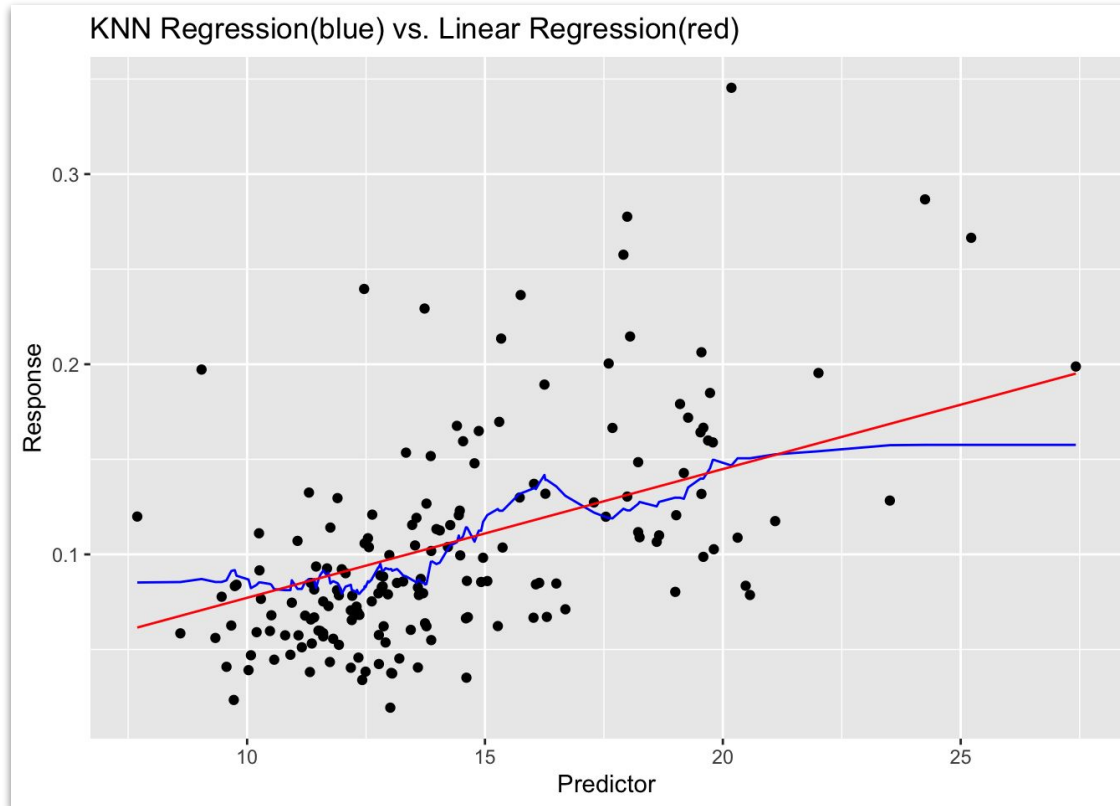
¿Cuántos puntos consideramos? Depende del valor de  $k$ .

# KNN con sklearn

```
X = data_roundup[["RU"]]  
Y = data_roundup["ID"]  
  
modelo_knn = KNeighborsRegressor()  
modelo_knn.fit(X,Y)  
  
Y_pred = modelo_knn.predict(X)  
mse = mean_squared_error(Y, Y_pred)
```



# KNN vs Regresión Lineal



# Ejercicio

Ajustar un modelo de regresión con knn para los **datos de altura** relevados con la planilla. Probar con  $k = 5$ .

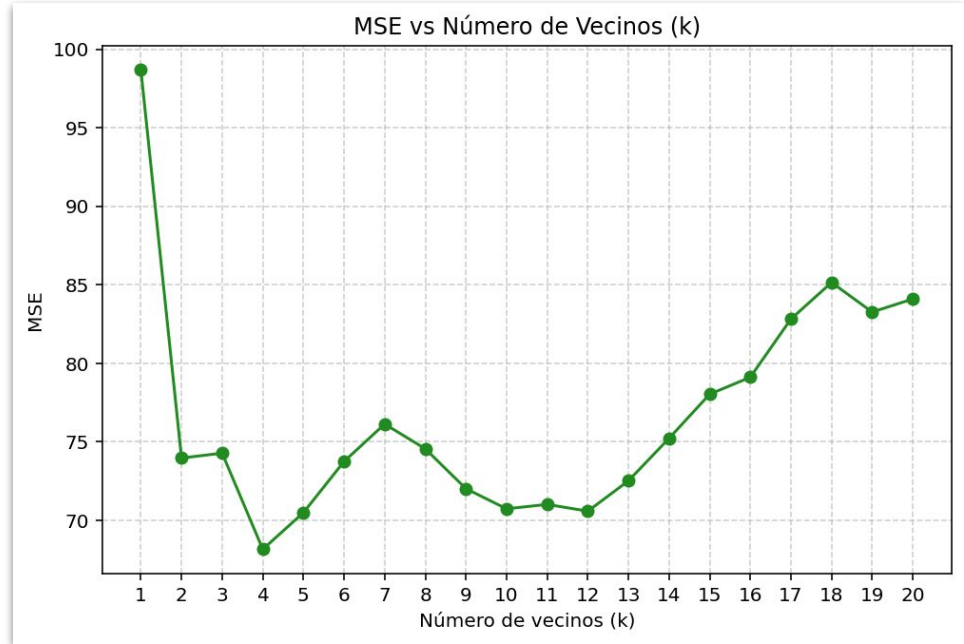
¿Cuán bien funciona?

Reportar el MSE.



# Ejercicio

1. Repetir con distintos valores de  $k$ .
2. Graficar el MSE para cada valor elegido.



# Datos de autos - mpg

La clase pasada trabajamos con un conjunto de datos sobre autos.

Variable a explicar: mpg.

Variables explicativas:

cylinders	int64
displacement	float64
horsepower	int64
weight	int64
acceleration	float64
model year	int64
origin	int64
car name	object



Vieron cómo se puede modelar con regresión lineal, de distinto tipo según la/s variable en cuestión.

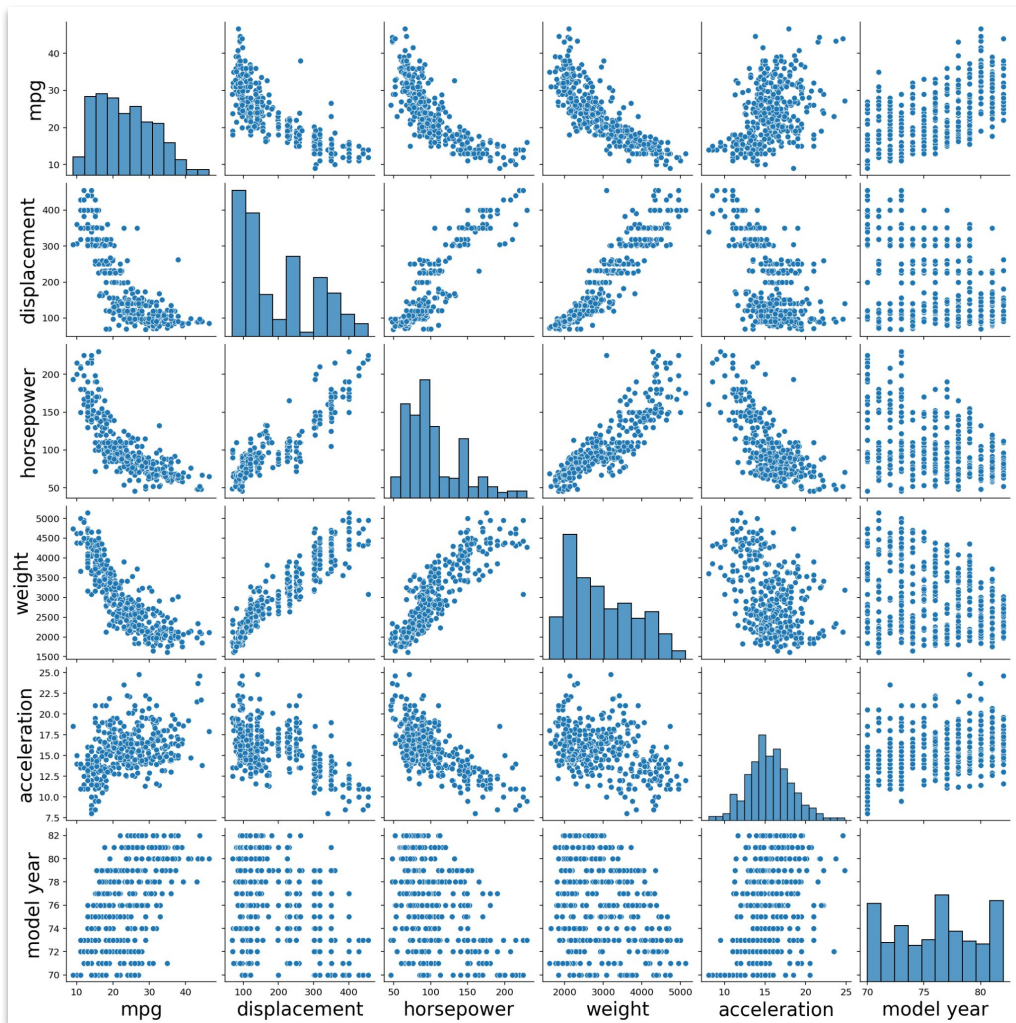
Veamos cómo modelarlo con KNN.

## Pairplot:

- Scatterplot entre cada par de variables
- Histograma de cada variable

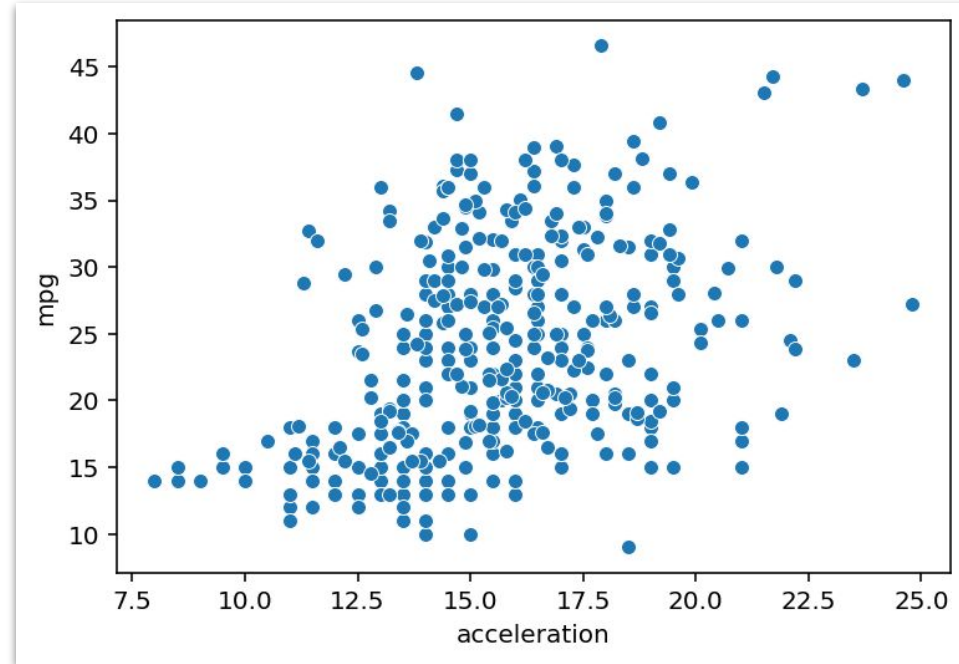
Sirve para ver de manera panorámica las interacciones entre variables.

¿Entre cuáles se ve una tendencia clara?



# Ejemplo

Mpg respecto de acceleration.



# Ejercicio

Utilizar knn para ajustar un modelo que prediga mpg en función de acceleration.

Probar con distintos valores de  $k$ , y ver el error en cada caso (MSE).

# Ejercicio

Repetir, pero

- + considerando varias variables (elegir cuáles). Antes de hacerlo, ¿sería conveniente reescalar los datos?
- + considerar distintos valores de  $k$  y comparar

# Ejercicio

- Considerar las variables: 'displacement', 'horsepower', 'weight', 'acceleration', 'model year'
- Separar el conjunto en train (80%) y test.
- Ajustar un modelo knn con  $k$  entre 1 y 20, y graficar MSE en función de  $k$ , diferenciando train de test.

