

# Ejemplo: Predicción de la población de Colombia para 2020.

**Autor:** Marlon Algarra.

En este ejemplo veremos cómo se puede usar Data Science para responder a la pregunta ¿Cuál será la población en Colombia para 2020? Utilizaremos los pasos vistos en la presentación.

## Obtención de los datos.

Así, añadiremos los datos del número de habitantes en Colombia de los años pares entre 1998 y 2018.

Datos tomados de:

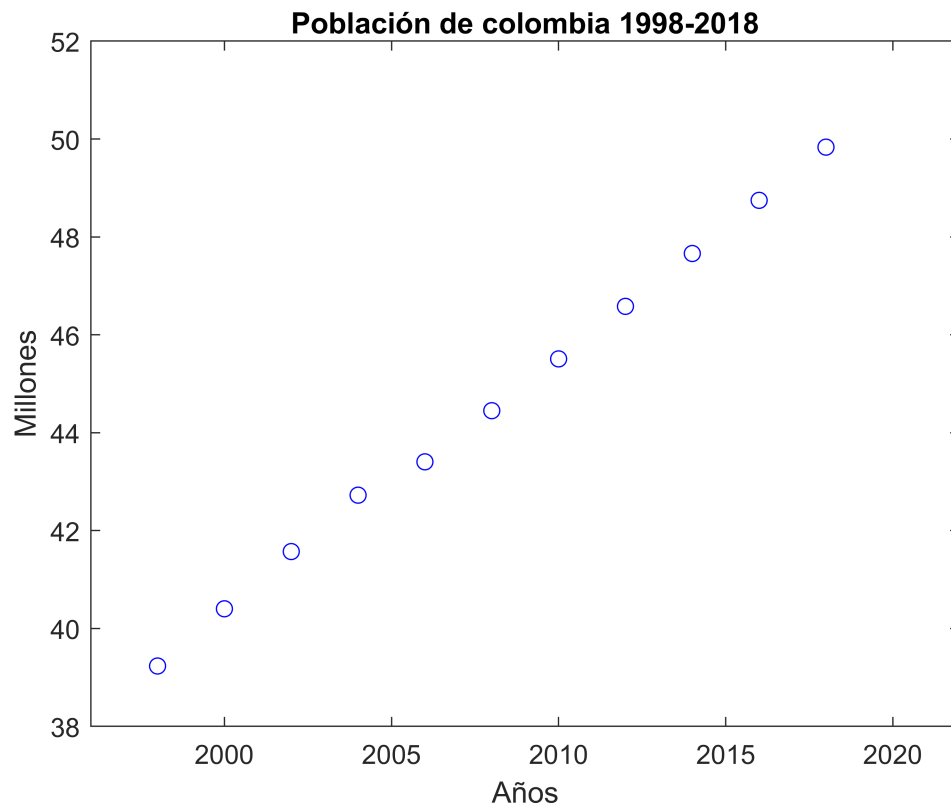
- <https://www.colombia.com/colombia-info/estadisticas/poblacion/>
- <https://www.populationpyramid.net/colombia>

```
years=(1998:2:2018);  
poblacion=[39.234 40.403 41.572 42.724 43.405 44.450 45.508 46.581 47.661 48.747 49.834];
```

## Visualización de los datos

Veamos cómo lucen nuestros datos en una gráfica:

```
plot(years,poblacion,'bo');  
title('Población de colombia 1998-2018');  
ylabel('Millones');  
xlabel('Años')  
xlim([1996 2022])  
ylim([38 52])
```



¿Con estos datos podremos predecir el número de habitantes en Colombia para 2020?

### Ajustando los datos.

Ajustar los datos en este caso no parece difícil, pues parece que crecen polinómicamente. Usaremos la función `polyfit` para obtener coeficientes de polinomios.

Las ecuaciones de ajuste son:

$y = ax + b$	lineal
$y = ax^2 + bx + c$	cuadrática
$y = ax^3 + bx^2 + cx + d.$	cúbica

```
x = (years-1998)/(2);
coef1 = polyfit(x,poblacion,1)
```

```
coef1 =
    1.0405    39.3535
```

```
coef2 = polyfit(x,poblacion,2)
```

```
coef2 =
    0.0013    1.0273   39.3735
```

```
coef3 = polyfit(x,poblacion,3)
```

```
coef3 =
    0.0031   -0.0448    1.2031   39.2628
```

## Graficando los ecuaciones de ajuste.

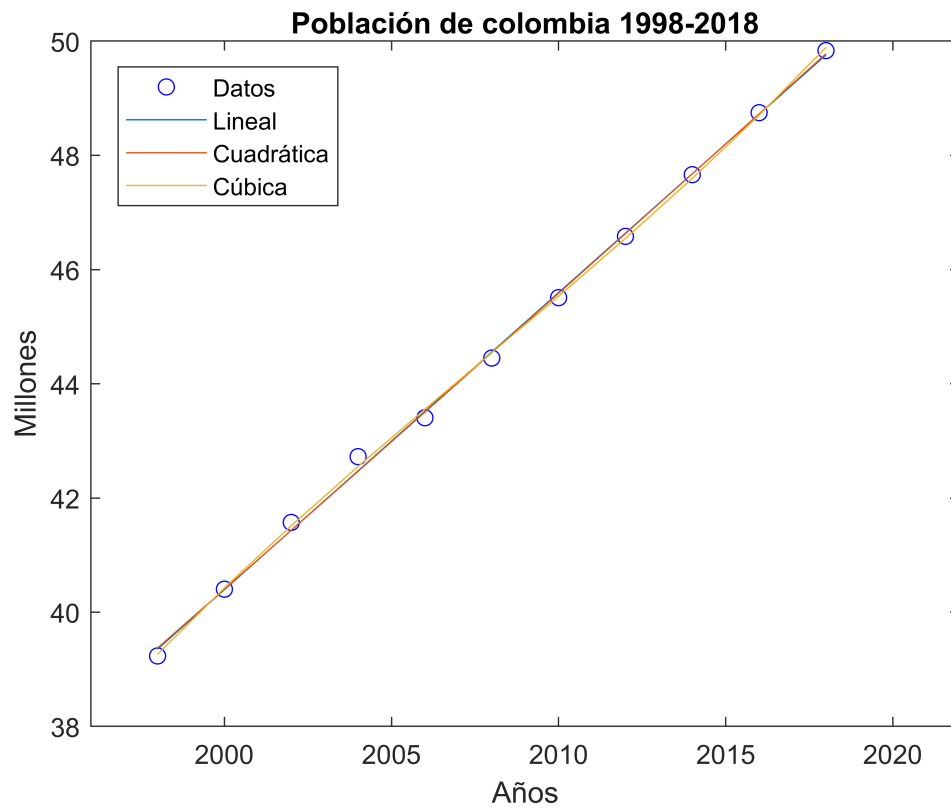
Para ver qué tan bien se ajustan las ecuaciones a nuestro conjunto de datos inicial, usaremos la función `polyval` para evaluar las ecuaciones de ajuste en los puntos de `x`.

```
pred1 = polyval(coef1,x);
pred2 = polyval(coef2,x);
pred3 = polyval(coef3,x);
[pred1; pred2; pred3]
```

```
ans =
    39.3535    40.3941    41.4346    42.4752    43.5157    44.5563    45.5968    46.6374 ...
    39.3735    40.4021    41.4333    42.4672    43.5038    44.5430    45.5849    46.6294
    39.2628    40.4242    41.5145    42.5521    43.5554    44.5430    45.5332    46.5445
```

Ahora veamos las gráficas de las ecuaciones de ajuste para observar qué tal se acomodan a nuestros datos.

```
hold on
plot(years,pred1)
plot(years,pred2)
plot(years,pred3)
ylim([38 50])
legend({'Datos' 'Lineal' 'Cuadrática' 'Cúbica'},'Location', 'NorthWest')
hold off
```



## Prediciendo la población

Así, podremos predecir la población para 2020 utilizando las tres ecuaciones que obtuvimos antes.

```
x2020 = (2020-1998)/2;
pred1 = polyval(coef1,x2020);
pred2 = polyval(coef2,x2020);
pred3 = polyval(coef3,x2020);
[pred1 pred2 pred3]
```

```
ans =
    50.7995    50.8341    51.1698
```

Podemos observar que la predicción está en un rango entre 50.7 millones a 51.1 millones de personas en Colombia para 2020.