

# Práctica 2 - Regresión

## Visitas a StatForecasting

---

### APRENDIZAJE AUTOMÁTICO



#### **Profesor**

José Anastacio  
Hernández Saldaña

#### **Alumna**

Judith Carolina  
Lucio Garza  
1877415

# Base de Datos

La base de datos trabajada “Daily Website Visits” se encuentra disponible en el siguiente link: [StatForecasting.Visits](https://statforecasting.com/visits).

StatForecasting es un sitio web que ha publicado diversos tópicos de Estadística. Durante 5 años se estuvieron registrando las veces que se consultaba la página, así como visitas únicas y recurrentes a la misma. Fue posible identificar a los usuarios a través de Cookies.

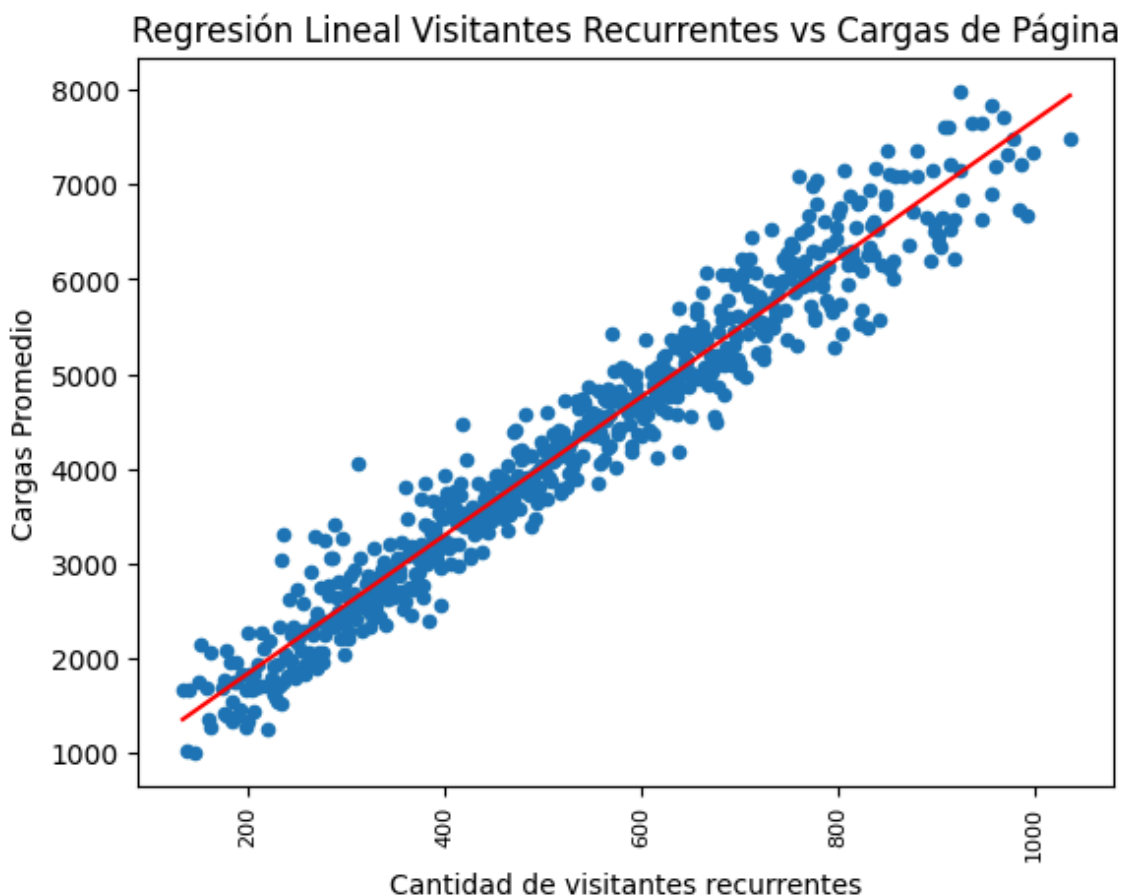
- Los registros abarcan del 14 de septiembre del 2014 al 19 de agosto del 2020.
- En total 6,378,882 usuarios visitaron el sitio web.
- Se tuvo un total de 8,921,516 cargas de página.
- El promedio de consultas diarias fue de 4,116.99.



# Regresión Lineal

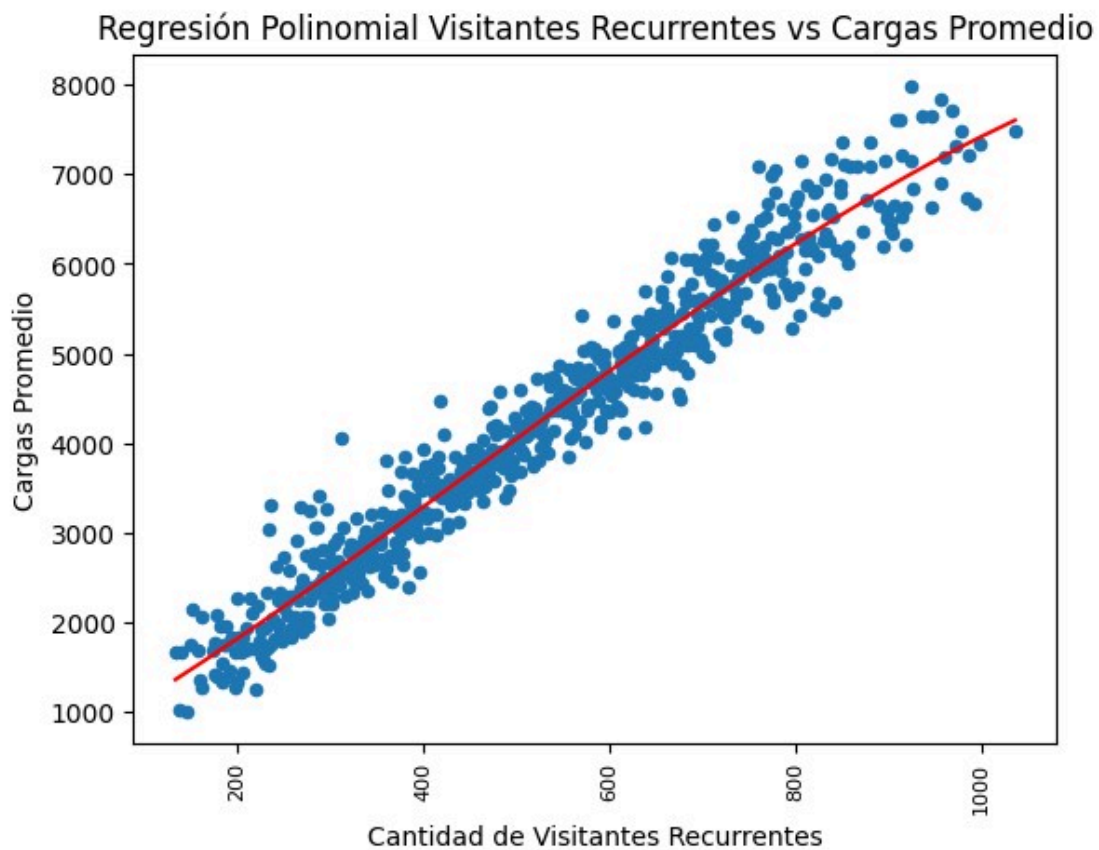
Se promediaron la cantidad de cargas en la página por total de visitantes que regresaron al sitio web (es decir, no era su primera visita). Así se establecieron las variables para regresión lineal, en la cual podemos observar una tendencia positiva.

Se destaca que la mayoría de los puntos se encuentran muy concentrados sobre la línea. Este modelo tuvo un ajuste de R cuadrada del 94.83%.



# Regresión Polinomial

A continuación se realizó una regresión polinomial de tercer grado. El modelo se muestra en la siguiente gráfica:

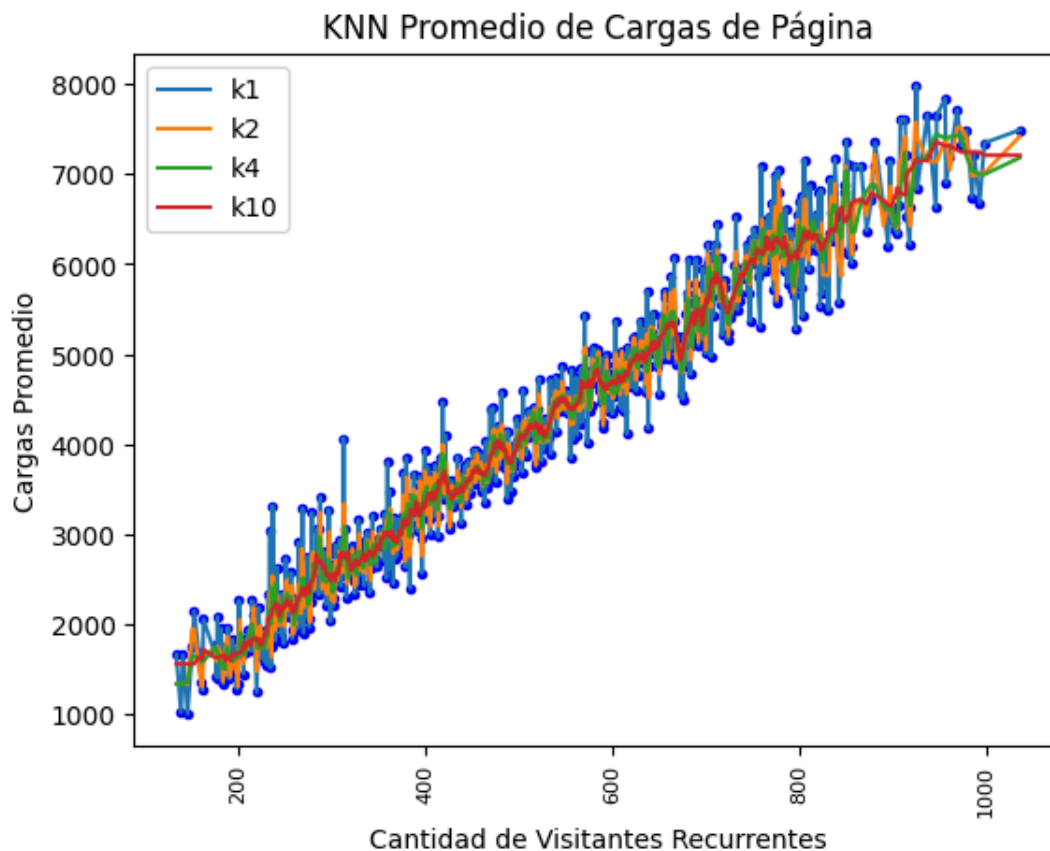


Evaluando su desempeño se tiene una R cuadrada ajustada del 94.90%. Ligeramente superior al ajuste de la regresión lineal.

# K Nearest Neighbors

Finalmente, se realizó un modelo KNN, considerando distintos valores para k y comparando su desempeño, tal como se muestra a continuación:

Neighbors	R Ajustada
1	100%
2	97.29%
4	96.09%
10	95.37%



Con los resultados obtenidos, se elige el modelo KNN con  $k=2$ , ya que es el que tiene un mayor ajuste (97.29%) sin estar completamente apegado a los datos.