



**Google Maps Scraper**

## Resumen ejecutivo

**Google Maps Scraper** es una automatización creada en **n8n** que permite **buscar y recopilar datos de empresas desde Google Maps** de forma automática, con el fin de construir una **base de datos actualizada de posibles contactos comerciales** (por ejemplo, clínicas, centros de salud, negocios locales, etc.).

En lugar de realizar búsquedas manuales, abrir webs una por una y copiar los datos a una hoja de cálculo, esta automatización **hace todo ese trabajo de forma autónoma**, rápida y estructurada.

Su función principal es **identificar empresas relevantes, extraer su información pública (nombre, dirección, teléfono, web, correo, etc.) y guardarla en una hoja de Google Sheets**, lista para ser usada en campañas de comunicación, prospección o análisis.

---

## Objetivo de la automatización

El objetivo es agilizar y automatizar la recopilación de datos de negocios desde Google Maps, transformando una tarea manual y lenta en un proceso automatizado que:

- Ahorra tiempo y recursos, eliminando la necesidad de búsquedas manuales.
  - Crea una base de datos limpia y actualizable, con información útil para marketing o ventas.
  - Evita duplicados y garantiza que los datos estén siempre actualizados.
  - Extrae información adicional (como correos electrónicos) directamente desde las webs de los negocios.
- 

## Funcionamiento paso a paso

La automatización se ejecuta manualmente o puede ser llamada desde otro flujo. Su funcionamiento se divide en tres grandes fases:

---

## Fase 1 – Búsqueda en Google Maps

El sistema comienza con una lista de términos de búsqueda (por ejemplo, “*clínica Madrid*”, “*clínica Majadahonda*”, etc.). Para cada término, envía consultas a **Google Maps** a través de la API de **Serper.dev**, que devuelve los resultados de negocios reales. Además, consulta las **páginas 1 a 4 de resultados**, para asegurarse de captar el máximo número de empresas relevantes.

---

## Fase 2 – Extracción y filtrado de información

De cada negocio encontrado, la automatización recoge los siguientes datos:

- Nombre del negocio
- Teléfono
- Página web
- Dirección
- Valoración media y número de reseñas
- Tipo o categoría del negocio
- Coordenadas geográficas

Una vez obtenida esta información, **visita cada sitio web** para localizar un **correo electrónico de contacto**, aplicando un sistema de filtrado que descarta correos genéricos, falsos o irrelevantes (como los de Gmail, Yahoo o páginas de prueba).

El resultado final es una ficha completa y limpia de cada empresa.

---

## Fase 3 – Guardado y actualización en Google Sheets

Finalmente, todos los datos procesados se **guardan automáticamente en una hoja de Google Sheets** llamada *BDAutomatizada*.

Si el negocio ya existía (se identifica por su página web), la automatización **actualiza su información existente**. Si es nuevo, **lo añade como un nuevo registro**.

De esta manera, la hoja de cálculo se convierte en una **base de datos viva**, en constante crecimiento y actualización, que puede utilizarse en campañas, análisis o procesos comerciales.

---

## Detalle técnico del flujo en n8n

El flujo de la automatización “Google Maps Scraper” está construido en **n8n** y se compone de varios nodos conectados entre sí, que permiten ejecutar de manera ordenada la búsqueda, el procesamiento y el almacenamiento de los datos obtenidos desde Google Maps.

A continuación, se explica cada parte del flujo, el papel que cumple y cómo interactúa con el resto de la automatización.

El proceso comienza con los **nodos disparadores**. El primero, llamado “*When clicking ‘Execute workflow’*”, permite iniciar la automatización manualmente desde el entorno de n8n. Es útil para realizar pruebas o ejecuciones puntuales, ya que contiene una lista de términos de búsqueda (por ejemplo, “clínica Madrid”, “clínica Majadahonda”, etc.) que serán utilizados por el sistema.

El segundo disparador, “*When Executed by Another Workflow*”, está diseñado para que este flujo pueda ser **ejecutado como subprocesso desde otros workflows**. Esto lo hace modular y permite integrarlo dentro de automatizaciones mayores, como una que programe ejecuciones periódicas o agrupe varias búsquedas diferentes.

Una vez activado el flujo, se inicia la **fase de procesamiento de las búsquedas**.

El nodo “*Split Out*” se encarga de dividir la lista de términos de búsqueda (queries) en elementos individuales, de modo que cada término se procese de manera independiente.

Después, el nodo “*Loop Over Items*” crea un bucle que recorre cada búsqueda una por una, garantizando un control preciso del flujo y evitando que se lancen todas simultáneamente.

Para prevenir bloqueos o límites de la API, el nodo “*Wait*” introduce una breve pausa, aproximadamente cuatro segundos, entre cada ejecución, lo que mejora la estabilidad y la eficiencia del sistema.

A continuación, comienza la **fase de obtención de datos desde Google Maps**. El nodo “*Search Google Maps page 1*” realiza la primera llamada a la API de **Serper.dev**, que actúa como intermediario para obtener los resultados de Google Maps.

El nodo siguiente, “*Search Google Maps pages 2–4*”, amplía la búsqueda solicitando también las páginas 2, 3 y 4 de resultados, con el fin de obtener un mayor número de negocios por cada término buscado.

Ambos nodos se autentican mediante un **header API key** proporcionado por Serper, que garantiza el acceso seguro y controlado al servicio externo.

Una vez descargados los resultados, el nodo “*Format maps results*” procesa toda la información recibida. Este nodo está programado con código JavaScript que combina los datos de las distintas páginas y **extrae los campos más relevantes** de cada negocio: nombre, teléfono, sitio web, dirección, tipo, valoración, número de reseñas y coordenadas geográficas.

Además, el script filtra los resultados para descartar aquellos negocios que no tengan una página web, ya que el siguiente paso requiere esa información para poder buscar correos de contacto.

Con los negocios ya formateados, comienza la **fase de análisis web y detección de correos electrónicos**. El nodo “*Scrape website*” accede al sitio web de cada empresa y obtiene el contenido HTML de la página. Este nodo está configurado para permitir certificados no verificados, evitando errores en sitios que no usan HTTPS correctamente.

A continuación, el nodo “*Code in JavaScript*” convierte ese HTML en texto plano, limitando la longitud máxima del contenido a 20.000 caracteres para no sobrecargar la memoria ni ralentizar el flujo.

Luego, el nodo “*Extract emails*” analiza el texto con una expresión regular que identifica direcciones de correo electrónico. Este script también aplica varios **filtros de limpieza**, eliminando correos falsos, direcciones genéricas (como Gmail o Yahoo) o cualquier email asociado a imágenes o contenido no válido. El resultado es un único correo electrónico válido y depurado por cada negocio encontrado.

Finalmente, la automatización pasa a la **fase de guardado de datos**. El nodo “*Save emails to Google Sheet*” se conecta mediante OAuth2 a la hoja de cálculo de Google llamada *BDAutomatizada*. Su operación está configurada en modo “append or update”, lo que significa que cada registro nuevo se añadirá a la hoja, mientras que los negocios ya existentes (identificados por su columna “Website”) se actualizarán con la información más reciente. Los campos que se almacenan incluyen nombre, correo, teléfono, dirección, web, valoración, número de reseñas, tipo de negocio y latitud.

Gracias a esta estructura, la automatización mantiene **una base de datos limpia, sin duplicados y permanentemente actualizada**, garantizando la coherencia entre las ejecuciones y la fiabilidad de la información recopilada.

---

## Conexiones y dependencias

- **API utilizada:** Serper.dev (para búsqueda en Google Maps).
  - **Autenticaciones:**
    - Serper API → `httpHeaderAuth`
    - Google Sheets → `OAuth2`
  - **Documento destino:** BDAutomatizada
  - **Columna clave para coincidencia:** `Website`
  - **Columnas registradas:**

Nombre, Email, Phone, Website, Address, Rating, Rating Count, Type, Latitude
- 

## Beneficios clave

- Automatiza la prospección de negocios.
  - Centraliza la información en una base de datos online.
  - Reduce errores humanos y evita duplicidades.
  - Permite escalar fácilmente añadiendo nuevos términos o zonas.
  - Base de datos siempre actualizada, ideal para acciones comerciales o mailing.
- 

## Posibilidades de integración futura

- Añadir validación automática de webs caídas.
- Incluir control de fecha de última actualización en la hoja.
- Extender la extracción a más páginas de resultados.
- Guardar varios correos por negocio si existen.
- Implementar alertas en caso de fallo de conexión o límite de API.