

## PIA – ETAPA 1

### Equipo:

- Jaime Sebastián González Ontiveros
- Mauro Jesús López Cerda
- Raziel Azarias Torres Garza

### Planteamiento del problema:

- El objetivo de este proyecto es analizar las empresas de “camiones” (transporte de carga, lavado, mantenimiento, etc.), utilizando datos del Directorio Estadístico Nacional de Unidades Económicas (DENUE) del INEGI. La finalidad es identificar zonas con mayor o menor concentración de estas empresas, entender su dispersión geográfica y visualizar patrones que podrían ser relevantes para estudios económicos, logísticos o de infraestructura.
- **¿Qué fenómeno o situación quieren analizar?**  
La distribución geográfica y características de las de “camiones” en distintas regiones de México. El enfoque está en identificar patrones de localización y estructura de estas empresas.
- **¿Por qué es relevante?**  
Este análisis es relevante porque el transporte de mercancías por camión representa una parte fundamental de la economía mexicana. Conocer dónde están ubicadas estas empresas puede ayudar a entender la logística, identificar posibles zonas de saturación o baja cobertura, y apoyar decisiones de inversión o desarrollo de infraestructura.
- **¿Qué datos necesitan obtener?**  
Se requiere información de la API del INEGI sobre empresas relacionadas con “camiones”, incluyendo nombre, actividad económica, ubicación geográfica (latitud y longitud), razón social y dirección. Esta información se obtiene consultando la API con diferentes coordenadas geográficas para cubrir diversas zonas del país.

### Descripción de la API elegida

- La API utilizada es la **API del DENUE (INEGI)**, la cual permite consultar unidades económicas por nombre, actividad o ubicación geográfica.
- Se usa el endpoint:  
<https://www.inegi.org.mx/app/api/denue/v1/consulta/Buscar/{palabraClave}/{latitud},{longitud}/{radio}/{token}>
- **Parámetros principales:**
  - Palabra clave: término de búsqueda (en este caso, "camiones").
  - Latitud, longitud: coordenadas del punto de búsqueda.
  - Radio: distancia en metros (hasta 2500).

- Token: clave personal para autenticación.
- Se eligieron varias coordenadas de distintas ciudades para ampliar la cobertura del análisis y obtener más diversidad de empresas.

#### Descripción de la estructura de datos utilizada

- Los datos extraídos se almacenan en una estructura compuesta por:
  - Una **lista de diccionarios** (transportes), donde cada diccionario representa una empresa con los campos:
    - 'nombre'
    - 'actividad'
    - 'latitud'
    - 'longitud'
    - 'razon\_social'
    - 'direccion'
- Esta estructura permite almacenar múltiples empresas de forma organizada y es compatible con la exportación a formatos como JSON y Excel, así como con bibliotecas de análisis como pandas.

#### Justificación del tratamiento de datos aplicado

- Se realiza **limpieza básica**, eliminando entradas sin nombre o con campos vacíos.
- Se aplican **expresiones regulares** para validar texto y evitar registros corruptos.
- Se estructura la información con nombres consistentes.
- Se calcula estadística descriptiva sobre las coordenadas geográficas.
- Se exportan los datos en formato “.json” y “.xlsx” para uso posterior.

Este tratamiento es fundamental para evitar errores en el análisis estadístico y para asegurar la integridad de los datos exportados.

#### Minutas de trabajo:

Fecha	Actividad realizada	Responsables	Observaciones
28-abr-2025	Revisión de documentación de la API del INEGI	Jaime Gonzalez, Mauro López, Raziél Torres	Se generó token de acceso
29-abr-2025	Pruebas iniciales con solicitudes a una coordenada	Jaime Gonzalez, Mauro López, Raziél Torres	Se confirmó funcionamiento
30-abr-2025	Implementación de múltiples coordenadas y limpieza	Jaime Gonzalez, Mauro López, Raziél Torres	Se mejoró cobertura de datos
01-may-2025	Análisis estadístico y visualización	Jaime Gonzalez, Mauro López, Raziél Torres	Se implementaron gráficos
02-may-2025	Exportación a Excel y validación final	Jaime Gonzalez, Mauro López, Raziél Torres	Documento final