Documentación del Proyecto de Análisis de Datos

**Análisis Empresarial Mexicano SIEM 2024**

Fecha: 19/05/2025

Autor(es):

Said Mariano Sánchez

Organización: Independiente

Propósito del Documento:

Resúmenes Ejecutivos:

Contenido

[1. Introducción 3](#_Toc198547850)

[2. Descripción del Proyecto 4](#_Toc198547851)

[2.1 Antecedentes 4](#_Toc198547852)

[2.2 Objetivos 5](#_Toc198547853)

[2.2.1 Objetivo Principal 5](#_Toc198547854)

[2.2.2 Objetivos Específicos 5](#_Toc198547855)

[2.3 Alcance 5](#_Toc198547856)

[2.4 Justificación 5](#_Toc198547857)

[2.5 Preguntas de Investigación 5](#_Toc198547858)

[3. Datos 6](#_Toc198547859)

[3.1 Fuentes de Datos 6](#_Toc198547860)

[3.1.1 Fuente 1 6](#_Toc198547861)

[3.1.2 Fuente 2 6](#_Toc198547862)

[3.2 Recolección de Datos 6](#_Toc198547863)

[3.3 Descripción de los Datos 6](#_Toc198547864)

[3.3.1 Variables: (Nombre, Tipo, Descripción) 6](#_Toc198547865)

[3.4 Calidad de los Datos 7](#_Toc198547866)

[3.5 Consideraciones Éticas 7](#_Toc198547867)

[4. Metodología 8](#_Toc198547868)

[4.1 Enfoque del Análisis 8](#_Toc198547869)

[4.2 Herramientas y Técnicas 8](#_Toc198547870)

[4.3 Preparación de Datos 8](#_Toc198547871)

[4.4 Análisis Exploratorio de Datos (EDA) 8](#_Toc198547872)

[4.5 Modelado 8](#_Toc198547873)

[4.5.1 Selección del Modelo Se aplicarán pruebas estadísticas para evaluar hipótesis específicas: 8](#_Toc198547874)

[4.5.2 Entrenamiento del Modelo 8](#_Toc198547875)

[4.5.3 Evaluación del Modelo 8](#_Toc198547876)

[4.6 Validación de Resultados 8](#_Toc198547877)

[5. Resultados 9](#_Toc198547878)

[5.1 Hallazgos Clave 9](#_Toc198547879)

[5.2 Visualizaciones 9](#_Toc198547880)

[5.3 Interpretación de Resultados 9](#_Toc198547881)

[6. Conclusiones y Recomendaciones 10](#_Toc198547882)

[6.1 Conclusiones 10](#_Toc198547883)

[6.2 Limitaciones del Estudio 10](#_Toc198547884)

[6.3 Recomendaciones 10](#_Toc198547885)

[6.4 Próximos Pasos 10](#_Toc198547886)

[7. Apéndices 11](#_Toc198547887)

[7.1 Glosario 11](#_Toc198547888)

[7.2 Código 11](#_Toc198547889)

[7.3 Datos Adicionales 11](#_Toc198547890)

[7.4 Referencias 11](#_Toc198547891)

# Introducción

El **Sistema de Información Empresarial Mexicano (SIEM)** es una herramienta oficial del Estado mexicano que recopila y difunde información sobre empresas de los sectores industrial, comercial, de servicios y turismo . Este proyecto tiene como objetivo analizar los datos del SIEM correspondientes al año 2024 para identificar patrones en la distribución, tamaño y sector económico de las empresas registradas en México.

# 2. Descripción del Proyecto

## 2.1 Antecedentes

El SIEM ofrece una base de datos pública y actualizada que permite conocer las características de las empresas en México. Esta información es esencial para estudios económicos, planificación de políticas públicas y análisis de mercado.

## 2.2 Objetivos

## 2.2.1 Objetivo Principal

Analizar la distribución y características de las empresas registradas en el SIEM 2024 para obtener insights sobre su localización, tamaño, sector económico predominante y diferencias regionales, utilizando técnicas estadísticas y visualización de datos.

## 2.2.2 Objetivos Específicos

* Determinar la distribución geográfica de las empresas por estado y municipio.
* Identificar los sectores económicos predominantes en cada entidad federativa.
* Analizar la relación entre el tamaño de las empresas y su sector económico.
* Evaluar diferencias regionales en la proporción de empresas grandes, medianas y pequeñas.

## 2.3 Alcance

El análisis se centrará en los datos del SIEM correspondientes al año 2024, abarcando todas las entidades federativas de México y considerando variables como ubicación, sector económico y tamaño de las empresas.

## 2.4 Justificación

Comprender la estructura empresarial del país es vital para la toma de decisiones estratégicas en sectores públicos y privados. Este análisis proporcionará información valiosa para identificar oportunidades de desarrollo económico y áreas que requieren atención específica.

## 2.5 Preguntas de Investigación

* ¿Existen diferencias significativas en el número de empresas registradas entre regiones del país?
* ¿Qué sectores dominan por entidad federativa?
* ¿Los estados con mayor número de empresas tienen también mayor proporción de empresas grandes?
* ¿Cuál es la relación entre el tamaño de empresa y el sector económico?
* ¿Las empresas del sector industrial tienden a ser más grandes que las del sector comercial?

# 3. Datos

## 3.1 Fuentes de Datos

## 3.1.1 Fuente 1

* **Nombre:** Sistema de Información Empresarial Mexicano (SIEM)
* **Descripción:** Base de datos oficial que contiene información sobre empresas registradas en México.
* **URL**: <https://www.datos.gob.mx/dataset/sistema_informacion_empresarial_mexicano>

## 3.1.2 Fuente 2

No aplica

## 3.2 Recolección de Datos

Los datos se descargarán directamente desde el portal de datos abiertos del gobierno de México. Se verificará la integridad y actualidad de los archivos, y se documentará el proceso de descarga y almacenamiento.

## 3.3 Descripción de los Datos

### 3.3.1 Variables: (Nombre, Tipo, Descripción)

|  |  |  |
| --- | --- | --- |
| Variable | Descripción | Tipo de dato |
| \_id | Identificador único del registro | Numérico |
| razon\_social | Nombre de la empresa o establecimiento | Texto |
| estado | Estado de la República Mexicana | Texto |
| municipio | Municipio del establecimiento | Texto |
| domicilio | Dirección completa | Texto |
| colonia | Colonia donde se ubica el establecimiento | Texto |
| cp | Código postal | Numérico |
| telefono | Teléfono de contacto | Numérico |
| e\_mail | Correo electrónico de contacto | Texto |
| giro | Giro comercial o actividad económica principal | Texto |
| scian | Código SCIAN correspondiente al giro | Numérico |
| rango\_de\_empleados | Rango estimado de empleados en el establecimiento | Texto |
| registrado\_por | Organización o cámara que registró el negocio | Texto |
| fn\_razon\_social | Nombre comercial final, si aplica | Texto |
| fn\_telefono | Teléfono adicional o actualizado | Numérico |
| fn\_e\_mail | Correo electrónico adicional o actualizado | Texto |
| fn\_giro | Giro comercial final corregido, si aplica | Texto |
| estado\_etq | Etiqueta o formato estándar del estado (Ej. Estado de México) | Texto |

## 3.4 Calidad de los Datos

Se evaluará la **calidad** de los datos mediante:

* Detección y eliminación de duplicados.
* Identificación de valores faltantes o atípicos.
* Verificación de la consistencia en la codificación de variables categóricas.

## 3.5 Consideraciones Éticas

Aunque los datos del SIEM son de acceso público, se garantizará la confidencialidad de la información sensible y se utilizarán exclusivamente con fines de análisis estadístico y académico.

# 4. Metodología

## 4.1 Enfoque del Análisis

Se empleará un enfoque cuantitativo, utilizando técnicas de estadística descriptiva e inferencial para analizar las características de las empresas registradas en el SIEM.

## 4.2 Herramientas y Técnicas

* **Lenguaje de programación:** Python
* **Bibliotecas:** pandas, numpy, matplotlib, seaborn, plotly, scipy
* **Entorno de desarrollo:** Jupyter Notebook

## 4.3 Preparación de Datos

* Limpieza y transformación de datos.
* Codificación de variables categóricas.
* Creación de variables derivadas, como regiones geográficas agrupadas.

## 4.4 Análisis Exploratorio de Datos (EDA)

* Análisis de distribución de empresas por estado, municipio, sector y tamaño.
* Visualización de datos mediante gráficos de barras, histogramas y mapas de calor.
* Identificación de patrones y tendencias en los datos.

## Modelado

## 4.5.1 Selección del Modelo Se aplicarán pruebas estadísticas para evaluar hipótesis específicas:

* **Prueba t de muestras independientes:** Para comparar promedios entre regiones.
* **Prueba de chi-cuadrada:** Para evaluar la independencia entre variables categóricas.

## 4.5.2 Entrenamiento del Modelo

No aplica, ya que el análisis se basa en pruebas estadísticas y no en modelos predictivos.

## 4.5.3 Evaluación del Modelo

Se interpretarán los resultados de las pruebas estadísticas, considerando niveles de significancia y tamaños del efecto.

## 4.6 Validación de Resultados

Se realizarán análisis de sensibilidad y se verificarán los supuestos de las pruebas estadísticas para asegurar la validez de los resultados.

# 5. Resultados

## 5.1 Hallazgos Clave

Aplica para la siguiente versión del documento

## 5.2 Visualizaciones

Se presentarán gráficos y mapas interactivos que ilustren los hallazgos clave, facilitando la comprensión de los patrones identificados (En la siguiente versión del documento).

## 5.3 Interpretación de Resultados

Aplica parala siguiente versión del documento

# 6. Conclusiones y Recomendaciones

## 6.1 Conclusiones

## 6.2 Limitaciones del Estudio

## 6.3 Recomendaciones

## 6.4 Próximos Pasos

# 7. Apéndices

## 7.1 Glosario

* **SIEM:** Sistema de Información Empresarial Mexicano.
* **EDA:** Análisis Exploratorio de Datos.

## 7.2 Código

El código utilizado para el análisis está disponible en el repositorio de GitHub: https://github.com/SaitoM17/analisis\_empresarial\_mexico\_siem\_2024

## 7.3 Datos Adicionales

## 7.4 Referencias

Secretaría de Economía. (s.f.). Conoce el Sistema de Información Empresarial Mexicano. Recuperado de https://www.gob.mx/se/articulos/conoce-el-sistema-de-informacion-empresarial-mexicano