



# Automatic Generation of Informative Documents Based on Official Statistics

Sprint on Artificial Intelligence and Data Science for Economic Statistics



DATA SCIENCE LAB

---

12/12/24

---

# Insights into the Industry

- Insights into the Industry: Sectoral and regional studies by INEGI.
- Developed with industrial chambers.
- Covers activities, employment, wages, production chains, and inflation.
- Presented in simple language with graphs and tables.



Housing industry



Automotive industry



Aerospace industry

# Example of LLM Chatbot



Chatbot LLM  
Mixtral 8x7B



Expert Instructions

# Example of LLM Chatbot

## llama.cpp

Reset all to default☒ Chat ☐ Completion

**Prompt**  

Esta es una interacción automatizada con un asistente virtual especializado en análisis económico y datos estadísticos del INEGI. El asistente está diseñado para ofrecer soporte informativo, ser eficiente y preciso en sus respuestas, y proporcionar asistencia inmediata con un enfoque en la exactitud y la claridad. Por favor, indique su consulta para recibir asistencia detallada y específica.

User name

Bot name

Usuario

MixtralStatBot

**Prompt template**  

{{prompt}}

**Chat history template**  

{{name}}: {{message}}

**Grammar**  

Use gbnf or JSON Schema+convert

order: prop1,prop2,prop3 Convert JSON Schema

**Predictions**

1010

**Temperature**

0.28

**Penalize repeat sequence**

1.55

**Consider N tokens for penalize**

422

**Top-K sampling**

52

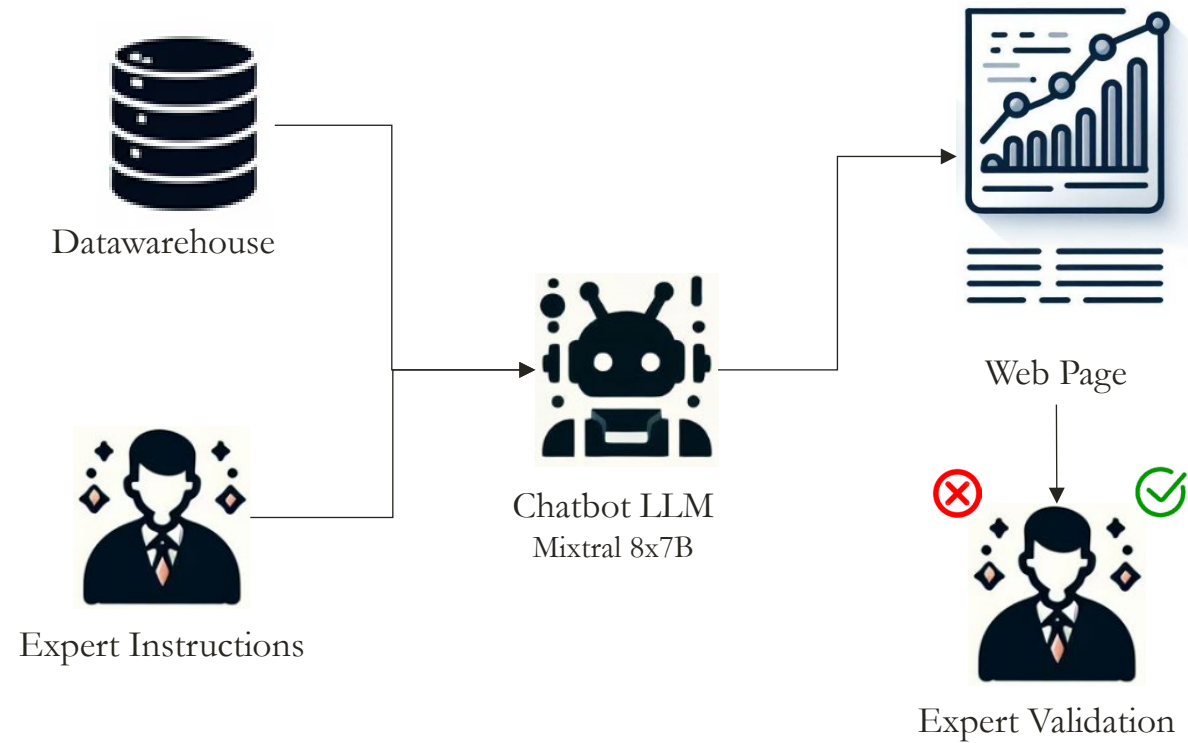
**Top-P sampling**

0.4

► More options

Say something...

# API Example Insights into the Industry





# Insights into the Industry

## Industria textil y del vestido

### Introducción



La industria textil y del vestido en México es un sector económico importante que ha experimentado un crecimiento significativo en los últimos años. Con una larga tradición en la producción de textiles, México se ha posicionado como el quinto productor mundial de algodón y el décimo en hilados y tejidos. Esta industria emplea directamente a más de 600 mil personas y genera ingresos por más de 25 mil millones de dólares anuales. La industria se caracteriza por su diversidad, abarcando desde la producción de materias primas hasta la confección de prendas de vestir y artículos textiles para el hogar. A pesar de los desafíos globales, como la competencia de países asiáticos y las fluctuaciones en el precio de las materias primas, la industria textil y del vestido en México continúa siendo un pilar fundamental de la economía nacional.

#### Producto Interno Bruto

**94,908**

Millones de pesos a precios reales

**-5.4**

Variación con respecto al mismo trimestre del año anterior

4° trimestre 2023

#### Puestos de trabajo

**519,313**

puestos de trabajo

**-0.2**

Variación porcentual anual

Año 2022

#### Inflación

**3.3%**

Prendas de vestir  
Variación con respecto al mismo mes del año anterior

Enero 2024

#### Comercio exterior

Exportaciones

**841**

millones de dólares

**0.1%**

Variación con respecto al mismo mes del año anterior

Diciembre 2023

#### Comercio exterior

Importaciones

**1,025**

millones de dólares

**3.2%**

Variación con respecto al mismo mes del año anterior

Diciembre 2023

### Índice

[Delimitación](#)

[Producción](#)

# Insights into the Industry

## Steps to create an AI generated document

1. Define relevant indicators for each industry: Contribution to GDP, Inflation, Exports, Imports, Salaries, Workforce, Supply chain, number of economic units per state, etc.
2. Use INEGI internal API to retrieve necessary data and calculate indicators
3. Interpret indicators using LLM
4. Create a website to display the data retrieved alongside interactive visualizations and AI generated insights.

# Insights into the Industry

## Steps to create an AI generated document

1. Define relevant indicators for each industry: GDP, Contribution to GDP, Inflation, Exports, Imports, Salaries, Workforce, Supply chain, number of economic units per state, etc.

### Industria textil y de la confección

#### Introducción



La industria textil y de la confección se refiere al conjunto de actividades económicas relacionadas con la producción de textiles y prendas de vestir. Esta industria abarca desde la fabricación de telas y materiales textiles hasta la confección de ropa. Aquí se ofrece un panorama económico de esta industria a través de diversos indicadores tales como: producción, empleo, Remuneraciones, cadenas productivas, precios, Exportaciones e Importaciones, entre otros temas, lo que permite dimensionar la importancia de la industria textil y de la confección en la economía.

#### Contenido

- [Delimitación](#)
- [Producto Interno Bruto \(PIB\)](#)
- [Cadenas de valor](#)
- [Puestos de trabajo](#)
- [Remuneraciones](#)
- [Inflación y precios](#)
- [Comercio exterior](#)
- [Anexo](#)



### Industria automotriz

#### Introducción



La Industria automotriz abarca un conjunto de actividades económicas, desde la producción de componentes y piezas hasta la fabricación de automóviles y camiones; esta industria desempeña un papel fundamental en México con conexiones relevantes en el mercado exterior. A continuación, se presentan diversos indicadores importantes como la Producción bruta, los Puestos de trabajo, las Remuneraciones, las Cadenas de valor, así como las Exportaciones e Importaciones lo que permite dimensionar la importancia de la Industria automotriz en la actividad económica de México y a nivel internacional.

#### Contenido

- [Delimitación](#)
- [Producto Interno Bruto \(PIB\)](#)
- [Entidades federativas](#)
- [Cadenas de valor](#)
- [Puestos de trabajo](#)
- [Remuneraciones](#)
- [Inflación y precios](#)
- [Comercio exterior](#)
- [Anexo](#)





# Insights into the Industry

## Steps to create an AI generated document

2. Use INEGI internal API to retrieve necessary data and calculate indicators

```
def get_data(lista_ids, token, fecha=None):
    url = "http://api.inegi.com.mx/service/bie"

    payload = json.dumps({
        "nombreServicio": "CONOCIENDO_INDUSTRIA",
        "P_LISTA_NUM_SERIE": lista_ids,
        "P_ANO": fecha
    })

    headers = {
        'Content-Type': 'application/json',
        'Authorization': 'Bearer ' + token
    }

    response = requests.request("POST", url, headers=headers, data=payload)
    if response.ok:
        return response.json()
    else:
        return response
```

Python notebooks are used to retrieve necessary data using the infrastructure from the Data Science Lab.

Expert economists provide the list of variables to retrieve, specifying granularity and time span, as well as the definition of variables (ratios, percentages, growth rates, etc.)

# Insights into the Industry

## Steps to create an AI generated document

### 3. Interpret indicators using LLM



2x 32GB Tesla V100S GPUs

1TB RAM

224 Cores

Mistral AI is a French company that released Mixtral 8x7B in December 2023. It is a sparse mixture of experts model (SMoE) with open weights . At the time of release it was the best overall model regarding cost/performance. It is similar to GPT3.5 in most standard benchmarks.

- Up to 32k tokens
- Handles English, French, Italian, German and Spanish
- It can be fine tuned to achieve 8.3 on MT Bench (engage in coherent, informative, and engaging conversations)

# Insights into the Industry

## 3. Interpret indicators using LLM

### Main issues

Mixtral 8x7B is not as good as we needed it to be when using regular prompts. Prompt engineering techniques were used to achieve the results we wanted.

### Arithmetics and object interpretation

Mixtral has issues doing basic arithmetics (sums, percentages of total, comparing values), as well as providing information about tables (total, min, max, mean, last and first elements).

|   | table |       |
|---|-------|-------|
|   | Year  | Value |
| 0 | 2020  | 140   |
| 1 | 2021  | 150   |
| 2 | 2022  | 90    |
| 3 | 2023  | 115   |

Prompt

Analyze the table {table}. Calculate the annual variation for the two most recent years. The indicator refers to the annual production of automobiles. Provide the result in a short paragraph of no more than three lines in Spanish.

Result:

The annual variation for the two most recent years is **19.4%** between 2022 and 2023 (calculated as  $(115-90)/90$ ) and **7.3%** between 2021 and 2022 (calculated as  $(90-150)/150$ ). Therefore, automobile production increased by 7.3% in 2022 compared to 2021 but decreased by 19.4% in 2023 compared to 2022.

**27.8% and -40% are the correct values**

# Insights into the Industry

## 3. Interpret indicators using LLM

### Main issues

#### Arithmetics and object interpretation

Mixtral has issues doing basic arithmetics (sums, percentages of total, comparing values), as well as providing information about tables (total, min, max, mean, last and first elements).

**Solution:** calculate indicators beforehand and include them in the prompt.

#### New prompt:

Interpret the variation from 2021 to 2022 contained in {var1} and the variation from 2022 to 2023 contained in {var2} in a short paragraph. Use percentages.

#### New result:

The variation of -40.0% between 2021 and 2022 indicates a significant decrease in the measured value or quantity. On the other hand, the variation of 27.8% between 2022 and 2023 represents a considerable increase in the same value or quantity. These percentages reflect the fluctuations that have occurred during these periods.

# Insights into the Industry

## 3. Interpret indicators using LLM:

### Main issues

#### Language defaults to English when not strictly specified

When not told to provide a response in Spanish, even when the main instruction is in Spanish, it will write some of the answer in English.

Original prompt in Spanish:

Interpreta la variación interanual para 2021 y 2022 que está en {var1}, y para 2022 y 2023 que está en {var2} en un párrafo corto. Escribe otro párrafo corto y menciona qué año tuvo el menor valor de {tabla}. Haz un comentario breve al respecto. El indicador se refiere a la producción de ropa y calzado en México.

Prompt (translation):

Interpret the year-on-year variation for 2021 and 2022, which is in {var1}, and for 2022 and 2023, which is in {var2}, in a short paragraph. Write another short paragraph mentioning which year had the lowest value in {table}. Add a brief comment about it. The indicator refers to the production of clothing and footwear in Mexico.

Response:

La variación interanual para 2021 y 2022 fue de -40.0, indicating a significant decrease in the production of clothing and footwear in Mexico during this period. However, the situation improved in 2022 and 2023, with a variation of 27.8, suggesting a substantial recovery in the sector.

Looking at the Year and Value data provided, it is evident that the year 2022 had the lowest value at 90. This suggests that the production of clothing and footwear in Mexico experienced a considerable decline in that year, despite the recovery seen in the following year. It would be interesting to investigate the factors contributing to this decrease, such as economic conditions, supply chain disruptions, or changes in consumer behavior.

# Insights into the Industry

## 3. Interpret indicators using LLM

### Main issues

#### Language defaults to English when not strictly specified

When not told to provide a response in Spanish, even when the main instruction is in Spanish, it will write some of the answer in English.

**Solution:** add context, indicate the desired output language.

#### Main instruction :

Interpreta la variación interanual para 2021 y 2022 que está en {var1}, y para 2022 y 2023 que está en {var2} en un párrafo corto.  
Escribe otro párrafo corto y menciona qué año tuvo el menor valor de {tabla}.  
Haz un comentario breve al respecto. El indicador se refiere a la producción de ropa y calzado en México.

#### Context:

Dame el resultado en español y con una redacción formal para publicarlo en el sitio web de la oficina nacional de estadística de México (buscar: INEGI)

#### Main instruction:

Interpret the year-on-year variation for 2021 and 2022, which is in {var1}, and for 2022 and 2023, which is in {var2}, in a short paragraph. Write another short paragraph mentioning which year had the lowest value in {table}. Add a brief comment about it. The indicator refers to the production of clothing and footwear in Mexico.

#### Context:

Provide the result in Spanish, written in a formal tone suitable for publication on the website of Mexico's National Statistics Office (search: INEGI).

#### Response:

La variación interanual de la producción de ropa y calzado en México experimentó una disminución significativa de 40.0% entre los años 2021 y 2022. Sin embargo, este indicador se recuperó notablemente alcanzando un incremento de 27.8% durante el periodo posterior, es decir, entre 2022 y 2023.

Para contextualizar estas cifras, resulta relevante señalar que el año 2022 registró el menor valor en la producción de ropa y calzado en México en los últimos cuatro años, con un total de 90 unidades. Este dato sugiere un desafío temporal en la industria durante dicho periodo. No obstante, el aumento del 27.8% en la producción entre 2022 y 2023 indica una clara tendencia a la recuperación y fortalecimiento de este sector en el país.



# Insights into the Industry

## 3. Interpret indicators using LLM:

### Main issues

#### Results can vary and sometimes are not good enough

The same prompt can produce different results. Even when changing the temperature parameter (randomness or reactivity of the response) the results can be unsatisfactory.

**Solution:** iterate the prompt and create a secondary prompt to ignore outlier responses and summarise satisfactory ones.

#### First prompt:

The sum of the percentage values of the "datos" results in the "indicador," which has the value "total." In turn, "total" represents the value of "sector" in the country's manufacturing production. Create a text that mentions these are the data for the year "anio."

Consider the following points for the text:

1. Provide a short, single-paragraph text.
2. Write the text in the style of an expert Mexican economist, using past tense.
3. Do not put the names in quotes; integrate them into the text.
4. Write the text in Spanish.
5. Do not make additional interpretations.
6. Do not mention anything about the objects used.

Use the following example:

"In 2019, the vehicle industry represented 30% of the country's manufacturing production. This, in turn, was made up of production services at 10% and manufacturing services at 20%."  
{datos} {anio} {indicador} {total} {sector}

#### Summarizing prompt:

Summarize the "texto" object.

Ignore paragraphs that differ significantly from the rest. Provide a concise summary. {texto}

#### Context:

Provide the result in Spanish, written in a formal style for publication on the website of the National Statistics Office of Mexico (search: INEGI).

# Insights into the Industry

## 3. Interpret indicators using LLM

### Main issues

#### Response of the first prompt 10 times

En 2023, la fabricación de equipo de comunicación representó un 1.5% del total de la producción manufacturera del país. Este sector se divide en dos categorías: la fabricación de equipo telefónico con un 1.2% y la fabricación de equipo de transmisión y recepción de señales de radio y televisión, y equipo de comunicación inalámbrico con un 0.4%.

En 2023, la fabricación de equipo telefónico representó un 1.5% del total de la producción manufacturera del país, mientras que la fabricación de equipo de transmisión y recepción de señales de radio y televisión, y equipo de comunicación inalámbrico alcanzó un 1.2%. Estos sectores forman parte de la industria de fabricación de equipo de computación, comunicación, medición y de otros equipos, componentes y accesorios electrónicos, que en su conjunto representa una proporción significativa del sector manufacturero nacional.

En 2023, la Fabricación de equipo telefónico y la Fabricación de equipo de transmisión y recepción de señales de radio y televisión, y equipo de comunicación inalámbrico representaron el 1.5% de la producción manufacturera total del país. Este sector se dividió en Fabricación de equipo de computación, comunicación, medición y de otros equipos, componentes y accesorios electrónicos con un 1.2%, y Fabricación de equipo de comunicación con un 0.4%.

En 2023, la Fabricación de equipo telefónico representó un 1.5% del total de la producción manufacturera del país, mientras que la Fabricación de equipo de transmisión y recepción de señales de radio y televisión, y equipo de comunicación inalámbrico alcanzó un 1.2%. Estos sectores contribuyeron en su conjunto con un 2.7% a la producción manufacturera nacional del año.

En 2023, la Fabricación de equipo telefónico representó un 1.5% del total de la producción manufacturera del país, mientras que la Fabricación de equipo de transmisión y recepción de señales de radio y televisión, y equipo de comunicación inalámbrico alcanzó un 1.2%. Estos sectores contribuyeron en su conjunto con un 2.7% a la producción manufacturera nacional del año.

En 2023, la Fabricación de equipo telefónico representó un 1.5% del total de la producción manufacturera del país, mientras que la Fabricación de equipo de transmisión y recepción de señales de radio y televisión, y equipo de comunicación inalámbrico alcanzó un 1.2%. Estos sectores contribuyeron en su conjunto con un 2.7% a la producción manufacturera nacional del año.

En 2023, la fabricación de equipo telefónico y la fabricación de equipo de transmisión y recepción de señales de radio y televisión, así como equipo de comunicación inalámbrico, representaron el 1.5% de la producción manufacturera total del país. Este sector se dividió en dos categorías: la fabricación de equipo de computación, comunicación, medición y de otros equipos, componentes y accesorios electrónicos con un 1.2%, y los servicios de producción con un 0.4%.

En 2023, la Fabricación de equipo telefónico representó un 1.5% del total de la producción manufacturera del país, mientras que la Fabricación de equipo de transmisión y recepción de señales de radio y televisión, y equipo de comunicación inalámbrico alcanzó un 1.2%. Estos sectores contribuyeron en su conjunto con un 2.7% a la producción manufacturera nacional del año.

En 2023, la Fabricación de equipo telefónico representó un 1.5% del total de la producción manufacturera del país, mientras que la Fabricación de equipo de transmisión y recepción de señales de radio y televisión, y equipo de comunicación inalámbrico alcanzó un 1.2%. Estos sectores contribuyeron en su conjunto con un 2.7% a la producción manufacturera nacional del año.

En 2023, la fabricación de equipo telefónico y la fabricación de equipo de transmisión y recepción de señales de radio y televisión, y equipo de comunicación inalámbrico representaron conjuntamente el 1.5% de la producción manufacturera total del país. Este sector se dividió en la fabricación de equipo de computación, comunicación, medición y de otros equipos, componentes y accesorios electrónicos, con un 1.2%, y los servicios de fabricación de equipo de comunicación, que alcanzaron el 0.4%.

#### Response of the summarizing prompt:

En 2023, la fabricación de equipo de comunicación representó un 1.6% del total de la producción manufacturera en México. Este sector se subdivide en dos categorías: la fabricación de equipo telefónico con una participación del 1.2%, y la fabricación de equipo de transmisión y recepción de señales de radio y televisión, y equipo de comunicación inalámbrico con un 0.4%. Estas industrias contribuyen significativamente a la producción manufacturera nacional.

#### Translation:

In 2023, the manufacture of communication equipment accounted for 1.5% of the total manufacturing production in Mexico. This sector is subdivided into two categories: the manufacture of telephone equipment, with a 1.2% share, and the manufacture of radio and television signal transmission and reception equipment, as well as wireless communication equipment, with 0.4%. These industries make a significant contribution to the national manufacturing production.

# Insights into the Industry

## 3. Interpret indicators using LLM

### New strategies

- Create prompts in English, translate to Spanish at the end.

#### Prompt:

Do the following:

1. Calculate YoY variation for the column "Value" in "table" for the three most recent values in column Year. Use the formula  $(t_2 - t_1) / t_1$ .

Interpret the results in a short paragraph. Verify your results.

From the same table, in a different paragraph, look for the lowest and the highest value

and write a short paragraph. The values refer to the coffee industry in Mexico.

Don't make references to the objects used. Verify your results.

2. Take the paragraphs in step 1 and make a coherent and concise interpretation without formulas or steps followed.

Try not to include a lot of technical terminology.

3. Translate the result of the second step to formal Spanish.  
{table}

#### Response:

1. The Year-over-Year (YoY) variation for the "table" in the coffee industry of Mexico for the two most recent years is calculated as follows:

- For 2021 to 2022:  $(90 - 150) / 150 = -40\%$
- For 2022 to 2023:  $(115 - 90) / 90 = 27.78\%$

The coffee industry in Mexico saw a significant decrease of 40% from 2021 to 2022, followed by a substantial increase of approximately 27.78% from 2022 to 2023.

Verification: The calculations were performed using the formula  $(t_0 - t_1) / t_1$  where  $t_0$  is the value in the current year and  $t_1$  is the value in the previous year.

The lowest value recorded was 90, which occurred in 2022, while the highest value of 150 was observed in 2021. This indicates a decline in the industry between these two years followed by a recovery in the subsequent year.

2. In the Mexican coffee industry, there was a considerable decrease of about 40% from 2021 to 2022 and then an approximate recovery of 27.78% from 2022 to 2023. The lowest value recorded was in 2022, while the highest was seen in 2021.

3. En la industria cafetera de México, se observó una disminución considerable de aproximadamente el 40% entre 2021 y 2022, seguida de una recuperación aproximada del 27.78% entre 2022 y 2023. El valor más bajo se registró en 2022, mientras que el más alto se observó en 2021.

# Insights into the Industry

## 3. Interpret indicators using LLM

### New strategies

- **Create meta-prompts: identify what the section is about and choose an adequate prompt.**

#### **Prompt:**

Based on {doc\_section}, select the adequate prompt to create a short text based on the available data:

#### 1. **\*\*Contribution to GDP Prompt:\*\***

Interpret the contribution to GDP shown in {gdp\_data} in a short paragraph. Analyze how this data reflects the economic impact of the respective sector in the country's overall economic performance.

#### 2. **\*\*Employment Prompt:\*\***

Interpret the employment data presented in {employment\_data} in a short paragraph. Discuss the trends in job creation or loss within the sector, highlighting the overall impact on the labor market.

#### 3. **\*\*Growth Rate Prompt:\*\***

Interpret the growth rate presented in {growth\_rate\_data} in a short paragraph. Analyze how this rate indicates the sector's performance over time and its potential future trajectory in the economy.

# Insights into the Industry

## Steps to create an AI generated document

4. Create a website to display the data retrieved alongside interactive visualizations and AI generated insights.

## Main advantages:

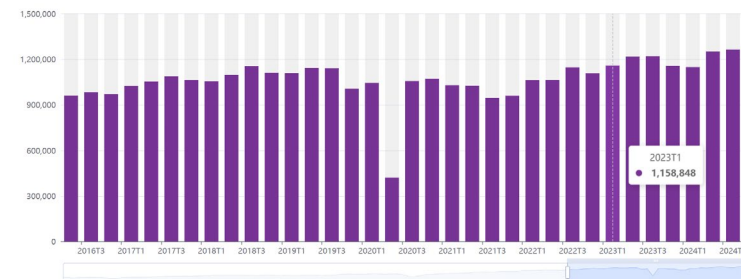
- Text, visualizations and downloadable data can be updated mere hours after new information is available. This process used to take at least three months.
- Interested parties can stay up to date with the relevant indicators and information produced by INEGI in a single website, without having to aggregate information by their NAICS code for sectors and subsectors.
- A note indicating that the document contains AI generated content supervised by experts will be displayed on the website.

### Producto Interno Bruto (PIB)

A lo largo de su historia, el Producto Interno Bruto (PIB) de la Industria automotriz ha experimentado incrementos, a pesar de haber caído en los años 1995, 2009 y 2020.

La Industria automotriz experimentó una variación del 3.5 % en el tercer trimestre del año 2024 respecto al mismo trimestre del año anterior.

#### Evolución del PIB de la Industria automotriz (Millones de pesos a precios de 2018)



Fuente: INEGI, Sistema de Cuentas Nacionales de México (SCNM), base 2018.

La producción de automóviles y camiones en el tercer trimestre del año 2024 tuvo una variación de 7.9 % respecto a la del mismo trimestre del año anterior.

### Entidades federativas

#### La Industria automotriz en las entidades federativas

En 2018, la mayor Producción bruta en la Industria automotriz se concentró en seis entidades federativas, las cuales representaron el 78.1 % de la producción nacional. Estas entidades Coahuila de Zaragoza, Nuevo León, Puebla, México, Aguascalientes y Guanajuato.



Nota: Los datos de las unidades económicas se obtuvieron del Directorio Estadístico Nacional de Unidades Económicas (DENUE), a noviembre de 2024.  
El personal ocupado y la producción bruta de los Censos Económicos, 2018.  
Para mayor detalle, consulte la ubicación de las unidades económicas de la industria automotriz.

Fuente: INEGI, Censos Económicos (CE), 2018; Directorio Estadístico Nacional de Unidades Económicas (DENUE), a noviembre de 2024.

### Principales proveedores de insumos de la rama 3361 Fabricación de automóviles y camiones (Participación porcentual)



# Insights into the Industry

## INEGI Data Science Lab Team

|                    |  |
|--------------------|--|
| Elio Villaseñor    | <a href="mailto:elio.villaseñor@inegi.org.mx">elio.villaseñor@inegi.org.mx</a>       |
| Abel Coronado      | <a href="mailto:abel.coronado@inegi.org.mx">abel.coronado@inegi.org.mx</a>           |
| Víctor Silva       | <a href="mailto:victor.silvac@inegi.org.mx">victor.silvac@inegi.org.mx</a>           |
| Alejandra Figueroa | <a href="mailto:alejandra.figueroa@inegi.org.mx">alejandra.figueroa@inegi.org.mx</a> |
| Irving Cabrera     | <a href="mailto:irving.cabrera@inegi.org.mx">irving.cabrera@inegi.org.mx</a>         |
| Amado Esquer       | <a href="mailto:amado.esquer@inegi.org.mx">amado.esquer@inegi.org.mx</a>             |

## INEGI Direction of Economic Studies

|               |  |
|---------------|--|
| Jorge Quevedo | <a href="mailto:jorge.quevedo@inegi.org.mx">jorge.quevedo@inegi.org.mx</a> |
| Natalia Anaya | <a href="mailto:eduardo.anaya@inegi.org.mx">eduardo.anaya@inegi.org.mx</a> |
| Felipe García | <a href="mailto:felipe.garcia@inegi.org.mx">felipe.garcia@inegi.org.mx</a> |



# GRACIAS

**Conociendo  
México**

800 111 46 34  
[www.inegi.org.mx](http://www.inegi.org.mx)  
[atencion.usuarios@inegi.org.mx](mailto:atencion.usuarios@inegi.org.mx)

    **INEGI** Informa