

## ▼ Import important libraries

```
import pandas as pd
import numpy as np
```

## ▼ Initial Loading and Cleanup

Load the CSV file.

Rename the columns for easier use in code.

Convert numeric columns to float to avoid errors in subsequent calculations.

```
df = pd.read_csv('temperature_012024_092025.csv', encoding='utf-8')
df.columns = ['longitude', 'latitude', 'code', 'state', 'station', 'temperature_min', 'temperature_max', 'year', 'month']
df[['longitude', 'latitude', 'temperature_min', 'temperature_max']] = df[['longitude', 'latitude', 'temperature_min', 'temperature_max']].apply(pd.to_numeric)
```

## ▼ Replace State Abbreviations

```
reemplazos = {
    'AGS': 'Aguascalientes',
    'BC': 'Baja California',
    'BCS': 'Baja California Sur',
    'CAMP': 'Campeche',
    'CDMX': 'Ciudad de México',
    'CHIH': 'Chihuahua',
    'CHIS': 'Chiapas',
    'COAH': 'Coahuila',
    'COL': 'Colima',
    'DF': 'Ciudad de México',
    'DGO': 'Durango',
    'GRO': 'Guerrero',
    'GTO': 'Guanajuato',
    'HGO': 'Hidalgo',
    'JAL': 'Jalisco',
    'MEX': 'Estado de México',
    'MICH': 'Michoacán',
    'MOR': 'Morelos',
    'NAY': 'Nayarit',
    'NL': 'Nuevo León',
    'OAX': 'Oaxaca',
    'PUE': 'Puebla',
    'QRO': 'Querétaro',
    'QROO': 'Quintana Roo',
    'SIN': 'Sinaloa',
    'SLP': 'San Luis Potosí',
    'SON': 'Sonora',
    'TAB': 'Tabasco',
    'TAMP': 'Tamaulipas',
    'TAMPS': 'Tamaulipas',
    'TLAX': 'Tlaxcala',
    'VER': 'Veracruz',
    'YUC': 'Yucatán',
    'ZAC': 'Zacatecas'
}

df = df.replace(reemplazos)
```

## ▼ Add region column to group states

```
# Diccionario de regiones por estado
regiones = {
```

```
'Norte': ['Baja California', 'Chihuahua', 'Coahuila', 'Durango', 'Nuevo León', 'Sonora', 'Tamaulipas'],
'Centro-Norte': ['Aguascalientes', 'Zacatecas', 'San Luis Potosí', 'Guanajuato', 'Querétaro'],
'Centro': ['Ciudad de México', 'Estado de México', 'Hidalgo', 'Morelos', 'Puebla', 'Tlaxcala'],
'Occidente': ['Jalisco', 'Colima', 'Michoacán', 'Nayarit'],
'Oriente': ['Veracruz'],
'Sur': ['Guerrero', 'Oaxaca', 'Chiapas'],
'Sureste': ['Tabasco', 'Campeche', 'Yucatán', 'Quintana Roo'],
'Noroeste': ['Baja California Sur', 'Sinaloa']
}

# Función para asignar región
def asignar_region(state):
    for region, lista in regiones.items():
        if state in lista:
            return region
    return 'desconocido'

# Aplicar al DataFrame
df['region'] = df['state'].apply(asignar_region)
```

```
print(df['state'].nunique())
print(df['state'].unique())
```

```
32
['Aguascalientes' 'Baja California' 'Baja California Sur' 'Campeche'
'Ciudad de México' 'Chihuahua' 'Chiapas' 'Coahuila' 'Colima' 'Durango'
'Guerrero' 'Guanajuato' 'Hidalgo' 'Jalisco' 'Estado de México'
'Michoacán' 'Morelos' 'Nayarit' 'Nuevo León' 'Oaxaca' 'Puebla'
'Querétaro' 'Quintana Roo' 'Sinaloa' 'San Luis Potosí' 'Sonora' 'Tabasco'
'Tamaulipas' 'Tlaxcala' 'Veracruz' 'Yucatán' 'Zacatecas']
```

```
df.head(5)
```

	longitude	latitude	code	state	station	temperature_min	temperature_max	year	month	region
0	-102.71	21.85	CALVILLO	Aguascalientes	Calvillo Ags. SMN*	5.4	26.8	2024	1	Centro-Norte
1	-102.32	21.70	MMAS	Aguascalientes	Aeropuerto Internacional de Aguascalientes Ags.*	6.7	23.4	2024	1	Centro-Norte
2	-102.37	21.78	NGRAG	Aguascalientes	El Niño García Ags.	3.6	24.1	2024	1	Centro-Norte

```
df_grouped = df.groupby(['region', 'state', 'year', 'month'])[['temperature_min', 'temperature_max']].mean().reset_index()
```

```
df_grouped.shape
```

```
(672, 6)
```

## ▼ Create a data frame with clean data

```
df.to_csv('temperature_data_clean.csv', index=False, encoding='utf-8')
df_grouped.to_csv('temperature_data_clean_grouped.csv', index=False, encoding='utf-8')
```