Shared Chat: Cleaned, Transposed CO2 Data

September 8, 2023

This chat contains files or images produced by Advanced Data Analysis which are not yet visible in Shared Chats.

User:

WorldBankEmissions.csv Spreadsheet

Please read in this dataset, describe it, and clean it by removing any years that contain all null values. Then please transpose this data set such that it forms a panel, with columns for country, years, and emissions. Only include non-null values in this preset.

ChatGPT:

Finished working

Show work[1]

The dataset appears to represent CO2 emissions (in metric tons per capita) for various countries and regions over multiple years.

Dataset Description:

- The dataset has 266 rows.
- The columns include "Country Name", "Country Code", "Indicator Name", "Indicator Code", and years from 1960 to 2020.

1

```
import pandas as pd

# Read the dataset
data = pd.read_csv('/mnt/data/WorldBankEmissions.csv')

# Describe the dataset
description = data.describe(include='all')

# Remove columns (years) that contain all null values
data_cleaned = data.dropna(axis=1, how='all')

description, data_cleaned.head()
```

- Most of the years from 1960 to 1989 had all null values, so they were removed during the cleaning process.
- The dataset now starts from the year 1990 and goes up to 2020.

Next, I'll transpose the dataset to form a panel with columns for country, years, and emissions. I'll only include non-null values in this presentation.

Finished working

Show work[2]

The dataset has been successfully transposed into a panel format. Here's a sample of the transformed dataset:

Country Name	Country Code	Indicator Name	Indicator Code	Year	Emissions
Africa Eastern and Southern	AFE	CO2 emissions (metric tons per capita)	EN.ATM.CO2E.PC	1990	0.982975
Afghanistan	AFG	CO2 emissions (metric tons per capita)	EN.ATM.CO2E.PC	1990	0.191389
Africa Western and Central	AFW	CO2 emissions (metric tons per capita)	EN.ATM.CO2E.PC	1990	0.470111
Angola	AGO	CO2 emissions (metric tons per capita)	EN.ATM.CO2E.PC	1990	0.554941
Albania	ALB	CO2 emissions (metric tons per capita)	EN.ATM.CO2E.PC	1990	1.844035

This format has columns for country name, country code, indicator name, indicator code, year, and emissions. Only non-null values for emissions are included in this presentation.

---

[2]

```
# Melt the cleaned dataframe to long format
melted_data = data_cleaned.melt(id_vars=["Country Name", "Country Code", "Indicator Name", "Indicator Code"],
                                var_name="Year", value_name="Emissions")

# Remove rows with null values in Emissions column
melted_data = melted_data.dropna(subset=["Emissions"])

# Reset index
melted_data.reset_index(drop=True, inplace=True)

melted_data.head()
```

User:

Please provide a download link for the cleaned csv dataset

ChatGPT

Finished working

Show work[3]

I apologize for the inconvenience. There seems to have been a technical hiccup. However, I've saved the cleaned dataset. You can download the cleaned CSV dataset using the link below:

Download the cleaned WorldBankEmissions dataset

---

[3]

```
# Save the cleaned dataset to CSV
output_path = "/mnt/data/cleaned_WorldBankEmissions.csv"
melted_data.to_csv(output_path, index=False)

output_path
```