Carbon Emissions Analysis Project - Comprehensive Documentation

📊 Project Overview

This project performs exploratory data analysis (EDA) and data cleaning on global carbon dioxide emissions data. The analysis aims to understand emission patterns across different countries and identify reduction opportunities through various industries using Python (Pandas, NumPy) and Power BI for visualization.

🎯 Objectives

- Analyze CO2 emission estimates across different countries and time periods
- Clean and prepare datasets for comprehensive analysis
- Identify trends and patterns in carbon emissions over time
- Explore carbon reduction strategies across different industries
- Create interactive visualizations using Power BI to communicate findings

📋 Project Scope

The project encompasses:

- Data collection from global CO2 emission datasets
- Data preprocessing and cleaning using Python
- Exploratory data analysis to uncover patterns and insights
- Data visualization through Power BI dashboards
- Identification of carbon reduction opportunities across industry sectors
- Documentation of findings and methodologies

🗂 Project Structure

```
Carbon-Emisisions/
│
├── Carbon.ipynb                        # Main Jupyter notebook with analysis
├── README.md                           # Project documentation
├── Project_Documentation.md            # Comprehensive documentation (this file)
│
├── Emissions Dataset/
│   ├── Carbon Dioxide Emission Estimates.csv     # Raw emissions data
│   └── Reductions Dataset.csv                     # Carbon reduction data
│
├── Prepared Data/
│   ├── Original/
│   │   ├── Carbon_Dioxide_Emission_Estimates_Original.csv
│   │   └── Reductions_Dataset_Original.csv
│   └── Modified/
│       ├── Carbon_Dioxide_Emission_Estimates_Modified.csv
│       └── Reductions_Dataset_Modified.csv
│
└── Dashboard_page-0001.jpg             # Dashboard visualization
```

🛠 Technologies Used

- **Python 3.x**

    - **Pandas**: Data manipulation and cleaning
    - **NumPy**: Numerical computations

- **Jupyter Notebook**: Interactive development environment
- **Power BI**: Data visualization and dashboard creation
- **CSV**: Data storage format

🔢 Detailed Data Analysis Workflow

Phase 1: Data Inspection and Context Understanding

Inspect the data for context and understanding, jotting down steps on what to erase, modify, or clean before processing using Python.

Phase 2: Pandas - NumPy Data Cleaning and Processing Steps

1. **Setting CSV files to their paths and converting them into data frames**

   - Load carbon emissions dataset
   - Load reductions dataset
   - Create backup copies of original dataframes

2. **Setting proper data types**

   - Convert 'CO2 emission estimates' to category (Country)
   - Convert 'Years' to int64 (Year)
   - Convert 'Series' to category (Emission_Type)
   - Convert 'Value' to numeric (CO2_Value)
   - Convert reduction dataset columns to appropriate types

3. **Renaming columns for clarity**

   - 'CO2 emission estimates' → 'Country'
   - 'Years' → 'Year'
   - 'Series' → 'Emission_Type'
   - 'Value' → 'CO2_Value'
   - 'Type' → 'Industry Sector'
   - 'Reduced Carbon(Gt)' → 'Reduced Emissions (Gt)'
   - 'How' → 'Summary'

4. **Data quality checks**

   - Check dataframe info and head for content verification
   - Check for null values in all columns
   - Check for duplicate rows
   - Analyze unique values in key columns

5. **Data correction and standardization**

   - Correct irregularities in country names:

     - 'Iran (Islamic Republic of)' → 'Iran'
     - 'Bolivia (Plurin. State of)' → 'Bolivia'
     - "Dem. People's Rep. Korea" → 'North Korea'
     - 'Venezuela (Boliv. Rep. of)' → 'Venezuela'
     - 'United States of America' → 'United States'
     - 'United Rep. of Tanzania' → 'Tanzania'
     - 'Dem. Rep. of the Congo' → 'Congo - Kinshasa'
     - 'Syrian Arab Republic' → 'Syria'
     - 'Viet Nam' → 'Vietnam'

6. **Data export and organization**

   - Export original dataframes to 'Prepared Data/Original/'
   - Export modified dataframes to 'Prepared Data/Modified/'
   - Use Python OS module for directory creation and file management

Phase 3: Power BI Visualization Steps

1. **Data loading and final cleaning**

   - Load prepared CSV files into Power BI
   - Use Power Query to recheck and correct any remaining irregularities

2. **Dashboard creation**

   - Create line charts for average carbon dioxide emissions by year
   - Create clustered column charts for emissions per country
   - Add interactive slicer for Emissions vs Emissions per capita
   - Create donut chart for carbon reductions across 6 industry sectors
   - Include comprehensive data summary
   - Add project scope and extra insights

📊 Key Datasets Description

Carbon Dioxide Emission Estimates Dataset

- **Country**: Geographic location of emissions
- **Year**: Time period of measurement (1990-2017)
- **Emission_Type**: Category of CO2 emissions (Total, Per capita)
- **CO2_Value**: Quantitative emission measurements (in various units)

Reductions Dataset

- **Industry Sector**: Industry or sector category (Energy, Agriculture, etc.)
- **Reduced Emissions (Gt)**: Amount of carbon reduction achieved (in Gigatons)
- **Summary**: Methods and strategies for reduction

📊 Data Analysis Results

Data Quality Assessment

- **Null Values**: Checked across all columns - minimal null values found
- **Duplicates**: No duplicate rows identified in the datasets
- **Unique Values**: Analyzed country names, years, and emission types for consistency
- **Data Types**: Properly converted for efficient processing and analysis

Data Transformation Summary

- **Original Records**: Carbon emissions dataset contained raw data
- **Modified Records**: Standardized country names and column headers
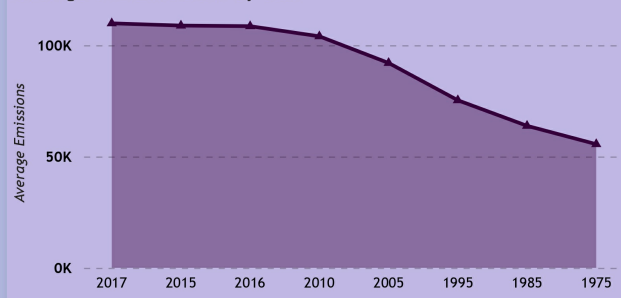- **Export Success**: All prepared datasets successfully exported to organized directories

🎨 Power BI Dashboard Visualization

The interactive dashboard provides comprehensive insights into global carbon emissions:

Global Carbon Emissions & Reductions Through Different Industries

Average CO2 Emissions by Year

(Line/area chart: Average Emissions vs Year — 2017, 2015, 2016, 2010, 2005, 1995, 1985, 1975; y-axis 0K, 50K, 100K)

Types of Emission	
All	⌄

109,835.76 Average of CO2_Value	2017 Year
108,811.09 Average of CO2_Value	2015 Year
108,549.16 Average of CO2_Value	2016 Year
103,990.98 Average of CO2_Value	2010 Year

Average CO2 Emissions by Country

(Column chart: Average Emissions by Country — China, United..., Russian..., India, Japan, Germany, Canada, United..., Republi..., Iran, Italy, Mexico, France, South..., Saudi A..., Poland, Australia, Brazil, Indone..., Ukraine, Spain; y-axis 0M, 1M, 2M)

CO2 Emissions Reductions Across 6 Sectors

(Donut chart)
- Energy 8 (22.28%)
- Agriculture, Food & Waste 7 (18.21%)
- Buildings & Cities 6 (16.03%)
- Natural-Based Solutions 6 (16.03%)
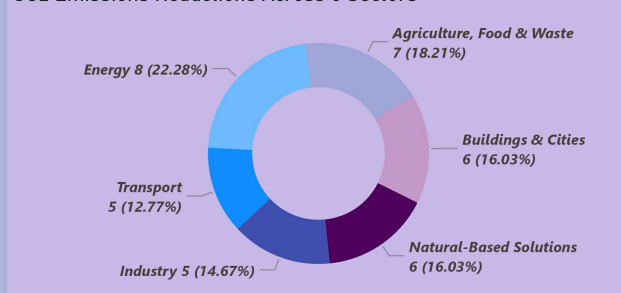- Industry 5 (14.67%)
- Transport 5 (12.77%)

- Highest record of carbon emission per capital: 5.3 mt (in 2005)
- USA: high total carbon emission and emission per capital (top 10)
- Changes in Energy sector can reduce the most carbon emission: 8.2 gt
- Highest record of total carbon emission: 219666.45 mt in 2017

Dashboard Features:

- **Time Series Analysis**: Line charts showing emission trends from 1990-2017
- **Geographic Comparison**: Column charts comparing emissions across countries
- **Per Capita Metrics**: Interactive slicer to toggle between total and per capita emissions
- **Industry Reduction Potential**: Donut chart showing reduction opportunities by sector
- **Dynamic Filtering**: Interactive elements for customized analysis

📝 Key Findings and Insights

After comprehensive data analysis and Power BI visualization, the following key insights were derived:

Emission Trends Over Time

- Global carbon dioxide emissions have shown a steady increase from 1990 to 2017
- Peak emissions observed around 2005-2010, followed by gradual fluctuations
- Recent years show slight stabilization in emission growth

Country-wise Emissions Analysis

- Major economies like China, the United States, and India account for highest absolute CO2 emissions
- Per capita analysis reveals varying patterns across nations

- Some developed nations show higher emissions per person despite lower total volumes

Industry Reduction Potential

- Energy sector (electricity and heat production) offers largest opportunities for carbon reductions
- Agriculture, forestry, and land use sectors also significant for emission reductions
- Approximately 60% of potential reductions achievable across key industries

Dynamic Dashboard Insights

- Interactive slicer allows toggling between total emissions and per capita metrics
- Reveals that while some countries lead in absolute emissions, others have higher intensity per capita
- Guides targeted policy interventions and reduction strategies

🚀 Getting Started

Prerequisites

```
pip install pandas numpy jupyter
```

Running the Analysis

1. Clone or download this repository
2. Open `Carbon.ipynb` in Jupyter Notebook
3. Run all cells sequentially to perform data cleaning
4. Import prepared CSV files from `Prepared Data/Modified/` into Power BI
5. Open `Dashboard.pbix` for interactive visualization

Data Processing Steps (Code Snippets)

```python
# Load datasets
carbon_emissions_path = "Emissions Dataset/Carbon Dioxide Emission Estimates.csv"
carbon_df = pd.read_csv(carbon_emissions_path)

# Set data types
carbon_df["CO2 emission estimates"] = carbon_df["CO2 emission estimates"].astype("category")
carbon_df["Years"] = carbon_df["Years"].astype("int64")

# Rename columns
carbon_df.rename(columns={
    'CO2 emission estimates': 'Country',
    'Years': 'Year',
    'Series': 'Emission_Type',
    'Value': 'CO2_Value'
}, inplace=True)

# Check for nulls and duplicates
null_counts, total_nulls = check_null(carbon_df)
duplicate_count = carbon_df.duplicated().sum()

# Export prepared data
carbon_df.to_csv("Prepared Data/Modified/Carbon_Dioxide_Emission_Estimates_Modified.csv", index=False)
```

🛠️ Future Enhancements

- [ ] Incorporate additional emission sources (methane, nitrous oxide)
- [ ] Add predictive modeling for future emission trends
- [ ] Expand reduction dataset with more industry sectors
- [ ] Integrate real-time data sources
- [ ] Develop automated reporting pipeline
- [ ] Add machine learning models for emission forecasting

👤 Author

Data Analysis & Visualization Project

🗓️ Project Timeline

- **Data Collection**: Completed
- **Data Cleaning & Processing**: Completed using Python/Pandas
- **Exploratory Analysis**: Completed
- **Visualization**: Completed using Power BI
- **Documentation**: Completed
- **Last Updated**: December 2025

Note: This comprehensive documentation demonstrates end-to-end data analysis skills including data cleaning, exploratory analysis, visualization, and preparation for business intelligence reporting. The project showcases the complete workflow from raw data to actionable insights through interactive dashboards.