Multiple Grid Energy Consumption of Data Centers

Course: Data Integration
Philipps University Marburg

Team members:
Indrit Berbiu
Amra Dadic

Agenda

Data Cleaning

- Data filtering
- Data preprocessing

Showcase

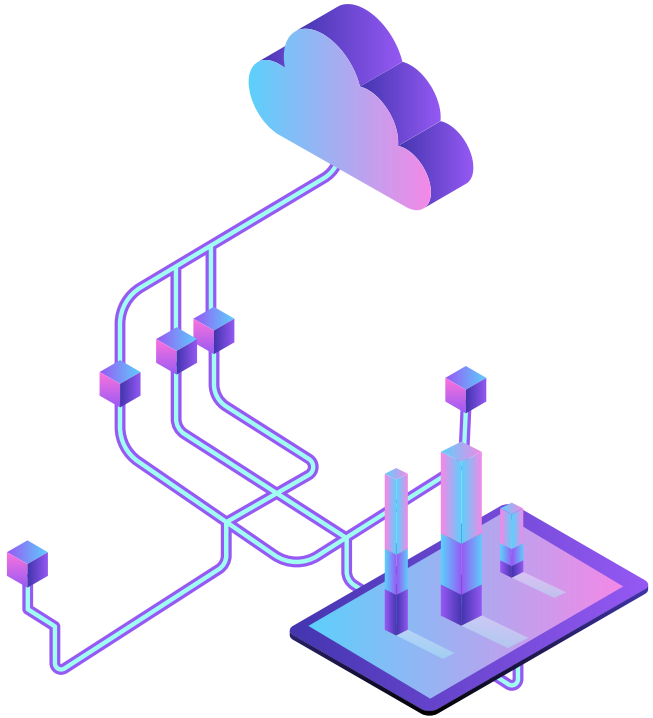
- Analysis of: Energy production, consumption in US
- Analysis of Solar Farm energy production
- Analysis of Data Center Energy consumption

Data Cleaning

- Data cleaning was performed also in the previous part
- Now it has been extended
- Datasets/tables that needed to be cleaned:
  - Location
  - Energy production

Data Cleaning: Energy Production

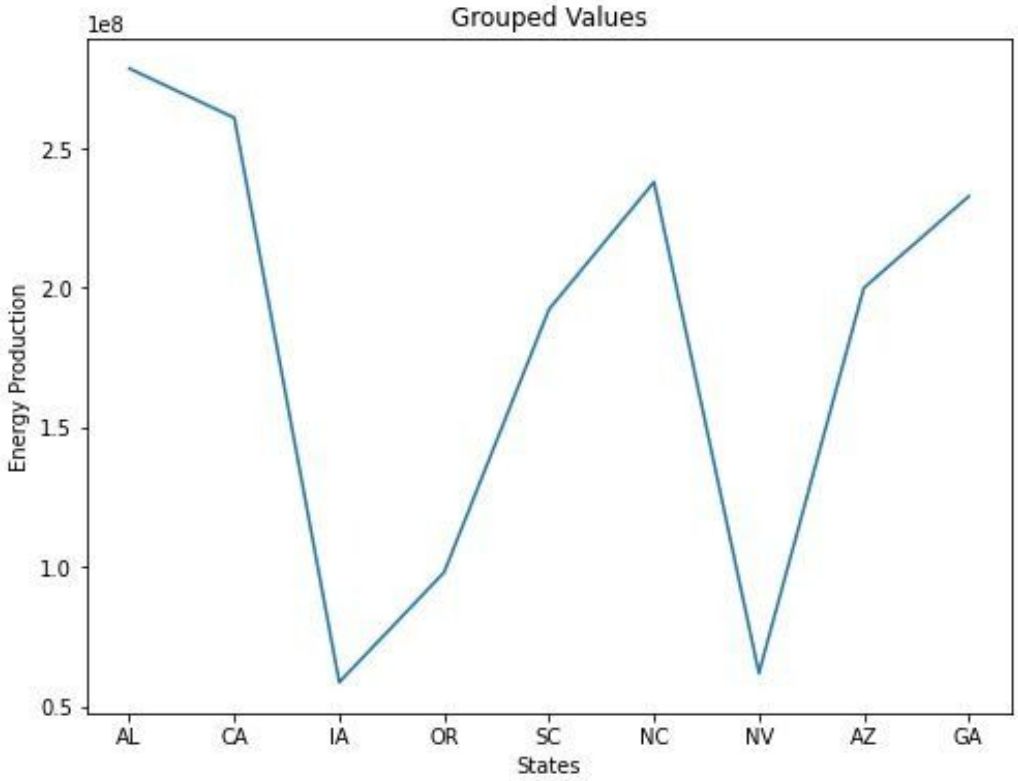
- Original dataset had around 50 000 rows
- It had all data between 2003 and 2022 year for energy production for each month, for each state
- We extracted data:
  - Year 2021
  - Certain types of fuels
  - Summed up fuel subtypes into one type
  - US states that we needed
- At the end: 54 rows



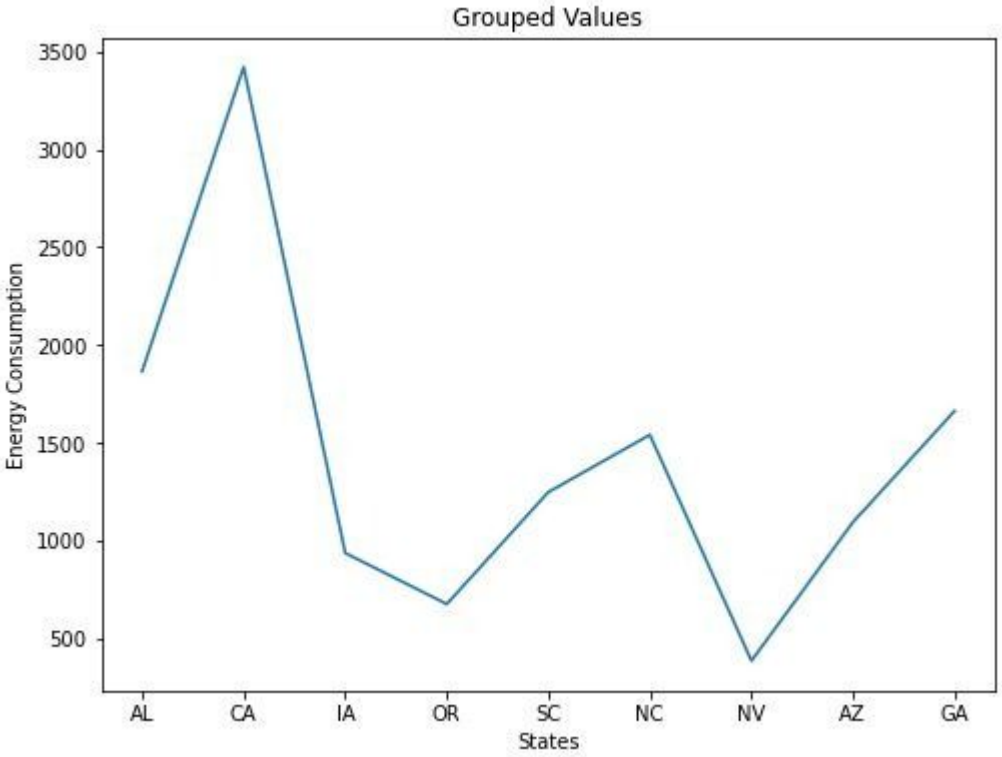
Data Cleaning: Location

- Dataset was created by us
- To simplify the problem we dropped:
  - Geo Coordinates
  - Precise Location name
- Data is connected by states
  - Explanation: if energy is produced in one state - it can be transported to whole state
  - Geo Coordinates can be used in future work to calculate the distances

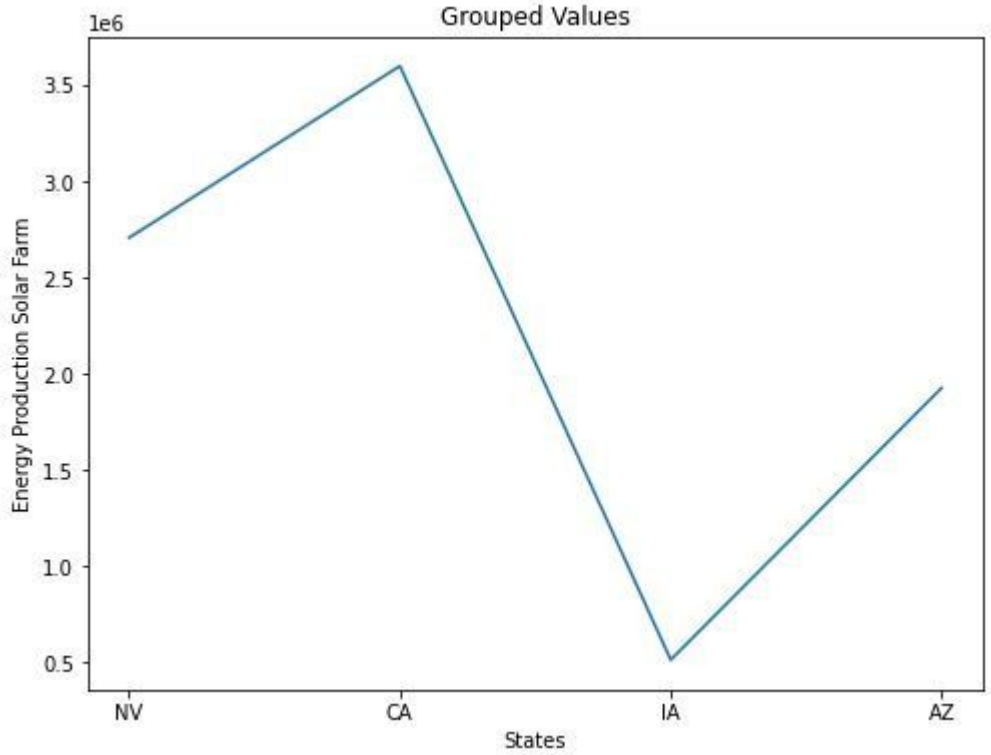
Data analysis: Energy production in US



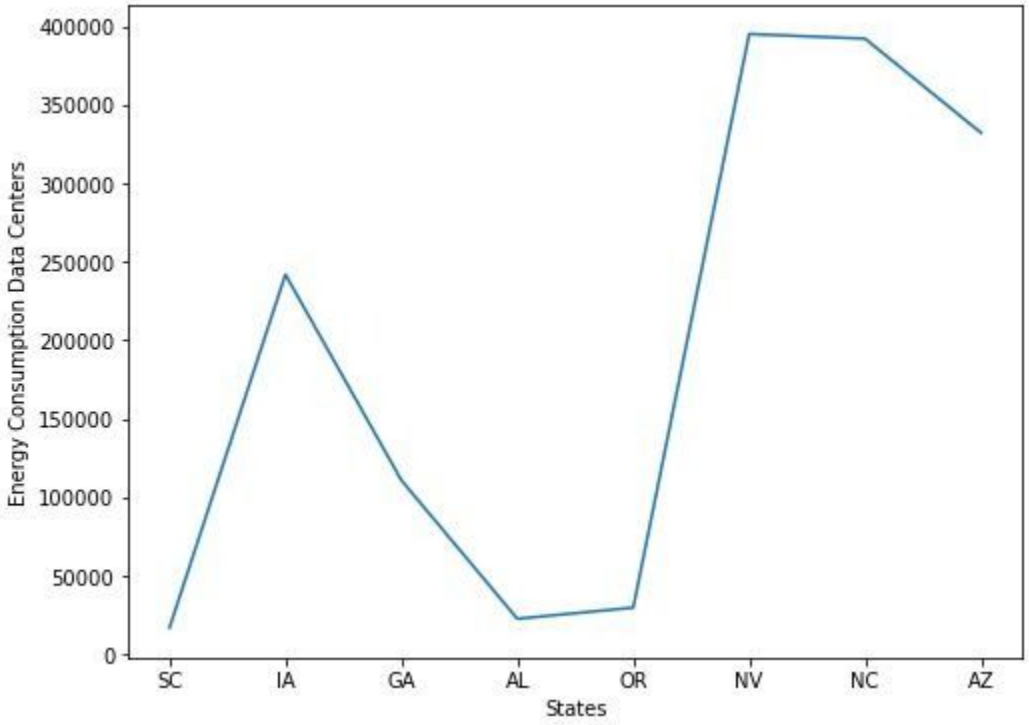
Data analysis: Energy consumption in US



Data analysis: Energy production by solar farms in US



Data analysis: Energy consumption by Data Centers



Thank you for your attention!