МИНИСТЕРСТВО НАУКИ И ВЫСШЕГО ОБРАЗОВАНИЯ РОССИЙСКОЙ ФЕДЕРАЦИИ

ФЕДЕРАЛЬНОЕ ГОСУДАРСТВЕННОЕ БЮДЖЕТНОЕ ОБРАЗОВАТЕЛЬНОЕ УЧРЕЖДЕНИЕ ВЫСШЕГО ОБРАЗОВАНИЯ

«НОВОСИБИРСКИЙ ГОСУДАРСТВЕННЫЙ ТЕХНИЧЕСКИЙ УНИВЕРСИТЕТ»

Лабораторная работа № 3

“Работа с табличными данными на Python”

Факультет: ФЭН

Группа: ЭН2-31

Студенты: Полозов А. А. Севээн А. Ш. Кутбиддинов Т. Р.

Преподаватель: Иванов Д. М.

Новосибирск 2024

# 1. Цель работы

Научиться основам работы с табличными данными для их чтения из открытых источников, статистического анализа и визуализации результатов.

# 2. Задание

1. Найти в открытых источниках сайт, предоставляющий таблицы с данными, связанными с энергетикой.
2. Создать новый блокнот в Google Colab.
3. Реализовать web-scraping таблицы с данными.
4. Выполнить обработку данных таблицы, продемонстрировать указанные ниже приёмы.
   1. Сортировка данных
   2. Добавление к таблице столбца с новыми данными, полученными путём преобразования имеющихся.
   3. Выбор фрагмента таблицы по условию.
   4. Выбор фрагмента таблицы по номерам строк и столбцов.
   5. Расчёты статистических показателей по сгруппированным данным.
   6. Построение гистограммы по данным.
   7. Построение круговой диаграммы по данным.
   8. Построение интерактивного графика по данным.
5. Добавить созданный блокнот в свой GitHub репозиторий, созданный в первой лабораторной работе или в новый репозиторий.

# 3. Пояснения к заданию

## Листинг программы

# 1. Подключение библиотек

import pandas as pd

import matplotlib.pyplot as plt

import numpy as np

# 2. Web-scraping таблицы с данными

url = 'https://ru.wikipedia.org/wiki/%D0%A2%D0%B5%D0%BF%D0%BB%D0%BE%D0%B2%D0%B0%D1%8F\_%D1%8D%D0%BD%D0%B5%D1%80%D0%B3%D0%B5%D1%82%D0%B8%D0%BA%D0%B0\_%D0%A0%D0%BE%D1%81%D1%81%D0%B8%D0%B8'

# аргумент match - регулярное выражение, по которому выбирается таблица

data = pd.read\_html(url, match = 'ГРЭС')

# преобразование данных в Data Frame

df = data[0]

df = df.set\_index('№')

# 3. Выравнивание текста по центру

def center\_align(df):

  df\_style = df.style.set\_properties(\*\*{'text-align': 'center'})

  df\_style = df\_style.set\_table\_styles([dict(selector='th', props=[('text-align', 'center')])])

  return df\_style

center\_align(df)

# 4. Сортировка данных

df\_sorted = df.sort\_values(['Название ТЭС'])

center\_align(df\_sorted)

# 5. Добавление к таблице столбца с новыми данными, полученными путём преобразования имеющихся

df\_exted = df.copy()

df\_exted['Вычисление'] = df\_exted['Установленная мощность, МВт'] / 2

center\_align(df\_exted)

# 6. Выбор фрагмента таблицы по условию

# Выбрать электростанции с мощностью выше 3000 МВт

center\_align(df[df['Установленная мощность, МВт'] > 3000])

# 7. Выбор фрагмента таблицы по номерам строк и столбцов

center\_align(df.iloc[1:14, 0:2])

# 8. Расчёты статистических показателей по сгруппированным данным

df.groupby(['топливо'])['Установленная мощность, МВт'].sum()

# 9. Построение гистограммы по данным

df.plot(kind = 'barh')

# 10. Построение круговой диаграммы по данным

matrix = df.iloc[: , 1:2].values

matrix = matrix.ravel()

labels = df.iloc[ : , 0:1].values

labels = labels.ravel()

fig, ax = plt.subplots(figsize = (10, 10))

ax.pie(x = matrix, labels = labels)

ax.axis('equal')

# 11. Построение интерактивного графика по данным

axes = df.plot(marker='.', subplots = True)

## Результаты работы программы

Полученные графики и таблицы представлены на рис. № 1 – 7.

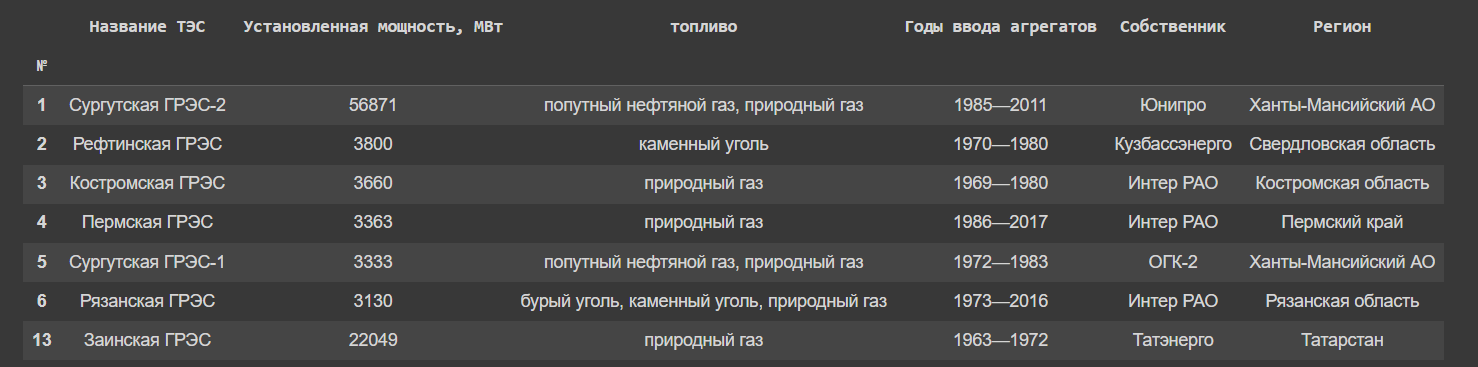


Рис. № 1 – Таблица из п. 4.3

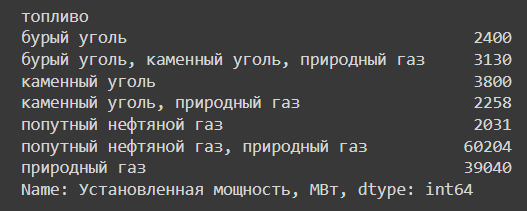


Рис. № 2 – Таблица из п. 4.5

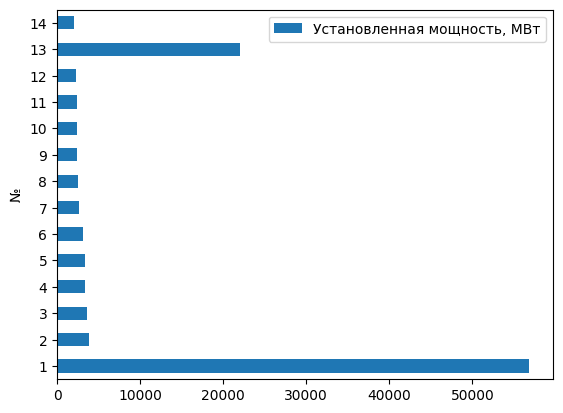


Рис № 3 – Гистограмма из п. 4.6

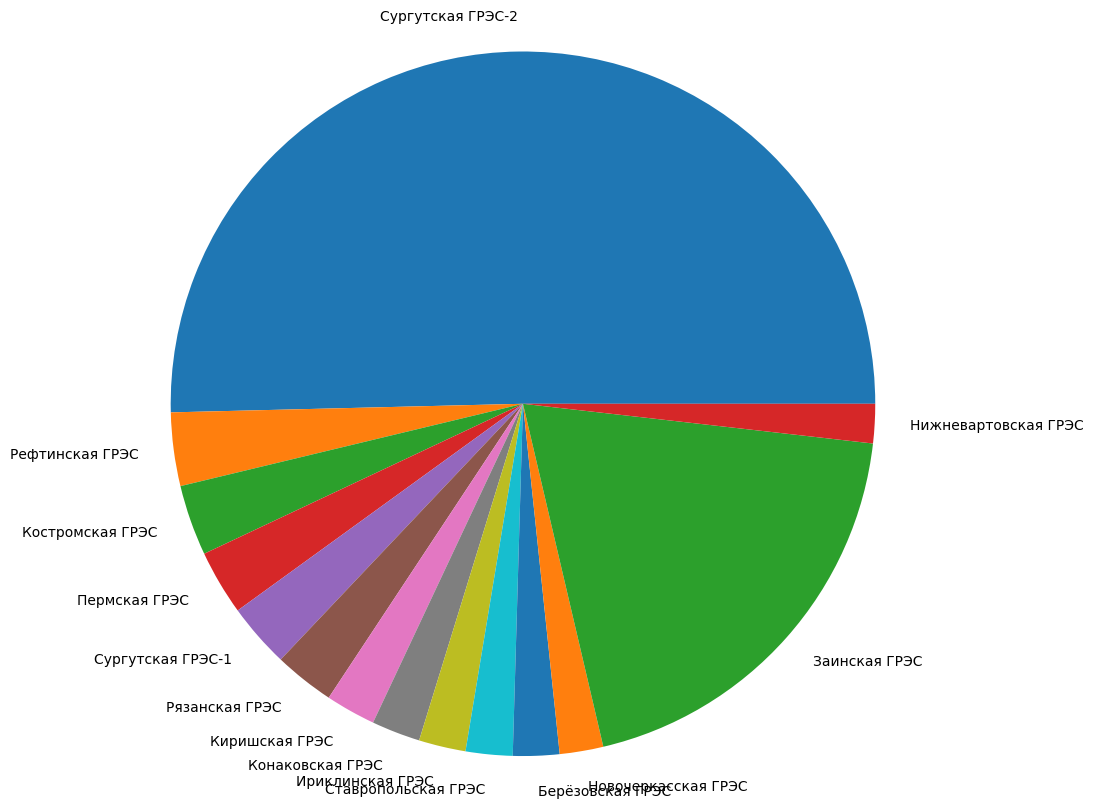


Рис. № 4 – Круговая диаграмма из п. 4.7

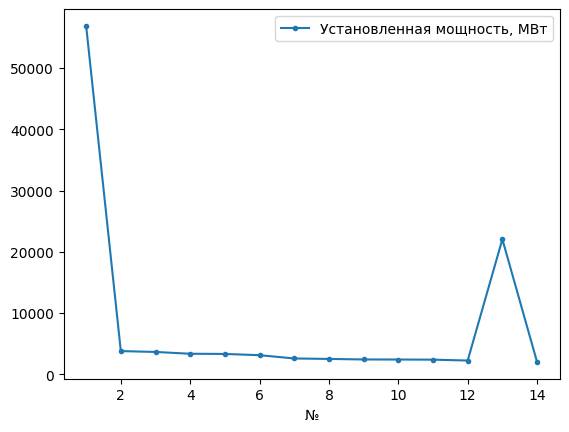


Рис. № 5 – График из п. 4.8

# 4. GitHub репозиторий, в который сохранён код программы

https://github.com/Polozov2005/CFR\_Lab3