PRESENTED BY
Teodora Ćurčić

Python without the **pain**

WRITE CODE WITH LLMS

case study

PARTY FUNDS

PARTY FUNDS

Party Funds is a database on the financing of political parties in Serbia. It includes **all reported revenues and expenditures** of 40 parties and citizen groups that have had representatives at any level of the Assembly (national, regional, or local) over the past nine years.

PARTY FUNDS

- **40** entities
- **15.5 billion RSD** from the budget
- **5.8 billion RSD** spent on advertising
- **9** years of data

STARTING POINT

Around 150 xlsx files containing data on incomes and expenses across 14 categories, complemented by the first edition of the database featuring over 15k entries.

FINAL PRODUCT

One xlsx file comprising over 30k entries detailing incomes and expenses for 40 different parties over a span of 9 years.

TECHNICAL PART

- merge files
- clean
- analyse

AI TO THE RESCUE

simplifies tasks without requiring extensive coding knowledge

CHECKING

- checking totals
- summary
- looking for mistakes

- Give inputs, write like you would to a colleague
- Share as many details as possible
- Include examples where necessary
- Break down tasks into clear, actionable steps
- You can ask before writing a prompt

TASK EXAMPLES

- Merge files
- Compare two different files
- Summarize based on a specific request
- Look for spelling mistakes or similarities (Pink and Pink media group)
- Convert exchange rates

i have a folder with over 100 xlsx files, each has 14 sheets, give me a script that will merge them while preserving their structure, also add a column with source info, so i know which entry originated from which file

i have a folder with over 100 xlsx files, each has 14 sheets, give me a script that will merge them while preserving their structure, also add a column with source info, so i know which entry originated from which file

Here's a script that merges multiple `.xlsx` files from a folder while preserving the sheet structure. Additionally, it adds a column in each sheet containing the filename as the source information.

Script Explanation:

1. **Libraries**: Uses `pandas` for data manipulation and `openpyxl` for Excel file handling.
2. **File Handling**: Reads all `.xlsx` files in a specified folder.
3. **Sheet Handling**: Iterates over all sheets in each file.
4. **Data Tracking**: Adds a new column, `Source`, to indicate the originating file.
5. **Output**: Writes the combined sheets back to a new Excel file.

i have a xlsx file with political party expenses and incomes, data is in these sheets (each sheet is another category name)

incomes:
'Prihodi iz budzeta'
'Prihodi od firmi'
'Clanarine i ostali prihodi'

expenses:
'Rezijski troskovi'
'Troskovi oglasavanja'
'Reklamni materijal i dogadjaji'
'Plate i ostali troskovi'

data i want to analyse is in these columns
"Stranka" for the party name
"Godina" for the year
"Iznos" for the amount

give me a script that will analyse this:
1. sum of all incomes per year per party
2. sum of all expenses per year per party

i have a xlsx file with political party expenses and incomes, data is in these sheets (each sheet is another category name)

incomes:
'Prihodi iz budzeta'
'Prihodi od firmi'
'Clanarine i ostali prihodi'

expenses:
'Rezijski troskovi'
'Troskovi oglasavanja'
'Reklamni materijal i dogadjaji'
'Plate i ostali troskovi'

data i want to analyse is in these columns
"Stranka" for the party name
"Godina" for the year
"Iznos" for the amount

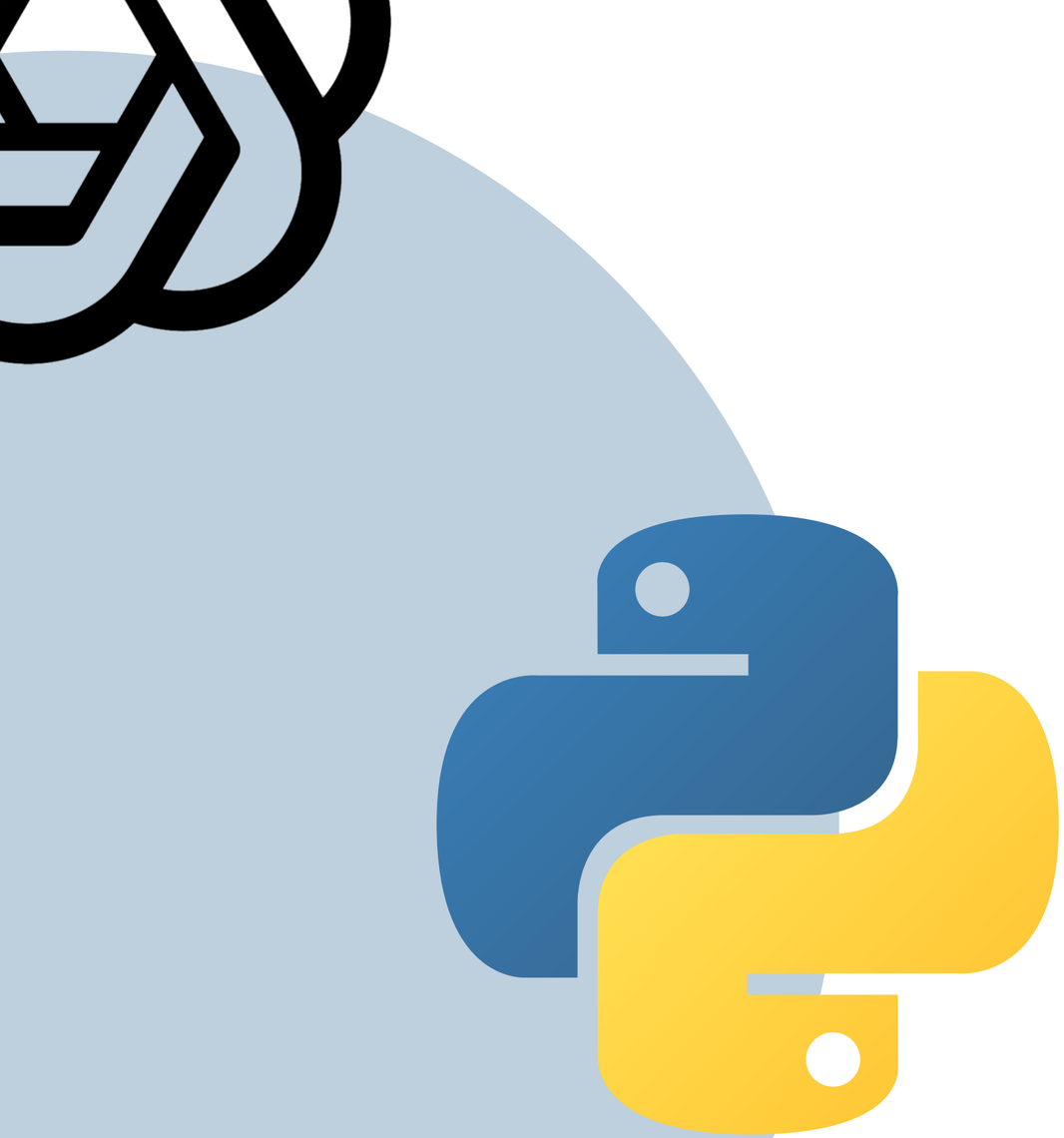
give me a script that will analyse this:
1. sum of all incomes per year per party
2. sum of all expenses per year per party

also, these sums are in serbian dinars, i need you to transfer it to euros and add a new column with euro value next to RSD value, here is the exchange rate for each year

2023: 117,2513
2022: 117,4588
2021: 117,5733
2020: 117,5777
2019: 117,8524
2018: 118,2716
2017: 121,3367
2016: 123,1179
2015 120,7328

i have a xlsx file with political party incomes and expenses spread across 7 sheets (for 7 categories) and subcategories, also two different purposes (general spending and elections) and i want to analyse the data, what information do you need from me to give me appropriate script

- Excecute the script in Python (or Colab)
- If any errors occur, copy and paste them into ChatGPT and see why
- Review and resolve issues as needed
- Check



Here is the script:

python

Copy code

```python
import os
import pandas as pd

def merge_excel_files(folder_path, output_file):
    # List to hold the combined data for each sheet
    merged_sheets = {}

    # Iterate over all files in the folder
    for filename in os.listdir(folder_path):
        if filename.endswith('.xlsx'):
            file_path = os.path.join(folder_path, filename)

            # Load the Excel file
            xls = pd.ExcelFile(file_path)

            # Iterate through each sheet in the Excel file
            for sheet_name in xls.sheet_names:
                # Read the sheet into a DataFrame
                df = pd.read_excel(file_path, sheet_name=sheet_name)

                # Add a column for source filename
                df['Source'] = filename

                # Append to the respective sheet in the merged_sheets
                if sheet_name not in merged_sheets:
                    merged_sheets[sheet_name] = [df]  # Start a new li
                else:
                    merged_sheets[sheet_name].append(df)  # Append to

    # Write the combined sheets into a single Excel file
    with pd.ExcelWriter(output_file, engine='openpyxl') as writer:
        for sheet_name, sheet_data in merged_sheets.items():
            # Concatenate all DataFrames for the same sheet
            combined_df = pd.concat(sheet_data, ignore_index=True)
```

- Data is prepared for further analysis by a journalist
- Data is also ready to be used for visualizations

	A	B	C	D	E
1	Stranka	Godina	Iznos	EUR	Type
2	Srpska napredna stranka (SNS)	2022			Income
3	Srpska napredna stranka (SNS)	2022			Expense
4	Srpska napredna stranka (SNS)	2023			Expense
5	Srpska napredna stranka (SNS)	2017			Income
6	Srpska napredna stranka (SNS)	2020			Income
7	Srpska napredna stranka (SNS)	2016			Expense
8	Srpska napredna stranka (SNS)	2021			Income
9	Srpska napredna stranka (SNS)	2017			Expense
10	Srpska napredna stranka (SNS)	2020			Expense
11	Srpska napredna stranka (SNS)	2023			Income
12	Srpska napredna stranka (SNS)	2016			Income
13	Srpska napredna stranka (SNS)	2019			Income
14	Srpska napredna stranka (SNS)	2018			Expense
15	Srpska napredna stranka (SNS)	2021			Expense
16	Srpska napredna stranka (SNS)	2018			Income
17	Socijalistička partija Srbije (SPS)	2016			Expense
18	Socijalistička partija Srbije (SPS)	2023			Expense
19	Socijalistička partija Srbije (SPS)	2022			Income
20	Socijalistička partija Srbije (SPS)	2022			Expense
21	Socijalistička partija Srbije (SPS)	2023			Income
22	Socijalistička partija Srbije (SPS)	2020			Expense
23	Srpska napredna stranka (SNS)	2015			Income
24	Socijalistička partija Srbije (SPS)	2020			Income
25	Socijalistička partija Srbije (SPS)	2019			Income
26	Stranka slobode i pravde (SSP)	2023			Expense

STRANAČKA KASA

Baza podataka Centra za istraživačko novinarstvo Srbije (CINS) o finansiranju stranaka i grupa građana u Srbiji

Srpska napredna stranka (SNS)

Socijalistička partija Srbije (SPS)

Srpska radikalna stranka (SRS)

Stranka slobode i pravde (SSP)

Demokratska stranka (DS)

Prikaži ostatak

Pretraži bazu

REMEMBER

- **safety**
- that's why it is better to use Jupyter
- ask for a script, not an analysis
- ask for improvement ideas
- ask about the mistakes
- be as specific as possible: source, categories, operations, output, fonts, colour codes, specific images, text size, map layer, etc.
- experiment, but keep a **copy** of your file
- **don't limit yourself to simple tasks**

Q&A