OPENPYXL FOR DATA ANALYSTS – COMPLETE PRACTICAL GUIDE

This guide explains what openpyxl is, why a Data Analyst needs it, and how to use it correctly for Excel automation and reporting.

1. What is openpyxl?

openpyxl is a Python library used to read, write, and modify Excel (.xlsx) files. Data Analysts use it mainly to format Excel outputs generated from pandas.

2. Why pandas alone is not enough

Pandas handles data manipulation well, but it cannot properly format Excel reports. openpyxl is used for formatting, freezing headers, column widths, and presentation.

3. Excel structure

Workbook → Worksheet → Cell. Workbook is the Excel file, Worksheet is a sheet, and Cell is a single value like A1.

4. Creating and loading workbooks

Create workbook: wb = Workbook()

Load workbook: wb = load_workbook('report.xlsx', data_only=True)

Save workbook: wb.save('final_report.xlsx')

5. Working with worksheets

Active sheet: ws = wb.active

Sheet names: wb.sheetnames

Access sheet: ws = wb['Sales']

Create sheet: wb.create_sheet('Summary', index=0)

Rename sheet: ws.title = 'Cleaned_Data'

Delete sheet: del wb['Sheet']

6. Working with cells

Read cell: ws['A1'].value

Write cell: ws['A1'] = 'Total Sales'

Row-column access: ws.cell(row=1, column=1).value

7. Iterating data

for row in ws.iter_rows(min_row=2, values_only=True): print(row)

8. Writing pandas DataFrame to Excel

with pd.ExcelWriter('report.xlsx', engine='openpyxl') as writer: df.to_excel(writer, sheet_name='Data', index=False)

9. Formatting (core analyst usage)

Font: Font(bold=True, size=12)

Fill: PatternFill(fill_type='solid', start_color='4F81BD')

Alignment: Alignment(horizontal='center')

Borders: Border(Side(style='thin'))

10. Rows, columns, filters

Column width: ws.column_dimensions['A'].width = 20

Freeze header: ws.freeze_panes = 'A2'

Filters: ws.auto_filter.ref = ws.dimensions

Conclusion

If you can use everything in this guide, you have mastered openpyxl for Data Analyst work.