

Instructor Guide: Campus Energy Audit — Analysis with Python in Excel

CONFIDENTIAL — FOR INSTRUCTOR AND TA USE ONLY

Assignment Overview

This assignment teaches students to summarize and visualize a campus utility dataset using a progression of Excel skills: basic aggregation functions (SUM, AVERAGE), conditional aggregation (SUMIF), PivotTables, native charting, and Python in Excel for DataFrame manipulation, correlation analysis, and matplotlib visualization. By the end, students should be able to cross-reference data across sheets using absolute references, build an interactive dashboard from a PivotTable, interpret Pearson correlations in an energy-management context, and produce publication-quality plots directly inside Excel using pandas and matplotlib. The assignment also reinforces worksheet hygiene — correct sheet names, correct target cells, and formula-driven (not hard-coded) outputs.

Partial Credit Policy

General principle: award partial credit whenever a student demonstrates understanding of the correct technique even if execution is incomplete. A correct formula in the wrong cell earns at most 50% of the task points — the technique is demonstrated but the placement instruction was not followed. A correct numeric value typed as a hard-coded constant (no formula) earns at most 40% of the task points for formula tasks (1–4) and 0% for Python tasks (7–9) where the entire learning objective is the code. For multi-step tasks (3, 5, 6, 9), award credit proportionally: if a task has four discernible sub-steps and two are correct, award approximately 50% unless the instructions specify otherwise. Never award negative points for an attempt. For Python tasks, if the code is syntactically correct and shows clear intent but fails due to a single misspelled column name, award no less than 60% of the task points and provide specific feedback on the column name error. Document all partial credit decisions with a brief note so students understand exactly what was correct and what needs improvement.

General Grading Tips

- Always open the student's workbook before running any automated checker — some checks (PivotTable field configuration, chart series count, Python cell output) require visual inspection that no formula-scanning script can fully replicate.
- Check sheet tab names carefully: 'Readings', 'Analysis', and 'Dashboard' must be spelled exactly (case-insensitive in Excel but enforce the exact names stated in the assignment); a sheet named 'Analyis' or 'analysis' may cause cross-sheet formula references to break.
- For Python in Excel tasks (7, 8, 9), verify that the formula bar shows =PY(...) and not just a static value — students who copy a displayed result and paste as a value will appear to have the right answer but have not used the required technique.

- When grading Tasks 3 and 9, compare the student's building-level totals against your own reference solution to catch calculation errors that a cell-formula type check would miss.
 - Remind TAs that in Excel's charting vocabulary, 'Bar' is horizontal and 'Column' is vertical — this is a perennial source of confusion and should be checked explicitly for Task 6.
 - For Task 5 (PivotTable), right-click any value cell in the PivotTable and choose 'Value Field Settings' to verify Sum vs Count — do not rely solely on the displayed numbers looking plausible.
 - Python in Excel requires a Microsoft 365 subscription with the Python feature enabled; confirm all students have access before the assignment is due and have a contingency plan (e.g., alternative pandas-based task) for students who cannot access the feature.
 - Keep a reference workbook with all nine tasks completed; open it alongside each student submission to do rapid visual comparisons of expected vs actual chart appearance and DataFrame output.
-

Task-by-Task Guide

Task 1

Common Student Mistakes

- Typing the numeric total directly into B2 instead of writing a SUM formula.
- Referencing the wrong column (e.g., Readings!G2:G121 for GasTherms instead of F2:F121 for ElectricityKWh).
- Writing the formula on the Readings sheet instead of the Analysis sheet.
- Including the header row in the range (F1:F121) which adds a text cell and may cause a #VALUE! error or silently ignore the header.
- Using =SUM(F2:F121) without the sheet reference, which sums the Analysis sheet's column F (empty) rather than the Readings data.

Socratic Questions (When Students Are Stuck)

- If a student is stuck, ask: Where exactly in Excel does the task tell you to place the result, and which sheet are you currently on?
- If a student is stuck, ask: How does Excel know to look at data on a different sheet — what syntax does it use to reference another sheet's cells?
- If a student is stuck, ask: Look at the column headers in row 1 of Readings — which column letter holds ElectricityKWh?

Solution Walkthrough

1. Click the 'Analysis' sheet tab at the bottom. 2. Click cell B2. 3. Type: =SUM(Readings!F2:F121) and press Enter. 4. Verify the result is a large positive number (total kWh across all 120 rows). The formula navigates to the Readings sheet via the sheet-prefix syntax (SheetName! followed by the cell range). F is the ElectricityKWh column; row 2 starts after the header; row 121 is the last data row (120 rows of data + 1 header row = row 121).

Grading Tips

Award full 5 points only if a SUM formula is present in B2 on the Analysis sheet and references Readings!F2:F121 (or an equivalent full-column or named-range reference that captures exactly the

120 data rows). Award 3/5 if the formula uses SUM but references the wrong range (e.g., off-by-one rows or wrong column) yet produces a plausible number. Award 2/5 if the correct numeric value appears in B2 but was typed rather than calculated — note 'hard-coded value, no formula' in feedback. Award 0/5 if the formula is placed in the wrong cell or on the wrong sheet, even if the formula itself is correct; redirect the student to resubmit in the correct location.

Discussion Prompt

If a new building were added and its readings appended to rows 122 onward, would this formula automatically include it? What change to the formula or dataset structure would make the summary self-updating?

Edge Cases

- Student uses `=SUM(Readings!F:F)` — includes the header text cell but Excel's SUM ignores text, so the numeric result is correct; accept for full credit but note the less precise range.
- Student uses a named range or Table reference that resolves to the same 120 rows — accept for full credit.
- Student's formula returns 0 because they are on the wrong sheet or referenced an empty column — verify which sheet is active before penalizing.

Task 2

Common Student Mistakes

- Using SUM instead of AVERAGE, producing the total cost rather than the mean.
- Placing the formula in B2 (overwriting Task 1) instead of B3.
- Referencing the wrong column (e.g., ElectricityKWh column F instead of TotalCost column J).
- Hard-coding the average value typed as a number.
- Dividing SUM by 120 manually (`=SUM(...)/120`) rather than using AVERAGE — this works numerically but does not use the required function.

Socratic Questions (When Students Are Stuck)

- If a student is stuck, ask: What Excel function is specifically designed to compute an arithmetic mean, and how does its syntax differ from SUM?
- If a student is stuck, ask: Which column letter in the Readings sheet holds TotalCost — have you checked the header row?
- If a student is stuck, ask: The task says cell B3 — are you sure you haven't placed the formula one row too high or too low?

Solution Walkthrough

1. Click the 'Analysis' sheet tab. 2. Click cell B3. 3. Type: `=AVERAGE(Readings!J2:J121)` and press Enter. 4. Confirm the result is a reasonable per-reading average cost (should be much smaller than the SUM in B2). Column J is TotalCost; the range J2:J121 covers all 120 data rows. AVERAGE sums the values and divides by the count of numeric cells automatically.

Grading Tips

Award full 5 points if AVERAGE is used in cell B3 on the Analysis sheet referencing Readings!J2:J121 (or equivalent). Award 3/5 if the formula is in B3 but uses the wrong column yet the AVERAGE function is correctly applied. Award 2/5 for a correct hard-coded value in B3 with no formula. Award 1/5 if =SUM(...)/120 produces the correct value — the result is right but the required function was not used; leave feedback. Award 0/5 if placed in the wrong cell.

Discussion Prompt

The AVERAGE here mixes readings from buildings of very different sizes. What does this number actually represent, and what might be a more meaningful measure of 'typical' monthly cost — mean, median, or a size-normalized metric?

Edge Cases

- Student uses =AVERAGE(Readings!J:J) including the header — AVERAGE ignores text cells, so the result is numerically correct; accept for full credit.
- Student writes =AVERAGEIF with no condition that effectively averages all rows — technically uses a different function; accept the numeric result but note the unnecessary complexity.

Task 3

Common Student Mistakes

- Forgetting to add absolute references (\$) to the Readings ranges, so filling down shifts the lookup or sum range and produces wrong results for C9:C13.
- Making the B-column reference absolute (\$B8) so all rows look up B8 instead of advancing down the column.
- Misspelling one or more building names in B8:B13 so SUMIF finds no matches and returns 0.
- Writing the building labels in the wrong column or starting in the wrong row.
- Using COUNTIF instead of SUMIF.
- Using SUMIFS (plural) with only one criterion pair — this works but is unnecessarily complex; note it but accept.
- Putting the sum_range as the second argument and the criteria range as the third (swapped order).

Socratic Questions (When Students Are Stuck)

- If a student is stuck, ask: When you fill a formula down, what happens to cell references that do not have dollar signs — and which references in your formula should stay fixed versus move?
- If a student is stuck, ask: SUMIF has three arguments in a specific order — can you describe what each argument is supposed to represent in plain English before writing the formula?
- If a student is stuck, ask: If C9 is returning 0 for Engineering, what could you check to verify whether the text in B9 exactly matches what appears in the Building column of the Readings sheet?

Solution Walkthrough

1. On the Analysis sheet, click B8 and type 'Wells Hall', then B9 through B13 with 'Engineering', 'Library', 'Union', 'Chemistry', 'Recreation' — one per row, matching the exact case in the Readings data. 2. Click C8. 3. Type: =SUMIF(Readings!\$C\$2:\$C\$121,B8,Readings!\$F\$2:\$F\$121) and press Enter. 4. Confirm a large positive number appears. 5. Click C8 again, then copy the cell (Ctrl+C). 6.

Select C9:C13 and paste (Ctrl+V) or use the fill handle and drag down to C13. 7. Click C9 and verify its formula reads =SUMIF(Readings!\$C\$2:\$C\$121,B9,Readings!\$F\$2:\$F\$121) — the Readings ranges stayed fixed, only B8 advanced to B9. 8. Spot-check: the six building totals in C8:C13 should sum to the same value as B2 from Task 1.

Grading Tips

Award full 10 points if all six cells C8:C13 contain SUMIF formulas with correct absolute references on the Readings ranges and the building labels in B8:B13 are spelled correctly and match the data. Award 7/10 if formulas are correct in C8:C13 but one or two building labels are misspelled causing one or two cells to return 0 — deduct per incorrect label. Award 5/10 if the formula in C8 is correct but the student manually typed different totals in C9:C13 instead of copying the formula. Award 4/10 if SUMIF is used but absolute references are missing and cells C9:C13 produce wrong results due to range shift. Award 2/10 if the formula is placed in the correct cells but references the wrong column (e.g., TotalCost instead of ElectricityKWh). A correct formula placed in D8:D13 instead of C8:C13 earns no more than 5/10 — note wrong cell location.

Discussion Prompt

The sum in C8:C13 should equal the total in B2 from Task 1. If they don't match, what are the possible causes? How would you use this cross-check as an auditing technique in a real finance or operations role?

Edge Cases

- Student types building names with extra spaces or different capitalization (e.g., 'wells hall') — SUMIF is case-insensitive but extra spaces will cause a zero result; deduct per affected row and provide feedback.
- Student enters building names in a different order than the task specifies — the formulas are still valid if they reference the correct B-column cell; accept for full credit as long as results are correct.
- Student uses SUMIFS with a single criterion pair — functionally equivalent; accept for full credit.
- Student sums C8:C13 in a separate cell and verifies it equals B2 — praise this as good auditing practice even though it was not required.

Task 4

Common Student Mistakes

- Using a cell reference for the criteria (e.g., referencing a cell containing 'Academic') when the task asks for the text string directly in the formula — this works but note it.
- Forgetting the quotes around 'Academic' in the formula, causing a NAME error.
- Referencing the Building column (C) instead of the BuildingType column (D) as the criteria range.
- Referencing ElectricityKWh (column F) instead of TotalCost (column J) as the sum range.
- Placing the formula in the wrong cell (e.g., B15 or B17 instead of B16).
- Hard-coding the dollar amount instead of using a formula.

Socratic Questions (When Students Are Stuck)

- If a student is stuck, ask: Look at the Readings sheet headers — which column holds the BuildingType category, and which holds TotalCost? Which column letter is each?

- If a student is stuck, ask: How do you tell Excel that the criteria is the literal text 'Academic' — what punctuation do you need inside the formula?
- If a student is stuck, ask: SUMIF needs three pieces of information: where to look for the condition, what condition to look for, and where to get the numbers to add. Can you identify each of those three things for this task?

Solution Walkthrough

1. On the Analysis sheet, click cell B16. 2. Type:

=SUMIF(Readings!D2:D121,"Academic",Readings!J2:J121) and press Enter. 3. Verify the result is a positive dollar amount less than the overall SUM of TotalCost. Breakdown: Readings!D2:D121 is the BuildingType column (the range to test); "Academic" is the text criteria (quotes required inside the formula); Readings!J2:J121 is the TotalCost column (the range to sum). Note: absolute references are not strictly necessary here since the formula is not being copied, but they are harmless.

Grading Tips

Award full 10 points if a SUMIF formula in B16 correctly references column D as the criteria range, 'Academic' as the criteria, and column J as the sum range. Award 7/10 if the formula is in B16, uses SUMIF, but references the wrong sum column (e.g., F instead of J) — the function is used correctly but the wrong data was aggregated. Award 5/10 if the correct value is typed as a hard-coded number. Award 3/10 if SUMIF is used but both the criteria range and sum range are wrong. Award 0/10 if placed in any cell other than B16 on the Analysis sheet, even if the formula is perfect.

Discussion Prompt

Academic buildings make up some fraction of total cost — how would you compute the percentage, and what follow-up question would you ask leadership if academic buildings account for an unexpectedly high or low share?

Edge Cases

- Student writes ="Academic" as the criteria using a cell reference pointing to a cell with that text — accept for full credit since the result is identical.
- Student uses SUMIFS(Readings!J2:J121,Readings!D2:D121,"Academic") — note argument order is different from SUMIF; if the result is correct, accept for full credit.
- Student sums all three BuildingType groups and verifies they equal the grand total — excellent practice; full credit plus commendation.

Task 5

Common Student Mistakes

- Creating the PivotTable on the Readings sheet or Dashboard sheet instead of the Analysis sheet.
- Setting the value fields to Count instead of Sum for ElectricityKWh and GasTherms.
- Using only one value field (e.g., only ElectricityKWh) and omitting GasTherms.
- Placing the PivotTable starting at a cell other than E2, causing it to overlap with formula work in columns A–D.
- Including extra fields such as Month or BuildingType in the Row Labels area, creating a more granular table than requested.

- Sourcing the PivotTable from the wrong range (e.g., only a subset of rows).

Socratic Questions (When Students Are Stuck)

- If a student is stuck, ask: In the PivotTable Fields pane, there are four areas — Filters, Columns, Rows, and Values. Which area should Building go into, and which area should ElectricityKWh and GasTherms go into?
- If a student is stuck, ask: After dragging ElectricityKWh to the Values area, how do you check whether it is summing or counting — and where do you click to change it?
- If a student is stuck, ask: The task says 'beginning at cell E2' — how do you control where a PivotTable is placed when you insert it?

Solution Walkthrough

1. On the Analysis sheet, click cell E2. 2. Go to the Insert ribbon tab and click PivotTable. 3. In the dialog, set the Table/Range to Readings!\$A\$1:\$J\$121. Choose 'Existing Worksheet' and set Location to Analysis!\$E\$2. Click OK. 4. In the PivotTable Fields pane, drag 'Building' to the Rows area. 5. Drag 'ElectricityKWh' to the Values area — it should default to Sum; if it shows Count, click the field, choose Value Field Settings, and switch to Sum. 6. Drag 'GasTherms' to the Values area and verify it is also set to Sum. 7. Confirm the PivotTable shows six building rows plus a Grand Total, with two value columns.

Grading Tips

Award full 15 points if a PivotTable exists on the Analysis sheet starting near E2, has Building as row labels, and shows Sum of ElectricityKWh and Sum of GasTherms as two value columns. Award 10/15 if the PivotTable is correct but placed starting at a noticeably different cell (e.g., E5 or H2) — deduct for not following placement instructions. Award 8/15 if only one value field is present. Award 5/15 if the PivotTable is on the wrong sheet but otherwise correctly configured. Award 3/15 if a PivotTable exists but has Count instead of Sum for the value fields. A PivotTable cannot be graded by formula checks alone — open the file and inspect the PivotTable Fields pane manually.

Discussion Prompt

PivotTables update when you click 'Refresh' after source data changes, but formulas like SUMIF in Tasks 3 and 4 update automatically. When would you prefer each approach, and what are the risks of forgetting to refresh a PivotTable?

Edge Cases

- Student adds a Columns field (e.g., BuildingType) creating a matrix layout — the required information is present but the layout differs; award 10/15 and note the instructions asked for no column field.
- Student uses a Pivot Chart instead of a PivotTable — a PivotChart creates a backing PivotTable; if the PivotTable with the correct fields exists, award full credit.
- Student's PivotTable source range excludes some rows — if the totals are wrong compared to Tasks 1 and 3, note the discrepancy and deduct accordingly.

Task 6

Common Student Mistakes

- Inserting a Column chart instead of a Bar chart (in Excel, 'Bar' means horizontal bars; 'Column' means vertical bars) — the visual intent is similar but the wrong chart type is selected.
- Placing the chart on the Analysis sheet instead of the Dashboard sheet.
- Not giving the chart a descriptive title, or leaving it as the default 'Chart Title'.
- Selecting data that does not include both ElectricityKWh and GasTherms series, resulting in a single-series chart.
- Sourcing the chart from the raw Readings data instead of the PivotTable summary.
- Creating a stacked bar instead of a clustered bar.

Socratic Questions (When Students Are Stuck)

- If a student is stuck, ask: In Excel's Insert > Charts group, there are separate icons for Bar and Column — which orientation does each produce, and which one does the task require?
- If a student is stuck, ask: After creating the chart, it may have appeared on the Analysis sheet — how do you move a chart to a different sheet?
- If a student is stuck, ask: When you select your data range before inserting the chart, have you included both the building names column and both energy columns from your PivotTable?

Solution Walkthrough

1. On the Analysis sheet, select the PivotTable data: click the first building name cell (likely E3 or the first data row under the row labels), then extend the selection to include the building names and both value columns (ElectricityKWh and GasTherms) through the last building row, excluding the Grand Total row. 2. Go to Insert > Charts and click the Bar Chart icon (horizontal bars), then select Clustered Bar. 3. The chart appears on the Analysis sheet. 4. Right-click the chart border and choose 'Move Chart'. Select 'Object in' and choose Dashboard, or choose 'New Sheet' and rename. Alternatively, cut the chart (Ctrl+X), navigate to the Dashboard sheet, and paste (Ctrl+V). 5. On the Dashboard sheet, click the chart title and type 'Energy Consumption by Building'. 6. Resize and reposition the chart to fit neatly within the Dashboard sheet by dragging the corners.

Grading Tips

Award full 15 points if a clustered bar chart is on the Dashboard sheet, shows both ElectricityKWh and GasTherms series, has building names as categories, and has a descriptive title. Award 10/15 if the chart is a Column chart (vertical) instead of Bar (horizontal) but is otherwise correct — the task explicitly says Bar. Award 10/15 if the chart is correctly configured but placed on the Analysis sheet instead of Dashboard. Award 8/15 if only one series is present. Award 5/15 if the chart is on Dashboard but uses random or incorrect source data. Award 2/15 if a chart exists on Dashboard but is the wrong type (e.g., pie or line) and has incorrect or missing labels. A correct chart in the wrong cell location on the Dashboard does not lose points for cell placement since the task does not specify a target cell — only sheet placement matters.

Discussion Prompt

A clustered bar chart shows absolute totals. What additional chart or transformation would let you compare buildings fairly given that they have different square footages — and how does Task 9's Python chart address this limitation?

Edge Cases

- Student creates the chart from a PivotChart (which is attached to the PivotTable) and moves it to Dashboard — accept for full credit if it displays both series correctly.
- Student selects the Grand Total row as part of the data, creating a 'Grand Total' bar that dwarfs the others — deduct 3 points and note the issue.
- Student creates two separate single-series bar charts instead of one clustered chart — partial credit 8/15; the task requires a single chart with both series.

Task 7

Common Student Mistakes

- Placing the Python formula in a cell other than A2 on the Analysis sheet (common: placing it on Dashboard or on Readings).
- Using `xl()` without `headers=True`, causing the first row to be treated as data rather than column names, so column-name selection fails.
- Using single brackets `df["Building"]` instead of double brackets `df[["Building", ...]]` to select multiple columns, returning a Series instead of a DataFrame.
- Misspelling a column name (e.g., 'Electricity_KWh' or 'Total_Cost') causing a `KeyError`.
- Not returning the DataFrame as the last expression, so the cell shows a Python object descriptor instead of spilling data.
- Forgetting to type the code inside the `PY()` wrapper in Excel's Python mode (i.e., typing Python syntax in a regular formula bar).

Socratic Questions (When Students Are Stuck)

- If a student is stuck, ask: What does `headers=True` do in the `xl()` function — what would happen to your column selection if you omitted it?
- If a student is stuck, ask: In pandas, what is the difference between `df["Building"]` and `df[["Building"]]` — what type of object does each return, and why does that matter here?
- If a student is stuck, ask: Python in Excel requires a specific way to enter Python code — have you activated the Python cell mode using the `PY()` function or the Python toggle before typing your code?

Solution Walkthrough

1. Navigate to the Analysis sheet and click cell A2. 2. In the formula bar, type `=PY()` to open Python in Excel mode (or use the Python cell button on the Formulas ribbon if available). 3. Inside the `PY` cell, type the following two lines: `df = xl("Readings!A1:J121", headers=True)` then on a new line: `df[["Building", "ElectricityKWh", "GasTherms", "TotalCost"]]` 4. Press `Ctrl+Enter` to confirm. 5. The cell should display a Python DataFrame object icon or spill the data. 6. Verify that the spilled output shows four columns: Building, ElectricityKWh, GasTherms, TotalCost with 120 data rows. Note: if the cell shows an error about the range overlapping with existing content (since B2 and B3 from Tasks 1–2 are in use), the spill will be blocked — students may need to place this differently or the instructor should confirm that A2 is the intended location given the layout.

Grading Tips

Award full 10 points if cell A2 on the Analysis sheet contains a PY formula that uses `xl()` with `headers=True` and returns a DataFrame with the four specified columns. Award 7/10 if `xl()` is used correctly but the column selection is missing one of the four required columns. Award 5/10 if `xl()` is used without `headers=True` but the student works around it (e.g., skipping the first row manually) and still returns a four-column result. Award 3/10 if the code is syntactically correct but placed in the wrong cell or wrong sheet. Award 1/10 if the student writes Python syntax in a regular Excel formula bar (i.e., not in a PY cell) — the code is written but not functional. A correct value typed manually earns 0 points; this task is purely about Python in Excel.

Discussion Prompt

This Python cell creates a filtered view of the Readings data. How does this compare to using Excel's native column-hiding or filtering features? In what scenarios would the Python approach be more powerful or more brittle?

Edge Cases

- The spill from A2 may conflict with content in B2 or B3 from Tasks 1 and 2 — if a student reports a spill error, check whether the four-column DataFrame actually extends into column B; if Tasks 1 and 2 are already in B2:B3, the spill may be blocked. Accept the formula as correct and note the layout conflict; do not penalize for the spill block.
- Student returns `df.copy()[[...]]` — functionally identical; accept.
- Student uses `xl("Readings!A1:J121")` without `headers=True` and then manually does `df.columns = df.iloc[0]`; `df = df[1:]` — creative workaround; award 7/10 and note the simpler approach.

Task 8

Common Student Mistakes

- Using `.corr()` on the entire DataFrame (`df.corr()`) and returning the full correlation matrix instead of two specific values.
- Swapping the pairs — computing HDD vs ElectricityKWh and CDD vs GasTherms instead of the specified pairings.
- Returning two separate scalar values without assembling them into a DataFrame, causing only one value to spill.
- Forgetting to import pandas (`pd`) — in Python in Excel, pandas is pre-imported as `pd`, so this is usually not an issue, but students may write `import pandas` as `pd` unnecessarily or forget that `pd` is already available.
- Misspelling column names such as 'HeatingDegreesDay' or 'GasTherm'.
- Placing the formula in the wrong cell (e.g., A24 or A26 instead of A25).

Socratic Questions (When Students Are Stuck)

- If a student is stuck, ask: The pandas Series `.corr()` method takes another Series as its argument — how would you pull a single column out of a DataFrame as a Series to pass into that method?
- If a student is stuck, ask: You need to return two correlation values in one output — what data structure in Python could hold both a label and a number for each of two comparisons, and how would you create it with pandas?

- If a student is stuck, ask: Look carefully at the task — which weather variable is paired with which energy variable? Are you computing HDD vs GasTherms or HDD vs ElectricityKWh?

Solution Walkthrough

1. Navigate to the Analysis sheet and click cell A25. 2. Enter Python mode (PY cell). 3. Type the following code: `df = xl("Readings!A1:J121", headers=True)` then `r_heat = df["HeatingDegreeDays"].corr(df["GasTherms"])` then `r_cool = df["CoolingDegreeDays"].corr(df["ElectricityKWh"])` then `pd.DataFrame({"Comparison": ["HDD vs GasTherms", "CDD vs ElectricityKWh"], "Correlation": [r_heat, r_cool]})` 4. Press Ctrl+Enter. 5. The cell should spill a two-row DataFrame with columns 'Comparison' and 'Correlation'. 6. Verify both correlation values are between -1 and 1. A positive `r_heat` near 0.7–0.9 is plausible (more heating degree days → more gas use). `r_cool` may be moderate.

Grading Tips

Award full 15 points if cell A25 on the Analysis sheet contains a PY formula that computes both specified Pearson correlations and returns a two-row DataFrame with 'Comparison' and 'Correlation' columns. Award 10/15 if both correlations are computed correctly but returned as two separate scalars (two cells) rather than a DataFrame. Award 8/15 if only one of the two correlations is computed correctly. Award 6/15 if the full correlation matrix is returned (`df.corr()`) rather than the two specific values — the information is present but the student did not extract the required pairs. Award 3/15 if the formula is in the wrong cell but otherwise correct. Award 0/15 if a hard-coded number is entered.

Discussion Prompt

The Pearson correlation assumes a linear relationship. If you plotted HeatingDegreeDays vs GasTherms in a scatter chart, what patterns might suggest the relationship is non-linear, and how would that affect the validity of using `r` as a summary statistic here?

Edge Cases

- Student uses numpy's `np.corrcoef()` instead of pandas `.corr()` — if the correct values are returned in a DataFrame, accept for full credit.
- Student computes `df.corr()` and then indexes the specific cells (e.g., `df.corr().loc["HeatingDegreeDays", "GasTherms"]`) — this is correct and elegant; award full credit.
- Correlation values may differ slightly from the expected answer due to floating-point precision — accept any result within 0.001 of the reference value.
- Student labels the DataFrame columns differently (e.g., 'Pair' and 'r') — accept for full credit if the content is correct; the task did not mandate exact column naming.

Task 9

Common Student Mistakes

- Dividing ElectricityKWh by SquareFeet at the row level before grouping, rather than summing ElectricityKWh per building and dividing by the building's square footage — this incorrectly weights months.
- Using `groupby` and `mean()` on ElectricityKWh/SquareFeet instead of the correct `sum(ElectricityKWh)/SquareFeet` calculation.

- Creating a vertical bar chart (`plt.bar`) instead of a horizontal bar chart (`plt.barh`).
- Forgetting to return the fig object as the last expression, causing the cell to show `None` or a text descriptor instead of the chart image.
- Placing the formula in a cell other than A2 on the Dashboard sheet (e.g., on Analysis sheet or in A1).
- Not setting the x-axis label or chart title, or setting the y-axis label instead of x-axis.
- Using `.iloc[0]` on `SquareFeet` without understanding that all rows for a building have the same `SquareFeet` value — students may try `.mean()` or `.sum()` on `SquareFeet` instead, which also work and are acceptable.

Socratic Questions (When Students Are Stuck)

- If a student is stuck, ask: Energy intensity is total kWh divided by square footage — if you use `groupby` on `Building`, how do you compute both the total kWh and the building's square footage within the same lambda?
- If a student is stuck, ask: matplotlib has both `plt.bar()` for vertical bars and `plt.barh()` for horizontal bars — which one puts buildings on the y-axis and kWh on the x-axis?
- If a student is stuck, ask: In Python in Excel, what must be the very last expression in your code cell for Excel to render it as output — and what type of object does matplotlib produce that Excel can display as an image?

Solution Walkthrough

1. Navigate to the Dashboard sheet and click cell A2. 2. Enter Python mode (PY cell). 3. Type the code: `df = xl("Readings!A1:J121", headers=True)` then `intensity = df.groupby("Building").apply(lambda g: g["ElectricityKWh"].sum() / g["SquareFeet"].iloc[0]).sort_values()` then `fig, ax = plt.subplots(figsize=(8, 5))` then `ax.barh(intensity.index, intensity.values)` then `ax.set_xlabel("kWh per Square Foot")` then `ax.set_title("Electricity Intensity by Building (kWh / sq ft)")` then `plt.tight_layout()` then `fig` 4. Press `Ctrl+Enter`. 5. The cell should render a horizontal bar chart image embedded in the Dashboard sheet. 6. Verify that six buildings appear on the y-axis, sorted from lowest to highest intensity, and that the x-axis label and title are present. Note: `plt` is pre-imported in Python in Excel; matplotlib does not need to be imported explicitly.

Grading Tips

Award full 15 points if cell A2 on the Dashboard sheet contains a PY formula that loads the data, computes a per-building electricity intensity (total kWh / square footage), and returns a matplotlib horizontal bar chart with the correct title and x-axis label. Award 11/15 if the chart is correct but uses `plt.bar` (vertical) instead of `plt.barh` (horizontal). Award 10/15 if the chart is horizontal and titled correctly but the intensity calculation is wrong (e.g., row-level division then averaged). Award 8/15 if the chart is produced and placed correctly but lacks the title or the x-axis label. Award 5/15 if a chart image is produced but the underlying calculation is entirely wrong (e.g., plotting raw `ElectricityKWh` totals instead of intensity). Award 3/15 if the formula is in the wrong cell or wrong sheet but otherwise correct. Do not award points if the result is a `DataFrame` rather than a chart (the student may have forgotten to include the `fig` object).

Discussion Prompt

The chart sorts buildings by intensity from lowest to highest. Which building appears worst, and what follow-up investigation would you recommend — is high intensity always a problem, or could building function (e.g., a 24/7 recreation facility vs a lecture hall) justify higher energy use per square foot?

Edge Cases

- Student uses `.mean()` on `SquareFeet` within the `groupby` instead of `.iloc[0]` — since all readings for a building share the same `SquareFeet` value, the mean equals the value; accept for full credit.
- Student uses `df.groupby("Building").agg({"ElectricityKWh": "sum", "SquareFeet": "first"})` and then computes the ratio as a separate step — correct and arguably cleaner; award full credit.
- Student forgets `.sort_values()` so bars are in arbitrary order — deduct 2 points for not sorting as the task implies (the word 'highlight worst-performing' suggests sorting).
- Student adds additional styling (color palette, grid lines, value labels on bars) — bonus effort; award full credit and commend.
- Student places a second chart on the Dashboard sheet in addition to the Task 6 clustered bar chart — both charts can coexist; confirm A2 contains the Python chart and grade accordingly.