

Instructor Guide: Riverbend Transit — Ridership Analysis with Python in Excel

CONFIDENTIAL — FOR INSTRUCTOR AND TA USE ONLY

Assignment Overview

This assignment assesses students' ability to perform transit operations analysis using Excel across four progressive skill tiers: basic aggregation (SUM, AVERAGE), conditional aggregation (SUMIF), PivotTables and charts for visualisation, and Python in Excel for programmatic data wrangling, statistics, and matplotlib visualisation. Students work across three sheets — Operations (raw data), Analysis (formula and PivotTable work), and Dashboard (charts and Python outputs) — reinforcing cross-sheet referencing, absolute vs. relative addressing, and the distinction between Excel-native and Python-native workflows. By the end, students should be able to build a complete analytical pipeline from raw operational records to a visual executive dashboard.

Partial Credit Policy

In general, award partial credit whenever a student demonstrates meaningful understanding of the required technique even if the implementation is incomplete. A correct formula placed in the wrong cell earns approximately 60% of the question points — the student understood the formula but misread the instructions. A correct numeric value typed as a hard-coded constant always earns 0 points, because the core learning objective is formula construction, not arithmetic. For multi-part questions (Task 3 with four cells, Task 9 with multiple chart elements), award proportional partial credit based on the fraction of correct sub-components. For Python in Excel tasks, award partial credit if the code is structurally correct but produces an error due to a minor typo in a column name, as this demonstrates understanding of the approach. Do not award partial credit for completely blank cells or for answers that reference the wrong sheet entirely with no evident attempt to cross-reference Operations data.

General Grading Tips

- Before grading, open the student's file and enable macros/content if prompted, then verify that both Analysis and Dashboard sheets exist — if either sheet is missing, Tasks 1-6 and 7-9 respectively cannot earn points.
- For cross-sheet formula tasks (1-4), click into the formula cell and check the formula bar directly; do not rely solely on the displayed value, which could match the correct answer even if hard-coded.
- For Python in Excel tasks (7-9), look for the yellow 'PY' badge in the cell's bottom-left corner to confirm a Python formula is present; a cell without this badge is either a regular formula or a hard-coded value.
- If a student's Python in Excel cells show a spinning icon or 'Calculating' indefinitely, the Python runtime may have timed out; ask the student to save, close, reopen, and try refreshing before marking as incorrect.

- Check absolute vs. relative references in Tasks 3 and 4 by clicking into each filled cell (B7, B8, B9) and inspecting the formula bar — do not assume fill-down worked correctly just because B6 is right.
 - For the PivotTable (Task 5), right-click a value cell inside the PivotTable and select 'Value Field Settings' to confirm the aggregation is Average, not Sum — the displayed numbers alone may not make this obvious.
 - For the chart (Task 6), click the chart on the Dashboard sheet, then go to Chart Design > Select Data to verify the data source is Analysis!A6:B9 and not some other range.
 - Grade Python in Excel tasks by running the formula yourself in a reference copy of the dataset if you are unsure whether the student's output values are correct — small discrepancies may indicate a logic error vs. a rounding difference.
 - Note that Python in Excel requires Microsoft 365 (certain subscription tiers) and may not be available to all students; confirm software availability issues with the department before marking Tasks 7-9 as zero.
-

Task-by-Task Guide

Task 1

Common Student Mistakes

- Typing the numeric total directly into B2 instead of using a SUM formula.
- Writing =SUM(K2:K81) without the sheet reference, which sums the Analysis sheet's own column K (likely empty) and returns 0.
- Using =SUM(Operations!K2:K80), missing the last data row (row 81).
- Using =SUM(Operations!K:K), which includes the header cell and may cause issues if the header is non-numeric — technically still works but is imprecise.
- Placing the formula in the wrong cell, e.g. B3 instead of B2.

Socratic Questions (When Students Are Stuck)

- If a student is stuck, ask: Which sheet holds the TotalRiders data, and how do you tell Excel to look at a different sheet inside a formula?
- If a student is stuck, ask: The assignment says 80 log entries starting at row 2 — what is the last row number that contains data?
- If a student is stuck, ask: How would you check that your formula is actually reading from the Operations sheet and not from the current sheet?

Solution Walkthrough

1. Click the Analysis sheet tab. 2. Click cell B2. 3. Type =SUM(then click the Operations sheet tab. 4. Select cells K2 through K81 (all 80 TotalRiders values) using the mouse or by typing the range. 5. Close the parenthesis and press Enter. The formula bar should read =SUM(Operations!K2:K81). 6. Verify the result is a plausible total (in the tens of thousands range for 80 routes).

Grading Tips

Award full 5 points only if a SUM formula referencing Operations!K2:K81 (or an equivalent full-column range that correctly captures all 80 rows) is present in B2. Award 3 out of 5 if the formula is correct but

placed in the wrong cell (e.g. B3). Award 0 points if the cell contains a hard-coded number with no formula, regardless of whether the value is correct. Award 4 out of 5 if the range is off by one row (e.g. K2:K80) but the formula structure is otherwise correct.

Discussion Prompt

Why is it important to use a formula rather than typing the total directly, especially in a live transit operations dashboard that might be refreshed with new data each week?

Edge Cases

- Student uses `=SUM(Operations!K:K)` — this will include the header row label, but since 'TotalRiders' is text it is ignored by SUM; accept for full credit but note in feedback that explicit row ranges are better practice.
- Student uses a named range if they defined one — accept as full credit if the named range correctly covers K2:K81 on Operations.

Task 2

Common Student Mistakes

- Using AVERAGE without wrapping in ROUND, e.g. `=AVERAGE(Operations!N2:N81)`, missing the rounding requirement.
- Reversing the argument order in ROUND, e.g. `=ROUND(2, AVERAGE(Operations!N2:N81))`.
- Rounding to the wrong number of decimal places, e.g. `ROUND(...,0)` or `ROUND(...,1)`.
- Hard-coding the computed average as a decimal number instead of using a formula.
- Forgetting the sheet reference and computing AVERAGE on an empty range in the Analysis sheet.

Socratic Questions (When Students Are Stuck)

- If a student is stuck, ask: ROUND takes two arguments — what does each argument represent, and which one controls the number of decimal places?
- If a student is stuck, ask: How would you nest two functions together in Excel so that the result of AVERAGE becomes the input to ROUND?
- If a student is stuck, ask: Which column letter in the Operations sheet holds FuelUsedGallons, and what row does the data end on?

Solution Walkthrough

1. Click cell B3 on the Analysis sheet. 2. Begin typing `=ROUND(` 3. Inside the ROUND, type AVERAGE(4. Click the Operations sheet tab and select N2:N81. 5. Close the AVERAGE parenthesis with). 6. Type a comma and then 2. 7. Close the ROUND parenthesis and press Enter. The formula bar should read `=ROUND(AVERAGE(Operations!N2:N81),2)`. 8. Confirm the result displays exactly two decimal places.

Grading Tips

Award full 5 points for `=ROUND(AVERAGE(Operations!N2:N81),2)` or equivalent range in B3. Award 3 out of 5 if AVERAGE is present and correctly references Operations!N2:N81 but ROUND is missing. Award 3 out of 5 if both functions are present but placed in the wrong cell. Award 0 points for a

hard-coded value even if numerically correct. Deduct 1 point if the decimal places argument is wrong (e.g. 0 or 1 instead of 2).

Discussion Prompt

When reporting fuel efficiency metrics to a transit authority, why might you choose to round averages to 2 decimal places specifically, and could over-rounding or under-rounding ever mislead stakeholders?

Edge Cases

- Student writes `=ROUND(AVERAGE(Operations!N2:N81),2)` but formats the cell to show 0 decimals — the formula is correct so award full credit; note the display formatting issue separately.
- Student uses `AVERAGEIF` with no criteria (effectively computing the same result) — award full credit if the result is correct and a formula is used.

Task 3

Common Student Mistakes

- Forgetting absolute references on the Operations ranges, so when the formula is filled down from B6 to B9 the criteria range and sum range shift incorrectly.
- Using relative references for the criteria (e.g. A6 becomes absolute as `$A6` or `$A$6`), preventing the criteria from changing as the formula fills down — the criteria cell A6:A9 must remain relative in the row.
- Hard-coding the zone name as a string literal instead of referencing A6:A9, e.g. `=SUMIF(Operations!$C$2:$C$81,"North",Operations!$K$2:$K$81)` for all four cells.
- Referencing the wrong column for the sum range (e.g. column J OnTimeTrips instead of column K TotalRiders).
- Using `SUMIFS` instead of `SUMIF` (acceptable but unusual for a single criterion — still award full credit if correct).
- Placing all four formulas in the wrong rows (e.g. B5:B8 instead of B6:B9).

Socratic Questions (When Students Are Stuck)

- If a student is stuck, ask: `SUMIF` has three arguments — what does each one represent in the context of this problem?
- If a student is stuck, ask: If you lock a reference with dollar signs, what happens to that part of the reference when you copy the formula to another cell?
- If a student is stuck, ask: Look at the formula you have in B6 — if you copy it to B7, which parts of the formula should change and which parts should stay the same?

Solution Walkthrough

1. Click cell B6 on the Analysis sheet. 2. Type `=SUMIF(` 3. Click the Operations sheet, select C2:C81, then press F4 to make it absolute: `Operations!$C$2:$C$81`. 4. Type a comma. 5. Click back to Analysis sheet and click cell A6 (leave this relative — no dollar signs). 6. Type a comma. 7. Click Operations sheet, select K2:K81, press F4: `Operations!$K$2:$K$81`. 8. Close parenthesis and press Enter. 9. Click B6 again, copy (Ctrl+C). 10. Select B7:B9 and paste (Ctrl+V). Each cell should reference the corresponding zone label in A7, A8, A9. 11. Verify that B6 through B9 each show different non-zero totals.

Grading Tips

Award full 10 points if all four cells B6:B9 contain SUMIF formulas with absolute references on Operations ranges and relative references on A6:A9, and produce correct zone totals. Award 7 out of 10 if the formula logic is correct but absolute references are missing and only B6 is correct (the fill-down produced wrong results). Award 5 out of 10 if all four cells contain hard-coded zone names as string literals in the criteria argument but the rest of the formula is correct. Award 3 out of 10 if only one correct formula exists in B6 and the others are missing or wrong. A correct formula placed in wrong rows (e.g. B5:B8) earns at most 6 out of 10.

Discussion Prompt

What are the trade-offs between using SUMIF with zone labels in helper cells (A6:A9) versus hard-coding the zone names as strings directly in the formula? How does the helper-cell approach improve maintainability?

Edge Cases

- Student types zone names in A6:A9 with extra spaces or different capitalisation (e.g. 'north' or 'North ') — SUMIF is case-insensitive but trailing spaces will cause a mismatch; check the labels in A6:A9 carefully before penalising the formula.
- Student uses SUMIFS with a single criterion — this is functionally identical and should receive full credit.
- Student copies the formula correctly but zones in A6:A9 are in a different order than North/South/East/West — award full credit if each formula correctly references its adjacent zone label, regardless of order.

Task 4

Common Student Mistakes

- Using AVERAGEIF instead of two SUMIF functions — AVERAGEIF on OnTimeTrips gives the average trip count, not the ratio of totals, which is a conceptually different (and incorrect) metric.
- Dividing SUMIF(OnTimeTrips) by SUMIF(ScheduledTrips) instead of SUMIF(CompletedTrips), misreading the on-time rate definition.
- Forgetting to format the cell as a percentage, leaving a raw decimal (e.g. 0.72 instead of 72%).
- Hard-coding the criteria string without quotes, causing a formula error.
- Using the wrong column for OnTimeTrips (column J) or CompletedTrips (column I) — swapping these columns.
- Not using absolute references on Operations ranges, which doesn't matter for a single cell but is still poor practice.

Socratic Questions (When Students Are Stuck)

- If a student is stuck, ask: The on-time rate is a ratio — what two quantities form the numerator and denominator, and which column in Operations holds each one?
- If a student is stuck, ask: How would you use SUMIF to add up values from one column only for rows that meet a condition in a different column?

- If a student is stuck, ask: After you calculate the decimal ratio, what step do you need to take in Excel to display it as a percentage?

Solution Walkthrough

1. Click cell B12 on the Analysis sheet. 2. Type =SUMIF(3. Select Operations!P2:P81 (IncidentReported column), press F4, type , "Yes", 4. Select Operations!J2:J81 (OnTimeTrips), press F4, close parenthesis. 5. Type / 6. Type SUMIF(7. Select Operations!P2:P81 again (absolute), type , "Yes", 8. Select Operations!I2:I81 (CompletedTrips), absolute, close parenthesis. 9. Press Enter. The formula bar should read =SUMIF(Operations!\$P\$2:\$P\$81,"Yes",Operations!\$J\$2:\$J\$81)/SUMIF(Operations!\$P\$2:\$P\$81,"Yes",Operations!\$I\$2:\$I\$81). 10. With B12 selected, go to Home > Number group and click the Percentage (%) button to format as a percentage.

Grading Tips

Award full 10 points for the correct two-SUMIF ratio in B12, formatted as a percentage. Award 7 out of 10 if the formula logic is correct but the cell is not formatted as a percentage (raw decimal shown). Award 5 out of 10 if only one SUMIF is used and the denominator is a total count (e.g. 80) rather than a conditional sum of CompletedTrips. Award 3 out of 10 if AVERAGEIF is used instead of the ratio approach but references the correct columns. A hard-coded percentage value earns 0 points regardless of accuracy.

Discussion Prompt

Why might an on-time performance rate calculated only for incident days be a useful KPI for a transit authority, and what decisions might it inform compared to the overall on-time rate?

Edge Cases

- If the Operations data contains no rows with IncidentReported='Yes', the formula will produce a division-by-zero error (#DIV/0!). In the context of this specific dataset this should not occur, but if it does, acknowledge the formula is structurally correct and award full marks.
- Student wraps the entire formula in IFERROR — accept for full credit; the added error handling is good practice.
- Student uses 'yes' (lowercase) as the criteria — SUMIF is case-insensitive so this still produces correct results; award full credit.

Task 5

Common Student Mistakes

- Placing the PivotTable starting at a cell other than D2, e.g. A1 (which conflicts with existing formulas) or D1.
- Choosing Sum instead of Average as the value field aggregation for FuelUsedGallons.
- Dragging VehicleType to Rows and RouteCode to Columns instead of the specified orientation.
- Pointing the data source at only part of the Operations table (e.g. leaving out the header row or missing some columns).
- Placing the PivotTable on the Operations sheet or Dashboard sheet instead of Analysis.
- Not formatting the PivotTable values to 2 decimal places.

Socratic Questions (When Students Are Stuck)

- If a student is stuck, ask: Where in the Excel ribbon do you find the option to insert a PivotTable, and what information does Excel need from you before it creates one?
- If a student is stuck, ask: After the PivotTable field list appears, which area (Rows, Columns, Values, Filters) would you drag each of the three fields into?
- If a student is stuck, ask: Once FuelUsedGallons is in the Values area, how do you change whether it is summarised as Sum or Average?

Solution Walkthrough

1. Click anywhere in the Operations sheet data or select the full range A1:P81. 2. Go to Insert tab > PivotTable > From Table/Range. 3. Confirm the range is Operations!\$A\$1:\$P\$81. 4. Select 'Existing Worksheet' and type or click Analysis!\$D\$2 as the location. 5. Click OK. 6. In the PivotTable Fields pane, drag RouteCode to the Rows area. 7. Drag VehicleType to the Columns area. 8. Drag FuelUsedGallons to the Values area. 9. Click the dropdown arrow on 'Sum of FuelUsedGallons' in the Values area, select 'Value Field Settings', and change the summarise function to Average. Click OK. 10. Right-click any value cell inside the PivotTable, select 'Number Format', choose Number with 2 decimal places, click OK.

Grading Tips

Award full 15 points if a PivotTable exists on the Analysis sheet with RouteCode in rows, VehicleType in columns, Average of FuelUsedGallons in values, formatted to 2 decimal places, starting at or very near D2. Award 10 out of 15 if the PivotTable is correctly configured but placed in the wrong starting cell (not D2) or on the wrong sheet. Award 8 out of 15 if Sum is used instead of Average but all other aspects are correct. Award 5 out of 15 if a PivotTable exists but rows and columns are swapped. No partial credit for a manually typed cross-tab table that mimics a PivotTable.

Discussion Prompt

How does a PivotTable's ability to quickly swap rows and columns help a transit analyst explore hypotheses about fuel usage patterns compared to writing SUMIF or AVERAGEIF formulas manually?

Edge Cases

- PivotTable starts at D3 or E2 instead of D2 — deduct 2 points for positioning error but award credit for correct configuration.
- Student creates the PivotTable correctly but it overlaps with formula cells B2:B9 — this typically indicates the PivotTable was placed in the wrong area; penalise placement but not configuration.
- Student uses a newer Excel version where the PivotTable field list UI looks different (e.g. Microsoft 365 new design) — the functional result is the same; grade on output, not UI path.

Task 6

Common Student Mistakes

- Creating a Column (vertical bar) chart instead of a Bar (horizontal bar) chart — these are different chart types in Excel.
- Using the wrong data source, e.g. pointing to all of Operations instead of the Analysis!A6:B9 zone summary.

- Omitting or misspelling the chart title ('Total Ridership by Zone' must match exactly or be very close).
- Leaving the chart on the Analysis sheet instead of moving it to the Dashboard sheet.
- Placing the chart outside the specified A2:H20 range on the Dashboard sheet.
- Creating a Line or Pie chart instead of a clustered Bar chart.

Socratic Questions (When Students Are Stuck)

- If a student is stuck, ask: In Excel's Insert Chart dialog, what is the difference between a 'Bar' chart and a 'Column' chart, and which one does the assignment ask for?
- If a student is stuck, ask: After you create the chart on the Analysis sheet, how do you move it to a different sheet entirely?
- If a student is stuck, ask: Where in the chart interface do you click to change or add the chart title?

Solution Walkthrough

1. Click the Analysis sheet tab. 2. Select cells A6:B9 (zone labels and totals). 3. Go to Insert tab > Charts group > Insert Bar or Column Chart > Clustered Bar (the horizontal bar option, first in the Bar row). 4. A chart will appear on the Analysis sheet. 5. Click the chart to select it. 6. Go to Chart Design tab > Move Chart > Object in 'Dashboard'. Click OK. 7. On the Dashboard sheet, the chart will appear. Click and drag it so the top-left corner aligns with cell A2. 8. Drag the bottom-right resize handle to fit within H20. 9. Click the chart title text box (default may say 'TotalRiders' or similar) and type 'Total Ridership by Zone'. 10. Click outside the chart to deselect.

Grading Tips

Award full 15 points if a clustered Bar chart (horizontal bars) exists on the Dashboard sheet within A2:H20, sourced from the zone summary data, with the correct title. Award 10 out of 15 if a Column chart (vertical bars) is used instead of a Bar chart but all other requirements are met — the student chose the wrong orientation. Award 8 out of 15 if the chart type and data are correct but the chart is on the Analysis sheet instead of Dashboard. Award 5 out of 15 if the chart exists on Dashboard but uses the wrong data source (e.g. raw Operations data). Deduct 2 points if the title is missing or significantly misspelled. Do not penalise for minor chart formatting choices (colours, gridlines).

Discussion Prompt

When would you choose a horizontal bar chart over a vertical column chart for communicating data to a transit authority board, and does the number of categories or the length of category labels influence this decision?

Edge Cases

- Student creates a 3-D Bar chart — this is technically a Bar chart variant; award full credit but note that 3-D charts can distort perception of values.
- Chart extends slightly beyond H20 (e.g. to H21 or I20) — deduct 1 point for positioning but do not penalise heavily for minor overflow.
- Student uses a Stacked Bar chart — this would only have a single series (TotalRiders) so it looks identical to a Clustered Bar; accept for full credit.
- Chart data source references the correct values but via a different path (e.g. directly from Operations using a SUMIF-equivalent range) — award full credit if the chart displays the correct

four-zone totals.

Task 7

Common Student Mistakes

- Using `xl()` without `headers=True`, causing column names to default to generic labels (Column1, Column2, etc.) instead of the actual field names.
- Forgetting `as_index=False` in `groupby`, resulting in `RouteCode` becoming the DataFrame index rather than a regular column, so it does not appear in the output as 'RouteCode'.
- Using the wrong aggregation key names in `agg()`, causing column naming errors (e.g. naming the column 'FuelUsedGallons' instead of 'AvgFuelGallons').
- Not calling `.reset_index(drop=True)` after `sort_values()`, leaving a non-sequential index that can confuse Excel's rendering.
- Calling `plt.show()` or returning a string description instead of the DataFrame object.
- Using `=PY()` syntax incorrectly — e.g. forgetting to open the Python environment with `Ctrl+Shift+Alt+P` instead of typing `=PY(` directly.

Socratic Questions (When Students Are Stuck)

- If a student is stuck, ask: What does the `headers=True` argument in `xl()` do, and what would happen to your column references if you left it out?
- If a student is stuck, ask: In a pandas `groupby().agg()` call, how do you specify both the source column name and the output column name for each aggregation?
- If a student is stuck, ask: After `sort_values()`, what does `reset_index(drop=True)` do, and why might Excel display results differently without it?

Solution Walkthrough

1. Click cell A22 on the Dashboard sheet. 2. Press `Ctrl+Shift+Alt+P` (or go to Formulas > Python) to open a Python in Excel cell. 3. Type the following Python code: `df = xl("Operations!A1:P81", headers=True)` — this loads all Operations data into a pandas DataFrame. 4. On a new line, write the `groupby` and `agg` call: `result = df.groupby("RouteCode", as_index=False).agg(CompletedTrips=("CompletedTrips", "sum"), TotalRiders=("TotalRiders", "sum"), AvgFuelGallons=("FuelUsedGallons", "mean"))`. 5. Round: `result["AvgFuelGallons"] = result["AvgFuelGallons"].round(2)`. 6. Return: `result.sort_values("RouteCode").reset_index(drop=True)`. 7. Press `Ctrl+Enter` to execute. 8. Excel should spill a DataFrame table starting at A22 with columns `RouteCode`, `CompletedTrips`, `TotalRiders`, `AvgFuelGallons`.

Grading Tips

Award full 15 points if: (a) a PY formula exists in A22, (b) the output DataFrame has exactly the four specified columns in any order, (c) all 8 RouteCodes are present, (d) AvgFuelGallons is rounded to 2 decimal places, (e) results are sorted ascending by RouteCode. Award 10 out of 15 if the DataFrame is correct but not sorted or not rounded. Award 8 out of 15 if the column names do not match exactly (e.g. 'FuelUsedGallons' instead of 'AvgFuelGallons') but the data is correct. Award 5 out of 15 if a PY formula exists but returns a Python object preview (not spilled as a DataFrame) — the student may need to right-click and choose 'Show as Array'. A hard-coded table earns 0 points.

Discussion Prompt

How does Python in Excel's `xl()` function bridge the gap between Excel's grid and pandas DataFrames, and what advantages does this offer compared to building the same summary with PivotTables or SUMIF formulas?

Edge Cases

- Student writes the agg dictionary with lambda functions instead of named tuples (e.g. `agg({'CompletedTrips': 'sum'})`) — this produces correct values but column names will be the original names, not the renamed ones; deduct 3 points for column naming mismatch.
- Student returns result without `sort_values` — accept with minor deduction (1-2 points) if all other requirements are met.
- Student uses a newer pandas syntax or alternative groupby approach (e.g. `pivot_table`) — award full credit if the output DataFrame content is correct.
- Python in Excel is not enabled in the student's Excel version — if confirmed as a software access issue rather than student error, consult the department's accommodation policy.

Task 8

Common Student Mistakes

- Including non-numeric columns in the correlation call (e.g. passing the entire DataFrame to `.corr()`), which in newer pandas raises an error or silently drops non-numeric columns.
- Misspelling a column name in the cols list (e.g. 'AvgSpeed' instead of 'AvgSpeedMPH'), causing a `KeyError`.
- Rounding to the wrong number of decimal places (e.g. `.round(2)` instead of `.round(3)`).
- Calling `.corr()` on a single column (Series) rather than a subset DataFrame, producing a scalar 1.0 instead of a matrix.
- Returning the result as a string (e.g. using `print()`) instead of the DataFrame object.

Socratic Questions (When Students Are Stuck)

- If a student is stuck, ask: How do you select multiple specific columns from a pandas DataFrame using a list of column names?
- If a student is stuck, ask: What does pandas' `.corr()` method return — a single number, a Series, or a DataFrame — and why does that matter for how Excel renders it?
- If a student is stuck, ask: The assignment asks for 3 decimal places — where in the code do you control how many decimal places appear in the output?

Solution Walkthrough

1. Click cell A42 on the Dashboard sheet. 2. Open a Python in Excel cell (`Ctrl+Shift+Alt+P`). 3. Type: `df = xl("Operations!A1:P81", headers=True)`. 4. On the next line, define the column list: `cols = ["ScheduledTrips", "CompletedTrips", "TotalRiders", "FuelUsedGallons", "AvgSpeedMPH"]`. 5. On the final line, return: `df[cols].corr().round(3)`. 6. Press `Ctrl+Enter`. 7. Excel should spill a 5x5 correlation matrix with row and column labels matching the five variable names. Diagonal values should all be 1.000.

Grading Tips

Award full 10 points if a PY formula in A42 returns a 5x5 correlation matrix with the correct five column/row labels and values rounded to 3 decimal places. Award 7 out of 10 if the matrix is correct but rounded to 2 decimal places. Award 5 out of 10 if only a subset of the five columns is included (e.g. only 3 columns). Award 3 out of 10 if the formula returns a Python object preview rather than a spilled matrix — student needs to choose 'Show as Array'. A hard-coded matrix earns 0 points.

Discussion Prompt

Looking at the correlation matrix output, which pairs of operational variables show the strongest relationships, and what might that imply for route planning or resource allocation decisions at TVRTA?

Edge Cases

- Newer pandas versions (2.0+) deprecated calling `.corr()` on DataFrames containing non-numeric types; if the student passes `df.corr()` without subsetting, they may get a warning or error — guide them to subset first rather than penalising the pandas version difference.
- Student uses `method='spearman'` or `method='kendall'` inside `.corr()` — these are valid alternative correlation methods; accept for full credit and note the difference in the feedback.
- The 5x5 matrix spills over into cells occupied by other content — this is a layout issue, not a formula error; award full formula credit.

Task 9

Common Student Mistakes

- Calling `plt.show()` instead of returning the fig object — `plt.show()` returns None, so Excel renders nothing.
- Not importing `matplotlib.pyplot` (missing `import matplotlib.pyplot as plt` at the top of the Python cell).
- Using `pivot_table` incorrectly or forgetting to call `.unstack()`, resulting in a Series rather than a wide DataFrame, which does not produce a grouped bar chart.
- Setting `kind='bar'` on the original long-form grouped DataFrame rather than the unstacked pivot, which produces a stacked or incorrect chart.
- Omitting axis labels or the chart title, missing partial credit points.
- Returning `plt` instead of `fig`, or returning `ax` instead of `fig`.

Socratic Questions (When Students Are Stuck)

- If a student is stuck, ask: In matplotlib, what is the difference between `fig` and `ax`, and which one does Excel need you to return to render a chart as an image?
- If a student is stuck, ask: After you group by two columns and compute the mean, what does `.unstack()` do to the resulting data structure, and why is that useful for a grouped bar chart?
- If a student is stuck, ask: If `plt.show()` displays a chart in Jupyter but returns None — and Excel needs a return value — what should you return instead?

Solution Walkthrough

1. Click cell A55 on the Dashboard sheet. 2. Open a Python in Excel cell. 3. First line: `import matplotlib.pyplot as plt` (note: in Python in Excel, matplotlib may be pre-imported, but explicit import is safer). 4. Load data: `df = xl("Operations!A1:P81", headers=True)`. 5. Create pivot: `pivot =`

`df.groupby(["Zone", "VehicleType"])["TotalRiders"].mean().unstack()`. This produces a DataFrame with Zone as index, VehicleType columns (Articulated, MiniBus, Standard), and mean TotalRiders as values. 6. Create figure: `fig, ax = plt.subplots(figsize=(8, 5))`. 7. Plot: `pivot.plot(kind="bar", ax=ax)`. 8. Set title: `ax.set_title("Avg Ridership by Zone and Vehicle Type")`. 9. Set x label: `ax.set_xlabel("Zone")`. 10. Set y label: `ax.set_ylabel("Average Riders")`. 11. Set legend: `ax.legend(title="VehicleType")`. 12. Layout: `plt.tight_layout()`. 13. Final line (return): `fig`. 14. Press Ctrl+Enter. Excel should embed a bar chart image in A55.

Grading Tips

Award full 15 points if: (a) a PY formula is in A55, (b) Excel renders an embedded chart image (not a Python object preview), (c) the chart is a grouped bar chart showing Zone on x-axis and Average Riders on y-axis, (d) chart title is 'Avg Ridership by Zone and Vehicle Type' or very close, (e) x and y axis labels are present, (f) a legend is present. Award 10 out of 15 if the chart renders correctly but title, axis labels, or legend is missing (deduct 1-2 points per missing element). Award 7 out of 15 if a chart image appears but it is not grouped by VehicleType (e.g. only one series per zone). Award 3 out of 15 if the code runs without error but returns a Python object preview rather than a rendered image — student needs to right-click and select 'Show as Image'. Hard-coded or screenshot-pasted images earn 0 points.

Discussion Prompt

What additional customisations to this matplotlib chart (e.g. colour palette, value labels on bars, figure size, error bars) would make it more suitable for inclusion in an executive report, and how do these differ from what Excel's built-in charting tools offer?

Edge Cases

- Student uses seaborn instead of matplotlib directly (e.g. `sns.barplot()`) — accept for full credit if the returned fig object renders correctly in Excel, as seaborn is built on matplotlib.
- The `unstack()` call results in NaN values for Zone-VehicleType combinations with no data — matplotlib will simply omit those bars; this is correct behaviour and should not be penalised.
- Student uses figsize values different from (8,5) — accept any reasonable figsize; this does not affect correctness.
- Python in Excel renders a generic 'Python Object' card rather than an image — instruct the student to right-click the cell and choose 'Show as Image (In Cell)' or 'Plot Output'; this is a display setting, not a code error, so do not penalise if the code is correct.