CURRICULUM • 6 WEEKS

Data Analysis with Python

06

WEEKLY MODULES

From Python code to evidence-based reporting

Google Colab • pandas • matplotlib • capstone project

An instructor-ready lesson plan

COURSE OVERVIEW

What students will learn

A coding-based course focused on asking, testing, and communicating fair comparisons.

AUDIENCE

Basic Python fluency

Variables, loops, functions, and lists

TOOLKIT

Browser-based

Python, pandas, and matplotlib in Colab

END PRODUCT

Evidence-based report

Student analysis presented in Week 6

DESIGNED BACKWARD FROM THE CAPSTONE

Data analysis is more than pandas syntax. Every week teaches a report skill, from describing the data to defending a claim and acknowledging limits.

Scheduling note: the overview says ~45 minutes per weekly session, but the detailed class agendas each run 90 minutes. The slides retain the 90-minute agendas.

COURSE OVERVIEW

Six weeks, one analytical story

One practice dataset for Weeks 1–4, then a capstone dataset for the final project.

01 From code to data

Describe a dataset

02 Cleaning + new columns

Document data decisions

03 Summarize + compare

Create fair summary tables

04 Visualization

Create charts with findings

05 Analysis to argument

Claims, evidence, limitations

06 Capstone presentations

Present the complete report

BEFORE WEEK 1

Launch in Google Colab

Students need only a Google account and a browser. Nothing must be installed.

REQUIRED TOOLS

1. **Google account**
   Sign in and save work to Drive
2. **Google Colab**
   Run Python, pandas, and matplotlib
3. **Google Sheets / Excel**
   Preview CSV files
4. **Optional: Python + JupyterLab**
   Use for offline work

STUDENT SETUP • ~10 MINUTES

1. Choose a Google account. If school access is restricted, use a personal one.
2. Go to colab.research.google.com and open a new notebook.
3. Run the test cell with Shift + Enter. Expect two version numbers and a line chart.
4. Rename the notebook week0_test. Colab saves it in Drive automatically.
5. If the test fails, send a screenshot to the instructor before Week 1.

BEFORE WEEK 1

Repeatable Colab workflows

Practice importing data once, then use the same routine throughout the course.

SMOKE TEST CODE

```
import pandas as pd
import matplotlib
import matplotlib.pyplot as plt

print("pandas", pd.__version__)
print("matplotlib",
matplotlib.__version__)

pd.Series([1, 3, 2, 5]).plot(title="Setup
works")
plt.show()
```

Successful output: two version numbers and a small line chart.

CSV + NOTEBOOK CHECKLIST

IMPORT A CSV

- Upload CSV from the left-hand Files panel.

```
df = pd.read_csv("filename.csv")
```

- Session uploads disappear after the session ends. Re-upload or use a direct CSV URL.

OPEN + SUBMIT A NOTEBOOK

- Open starter link, choose File > Save a copy in Drive, and work in your copy.
- Share a viewable link to turn in the notebook.

From code to data

Load a CSV, inspect it, and answer factual questions using code.

THE BIG IDEA

Start with the unit of observation: what does one row represent?

PYTHON SKILLS

read_csv · head · info · describe · shape · select columns · filter · sort_values · value_counts

CLASS FLOW • 90 MINUTES

0:00–0:10	Setup check: run the Week 0 test cell.
0:10–0:25	Introduce DataFrames through lists of dictionaries.
0:25–0:50	Load and inspect the practice dataset. Identify each row.
0:50–1:15	Pair work: five guided filtering and sorting questions.
1:15–1:30	Share answers and resolve any differing results.

HOMEWORK Five factual questions, each answered with one code cell and one English sentence.

Cleaning and building new columns

Fix messy data and construct calculated columns for fair comparisons.

THE BIG IDEA

Raw totals can mislead. Compare rates, percentages, or per-pupil figures.

PYTHON SKILLS

isna · dropna · fillna · astype · drop_duplicates · rename · .str methods · arithmetic columns · % change

CLASS FLOW • 90 MINUTES

0:00–0:10	Two students review their previous answers.
0:10–0:35	Fix missing data, text numbers, duplicates, and names.
0:35–0:55	Create per-unit and rate columns; compare rankings.
0:55–1:20	Pairs clean a file with at least four planted issues.
1:20–1:30	Debate when dropping rows may bias results.

HOMEWORK Clean a messy dataset and justify every cleaning decision in one sentence.

Summarizing and comparing

Summarize groups, combine datasets, and choose honest statistics.

THE BIG IDEA

Mean and median tell different stories. Check the spread and outliers.

PYTHON SKILLS

groupby · agg · merge · mean · median · std · min · max · quantile · outlier checks

CLASS FLOW • 90 MINUTES

0:00–0:10	Compare students' cleaning decisions.
0:10–0:35	Use groupby + agg to create a category table.
0:35–0:50	Predict and test the effect of an outlier.
0:50–1:05	Merge a second table; verify row counts.
1:05–1:30	Compare two or three groups and explain the numbers.

HOMEWORK Submit one summary table and a paragraph citing specific numbers.

Choose the chart that answers the question

Use matplotlib or df.plot, then label clearly and avoid visual distortion.

BAR



How do categories compare?

HISTOGRAM



How is a variable distributed?

LINE



How did something change over time?

SCATTER



Do two variables move together?

MAKE EVERY CHART HONEST

Use meaningful titles that state findings, clear axis labels, readable units, sorted bars, and non-misleading axis ranges.

Build charts that communicate a finding

The capstone dataset is released at the end of class.

CLASS FLOW • 90 MINUTES

0:00–0:10	Review a student summary table and paragraph.
0:10–0:40	Demo bar, histogram, line, and scatter charts.
0:40–0:55	Bad-chart clinic: correct truncation, missing units, or clutter.
0:55–1:15	Pairs convert their summary table into a finding-led chart.
1:15–1:30	Launch capstone: distribute data, starter questions, and rubric.

HOMEWORK Explore capstone data and submit a question plus one sentence on its feasibility.

Support the claim. Name the limits.

Build a report that says what the evidence shows and what it cannot prove.

CORE IDEA

Correlation does not establish causation.

- A hidden confounder may affect both variables.
- Small samples may produce chance patterns.

Skill focus: corr, scatter plots, confounders, supported claims, report structure

REPORT ANATOMY

01	**Question**	What are you testing?
02	**Data**	What is each row? How did you clean it?
03	**Findings**	Which tables and charts support the answer?
04	**Claim**	Which exact numbers support it?
05	**Limitations**	What alternative explanations remain?

Draft, review, and validate capstone claims

Individual instructor check-ins are the most important work session.

0:00–0:25	Use practice examples: correlation, confounders, small samples.
0:25–0:40	Read a model report: question, data, findings, claim, limits.
0:40–0:50	Rate five sample claims: supported, overstated, unsupported.
0:50–1:25	Capstone work time; check each question and first chart.
1:25–1:30	Explain presentation logistics.

HOMEWORK Finish the written report and prepare a 5-minute presentation.

Present the story your data can support

Five-minute presentation, two-minute Q&A, and a completed written report.

PRESENTATION DAY • 90 MINUTES

0:00–0:05	Set the order and explain peer feedback.
0:05–1:15	Presentations + questions (≈10 students).
1:15–1:30	Reflect: what was surprising, and what comes next?

For groups larger than 10: split into two rooms or use 4-minute presentations.

PRESENTATION CONTENT

1. Your analytical question
2. One or two key charts
3. The main claim with numbers
4. One meaningful limitation

Peer feedback: one clear finding + one open question

CAPSTONE

The five parts of the final report

Submit a Colab notebook with written sections or a 2–3 page document.

01	QUESTION	One clear question that the data can answer.
02	DATA	Source, unit of observation, cleaning decisions, and reasons.
03	FINDINGS	At least two summary tables and two finding-titled charts.
04	CLAIM	A main conclusion supported by specific numbers.
05	LIMITATIONS	What the data cannot show and alternative explanations.

Grading rubric • 20 points

Five dimensions, each scored at 4, 2, or 0 points.

DIMENSION	4 POINTS	2 POINTS	0 POINTS
Question clarity	Specific and answerable	Vague or partly answerable	No clear question
Correctness of analysis	Code runs; fair, normalized comparisons	Minor errors or unfair comparison	Major errors change conclusion
Chart quality	Right type; labeled, honest axes; finding title	Missing labels or weak chart choice	Misleading or unreadable
Claims match evidence	Every claim cites numbers; no overreach	Missing numbers or overstatement	Unsupported claims
Limitations	Specific gaps + alternative explanations	Generic need for more data	Missing

Score each dimension using the descriptors above. Maximum score: 20.

Use real Washington K–12 data

The source curriculum recommends public OSPI datasets for both practice and capstone.

WEEKS 1–4 • PRACTICE

Report Card Enrollment

Simple structure. Clear categories. Suitable for filtering, grouping, and repeated practice.

REUSE ONE DATASET

WEEKS 4–6 • CAPSTONE

Graduation or Assessment + finance

Merge district-level outcomes with finance data to investigate spending, graduation, and equity.

RELEASE AT WEEK 4

WHERE TO FIND DATA	OSPI data portal → data.wa.gov → search “ReportCard” (typically published by school year)

Offer a three-level capstone question ladder

Students may select a starter question or propose one of their own.

EASIER **Which 10 districts have the highest and lowest graduation rates, and how large is the gap?**

Rank and calculate differences.

MEDIUM **Do graduation rates differ between small and large districts when grouped by enrollment?**

Group, summarize, and compare.

HARDER **Is per-pupil spending related to graduation rate, and what else could explain the pattern?**

Compare, visualize, and discuss confounders.

Before teaching, prepare the data and notebooks

Keep enough complexity for learning without overwhelming the students.

INSTRUCTOR CHECKLIST

- ○ Create a shared Drive folder, with one subfolder per week.
- ○ Set starter notebooks to view-only, so students make copies.
- ○ Run every starter notebook in a fresh session the day before.
- ○ Host datasets at a direct CSV link, such as a public GitHub file.
- ○ Trim large OSPI data to district-level rows and 10–15 columns.
- ○ Leave some real issues in capstone data: suppressed values, text numbers.
- ○ Pre-test all three starter questions for answerability.

TEACHING PRINCIPLE

Make every step build toward the capstone.

Give each student a question they can answer, and require every conclusion to match the evidence.