Python Data Analysis Projects PDF

Python data analysis projects teach you to turn a raw CSV into insight with pandas. This page gives six project briefs: sales trends, customer segments, survey analysis, delivery delays, a movie ratings study and a budget tracker. Each has a question, steps, pandas hints and a way to write it up for a portfolio.

Key takeaways

- Each project starts from a question, not from a dataset.
- The common steps are load, inspect, clean, explore, answer, chart and write up.
- Use public or made-up data and say where it came from.
- A short README with the question and findings makes the project readable.
- Six small finished projects beat one unfinished large one.

What is Python data analysis projects?

A Python data analysis project is a small piece of work where you load a dataset, clean it, explore it, answer a clear question with pandas and charts, and explain the result in writing.

What steps does every project follow?

1. **Question:** write one sentence you want to answer.
2. **Load and inspect:** `read_csv`, `head`, `info`, `describe`.
3. **Clean:** types, missing values, duplicates, inconsistent labels.
4. **Explore:** groups, distributions, trends.
5. **Answer:** a table or chart that responds to the question.
6. **Write up:** findings, limits and next steps in plain language.

Project 1: monthly sales trends

Question: which months and products drive revenue? **Data:** order date, product, quantity, price. **Steps:** create a revenue column, convert dates, resample by month, rank products. **Pandas hints:** `to_datetime`, `groupby`, `resample`. **Chart:** a line for revenue by month and a bar for top products. **Write-up:** describe the peak month and one possible reason, noting that the reason is a hypothesis.

Project 2: customer segments

Question: which customers are most valuable? **Data:** customer ID, order date, amount. **Steps:** compute recency, frequency and monetary value per customer, then group into segments with simple rules or quantiles. **Pandas hints:** `groupby`, `agg`, `qcut`. **Chart:** a bar of customers per segment. **Write-up:** suggest one action per segment.

Project 3: survey analysis

Question: what do respondents say about satisfaction? **Data:** a survey with scales and free text. **Steps:** clean category labels, compute shares, compare groups, count common words in comments. **Pandas hints:** `value_counts`, `crosstab`. **Chart:** stacked bars by group. **Write-up:** report sample size and warn about small groups.

Project 4: delivery delays

Question: which routes or days have the most late deliveries? **Data:** ordered date, promised date, delivered date, city. **Steps:** compute delay in days, flag late orders, compare by city and weekday. **Pandas hints:** date subtraction, `dt.day_name()`. **Chart:** a heatmap of late share by city and weekday. **Write-up:** suggest where to investigate first.

Project 5: movie ratings study

Question: do longer movies get higher ratings? **Data:** title, genre, runtime, rating. **Steps:** clean runtime, bin it, compare average rating per bin, check correlation. **Pandas hints:** `cut`, `corr`. **Chart:** a scatter plot with a trend line. **Write-up:** point out that correlation is not causation.

Project 6: personal budget tracker

Question: where does my money go? **Data:** date, category, amount (made up). **Steps:** group by month and category, compute savings rate, flag unusual spends. **Pandas hints:** `pivot_table`, `rolling`. **Chart:** a stacked bar by month. **Write-up:** set one realistic goal.

How do I present a project?

Create a folder with a README, the notebook, a small sample of the data and a chart image. The README should have the question, the data source, steps, key findings in three bullets, limits and how to run the notebook. Keep code readable with clear variable names and short comments.

What mistakes should I avoid?

Skipping the question, copying a tutorial without changing the data, ignoring missing values, showing many charts with no message and publishing private data. Check licences for any public dataset.

How do I choose between projects?

Pick the project closest to the job you want. For a business analyst role choose sales trends and customer segments; for operations choose delivery delays; for marketing choose survey analysis. Finish two before starting a third, and make sure each project has a different type of question so that your portfolio shows range.

What does a good notebook look like?

It starts with the question and a short summary of findings, then shows the steps in order, with a text cell before each code cell explaining why. Charts have titles and labels. The notebook runs from top to bottom without errors, and any file paths are relative so that others can run it. Remove unused code and keep outputs short.

How do I check my numbers?

Compare totals after each transformation, test a few rows by hand, and confirm that the number of rows after a merge is what you expect. Write these checks as small assertions in the notebook. A reviewer who sees checks will trust the result more.

How can I extend a project?

Add a second dataset, build a small dashboard, schedule the notebook to run weekly or write a short blog post about what you learned. Each extension is a good talking point in an interview.

How should I name and organise files?

Use a folder per project with `data`, `notebooks` and `outputs` subfolders, and name files with lowercase words and underscores. Keep raw data unchanged in `data` and save cleaned data under a new name, so that anyone can repeat your steps from the start.

Frequently asked questions

What are good Python data analysis projects for beginners?
Sales trends, customer segments, survey analysis, delivery delays, a ratings study and a budget tracker are good starting points.

Which library do I need?
pandas for data handling and Matplotlib or Seaborn for charts, plus a notebook editor.

Where can I find datasets?
Use public datasets with a clear licence, or create a small made-up dataset, and always state the source.

How should I present a project?

Publish a folder with a README, the notebook, a data sample and a chart, and describe the question, steps, findings and limits.

How many projects do I need?

Three to five finished, well-explained projects are usually more useful than many unfinished ones.

Sources and further reading

- Python documentation
- pandas documentation
- Matplotlib documentation

About this PDF

These PDFs are study material for learning and revision. They are not a certificate, a degree or a promise of a job, and technology changes, so check official documentation for the latest details. Spot an error? Write to info@uncodemy.com and the trainer team will review it.

More free resources and courses: uncodemy.com | info@uncodemy.com | +91 98183 66550