PYTHON FOR DATA SCIENCE · LESSON 20 OF 29

Data Visualisation I: Matplotlib and Choosing the Right Chart

Pick the right chart for the question and build it with Matplotlib's object-oriented API. Line, bar, histogram, scatter, box and pie charts, subplots, annotations, styling and saving, plus design rules that make charts honest and readable.

Python class 16 | Matplotlib 3.10 | Intermediate | 110 min

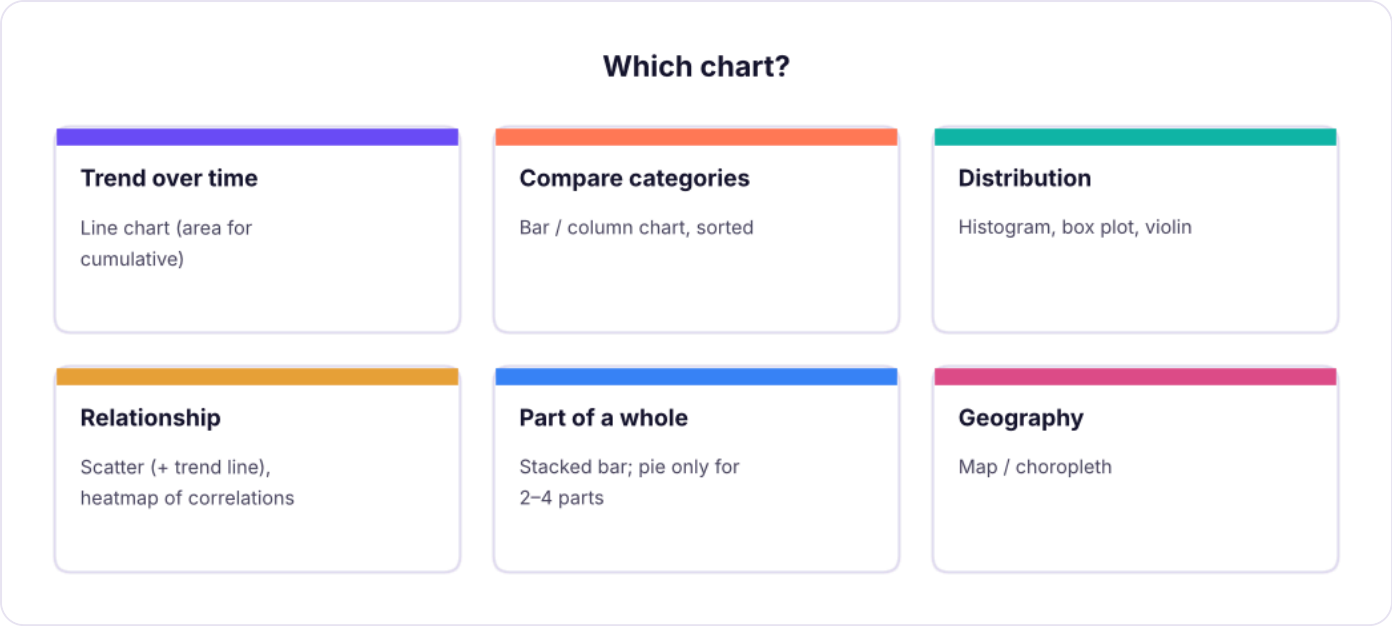
By the end of this lesson you can

- Choose a chart type from the analytical question (trend, comparison, distribution, relationship, composition)
- Build charts with fig, ax = plt.subplots() and customise titles, labels, ticks, legends and colours
- Combine several charts with subplots and annotate key points
- Plot directly from pandas and save figures at publication quality
- Apply design rules: start bars at zero, remove clutter, label directly, use colour with purpose

Datasets used

`superstore_sales.csv` `house_prices.csv`

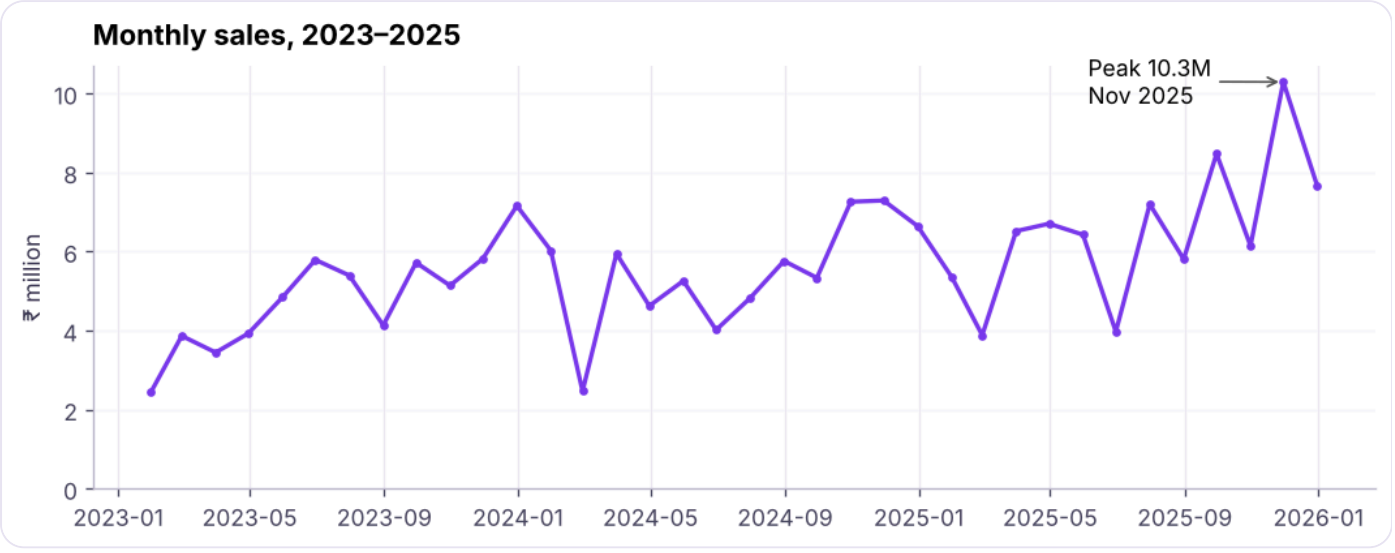
Choose the chart from the question



Start from the question, not the chart type.

The object-oriented API

Matplotlib has two styles. Use the **object-oriented** one: create a **Figure** (the canvas) and one or more **Axes** (plots), then call methods on the axes.



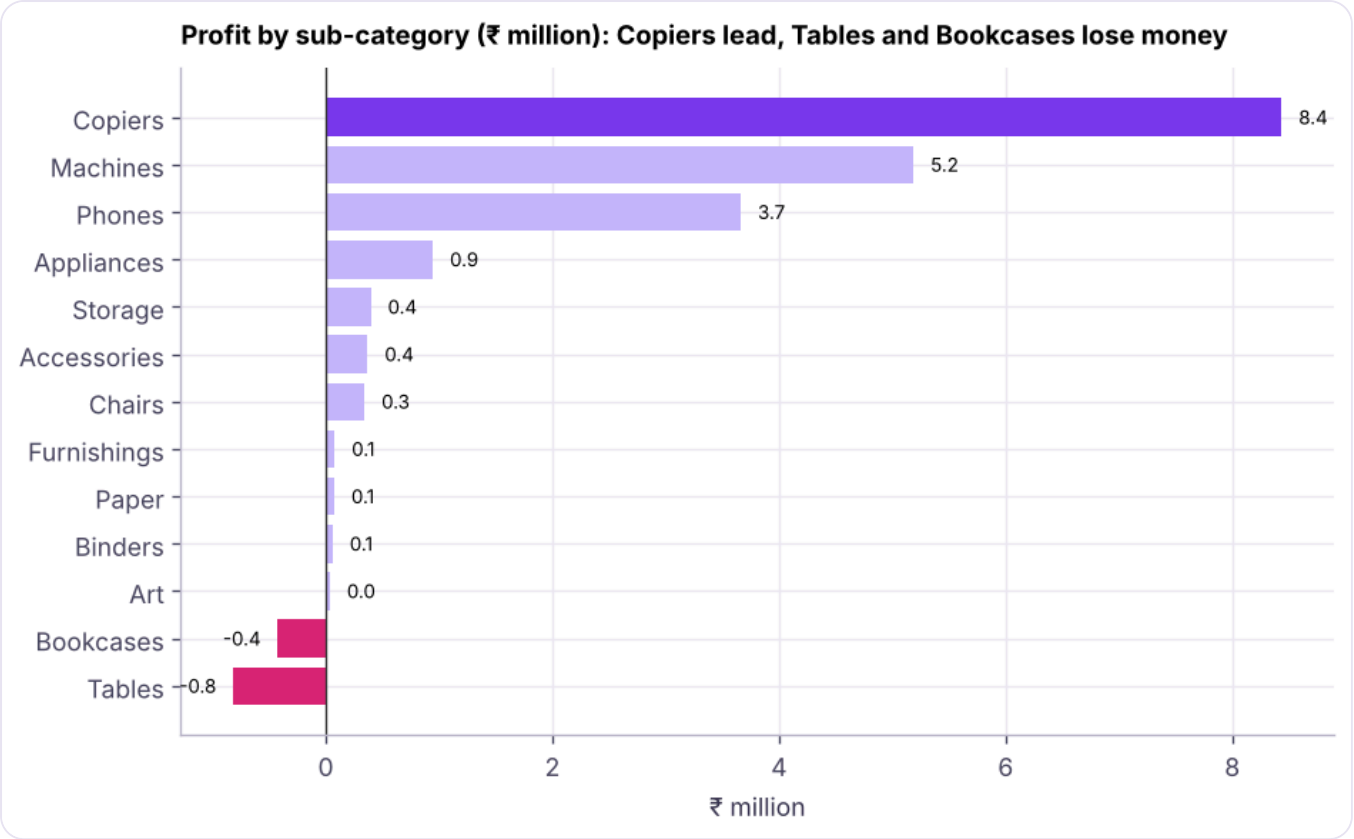
Monthly Superstore sales: a line chart shows the trend and the seasonal peaks.

Python

```python
import matplotlib.pyplot as plt
fig, ax = plt.subplots(figsize=(9, 3.6))   # one figure, one axes
ax.plot(x, y)                              # draw
ax.set_title("..."); ax.set_xlabel("..."); ax.set_ylabel("...")
fig.savefig("chart.png", dpi=200, bbox_inches="tight")
plt.show()                                 # in scripts; notebooks display automatically
```

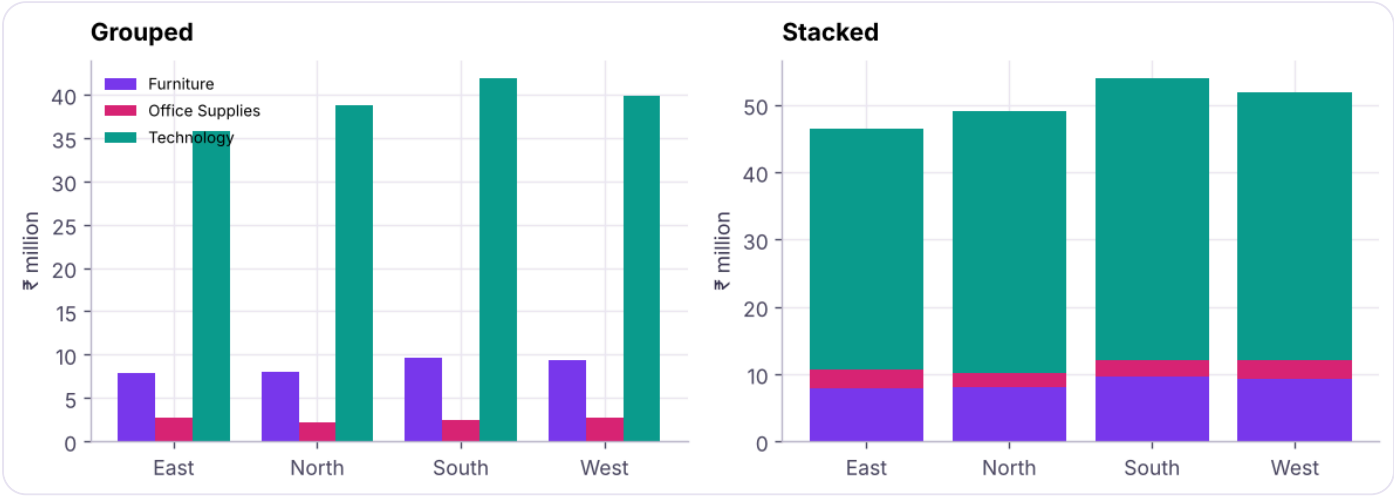
Element	Method
Title, axis labels	`ax.set_title` , `ax.set_xlabel` , `ax.set_ylabel`
Limits, ticks	`ax.set_ylim(0)` , `ax.set_xticks` , `ax.tick_params(rotation=45)`
Legend	`label="..."` in the plot call, then `ax.legend()`
Grid, spines	`ax.grid(axis="y")` , `ax.spines[["top","right"]].set_visible(False)`
Text, arrows	`ax.text` , `ax.annotate`
Reference lines	`ax.axhline(y)` , `ax.axvline(x)` , `ax.axvspan`
Number format	`ax.yaxis.set_major_formatter(...)`

Bar charts



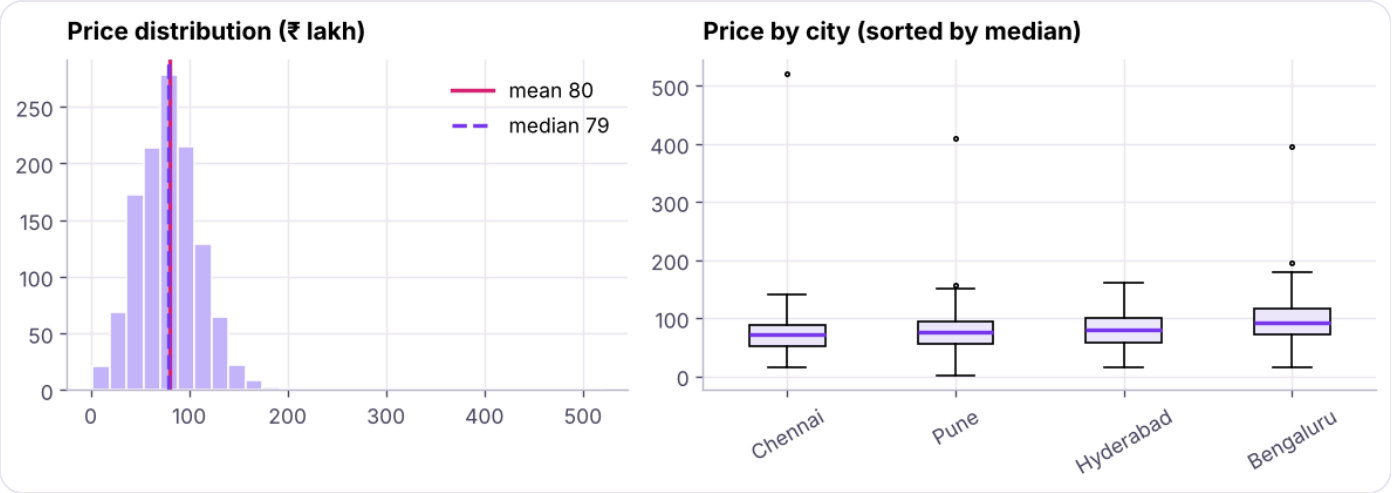
Sorted horizontal bars make category comparisons easy; one highlight colour draws the eye.

Grouped and stacked bars compare two dimensions:



Sales by region and category, as grouped bars (compare within region) and stacked bars (compare totals).

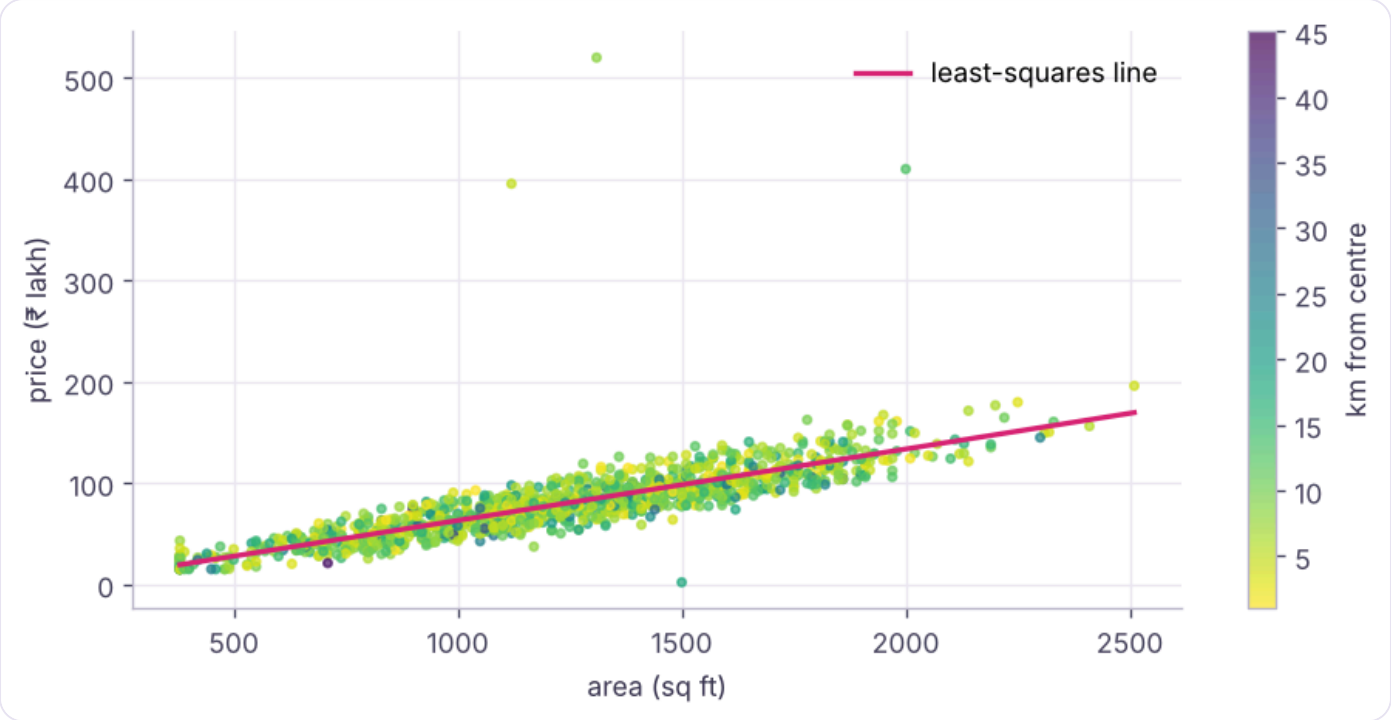
Distributions: histogram and box plot



House prices are right-skewed: the mean sits above the median. Box plots compare distributions across cities.

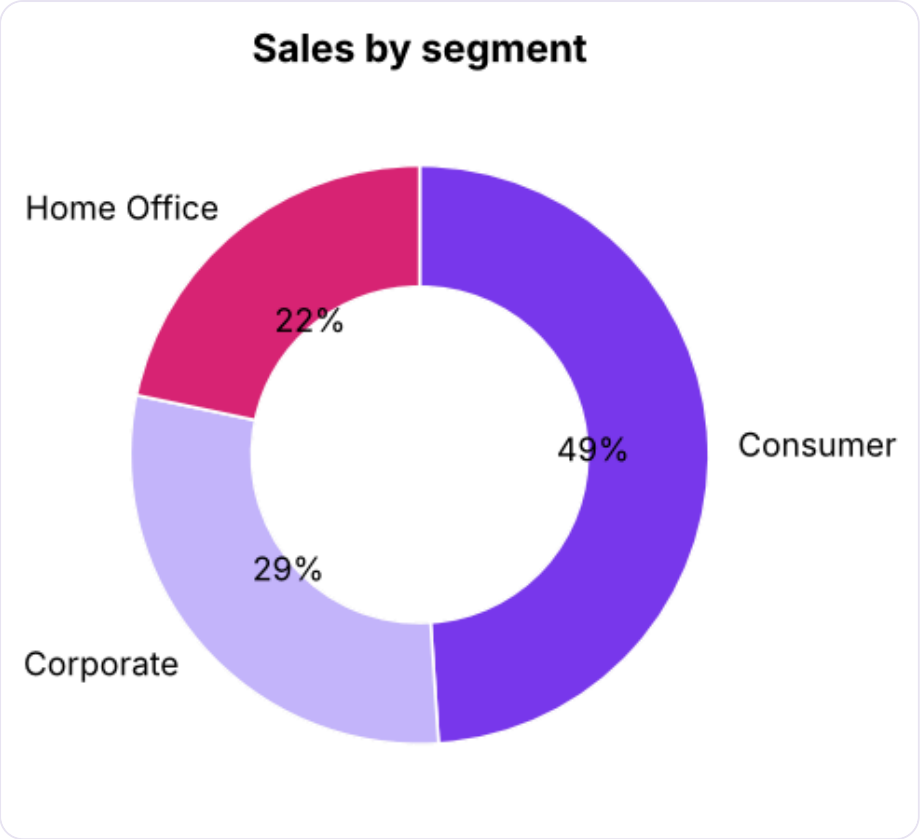
Bins matter: too few hide the shape, too many look noisy. Try `bins="auto"` or a few values.

Relationships: scatter



Larger houses cost more (r ≈ 0.8); colour shows a third variable, the distance to the centre.

Part of a whole

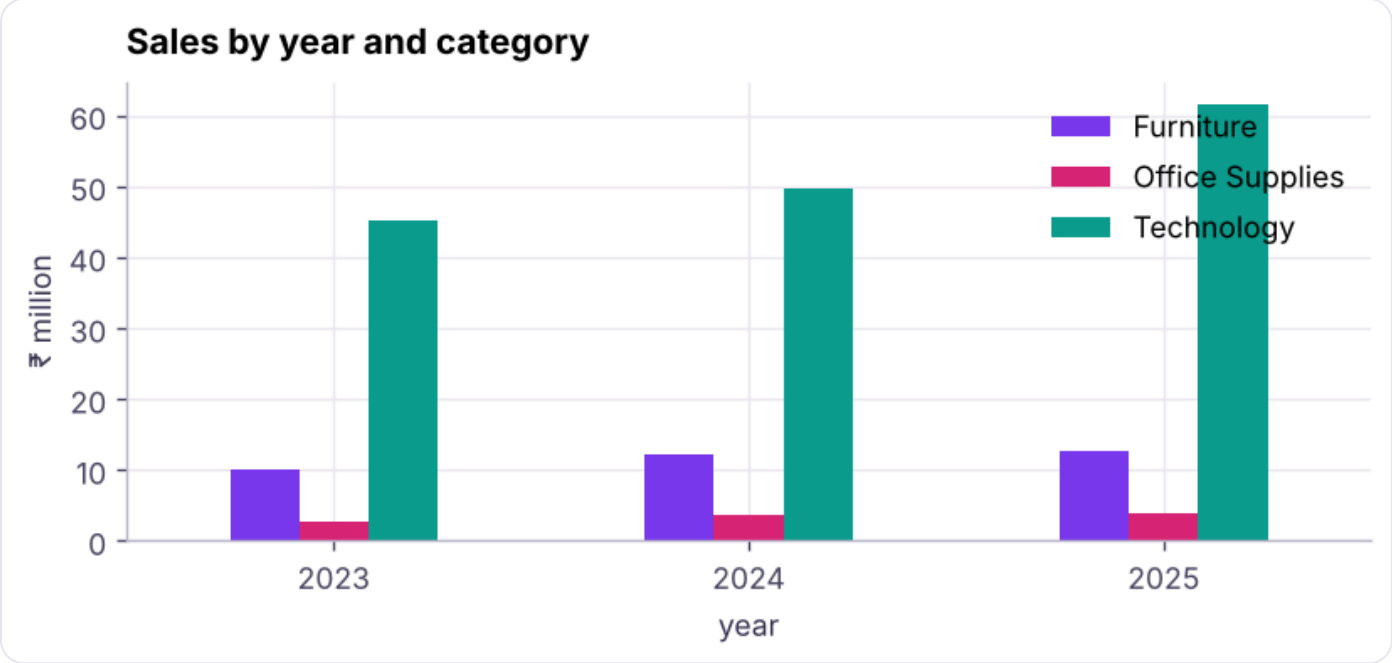


Pie charts work for a few parts; the donut variant leaves room for a label.

For more than 4 parts, or to compare parts across groups, use (stacked) bars.

Plotting from pandas

DataFrames and Series have `.plot()`, a quick wrapper around Matplotlib that returns an `ax` you can keep styling.



One line of pandas plotting, then a little Matplotlib styling.

`kind=` can be `line`, `bar`, `barh`, `hist`, `box`, `scatter` (with `x=`, `y=`), `area`, `pie`.

Subplots dashboards and saving

Python

```
fig, axes = plt.subplots(2, 2, figsize=(11, 7), constrained_layout=True)
axes[0, 0].plot(...); axes[0, 1].barh(...); axes[1, 0].hist(...); axes[1, 1].scatter(...)
fig.suptitle("Sales overview", fontsize=15, weight="bold")
fig.savefig("overview.png", dpi=200, bbox_inches="tight")   # also .svg / .pdf for crisp
vector output
```

Design rules that make charts honest

Rule	Why
Bars start at zero	Truncated bars exaggerate differences
Sort categories by value	Faster comparison
Title states the insight	"Tables lose money" beats "Profit by sub-category"
Label directly, remove chart junk	Fewer eye movements
One highlight colour, grey for context	Colour carries meaning
Colour-blind safe palettes (viridis, Okabe-Ito)	About 1 in 12 men are colour-blind
Same scale when comparing panels	Avoid misleading comparisons
Avoid 3-D and dual axes when possible	They distort perception

Summary

KEY TAKEAWAYS

- Pick the chart from the question: trend → line, comparison → sorted bar, distribution → histogram/box, relationship → scatter, composition → stacked bar (pie only for a few parts).
- Use `fig, ax = plt.subplots()` and style the axes: titles, labels, limits, spines, legends, annotations.
- `plt.subplots(rows, cols)` builds small multiples; `fig.savefig(..., dpi=200, bbox_inches="tight")` exports.
- pandas `.plot()` is a fast start that returns a Matplotlib axes.
- Honest design: zero-based bars, sorted categories, insight titles, purposeful colour.

Quick quiz

6 questions · tap an answer to check it

1. Which chart best shows a trend over 36 months?

A Pie

B Line

C Histogram

D Box plot

Answer: B. Lines connect ordered time points.

2. Which chart shows the distribution of one numeric variable?

A Scatter

B Histogram

C Pie

D Stacked bar

Answer: B. Box and violin plots also work.

3. What does `fig, ax = plt.subplots(1, 2)` create?

A One chart

B One figure with two axes side by side

C Two separate windows

D A 2 × 2 grid

Answer: B. `ax` becomes an array of two axes.

4. Why should bar charts start at zero?

A Matplotlib requires it

B Bar length encodes value; a cut axis exaggerates differences

C It looks better

D It saves space

Answer: B. Line charts can zoom; bars should not.

5. When is a pie chart acceptable?

A For 12 categories

B For a few parts of one whole, when exact comparison is not needed

C For trends

D Never

Answer: B. Otherwise use bars.

6. How do you save a sharp image without cut-off labels?

A `plt.show()`

B `fig.savefig("c.png", dpi=200, bbox_inches="tight")`

C Screenshot

D `fig.close()`

Answer: B. SVG/PDF give vector output.

EXERCISE

Chart practice

(1) Line chart of monthly profit with a horizontal line at zero. (2) Sorted horizontal bar of sales by state (top 10). (3) Histogram of `discount`. (4) Scatter of `discount` vs `profit` (what do you see?). (5) A 2 × 2 dashboard of these four charts saved as PNG.

Show solution (worked answer)

Python

```
mp = s.set_index("order_date")["profit"].resample("ME").sum()
fig, axes = plt.subplots(2, 2, figsize=(11, 7), constrained_layout=True)
axes[0, 0].plot(mp.index, mp / 1e6); axes[0, 0].axhline(0, color="grey")
st = s.groupby("state")["sales"].sum().nlargest(10).sort_values(); axes[0,
1].barh(st.index, st / 1e6)
axes[1, 0].hist(s.discount, bins=9)
axes[1, 1].scatter(s.discount, s.profit, s=4, alpha=.3)   # profit falls as discount
rises
fig.savefig("dashboard.png", dpi=150, bbox_inches="tight")
```

INTERVIEW CORNER

Questions you may be asked

Q: How do you choose a chart? Start from the question (trend, comparison, distribution, relationship, composition), the number of variables and the audience, then remove anything that does not help the message.

Q: What makes a chart misleading? Truncated bar axes, inconsistent scales, 3-D effects, cherry-picked ranges, dual axes with arbitrary scaling, and areas that do not match values.

Q: Matplotlib vs Seaborn vs Plotly? Matplotlib is the low-level foundation with full control; Seaborn adds statistical charts with good defaults on top; Plotly makes interactive web charts.

Practise more online

Free, trusted places to test yourself on this topic. Links open the original site in a new tab.