DATA ANALYST ROADMAP

120 concepts to become a data analyst

8 areas, 15 concepts each. Every concept explained in one line, the order to study each area, and two practice questions to test yourself.

How to use it: one area at a time, tick each concept only when you can explain it out loud in under a minute, then answer the two practice questions without notes.

@chalkesahil · chalkesahil.ai

AREA 01 / 08

Excel

Study order: Every analyst job still opens Excel on day one; get fast at it first.

01	**Cell references**	Relative, absolute and mixed references with $.
02	**Sorting and filtering**	Find the rows you need in seconds.
03	**SUMIFS and COUNTIFS**	Totals and counts with conditions.
04	**XLOOKUP**	Pull a value from another table by key.
05	**INDEX MATCH**	The older lookup you'll still see in shared sheets.
06	**IF and nested logic**	Label rows based on rules.
07	**Text functions**	LEFT, MID, TRIM and TEXTSPLIT to clean text.
08	**Date functions**	Months, weeks and days between dates.
09	**Pivot tables**	Summarise thousands of rows by any column.
10	**Charts**	Pick the chart that answers the question.
11	**Conditional formatting**	Make outliers jump out.
12	**Data validation**	Dropdowns that stop bad input.
13	**Power Query**	Repeatable cleaning steps instead of manual edits.
14	**Named ranges and tables**	Formulas that grow with the data.
15	**Keyboard shortcuts**	Speed is part of the job.

Practice

1. Find each region's top-selling product from a sales sheet.
2. Two sheets have customer IDs. Find who is in one but not the other.

AREA 02 / 08

SQL

Study order: Practise on a real sample database every day; analysts live in SQL.

01	**SELECT and WHERE**	Pick columns, filter rows.
02	**ORDER BY and LIMIT**	Top N and bottom N.
03	**Joins**	Inner, left and when each drops rows.
04	**GROUP BY and HAVING**	Aggregate per group, filter after grouping.
05	**Aggregate functions**	SUM, AVG, COUNT, MIN, MAX.
06	**CASE WHEN**	Bucket values into labels.
07	**Subqueries**	Use one query's result inside another.
08	**CTEs**	Break long queries into readable steps.
09	**Window functions**	Running totals and ranks without collapsing rows.
10	**LAG and LEAD**	Compare a row with the previous one.
11	**Date handling**	Truncate to day, week, month.
12	**NULL handling**	COALESCE and why NULL breaks comparisons.
13	**DISTINCT and duplicates**	Count unique users correctly.
14	**UNION**	Stack results from two queries.
15	**Reading query errors**	Fix the query from the error message.

Practice

1. Find month-over-month revenue growth.
2. Find users who ordered in January but not in February.

AREA 03 / 08

Statistics

Study order: Learn the ideas with real examples, not formulas first.

01	**Mean, median and mode**	Which average to report and why.
02	**Spread**	Range, variance and standard deviation.
03	**Percentiles**	What the 90th percentile delivery time means.
04	**Distributions**	Normal, skewed and long-tail data.
05	**Outliers**	Find them and decide whether to keep them.
06	**Sampling**	Why a small sample can mislead you.
07	**Correlation**	Two things moving together, not causing each other.
08	**Causation**	What it takes to claim one thing caused another.
09	**Probability basics**	Chances, independent events, conditional probability.
10	**Hypothesis testing**	Is a difference real or just noise?
11	**p-values**	What they say and what they don't.
12	**Confidence intervals**	A range instead of a single guess.
13	**A/B testing**	Compare two versions with a fair split.
14	**Simpson's paradox**	A trend that flips when you split the data.
15	**Linear regression**	A first look at predicting one number from others.

Practice

1. Average order value went up but revenue fell. How?
2. An A/B test shows a 2% lift. How do you know it's real?

AREA 04 / 08

Python and pandas

Study order: Use Python when Excel gets too slow or too manual.

01	**Python basics**	Variables, lists, dicts and loops.
02	**Functions**	Reuse the same steps across reports.
03	**Jupyter notebooks**	Code, output and notes in one place.
04	**Reading files**	CSV, Excel and JSON into a DataFrame.
05	**Selecting data**	loc, iloc and boolean filters.
06	**groupby**	Pivot-table logic in code.
07	**merge**	Joins between DataFrames.
08	**Handling missing values**	fillna, dropna and when to use each.
09	**apply and vectorising**	Transform columns without slow loops.
10	**Dates in pandas**	Parse dates, resample by week or month.
11	**pivot_table**	Wide summaries in one line.
12	**Matplotlib and Seaborn**	Quick charts from a DataFrame.
13	**Reading from SQL**	Pull query results straight into pandas.
14	**Exporting results**	Save clean output for others to use.
15	**Automating a report**	Turn a weekly manual task into a script.

Practice

1. Merge orders and customers, then find revenue by city.
2. Automate a weekly Excel report that takes you an hour by hand.

AREA 05 / 08

Data cleaning

Study order: Most of the job is cleaning; get good at it and you stand out.

01	**Spotting bad data**	Profile every column before you trust it.
02	**Duplicates**	Find and remove repeated records.
03	**Missing values**	Fill, flag or drop, and say which you did.
04	**Inconsistent text**	Mumbai, mumbai and Bombay are one city.
05	**Data types**	Numbers stored as text break every sum.
06	**Date formats**	DD/MM vs MM/DD mistakes.
07	**Outlier rules**	Decide a rule and apply it every time.
08	**Joining messy keys**	Trim, lowercase and match IDs.
09	**Validation checks**	Totals should match the source system.
10	**Units and currencies**	Never add rupees to dollars.
11	**Time zones**	Orders at midnight can land on the wrong day.
12	**Documenting steps**	Write down every cleaning choice.
13	**Reproducibility**	Anyone should get the same result from your steps.
14	**Raw vs clean copies**	Never overwrite the original data.
15	**Sanity checks**	Does the number make sense before you share it?

Practice

1. A dashboard shows 3x the real users. Where do you look first?
2. Clean a customer list with mixed phone number formats.

AREA 06 / 08

Power BI and Tableau

Study order: Pick one tool and build three dashboards on public data.

01	**Connecting data**	Excel, SQL and CSV sources.
02	**Data model**	Tables and relationships between them.
03	**Star schema basics**	Facts in the middle, dimensions around them.
04	**Measures vs columns**	Calculated on the fly vs stored per row.
05	**DAX basics**	CALCULATE, SUM and FILTER in Power BI.
06	**Calculated fields**	The Tableau version of a measure.
07	**Choosing charts**	Bars for comparing, lines for time.
08	**Filters and slicers**	Let users ask their own questions.
09	**Drill-down**	From country to state to city.
10	**KPI cards**	The three numbers a manager checks first.
11	**Dashboard layout**	Most important number top left.
12	**Colour with purpose**	Highlight one thing, grey the rest.
13	**Refresh schedules**	Keep the dashboard up to date.
14	**Sharing and access**	Who sees which data.
15	**Storytelling**	Lead with the answer, then the evidence.

Practice

1. Build a sales dashboard a store manager can read in 10 seconds.
2. When would you use a table instead of a chart?

AREA 07 / 08

Business metrics

Study order: Learn the numbers a business runs on; this is what interviews test.

01	**Revenue and profit**	What the business earns and keeps.
02	**Growth rates**	Month over month and year over year.
03	**Conversion rate**	Visitors who actually buy.
04	**Funnels**	Where users drop off step by step.
05	**Retention**	How many users come back.
06	**Cohort analysis**	Group users by start month and track them.
07	**Churn**	Customers you lost and why.
08	**Customer acquisition cost**	What it costs to win one customer.
09	**Lifetime value**	What one customer is worth over time.
10	**Average order value**	Revenue per order.
11	**Active users**	Daily and monthly active users.
12	**North star metric**	The one number a team optimises.
13	**Root cause analysis**	Why did the metric move?
14	**Guesstimates**	Estimate a market size step by step.
15	**Writing insights**	What happened, why, and what to do next.

Practice

1. Orders dropped 20% yesterday. How do you find out why?
2. How many cups of chai are sold in Mumbai every day?

AREA 08 / 08

AI tools

Study order: AI now drafts the SQL and the charts; your job is checking it is right.

01	**ChatGPT for SQL**	Draft queries, then verify every join.
02	**Explaining errors**	Paste an error, get a fix, understand it.
03	**Copilot in Excel**	Ask questions of a sheet in plain English.
04	**Power BI Copilot**	Generate visuals and summaries from your model.
05	**Python with AI**	Write pandas code faster, review it line by line.
06	**Prompting for analysis**	Give the table schema and the question.
07	**Checking AI output**	AI makes confident mistakes in numbers.
08	**Data privacy**	Never paste customer data into public tools.
09	**Summarising results**	Turn findings into a short note for managers.
10	**Text data**	Classify reviews and feedback at scale.
11	**Automating reports**	Scripts plus AI to write the weekly summary.
12	**When not to use AI**	Small tasks you understand faster by hand.
13	**Portfolio project**	One end-to-end analysis on a public dataset.
14	**Resume bullets**	Show impact with numbers, not tool lists.
15	**Interview prep**	Practise SQL and case questions out loud.

Practice

1. AI wrote a query that returns double the revenue. How do you check it?
2. Present a finding to a manager in under one minute.