

Data analyst

Roadmap

1) Foundations (What to understand first)

- **Problem mindset: Business questions → data questions → actionable insights.**
 - **Basic statistics: mean, median, variance, distributions, hypothesis testing, confidence intervals, p-values.**
 - **Data literacy: data types, data cleaning, missing values, outliers, basic ETL concepts.**
 - **SQL basics: SELECT, WHERE, JOIN, GROUP BY, aggregates.**
 - **Excel / Google Sheets: formulas, pivot tables, VLOOKUP/XLOOKUP, charts.**
-

2) Core Technical Skills (must-have)

- **SQL (essential): querying, joins, window functions, CTEs, performance basics.**
- **Python (or R): for Python — pandas, numpy, matplotlib, seaborn; basics of scripting and functions. (R is alternate for statistics-heavy paths.)**
- **Data Visualization: Tableau / Power BI (interactive dashboards), plus charting libraries in Python.**
- **Statistics & A/B testing: designing tests, analysing results, interpreting lift and significance.**

- **Data cleaning / ETL:** string cleaning, date handling, type conversion, merging datasets.
 - **Version control basics:** Git (commits, branches) for code traceability.
 - **Basic machine learning concepts (optional but valuable):** supervised vs unsupervised, simple regression/classification to build predictive features.
-

3) Tools & Tech Stack (practical set)

- **Databases:** MySQL, PostgreSQL, BigQuery or Redshift (cloud).
 - **Notebook & IDE:** Jupyter Notebook / VSCode.
 - **BI Tools:** Tableau or Power BI (pick one deeply).
 - **Python libs:** pandas, numpy, matplotlib, seaborn, scikit-learn (intro).
 - **Cloud basics:** working knowledge of GCP/AWS/Azure (how to run SQL queries on BigQuery / Athena).
 - **Spreadsheet:** Excel (advanced) / Google Sheets.
 - **Collaboration & tracking:** GitHub, Google Drive, Slack.
-

4) Learning / Practice Roadmap (6 months plan — self-study + projects)

Month 0 (setup)

- **Install Python, set up GitHub, create LinkedIn, choose BI tool trial (Tableau/Power BI).**
- **Learn basic SQL queries and Excel pivot tables.**

Month 1 (data wrangling & SQL)

- Deep SQL: joins, aggregations, window functions.
- Practice on sample datasets (Kaggle/public).
- Small project: Sales summary dashboard using SQL + Excel.

Month 2 (Python & EDA)

- Learn pandas + numpy; practice data cleaning (missing, duplicates).
- Exploratory Data Analysis (EDA) — plots, distributions, correlations.
- Project: EDA + cleaned dataset + EDA notebook.

Month 3 (Data viz & dashboards)

- Learn Tableau/Power BI: calculated fields, filters, parameters, publishing dashboards.
- Build 1–2 interactive dashboards (marketing funnel, sales by region).
- Optimize dashboard UX (filters, story).

Month 4 (Statistics & A/B testing)

- Hypothesis testing, t-tests, chi-square, confidence intervals.
- Design & interpret an A/B test.
- Project: A/B test analysis notebook.

Month 5 (Advanced SQL + basic ML)

- Window functions, performance tuning, CTEs, analytical queries.
- Intro to regression/classification with scikit-learn (feature engineering).
- Project: Predict churn probability or sales forecasting prototype.

Month 6 (Portfolio, case studies & interview prep)

- **Finalise 3 polished projects (SQL report, EDA notebook, interactive dashboard).**
 - **Write 1-page case studies for each (problem, approach, tools, results, business impact).**
 - **Mock interviews, common questions, take-home assignment practice.**
-

5) Real Projects to Add to Portfolio (Practical ideas)

- **Sales Performance Dashboard: region, product, channel breakdown + KPIs.**
 - **Customer Churn Analysis: EDA, feature importance, simple model + action recommendations.**
 - **Marketing Funnel & A/B Test: compute conversion rates, run significance tests, recommend next steps.**
 - **Product Usage Analysis: cohort analysis, retention curves, LTV estimation.**
 - **Automated Report: scheduled SQL + Python script to generate weekly KPI report.**
-

6) How to Present Work (Portfolio & Resume)

- **GitHub repo: clean notebooks + README.**
- **Dashboard links: Tableau Public / Power BI reports (if allowed).**
- **One-page case study per project: goal, dataset, approach (SQL/Python/Tableau), key metrics & business impact.**

- **Resume bullets: quantify impact (e.g., “Improved report generation time by 60% via SQL automation”; “Identified churn drivers leading to 12% retention uplift recommendation”).**
 - **LinkedIn: short project posts with screenshots and top insights.**
-

7) Interview & Hiring Prep (common topics)

- **SQL live tests (joins, aggregates, window functions).**
 - **Coding/Notebook: data cleaning tasks in Python.**
 - **Take-home case: analyse dataset + present findings + dashboard.**
 - **Behavioral: explain projects, tradeoffs, how you measured business impact.**
-

8) Metrics & KPIs Data Analysts Own or Track

- **Data quality metrics: % missing, duplication rate, freshness.**
 - **Business KPIs depending on domain: conversion rate, churn rate, retention, average order value, LTV, MRR.**
 - **Dashboard adoption: active users, report load time, time-to-insight.**
-

9) Tips & Best Practices

- **Start small, ship often: finish tiny projects end-to-end.**
- **Document your assumptions and data lineage for every analysis.**

- **Focus on business impact: always answer “so what?” — what should business do next.**
 - **Automate repetitive tasks (scripts, scheduled queries).**
 - **Read others’ notebooks on GitHub/Kaggle to learn idioms.**
 - **Communicate clearly: visualizations + 1–2 slide summary for stakeholders.**
 - **Keep learning: domain knowledge matters (finance, e-commerce, health etc).**
-

10) Next steps I can do for you (pick any)

- **Bana du 6-month personalised weekly study plan with resources & exercises.**
- **Ek sample resume + 3 case-study templates tailored to freshers.**
- **2 portfolio project outlines (with datasets and step-by-step checklist)**

[Techknitia | Online Courses & leaning | techknitia](#)

@Techknitia