Complete Data Analyst Roadmap for Beginners 2025

Your Step-by-Step Guide to Landing Your First Data Analyst Job

---

📊 OVERVIEW

This comprehensive roadmap will guide you from complete beginner to job-ready data analyst in **5 months**. With the rise of AI and machine learning, 2025 is the perfect time to enter this field.

Duration: 5 months
Daily Commitment:

- Part-time learners (working professionals): 2-3 hours daily
- Full-time learners: 4-6 hours daily

Expected Timeline by Background:

- Freshers/Non-IT professionals: 5-6 months
- Career changers with transferable skills: 4-5 months

🎯 PHASE 1: FOUNDATION BUILDING (MONTH 1-2)

Week 1-2: Statistics & Probability

Learning Objectives:

Descriptive Statistics

- Measures of central tendency: Mean, Median, Mode
- Measures of dispersion: Range, Variance, Standard Deviation, Percentiles, Quartiles
- Frequency distributions (Relative and Cumulative)
- Graphical representations: Boxplots, Histograms, Scatterplots
- Outlier detection and treatment
- Correlation and Covariance

Probability

- Basic probability concepts: Sample space, Events
- Conditional Probability
- Bayes Theorem

- Probability Distributions
- Normal Distribution and Standard Normal Distribution
- Empirical Rule

Inferential Statistics

- Confidence Intervals (Z and T distributions)
- Hypothesis Testing
- Null and Alternative Hypotheses
- p-values and significance levels
- Types of tests: t-test, z-test, ANOVA, Chi-square

Resources:

- Video: Statistics Crash Course Tutorial (2.5 hours) https://www.youtube.com/watch?v=S7LvZZNq4ys&t=8317s
- Platform: Khan Academy Statistics & Probability https://www.khanacademy.org/math/statistics-probability
- Practice: Complete 20-30 basic statistics problems

Practical Projects:

1. Calculate descriptive statistics for any dataset (e.g., student grades)
2. Perform hypothesis testing on sample data
3. Create visualizations for data distributions

Time Allocation:

- Theory: 8-10 hours
- Practice: 10-12 hours
- Projects: 5-6 hours

Week 3-4: Excel Fundamentals

Learning Objectives:

Basic Functions

- Mathematical: SUM, AVERAGE, COUNT, MIN, MAX, ROUND
- Text functions: LEFT, RIGHT, MID, LEN, CONCATENATE, TRIM, PROPER

- Date functions: TODAY, DATE, DATEVALUE, EMONTH, EDATE, NETWORKDAYS

Conditional & Lookup Functions

- IF, SUMIF, SUMIFS, COUNTIF, COUNTIFS
- VLOOKUP, HLOOKUP, INDEX-MATCH, XLOOKUP
- AND, OR logical operators

Data Management

- Data Cleaning: Remove duplicates, handle missing values
- Data Validation: Input restrictions, dropdown lists
- Conditional Formatting: Highlight rules, color scales, data bars
- Sorting and Filtering

Advanced Features

- Pivot Tables and Pivot Charts
- Slicers and Timelines
- Power Query basics
- What-If Analysis: Goal Seek, Data Tables, Scenario Manager
- Basic Macros (recording)

Resources:

- Playlist: Excel Complete Tutorial
  https://www.youtube.com/playlist?list=PLUaB-1hjhk8Hyd5NiPQ9CND82vNodlFF5
- Project Tutorial: Sales Dashboard using Excel
  https://www.youtube.com/watch?v=6OMR81faW54&t=58s
- Project Tutorial: End-to-End Excel Project
  https://www.youtube.com/watch?v=gTK5rNhWJyA&t=5s
- Practice Platform: Excel Practice Online
  https://www.excelpracticeonline.com/
- Additional Practice: Excel Easy
  https://www.excel-easy.com/

Practical Projects:

1. Sales Dashboard with Pivot Tables and Charts
2. Budget Tracker with advanced formulas

3. Employee Database with data validation and conditional formatting
4. KPI Dashboard with dynamic charts

Time Allocation:

- Theory & Tutorials: 10-12 hours
- Practice: 12-15 hours
- Projects: 8-10 hours

Week 5-6: SQL Basics

Learning Objectives:

Database Fundamentals

- Understanding databases, tables, rows, columns
- Data types: Numeric, String, Date/Time
- Primary Keys, Foreign Keys, Constraints
- Database normalization basics (1NF, 2NF, 3NF)

Basic SQL Operations

- SELECT statements
- WHERE clause and filtering
- ORDER BY and sorting
- LIMIT clause
- DISTINCT values

Aggregate Functions

- COUNT, SUM, AVG, MIN, MAX
- GROUP BY clause
- HAVING clause

Basic Joins

- INNER JOIN
- LEFT JOIN
- RIGHT JOIN

- Understanding join logic

Resources:

- Video: SQL Crash Course
  https://www.youtube.com/watch?v=On9eSN3F8w0
- Tutorial: Mode SQL Tutorial (Interactive)
  https://mode.com/sql-tutorial/
- Practice: LeetCode SQL 50 Study Plan
  https://leetcode.com/studyplan/top-sql-50/
- Practice: HackerRank SQL
  https://www.hackerrank.com/domains/sql
- Practice: LeetCode SQL Problems
  https://leetcode.com/problemset/database/

Practical Projects:

1. Query a sample database (Northwind or Chinook)
2. Create sales reports using GROUP BY
3. Customer analysis using JOINs
4. Product inventory analysis

Time Allocation:

- Theory: 8-10 hours
- Practice Problems: 15-20 hours (2-3 problems daily)
- Projects: 5-7 hours

Important Note: Start practicing SQL daily from week 5. Consistency is key for SQL mastery.

Week 7-8: Python Basics

Learning Objectives:

Python Fundamentals

- Python syntax and basics
- Variables and data types (int, float, string, boolean)
- Basic operators (arithmetic, comparison, logical)
- Input/Output operations

Data Structures

- Lists: Creation, indexing, slicing, methods
- Tuples: Immutable sequences
- Sets: Unique elements, set operations
- Dictionaries: Key-value pairs, methods
- List comprehensions

Control Flow

- if-elif-else statements
- for loops and while loops
- break, continue, pass statements
- range() function

Functions

- Defining functions
- Parameters and arguments
- Return values
- Lambda functions
- Scope (local vs global)

Error Handling

- try-except blocks
- Common exceptions
- Debugging basics

Introduction to Libraries

- Pandas basics: DataFrames and Series
- NumPy basics: Arrays and basic operations
- Why use these libraries?

Resources:

- Playlist: Python Tutorial by Krish Naik
  https://www.youtube.com/watch?v=bPrmA1SEN2k&list=PLZoTAELRMXVNUL99R4bDlVYsncUNvwUBB

- Video: Python for Data Analysis
  https://www.youtube.com/watch?v=wUSDVGivd-8
- Practice: HackerRank Python
  https://www.hackerrank.com/domains/python
- Practice: LeetCode Easy Problems
  https://leetcode.com/problemset/
- Tutorial: Analytics Vidhya Python Interview Questions
  https://www.analyticsvidhya.com/blog/2024/05/python-coding-interview-questions-for-beginners/

Practical Projects:

1. Simple calculator with functions
2. To-do list application
3. Data cleaning script using Pandas
4. Basic data analysis on CSV file

Time Allocation:

- Theory: 10-12 hours
- Practice: 15-18 hours
- Projects: 6-8 hours

🚀 PHASE 2: INTERMEDIATE SKILLS (MONTH 3)

Week 9-10: Advanced SQL

Learning Objectives:

Subqueries

- Single-row subqueries
- Multiple-row subqueries
- Correlated subqueries
- Subqueries in SELECT, FROM, WHERE

Common Table Expressions (CTEs)

- WITH clause syntax
- Recursive CTEs

- When to use CTEs vs subqueries

Window Functions

- ROW_NUMBER()
- RANK() and DENSE_RANK()
- LEAD() and LAG()
- PARTITION BY clause
- Running totals and moving averages

Advanced Techniques

- CASE statements
- Multiple table JOINs
- Self JOINs
- UNION and UNION ALL
- EXISTS and IN operators

Stored Procedures & Functions

- Creating procedures
- Creating functions
- Variables and parameters

Query Optimization

- Understanding indexes
- Query execution plans
- Performance best practices

Resources:

- Project: SQL and Python End-to-End Project https://www.youtube.com/watch?v=2VMAdlzNuTw&t=1350s
- Project: Airlines Analysis using SQL and Python https://www.youtube.com/watch?v=LcMjsqZiSjY&t=115s
- Documentation: PostgreSQL Window Functions https://www.postgresql.org/docs/current/tutorial-window.html
- Practice: Continue with LeetCode Medium/Hard SQL problems

Practical Projects:

1. Airlines data analysis with window functions
2. E-commerce database analysis with complex JOINs
3. Create a data warehouse schema
4. Employee hierarchy analysis with CTEs

Time Allocation:

- Theory: 8-10 hours
- Practice Problems: 20-25 hours (3-4 problems daily)
- Projects: 8-10 hours

Daily Practice Goal: Solve 2-3 medium-level SQL problems every day.

Week 11-12: Python for Data Analysis

Learning Objectives:

Pandas in Depth

- DataFrames and Series operations
- Reading/Writing data (CSV, Excel, SQL)
- Data selection: loc, iloc, boolean indexing
- Data cleaning: handling missing values, duplicates
- Data transformation: apply, map, replace
- Merging and joining DataFrames
- Grouping and aggregation
- Pivot tables and crosstabs
- Time series data handling

NumPy Mastery

- Array creation and manipulation
- Array indexing and slicing
- Mathematical operations
- Broadcasting

- Linear algebra operations
- Random sampling
- Statistical functions

Data Visualization

- Matplotlib: Line plots, bar charts, scatter plots, histograms
- Seaborn: Statistical visualizations, heatmaps, pair plots
- Customizing plots: colors, labels, legends, styles
- Subplots and figure management

Scipy Basics

- Statistical tests
- Optimization
- Signal processing basics

Resources:

- Project: Data Analysis End-to-End Python Project https://www.youtube.com/watch?v=obJZ1rB7TKc&t=2074s
- Project: Complete Python Project https://www.youtube.com/watch?v=KgCgpCIOkIs&t=1226s
- Documentation: Pandas Official Documentation https://pandas.pydata.org/docs/
- Documentation: NumPy Official Documentation https://numpy.org/doc/
- Documentation: Matplotlib Documentation https://matplotlib.org/

Practical Projects:

1. Exploratory Data Analysis on Titanic dataset
2. Sales analysis with visualizations
3. Housing price data cleaning and analysis
4. Customer segmentation analysis
5. Time series analysis on stock data

Time Allocation:

- Theory & Tutorials: 12-15 hours
- Practice: 18-20 hours
- Projects: 10-12 hours

📈 PHASE 3: VISUALIZATION & ADVANCED ANALYTICS (MONTH 4)

Week 13-14: Business Intelligence Tools (Choose ONE)

Important: Focus on mastering ONE tool (Power BI or Tableau) rather than learning both superficially.

Option A: Power BI

Learning Objectives:

Power Query (Data Preparation)

- Connecting to data sources (Excel, CSV, SQL, Web)
- Data transformation operations
- Merging and appending queries
- Unpivoting and pivoting data
- Creating custom columns
- Data type management

Data Modeling

- Creating relationships between tables
- Star schema and snowflake schema
- Database normalization
- Cardinality and cross-filter direction

DAX (Data Analysis Expressions)

- Calculated columns vs measures
- Basic functions: SUM, AVERAGE, COUNT
- Time intelligence functions
- CALCULATE and FILTER
- Variables in DAX
- Common DAX patterns

Visualization & Dashboard Design

- Chart types and when to use them
- Creating interactive reports
- Slicers and filters
- Drill-through and drill-down
- Bookmarks and buttons
- Mobile layout
- Best practices for dashboard design

Resources:

- Video: Power BI Complete Tutorial (3 hours)
  https://www.youtube.com/watch?v=bQ-HTp-tx40
- Project: HR Analytics Dashboard
  https://www.youtube.com/watch?v=6H4afhQeewU&t=6s
- Project: Blinkit Real-Time Dashboard
  https://www.youtube.com/watch?v=mmxVCFceQgU&t=30s
- Project: End-to-End Data Analyst Project
  https://www.youtube.com/watch?v=tT4V7zguCnc
- Platform: Microsoft Learn Power BI
  https://learn.microsoft.com/en-us/power-bi/

Practical Projects:

1. Sales Performance Dashboard
2. HR Analytics Dashboard (attrition, demographics, performance)
3. Financial KPI Dashboard
4. Customer Analytics Dashboard

Option B: Tableau

Learning Objectives:

Tableau Fundamentals

- Connecting to data sources
- Data pane and data types

- Dimensions vs measures
- Marks and properties
- Sorting and grouping

Calculations

- Calculated fields
- Table calculations
- Level of Detail (LOD) expressions
- Parameters

Visualizations

- Creating various chart types
- Dual-axis charts
- Trend lines and forecasting
- Clustering
- Reference lines and bands

Dashboard Design

- Dashboard layout and sizing
- Action filters
- Navigation and interactivity
- Stories
- Best practices

Advanced Features

- Data blending
- Joins and relationships
- Pivoting data
- Nested LODs
- Dynamic designs
- Extensions

Resources:

- Video: Tableau Complete Tutorial
  https://www.youtube.com/watch?v=j8FSP8XuFyk
- Project: Tableau End-to-End Project
  https://www.youtube.com/watch?v=KlAKAarfLRQ
- Project: Sales Dashboard using Tableau
  https://www.youtube.com/watch?v=dahrmqT5GD4
- Platform: Tableau Public (for hosting dashboards)
  https://public.tableau.com/
- Inspiration: Tableau Public Gallery
  https://public.tableau.com/app/discover

Practical Projects:

1. Sales Performance Dashboard
2. COVID-19 Data Visualization
3. E-commerce Analytics Dashboard
4. Marketing Campaign Dashboard

Time Allocation (for chosen tool):

- Theory & Tutorials: 12-15 hours
- Practice: 15-18 hours
- Projects: 10-12 hours

Week 15-16: Advanced Analytics

Learning Objectives:

Time Series Analysis

- Components: Trend, Seasonality, Cyclical, Irregular
- Stationarity testing
- Moving averages
- Exponential smoothing
- ARIMA basics
- Seasonal decomposition
- Forecasting techniques

Regression Analysis

- Simple linear regression
- Multiple linear regression
- Assumptions of linear regression
- Model evaluation: R^2, RMSE, MAE
- Residual analysis
- Feature selection

A/B Testing

- Hypothesis formulation
- Sample size calculation
- Statistical significance
- Power analysis
- Interpreting results
- Common pitfalls

Introduction to Machine Learning

- Supervised vs unsupervised learning
- Classification basics
- Clustering basics (K-Means)
- Model evaluation metrics
- Train-test split
- Cross-validation

Resources:

- Video: Time Series Analysis Crash Course https://www.youtube.com/watch?v=A3fowDMo8mM&t=8995s
- Project: Stock Price Prediction https://www.youtube.com/watch?v=IY8HZ2Z_sn4
- Video: A/B Testing and Regression Analysis Project https://www.youtube.com/watch?v=iCj4lT5KvJk&t=1057s
- Video: Machine Learning Tutorial by Krish Naik https://www.youtube.com/watch?v=JxgmHe2NyeY

- Playlist: Machine Learning Complete Playlist https://www.youtube.com/watch?v=ZftI2fEz0Fw&list=PLKnIA16_Rmvbr7zKYQuBfsVkjoLcJgxHH

Practical Projects:

1. Sales forecasting using time series
2. Stock price prediction
3. A/B test analysis for website conversion
4. Customer churn prediction (classification)
5. Customer segmentation (clustering)

Time Allocation:

- Theory: 12-14 hours
- Practice: 15-18 hours
- Projects: 10-12 hours

💼 PHASE 4: JOB PREPARATION & PORTFOLIO (MONTH 5)

Week 17-18: Build Your Professional Portfolio

Objective: Create 3-5 end-to-end projects that showcase your skills across different tools and domains.

Portfolio Project Requirements:

Project 1: Excel Dashboard Project

- Business problem: Sales analysis, financial dashboard, or HR analytics
- Include: Pivot tables, advanced formulas, charts, slicers
- Demonstrate: Data cleaning, conditional formatting, what-if analysis

Project 2: SQL Analysis Project

- Business problem: E-commerce analysis, customer behavior, or inventory management
- Include: Complex queries, window functions, CTEs, subqueries
- Demonstrate: Multi-table joins, aggregations, optimization

Project 3: Python EDA Project

- Business problem: Exploratory data analysis on real-world dataset
- Include: Data cleaning, visualizations, statistical analysis

- Demonstrate: Pandas proficiency, data storytelling

Project 4: BI Tool Dashboard

- Business problem: Business dashboard in Power BI or Tableau
- Include: Multiple data sources, DAX/calculated fields, interactivity
- Demonstrate: Data modeling, visualization best practices

Project 5: Advanced Analytics Project (Optional)

- Business problem: Predictive model or time series forecasting
- Include: Machine learning model or advanced statistics
- Demonstrate: End-to-end workflow from data to insights

Project Best Practices:

1. Choose problems from different industries (retail, finance, healthcare, marketing)
2. Use real or realistic datasets (Kaggle, UCI, government data)
3. Document your process thoroughly
4. Create a clear narrative: Problem → Analysis → Insights → Recommendations
5. Include visualizations at every step
6. Prepare a presentation or report (PDF/PPT)
7. Write comprehensive README files

Portfolio Hosting:

- GitHub: For code and project documentation
- Tableau Public / Power BI Service: For interactive dashboards
- Personal Website: GitHub Pages or portfolio site
- Medium/LinkedIn: Write articles about your projects

Resources:

- Guide: How to Create Data Analysis Projects
  https://www.youtube.com/watch?v=X-GRMfxNfrE&t=70s
- Project Example: End-to-End Data Analyst Project
  https://www.youtube.com/watch?v=tT4V7zguCnc

Time Allocation:

- Planning: 3-4 hours

- Execution: 40-50 hours
- Documentation: 10-12 hours
- Portfolio website: 5-6 hours

Week 19: Resume, LinkedIn & Professional Branding

Resume Building

Structure:

1. Header: Name, phone, email, LinkedIn, GitHub, portfolio
2. Professional Summary: 2-3 sentences highlighting your skills
3. Technical Skills: Categorized by tool type
4. Projects: 3-5 key projects with quantified results
5. Experience: Any relevant work (even non-analytics)
6. Education: Degrees, certifications, relevant coursework
7. Additional: Certifications, languages, volunteer work

ATS-Friendly Tips:

- Use standard fonts (Arial, Calibri, Times New Roman)
- Avoid tables, text boxes, and graphics
- Include relevant keywords from job descriptions
- Use standard section headings
- Save as .docx or PDF (check job posting)
- Use bullet points, not paragraphs
- Quantify achievements with numbers

Keywords to Include:

- Technical: SQL, Python, Excel, Power BI/Tableau, Pandas, Data Visualization
- Skills: Data Analysis, Data Cleaning, Statistical Analysis, ETL, Dashboard Creation
- Methods: A/B Testing, Regression Analysis, Forecasting, Reporting
- Soft Skills: Problem Solving, Communication, Stakeholder Management

LinkedIn Optimization

Profile Sections:

1. **Professional Photo:** High-quality, professional headshot
2. **Headline:** "Aspiring Data Analyst | SQL, Python, Power BI | Turning Data into Insights"
3. **About Section:** Tell your story, mention skills, add personality
4. **Experience:** Include projects as "experience"
5. **Skills:** Add 30+ relevant skills
6. **Certifications:** List any completed courses
7. **Projects:** Showcase portfolio work
8. **Recommendations:** Request from colleagues or mentors

LinkedIn Activity:

- Post about your learning journey (2-3 times/week)
- Share insights from projects
- Comment on industry posts
- Connect with data professionals (aim for 500+ connections)
- Turn on "Open to Work" badge
- Engage with content daily (5-10 minutes)

Cover Letter

Structure:

1. Introduction: Why you're interested in this company/role
2. Body Paragraph 1: Your relevant skills and projects
3. Body Paragraph 2: How you can add value
4. Closing: Call to action, thank you

Tips:

- Customize for each application
- Keep it to one page
- Show enthusiasm
- Include specific examples from your projects
- Proofread carefully

Resources:

- Video: Write Resume with No Experience
  https://www.youtube.com/watch?v=EXyO1WiVuZw
- Video: Cover Letter Writing Guide
  https://www.youtube.com/watch?v=a0ATCc6ytyw&t=29s
- Video: Finding Jobs on LinkedIn
  https://www.youtube.com/watch?v=NgdtWKtes6A
- Video: Common Data Analyst Mistakes to Avoid
  https://www.youtube.com/watch?v=W--TWiZPztU&t=85s

Time Allocation:

- Resume creation: 8-10 hours
- LinkedIn optimization: 6-8 hours
- Cover letter template: 3-4 hours
- Personal branding: 5-6 hours

Week 20: Interview Preparation & Job Applications

Technical Interview Preparation

SQL Practice (Priority #1):

- Solve 50+ SQL problems across difficulty levels
- Focus on: Joins, window functions, subqueries, aggregations
- Practice writing queries on a whiteboard
- Explain your thought process out loud
- Time yourself (10-15 minutes per problem)

Practice Platforms:

- LeetCode SQL (Easy: 20, Medium: 25, Hard: 5+)
- HackerRank SQL
- StrataScratch
- DataLemur

Python Practice:

- Solve 30+ Python problems

- Focus on: Pandas operations, data manipulation, basic algorithms
- Practice explaining code

Statistics Questions:

- Review 20+ statistics scenarios
- Practice explaining: Hypothesis testing, p-values, confidence intervals
- Be ready to interpret results

Case Study Preparation:

- Practice 10+ business case studies
- Framework: Clarify → Structure → Analyze → Recommend
- Focus on data-driven decision making

Common Technical Questions:

1. "What's the difference between LEFT JOIN and INNER JOIN?"
2. "How do you handle missing data?"
3. "Explain p-value in simple terms"
4. "What's the difference between correlation and causation?"
5. "How would you detect outliers?"
6. "Explain your approach to an A/B test"
7. "Walk me through a project from your portfolio"

Behavioral Interview Preparation

STAR Method:

- Situation: Set the context
- Task: What was your responsibility?
- Action: What did you do?
- Result: What was the outcome? (Quantify!)

Common Behavioral Questions:

1. "Tell me about yourself" (2-minute pitch)
2. "Why data analysis?"
3. "Describe a challenging project"

4. "Tell me about a time you made a mistake"
5. "How do you handle tight deadlines?"
6. "Describe a time you worked in a team"
7. "How do you explain technical concepts to non-technical people?"
8. "Why do you want to work here?"
9. "Where do you see yourself in 5 years?"
10. "What's your greatest strength/weakness?"

Prepare 5-7 STAR stories that can be adapted to different questions.

Questions to Ask Interviewers:

1. "What does a typical day look like in this role?"
2. "What tools and technologies does the team use?"
3. "How does the team approach collaboration?"
4. "What are the biggest challenges facing the team?"
5. "How do you measure success in this role?"
6. "What opportunities are there for growth and learning?"

Mock Interviews:

- Practice with friends or family
- Use Pramp or Interviewing.io
- Record yourself answering questions
- Join online communities for mock interviews

Job Application Strategy

Where to Apply:

- LinkedIn Jobs (turn on "Open to Work")
- Indeed
- Glassdoor
- Company career pages directly
- AngelList (for startups)
- Networking referrals

Application Process:

1. **Research:** Understand company and role (15-20 minutes)
2. **Customize Resume:** Tailor to job description (10-15 minutes)
3. **Write Cover Letter:** Personalize (15-20 minutes)
4. **Apply:** Submit application
5. **Follow Up:** After 1 week if no response
6. **Track:** Use spreadsheet to track applications

Application Targets:

- Apply to 10-15 jobs per week
- Mix of: Dream jobs, target jobs, stretch opportunities
- Prioritize roles that match 70%+ of requirements

Networking Strategy:

1. Connect with hiring managers and recruiters
2. Send personalized connection requests
3. Engage with company content
4. Ask for informational interviews
5. Request referrals from connections
6. Attend virtual career fairs and meetups

Post-Interview:

- Send thank-you email within 24 hours
- Reiterate interest and key qualifications
- Follow up if no response in 1-2 weeks

Handling Rejections:

- It's normal (expect 90%+ rejection rate initially)
- Ask for feedback when possible
- Learn from each interview
- Don't take it personally
- Keep applying and improving

Time Allocation:

- SQL Practice: 20-25 hours
- Python Practice: 10-12 hours
- Statistics Review: 8-10 hours
- Case Study Practice: 8-10 hours
- Behavioral Prep: 10-12 hours
- Mock Interviews: 6-8 hours
- Applications: 15-20 hours

🎓 MODERN SKILLS FOR 2025

Prompt Engineering & AI Tools

Why It Matters: AI tools are transforming data analysis. Being proficient in leveraging AI assistants will make you more productive and valuable.

What to Learn:

Prompt Engineering Basics:

- Anatomy of a good prompt: Clarity, context, specificity
- Types of prompts: Instructional, conversational, zero-shot, few-shot
- Using examples to guide AI
- Iterative prompting

AI Tools for Data Analysis:

- ChatGPT / Claude: Code generation, debugging, explanation
- GitHub Copilot: Code assistance
- NotebookLM: Research and analysis
- Data interpretation and insight generation

Practical Applications:

1. Generate SQL queries from natural language
2. Debug Python code
3. Create data visualizations
4. Write documentation
5. Formulate research questions

6. Summarize reports
7. Generate regex patterns
8. Explain complex concepts

Resources:

- Video: Prompt Engineering Tutorial https://www.youtube.com/watch?v=5i2Hn8OG94o
- Practice: Use ChatGPT/Claude for daily data tasks
- Platform: OpenAI Playground

Best Practices:

- Always verify AI-generated code
- Use AI as a learning tool, not a replacement
- Combine AI assistance with your knowledge
- Document your prompts for reuse

Time Investment: 5-8 hours spread throughout the roadmap

Machine Learning Basics (Optional but Recommended)

Why Include ML: Many data analyst roles now expect basic ML knowledge. It demonstrates your ability to work with advanced analytics.

What to Learn:

Supervised Learning:

- Linear Regression: Predicting continuous values
- Logistic Regression: Binary classification
- Decision Trees: Rules-based predictions
- Random Forest: Ensemble method
- Gradient Boosting: XGBoost, LightGBM basics

Unsupervised Learning:

- K-Means Clustering: Customer segmentation
- Hierarchical Clustering
- PCA: Dimensionality reduction

- Association Rules: Market basket analysis

Model Evaluation:

- Metrics: Accuracy, Precision, Recall, F1-Score, AUC-ROC
- Confusion Matrix
- Cross-validation
- Train-test split
- Overfitting and underfitting

ML Workflow:

1. Problem definition
2. Data collection and preparation
3. Feature engineering
4. Model selection and training
5. Model evaluation
6. Model tuning
7. Deployment (basic concept)

Python Libraries:

- scikit-learn: Main ML library
- pandas: Data manipulation
- matplotlib/seaborn: Visualization

Resources:

- Video: Machine Learning Tutorial by Krish Naik
  https://www.youtube.com/watch?v=JxgmHe2NyeY
- Playlist: Machine Learning Complete Playlist
  https://www.youtube.com/watch?v=ZftI2fEz0Fw&list=PLKnIA16_Rmvbr7zKYQuBfsVkjoLcJgxHH
- Project: End-to-End Machine Learning Project
  https://www.youtube.com/watch?v=S_F_c9e2bz4&list=PLZoTAELRMXVPS-dOaVbAux22vzqdgoGhG
- Course: Google's Machine Learning Crash Course
- Platform: Kaggle Learn

Project Ideas:

1. Customer churn prediction
2. House price prediction
3. Credit risk assessment
4. Sales forecasting with ML
5. Customer segmentation

Time Investment: 20-30 hours (can be done in Month 4 or parallel to other learning)

📚 SOFT SKILLS DEVELOPMENT (THROUGHOUT)

Soft skills are as critical as technical skills for data analyst success.

1. Communication Skills

Why It Matters: Data analysts must explain complex findings to non-technical stakeholders.

How to Develop:

- Present your projects to friends/family
- Explain technical concepts in simple terms
- Practice the "explain it to a 10-year-old" approach
- Record yourself presenting and review
- Join Toastmasters or public speaking groups

Practice Exercises:

- Explain p-values without using technical jargon
- Present a project insight in 1 minute
- Create an executive summary for each project