

# THE COMPLETE DATA ANALYST ROADMAP



**Go From Zero to a Data Analyst in 12 Months**

Mosh Hamedani



Hi! I am Mosh Hamedani, a software engineer with over 20 years of experience.

Over the past 10 years, I've had the privilege of teaching millions of people how to code and become professional software engineers through my YouTube channel and online courses.

It's my mission to make software engineering accessible to everyone. Join me on this journey and unlock your potential in the world of coding!

<https://codewithmosh.com>

# Table of Content

<b>Introduction</b>	<b>4</b>
Target Audience	4
Resources	4
<b>Roadmap Overview</b>	<b>5</b>
<b>Mathematics and Statistics</b>	<b>6</b>
<b>Excel</b>	<b>8</b>
<b>SQL</b>	<b>10</b>
<b>Python</b>	<b>12</b>
<b>Version Control (Git)</b>	<b>14</b>
<b>Data Collection and Preparation</b>	<b>15</b>
<b>Data Visualization</b>	<b>17</b>
<b>Machine Learning (Optional)</b>	<b>19</b>
<b>Big Data (Optional)</b>	<b>20</b>

## Introduction

This guide is designed to help you navigate the essential skills needed to become a successful data analyst. Whether you're just starting out or looking to enhance your existing skills, this roadmap will provide a clear and structured path.

## Target Audience

This guide is for:

- **Beginners** who want to know what they need to learn to land a data analyst job.
- **Experienced individuals** looking to level up their skills and fill in the gaps in their knowledge.

## Resources

For detailed tutorials and full courses, check out the following resources:

- **YouTube Channel:** <https://www.youtube.com/c/programmingwithmosh>
- **Full Courses:** <https://codewithmosh.com>

## Roadmap Overview

Below is a comprehensive table listing all the essential skills needed to become a proficient data analyst, along with the estimated time required to learn each skill.

Keep in mind that the time needed to learn each skill can vary for everyone. These estimates are based on dedicating 3 to 5 hours of study every day.

Use this roadmap to guide your learning journey and track your progress as you build a strong foundation in data analysis.

Skill	Est. Time	Learning Phase
Mathematics and Statistics	1-2 months	Beginner
Excel	2-3 weeks	Beginner
SQL	1-2 months	Beginner
Programming (Python)	1-2 months	Beginner
Version Control (Git)	1-2 weeks	Beginner
Data Collection and Preparation	1-2 months	Intermediate
Data Visualization	1-2 months	Intermediate
Machine Learning (Optional)	1-2 months	Advanced
Big Data (Optional)	1-2 months	Advanced
<b>Total</b>	<b>8-16 months</b>	

# Mathematics and Statistics

A strong foundation in mathematics and statistics is essential for data analysis. Concepts such as probability, statistical analysis, and algebra provide the theoretical underpinnings for understanding and interpreting data.

Estimated Time: 1-2 months

## Essential Concepts

- **Basic Algebra**
  - Basic algebraic operations
  - Solving equations
- **Calculus**
  - Basic concepts of differentiation and integration
  - Understanding of limits and functions
- **Linear Algebra**
  - Vectors and matrices
  - Matrix operations
  - Eigenvalues and eigenvectors
- **Probability**
  - Basic probability concepts
  - Probability distributions (normal, binomial, poisson)
  - Random variables
  - Bayes' theorem

- **Descriptive Statistics**

- Mean, median, mode
- Variance and standard deviation
- Skewness and kurtosis
- Quartiles and percentiles

- **Inferential Statistics**

- Hypothesis testing
- Confidence intervals
- p-values
- T-tests and chi-square tests
- Analysis of variance (ANOVA)

- **Regression Analysis**

- Simple linear regression
- Multiple linear regression
- Logistic regression

## Excel

Excel is a powerful tool for data analysis and manipulation. It's widely used for its simplicity, accessibility, and a broad range of built-in functions that make data analysis straightforward.

Estimated Time: 2-3 weeks

### Essential Concepts

- **Basics**

- Creating and managing workbooks
- Navigating and selecting cells
- Cell formatting (number, text, date, currency)

- **Formulas and Functions**

- Basic arithmetic operations (+, -, \*, /)
- Common functions: SUM, AVERAGE, COUNT, MAX, MIN
- Logical functions: IF, AND, OR, NOT
- Text functions: CONCATENATE, LEFT, RIGHT, MID, TRIM
- Date functions: TODAY, NOW, DATE, DATEDIF, YEAR, MONTH, DAY
- Lookup functions: VLOOKUP, HLOOKUP, INDEX-MATCH

- **Basic Data Analysis**

- Sorting data
- Filtering data
- Conditional formatting



- **Charts**

- Creating charts (bar charts, line charts, pie charts, scatter plots)
- Customizing chart elements (titles, labels, legends)

- **Pivot Tables**

- Creating pivot tables
- Customizing pivot tables

- **Advanced Formulas**

- IF
- SUMIF
- COUNTIF
- IFERROR

# SQL

SQL (Structured Query Language) is essential for querying and managing data in relational databases. It's a fundamental skill for any data analyst working with structured data.

Estimated Time: 1-2 months

Learning resources: [YouTube Tutorial](#) | [Full Course](#)

## Essential Concepts

- **Basic Operations**

- Querying data (SELECT)
- Modifying data (INSERT, UPDATE, DELETE)
- Filtering data (WHERE, IN, BETWEEN, LIKE, IS NULL, REGEXP)
- Logical operators (AND, OR, NOT)
- Sorting and limiting data (ORDER BY, LIMIT)

- **Complex Queries**

- Joins (INNER, OUTER, SELF, NATURAL, CROSS)
- Aggregate functions (MAX, MIN, AVG, SUM, COUNT)
- Grouping data (GROUP BY, HAVING, ROLLUP)
- Subqueries

- **Views**

- **Stored Procedures and Functions**

- **Triggers and Events**

- **Transactions**

- Transaction isolation levels
- BEGIN, COMMIT, ROLLBACK

- **Database Design**

- Normalization
- Database integrity with primary keys, foreign keys, and constraints

- **Indexes**

- **Security and Permissions**

- Managing users and privileges

# Python

Python is a versatile and widely-used programming language in data analysis due to its simplicity and extensive library support.

Estimated Time: 1-2 months

Learning resources: [YouTube Tutorial](#) | [Full Course](#)

## Essential Concepts

- **Python Fundamentals**
  - Variables and data types
  - Loops (for, while) and conditional statements (if, elif, else)
  - Functions and scope
- **Data Structures**
  - Arrays, lists, tuples and sets
  - Stacks and queues
  - Dictionaries
  - Comprehensions
  - Generator expressions
- **Exception Handling**
  - Handling exceptions with try/except
  - Raising exceptions
- **Functional Programming**
  - Lambda functions
  - Map, reduce, filter

- **Object-oriented Programming**
  - Classes and objects
  - Inheritance and polymorphism
- **Modules and packages**
  - Creating modules
  - Managing packages with pip and pipenv
  - Virtual environments
- **Python Standard Library**
  - Working with paths, files, and directories
  - Working with CSV and JSON files
  - Working with Date/time
  - Generating random values
- **Data Analysis Libraries**
  - Pandas
  - NumPy
  - SciPy

## Version Control (Git)

Git is a version control system that is crucial for managing code and collaboration. It allows you to track changes, collaborate with others, and maintain the integrity of your codebase, making it an essential tool for any data analyst

Estimated Time: 1-2 weeks

Learning resources: [YouTube Tutorial](#) | [Full Course](#)

### Essential Concepts

- **Setup and Configuration:** init, clone, config
- **Staging:** status, add, rm, mv, commit, reset
- **Inspect and Compare:** log, diff, show
- **Branching:** branch, checkout, merge
- **Remote Repositories:** remote, fetch, pull, push
- **Temporary Commits:** stash
- **GitHub:** fork, pull request, code review

## Data Collection and Preparation

Data collection, cleaning, and preparation are critical steps in data analysis. This involves gathering raw data, transforming it into a usable format, and ensuring its quality and accuracy for reliable analysis.

Estimated Time: 1-2 months

### Essential Concepts

- **Python Libraries**
  - Numpy (numerical computing)
  - Pandas (data manipulation)
  - BeautifulSoup (web scraping)
  - Scrapy (web scraping)
- **Data Collection**
  - Importing data from CSV, Excel, and JSON files
  - Connecting to databases
  - Using APIs to collect data
  - Web scraping
- **Data Cleaning**
  - Handling missing values
  - Removing duplicates
  - Finding outliers
  - Data transformation
    - Converting data types (text to numbers, dates, etc)

- Parsing and splitting data (text to columns)
- Standardizing data formats (uppercase, lowercase, date formats)
- **Data Integration**
  - Combining data from multiple sources



## Data Visualization

Data visualization involves creating graphical representations of data to identify trends, patterns, and insights. It is a crucial skill for communicating results effectively.

Estimated Time: 1-2 months

### Essential Concepts

- **Python Libraries**
  - Matplotlib
  - Seaborn
- **Visualization Tools**
  - Basics of Tableau or Power BI (growing in popularity)
- **Charts and Graphs**
  - Bar charts
  - Line charts
  - Scatter plots
  - Funnel charts
  - Histograms
  - Stacked charts
  - Heatmaps
  - Pie charts
- **Dashboards**
  - Creating interactive dashboards

- Dynamic dashboards
- **Storytelling with Data**
  - Creating narratives
  - Insights through visualization

## Machine Learning (Optional)

Machine learning provides data analysts with powerful tools to create predictive models and uncover deeper insights from data. Understanding the basics of machine learning can significantly enhance your ability to analyze complex datasets and provide actionable insights.

Estimated Time: 1-2 months

### Essential Concepts

- **Python Libraries**
  - Scikit-learn
- **Supervised Learning**
  - Regression algorithms (e.g., linear regression, logistic regression)
  - Classification algorithms (e.g., decision trees, k-nearest neighbors, support vector machines)
- **Unsupervised Learning**
  - Clustering algorithms (e.g., K-means, hierarchical clustering)
  - Dimensionality reduction techniques (e.g., PCA, LDA)
- **Model Evaluation**
  - Confusion matrix
  - Precision
  - Recall
  - F1 score
  - ROC curves

## Big Data (Optional)

Big data technologies enable data analysts to handle and process vast amounts of data efficiently. Understanding big data concepts and tools is essential for working with large datasets and gaining insights from them.

Estimated Time: 1-2 months

### Essential Concepts

- **Big Data Frameworks**

- Hadoop
- Apache Spark

- **Data Storage**

- HDFS
- NoSQL databases (Cassandra, MongoDB)

- **Data Processing**

- MapReduce programming model
- Batch processing with Spark
- Real-time processing with Spark Streaming

- **Data Ingestion**

- Data collection tools (Kafka, Flume)

*Learning to code is a journey. Be patient with yourself and stay persistent, even when things get tough.*

*- Mosh*