

# Data Analyst → Data Engineer Roadmap

This roadmap is designed for freshers who want to start as a Data Analyst and later grow into Data Engineering roles. It explains what to learn, in what sequence, and why. Following this plan step by step will give you strong portfolio projects, a resume that stands out, and the flexibility to apply to Data Analyst, BI, and Junior Data Engineer roles.

## Step 1: Python Automation (Start Here)

- Automate Excel/CSV cleaning and reporting using Python (pandas, openpyxl).
- Create scripts that refresh reports automatically instead of manual work.
- Mini project idea: Automate daily sales Excel file cleanup and summary report.

## Step 2: Cloud Basics

- Learn a cloud platform (AWS recommended, but GCP/Azure also good).
- Focus first on: Storage (S3/GCS), Compute basics (EC2/VMs), IAM (permissions).
- Practice uploading cleaned data from Python to cloud storage.
- Mini project idea: Upload cleaned CSV files to AWS S3 via Python.

## Step 3: Data Warehousing

- Learn concepts: OLTP vs OLAP, star schema, fact vs dimension tables.
- Pick one DW: AWS Redshift, GCP BigQuery, or Snowflake.
- Load cloud data into DW, query it with SQL.
- Mini project idea: CSV → S3 → Redshift → Power BI dashboard.

## Step 4: Workflow Automation

- Learn Apache Airflow or dbt for scheduling pipelines.
- Automate end-to-end flow: data ingestion → transformation → dashboard refresh.
- Mini project idea: Airflow DAG that pulls raw CSV → uploads to S3 → loads into Redshift → refreshes Power BI.

## Portfolio & Resume

- Show 4 main projects (1 from each step) on GitHub with clear READMEs + screenshots.
- Resume bullets should highlight automation, cloud, and end-to-end pipeline experience.
- Apply to Data Analyst, BI Analyst, Junior Data Engineer, and Analytics Engineer roles.

■ Recommended Sequence: Python Automation → Cloud Basics → Data Warehousing → Workflow Automation. This order ensures early results, strong portfolio projects, and a resume that stands out in off-campus hiring.