# 📘 Data Analytics – Full Course Syllabus (16 Weeks, Updated)

**Audience:** 3rd-year Computer Science students
**Format:** 2 hrs lecture + 2 hrs lab each week
**Final Deliverable:** Capstone project where students collect, clean, analyze, and visualize real-world data using Python, SQL, and one BI tool (Excel, Tableau, or Power BI).

---

### Week 1 – Introduction to Data Analytics

- What is data analytics? Types (descriptive, diagnostic, predictive, prescriptive)
- Data lifecycle: collection → cleaning → analysis → visualization → decision-making
- Roles: Data Analyst vs Data Scientist vs Business Analyst
- Tools overview: Python, SQL, Excel, Tableau, Power BI
  **Lab:** Install Anaconda, Jupyter; "Hello Data" in Python (load CSV, summary stats).

### Week 2 – Excel for Data Analytics

- Cleaning data in Excel
- Pivot tables & pivot charts
- Lookup functions (VLOOKUP, XLOOKUP)
- Conditional formatting
  **Lab:** "Company Sales" dataset; pivot: sales by region/month; conditional formatting for performance.

### Week 3 – Python for Data Analytics (NumPy)

- NumPy arrays vs lists (performance)
- Array creation: `arange`, `linspace`, `random`
- Indexing, slicing, boolean filtering
- Vectorized operations, broadcasting
- Aggregations: `sum`, `cumsum`, `argmax`, `argmin`
- Reshaping & stacking (`reshape`, `hstack`, `vstack`)
  **Lab:** Simulate sales with `np.random`; compute monthly averages; compare loop vs vectorized sum; reshape daily → weekly sales.

### Week 4 – Pandas DataFrames (Exploration & Manipulation)

- Creating Series & DataFrames
- Importing CSV, Excel, JSON
- Overview: `head`, `info`, `describe`
- Selecting & filtering (`loc`, `iloc`)
- Sorting by multiple columns
- Adding calculated columns (profit, margin)
  **Lab:** Load "Retail Sales" dataset; compute profit margin; sort products by revenue; filter products > $10K revenue.

### Week 5 – Data Preparation for Analysis

- Handling missing values (drop, fill simple methods)
- Duplicates & inconsistent entries
- Type conversions (dates, numbers, categories)
- Working with dates (extract month/year, resampling)
- Renaming/reordering columns
- Combining datasets (`merge`, `concat`)
- Simple outlier handling (3σ, IQR rules)
  **Lab:** Clean "HR Employee" dataset; parse hire dates; fill missing salaries; remove duplicates; merge HR + Department datasets.

### Week 6 – Business-Oriented Data Analysis (Pandas)

- Descriptive stats: totals, averages, growth %, contribution
- GroupBy & aggregation (single & multi-level)
- Crosstabs (`pd.crosstab`)
- Pivot tables
- Ranking & sorting (top-N customers/products)
- Window operations (rolling averages, cumulative sums)
- Reshaping: `melt`, `pivot`
- Exporting results (Excel/CSV for BI)
  **Lab:** Analyze "E-commerce Orders"; revenue by region & category; top 3 products per region; monthly revenue growth (rolling avg); pivot table by region/year.

### Week 7 – Data Visualization for Analytics

- Choosing the right chart for the question (comparison, trend, distribution, composition)
- Matplotlib basics for quick insights
- Seaborn advanced visuals (heatmaps, pairplots, boxplots)
- Analytics-focused charts:
  - Trendlines (monthly revenue, user growth)
  - Category comparisons (top products, regional sales)
  - Distribution analysis (customer age, order value)
  - Pareto chart (80/20 rule: products driving most revenue)
- Combining multiple visuals into mini-dashboard (subplot grids)
  **Lab:** Visualize "E-commerce Sales" dataset:
  - Revenue trendline (line plot)
  - Top 5 products (bar chart)
  - Sales distribution (histogram/boxplot)
  - Heatmap (region × category sales)
  - Pareto analysis of product revenue contribution

### Week 8 – SQL Basics

- Relational databases & schemas
- SELECT, WHERE, ORDER BY, LIMIT
  **Lab:** Query "Employee" database; retrieve top salaries; employees hired after 2020.

### Week 9 – Intermediate SQL

- Aggregations (SUM, AVG, COUNT, GROUP BY)

- Joins (INNER, LEFT, RIGHT, FULL)
  **Lab:** "HR + Department" DB; avg salary by department; employee counts; departments exceeding budget.

## Week 10 – Advanced SQL

- Subqueries & nested queries
- Window functions (ROW_NUMBER, RANK, PARTITION BY)
  **Lab:** "Sales + Customer Churn" DB; rank customers by revenue; top 3 products in each category.

## Week 11 – Tableau Basics

- Tableau interface & workflow
- Connecting to data (Excel, CSV, SQL)
- Filters, groups, sorting
- Simple charts (bar, line, scatter, pie)
  **Lab:** "Global Superstore" dataset; dashboard with sales by category, region, trendline.

## Week 12 – Tableau Advanced

- Joins & blending
- Calculated fields, parameters
- Maps & geospatial analysis
- Dashboards & storytelling (filters, interactivity)
  **Lab:** Build interactive COVID or UN dataset dashboard; filter by country/year; storytelling presentation.

## Week 13 – Power BI Basics & Advanced

- Power BI interface & importing data
- Visualizations: bar, line, map, gauge
- DAX basics (calculated columns, measures)
- Combining multiple data sources
  **Lab:** "HR + Finance" dataset; salary distribution, employee count by department, expenses trend.

## Week 14 – Integrating Tools & End-to-End Workflow

- When to use Excel vs Tableau vs Power BI
- Workflow: SQL → Python → BI tool
- Case study: e-commerce pipeline
  **Lab:** Mini-project: query sales from SQL, clean in Python, dashboard in BI tool (student's choice).

## Week 15 – Ethics & Case Studies

- Data privacy (GDPR, HIPAA)
- Responsible analytics & bias
- Case studies: Cambridge Analytica, healthcare misuse
  **Lab:** "Census" dataset → detect possible bias in gender/age representation; discuss ethical interpretation.

## Week 16 – Capstone Project Presentations

- Projects must combine Python + SQL + 1 BI tool (Excel, Tableau, or Power BI).
- Example topics:
    - COVID-19 global trends
    - E-commerce customer retention
    - Financial market analytics
    - Social media sentiment

## Grading and Exams

- Midterm (20%)
- Final (20%)
- Lab Assignments (20%)
- Project (40%)

Late submissions won't be accepted. No exam retakes.