### Trientation Session: Core Focus Areas

Here's how to structure the official course orientation for your Data Analytics subject:

### 1. Course Overview

• Name: **DATA ANALYTICS** 

• Semester: Fifth (Diploma – K Scheme)

• Course Code: 315326

• Applicable Programmes: CM, CO, CW, IF, IH, SE, TE

Total Marks: 150 (70 Theory, 25 Practical External, 25 Practical Internal, 10 SLA, 20 CT)

## 2. Syllabus Breakdown (High-Level)

Unit	Title	Key Topics	COs Covered
I	Introduction to Data Analytics	Types, sampling, lifecycle, central tendency	CO1
II	Statistical Analysis	Box plots, correlation, ANOVA, hypothesis testing	CO2
III	Analytics with Excel	Dashboards, pivot tables and charts	CO3
IV	Data Visualization	Excel charts, formatting, legends	CO4
V	Python Visualization	Matplotlib, customization, exporting	CO5

Make sure you show how **each unit aligns with a Course Outcome (CO)** to give students a sense of progression.

### u 3. Reference Books & Learning Material

Here's your curated list:

Title	Author	ISBN
Excel Data Analysis	Jinjer Simon	978-0470591604
Python Data Analytics	Fabio Nelli	978-1484239124
Python Data Science Handbook	Jake VanderPlas	978-9355422552

Data Analysis with Excel	A.J. Smalley	9780070139909
Business Analytics (MindTap)	Camm, Cochran, Fry	9789360533533

#### Also include key portals like:

• NPTEL Course FreeCodeCamp Matplotlib Simplificarn Excel Tutorial

# 4. Assessment Strategy

Break down Formative (FA) and Summative (SA) components:

- Theory: Class tests, final exam
- Practicals: Lab experiments aligned with COs
- SLA: Microprojects, assignments, dashboards, interactive visualizations

Emphasize "**Process vs. Product**" scoring — practicals assess both how students work and what they produce.

# **5. Lab Resources**

### Requirement Details

Software Excel, Office 365, Python setup

Hardware i5 system, 8GB RAM, Windows 10+

Learning Tools Jupyter Notebook, Matplotlib, Slicers

### 6. Closing Tips for Students

- Engage deeply in hands-on labs
- Don't just learn tools focus on problem-solving
- Apply statistical concepts to real scenarios
- Explore dashboards & Python plots beyond templates

In the **next session**, that's when you can **bring the concepts alive** with examples like Swiggy analytics, Amazon forecasting, and YouTube retention curves.

## What Is Data?

**Data** is raw information — numbers, text, images, or facts — collected from various sources. It can be:

- **Structured** (like rows in Excel or SQL databases)
- **Unstructured** (like tweets, videos, or customer reviews)

Example: Every time someone places an order on Swiggy, it generates data — what they ordered, when, from which restaurant, and how much they paid.

# What Is Analytics?

**Analytics** is the process of examining data to find patterns, trends, and insights. It helps answer questions like:

- What happened?
- · Why did it happen?
- What might happen next?
- What should we do about it?

Example: YouTube uses analytics to track how many people watched a video, how long they stayed, and what made them click away.

# What Is Data Analytics?

**Data Analytics** is the science of analyzing raw data to make decisions. It combines:

- Statistics (mean, median, regression)
- **Programming** (Python, R, SQL)
- Visualization (charts, dashboards)

Example: Amazon uses data analytics to predict what products you might want to buy next based on your browsing and purchase history.

#### Python

R: **R** is a programming language designed specifically for **statistical computing**, **data analysis**, and **visualization**.

SQL: **SQL** (Structured Query Language) is used to interact with databases—to store, retrieve, and manipulate data.

# Types of Data Analytics

Туре	What It Does	Example Use Case
Descriptive	Tells you what happened	Swiggy shows total orders last month
Diagnostic	Explains why it happened	YouTube analyzes why viewers dropped off
Predictive	Forecasts what might happen	Amazon predicts future buying behavior
Prescriptive	Suggests what to do next	Swiggy recommends delivery route changes

# What Is MIS Reporting?

MIS (Management Information System) Reporting is a structured way to present business data to decision-makers. It includes:

- · Sales reports
- Inventory status
- Financial summaries
- Customer feedback

Example: A Swiggy manager might use an MIS report to see which restaurants had delayed deliveries last week.

# **ii** What Is a Dashboard?

A **dashboard** is a visual interface that shows key metrics in real time. It uses charts, graphs, and tables to make data easy to understand.

- Example: YouTube Studio dashboard shows creators:
  - Views
  - Watch time
  - Subscriber growth
  - Revenue

- Example: Swiggy's internal dashboard might show:
  - · Orders per hour
  - Delivery time
  - Customer ratings

### YouTube Views Analytics

#### YouTube tracks:

- Impressions (how often a thumbnail is shown)
- Click-through rate (how often people click)
- Watch time (how long they stay)
- Audience retention (where they drop off)
- Creators use this to improve thumbnails, titles, and video length.

### Swiggy Pattern Analytics

#### Swiggy analyzes:

- Order trends (which food is popular at what time)
- Customer behavior (repeat orders, ratings)
- **Delivery performance** (time, route efficiency)
- Revenue breakdown (by restaurant, cuisine, region)
- Example: If biryani orders spike on weekends, Swiggy might promote biryani deals on Fridays.

#### Amazon Predictions

#### Amazon uses predictive analytics and machine learning to:

- Recommend products
- Forecast demand
- Optimize inventory
- Personalize ads
- Example: If you buy a phone, Amazon might suggest a case, charger, or screen protector based on millions of similar purchases.

Absolutely, Akash! Let's start with the basics of **R** and **SQL**, two powerful tools in the data analytics world. I'll explain them in a way that connects with your academic and practical goals.

# What Is R?

R is a programming language designed specifically for statistical computing, data analysis, and visualization.

# Key Features of R

- Built for data manipulation, statistical modeling, and graphing
- Rich ecosystem of packages like ggplot2, dplyr, tidyr, and caret
- Ideal for academic research, machine learning, and data science

# 🙀 Example Use Case

You have a dataset of student scores. In R, you can:

```
scores <- c(85, 90, 78, 92, 88)
mean(scores)  # Calculates average
sd(scores)  # Standard deviation
boxplot(scores)  # Visualizes distribution</pre>
```

### Why R Is Useful for You

- Great for statistical analysis like ANOVA, regression, and hypothesis testing
- Supports data wrangling and visual storytelling
- Widely used in academia, healthcare, and finance

You can explore more on R-Project's official site or try tutorials at W3Schools and GeeksforGeeks.

# ₩hat Is SQL?

**SQL (Structured Query Language)** is used to **interact with databases**—to store, retrieve, and manipulate data.

## Key Features of SQL

- Works with relational databases like MySQL, PostgreSQL, Oracle, and SQL Server
- Lets you query, filter, join, and aggregate data
- Essential for MIS reporting, dashboards, and backend data operations

## **ii** Example Use Case

You have a table of student records. In SQL, you can:

```
Sql

SELECT name, score FROM students WHERE score > 80;
```

This retrieves names and scores of students who scored above 80.

#### Why SQL Is Useful for You

- Core skill for data analysts, BI developers, and backend engineers
- Used in Swiggy, Amazon, and YouTube to manage user data
- Integrates with Excel, Power BI, and Python for full-stack analytics

You can start learning SQL at <u>W3Schools</u> or dive deeper with <u>GeeksforGeeks SQL</u> Tutorial.

R and SQL are a powerful duo in data analytics, and they're often used together in real-world projects to handle everything from data extraction to statistical modeling and visualization. Here's a breakdown of common project types that leverage both:

# Real-World Projects Using R

Project Type	What R Does Best	Example Use Case
Customer Churn Analysis	Predictive modeling, logistic regression	Telecom companies predicting cancellations
Sentiment Analysis	Text mining, NLP with tidytext	Analyzing product reviews or tweets
Sales Forecasting	Time series analysis with forecast package	E-commerce platforms predicting demand
Healthcare Risk Prediction	Classification models, ROC curves	Predicting diabetes or heart disease
Credit Card Fraud Detection	Anomaly detection, decision trees	Flagging suspicious transactions
Recommendation Systems	Collaborative filtering, clustering	Movie or music suggestions
Market Basket Analysis	Association rules with arules	Retail product bundling
Data Visualization Dashboards	ggplot2, shiny for interactive visuals	Executive dashboards for KPIs

# Real-World Projects Using SQL

Project Type	What SQL Handles	Example Use Case
Inventory Management System	CRUD operations, joins, triggers	Retail stock tracking
Library or Hospital Database	Relational schema design, indexing	Managing books or patient records
Sales Data Analysis	Aggregations, filtering, grouping	Monthly revenue breakdowns
Customer Segmentation	Querying demographics, purchase history	Targeted marketing campaigns

E-commerce Order Tracking	Real-time updates, foreign key relationships	Amazon-style order systems
Flight or Railway Booking	Reservation logic, availability checks	Ticketing platforms
MIS Reporting	Scheduled queries, report generation	Business intelligence dashboards

For more SQL project ideas, check out GeeksforGeeks and upGrad's curated list.

# Projects That Use Both R and SQL Together

These are especially relevant for data analysts and BI professionals:

- ETL Pipelines: SQL extracts data → R cleans and models it
- Dashboard Reporting: SQL pulls data → R visualizes it with ggplot2 or shiny
- Data Warehousing: SQL stores structured data → R performs statistical analysis
- Survey Analysis: SQL organizes responses → R runs ANOVA, regression, or clustering

# ETL Pipelines (Extract, Transform, Load)

ETL pipelines are **data workflows** used to move and process data from one system to another:

- Extract: Pull data from sources like databases, APIs, or files.
- **Transform**: Clean, format, and enrich the data (e.g., convert types, remove duplicates).
- **Load**: Push the transformed data into a target system like a data warehouse.

\* Example: Pulling sales data from multiple stores, cleaning it, and loading it into a central dashboard for analysis.

# CRUD Operations (Create, Read, Update, Delete)

CRUD defines the **basic operations** for interacting with databases or APIs:

- **Create**: Add new records (e.g., INSERT in SQL).
- **Read**: Retrieve data (e.g., SELECT).

- **Update**: Modify existing records (e.g., UPDATE).
- **Delete**: Remove records (e.g., DELETE).

\* Example: A user management system where you add new users, view their profiles, update details, or delete accounts.

# 🙀 arules (R Package)

arules is an R package for association rule mining and frequent itemset discovery:

- Used in **market basket analysis** to find patterns like "If a customer buys bread and butter, they're likely to buy jam."
- Implements algorithms like **Apriori** and **Eclat**.
- Outputs rules with metrics like support, confidence, and lift.
- 📌 Example: Analyzing grocery transactions to recommend product bundles.

# ggplot2 (R Package)

ggplot2 is a powerful R package for **data visualization**, based on the **Grammar of Graphics**:

- Allows layering of plots: ggplot(data) + aes(...) + geom\_point()
- Supports histograms, scatter plots, boxplots, and more.
- Highly customizable with themes, scales, and facets.
- \* Example: Visualizing the relationship between horsepower and fuel efficiency in cars.

# Shiny (R Package)

Shiny lets you build **interactive web apps** using R:

- Combines **UI** and **server logic** in one app.
- Supports sliders, inputs, dynamic plots, and dashboards.
- Great for sharing data insights without needing HTML/JS.
- \* Example: A dashboard where users select a date range and view updated sales trends.

# Time Series Forecasting (Quick Revision)

Since you've touched this before, here's a refresher:

- Models: ARIMA, Exponential Smoothing, Prophet.
- Key Concepts: Trend, seasonality, autocorrelation.
- **Tools in R**: forecast, ts, fable, prophet.
- ★ Example: Forecasting monthly electricity demand based on historical usage.

A **time series** is a sequence of data points collected or recorded at regular time intervals — like daily, monthly, or yearly. What makes it special is that **time itself is a key variable**, and the order of data matters. You're not just looking at values, you're looking at how those values evolve over time.

# **Real-Life Examples of Time Series**

Domain	Time Series Example	What It Tracks Over Time
Finance	Stock prices	Daily closing values
Weather	Temperature readings	Hourly or daily temperatures
Retail	Sales data	Monthly revenue or product demand
Healthcare	ECG or heart rate monitoring	Beats per minute over time
Tech Platforms	YouTube views or Swiggy orders	Viewer count or food orders per hour

# Key Components of Time Series

- 1. **Trend**: Long-term increase or decrease in data *E.g., rising smartphone sales over years*
- 2. **Seasonality**: Repeating patterns at fixed intervals *E.g.*, *ice cream sales peaking in summer*
- 3. **Cyclic Patterns**: Irregular, long-term fluctuations *E.g.*, economic booms and recessions
- 4. **Noise (Irregularity)**: Random variations *E.g., sudden spike in Swiggy orders due to a cricket match*

# Why Time Series Matters in Analytics

- Helps **forecast future values** (like predicting next month's sales)
- Detects **anomalies** (e.g., fraud detection)
- Supports decision-making (e.g., staffing based on demand trends)

# 樳 Bonus: Time Series vs. Other Data

Туре	Description
Time Series	Data ordered by time (e.g., hourly sales)
Cross-sectional	Data at a single point in time (e.g., survey)
Panel Data	Combines time series + cross-sectional

Amazon is a masterclass in using **time series forecasting** to optimize nearly every corner of its operations. Let's break down how it works and where it's applied:

# **1. Inventory Forecasting**

Amazon predicts how many units of a product will be needed in each warehouse, at each time interval (daily, weekly, monthly).

- Time Series Input: Historical sales data per SKU (product ID), timestamped.
- Forecast Output: Expected demand for each product in the next 30 days.
- **Impact**: Reduces overstock and stockouts, saving millions in storage and logistics.

Example: If umbrellas sell more during monsoon season in Mumbai, Amazon forecasts a spike and pre-positions inventory accordingly.

#### 2. Delivery & Logistics Optimization

Amazon uses time series models to forecast delivery volumes and traffic patterns.

- Time Series Input: Past delivery counts, time of day, location, weather.
- Forecast Output: Expected number of deliveries per hour per region.
- Impact: Helps plan driver shifts, route assignments, and delivery windows.
- Example: During Prime Day, Amazon forecasts a surge in orders and adjusts staffing and routes in advance.

### 3. Sales Forecasting

Amazon predicts future sales for millions of products using deep learning models like **Amazon Forecast** and **Chronos**.

- Time Series Input: Historical sales, promotions, seasonality, competitor pricing.
- Forecast Output: Sales volume per product per day/week.
- **Impact**: Informs pricing, promotions, and procurement.

Example: If a laptop's sales spike every August (back-to-school season), Amazon adjusts pricing and stock accordingly.

#### 4. Personalized Recommendations

While not traditional time series, Amazon uses **temporal patterns** in user behavior to recommend products.

- Input: User's browsing and purchase history over time.
- Model: Sequence-based models (like RNNs or Transformers).
- Impact: Suggests what you're likely to buy next often eerily accurate!

Example: If you bought a phone yesterday, Amazon might recommend a case or screen protector today.

### Tools Amazon Uses

- Amazon Forecast: A fully managed ML service for time series forecasting
- SageMaker Canvas: No-code tool to build forecasts from tabular data
- Chronos: Pretrained time series models built by Amazon Science