

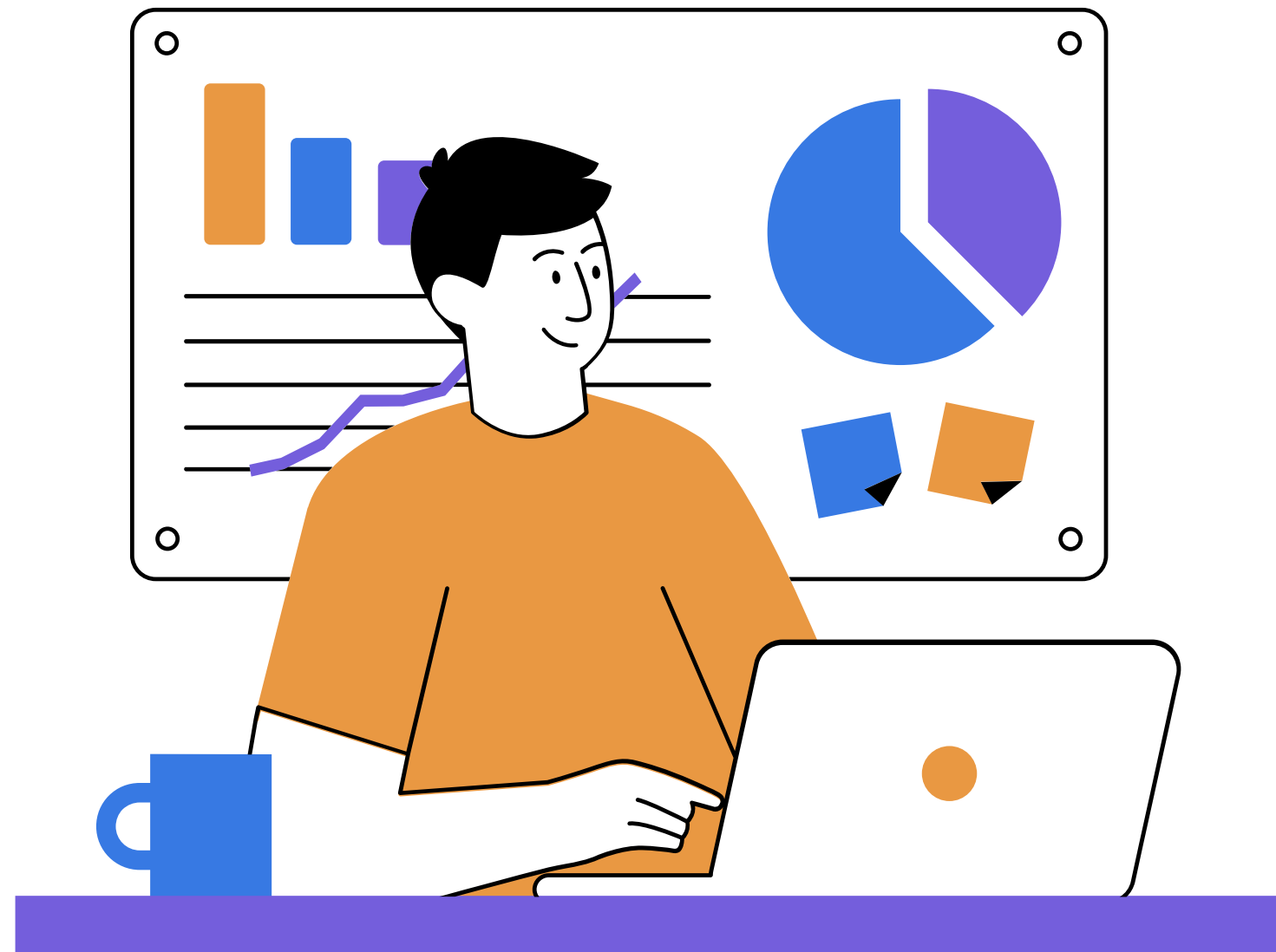
Learn

Data Science

Enroll Now



www.brightlighttutorials.co.za



What is a Data Scientist?

A **Data Scientist** is a multidisciplinary professional who leverages **statistical methods, programming, and domain expertise** to **extract actionable insights** and **build predictive models** from **complex and often large-scale data**. They bridge the gap between **raw data and strategic decision-making** by combining techniques from **statistics, machine learning, and data engineering**. A data scientist formulates the right questions and uses data to answer them: they design and execute experiments, **uncover patterns**, and **translate quantitative findings into clear, business-oriented recommendations**.

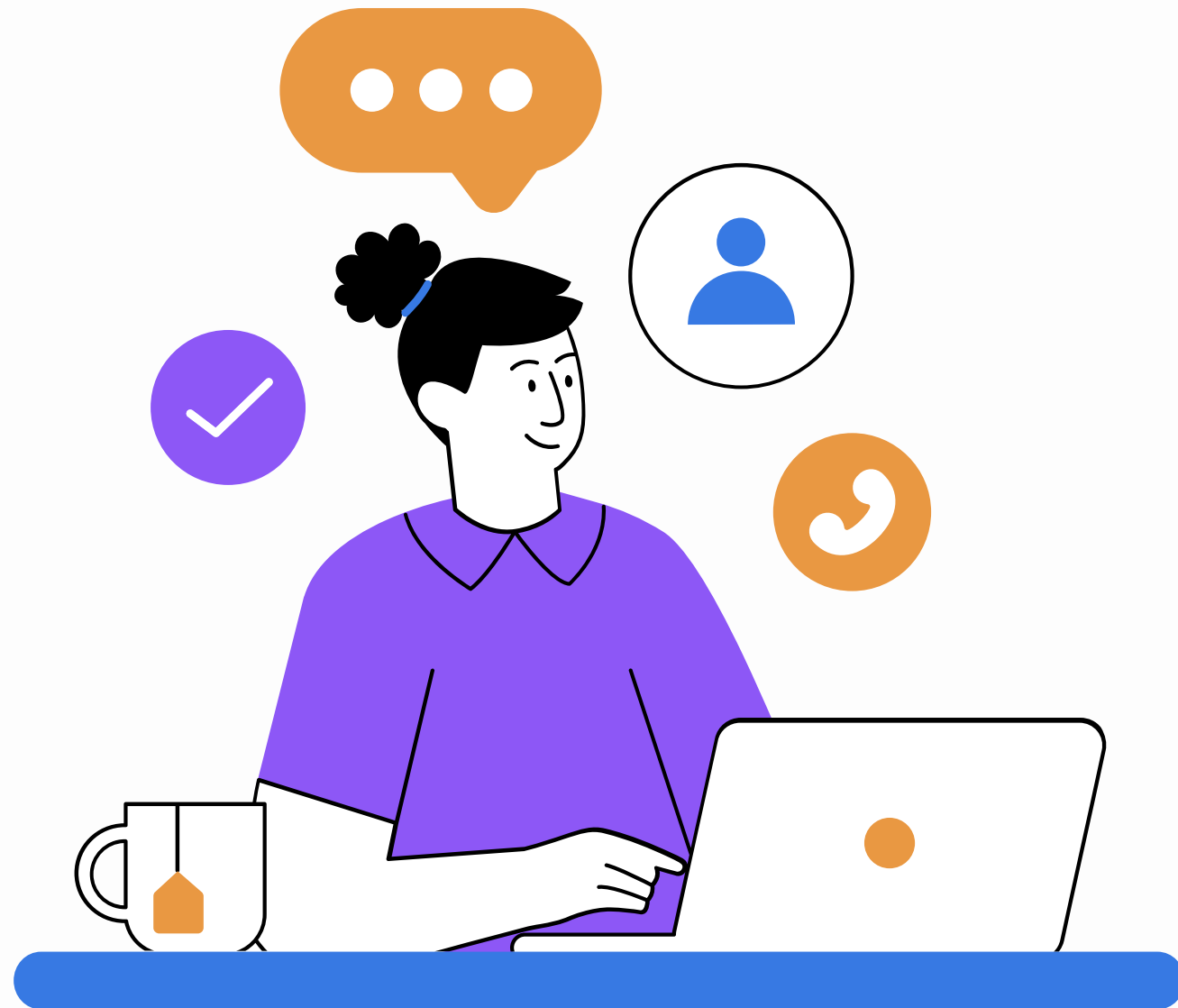
Impact & Outcomes

Business Optimization: Reduce costs, streamline operations, forecast demand, etc

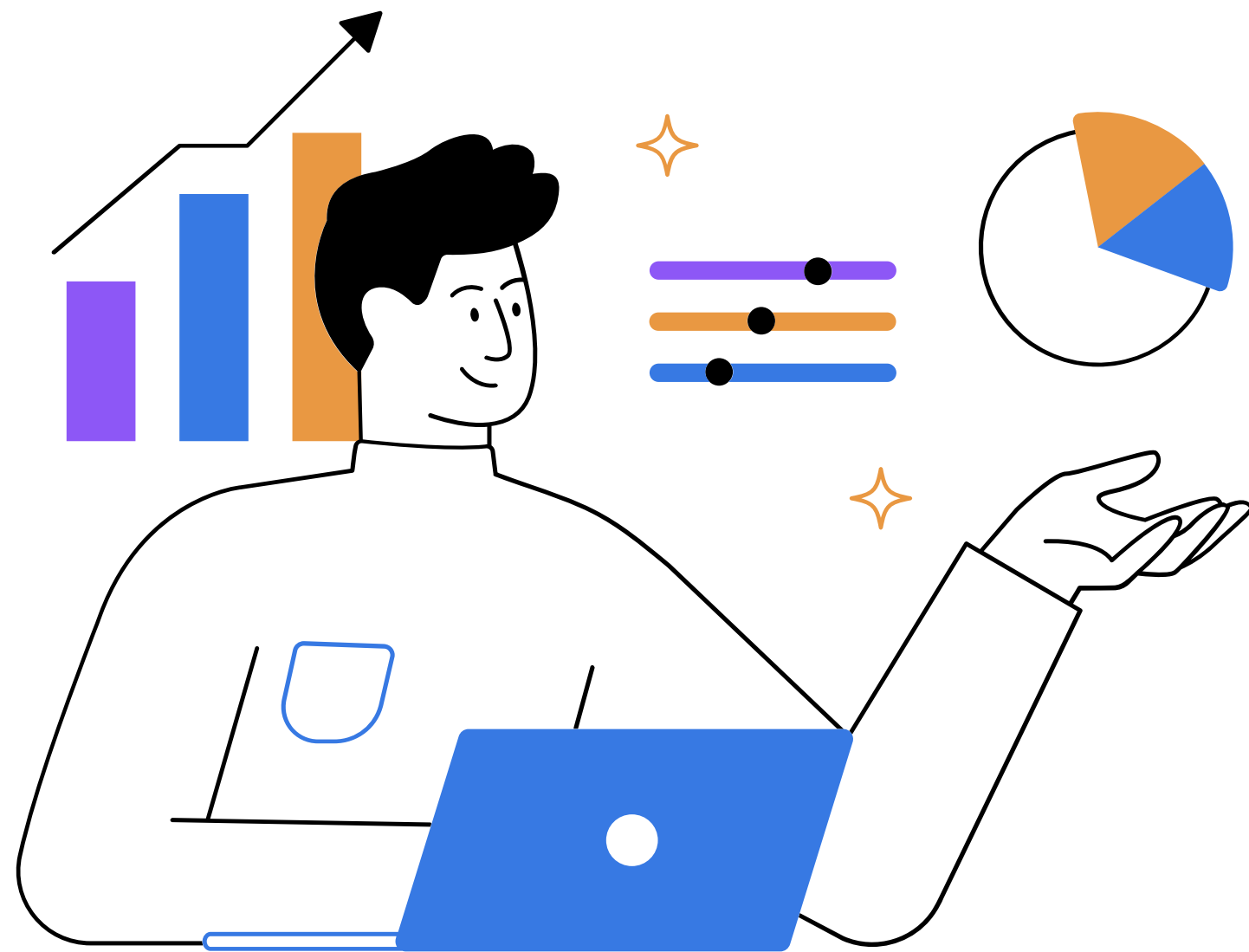
Product Innovation: Power recommendation engines, personalization features, fraud detection, etc

Strategic Insights: Inform executive decisions with data-driven reports and scenario analyses.

Key Responsibilities of a Data Scientist



- **Problem Framing:**
 - Collaborate with stakeholders to translate business challenges into analytical problems.
- **Data Acquisition & Wrangling:**
 - Gather data from multiple sources (databases, APIs, logs, third-party).
 - Clean, normalize, and integrate data to ensure quality and consistency.
- **Exploratory Data Analysis (EDA):**
 - Use descriptive statistics and visualization to understand distributions, relationships, and anomalies.
- **Model Development & Validation:**
 - Build and evaluate models (regression, classification, clustering, time-series forecasting, etc.).
 - Tune hyperparameters and assess performance via cross-validation, A/B testing, or backtesting.
- **Deployment & Monitoring:**
 - Collaborate with engineering to deploy models into production (APIs, batch pipelines).
 - Monitor model performance and retrain or recalibrate as data drifts.
- **Communication & Storytelling:**
 - Craft compelling narratives and dashboards (e.g., Power BI, Tableau) to convey insights to non-technical audiences.



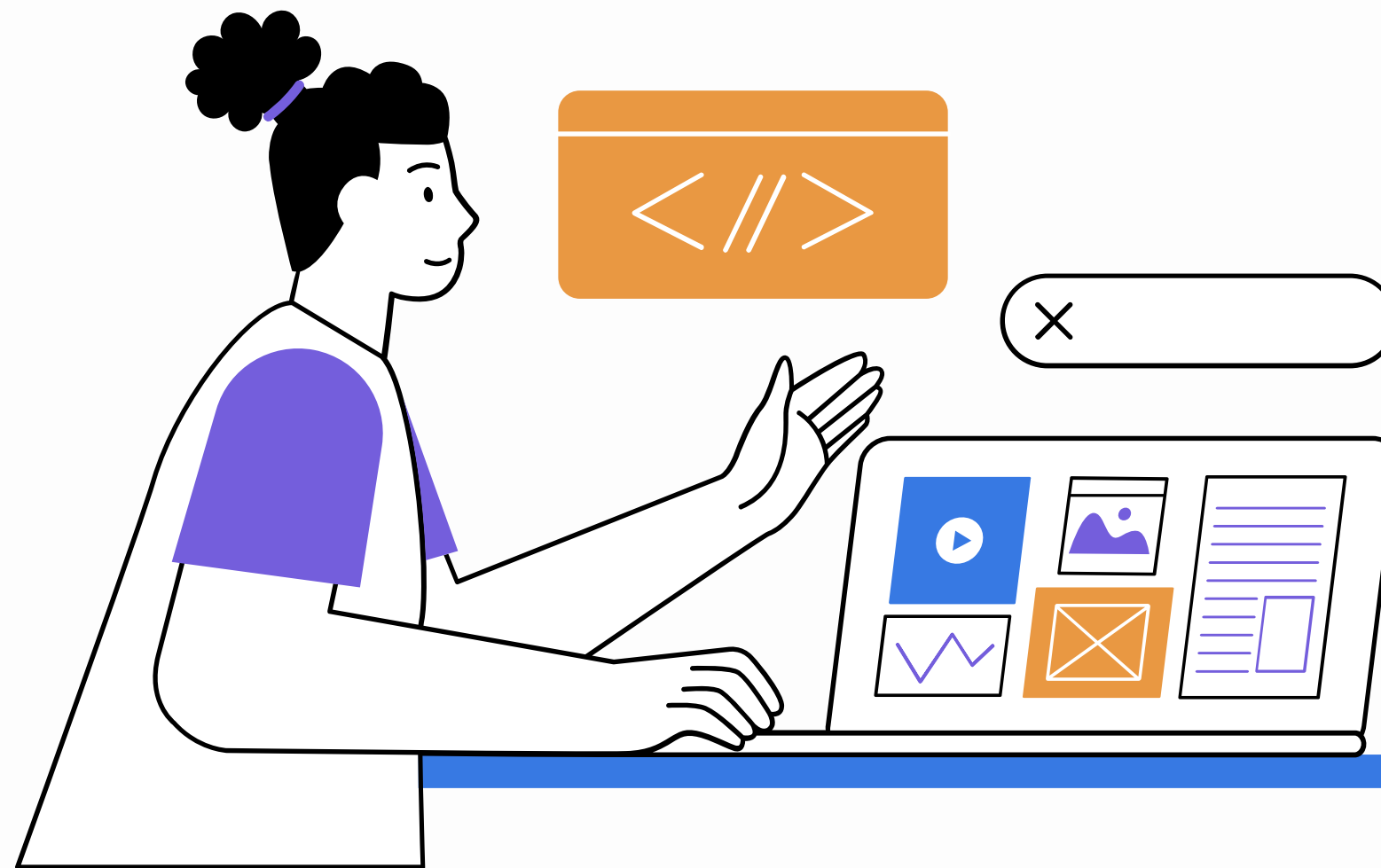
Essential Skills for Data Scientist

Technical Skills

- Data Manipulation
- Statistical Analysis
- Data Visualization
- Programming
- Database Management
- Data Integration and ETL Tools
- Version Control
- Artificial Intelligence
- Cloud Deployment
- Machine Learning
- Data Engineering
- Business Intelligence Tools

Soft Skills

- Problem-Solving
- Attention to Detail
- Communication Skills
- Collaboration and Teamwork
- Time Management
- Critical Thinking
- Adaptability
- Curiosity and Learning Mindset
- Presentation Skills
- Ethical Thinking and Integrity



Tools & Technologies

1. Programming:

- SQL (for querying relational databases)
- Python (Pandas, NumPy, SciPy, PySpark, Scikit-learn)

2. Data Visualization:

- Tableau, Google Looker Studio, Power BI
- Matplotlib, Seaborn, Plotly (Python library)

3. Cloud & Deployment:

- Google Cloud Platform (BigQuery)
- Amazon Web Services (AWS)
- Microsoft Azure
- Docker
- Kubernetes

4. Modelling:

- Hugging Face
- H2O
- Sciki Learn

5. Databases:

- MySQL
- Databricks
- Snowflake
- Microsoft SQL Server

6. Version Control & Collaboration:

- Git/GitHub

7. Data Engineering:

- ETL Pipelines (Airflow, SSIS, ADF)
- Hadoop

Typical Day-to-Day Activities

Morning Stand-up:

- Sync with the analytics team on progress, blockers, and priorities.

Data Exploration:

- Write SQL queries or notebook scripts (Python/R) to profile new datasets.

Feature Engineering:

- Create new variables to improve model accuracy (e.g., rolling averages, one-hot encodings, embeddings).

Modeling:

- Iterate on algorithms—train, validate, compare—and document findings.

Collaboration:

- Meet with product managers, marketers, or domain experts to refine use cases.

Reporting:

- Update dashboards, prepare slide decks, or draft technical blog posts.



Data Science RoadMap

Data Structures

Learn Data Fundamentals

- Arrays
- Linked list
- Stack (LIFO)
- Queue (FIFO)
- Tree
- Hash Table



Data Visualization

- Build interactive dashboards using tools like PowerBI & Google Looker Studio
- Data Analysis in Microsoft Excel



Machine Learning

- Learn Supervised Machine Learning (Classification and regression)
- Machine learning Processes covered includes EDA, Feature Engineering, Feature Selection, Model Training, etc



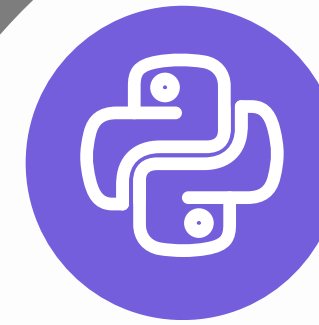
Introduction to Databases

- Relational Database
- Datawarehouse
- Cloud Data Systems
- Data Management Systems (Snowflake, Databricks, Microsoft SQL Server)



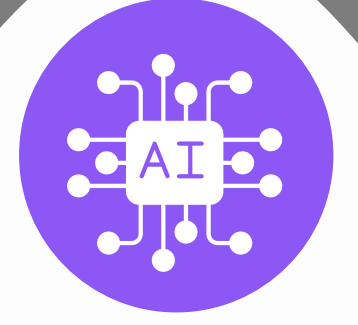
SQL Fundamentals & Advanced SQL

- Covering Basic to Advanced SQL syntax
- Data processing in Snowflake, Databricks, Big Query



Python Programming

- Python basic syntax,
- Python libraries such as Pandas, Matplotlib, Plotly express, pyspark.

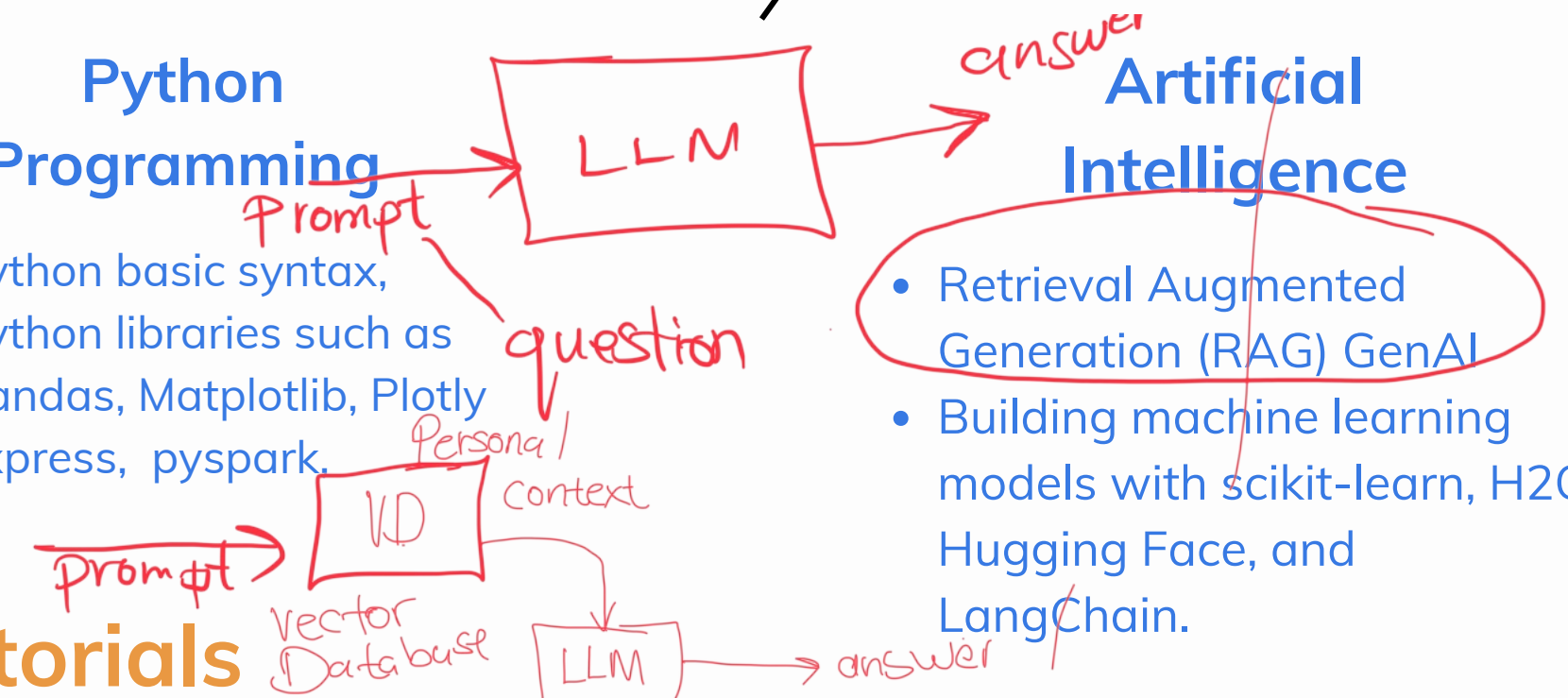


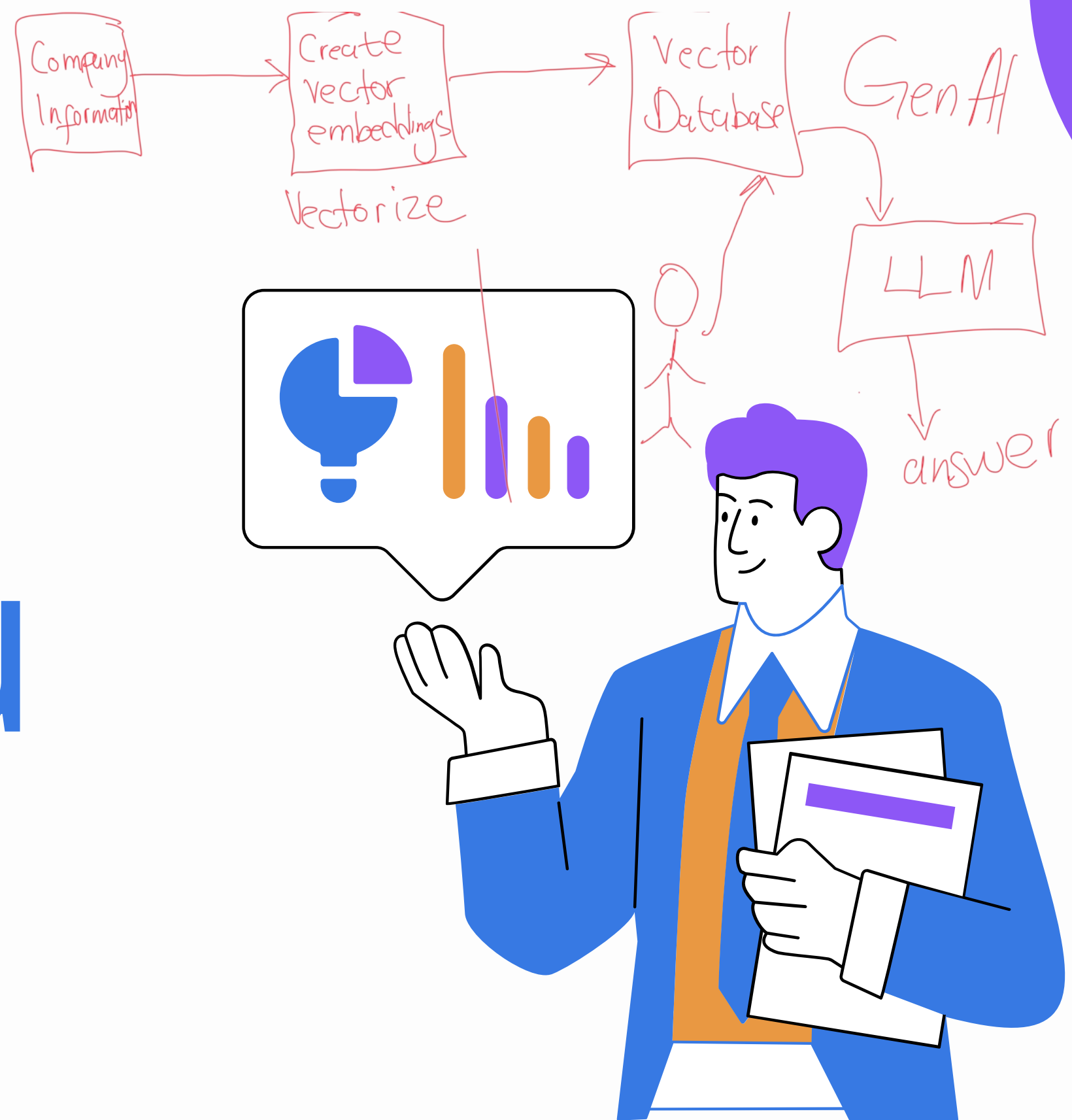
Artificial Intelligence

- Retrieval Augmented Generation (RAG) GenAI
- Building machine learning models with scikit-learn, H2O, Hugging Face, and LangChain.



BrightLight Tutorials





Thank You

● 078 737 2893

● rophiwa@brightlighttutorials.co.za

● www.brightlighttutorials.co.za