Data Analysis Process & Tools

What is data analysis?

Data analysis is the process of cleaning, changing, and processing raw data, and extracting actionable, relevant information that helps businesses make informed decisions.

- . The procedure helps reduce the risks inherent in decision-making by providing useful insights and statistics, often presented in charts, images, tables, and graphs.

- Several years ago, IBM predicted that the demand for data analysts would soar 28% by the end of 2020, resulting in more than 360,000 new positions this year. As per the reports of Salary.com, data analysts earn an annual average of USD 75,724, hitting over USD 85,000 in the high end of the range.

four types of data analytics

Statistical Analysis / Descriptive analytics

Descriptive analytics answers the question of what happened.

Diagnostic analytics

Descriptive analytics answers the question of Why is it happening?.
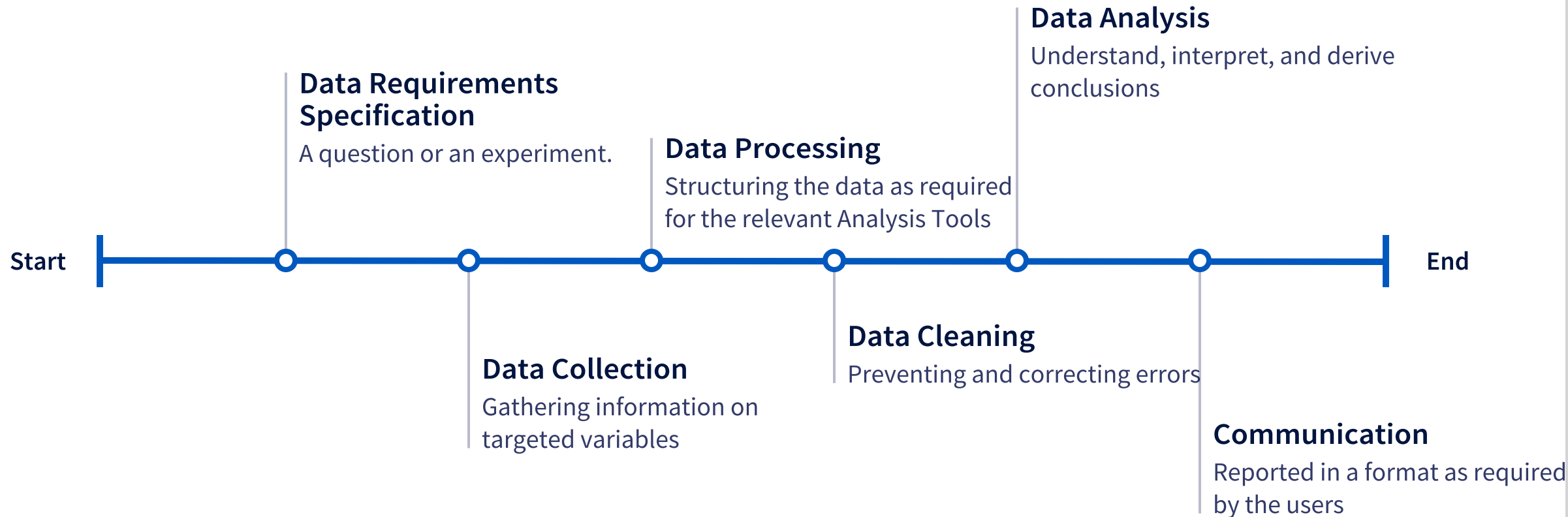
Predictive analytics

Predictive analytics tells what is likely to happen

Prescriptive analytics

The purpose of prescriptive analytics is to literally prescribe what action to take to eliminate a future problem or take full advantage of a promising trend.

AFRICA DATA SCHOOL

Data Analysis Process

Data Analysis

Understand, interpret, and derive conclusions

Data Requirements Specification

A question or an experiment.

Data Processing

Structuring the data as required for the relevant Analysis Tools

Start ———○———○———○———○———○———○——— End

Data Cleaning

Preventing and correcting errors

Data Collection

Gathering information on targeted variables

Communication

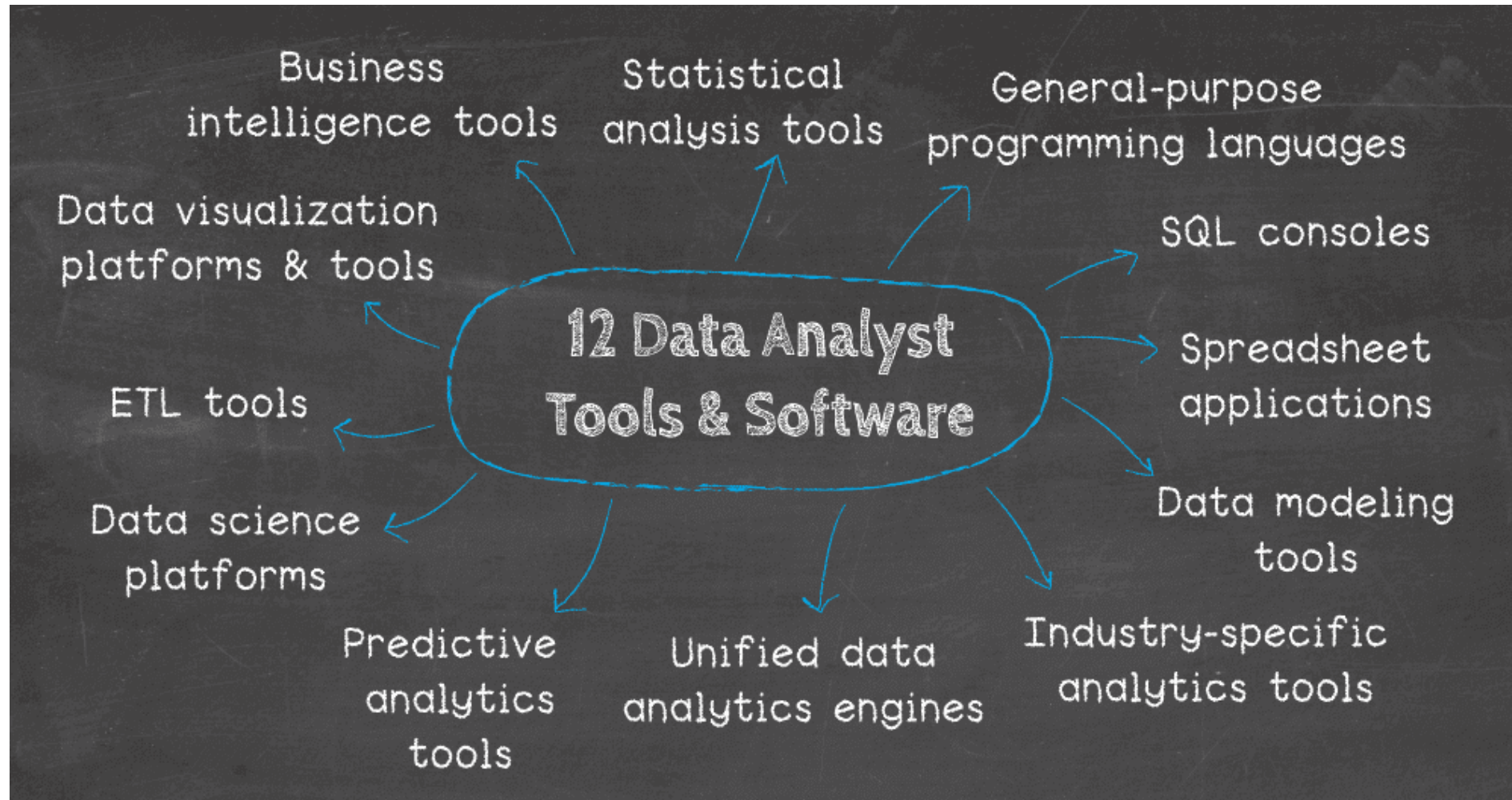Reported in a format as required by the users

AFRICA DATA SCHOOL

Data Analysis Methods

they all fall into one of two primary types: qualitative analysis and quantitative analysis.

- **Qualitative Data Analysis:** The qualitative data analysis method derives data via words, symbols, pictures, and observations. This method doesn't use statistics. The most common qualitative methods include
- **Quantitative Data Analysis:** Statistical data analysis methods collect raw data and process it into numerical data. Quantitative analysis methods include:

WHAT TOOLS DO DATA ANALYSTS USE?

TOOLS

Programming Language (Python)

Python has the highest popularity among data scientists. This is due to its wide range of uses. It is often the go-to choice for a range of tasks for domains, such as, machine learning, deep learning, artificial intelligence and other popular forms of technology.

PYPL PopularitY of Programming Language

Worldwide, Sept 2021 compared to a year ago:

Rank	Change	Language	Share	Trend
1		Python	29.48 %	-2.4 %
2		Java	17.18 %	+0.7 %
3		JavaScript	9.14 %	+0.8 %
4		C#	6.94 %	+0.6 %
5		PHP	6.49 %	+0.4 %
6		C/C++	6.49 %	+0.9 %
7		R	3.59 %	-0.5 %
8	↑↑↑	TypeScript	2.18 %	+0.3 %
9		Swift	2.1 %	-0.4 %
10	↓↓	Objective-C	2.06 %	-0.6 %

The PYPL PopularitY of Programming Language Index is created by analyzing how often language tutorials are searched on Google.

Code Editor （Jupyter Notebook & Pycharm）

- Jupyter Notebook - **Allows data analyst to create and share documents that integrate live code, equations, computational output, visualizations, and other multimedia resources**, along with explanatory text in a single document.

- Pycharm   - PyCharm is an integrated development environment (IDE) used in computer programming, specifically for the **Python languag**

Version Control (GIT & GITHUB)

- GIT - Git is a free and open source distributed version control system designed to handle everything from small to very large projects with speed and efficiency.
- GITHUB -   provider of Internet hosting for software development and version control using Git. It

DATA ANALYSIS (Pandas & SQL)

- Pandas - Pandas is a fast, powerful, flexible and easy to use open source data analysis and manipulation tool, built on top of the Python programming language.
- SQL - SQL (**Structured Query Language**) is a standardized programming language that's used to manage relational databases and perform various operations on the data in them

Data Visualization

- Seaborn - Seaborn is a Python data visualization library
- Plotly   -   Plotly provides **online graphing, analytics, and statistics tools for individuals and collaboration**, as well as scientific graphing libraries for Python,

Dashboard (Streamlit & Dash)

- Streamlit - Create beautiful data apps in hours, not weeks. All in pure Python

- Dash -   Dash analytics application framework is also free and open-source software, licensed under the MIT license

Practical

Download Anaconda and read CSV/ excel file