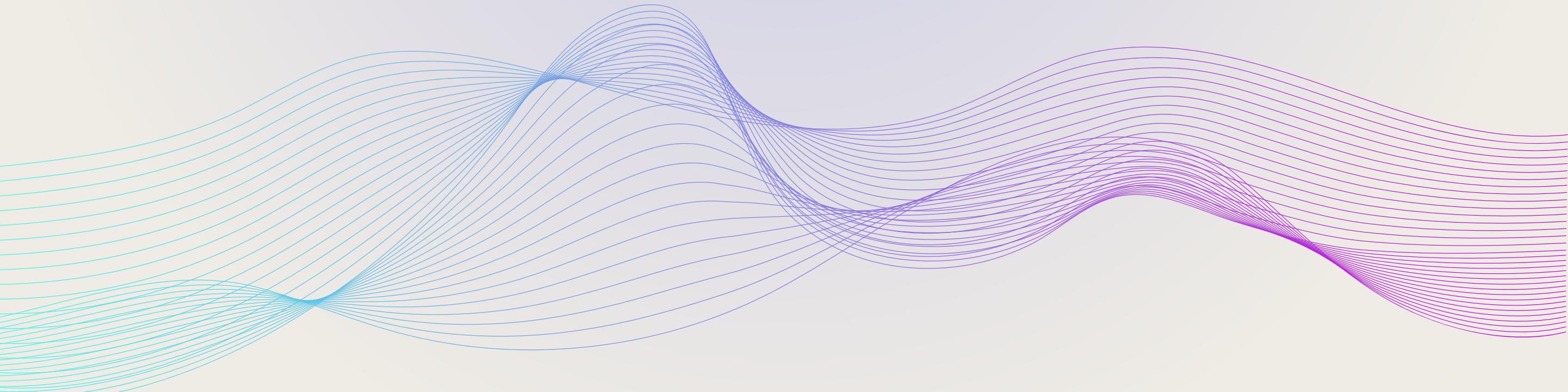


What is the
DATA SCIENCE
Process



I. BUSINESS UNDERSTANDING



Business understanding focuses on understanding objectives and requirements. In this phase, there are three other tasks that are universal to most projects.

- __ Determine business objectives.
- __ Assess situation
- __ Produce project plan

Establishing a strong business understanding is like building the foundation of a house, and it's absolutely essential.

II. DATA UNDERSTANDING



Next is the Data Understanding phase. The focus is added to identify, collect, and analyze data sets that can help you accomplish the project goals. This phase also has four tasks:

- _ Collect initial data
- _ Describe data
- _ Explore data
- _ Verify data quality

III. DATA PREPARATION



A common thumb rule is that 80% of the project is data preparation.

This phase, which is often referred to as “data munging”, prepares the final data set(s) for modeling.

It has five tasks:

- __ Select data
- __ Integrate data
- __ Clean data
- __ Format data
- __ Construct data

IV. MODELING

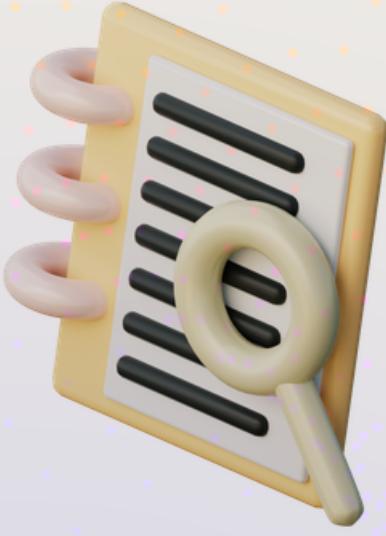


Data science's most exciting work is often the shortest phase of the project.

Here you'll likely build and assess various models based on several different modeling techniques. This phase has four tasks:

- _ Select modeling techniques
- _ Generate test design
- _ Build model
- _ Assess model

V. EVALUATION



The Modeling phase focuses on technical model assessment, whereas the Evaluation phase looks at which model best fits the business and what to do next. This phase has three tasks:

- __ Evaluate results
- __ Review process
- __ Determine next steps

VI. DEPLOYMENT



Depending on the requirements, the deployment phase can be simple or complex.

A model is not particularly useful unless the customer can access its results. The complexity of this phase varies widely. This final phase has four tasks:

- __ Plan deployment
- __ Plan monitoring and maintenance
- __ Produce final report
- __ Review project