



pandas

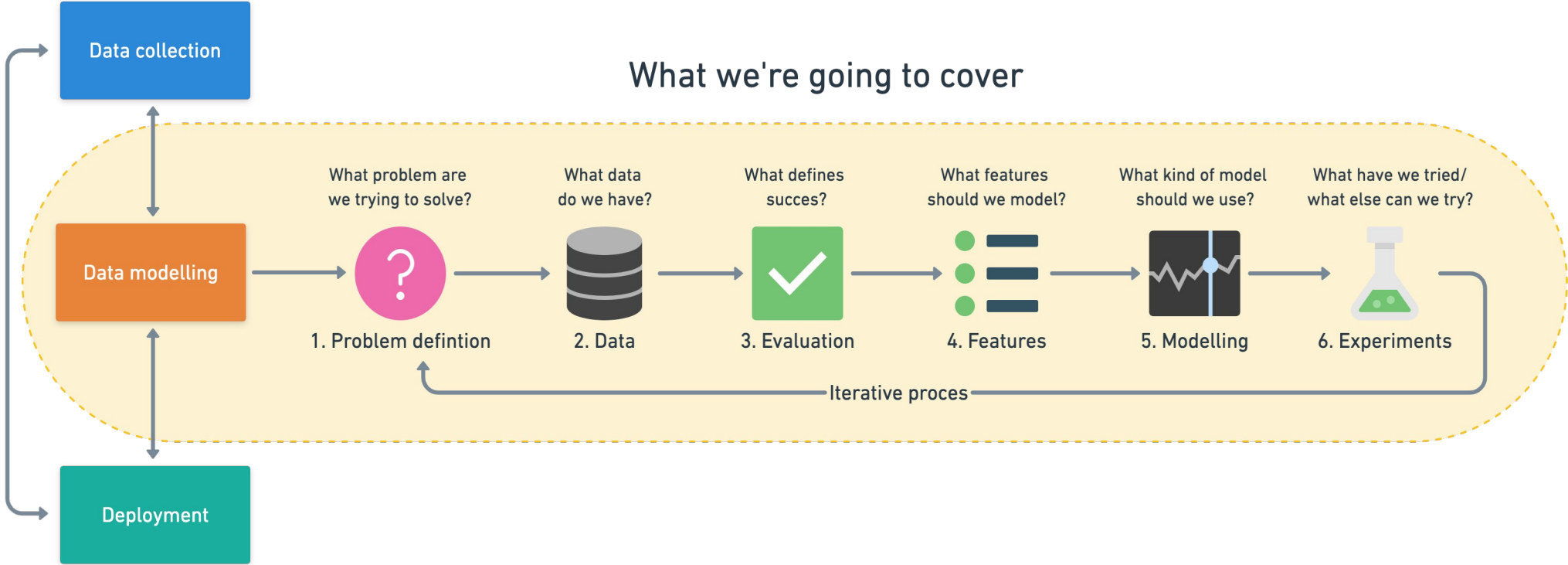


What is Pandas

- Pandas is a software library written for the Python programming language for data manipulation and analysis.
- Pandas is used to explore, analysed and manipulate data ready for machine learning.

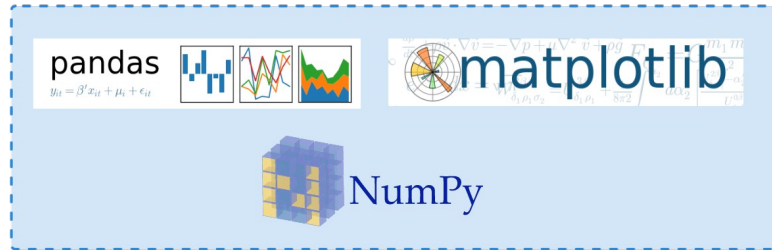
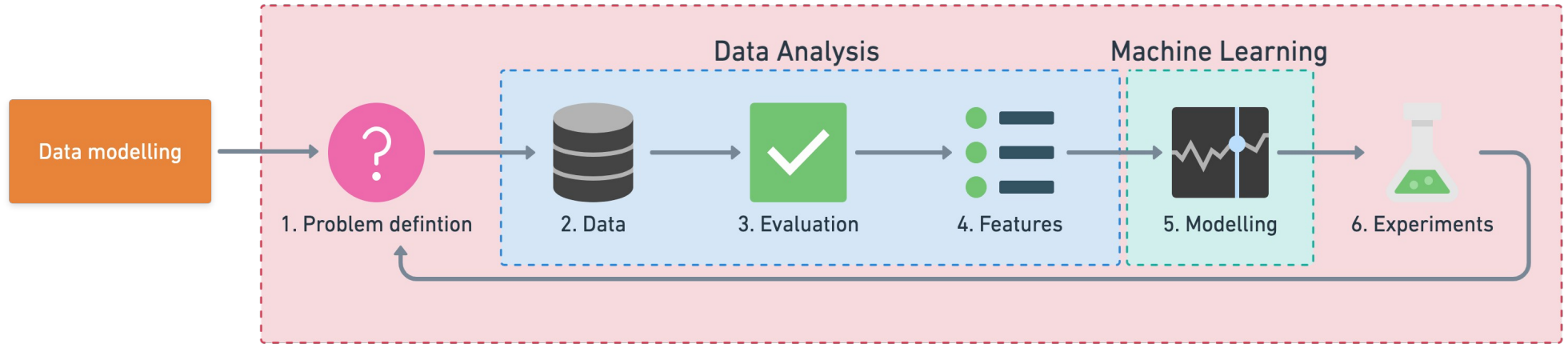


Steps in a full machine learning project



Tools you can use

Data Science



Why Pandas?

- It's simple to use
- It's written in python programming language
- Helps you set your data ready for machine learning algorithm
- Pandas gives your set of tools to do Data Analysis
- You can use it on various type of data file such as
Excel
Csv... etc



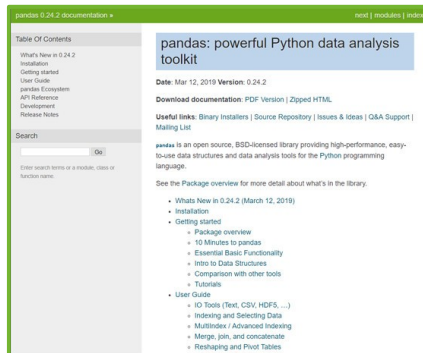
What we are going to cover

- Most useful functions
- Pandas Datatypes
- Importing and Exporting data
- Describing data
- Viewing and selection data
- Manipulating data



Where can I get help

- Follow along with the code
- Try it for yourself
- Stack overflow
- Pandas documentation



Let's code!



Anatomy of a DataFrame

The diagram illustrates the anatomy of a DataFrame using a table of car data. Annotations include:

- Column (axis = 1):** Points to the header row.
- Index number (starts at 0 by default):** Points to the first column of the data rows.
- Row (axis = 0):** Points to the first column of the data rows.
- Column name:** Points to the header row.
- Data:** Points to specific data cells in the first row.

| | Make | Colour | Odometer | Doors | Price |
|---|--------|--------|----------|-------|----------|
| 0 | Toyota | White | 150043 | 4 | \$4,000 |
| 1 | Honda | Red | 87899 | 4 | \$5,000 |
| 2 | Toyota | Blue | 32549 | 3 | \$7,000 |
| 3 | BMW | Black | 11179 | 5 | \$22,000 |
| 4 | Nissan | White | 213095 | 4 | \$3,500 |