

Python for Data Analysts

GETTING STARTED WITH PYTHON FOR DATA ANALYSIS



Janani Ravi

CO-FOUNDER, LOONYCORN

www.loonycorn.com

Overview

Choices for data analysts

Python's role in data analysis

Installing Python

Running your first Python program

Arithmetic and logical operations

Variables as basic building blocks

Basic data types

Single line and multi-line strings

Prerequisites and Course Outline

Prerequisites



**No prior programming experience
required**

Course Outline



Getting started with Python

Using built-in functions and complex data types

Using control structures

Code reuse using functions

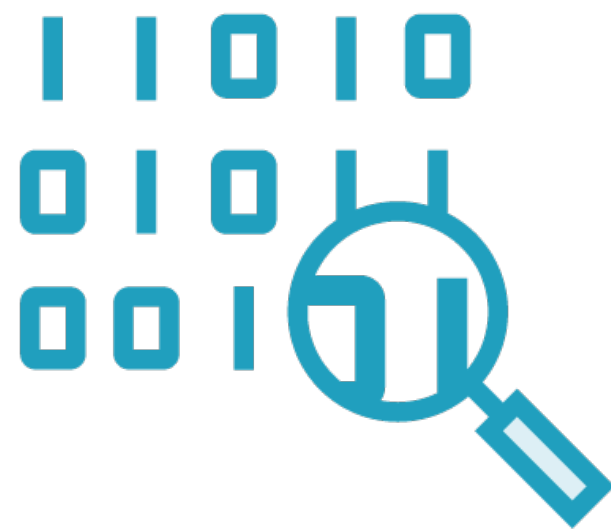
Loading and saving data

Python for Data Analysts

“My mind is made up. Don’t confuse me with the facts.”

Some powerful person

Thoughtful, Fact-based Point of View



Fact-based

Built with
painstakingly
collected data



Thoughtful

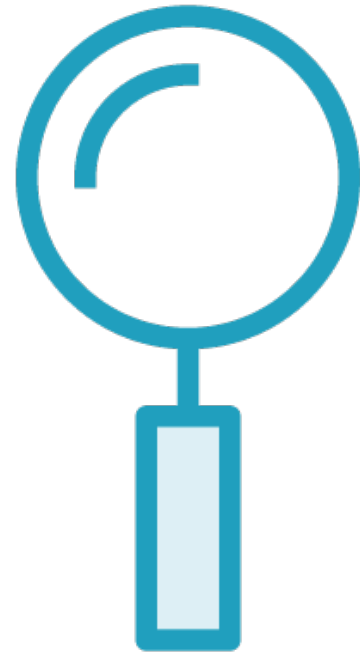
Balanced, weighing
pros and cons



Point of View

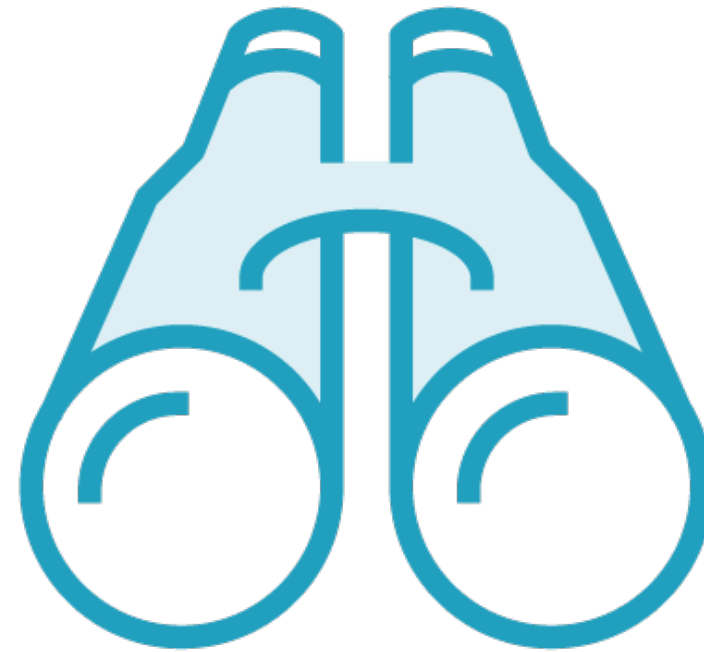
Prediction,
recommendation,
call to action

Two Sets of Statistical Tools



Descriptive Statistics

Identify important elements in a dataset



Inferential Statistics

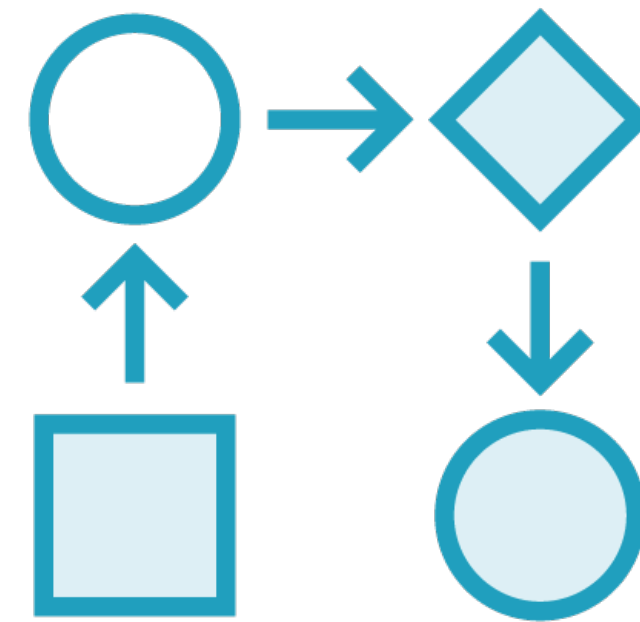
Explain those elements via relationships with other elements

Two Hats of a Data Professional



Find the Dots

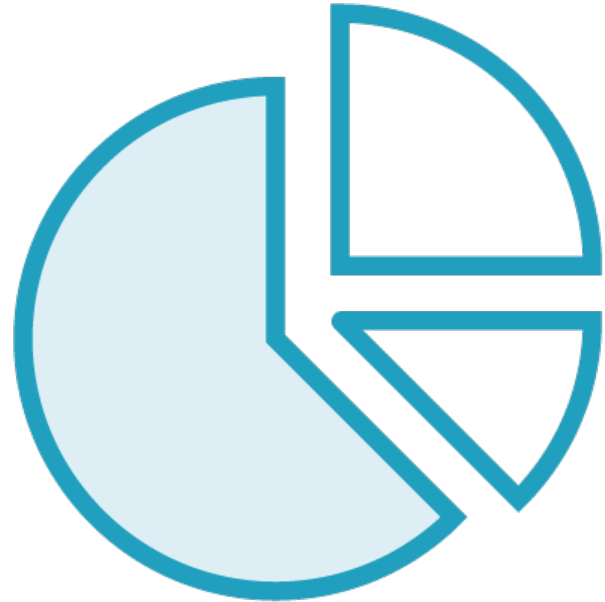
Identify important elements in a dataset



Connect the Dots

Explain those elements via relationships with other elements

Finding the Dots

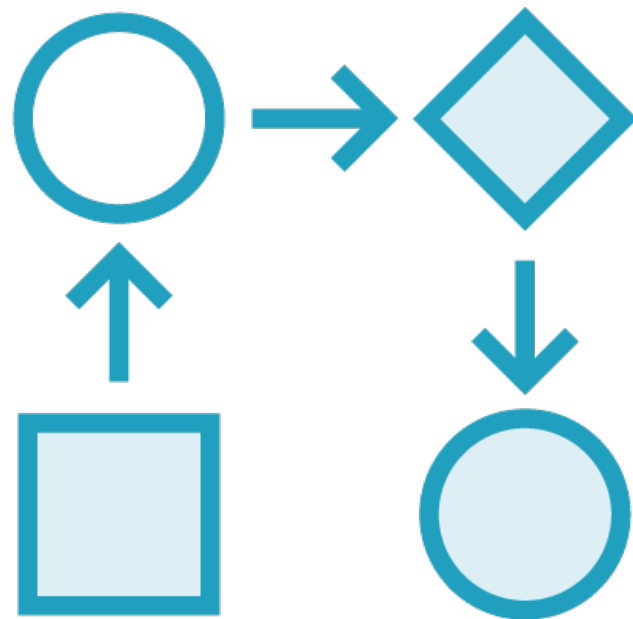


Data is more and more plentiful

However careful handling is needed

- Missing values
- Outliers
 - Genuine outliers
 - Erroneously measured points

Connecting the Dots



Spreadsheets

Programming languages

- In-memory processing
- Distributed processing

SQL

- Relational databases
- Data warehouses

Python has truly democratized
data analysis more than any
technology since Microsoft Excel

Choices of Technology

Microsoft Excel

Fast prototyping

Bad for production use

SQL Databases

Business users who can't code

Not yet Big Data; problem of silos

Data Warehouses

SQL for Big Data analytics

Streaming data, ML integrations

Python with Pandas

Fast prototyping in REPL environment

Still constrained to in-memory data

Python with Spark

Fast prototyping with Big Data

Truly powerful - still needs code to be written

Essential Analytical Building Blocks

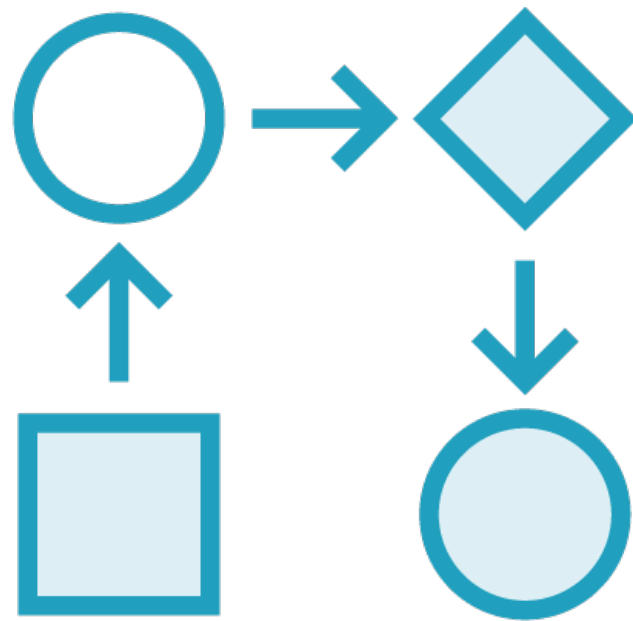
Conditional Execution

Interconnected Calculations

**Repeated Execution
(Iteration)**

**Re-use of Logic
(Composition)**

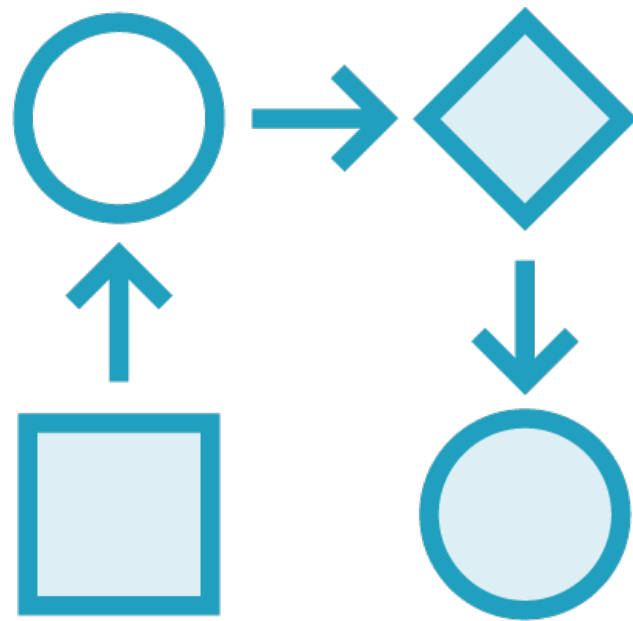
Choices of Technology



Spreadsheets are best for

- Complex inter-connected calculations
- Rapid prototyping

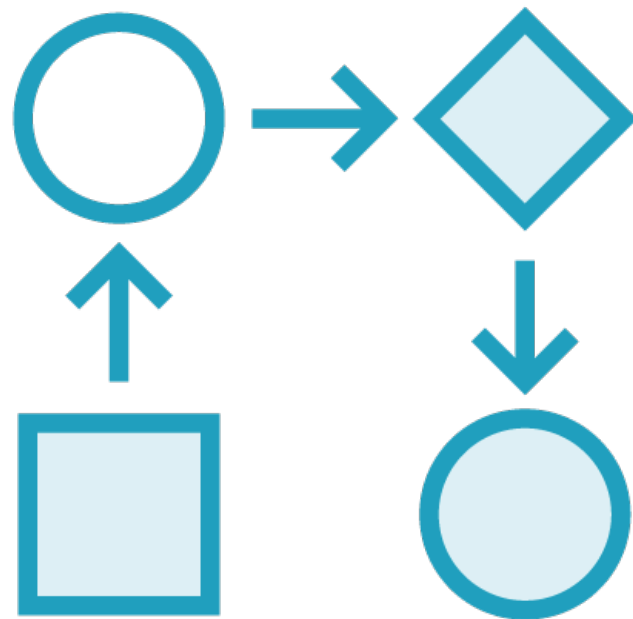
Choices of Technology



SQL is best for

- Iterating over independent rows
- Simple syntax

Spreadsheets for Analytics

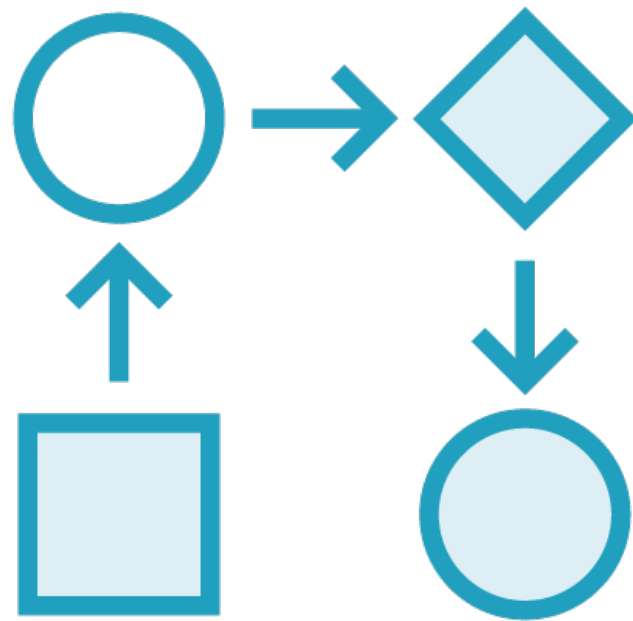


Conditionals: if() function within cells

Iteration: Copy-paste, or worse, macros

Composition: Not possible

SQL Databases for Analytics

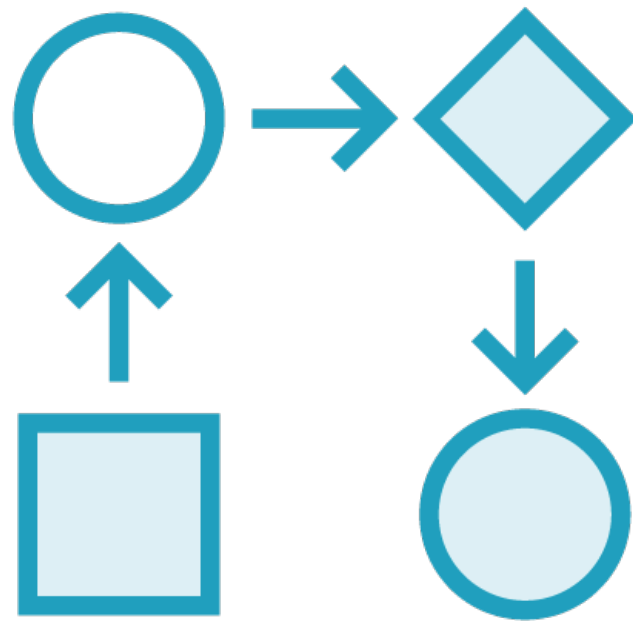


Conditionals: if() function within queries

Iteration: Queries, cursors

Composition: Views, stored procedures

Python for Analytics



Programming languages offer full support for analytical operations

Conditionals: If-else

Iteration: For and while loops

Composition: Functions

Python combines Excel's
ease-of-prototyping with
SQL's simple syntax

Demo

**Install and setup Python using
Anaconda on the Mac**

Demo

**Install and setup Python using
Anaconda on Windows**

Demo

**Executing simple arithmetic
operations in Python**

Demo

**Executing simple logical operations in
Python**

Demo

Using variables to store values

Demo

Working with basic types in Python

Demo

**Defining strings using single quotes,
double quotes, and triple quotes**

Summary

Choices for data analysts

Python's role in data analysis

Installing Python

Running your first Python program

Arithmetic and logical operations

Variables as basic building blocks

Basic data types

Single line and multi-line strings