Python

1991 → 2026

From a hobby language to PySpark —
do finance teams really need Spark
for 12 million rows?



Timeline part 1

1991
First public release

2000
Python 2

2006
NumPy

2008
pandas — born in a hedge fund

2008
Python 3

2014
Spark becomes an Apache top-level project

Timeline part 2

2018
Python in Power BI

2020
Python 2 retires

2024
Python in Excel

2025
Databricks Free Edition

2026
Python 3.15

May 2027
Python visuals in Power BI retire

Plain Python vs libraries



Plain Python — the language and its standard library

- pandas
- DuckDB
- PySpark
- NumPy
- matplotlib
- openpyxl

Plain Python is Excel without formulas.

Libraries are the SUMIFS, pivot tables and Power Query.

Where Python lives in 2026

Excel

(Python in a cell)

Power Query

(Python script)

Power BI visuals

(retiring May 2027)

Microsoft Fabric

(Python and Spark notebooks)

Databricks Free Edition

Your laptop

The test



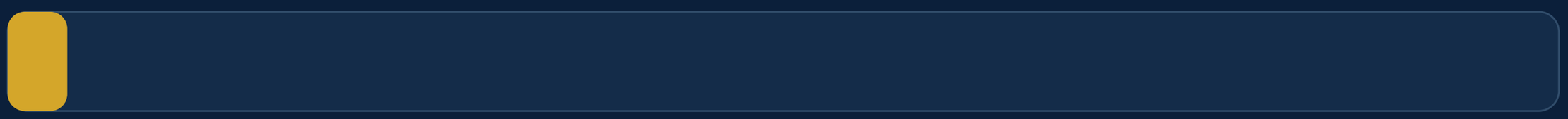
About 12 million GL lines

368 MB Parquet (1.8 GB as CSV)

Same 4 finance questions

4 engines — identical answers

The results



2-core cloud machine, 7 GB RAM. Your laptop will differ.

Why Spark came third

Spark spends about 5 seconds just starting up.
It is built for data spread across many machines
— not for one laptop.



Plain Python finished too — but only after the Parquet file was turned into a 1.8 GB CSV.

When to pick which

DuckDB

SQL on files, fastest on one machine

pandas

Tables in Python, the most tutorials
and examples

PySpark

When data no longer fits one machine,
or a team shares a lakehouse

Download the ledger and race them yourself

findi360.com/data-resources/python-history-pandas-duckdb-pyspark

Free dataset: Data Dare DD-MFG-005