



Bootcamp Data Science

Python, DuckDB & Streamlit

Pembelajaran Praktis dengan Dataset Nyata

Durasi: 2 Hari (16 Jam Total)

Selamat Datang!

Tentang Bootcamp

Bootcamp intensif ini dirancang untuk membekali Anda dengan keterampilan praktis dalam:

-  **Python & Pandas** untuk manipulasi dan analisis data
-  **DuckDB** untuk query analitik yang powerful
-  **Visualisasi Data** dengan Matplotlib, Seaborn, dan Plotly
-  **Streamlit** untuk membuat dashboard interaktif

Pembelajaran berbasis praktik menggunakan dataset nyata: **RUP 2025**



Dataset: RUP 2025

Rencana Umum Pengadaan Barang/Jasa Pemerintah

Karakteristik Dataset

- **Records:** 16,430 paket
- **Size:** ~1.3 MB (Parquet)
- **Tahun:** 2025
- **Source:** Data Pengadaan Pemerintah

Kolom Utama

- Nama paket & kode RUP
- Pagu anggaran
- Metode pengadaan
- Satuan kerja
- Timeline pengadaan
- Lokasi dan kategori

Data dunia nyata = Keterampilan dunia nyata!

Tujuan Pembelajaran

Setelah bootcamp ini, Anda akan mampu:

1. Analisis Data dengan Python & Pandas

-  Membaca dan mengeksplorasi data (melihat pola & tren)
-  Membersihkan data yang kotor/tidak lengkap
-  Mengolah data untuk mendapatkan informasi

2. Menulis Query dengan DuckDB

-  Membuat query SQL untuk analisis data
-  Menggabungkan DuckDB dengan Pandas
-  Membuat query yang lebih cepat



Tujuan Pembelajaran (lanjutan)

3. Membuat Visualisasi Data

- Membuat grafik statis dengan Matplotlib & Seaborn
- Membuat grafik interaktif dengan Plotly
- Menyajikan data dengan cara yang mudah dipahami

4. Membangun Dashboard dengan Streamlit

- Membuat aplikasi web tanpa perlu jago coding web
- Menambahkan filter interaktif di dashboard
- Membuat dashboard yang siap dipakai



Tujuan Pembelajaran (lanjutan)

5. Teknik Analisis Dasar & Lanjutan

- Menganalisis data berdasarkan waktu (time series)
- Melakukan uji statistik sederhana
- Dasar-dasar A/B testing

Target Peserta

Bootcamp ini cocok untuk:

-  **Data Analyst** yang ingin meningkatkan keterampilan
-  **Business Analyst** yang ingin lebih teknis
-  **Programmer** yang ingin masuk ke Data Analytics
-  **Fresh Graduate** yang ingin berkarir di bidang data
-  **Profesional** yang perlu mengolah & memvisualisasikan data

Prasyarat

Yang Perlu Anda Siapkan:

Pengetahuan

-  Pernah coding Python (walau sedikit) - **pemula boleh ikut!**
-  Pernah dengar SQL (tidak harus jago)
-  Suka berpikir logis dan memecahkan masalah

Hardware & Software

-  Laptop dengan minimal **8GB RAM**
-  **Python 3.8+** terinstal (akan dibantu saat setup)
-  Text editor seperti VS Code (gratis & mudah)
-  Github - untuk menyimpan kode

Tools & Libraries

Alat Utama

- **Python** - Bahasa pemrograman yang mudah
- **Jupyter Notebook** - Tempat nulis kode & lihat hasilnya langsung
- **uv** - Untuk install library (otomatis & mudah)
- **Git** - Untuk backup kode

Untuk Visualisasi

- **Plotly** - Buat grafik yang bisa di-klik & zoom
- **Matplotlib** - Buat grafik sederhana
- **Seaborn** - Buat grafik statistik cantik

Untuk Dashboard

- **Streamlit** - Buat web app tanpa perlu HTML/CSS/JS
- Mudah banget, cukup Python!

Tools & Libraries (lanjutan)

Untuk Olah Data

- **Pandas** - Untuk baca & olah data (seperti Excel tapi lebih powerful)
- **NumPy** - Untuk hitung-hitungan angka
- **DuckDB** - Untuk query data dengan SQL (cepat!)



Hasil Akhir yang Diharapkan

Portofolio Anda akan berisi:

- **✓ 7 Jupyter Notebooks** dengan analisis lengkap
 - EDA, DuckDB, Visualisasi, Pembersihan Data, Time Series, Analisis Statistik
- **✓ 4 Aplikasi Streamlit** interaktif
 - Hello Streamlit, Demo Komponen, Data Explorer, Dashboard RUP
- **✓ 1 Dashboard Siap Produksi**
 - Analisis komprehensif RUP 2025
- **✓ Pemahaman mendalam** alur kerja analisis data end-to-end



Agenda Bootcamp

Hari 1: Fundamental Analisis Data (8 Jam)

Waktu	Sesi	Topik
08:00 - 12:00	Sesi 1	Python & Pandas untuk Analisis Data
12:00 - 13:00	BREAK / ISHOMA	Istirahat & Makan Siang
13:00 - 15:00	Sesi 2	DuckDB untuk Query Analitik
15:00 - 15:30	BREAK	Coffee Break
15:30 - 17:00	Sesi 3	Visualisasi Data

JUL
17

Agenda Bootcamp

Hari 2: Analisis Lanjutan & Dashboard (8 Jam)

Waktu	Sesi	Topik
08:00 - 12:00	Sesi 4	Teknik Analisis Data Lanjutan
12:00 - 13:00	BREAK / ISHOMA	Istirahat & Makan Siang
13:00 - 15:00	Sesi 5	Dashboard Interaktif dengan Streamlit
15:00 - 15:30	BREAK	Coffee Break
15:30 - 17:00	Sesi 5 (lanjutan)	Deployment & Presentasi Dashboard



Struktur Project

```
data-science/
└── day1/
    ├── session1_python_pandas/
    │   └── notebooks/01_exploratory_data_analysis_rup.ipynb
    ├── session2_duckdb/
    │   └── notebooks/01_duckdb_intro.ipynb
    └── session3_visualization/
        └── notebooks/
            ├── 01_matplotlib_seaborn.ipynb
            └── 02_plotly_interactive.ipynb
└── day2/
    ├── session4_advanced_analysis/
    │   └── notebooks/ (3 notebooks)
    └── session5_streamlit/
        └── apps/ (4 aplikasi)
└── datasets/rup/
```

Setup Environment

Memulai Cepat

```
# 1. Clone repository
git clone <repository-url>
cd data-science

# 2. Install dependencies dengan uv
uv sync

# 3. Jalankan Jupyter Notebook
uv run jupyter notebook

# 4. Jalankan Streamlit App (nanti di hari 2)
uv run streamlit run day2/session5_streamlit/apps/rup_dashboard.py
```

Setup Environment (Alternatif)

Menggunakan pip

```
# 1. Buat virtual environment
python -m venv .venv

# 2. Aktivasi virtual environment
# Windows:
.venv\Scripts\activate
# Linux/Mac:
source .venv/bin/activate

# 3. Install dependencies
pip install pandas numpy duckdb streamlit plotly \
          matplotlib seaborn jupyter openpyxl pyarrow

# 4. Jalankan Jupyter
jupyter notebook
```



Metode Pembelajaran

Pembelajaran Hands-On

Bootcamp ini menggunakan pendekatan **belajar dengan praktik**:

1.  **Teori Singkat** (20%)
 - Konsep fundamental
 - Praktik terbaik
2.  **Live Coding** (30%)
 - Demonstrasi instruktur
 - Ikuti bersama



Metode Pembelajaran (lanjutan)

Pembelajaran Hands-On

3.  **Praktikum (50%)**
 - Latihan hands-on
 - Masalah dunia nyata
 - Bangun solusi Anda sendiri



Tips untuk Sukses

Belajar Maksimal

- **✓ Ikuti semua latihan hands-on**
 - Jangan hanya menonton, coding bersama!
- **✓ Bertanya jika ada yang tidak jelas**
 - Tidak ada pertanyaan yang bodoh
- **✓ Bereksperimen dengan kode**
 - Coba variasi, rusak sesuatu, belajar dari kesalahan
- **✓ Catat wawasan penting**
 - Catatan untuk nanti



Tips untuk Sukses (lanjutan)

Belajar Maksimal

- **Berkolaborasi dengan peserta lain**
 - Belajar dari rekan, berbagi pengetahuan

Sumber Belajar

Dokumentasi Resmi

- **Pandas:** <https://pandas.pydata.org/docs/>
- **DuckDB:** <https://duckdb.org/docs/>
- **Streamlit:** <https://docs.streamlit.io/>
- **Plotly:** <https://plotly.com/python/>

Cheat Sheets

- Pandas Cheat Sheet
- DuckDB SQL Reference
- Streamlit Cheat Sheet
- Python Graph Gallery



Hasil Pembelajaran - Detail

HARI 1

Sesi 1: Pandas

- Memuat & memeriksa data
- Penyaringan & pemilihan
- GroupBy & agregasi
- Penanganan nilai yang kosong
- Ringkasan statistik

Sesi 2: DuckDB

- Dasar-dasar SQL
- Window functions
- CTE & subquery
- Integrasi Pandas

Hasil Pembelajaran - Detail (lanjutan)

HARI 2

Sesi 3: Visualisasi

- Dasar-dasar Matplotlib
- Plot statistik Seaborn
- Interaktivitas Plotly
- Panduan pemilihan grafik

Sesi 4: Lanjutan

- Teknik pembersihan data
- Analisis time series
- Pengujian statistik

Sesi 5: Streamlit

- Komponen interaktif
- Membangun dashboard
- Deployment produksi

🏆 Apa yang Membuat Bootcamp Ini Berbeda?

Yang Membuat Bootcamp Ini Berbeda

1. Dataset Dunia Nyata

- Bukan dataset mainan, tapi data pengadaan barang dan jasa pemerintah asli

2. Belajar Lengkap dari Awal sampai Akhir

- Dari data mentah → wawasan → dashboard produksi

3. Tools Modern yang Banyak Dipakai

- DuckDB untuk performa, Streamlit untuk kesederhanaan

4. Keterampilan yang Langsung Bisa Dipakai Kerja

- Bukan hanya tutorial, tapi keterampilan untuk bekerja

🏆 Apa yang Membuat Bootcamp Ini Berbeda? (lanjutan)

Yang Membuat Bootcamp Ini Berbeda

5. Membangun Portofolio 📁

- 7 notebook + 4 aplikasi = portofolio yang mengesankan

Langkah Selanjutnya

Persiapan Sebelum Mulai

Daftar Periksa Pra-Bootcamp

- [] Instal Python 3.8+
- [] Instal VS Code atau editor favorit
- [] Instal Git (jika diperlukan)
- [] Clone repository
- [] Siapkan virtual environment
- [] Instal dependencies (`uv sync` atau `pip install`)
- [] Tes dengan `jupyter notebook`
- [] Unduh/verifikasi dataset RUP 2025

? FAQ - Pertanyaan yang Sering Diajukan

T: Apakah saya harus sudah mahir Python?

J: Tidak perlu mahir, tapi sebaiknya sudah familiar dengan sintaks dasar

T: Apakah perlu latar belakang statistik?

J: Tidak wajib, konsep statistik dasar

T: Apakah bisa menggunakan dataset sendiri?

J: Ya! Setelah paham alur kerja dengan RUP, bisa menggunakan dataset lain

T: Bagaimana jika tertinggal?

J: Semua materi ada di notebook, bisa direview nanti

T: Apakah ada sertifikat?

J: Yang penting adalah keterampilan dan portofolio yang dihasilkan!



Ready to Start?

Mari Mulai Perjalanan Data Analytics Anda!

Hari 1 - Sesi 1

Python & Pandas untuk Analisis Data

Let's dive in! 💪

Terima Kasih! 🙏

Good Luck & Happy Learning!

Pertanyaan sebelum kita mulai?

Tetap Terhubung

Resources & Support

-  **Nama:** [Kurnia Ramadhan,ST.,M.Eng]
-  **Email:** [kurnia@ramadhan.me]

We're here to support your journey! 