

# WORKSHOP

Data Analytics With Python & SQL

Bakti Siregar, M.Sc

2023-06-29



# Contents

<b>Kata Pengantar</b>	<b>5</b>
Maksud & Tujuan . . . . .	5
Jadwal Kegiatan . . . . .	5
Pemateri & Mediator . . . . .	7
Para Peserta . . . . .	8
Manfaat dan Fasilitas . . . . .	8
Syarat Peserta . . . . .	9
Masukan & Saran . . . . .	9
Ucapan Terima Kasih . . . . .	9
 <b>1 Pendahuluan</b>	 <b>11</b>
1.1 Apa itu SBD? . . . . .	11
1.2 Mengapa Python & SQL? . . . . .	13
1.3 MySQL vs PostgreSQL . . . . .	13
1.4 Instalasi Python (Anaconda) . . . . .	14
1.5 Instalasi MySQL (XAMPP) . . . . .	15
1.6 Instalasi PostgreSQL . . . . .	16
1.7 Praktikal (Hands On) . . . . .	18
1.8 Latihan . . . . .	19
 <b>2 Pemilihan Data</b>	 <b>21</b>
 <b>3 Pengelompokan Data</b>	 <b>23</b>

4	Perhitungan Waktu & Subkueri	25
5	Pemahaman 5W1H & Visualisasi	27
6	Persentasi Tugas Akhir	29
7	Penutup	31
		33

# Kata Pengantar

Di zaman serba digital seperti sekarang ini, data menjadi sumber rujukan utama dalam setiap perencanaan, pengendalian, dan pengembangan diberbagai aspek kehidupan kita. Dampak dari transformasi digital ini adalah semakin bertambahnya kuantitas data yang terus meningkat seperti air mengalir. Sehingga dibutuhkan pengendalian data dengan cepat, tepat dan akurat. Kemampuan dalam menggunakan bahasa pemrograman Python & SQL adalah jawaban bagi setiap individu (professional) yang bekerja dengan data tersebut. Perlu diketahui bahwa Python & SQL sudah terintegrasi dengan sangat baik, mudah dipelajari, fleksibel, dapat digunakan untuk mengakses, memperbarui, memodifikasi, menganalisis data, dan umumnya digunakan dalam basis data relasional yang sangat besar.

## Maksud & Tujuan

Maksud dari kegiatan ini adalah melanjutkan pelatihan data analitik sebelumnya yang berjudul Workshop Data Analytics with Python. Pada workshop kali ini akan digunakan Python dan SQL dengan melibatkan karyawan atau pimpinan instansi masing-masing di JNE. Tujuan yang ingin dicapai dari dilaksanakan pelatihan komputer aplikasi ini adalah :

- a. Mengenalkan kepada para karyawan pentingnya penguasaan Python dan SQL dalam mengolah data.
- b. Melatih untuk menggunakan Python dan SQL dalam praktek pengolahan data sesuai lingkup pekerjaan.
- c. Menerapkan penggunaan Python dan SQL dalam pekerjaan.

## Jadwal Kegiatan

Adapun rencana jadwal kegiatan dan topik pembahasan adalah sebagai berikut

**Pertemuan 1**

Waktu	Nama Kegiatan
09:00 ~ 09:15 WIB	Pembukaan
09:15 ~ 10:15 WIB	Materi I: Pengenalan Python+SQL
10:15 ~ 10:30 WIB	Coffee Break I
10:30 ~ 11:30 WIB	Materi II: Instalasi Python+SQL
11:30 ~ 13:00 WIB	ISOMA
13:00 ~ 14:00 WIB	Materi III : Koneksi Python+SQL
14:00 ~ 15:15 WIB	Materi IV : Pengantar Python+SQL
15:15 ~ 15:30 WIB	ISO
15:30 ~ 16:30 WIB	Kuis atau Praktikum oleh Peserta

**Pertemuan 2**

Waktu	Nama Kegiatan
09:00 ~ 10:15 WIB	Materi I: Pemilahan Data
10:15 ~ 10:30 WIB	Coffee Break I
10:30 ~ 11:30 WIB	Materi II: Pemilahan Data
11:30 ~ 13:00 WIB	ISOMA
13:00 ~ 14:00 WIB	Materi III : Meringkas Data
14:00 ~ 15:15 WIB	Materi IV : Meringkas Data
15:15 ~ 15:30 WIB	ISO
15:30 ~ 16:30 WIB	Kuis atau Praktikum oleh Peserta

**Pertemuan 3**

Waktu	Nama Kegiatan
09:00 ~ 10:15 WIB	Materi I: Mengelompokkan Data
10:15 ~ 10:30 WIB	Coffee Break I
10:30 ~ 11:30 WIB	Materi II: Mengelompokkan Data
11:30 ~ 13:00 WIB	ISOMA
13:00 ~ 14:00 WIB	Materi III : Menggabungkan Data
14:00 ~ 15:15 WIB	Materi IV : Menggabungkan Data
15:15 ~ 15:30 WIB	ISO
15:30 ~ 16:30 WIB	Kuis atau Praktikum oleh Peserta

**Pertemuan 4**

Waktu	Nama Kegiatan
09:00 ~ 10:15 WIB	Materi I: Konversi Waktu
10:15 ~ 10:30 WIB	Coffee Break I
10:30 ~ 11:30 WIB	Materi II: Menghitung Waktu
11:30 ~ 13:00 WIB	ISOMA
13:00 ~ 14:00 WIB	Materi III : Subkueri (Subqueries)
14:00 ~ 15:15 WIB	Materi IV : Subkueri (Subqueries)
15:15 ~ 15:30 WIB	ISO
15:30 ~ 16:30 WIB	Kuis atau Praktikum oleh Peserta

## Pertemuan 5

Waktu	Nama Kegiatan
09:00 ~ 10:15 WIB	Materi I: Pemahaman 5W1H
10:15 ~ 10:30 WIB	Coffee Break I
10:30 ~ 11:30 WIB	Studi Kasus: 5W1H
11:30 ~ 13:00 WIB	ISOMA
13:00 ~ 14:00 WIB	Studi Kasus: 5W1H + Visualisasi
14:00 ~ 15:15 WIB	Studi Kasus: 5W1H + Visualisasi
15:15 ~ 15:30 WIB	ISO
15:30 ~ 16:30 WIB	Kuis atau Praktikum oleh Peserta

## Pertemuan 6

Waktu	Nama Kegiatan
09:00-10:00 WIB	Persentasi Kelompok A
10:00 -10:30 WIB	Coffee Break I
10:30-11:30 WIB	Persentasi Kelompok B
11:30 -13:00 WIB	ISOMA
13:00-14:00 WIB	Persentasi Kelompok C
15:30-15:30 WIB	Pemabahasan Persentasi
15:30-Selesai	Penutupan Acara Pelatihan 2

## Pemateri & Mediator

Berikut ini adalah nama dan biografi singkat para penulis:

- **Bakti Siregar, M.Sc** Bakti Siregar seorang lulusan Magister Sains dari departemen matematika terapan di National Sun Yat-Sen University (Taiwan). Beliau adalah seorang Data Scientist profesional yang memiliki Kompetensi tidak dapat diragukan lagi dalam hal pengolahan data dengan menggunakan SQL dan bahasa pemrograman Python/Rstudio. Lebih dari pada itu, Dia juga akrab dalam memproses Big Data seperti pengenalan distributed computing, NoSQL Database, Hadoop, Spark dan lainnya. Dalam hal visualisasi data, Bakti Siregar mampu menginterpretasikan data dengan sangat baik menggunakan Tableau maupun Power BI. Beberapa project beliau dapat dilihat di link berikut: Rpubs, Github, Website, dan Kaggle.
- **Yosia** adalah salah satu mahasiswa terbaik di jurusan Statistika Universitas Matana. Dia juga memiliki minat dalam pembelajaran sains data dan akuturia khususnya melakukan komputasi dengan menggunakan R dan Python. Yosia bercita-cita suatu saat nanti akan menjadi seseorang yang ahli dibidang aktuaria maupun sains data. Dia memiliki minat teoretis yang luas serta minat dalam komputasi, Yosia pernah terlibat dalam menerbitkan di jurnal Pengabdian Kepada Masyarakat Salah satunya menjadi Assisten (**Workshop Data Analytics With Python & SQL**) yang juga aktif dalam berbagai aktifitas organisasi kampus.
- **Clara Della** adalah mahasiswi jurusan Statistik di Universitas Matana. Dia memiliki kemampuan dalam menggunakan beberapa perangkat lunak yang dibutuhkan dalam proses analisa data, seperti; R, Python, SQL, Tableau dan SPSS. Dia juga sudah pernah terlibat menjadi salah satu Assisten (**Workshop Data Analytics With Python & SQL**). Selain itu, Della juga aktif dalam berbagai aktifitas organisasi kampus.

## Para Peserta

Adun para peserta pelatihan adalah para karyawan JNE dengan jumlah maksimum peserta 15 orang, sebagai berikut:

## Manfaat dan Fasilitas

- Instruktur Berpengalaman di Bidang Analis Data (Akademisi & Praktisi)
- Pelatihan 6 hari (8 Jam Per-pertemuan)
- Modul Pelatihan (E-Book yang dapat diakses hanya untuk Peserta dari JNE)
- Sertifikat Kehadiran (Untuk peserta yang tidak lulus sertifikat kompetensi)



- Sertifikat Kompetensi dari Dsciencelabs (Analisis Data Lanjutan)
- Laporan Kegiatan Pelatihan

## Syarat Peserta

- Laptop yang sudah ter-install Anaconda dan XAMPP atau PostgreSQL
- Sudah terbiasa dengan data, minimal menggunakan Ms. Excel
- Berkemauan tinggi untuk belajar Koding

## Masukan & Saran

Akses untuk Ebook ini hanya tersedia secara gratis untuk para peserta Workshop di JNE. Meskipun banyak upaya telah dilakukan untuk pengembangan, tentu kami sangat mengharapkan masukan dari anda sekalian.

## Ucapan Terima Kasih

Kami juga ingin mengucapkan terima kasih yang sebesar-sebesar pada semua pihak yang terlibat dalam pengembangan buku ini, yakni; mahasiswa-i, dosen, dan Universitas Matana atas dukungan dalam upaya bersama kami untuk menyediakan materi untuk **Pelatihan Analisis Data Dengan Menggunakan Python dan SQL** ini.



# Chapter 1

## Pendahuluan

Sejak tahun 1970, **Structured Query Language (SQL)** telah digunakan oleh para programmer untuk membangun dan mengakses **Sistem Basis Data (SBD)**. Banyak sekali perdebatan mengenai cara penyebutan SQL ini, namun pada kenyataannya, kita dapat melafalkannya sebagai “sequel” ataupun “S.Q.L”. Mempelajari bahasa pemrograman umum seperti Python dan R adalah penting, namun tidak memiliki kemampuan SQL akan membuat kita amat sulit untuk mendapatkan pekerjaan dalam bidang pengolahan data.

Hampir semua nama besar di bidang teknologi menggunakan SQL seperti Uber, Netflix, dan Airbnb. Bahkan dalam perusahaan seperti Facebook, Google dan Amazon, yang telah membuat sendiri **SBD** berkemampuan tinggi, tetap menggunakan SQL untuk melakukan query dan analisis data.

### 1.1 Apa itu SBD?

Secara umum **SBD** dapat didefinisikan sebagai berikut:

#### 1.1.1 Komponen SBD

Adapun beberapa komponen dasar yang diperlukan dalam SBD adalah:

---

#### 1.1.2 Manfaat SBD

Manfaat atau kegunaan penerapan SBD cukup banyak dan cakupannya pun luas dalam mendukung keberadaan lembaga atau organisasi maupun perusahaan, diantaranya:

### 1.1.3 Definisi SQL vs NoSQL

Sebenarnya perbedaan antara SQL dan NoSQL secara mendasar sudah dapat dijelaskan dari akronimnya. *SQL* basis data relasional yang menggunakan ‘relasi’ (yang biasanya disebut tabel) untuk menyimpan data dan mencocokkan data tersebut dengan memakai karakteristik umum di setiap dataset.

Sedangkan, *NoSQL* adalah database yang menggunakan format JSON untuk setiap dokumennya sehingga mudah dibaca dan dimengerti. NoSQL banyak diminati karena memiliki performa yang tinggi dan bersifat non-relasional sehingga dapat memakai berbagai model data.

---

### 1.1.4 Perbedaan SQL vs NoSQL

Sebenarnya banyak perbedaan yang dimiliki di antara dua database tersebut tapi inilah perbedaan yang paling mencolok antara SQL dan NoSQL:

---

### 1.1.5 Top 7 SQL

Tercatat sampai bulan Februari 2020 ada 334 jenis database menurut db-engines.com. Berikut ini saya merangkum daftar 7 database terpopuler yang menggunakan SQL (Relasional):

---

### 1.1.6 Top 8 NoSQL

Kebanyakan basis data NoSQL digunakan dalam dunia aplikasi web waktu nyata (real-time web app). Berikut ini adalah ulasan 8 jenis basis data NoSQL yang paling populer digunakan diseluruh dunia:

---

## 1.2 Mengapa Python & SQL?

**Python** adalah sebuah bahasa pemrograman yang digunakan untuk membuat aplikasi, perintah komputer, dan melakukan analisis data. **SQL** adalah sintaks (bahasa) yang digunakan untuk mengatur dan mengambil informasi dari sistem basis data relasional dan yang juga dapat digunakan untuk membuat basis data sendiri. Jadi, kedua perangkat lunak ini mempunyai peran masing-masing. SQL ditujukan untuk mengakses dan mengambil data dari database. Sedangkan Python digunakan untuk menganalisis dan memanipulasi data dengan menggunakan regresi, time series maupun perhitungan lainnya.

Jika kita ingin bekerja di bidang pengolahan data, kita harus fokus kepada kemampuan yang diinginkan oleh para perusahaan pemberi kerja tersebut. Untuk posisi analis data, SQL termasuk kemampuan yang diutamakan, sekitar 57.4% dari keseluruhan lowongan pekerjaan. Berdasarkan survei yang dilakukan Dataquest dan StackOverflow pada gambar dibawah ini, dapat dilihat bahwa SQL bahkan lebih populer di kalangan data scientist dan data engineer dibandingkan Python dan R.

## 1.3 MySQL vs PostgreSQL

MySQL adalah sistem manajemen basis data relasional yang memungkinkan Anda untuk menyimpan data sebagai tabel dengan baris dan kolom. Sistem ini populer sehingga digunakan di banyak aplikasi web, situs web dinamis, dan sistem tertanam. PostgreSQL adalah sistem manajemen basis data relasional-objek yang menawarkan lebih banyak fitur daripada MySQL. Sistem ini memberi Anda lebih banyak fleksibilitas dalam tipe data, skalabilitas, konkurensi, dan integritas data.

MySQL dan PostgreSQL, Keduanya menyimpan data di dalam tabel yang terkait satu sama lain melalui nilai kolom umum. Namun keduanya sering dibandingkan karena terdapat beberapa perbedaan. Ingin mengenal lebih dalam? Simak penjelasan di bawah.

### 1.3.1 Kelebihan

MySQL	PostgreSQL
Integrasi bahasa pemrograman sangat luas;	Support framework website modern seperti Node.js dan Django; Support framework website modern seperti Node.js dan Django;

MySQL	PostgreSQL
Aplikasi ringan, tidak membutuhkan spesifikasi hardware yang tinggi;	Dirilis dengan lisensi PostgreSQL sendiri;
Struktur tabel dengan fleksibilitas tinggi;	Bersifat open source dan gratis;
Dibekali banyak administrative tools;	Skala besar, mampu memuat hingga ribuan transaksi data;
Bersifat open source dan gratis (versi basic);	Memiliki banyak fitur yang mumpuni;
Meski open source, MySQL menjamin keamanan tingkat tinggi;	Memiliki banyak fitur yang mumpuni;
Mendukung berbagai variasi Data Type;	Performa sangat baik meski menuntut query yang lebih kompleks;
Dapat digunakan banyak pengguna karena mendukung multi user.	Kecepatan analisis (read-write) sangat cepat; Keamanan yang lebih ketat.

### 1.3.2 Kekurangan

MySQL	PostgreSQL
Sistem manajemen database kurang cocok untuk aplikasi mobile dan game;	PostgreSQL tidak mendukung semua stack development;
Technical support MySQL dinilai kurang baik;	Meski memiliki integrasi dan skalabilitas tinggi, kecepatan PostgreSQL kalah unggul dibandingkan RDBMS lain;
Sulit diaplikasikan untuk manajemen database berskala besar.	Sistem kompatibilitas PostgreSQL menuntut pengguna untuk bekerja lebih keras dalam perbaikan dan perawatan.

## 1.4 Instalasi Python (Anaconda)

Berikut ini adalah proses langkah demi langkah tentang Cara Menginstal Anaconda di Windows:

### 1.4.1 Open site

Buka Situs Web Anaconda Lalu pilih Download pada logo Windows

### 1.4.2 Buka file exe yang telah didownload

Ketika layar di bawah ini muncul, klik Next.

### 1.4.3 Baca perjanjian lisensi dan klik I Agree.

### 1.4.4 Click on Next

### 1.4.5 Catat lokasi penginstalan Anda, lalu klik Next.

### 1.4.6 Customiza Anaconda

Ini adalah bagian penting dari proses instalasi. approach yang disarankan adalah tidak mencentang kotak untuk menambahkan Anaconda ke path Anda. Ini berarti Anda harus menggunakan Anaconda Navigator atau Anaconda Command Prompt (terletak di Start Menu di bawah “Anaconda”) ketika Anda ingin menggunakan Anaconda (Anda selalu dapat menambahkan Anaconda ke PATH Anda nanti jika Anda tidak mencentang kotak). Jika Anda ingin menggunakan Anaconda pada command prompt Anda (atau git bash, cmd, powershell, dll), gunakan alternative approach dan centang kotaknya.

### 1.4.7 Setup Complete

###Install Microsoft VSCode

Anda dapat menginstal Microsoft VSCode jika Anda mau, tetapi ini opsional.

### 1.4.8 Selesai

Click on Finish.

### 1.4.9 Video Instalasi

- Download Anaconda

## 1.5 Instalasi MySQL (XAMPP)

### 1.5.1 Download Aplikasi XAMPP

Silakan klik [disini](#) untuk mengunduh aplikasi XAMPP, pilih salah satu saja sesuai Operating System pada Komputer anda.

### 1.5.2 Install Aplikasi

Temukan file XAMPP.exe yang telah anda download, secara default biasanya disimpan di;

Selanjutnya, akan muncul Warning di klik **OK**

selanjutnya klik next

Klik next lagi, karena sudah dipilih secara default oleh XAMPP

### 1.5.3 Pilih Folder

Secara default akan membuat folder baru **C:\xampp**, lalu pilih next.

**note:** jika anda sudah pernah mendownload aplikasi xampp, perlu di hapus terlebih dahulu file xampp yang lama di file **C:\xampp**

### 1.5.4 Jalankan proses Instalasi

Tunggu proses instalasi selesai **Biasanya 5-10 menit, tergantung kecepatan komputer anda.**

### 1.5.5 XAMPP sudah terinstall

Setelah aplikasi terinstall, sudah bisa digunakan.

### 1.5.6 Video Instalasi XAMPP

## 1.6 Instalasi PostgreSQL

Berikut ini adalah proses langkah demi langkah tentang Cara Menginstal PostgreSQL di Windows:

### 1.6.1 Buka Browser

Klik <https://www.postgresql.org/download> and pilih Windows

### 1.6.2 Cek Option

Klik Download the installer Interactive Installer by EnterpriseDB



### 1.6.3 Pilih PostgreSQL version

Anda akan diminta untuk memilih versi PostgreSQL dan sistem operasi yang diinginkan. Pilih versi PostgreSQL terbaru dan OS sesuai dengan environment Anda, **klik tombol download**.

### 1.6.4 Open exe file

Setelah Anda mengunduh PostgreSQL, buka exe yang telah diunduh dan Klik berikutnya pada layar install welcome screen.

### 1.6.5 Pilih folder

Ubah direktori Instalasi jika diperlukan, jika tidak, biarkan default, **klik Next**.

### 1.6.6 Select components

Anda dapat memilih komponen yang ingin Anda instal di sistem Anda. Anda dapat menghapus centang pada Stack Builder (*disarankan ikuti secara default*), **klik Next**.

### 1.6.7 Check data location

Anda dapat mengubah lokasi data, **Klik Next**.

### 1.6.8 Masukkan Password

Masukkan kata sandi superuser. Catat kata sandi tersebut, **Klik Next**.

### 1.6.9 Cek opsi port

Biarkan nomor port menjadi default, **Klik Next**.

### 1.6.10 Cek Summary

Periksa pra-penginstalan summary, **Klik Next**

### 1.6.11 Ready to Install

Klik tombol Next

### 1.6.12 Check stack builder prompt

Setelah instalasi selesai, Anda akan melihat prompt Stack Builder. Hapus centang pada opsi tersebut. Kita akan menggunakan Stack Builder dalam tutorial selanjutnya, **Klik Finish**.

### 1.6.13 Launch PostgreSQL

Untuk launch PostgreSQL, buka Start Menu dan cari pgAdmin 4

### 1.6.14 Check pgAdmin

Anda akan melihat beranda pgAdmin

### 1.6.15 Cari PostgreSQL 15

Klik pada Servers > PostgreSQL 15 di sub sebelah kiri

### 1.6.16 Enter password

Masukkan kata sandi superuser yang ditetapkan selama instalasi, **Klik OK**

### 1.6.17 Cek Dashboard

Anda akan melihat Dashboard

### 1.6.18 Video Instalasi PostgreSQL

- Download PostgreSQL

## 1.7 Praktikal (Hands On)

Untuk memulai bagian ini, pastikan bahwa anda sudah meng-install Anaconda dan XAMPP di komputer yang akan digunakan. Kemudian, bukalah aplikasi Jupyterlab yang ada pada layar Anaconda Navigator, seperti diperlihatkan sebagai berikut:

Selanjutnya, akan muncul halaman baru pada browser seperti gambar berikut;

Selanjutnya, klik 'File > Open from URL'. Muncul tampilan sebagai berikut:

Terakhir, Copy-Paste dibawah ini pada kolom `Open URL`. Selanjutnya, ikuti tutorialnya.

- **Akses Databases MySQL dengan Python**
- **Akses Databases PostgreSQL dengan Python**

## 1.8 Latihan

Materi latihan akan diberikan setelah tutorial selesai.



## Chapter 2

# Pemilihan Data



## Chapter 3

# Pengelompokan Data





## Chapter 4

# Perhitungan Waktu & Subkueri



## Chapter 5

# Pemahaman 5W1H & Visualisasi



## Chapter 6

# Persentasi Tugas Akhir



## Chapter 7

# Penutup





