

MODUL PEMBELAJARAN

PENGENALAN SAINS DATA
MELALUI PELATIHAN ANALISIS DAN VISUALISASI DATA
UNTUK SISWA SMK ANTARTIKA 1 SIDOARJO

KATA PENGANTAR

Puji syukur kehadiran Tuhan Yang Maha Esa atas segala rahmat dan karunia-Nya, sehingga modul pembelajaran berjudul “Pengenalan Sains Data melalui Pelatihan Analisis dan Visualisasi Data untuk Siswa SMK Antartika 1 Sidoarjo” ini dapat diselesaikan dengan baik. Modul ini disusun sebagai bahan ajar dalam kegiatan Sains Data untuk Masyarakat, yang bertujuan untuk memperkenalkan konsep dasar sains data, analisis data, serta visualisasi data kepada siswa SMK. Melalui kegiatan ini, diharapkan peserta mampu memahami bagaimana data dikumpulkan, diolah, dan divisualisasikan menjadi informasi yang bermakna, sekaligus mengembangkan keterampilan berpikir logis, analitis, dan kritis yang dibutuhkan di era digital saat ini.

Materi dalam modul ini mencakup tiga bagian utama, yaitu:

1. Pengenalan Data dan Sains Data,
2. Analisis dan Visualisasi Data menggunakan Microsoft Excel, dan
3. Analisis dan Visualisasi Data menggunakan Python.

Penyusunan materi dilakukan dengan bahasa yang sederhana dan mudah dipahami oleh siswa SMK, disertai contoh praktikum serta mini project yang aplikatif agar peserta tidak hanya memahami teori, tetapi juga dapat menerapkannya secara langsung.

Ucapan terima kasih kami sampaikan kepada Ibu Amalia Nur Alifah, S.Si., M.Si. selaku dosen pembimbing atas arahan, bimbingan, dan dukungan yang telah diberikan selama proses penyusunan modul ini.

Kami juga mengucapkan terima kasih kepada pihak SMK Antartika 1 Sidoarjo atas kerja sama dan partisipasinya dalam kegiatan pelatihan ini.

Kami menyadari bahwa modul ini masih memiliki kekurangan dan memerlukan penyempurnaan. Oleh karena itu, kami sangat terbuka terhadap kritik dan saran yang membangun demi perbaikan modul ini di masa mendatang. Semoga modul ini dapat memberikan manfaat bagi siswa dalam meningkatkan kompetensi digital, serta menjadi langkah awal untuk mengenal dunia sains data yang terus berkembang.

Surabaya, 19 Oktober 2025

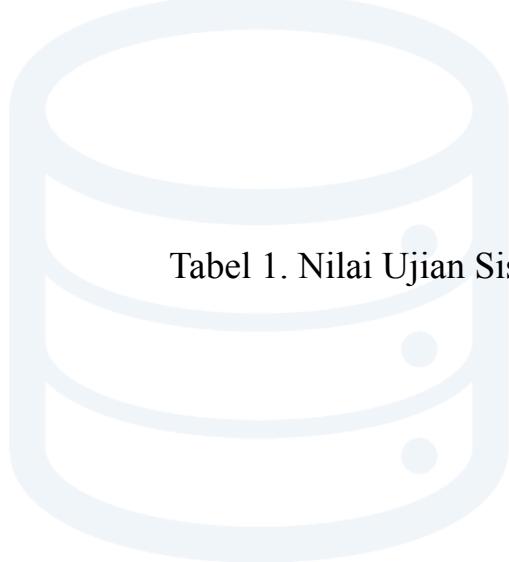
Tim Penyusun

DAFTAR ISI

Materi 1: Pengenalan Data dan Sains Data.....	1
1. Apa itu Data dan Sains Data?.....	1
2. Tahapan-tahapan dalam Sains Data:.....	3
3. Mengapa Visualisasi Data Penting?.....	4
Materi 2: Visualisasi Data dengan Excel.....	7
1. Mengenal Excel sebagai Alat Analisis.....	7
2. Analisis Data Sederhana.....	7
3. Membuat Visualisasi sederhana di Excel.....	10
4. Mini Project Excel (Study Case).....	14
Materi 3: Pengenalan Visualisasi Data dengan Python.....	15
1. Pengenalan Visualisasi Data dengan Python.....	15
2. Analisis Data dengan Pandas.....	15
3. Analisis Data dengan Numpy.....	19
4. Membuat Grafik dengan Matplotlib dan Seaborn.....	20
a. Membuat Grafik dengan Matplotlib.....	20
b. Membuat Grafik dengan Seaborn.....	22
5. Mini Project Python (Study Case).....	26

DAFTAR GAMBAR

Gambar 1. Ilustrasi Apa itu Data?.....	1
Gambar 2. Ilustrasi kenapa sains data penting?.....	2
Gambar 3. Tabel Angka.....	5
Gambar 4. Contoh Grafik Batang.....	5
Gambar 5. Mencari nilai Rata-rata.....	8
Gambar 6. Mencari nilai tertinggi.....	8
Gambar 7. Mencari nilai terendah.....	9
Gambar 8. Mencari nilai paling sering muncul.....	9
Gambar 9. Mencari nilai tengah.....	9
Gambar 10. Penerapan rumus IF ELSE.....	10
Gambar 11. Bar Chart perbandingan jumlah produk.....	11
Gambar 12. Pie Chart presentase ekstrakulikuler.....	11
Gambar 13. Line Chart Jumlah pengunjung kebun binatang.....	12
Gambar 13. Scatter Plot hubungan belajar dengan nilai ujian siswa.....	13
Gambar 14. Output tampilan data df.....	16
Gambar 15. Output tampilan df.head().....	17
Gambar 16. Hasil statistika deskriptif data.....	18
Gambar 17. Pengecekan nilai kosong.....	18
Gambar 18. Output tampilan data df setelah kolom baru.....	19
Gambar 19. Hasil visualisasi nilai ujian siswa.....	22
Gambar 20. Hasil visualisasi nilai ujian siswa dengan seaborn.....	24
Gambar 21. Persentase kelulusan visualisasi pie chart.....	25



DAFTAR TABEL

Tabel 1. Nilai Ujian Siswa.....	7
---------------------------------	---

Materi 1: Pengenalan Data dan Sains Data

1. Apa itu Data dan Sains Data?

Apa itu Data?



Gambar 1. Ilustrasi Apa itu Data?

Sebelum kita bahas “Sains Data”, kita harus tahu dulu apa itu **data**.

Data merupakan kumpulan fakta, angka, atau informasi yang bisa diolah untuk mendapatkan sebuah insight atau wawasan.

Menurut **Kalsum et al. (2025)**, data bisa berupa angka, teks, gambar, atau simbol yang menggambarkan suatu keadaan, dan baru akan bermakna setelah diolah menjadi informasi.

Singkatnya, **Data = bahan mentah**, sedangkan **informasi = hasil olahan data**.

Contoh Data:

- Nilai ujian tiap siswa di kelas.
- Jumlah pengunjung kantin per hari.
- Jumlah like dan comment di akun media sosial

Ketika data itu dikumpulkan, dibersihkan, dan dianalisis, barulah muncul insight seperti:

“Oh ternyata hari Selasa kantin paling ramai dikunjungi!”

“Atau rata-rata nilai Matematika Andi lebih tinggi dari nilai Bahasa Inggris.”

Apa Itu Sains Data?

Setelah mengenal apa itu data, sekarang kita akan membahas singkat apa itu **Sains Data**.

Sains Data atau *Data Science* itu bidang ilmu yang menggabungkan *statistika, matematika, dan ilmu komputer* buat mengolah data jadi informasi yang bisa bantu mengambil keputusan.

Menurut **Han, Kamber, & Pei (2022)** dalam bukunya *Data Mining: Concepts and Techniques*,

“Sains Data adalah ilmu interdisipliner yang menggabungkan statistik, ilmu komputer, untuk mengekstrak pola-pola bermakna dari data.”

Kenapa Sains Data Itu Penting?



Gambar 2. Ilustrasi kenapa sains data penting?

Coba teman-teeman perhatikan,

Pernahkah saat sedang *scrolling* di TikTok, tiba-tiba muncul banyak video tentang makanan, lalu semakin lama kamu menontonnya, semakin sering pula konten serupa muncul di beranda (*For You Page*) kamu?

Hal tersebut **bukan kebetulan**. Itu merupakan hasil kerja **algoritma dan penerapan sains data**.

Menurut Bishop (2023) dalam bukunya *Pattern Recognition and Machine Learning*,

“Sistem rekomendasi modern sangat bergantung pada teknik ilmu data untuk mempelajari preferensi pengguna dari data perilaku.”

Artinya, **platform seperti TikTok, YouTube, dan Spotify** menggunakan data perilaku pengguna untuk mempelajari kebiasaan dan kesukaan mereka.

Contohnya:

- Berapa lama kamu menonton suatu video,
- Konten apa yang kamu sukai (*like*),
- Akun mana yang sering kamu kunjungi,
- Serta topik apa yang sering kamu cari

Semua informasi tersebut direkam dan dianalisis menggunakan algoritma pembelajaran mesin (*machine learning*) agar sistem dapat menampilkan konten yang paling sesuai dengan minatmu.

2. Tahapan-tahapan dalam Sains Data:

1. Business Understanding

Tahap pertama adalah memahami tujuan. Misalnya, kita ingin mengetahui produk apa yang paling banyak terjual atau kapan waktu penjualan tertinggi terjadi.

2. Data Understanding

Setelah mengetahui tujuan, langkah berikutnya adalah mengumpulkan dan memahami data yang akan digunakan. Pada tahap ini, kita perlu memeriksa struktur data, format, serta sumbernya, untuk memastikan bahwa data tersebut relevan dengan tujuan analisis.

3. Data Preparation

Setelah data terkumpul, langkah selanjutnya adalah membersihkan dan menyiapkan data agar siap dianalisis. Proses ini mencakup kegiatan seperti menghapus data ganda, mengisi nilai yang hilang, dan menyesuaikan format data agar lebih rapi dan konsisten.

4. Exploratory Data Analysis

Tahap ini disebut juga analisis eksploratif, yaitu proses untuk menjelajahi dan memahami isi data secara mendalam. Di sini, kita mulai melakukan analisis awal dan membuat visualisasi sederhana untuk mencari pola, tren, atau hubungan menarik yang muncul dari data.

5. Data Modeling

Setelah memperoleh pemahaman dari data, langkah selanjutnya adalah membangun model statistik atau model machine learning. Model ini digunakan untuk melakukan prediksi atau menghasilkan wawasan baru dari data yang telah dianalisis.

6. Data Evaluation

Tahap ini bertujuan untuk menilai apakah model yang telah dibuat sudah sesuai dengan tujuan awal. Jika hasilnya belum optimal, maka model dapat diperbaiki atau disesuaikan kembali.

7. Model Deployment

Tahap terakhir adalah menerapkan model ke dunia nyata, misalnya digunakan dalam sistem rekomendasi produk, prediksi cuaca, atau analisis perilaku pengguna. Hasil dari penerapan model ini dapat membantu pengambilan keputusan yang lebih tepat dan efisien.

Setelah memahami tahapan-tahapan dalam sains data, ada satu bagian yang sangat menarik dan sering digunakan dalam praktik sehari-hari, yaitu *Exploratory Data Analysis* (EDA).

Melalui EDA, kita dapat menjelajahi data, menemukan pola, tren, serta hubungan antarvariabel.

Salah satu hal terpenting dalam EDA adalah visualisasi data. Dengan visualisasi, data yang awalnya tampak rumit dapat disajikan secara lebih jelas, mudah dibaca, dan cepat dipahami.

3. Mengapa Visualisasi Data Penting?

Bayangkan jika kamu diberikan sebuah tabel panjang yang berisi **ratusan angka**. Kira-kira, apakah kamu bisa langsung memahami makna dari data tersebut?

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q
	name	album	release_date	rack_number	id	uri	acousticness	danceability	energy	instrumentalness	liveness	loudness	speechiness	tempo	valence	popularity	duration_ms
1	right Post MITS DEPARTME	2024-04-19	1	rc0EuGzHkAavXqIdDwoeGvzH+	0.502	0.504	0.386	1.53E-05	0.0961	-10.976	0.0308	192.004	0.281	0.02	228965		
2	ght feat. Post MITS DEPARTME	2024-04-19	2	aGZub4lghChg-4PdLsGZub4k	0.0483	0.604	0.428	0	0.126	-8.441	0.0255	110.259	0.292	79	292048		
3	Tured Poets Dep:TS DEPARTME	2024-04-19	3	MwRy24dmnTRUuGYwRy24d	0.137	0.596	0.563	0	0.302	-7.362	0.0269	97.073	0.481	0.00	203801		
4	ly Breaks His Fa:TS DEPARTME	2024-04-19	4	BdfcgqdgLXCJ1kbEbBdfgqdgQ	0.56	0.541	0.366	1.00E-06	0.0946	-10.412	0.0748	159.707	0.168	0.02	261228		
5	Down Bad :TS DEPARTME	2024-04-19	5	jFSuJ27V8362M7wAkDFsJ27Vf	0.73	0.423	0.533	0.00264	0.0816	-11.388	0.322	160.218	0.248	0.00	262974		
6	So Long, London:TS DEPARTME	2024-04-19	6	jEHjQuKqPnCnrgf-4QMcgEHQuKqI	0.384	0.521	0.72	0	0.135	-7.684	0.104	79.943	0.438	0.01	340428		
7	X Daddy I Love HTS DEPARTME	2024-04-19	7	WOfloHoHhras97WcdWOfluoH5	0.624	0.33	0.483	0	0.111	-9.394	0.0399	87.115	0.34	0.00	210789		
8	in Out The Slam:TS DEPARTME	2024-04-19	8	lddhNryAdSAExOm0dh4NyRy	0.178	0.533	0.573	0	0.309	-7.123	0.138	150.088	0.398	79	215463		
9	at Florence + TTS DEPARTME	2024-04-19	9	pEbhZp0mHtA8799KpEbhZp0m	0.607	0.626	0.428	0	0.0921	-8.374	0.0261	94.95	0.487	0.02	254365		
10	Guilty as Sin:TS DEPARTME	2024-04-19	10	xVNhJnciuuaP-2d8UvVNhuic8	0.315	0.604	0.338	0	0.106	-10.555	0.048	119.88	0.238	0.01	334084		
11	Afraid of Little OITS DEPARTME	2024-04-19	11	fB9qMK3W1M41chnRTB9qMK3V	0.768	0.661	0.257	1.72E-06	0.176	-13.961	0.045	106.894	0.318	0.00	156297		
12	ix Him (No Really):TS DEPARTME	2024-04-19	12	lrAvbKNmwwEd:3YNKNiAvkNnn	0.858	0.423	0.272	0	0.108	-12.56	0.0489	78.539	0.195	0.00	277156		
13	lomf :TS DEPARTME	2024-04-19	13	IGGEZOKQJ9qX-2IPQIGGEZOK	0.123	0.701	0.751	0	0.15	-8.39	0.0502	129.994	0.22	0.02	218004		
14	o It With a Broke:TS DEPARTME	2024-04-19	14	krCR60wKwKx1tw1wCR60w2	0.52	0.426	0.41	0	0.338	-9.083	0.0581	142.377	0.263	0.01	245541		
15	lest Man Who E:TS DEPARTME	2024-04-19	15	OrtRQVHvONR81tuNuOrtRQVHv	0.422	0.64	0.421	0	0.103	-8.802	0.0273	100.056	0.353	0.00	196900		
16	The Alchemy :TS DEPARTME	2024-04-19	16	?IEVj9jW50uLnk-4d9jPIEVj9jW!	0.636	0.686	0.318	0.00144	0.106	-11.046	0.0263	104.088	0.214	0.00	216662		
17	Clara Bow :TS DEPARTME	2024-04-19	17	R0ad0dMSHyuy62E2nR0ad0dMS	0.707	0.527	0.27	0	0.102	-10.855	0.038	154.042	0.23	0.00	238857		
18	The Black Dog :TS DEPARTME	2024-04-19	18	PDQnqEgmezX1kvpPDQnqEg	0.608	0.784	0.391	0	0.252	-9.471	0.0633	110.007	0.15	0.01	222072		
19	agonnageyoubac:TS DEPARTME	2024-04-19	19	yONoQyblQxMxeF6lyONoQy0l	0.655	0.352	0.479	0.0935	-8.942	0.0583	172.013	0.292	0.00	183878			
20	The Albatross :TS DEPARTME	2024-04-19	20	iOewz2lmZTMA1mEsOewz2lm	0.78	0.516	0.451	0	0.105	-8.552	0.0446	70.266	0.218	0.00	213281		
21	How Did it End?:TS DEPARTME	2024-04-19	21	OsvoZD7RGnJK5BdnosVZD7f	0.785	0.492	0.379	0.00093	0.108	-8.859	0.0257	83.993	0.273	0.00	238706		
22	Sam or Sophia o:TS DEPARTME	2024-04-19	22	0jOpf4kuOnmc:7Ms0jOpf4k	0.0274	0.366	0.866	3.07E-06	0.106	-4.514	0.0466	176.07	0.293	0.02	228800		

Gambar 3. Tabel Angka

Sekarang, coba bayangkan data yang sama ditampilkan dalam bentuk grafik batang.

Hanya dengan satu kali pandang, kamu bisa langsung memahami informasi yang disajikan misalnya, pada grafik batang berikut yang menyajikan 10 lagu Taylor Swift dengan tingkat popularitas tertinggi.

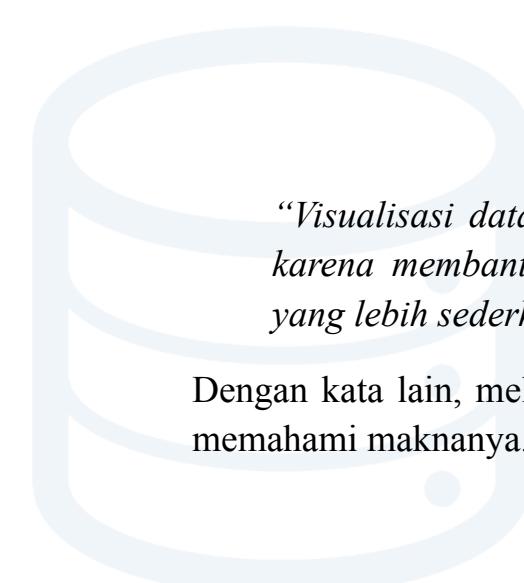


Gambar 4. Contoh Grafik Batang

Perbedaan ini menunjukkan betapa pentingnya visualisasi data.

Visualisasi membantu kita mengubah data mentah menjadi informasi yang mudah dibaca dan dipahami oleh mata.

Menurut Kalsum et al. (2025),



“Visualisasi data merupakan langkah penting dalam proses analisis karena membantu menyampaikan informasi kompleks dengan cara yang lebih sederhana dan intuitif.”

Dengan kata lain, melalui visualisasi, kita tidak hanya *melihat* data, tetapi juga memahami maknanya.

Materi 2: Visualisasi Data dengan Excel

1. Mengenal Excel sebagai Alat Analisis

Menurut **Albright & Winston (2022)** dalam buku *Data Analysis and Decision Making*,

“Microsoft Excel is one of the most accessible tools for statistical analysis and data visualization, especially for beginners.”

Excel bisa dipakai untuk:

- Mengelola dan menyimpan data
- Menghitung otomatis pakai rumus.
- Membuat grafik (Bar chart, Line Chart, Pie Chart, dan sebagainya)

2. Analisis Data Sederhana

Misalnya kamu punya data nilai ujian siswa seperti ini:

Nama	Nilai
Andi	73
Budi	80
Cantika	89
Dani	77
Eka	90
Fira	75
Galih	92
Hery	83
Indah	83
Jay	78

Tabel 1. Nilai Ujian Siswa

Dari data tersebut, kamu dapat melakukan analisis sederhana menggunakan beberapa fungsi di Excel, seperti:

1. Mencari Rata-Rata Nilai Ujian menggunakan AVERAGE()

	A	B	C	D	E	F	G
1	Nama	Nilai Ujian					
2	Andi	73					
3	Budi	80					
4	Cantika	89					
5	Dani	77					
6	Eka	90					
7	Fira	75					
8	Galih	92					
9	Hery	83					
10	Indah	83					
11	Jay	78					

Gambar 5. Mencari nilai Rata-rata

2. Mencari Nilai Tertinggi menggunakan MAX()

	A	B	C	D	E
1	Nama	Nilai Ujian			
2	Andi	73			
3	Budi	80			
4	Cantika	89	Rata-Rata Nilai Ujian	82	
5	Dani	77	Nilai Tertinggi	=MAX(B2:B11)	
6	Eka	90			
7	Fira	75			
8	Galih	92			
9	Hery	83			
10	Indah	83			
11	Jay	78			

Gambar 6. Mencari nilai tertinggi

3. Mencari Nilai Terendah menggunakan MIN()

	A	B	C	D	E
1	Nama	Nilai Ujian			
2	Andi	73			
3	Budi	80	Rata-Rata Nilai Ujian	82	
4	Cantika	89	Nilai Tertinggi	92	
5	Dani	77	Nilai Terendah	73	=MIN(B2:B11)
6	Eka	90			
7	Fira	75			
8	Galih	92			
9	Hery	83			
10	Indah	83			
11	Jay	78			

Gambar 7. Mencari nilai terendah.

4. Mencari Nilai yang paling sering muncul menggunakan MODE()

	A	B	C	D	E	F
1	Nama	Nilai Ujian				
2	Andi	73				
3	Budi	80	Rata-Rata Nilai Ujian	82		
4	Cantika	89	Nilai Tertinggi	92		
5	Dani	77	Nilai Terendah	73		
6	Eka	90	Nilai yang paling sering muncul (Modus)	73	=MODE(B2:B11)	
7	Fira	75				
8	Galih	92				
9	Hery	83				
10	Indah	83				
11	Jay	78				

Gambar 8. Mencari nilai paling sering muncul

5. Mencari Nilai tengah data menggunakan MEDIAN()

	A	B	C	D	E
1	Nama	Nilai Ujian			
2	Andi	73			
3	Budi	80	Rata-Rata Nilai Ujian	82	
4	Cantika	89	Nilai Tertinggi	92	
5	Dani	77	Nilai Terendah	73	
6	Eka	90	Nilai yang paling sering muncul (Modus)	83	
7	Fira	75	Nilai Tengah	83	=MEDIAN(B2:B11)
8	Galih	92			
9	Hery	83			
10	Indah	83			
11	Jay	78			

Gambar 9. Mencari nilai tengah

6. Membuat kolom baru bernama keterangan untuk melihat status siswa yang lulus dan tidak lulus menggunakan kondisi IF ELSE

C2		=IF(B2>75;"Lulus";"Tidak Lulus")				
A	B	C	D	E	F	G
1 Nama	Nilai Ujian	Keterangan				
2 Andi	73	Tidak Lulus				
3 Budi	80	Lulus				
4 Cantika	89	Lulus				
5 Dani	77	Lulus				
6 Eka	90	Lulus				
7 Fira	75	Tidak Lulus				
8 Galih	92	Lulus				
9 Hery	83	Lulus				
10 Indah	83	Lulus				
11 Jay	78	Lulus				

Gambar 10. Penerapan rumus IF ELSE

3. Membuat Visualisasi sederhana di Excel

beberapa tipe-tipe chart pada visualisasi, antara lain:

1. Bar chart

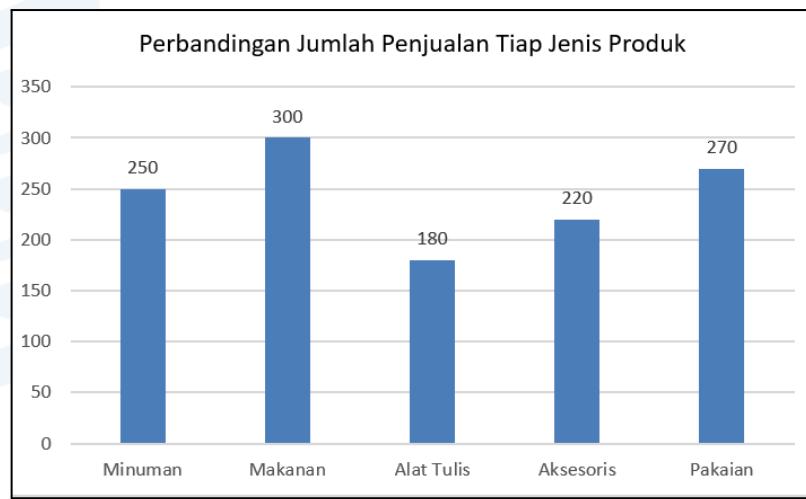
digunakan untuk menyajikan data berskala nominal atau ordinal dalam bentuk batang. Setiap batang mewakili frekuensi, jumlah, atau nilai dari suatu kategori tertentu. Grafik ini cocok digunakan untuk membandingkan antar kategori, seperti jumlah siswa per kelas, penjualan tiap produk, atau rata-rata nilai tiap mata pelajaran.

Biasanya:

- Sumbu X menunjukkan kategori (misal: nama produk, nama siswa).
- Sumbu Y menunjukkan nilai atau jumlah.

Contoh:

Perbandingan Jumlah Penjualan Tiap jenis produk.



Gambar 11. Bar Chart perbandingan jumlah produk

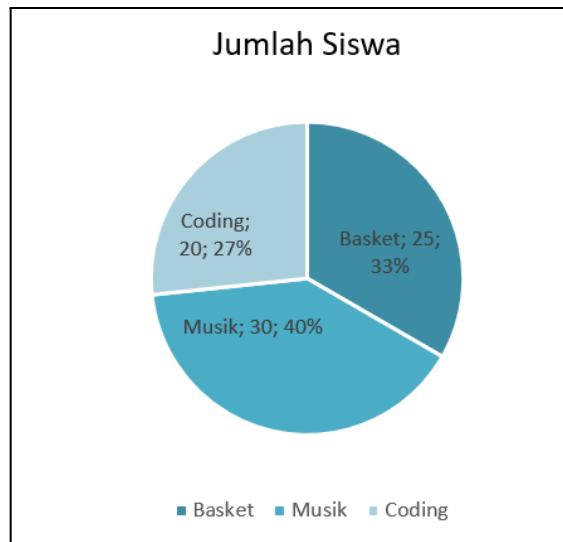
2. Pie chart

Pie Chart digunakan untuk menampilkan proporsi antar bagian dari suatu keseluruhan. Setiap potongan (slice) mewakili persentase dari total data, dan jika dijumlahkan semuanya akan menghasilkan 100%. Grafik ini cocok digunakan untuk memperlihatkan pembagian komposisi, misalnya persentase pengeluaran, jumlah siswa per jurusan, atau hasil survei pilihan.

Namun, hindari menampilkan terlalu banyak kategori di pie chart (**maksimal 5–7 bagian**) karena potongan kecil akan sulit dibaca dan dibandingkan.

Contoh:

Persentase siswa yang memilih ekstrakurikuler berbeda (Basket, Musik, dan Coding)



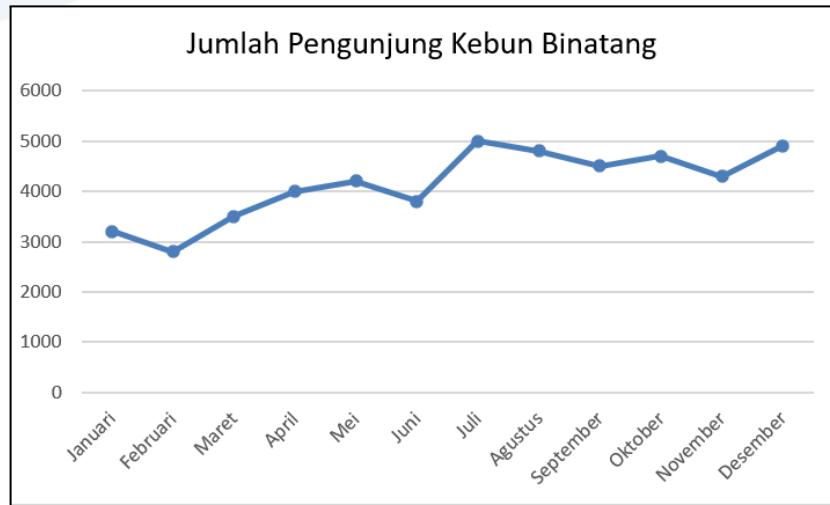
Gambar 12. Pie Chart presentase ekstrakurikuler

3. Line Chart (Grafik Garis)

Line chart digunakan untuk **menampilkan perubahan nilai dari waktu ke waktu (tren)**. Garis yang menghubungkan titik-titik data menggambarkan arah kenaikan atau penurunan dari suatu nilai. Jenis grafik ini sangat cocok untuk menunjukkan perkembangan data berkala, seperti jumlah pengunjung per bulan, nilai ujian per semester, atau pertumbuhan penjualan tahunan.

Contoh:

Grafik jumlah pengunjung Kebun Binatang Surabaya dari Januari hingga Desember.



Gambar 13. Line Chart Jumlah pengunjung kebun binatang

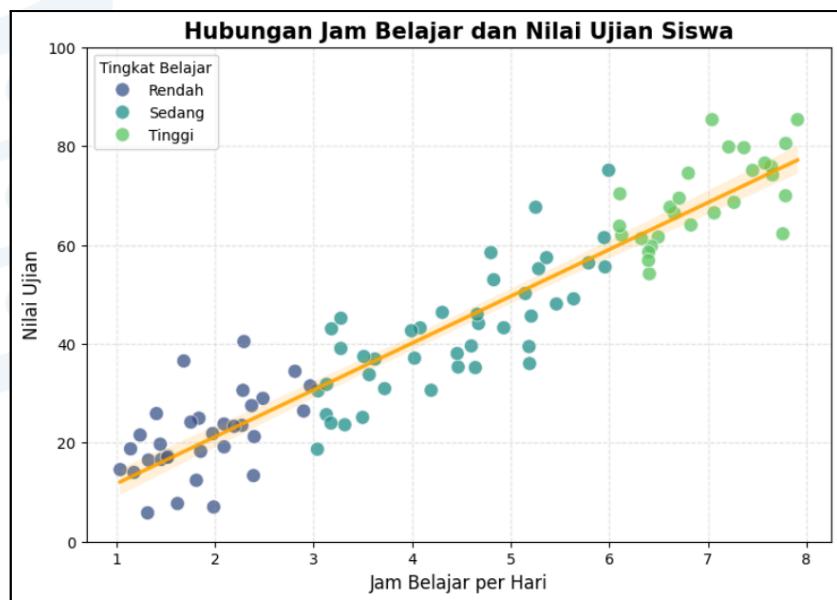
4. Scatter Plot

Scatter plot digunakan untuk **menunjukkan hubungan antara dua variabel numerik**.

Setiap titik pada grafik merepresentasikan satu data observasi dengan dua nilai (sumbu X dan Y). Grafik ini berguna untuk melihat pola korelasi, misalnya apakah dua variabel memiliki hubungan positif, negatif, atau tidak ada hubungan sama sekali.

Contoh:

Hubungan antara jam belajar dengan nilai ujian siswa.



Gambar 13. Scatter Plot hubungan belajar dengan nilai ujian siswa

4. Mini Project Excel (Study Case)

StudyCase

Materi 3: Pengenalan Visualisasi Data dengan Python

1. Pengenalan Visualisasi Data dengan Python

Python merupakan bahasa pemrograman yang populer dalam bidang *Data Science* karena mudah dipelajari, bersifat fleksibel, dan memiliki banyak pustaka (*library*) untuk mengolah serta menampilkan data secara visual.

Python sangat bermanfaat karena:

- Dapat mengolah data dalam jumlah besar dengan cepat.
- Dapat membuat grafik yang lebih fleksibel dan menarik.
- Dapat digunakan berulang kali tanpa perlu mengatur ulang seperti di Excel.
- Dapat diintegrasikan dengan berbagai sumber data (file CSV, database, API, dan lainnya).

2. Analisis Data dengan Pandas

Salah satu pustaka utama dalam Python yang digunakan untuk analisis data adalah Pandas.

Pandas berfungsi untuk mengimpor, mengolah, dan menganalisis data dalam bentuk tabel mirip seperti lembar kerja pada Excel.

Menurut **McKinney (2022)** dalam *Python for Data Analysis*,

“Python’s simplicity and powerful libraries make it the ideal language for data analysis and visualization.”

Data di dalam Pandas disimpan dalam struktur yang disebut **DataFrame**, yaitu tabel dua dimensi yang memiliki baris dan kolom.

Berikut contoh sederhana penggunaan Pandas menggunakan Google Colab:

- Membuat dan Membaca Dataset

```
import pandas as pd

# Membuat data sederhana
data = {'Nama': ['Andi', 'Budi', 'Cantika', 'Dwi', 'Eka', 'Fira', 'Galih', 'Herry', 'Indah', 'Jay'],
        'Nilai': [73, 80, 89, 77, 90, 75, 92, 83, 83, 78]}

# Mengubah data menjadi DataFrame
df = pd.DataFrame(data)

# Menampilkan data
df
```

Output:

	Nama	Nilai	
0	Andi	73	
1	Budi	80	
2	Cantika	89	
3	Dwi	77	
4	Eka	90	
5	Fira	75	
6	Galih	92	
7	Herry	83	
8	Indah	83	
9	Jay	78	

Gambar 14. Output tampilan data df

Penjelasan:

import pandas as pd berfungsi untuk mengaktifkan pustaka Pandas.

pd.DataFrame(data) mengubah data menjadi bentuk tabel.

`df` menampilkan data.

- Mengeksplorasi Data

```
df.head() # Melihat 5 data teratas
```

Output:

	Nama	Nilai
0	Andi	73
1	Budi	80
2	Cantika	89
3	Dwi	77
4	Eka	90

Gambar 15. Output tampilan `df.head()`

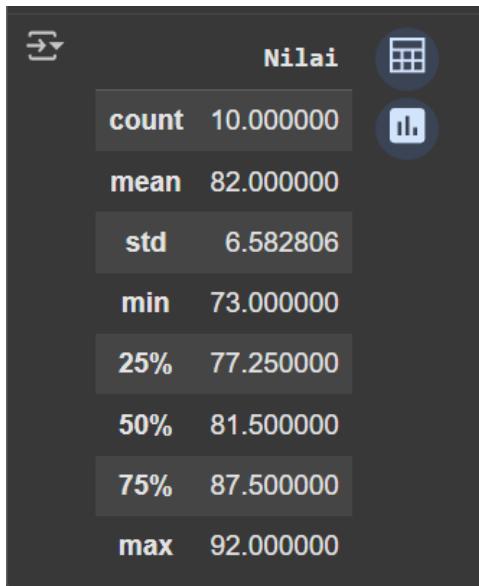
```
df.info() # Informasi tipe data
```

Output:

```
Data columns (total 2 columns):  
 # Column Non-Null Count Dtype  
 ---  
 0 Nama    10 non-null   object  
 1 Nilai   10 non-null   int64  
 dtypes: int64(1), object(1)
```

```
df.describe() # Statistik ringkas
```

Output:

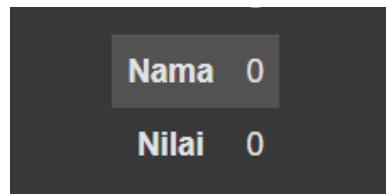


Gambar 16. Hasil statistika deskriptif data

- Pembersihan dan Persiapan Data (*Data Cleaning*)

```
# Cek nilai kosong  
df.isnull().sum()
```

Output:



Gambar 17. Pengecekan nilai kosong

```
# Hapus duplikat  
df.drop_duplicates(inplace=True)
```

- Membuat kolom baru untuk menambah insight:

```
# Menentukan kelulusan
df['Status'] = df['Nilai'].apply(lambda x: 'Lulus' if x >= 75 else 'Tidak Lulus')
df
```

Output:

	Nama	Nilai	Status		
0	Andi	73	Tidak Lulus		
1	Budi	80	Lulus		
2	Cantika	89	Lulus		
3	Dwi	77	Lulus		
4	Eka	90	Lulus		
5	Fira	75	Lulus		
6	Galih	92	Lulus		
7	Herry	83	Lulus		
8	Indah	83	Lulus		
9	Jay	78	Lulus		

Gambar 18. Output tampilan data df setelah kolom baru

Di sini kita menggunakan `apply()` dan `lambda` untuk membuat kolom “Status” yang menilai siswa lulus atau tidak.

3. Analisis Data dengan Numpy

NumPy (Numerical Python) adalah pustaka Python yang digunakan untuk menghitung data numerik, seperti operasi matematika, statistik, dan manipulasi data dalam bentuk array.

Menurut **Harris et al. (2020)** dalam jurnal *Nature*,

“NumPy is the fundamental package for scientific computing in Python, providing fast and efficient numerical operations.”

Berikut contoh sederhana penggunaan numpy menggunakan Google Colab:

```
import numpy as np

# Membuat array
nilai = np.array([80, 90, 75, 95, 85])

# Menampilkan hasil perhitungan
print("Rata-rata:", np.mean(nilai))
print("Nilai tertinggi:", np.max(nilai))
print("Nilai terendah:", np.min(nilai))
print("Median:", np.median(nilai))
```

Output:

```
Rata-rata: 85.0
Nilai tertinggi: 95
Nilai terendah: 75
Median: 85.0
```

4. Membuat Grafik dengan Matplotlib dan Seaborn

Python memiliki dua pustaka populer untuk visualisasi data, yaitu Matplotlib dan Seaborn. Keduanya digunakan untuk menampilkan data dalam bentuk grafik batang, garis, maupun lingkaran.

a. Membuat Grafik dengan Matplotlib

Matplotlib digunakan untuk membuat berbagai jenis grafik, seperti grafik batang (*bar chart*), garis (*line chart*), maupun lingkaran (*pie chart*). Pustaka ini merupakan dasar dari banyak pustaka visualisasi lainnya.

Langkah awal, pastikan kamu sudah menginstal pustaka matplotlib, seperti contoh di bawah ini:

```
pip install matplotlib
```

Berikut adalah bentuk implementasi sederhana menggunakan matplotlib pada Google Colab:

```
import pandas as pd

import matplotlib.pyplot as plt

# Data

siswa = ['Andi', 'Budi', 'Cantika', 'Dwi', 'Eka', 'Fira', 'Galih', 'Herry', 'Indah', 'Jay']

nilai = [73, 80, 89, 77, 90, 75, 92, 83, 83, 78]

# Membuat DataFrame

data = pd.DataFrame({ 

    'siswa': siswa, 

    'nilai': nilai

})

# Mengurutkan data berdasarkan nilai (ascending)

data_sorted = data.sort_values(by='nilai', ascending=True)

# Membuat grafik batang dari data yang sudah diurutkan

plt.bar(data_sorted['siswa'], data_sorted['nilai'], color='green')

plt.title('Nilai Ujian Siswa (Diurutkan)')

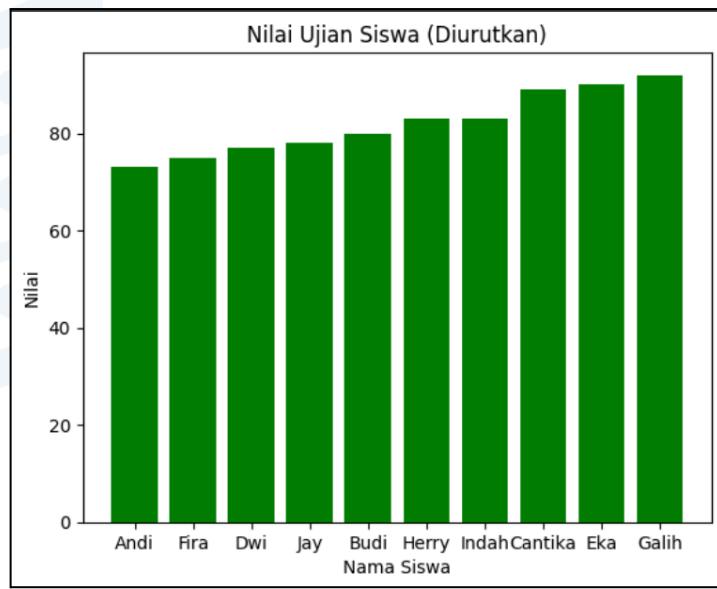
plt.xlabel('Nama Siswa')

plt.ylabel('Nilai')

plt.show()
```

Output:

Hasilnya adalah grafik sederhana yang menampilkan nilai ujian masing-masing siswa dari terendah ke tertinggi.



Gambar 19. Hasil visualisasi nilai ujian siswa

Penjelasan kode di atas:

- `pd.DataFrame()` digunakan untuk membuat tabel data dari daftar siswa dan nilai.
- `sort_values()` berfungsi mengurutkan data berdasarkan kolom `nilai` dari yang paling kecil ke paling besar.
- `plt.bar()` digunakan untuk membuat grafik batang berdasarkan nama siswa dan nilai.
- `plt.title()`, `plt.xlabel()`, dan `plt.ylabel()` menambahkan judul dan label pada sumbu X dan Y.
- `plt.show()` digunakan untuk menampilkan grafik yang telah dibuat.

b. Membuat Grafik dengan Seaborn

Seaborn merupakan pustaka visualisasi yang dikembangkan di atas Matplotlib.

Perbedaannya terletak pada tampilan grafik yang **lebih estetik** dan lebih mudah digunakan.

Sama seperti matplotlib, pastikan kamu sudah menginstal seaborn seperti contoh di bawah:

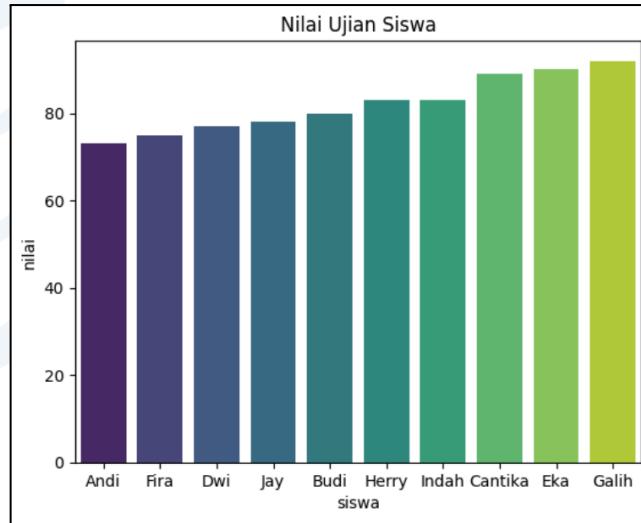
```
pip install seaborn
```

Berikut implementasi sederhana menggunakan pustaka seaborn pada Google Colab:

```
import seaborn as sns
```

```
import pandas as pd  
  
import matplotlib.pyplot as plt  
  
  
# Data  
  
siswa = ['Andi', 'Budi', 'Cantika', 'Dwi', 'Eka', 'Fira', 'Galih', 'Herry', 'Indah', 'Jay']  
  
nilai = [73, 80, 89, 77, 90, 75, 92, 83, 83, 78]  
  
# Membuat DataFrame  
  
data = pd.DataFrame({  
    'siswa': siswa,  
  
    'nilai': nilai  
})  
  
data_sorted = data.sort_values(by='nilai', ascending=True)  
  
# Membuat grafik batang dengan Seaborn  
  
sns.barplot(x='siswa', y='nilai', data=data_sorted, palette='viridis')  
  
plt.title('Nilai Ujian Siswa')  
  
plt.show()
```

Output:



Gambar 20. Hasil visualisasi nilai ujian siswa dengan seaborn

Penjelasan kode diatas:

- `pd.DataFrame()` digunakan untuk membuat tabel data dari daftar siswa dan nilai.
- `sort_values()` berfungsi mengurutkan data berdasarkan kolom `nilai` dari yang paling kecil ke paling besar.
- `sns.barplot()` digunakan untuk membuat grafik batang dengan tampilan lebih estetik menggunakan **Seaborn**.
- `palette='viridis'` memberikan warna gradasi pada grafik agar terlihat lebih menarik.
- `plt.title()` menambahkan judul grafik.
- `plt.show()` digunakan untuk menampilkan grafik yang telah dibuat.
- Melihat persentase kelulusan menggunakan pie chart.

```
df['Status'].value_counts().plot(kind='pie', autopct='%.1f%%',
```

```
colors=['lightgreen','lightcoral'])  
plt.title("Persentase Kelulusan Siswa")  
plt.ylabel("")  
plt.show()
```

Output:



Gambar 21. Persentase kelulusan visualisasi pie chart

Penjelasan kode:

`kind='pie'` → menentukan jenis grafik pie.

`autopct='%.1f%%'` → menampilkan persentase pada tiap potongan lingkaran.



5. Mini Project Python (Study Case)

Studycase

Referensi:

- Atlassian. (n.d.). *Essential chart types for data visualization*. Atlassian. <https://www.atlassian.com/data/charts/essential-chart-types-for-data-visualization>
- Bishop, C. M. (2023). *Pattern Recognition and Machine Learning*. Springer.
- Han, J., Kamber, M., & Pei, J. (2022). *Data Mining: Concepts and Techniques (4th ed.)*. Morgan Kaufmann...
- Harris, C. R., Millman, K. J., van der Walt, S. J., Gommers, R., Virtanen, P., Cournapeau, D., ... Oliphant, T. E. (2020). *Array programming with NumPy*. *Nature*, 585, 357–362. <https://doi.org/10.1038/s41586-020-2649-2>
- Kalsum, N., dkk. (2025). *Pengantar Sains Data: Teori dan Aplikasi*. Telkom University Press.
- Jaspersoft. (n.d.). *What is a Pie Chart?* <https://www.jaspersoft.com/articles/what-is-a-pie-chart>
- McKinney, W. (2022). *Python for data analysis (3rd ed.)*. O'Reilly Media. <https://wesmckinney.com/book/>
- University of Houston Libraries. (n.d.). *Types of data visualization*. University of Houston. <https://guides.lib.uh.edu/c.php?g=1142515&p=8335670>