

Quiz week 2 : Excel

- Silahkan kerjakan soal-soal dibawah ini menggunakan microsoft excel dan kumpulkan hasil jawaban dari setiap soal dalam satu buah file excel.
- Upload seluruh file yang anda gunakan untuk mengerjakan quiz week 2 ini dalam folder google drive sesuai ketentuan juknis. File yang diupload antara lain (**Lemonade.csv yang sudah diolah, file Quiz_2_pandas.pynb yang telah dijawab, Quiz week 2.pdf, dan Salaries.csv**)
- Soal quiz pandas dapat anda akses pada file ipynb yang terlampir. **SELAMAT MENGERJAKAN. Tetap semangat yaaa.**

Deskripsi : Dalam tugas week 1 telah terlampir file excel yang akan digunakan untuk mengerjakan soal dalam quiz kali ini. Data pertama merupakan data penjualan sebuah toko jus lemon selama satu tahun. Data terdiri dari 6 fitur kolom antara lain **Date, Day, Temperature, Rainfall, Flyers, Price, Sales**. Tugas anda kali ini adalah sebagai berikut.

1. Lakukan pengolahan pada data csv tersebut sehingga akan tampak seperti ini :

Date	Month	Day	Temperat	Rainfall	Flyers	Price	Sales	Revenue
01/01/2017	Januari	Sunday	27	2,00	15	0,3	10	\$3,00
02/01/2017	Januari	Monday	28,9	1,33	15	0,3	13	\$3,90
03/01/2017	Januari	Tuesday	34,5	1,33	27	0,3	15	\$4,50

2. Selanjutnya setting tabel penjualan tersebut sehingga memiliki fitur filter seperti yang tampak pada gambar berikut:

Date	Month	Day	Temperature	Rainfall	Flyers	Price	Sales	Revenue
01/01/2017	Januari	Sunday	27	2,00	15	0,3	10	\$3,00
02/01/2017	Januari	Monday	28,9	1,33	15	0,3	13	\$3,90
03/01/2017	Januari	Tuesday	34,5	1,33	27	0,3	15	\$4,50

3. Data yang telah anda olah terlihat lebih baik untuk di lihat. Namun terlihat cukup membosankan untuk melihat data tabulasi yang hanya menampilkan angka tanpa grafik. Oleh karena itu anda dapat menggunakan fitur dalam excel seperti styles atau condition formatting untuk membuat data anda lebih informatif dan lebih nyaman untuk dipandang. Sebagai contoh penjual ingin menandai temperatur tinggi dengan warna merah dan temperatur rendah dengan warna yang lebih putih kemudian mengurutkan data berdasarkan nilai temperatur terendah ke nilai tertinggi.

Maka hal yang perlu dilakukan adalah sebagai berikut :

- **Select data pada fitur kolom temperature**
- Pilih **Conditional Formatting** yang ada di menu bar
- Kemudian pilih menu **Color Scale**

- Kemudian pilih sesuai keinginan penjual tersebut.
- Terakhir **klik tombol filter** yang berada di pojok kanan fitur temperature dan pilih **sort Smallest to Largest**
- Hasilnya akan tampak seperti berikut :

1	Date	Day	Temperature	Rainfall	Flyers	Price	Sales
2	31/12/2017	Sunday	15.1	2.5	9	0.3	7
3	05/12/2017	Tuesday	22	1.82	11	0.3	10
4	06/01/2017	Friday	25.3	1.54	23	0.3	11
5	01/01/2017	Sunday	27	2	15	0.3	10

Data urutan teratas

	Date	Day	Temperature	Rainfall	Flyers	Price	Sales
363	18/07/2017	Tuesday	99.3	0.47	76	0.5	41
364	22/07/2017	Saturday	99.6	0.47	49	0.5	42
365	26/06/2017	Monday	102.6	0.47	60	0.3	42
366	01/07/2017	Saturday	102.9	0.47	59	0.5	43

Data urutan terbawah

TUGAS : sekarang tugas anda membuat kolom **Rainfall,Flyers,Price,Sales** agar menjadi lebih informatif lagi hingga tampak seperti ini.

	A	B	C	D	E	F	G
1	Date	Day	Temperature	Rainfall	Flyers	Price	Sales
2	01/01/2017	Sunday	27		2	15	0.3
3	02/01/2017	Monday	28.9		1.33	15	0.3
4	03/01/2017	Tuesday	34.5		1.33	27	0.3

Disini anda bisa melihat bahwa kolom rainfall menunjukkan curah hujan yang digambarkan oleh grafik biru semakin tinggi curah hujan maka grafik warna biru semakin memenuhi cell Rainfall. Kemudian jumlah flyers yang dibagikan digambarkan oleh grafik tanda panah seperti yang tampak pada gambar. Price 0.3 ditandai dengan tanda segitiga merah dan 0.5 ditandai oleh segitiga warna hijau. Terakhir jumlah penjualan harian lebih mudah terlihat dengan grafik warna kuning yang tampak di masing-masing cell-nya.

- Selanjutnya anda ditugaskan untuk meng-highlight 10% penjualan terendah dengan warna cell merah dan 10% penjualan tertinggi dengan warna cell hijau.

Hasilnya akan tampak seperti berikut untuk 10% **Terendah** :

	A	B	C	D	E	F	G
1	Date	Day	Temperature	Rainfall	Flyers	Price	Sales
2	31/12/2017	Sunday	15.1		2.5	9	0.3
3	05/12/2017	Tuesday	22		1.82	11	0.3
4	06/01/2017	Friday	25.3		1.54	23	0.3

Hasilnya akan tampak seperti berikut untuk 10% **Tertinggi** :

	A	B	C	D	E	F	G
1	Date	Day	Temperature	Rainfall	Flyers	Price	Sales
2	27/07/2017	Thursday	97.9	0.47	74	0.5	43
3	01/07/2017	Saturday	102.9	0.47	59	0.5	43
4	22/07/2017	Saturday	99.6	0.47	49	0.5	42

5. Kali ini anda ditugaskan untuk membuat laporan statistika sederhana pada data penjualan lemon yang sudah anda olah sebelumnya. Anda diminta untuk menampilkan nilai mean, median, mode, Varian dari populasi dan standar deviasi dari populasi (**untuk kolom Temperature, Rainfall, dan Sales**). Gunakan rumus yang sudah tersedia dalam excel. **Hasil output** : (beri penjelasan tentang apaitu mean, median, mode, variance of population dan std dev population)

Temperature Statistics

Mean
Median
Mode
Variance
Std Dev

X
X
X
X
X

Rainfall Statistics

Mean
Median
Mode
Variance
Std Dev

X
X
X
X
X

Sales Statistics

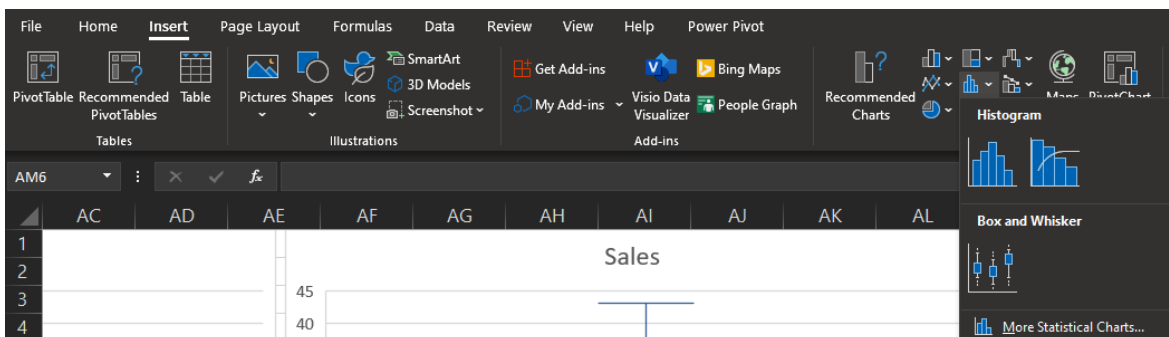
Mean
Median
Mode
Variance
Std Dev

X
X
X
X
X

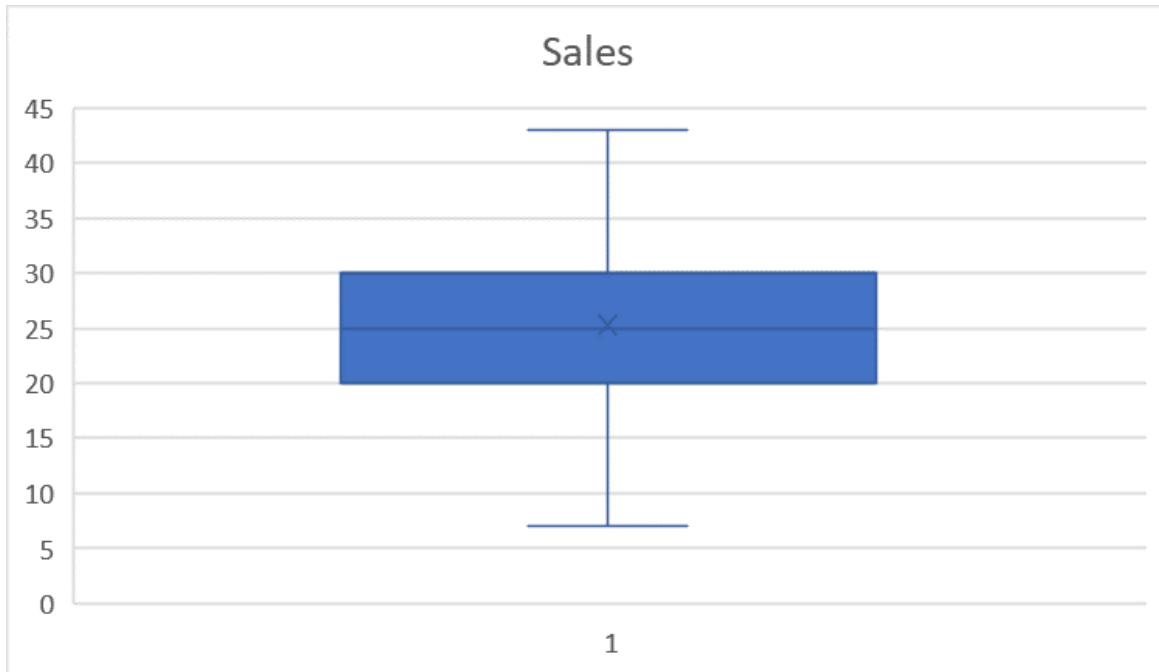
6. Sekarang kita akan mencoba membuat sebuah visualisasi sederhana pada data yang telah kita miliki.

Disini kita akan mencoba membuat boxplot dari fitur **Sales** perhatikan contoh berikut untuk membuat boxplot :

- Pertama select kolom **Sales**
- Kemudian pilih menu **insert**
- Kemudian pilih **Chart**
- Terakhir pilih menu **box and whisker**



Hasilnya akan tampak seperti ini :



Perhatikan bagan, dan pahami hal-hal berikut:

- Garis horizontal di tengah menunjukkan nilai median untuk penjualan. Ini adalah persentil 50% - dengan kata lain, 50% dari nilai lebih tinggi dari ini, dan 50% lebih rendah.
- Tanda **X** di kotak menunjukkan **mean** yang mana sedikit lebih tinggi dari median.
- Kotak yang dipenuhi dengan warna biru menunjukkan kisaran nilai di antara kuartil kedua dan ketiga. Dengan kata lain, dari persentil ke-25 hingga persentil ke-75. Nilai-nilai ini berkisar dari sekitar 20 hingga 30, menunjukkan bahwa jumlah penjualan pada separuh hari nya berada dalam rentang ini,
- Garis-garis memanjang dari kotak (dikenal sebagai kumis) menunjukkan kisaran untuk kuartil pertama dan keempat, di mana ada lebih banyak atau lebih sedikit penjualan daripada di kuartil kedua dan ketiga.
- Terlihat bahwa tidak ada percikan dalam kolom **Sales**.

Tugas : Anda diminta membuat boxplot padat fitur Rainfall dan lakukan penjabaran hasil analisa yang dapat diperlihatkan oleh boxplot yang anda buat. Hasil boxplot kurang lebih akan tampak seperti ini.

7. Dalam mengolah data sering kali perlu melihat nilai korelasi dari dua variabel atau fitur data. Misalnya disini kita ingin melihat korelasi dari fitur **Temperature dan Sales**. Dalam excel nilai korelasi dari dua fitur dapat dihitung menggunakan rumus berikut :

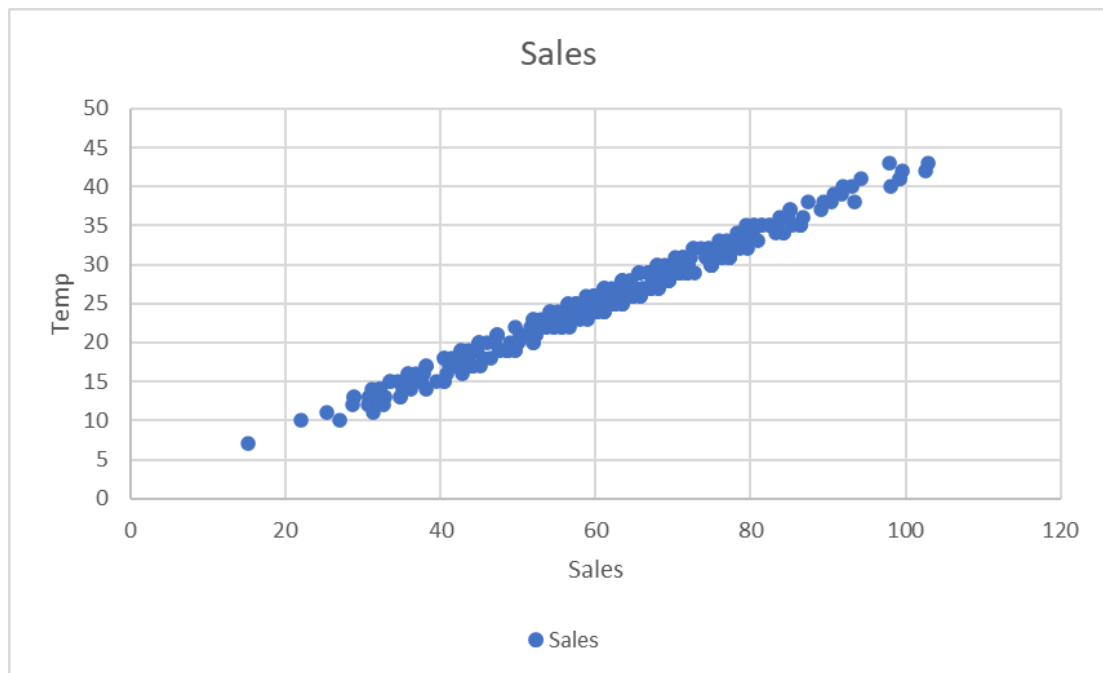
$$= CORREL(fitur1; fitur2)$$

Nilai korelasi dari fitur **Temperature** dan **Sales** akan tampak seperti ini :

Correlation
0.989832

Analisa : Hasil hitung nilai korelasi antara suhu dan jumlah penjualan harian menghasilkan nilai sekitar 0,989832. Hal ini menunjukkan sebuah korelasi positif yang memperlihatkan kekuatan hubungan secara statistik yang jelas dari dua variabel/fitur numerik dalam hal ini, suhu dan jumlah penjualan harian. Disini kita bisa sedikit menyimpulkan bahwa suhu memiliki pengaruh yang positif terhadap jumlah penjualan harian. Nilai penjualan harian jus limun akan tampak lebih tinggi ketika suhu sedang tinggi atau panas (*meskipun ini bukanlah alasan utama mengapa penjualan limun meningkat secara mutlak*)

Nilai korelasi dapat divisualkan menggunakan scatter plot seperti contoh dibawah :



Korelasi diukur sebagai nilai antara -1 dan 1. Nilai mendekati 1 menunjukkan korelasi positif; dengan kata lain, nilai tinggi untuk satu variabel tampaknya sesuai dengan nilai tinggi untuk variabel lain. Nilai mendekati -1, di sisi lain, menunjukkan korelasi negatif, di mana nilai-nilai tinggi untuk satu variabel sesuai dengan nilai-nilai rendah untuk variabel lain. Nilai mendekati 0 menunjukkan kurangnya hubungan yang dapat dilihat di antara variabel.

Para ahli statistik sering mengatakan “korelasi bukanlah sebab-akibat”. Kita dapat menggunakan korelasi untuk menentukan hari-hari dengan volume penjualan tinggi cenderung memiliki suhu tinggi, tetapi kita tidak bisa mengatakan bahwa Rosie menjual

banyak limun pada hari tertentu karena suhu tinggi. Sama halnya seperti kita tidak bisa mengatakan bahwa hari itu sangat panas karena Rosie menjual banyak limun. Faktanya, mungkin ada faktor ketiga "*hidden factor*" yang mempengaruhi penjualan limun dan suhu. Misalnya, selama musim panas, ketika lemon sedang musim, suhu cenderung lebih tinggi.

Tugas : Anda diminta membuat visualisasi scatter plot untuk memperlihatkan korelasi antara fitur Rainfall dan Sales. Gunakan rumus excel untuk mendapatkan nilai korelasi dari kedua fitur tersebut. Jabarkan hasil analisa yang anda dapatkan dari hasil perhitungan dan visualisasi yang anda dapatkan.