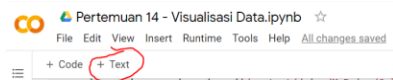


Proyek

Di dalam proyek kali ini, Anda akan diminta untuk melakukan pemrograman dan analisa data dengan menggabungkan semua keilmuan yang sudah Anda pelajari selama pertemuan 9-15. Ingat, proyek ini bernilai 30% dari total seluruh nilai Anda karena itu harap dikerjakan dengan sebaik mungkin. Mohon dikerjakan sendiri-sendiri karena adanya tindakan plagiarisme akan mengakibatkan kepada pemotongan nilai/pemberian nol.

Petunjuk pengerjaan:*

* Pastikan semua langkah petunjuk tercatat di dalam Jupyter Notebook + manfaatkan Text untuk memberi penjelasan apa yang Anda kerjakan



1. Download dataset yang akan digunakan di sini:
<https://www.kaggle.com/zynicide/wine-reviews>, yaitu: *winemag-data-130k-v2.csv*

Deskripsi:

Dataset ini adalah dataset yang berisikan informasi mengenai hasil review dari wine di dalam website https://www.winemag.com/?s=&drink_type=wine.

Deskripsi masing-masing kolom:

- points: besarnya nilai dari wine setelah dicicipi oleh winery
- descriptions: penjelasan dari wine
- designation: nama kebun wine
- price: harga wine
- country: asal negara dari wine
- province: asal provinsi dari wine
- region_1: asal daerah dari wine
- region_2: asal daerah dari wine
- variety: jenis anggur
- winery: reviewer dari wine

	country	description	designation	points	price	province	region_1	region_2	variety	winery
0	US	This tremendous 100% varietal wine hails from ...	Martha's Vineyard	96	235.0	California	Napa Valley	Napa	Cabernet Sauvignon	Heitz
1	Spain	Ripe aromas of fig, blackberry and cassis are ...	Carodorum Selección Especial Reserva	96	110.0	Northern Spain	Toro	NaN	Tinta de Toro	Bodega Carmen Rodríguez
2	US	Mac Watson honors the memory of a wine once ma...	Special Selected Late Harvest	96	90.0	California	Knights Valley	Sonoma	Sauvignon Blanc	Macaleay
3	US	This spent 20 months in 30% new French oak, an...	Reserve	96	65.0	Oregon	Willamette Valley	Willamette Valley	Pinot Noir	Ponzi
4	France	This is the top wine from La Bégude, named aft...	La Brûlade	95	66.0	Provence	Bandol	NaN	Provence red blend	Domaine de la Bégude

Gambar 1. Tampilan dataset *winemag-data-130k-v2.csv*

2. Panggil data ke dalam Python lalu simpan dengan nama variabel **wine_review**.
3. Buatlah fungsi yang dapat menghitung besarnya nilai **rata-rata** dan **standar deviasi** lalu gunakan fungsi tersebut untuk mencari nilai rata-rata dan standar deviasi dari **points** dan **price**

4. Dengan menggunakan library seaborn buatlah histogram untuk **points** dan **price** dengan menggunakan *distribution plot* + tambahkan garis yang memperlihatkan rata-rata dan rata-rata +- standar deviasi.
Petunjuk:
 - <https://seaborn.pydata.org/tutorial/distributions.html>
 - <https://stackoverflow.com/questions/36146684/using-seaborn-how-can-i-draw-a-line-of-my-choice-across-my-scatterplot>
5. Dengan menggunakan fungsi groupby() yang sudah dipelajari, hitunglah kontribusi jumlah wine dari masing-masing negara di dalam kolom **country**. Tentukan 3 negara dengan jumlah kontribusi jumlah paling besar
6. Filter data **wine_review** dengan syarat sebagai berikut: berdasarkan hasil langkah 3 dan nilai **points** di atas 96.
Petunjuk: <https://stackoverflow.com/questions/13611065/efficient-way-to-apply-multiple-filters-to-pandas-dataframe-or-series>
7. Dengan menggunakan library matplotlib buatlah *scatter plot* dengan nilai sumbu-x adalah **points** dan sumbu-y adalah **price** lalu tambahkanlah legenda berdasarkan negara dari kolom **country**
8. Perhatikan hasilnya lalu tuliskan kira-kira apa yang menjadi kekurangan dari penggunaan *scatter plot* untuk data **wine_review** yang sudah difilter pada langkah 4
9. Untuk menindaklanjuti kelemahan pada langkah 7, buatlah *swarm plot* dengan menggunakan library seaborn dengan legenda berdasarkan negara dari kolom **country**
Petunjuk: <https://seaborn.pydata.org/generated/seaborn.swarmplot.html>
10. Perhatikan hasilnya lalu tuliskan pendapat Anda mengenai negara mana yang memiliki kualitas wine paling baik dengan harga paling murah
11. Lanjutkanlah analisa Anda dengan melakukan filter untuk data **wine_review** pada langkah 4 berdasarkan negara yang menurut Anda memiliki kualitas wine paling baik lalu ulangi langkah 8 dengan legenda berdasarkan varietas wine dari kolom **variety**
12. Perhatikan hasilnya lalu tuliskan pendapat Anda mengenai varietas wine mana yang memiliki kualitas paling baik dengan harga paling murah

Nilai tambahan:

1. Jika Anda dapat menceritakan proses debugging yang Anda temui selama pengerjaan project + cara untuk menyelesaikannya. Tuliskan di dalam Text Jupyter Notebook
2. Jika Anda dapat membuat analisa lain selain dari langkah 1-12 dengan cerita yang menarik akan mendapat tambahan nilai

Pengumpulan: Kumpulkan tugas dengan dalam format .ipynb dalam kondisi semua baris sudah di run.

Format pengumpulan proyek:

<Nama>_<NIM>_Project.ipynb

Ex: Lucas_13312024_Project.ipynb