Projekt 1

Temat:

Wykonaj wykresy czasowe w formie interaktywnej z możliwością wyboru co najmniej dwóch serii danych na wykresie. Dla danych wykonaj wizualizację braków danych. Uzasadnij wybrane narzędzia, zinterpretuj rezultaty prac.

Alicja Osam-Gyaabin, Mikołaj Zawada, Karol Kociołek

Zbiór

Zbiór pobrany z kaggle - https://www.kaggle.com/datasets/alistairking/weather-long-term-time-series-forecasting. Dotyczy pogody rejestrowanej co 10 minut przez cały rok 2020 i obejmuje 20 mierzonych wskaźników meteorologicznych.

date – Data i godzina obserwacji.

p – Ciśnienie atmosferyczne w milibarach (mbar).

T – Temperatura powietrza w stopniach Celsjusza (°C).

Tpot – Temperatura potencjalna w Kelwinach (K), reprezentująca temperaturę, jaką miałby pakiet powietrza, gdyby został przeniesiony na standardowy poziom ciśnienia.

Tdew – Temperatura punktu rosy w stopniach Celsjusza (°C), wskazująca temperaturę, przy której powietrze staje się nasycone wilgocią.

rh – Wilgotność względna w procentach (%), pokazująca ilość wilgoci w powietrzu w stosunku do maksymalnej ilości, jaką może pomieścić przy tej temperaturze.

VPmax – Maksymalne ciśnienie pary wodnej w milibarach (mbar), reprezentujące maksymalne ciśnienie wywierane przez parę wodną przy danej temperaturze.

VPact – Rzeczywiste ciśnienie pary wodnej w milibarach (mbar), wskazujące bieżące ciśnienie pary wodnej w powietrzu.

VPdef – Niedobór ciśnienia pary wodnej w milibarach (mbar), mierzący różnicę między maksymalnym a rzeczywistym ciśnieniem pary wodnej, wykorzystywany do oceny potencjału osuszania.

sh – Wilgotność specyficzna w gramach na kilogram (g/kg), pokazująca masę pary wodnej na kilogram powietrza.

H2OC – Stężenie pary wodnej w milimolach na mol (mmol/mol) suchego powietrza.

rho – Gęstość powietrza w gramach na metr sześcienny (g/m³), odzwierciedlająca masę powietrza na jednostkę objętości.

wv – Prędkość wiatru w metrach na sekundę (m/s), mierząca poziomą prędkość ruchu powietrza.

max. wv – Maksymalna prędkość wiatru w metrach na sekundę (m/s), wskazująca najwyższą zanotowaną prędkość wiatru w danym okresie.

wd – Kierunek wiatru w stopniach (°), reprezentujący kierunek, z którego wieje wiatr.

rain – Łączna opad deszczu w milimetrach (mm), wskazująca ilość opadów w okresie obserwacji.

raining – Czas trwania deszczu w sekundach (s), rejestrujący czas, przez który występowały opady deszczu w okresie obserwacji.

SWDR – Krótkofalowe promieniowanie w dół w watach na metr kwadratowy (W/m²), mierzące napływające promieniowanie słoneczne.

PAR – Promieniowanie fotosyntetycznie aktywne w mikromolach na metr kwadratowy na sekundę (µmol/m²/s), wskazujące ilość światła dostępnego do fotosyntezy.

max. PAR – Maksymalne promieniowanie fotosyntetycznie aktywne zarejestrowane w okresie obserwacji w µmol/m²/s.

Tlog – Zarejestrowana temperatura w stopniach Celsjusza (°C), potencjalnie z drugiego czujnika lub loggera.

OT – Prawdopodobnie odnosi się do "znacznika czasowego operacyjnego" lub przesunięcia czasowego, ale może wymagać doprecyzowania w zależności od kontekstu zbioru danych.

Wykres czasowe

Wykres 1:

Obraz zawierający tekst, zrzut ekranu, Wykres, Czcionka

Opis wygenerowany automatycznie

Wykres 2:

Obraz zawierający zrzut ekranu, tekst, Wykres, Oprogramowanie graficzne

Opis wygenerowany automatycznie

Wnioski

* Wykorzystano narzędzie bokeh, czyli bibliotekę do tworzenia interaktywnych wizualizacji danych w Pythonie.
* Wykres pozwala na wizualizację dwóch wybranych zmiennych na osi czasu.

Wizualizacja braków danych

Braki zostały sztucznie wygenerowane.

Obliczenie procent braków:

Obraz zawierający tekst, zrzut ekranu, oprogramowanie, Czcionka

Opis wygenerowany automatycznie

Wizualizcja procentu braków przy pomocy wordcloud:

Obraz zawierający tekst, Czcionka, zrzut ekranu, Grafika

Opis wygenerowany automatycznie

Wykres macierzowy utworzony w pakiecie missingno:

Obraz zawierający Prostokąt, wzór, kwadrat, monochromatyzm

Opis wygenerowany automatycznie

Wykres słupkowy utworzony w pakiecie missingno:

Obraz zawierający tekst, linia, zrzut ekranu, czarne i białe

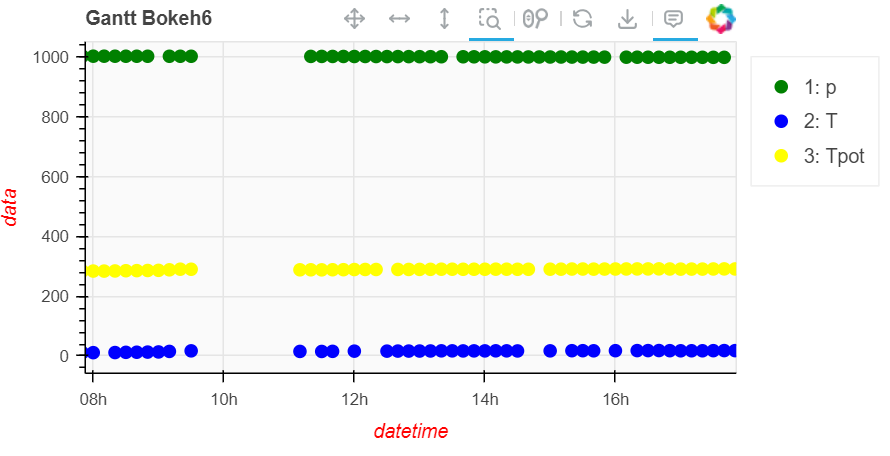
Opis wygenerowany automatycznie

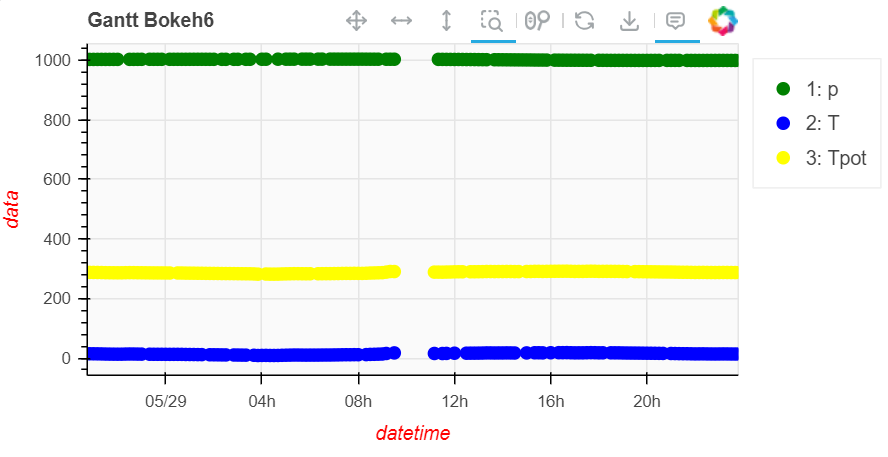
Wizualizacja z formatowaniem warunkowym i zapisanie do pliku excel:

Obraz zawierający tekst, zrzut ekranu, numer, Równolegle

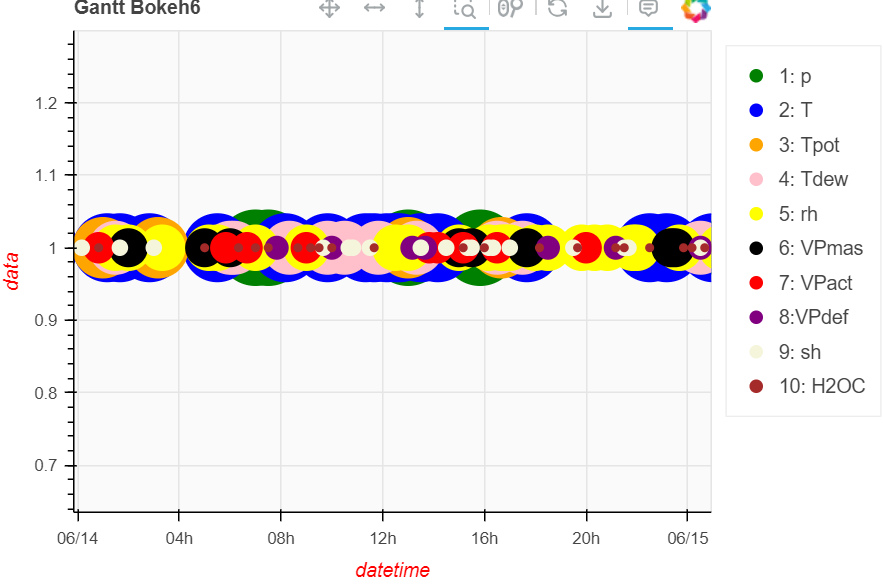
Opis wygenerowany automatycznie

Gantt punktowy prosty:





Gantt punktowy braków danych z nakładaniem serii dla większej liczby zmiennych



Wnioski

* W całym projekcie do manipulacji danymi została użyta biblioteka pandas. Użyto również worldcloud oraz missingno.
* Przy pomocy missingo wygenerowany został wykres macierzowy pozwalający no zobrazowanie ogólnej ilości braków oraz wskazanie ich miejsca jak również wykres słupkowy pokazujący stopień pełności każdej kolumny
* Przy pomocy worldcloud można w bardziej przyjemny dla oka sposób zobrazować różnice w brakach. Może nie być najlepsze przy dużej ilości kolumn lub małych różnicach.
* Bardzo dobrze i dokładnie pokazuje braki formatowanie warunkowe.
* Dołożenie narzędzi interaktywnych do Gantta punktowego prostego pozwala na zobrazowanie braków danych w skali mikro i makro.
* Przy dużej ilości danych Gantt może nie być najlepiej czytelny.