

Logboek groepje 16

Logboek Hidde Jessen 2541190

5-06 tot 9-06:

bekend worden met de data, eerste plotjes maken en missing values/outliers vinden

10-06:

bekend raken met Bokeh

11-06:

Nieuwe plattegrond gemaakt van het gebied, functies geschreven die lijnen tussen grafieken en sensoren maken en hier een range van + of - x graden omheen aftekenen

12-06:

oude grafieken omzetten naar Bokeh, begin maken van functie die effect van de windrichting onderzoekt: kijk of er een sensor tov een fabriek in de lijn van de windrichting ligt en of er op dit moment iets gebeurt met de gemeten waardes

13-06 en 14-06:

studeren voor tentamen voor andere studie

15-06 en 16-06

terug naar de EDA, grafiek voor elke sensor van elk stofje gemaakt, experimenteren met interactief maken van de grafieken.

17-06:

grafieken samenvoegen + eerste plot obv windrichting

18-06:

zo algemeen mogelijke functie schrijven voor plotter tov wind + documentatie bijwerken

19-06:

Proberen grafieken interactief te maken, fouten uit eerdere grafieken gehaald, verschillende functies geschreven die hoeken tussen fabrieken en sensoren berekent

20-06:

Momenten van overlap tussen windrichting en fabriek->sensor bepalen, op deze momenten de gemeten waardes bij de sensor bepalen en deze per fabriek plotten

21-06:

Registreren van windoverlap verbeteren. Terug naar de basis en opnieuw opgebouwd.

22-06:

Schrijven verslag

23-06 en 24-06:

outliers registreren obv Standard error of the mean icm windrichting. Stuk code herschrijven zodat het geen 25 minuten meer nodig heeft om te runnen. Per grafiek het gemiddelde toevoegen, wordt geprobeerd hier een betrouwbaarheidsinterval omheen weer te geven, lukt nog niet helemaal

25-06:

filter voor sensor 4 schrijven, plattegrond van de fabrieken en sensoren in bokeh schrijven ipv matplotlib

26-06:

zoeken naar beantwoording van vraag 3, fabrieken aan chemicalien koppelen. Verschillende maten geprobeerd, extra loop ingebouwd om de gemiddelde readings bij overlap te kunnen vergelijken bij de overall gemiddelde readings.

27-06 en 28-06:

comments bij code voegen, git opruimen verslag schrijven, resultaten verwerken en discussie schrijven. Klooien met tabellen in Latex

Logboek Sven Boogmans 11319666**- 04/06/18**

Overleg met de groep, kiezen van een dataset (gekozen voor de VAST challenge), vragen opgesteld voor de TA meeting.

- 05/06/18

TA meeting, definitief gekozen voor VAST challenge als dataset. Pre-processing begonnen, git werkend krijgen, juiste software downloaden voor het project, de dataset downloaden, en de verschillende datatypes uit de dataset in kaart brengen.

-06/06/18

Groepsoverleg over pre-processing. Ervoor gekozen dat ieder zijn eigen pre processing doet, om er voor te zorgen dat iedereen gevoel krijgt voor de data. Gewerkt met pandas, een pandas dataframe gemaakt van de dataset (twee dataframes, een voor elke dataset).

-07/06/18

TA meeting, debuggen van pre-processing, twee dataframes samenvoegen tot een dataframe, manueel data bewerken (oa. kolom namen logischer maken).

-08/06/18

Overleg met de groep, terugblikken op de pre processing week, en uitwisselen van gevonden rariteiten in de data. Functies schrijven waarmee scatterplots gemaakt kunnen worden (nog niet gesorteerd op chemical of sensor).

-11/06/18

Functies waarmee scatterplots gemaakt kunnen worden voor de windrichting en windsnelheid over de tijd. Vragen voorbereiden voor TA meeting volgende dag.

-12/06/18

TA meeting. EDA: verdelen van taken. Ik ga me bezig houden met de sensor functioning. Begin maken met het scatterplotten van readings over tijd per sensor.

-13/06/18

Scatterplotten van readings over tijd per sensor gelukt. Nu de analyse. Raar gedrag gezien voor sensor 4 en 9. Sensor 4 lijkt last te hebben van drifting, sensor 9 heeft een plotse sprijding in de data.

-14/06/18

De plotjes net maken (as namen, juiste schaal etc.). Ook raar gedrag vinden voor sensor 3 (grotere spreiding) en 5 (geleidelijk groter worden van spreiding).

-15/06/18

Present findings in TA-meeting, planning maken met groepje en resultaten vergelijken (plotjes uitwisselen).

-17/06/18

Resultaten verwerken in tekst voor verslag

-18/06/18

Resultaten verwerken in tekst voor verslag

-19/06/18

Het sorteren van de data in elke test periode (april, augustus, and december). Functie die gemiddelde kan berekenen voor elke testperiode gegeven een dataframe.

-20/06/18

Functies die voor elke chemical en sensor de standaard deviatie en het gemiddelde kan berekenen. De resultaten komen overeen met de resultaten uit de EDA, de standaard deviatie is groter voor sensor 3, gemiddelde stijgt bij sensor 4, standaard deviatie stijgt voor sensor 5, en stijgt plots voor sensor 9.

-21/06/18

Resultaten presenteren bij TA meeting, beginnen aan eerste versie verslag (inleiding en resultaten voor de functioning of the sensors).

-22/06/18

Inleiding verslag afmaken, en resultaten voor functioning sensors in het verslag verwerken.

-25/06/18

Beginnen aan de website. Opzetje gemaakt, experimenteren met bokeh importeren en op zoek gaan naar een hosting service (niet gevonden).

-26/06/18

TA-meeting, 1e versie verslag bespreken, website laten zien, website verder afmaken, jump to linkjes toevoegen aan website.

-27/06/18

Bokeh plots toevoegen aan de website.

-28/06/18

Feedback 1e versie verslag verwerken, laatste hand aan de website, python code opschonen.

Angelo Groot:

Invested time in the project

n'th of June	what?
3-10	Data exploration for cleaning.etc
11	tried to improve upon the data set format
12	temporarily gave up improving the data set, started trying to interpolate wind direction linearly
13	got the idea to interpolate wind direction according to its independent vectors, finished interpolation methods for Cubic rom-spline and linear interpolation
14	bugfixing
15	bugfixing
16	more bugfixing
17	Fixed "bugs" in interpolation methods, created a program that makes plots of all the combinations of sensors with their readings during wind speeds. also plotted all chems to the wind speed.
18	Worked on scatterplots for chemical abundancy, tried to use bokeh with markdown pages
19	Made cumulative abundancy plots for the chemicals at each sensor
20	Updated the final data
21	Worked on the report and plots
22	Worked on the report
23	-- (weekend be weekend)
24	made different correction functions for sensor 4
25	explored github pages and possibilities therein
26	worked on the report, exploring different visualisations for data
27	tried working on the website, errors everywhere
28	worked like mad on the still incomplete report.

Phillip

Week of 4th of June until 8th of June:

Within the first week I took a lot of time learning the basics of pandas. I played with the manipulation of dataframes such as adding, deleting and renaming columns, extracting rows and performing manipulations of values based upon indices. Also I got my thinking process up again and checked some basic linear algebra.

Monday 11th of June

We discussed the issue of interpolation and how we were to address the question of missing data. Linear interpolation was brought up and I said it was my goal to write a function which created a „smooth interval“. Additionally, some playful engagement with

Tuesday 12^h of June

Googling different interpolation techniques. Different versions of splining with respect to degrees of freedom, different dimensions within which to apply these.

Wednesday 13th of June

Tried to rewrite the Catmull-Rom splining technique from Wikipedia to fit our data. Failed at retrieving the values at specific points of the resulting graph. Then discovered basic spline functions in python libraries and got em working.

Thursday 14th of June

Thereafter proceeded to address the question of how to arrive at definite conclusions about the chemical release. Watched some youtube on different dispersion models and controlled fluid dynamic models.

Friday 15th of June

Continued watching videos on dispersion models, disregarded CFD models. Read the paper „*Spatial Interpolation of Fine Particulate Matter Concentrations Using the Shortest Wind-Field Path Distance*“.

Monday 18th of June

Set aside the dispersion models for a day and wrote a program that splits the data into monthly chunks, therewithin creating daily output releases of each sensor for all hours of the day.

Tuesday 19^h of June

Fixed bug within the program above and printed many, many plots.

Wednesday 20th of June

Turned to dispersion models again, trying to figure out how to compute the inverse of a Gaussian dispersion model, as we do not know the emission rate but do know the concentration measured in the field, iverse to what is regular „irl“.

Thursday 21st of June

Tried writing a function, failed writing a function.

Friday 22nd of June

Contemplated how to model the standard deviation of a wind field based upon singular input. Cam

up with using a time interval around the given datapoint to see overall fluctuations within this range. Kept working on the cost model function. Wrote function that removes outliers based upon a quantile range applied to each column individually.

Monday 25th of June

Coding my program, achieved to create a function that converts degrees to vectors, and a function that combines source_sensor paths to compute angle and downwind distance (polar coordinates).

Tuesday 26th of June

Coding my program, trying to apply this function with respect to all four factories at any given time within the dataframe. Then comparing the inverse of the necessary emission as an indication of likelihood of this source-sensor pair, adding them up and disregarding all values where likelihood was below .5. Then using this likelihood (squared to strengthen reliability of data used) times the computed emission as a weight to assign the factory/sensor pair with the highest explanatory value/lowest needed emission. Failed

Wednesday 27th of June

Failed.

Thursday 28th of June

Failed until 3:30pm. Then turned to the report and wrote some datacleaning functions and produced plots for all months regarding the hourly output of chemicals.