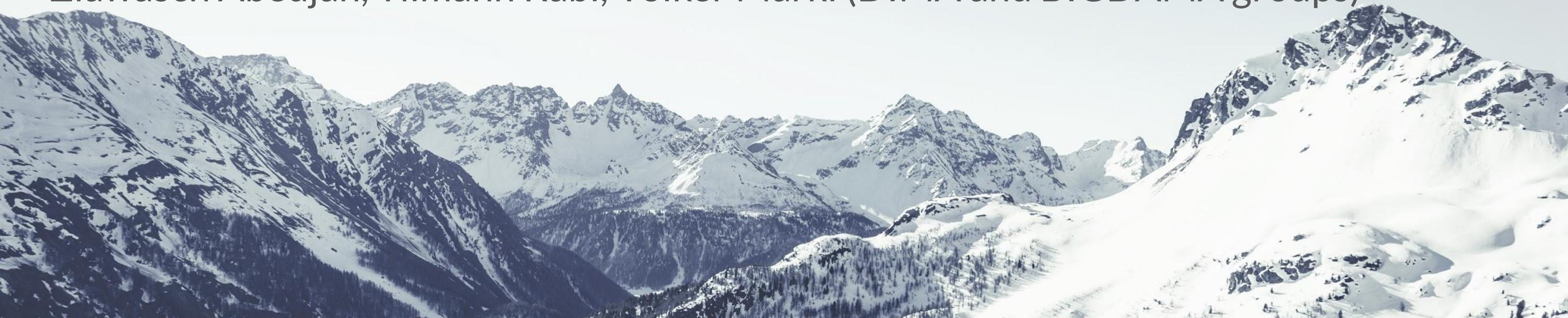
Technische Universität Berlin

Deutsches Forschungszentrum für Künstliche Intelligenz GmbH

Explanation of Air Pollution Using External Data Sources

Mahdi Esmailoghli, Sergey Redyuk, Ricardo Martinez, Ariane Ziehn, Ziawasch Abedjan, Tilmann Rabl, Volker Markl (DIMA and BIGDAMA groups)



BTW Data Science Challenge

LuftDaten (pollution sensor data)

Challenges:

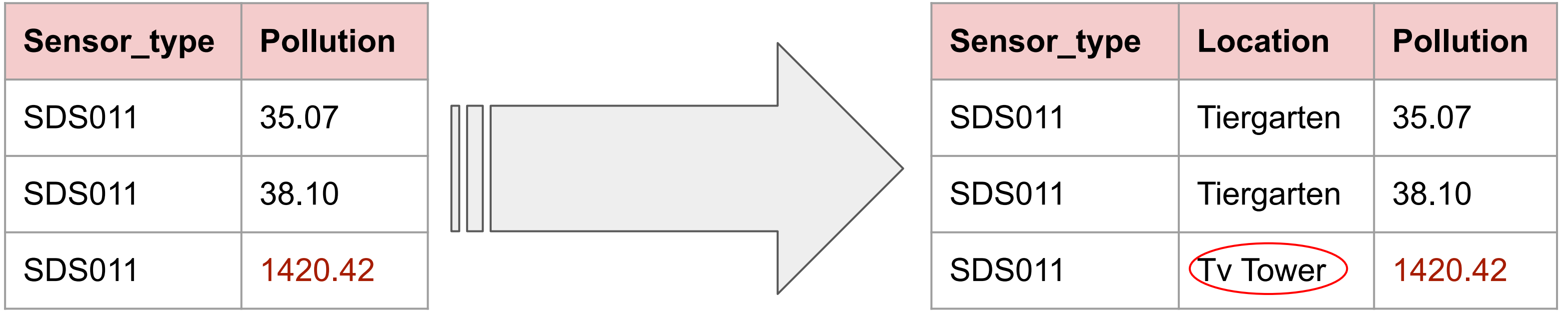
- Limited feature set
- Different schemas/sensors
- Malfunctioning sensors
- Stream nature of data

BTW Data Science Challenge - Our Goal

- Goal:
  - Explaining air pollution
  - Detecting the reasons of low air quality
- Problem:
  - Lack of information in provided data
  - Current ML algorithms cannot explain pollution based on provided data

BTW Data Science Challenge - Our Proposal

- Decision tree and Macrobase [Bailis’2017]*

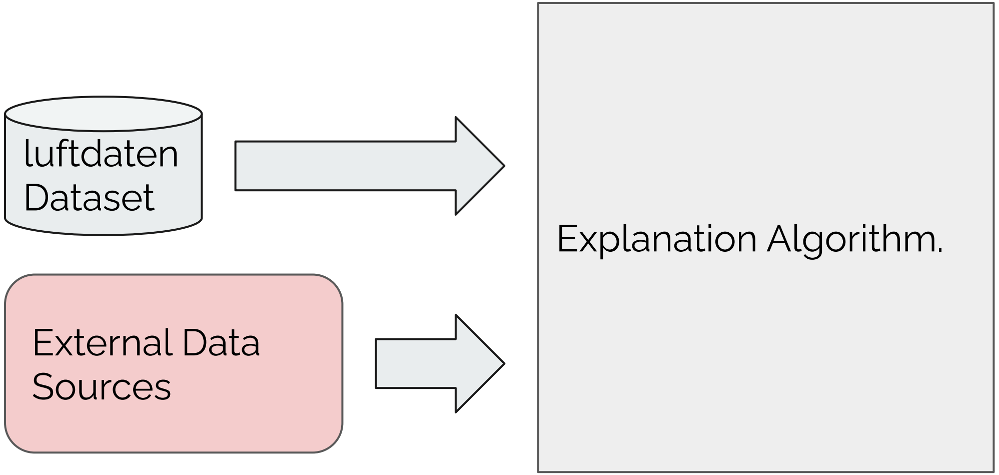


Sensor_type	Pollution
SDS011	35.07
SDS011	38.10
SDS011	1420.42

Sensor_type	Location	Pollution
SDS011	Tiergarten	35.07
SDS011	Tiergarten	38.10
SDS011	Tv Tower	1420.42

BTW Data Science Challenge - Our Proposal

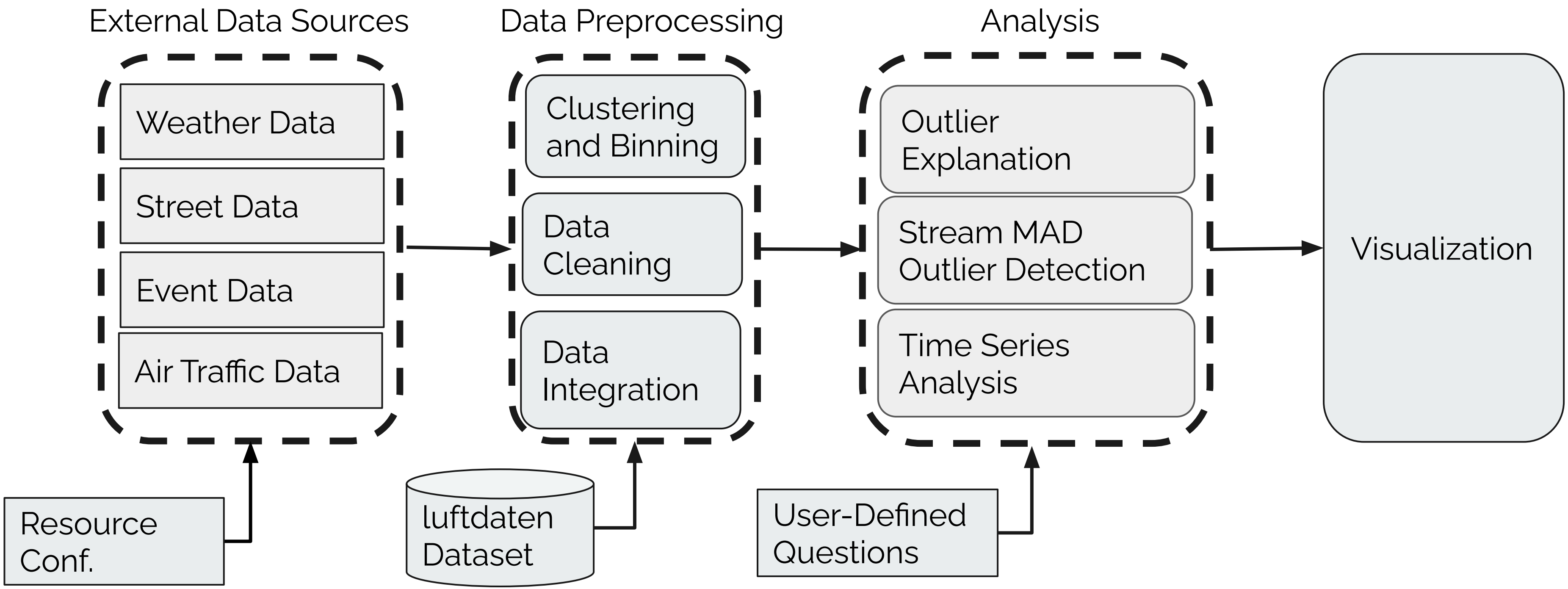
- Enriching the main dataset (Luftdaten) with extra information
- Adding features that correlate with air pollution



External Data Sources

- Air traffic data
  - Airplanes’ route
- Event data
- Weather data
  - Wind (speed and direction)/Temperature/Precipitation
- Openstreetmap data
  - Number of crossroads and streets/Train stations

System Architecture



External Data Sources

- Weather Data
- Street Data
- Event Data
- Air Traffic Data

Data Preprocessing

- Clustering and Binning
- Data Cleaning
- Data Integration

Analysis

- Outlier Explanation
- Stream MAD Outlier Detection
- Time Series Analysis

Visualization

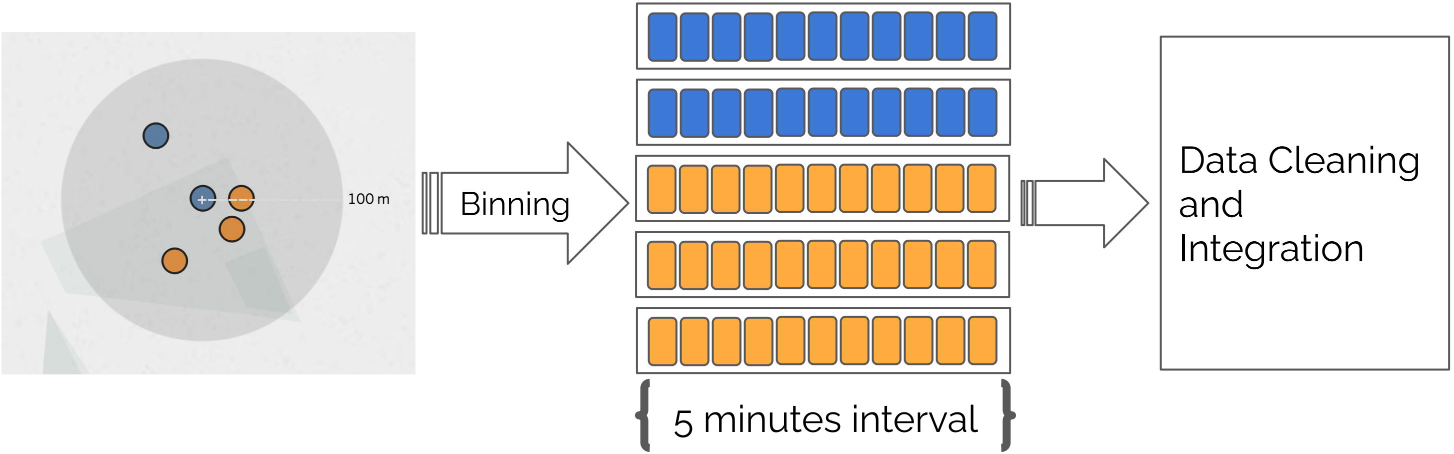
Resource Conf.

luftdaten Dataset

User-Defined Questions

Clustering and Binning

- Spatial: clustering, 100-meter radius
- Temporal: binning, 5 minute-interval



Binning

5 minutes interval

Data Cleaning and Integration

Data Cleaning

- Wrong readings - malfunctioning sensors / network
- Deviating readings - outliers within the cluster / time slot

TimeStamp	P1
11:17:31	3.5
11:17:59	1.9
11:18:26	100012.7
11:20:44	3.2
11:21:58	2.4

Observation error

Data Integration

Time	P1
11:15:31	3.5
11:16:59	2.5
11:17:26	3.0
11:18:12	3.1
11:19:00	2.9

Time	Temp.
11:16:06	18.1
11:18:44	18.2

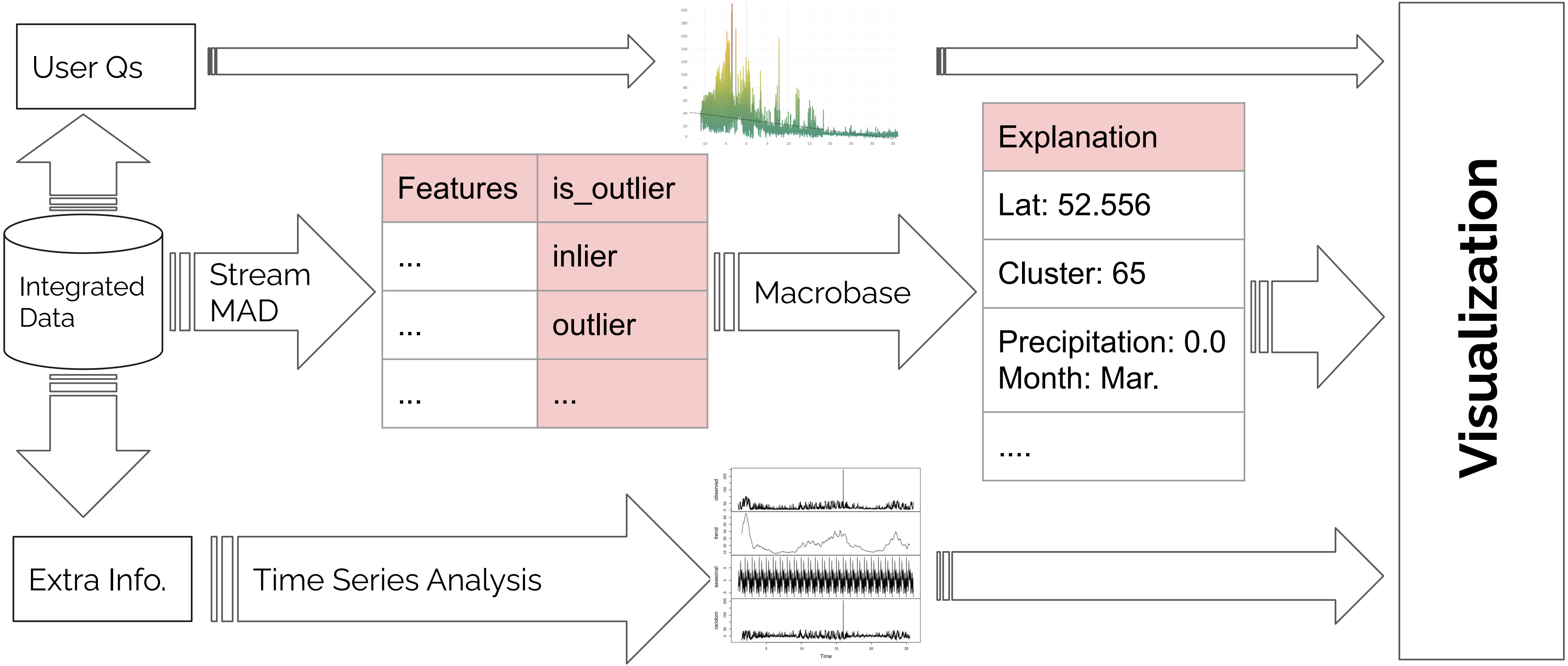
Time	Prec.
11:15:18	0.2
11:17:55	0.1
11:19:26	0.1

Time	Humid.
11:19:01	60%

Time	Wind	Degree
11:15:09	1.2	240
11:15:19	1.2	240
11:19:22	1.3	250

Time	P1	Temp.	Prec.	Humid.	Wind	Degree
11: [15 - 20]	3.0	18.15	0.1	60%	1.2	240

Analysis and Visualization



User Qs

Integrated Data

Stream MAD

Features	is_outlier
...	inlier
...	outlier
...	...

Macrobase

Explanation
Lat: 52.556
Cluster: 65
Precipitation: 0.0 Month: Mar.
....

Extra Info.

Time Series Analysis

Visualization

Results Based on External Data Sources

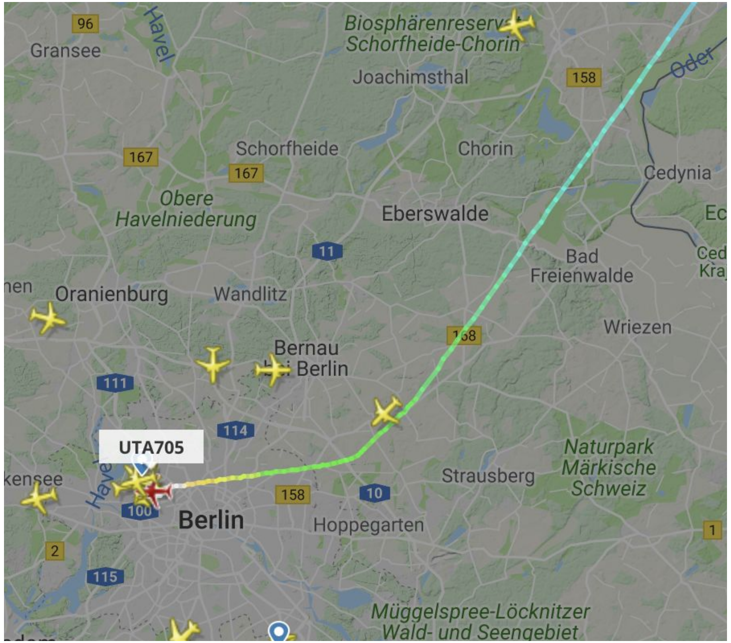
- Air traffic data
  - How does air traffic affect particulate matter pollution?
- Event data
  - Are there events that lead to short-term particulate matter pollution?
- Weather data
  - What is the correlation between weather data and air quality?
- Openstreetmap data
  - Do crossroads/roads/stations/diesel bans affect air pollution?

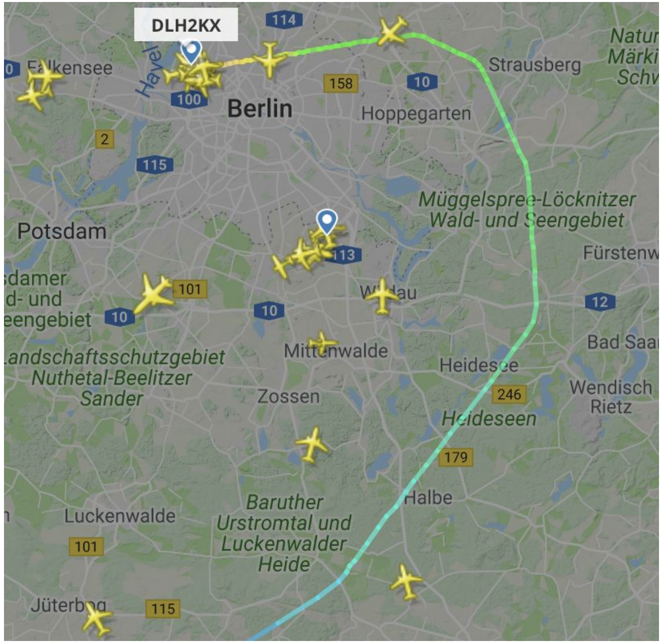
Berlin

Results (Air Traffic)

How Air Traffic Affects air quality?

Explanation: Latitude: 52.556 (TXL Airport)

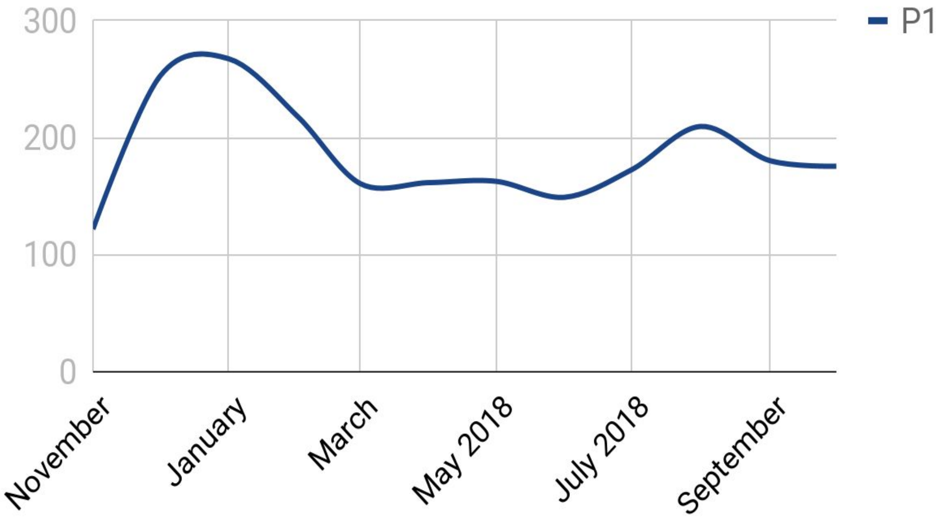




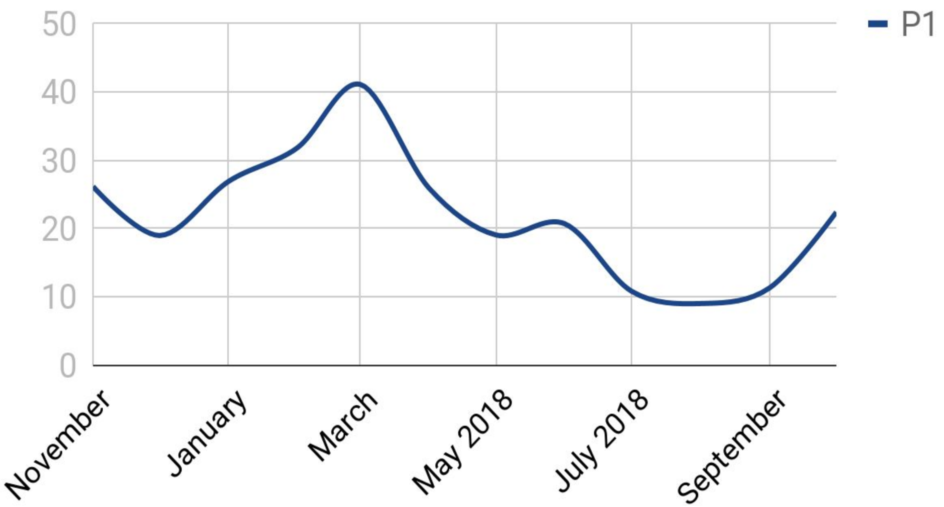
How Air Traffic Affects air quality?

Explanation: Latitude: 52.556 (TXL Airport)

Sensors with lat = 52.556



Other sensors

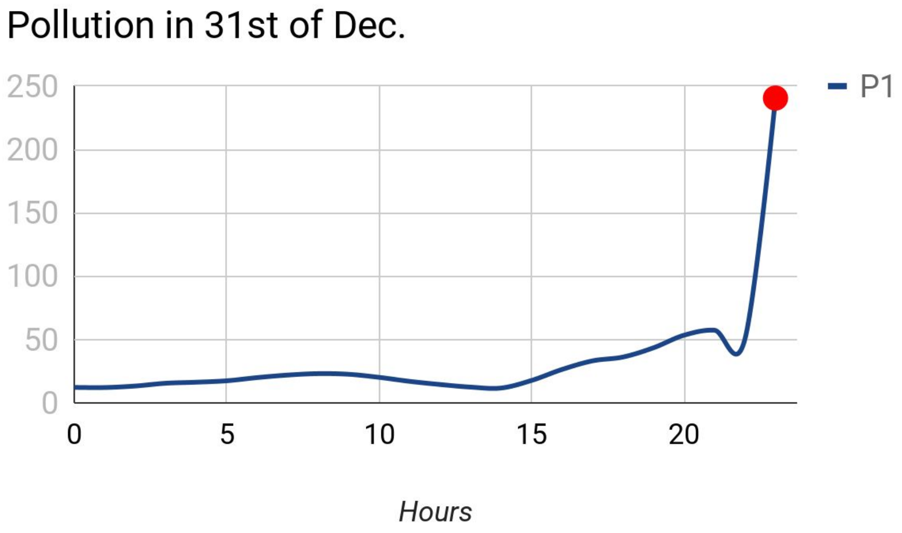


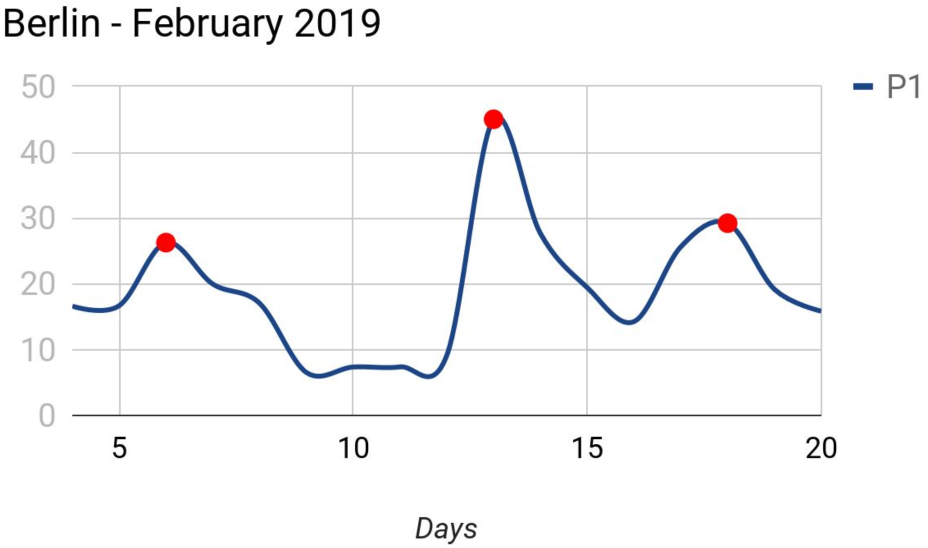
Results (Events)

How Events Play a Role in Pollution?

New Year's Eve

Berlin International Film Festival

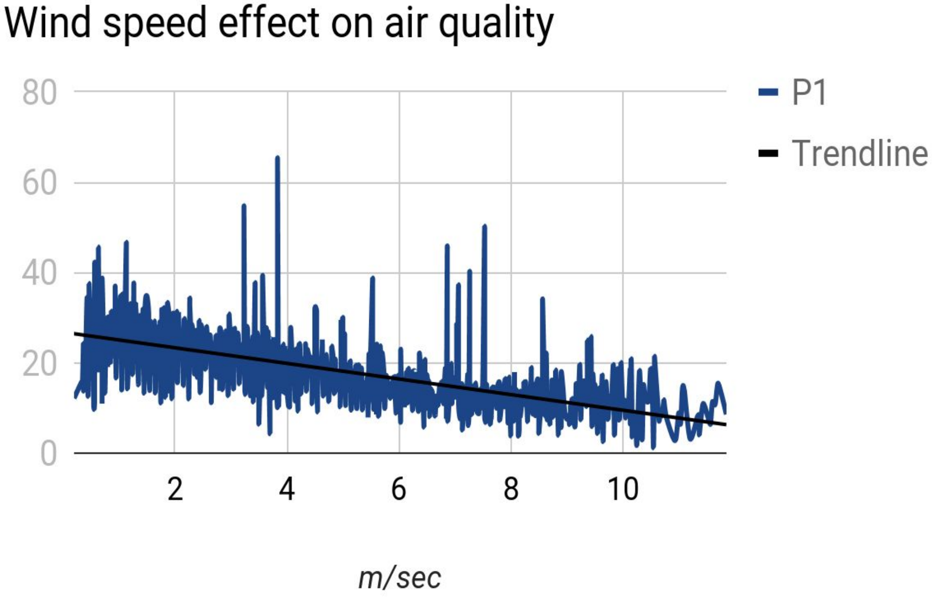


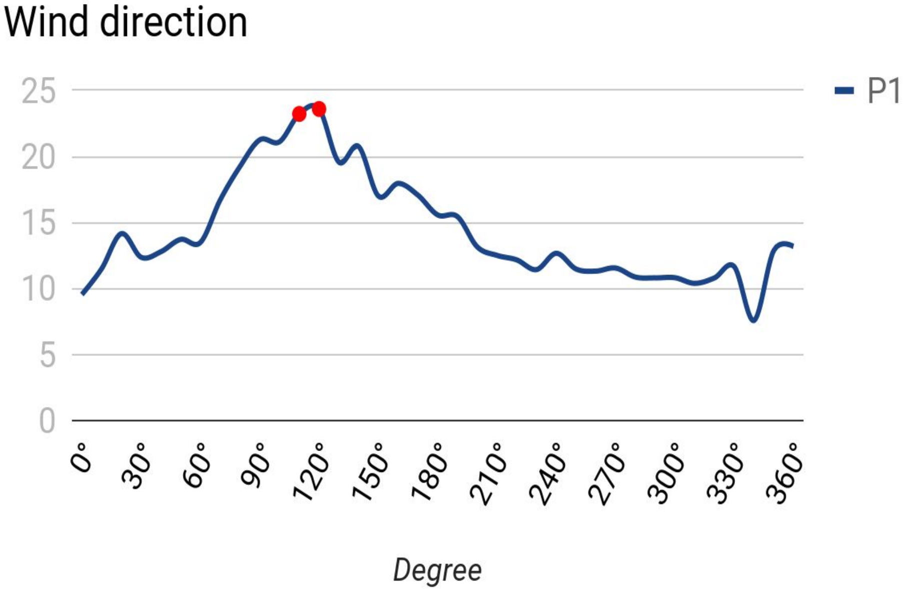


Results (Weather)

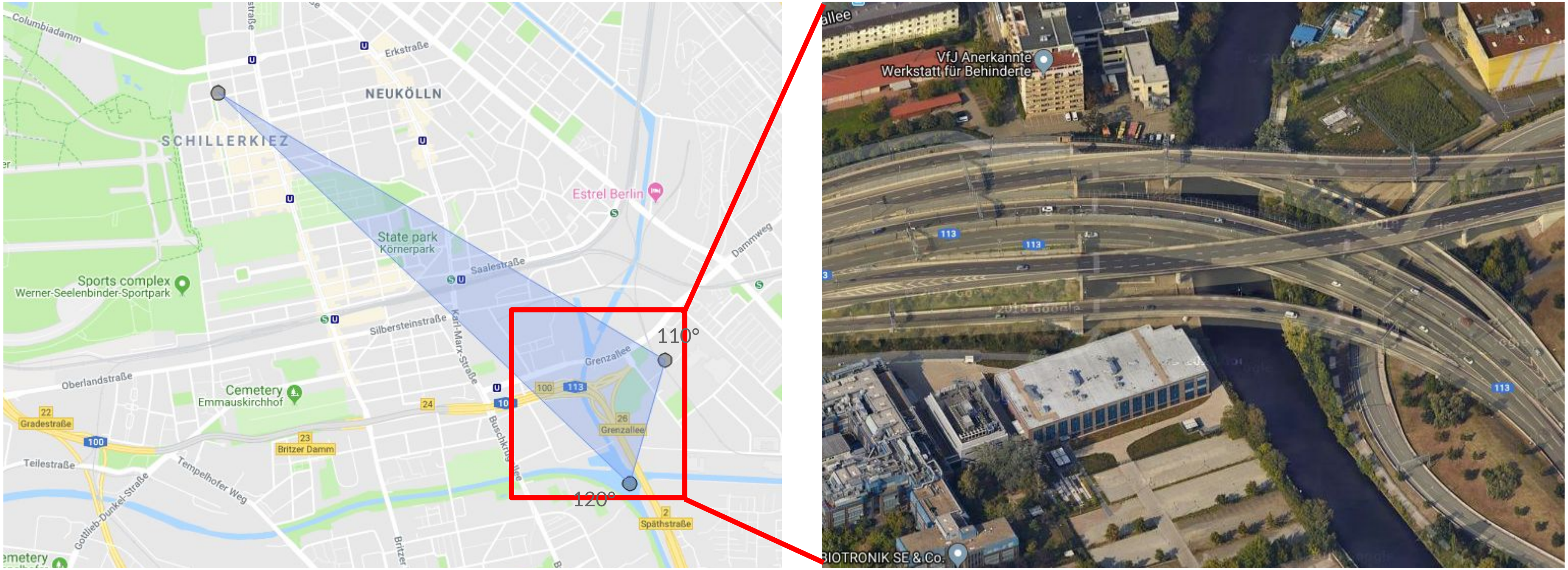
How Does The Weather Affect Air Pollution?

Explanation: Wind degree (cluster 104): 110 - 120





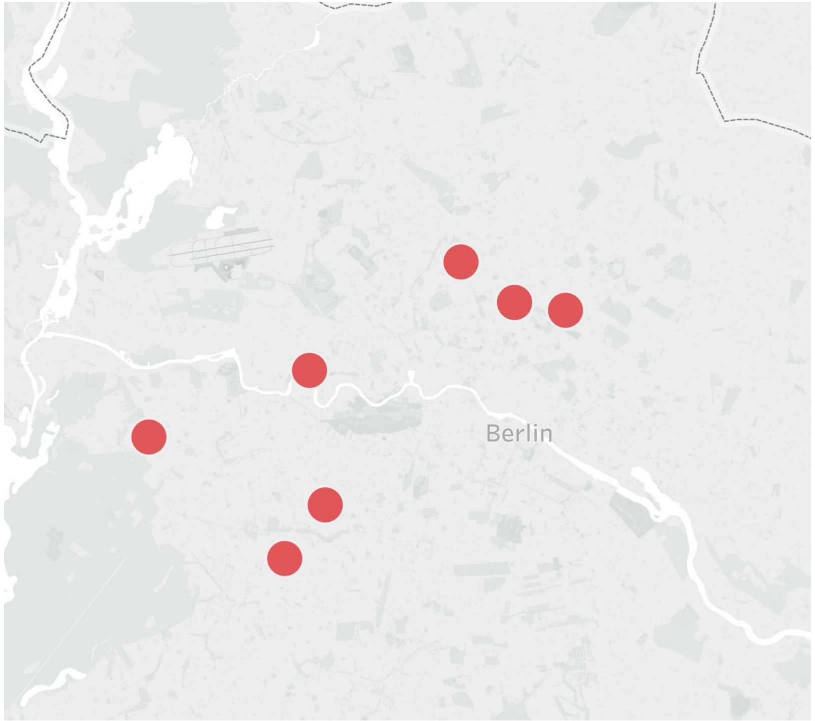
How Weather Data Affect Air Pollution?

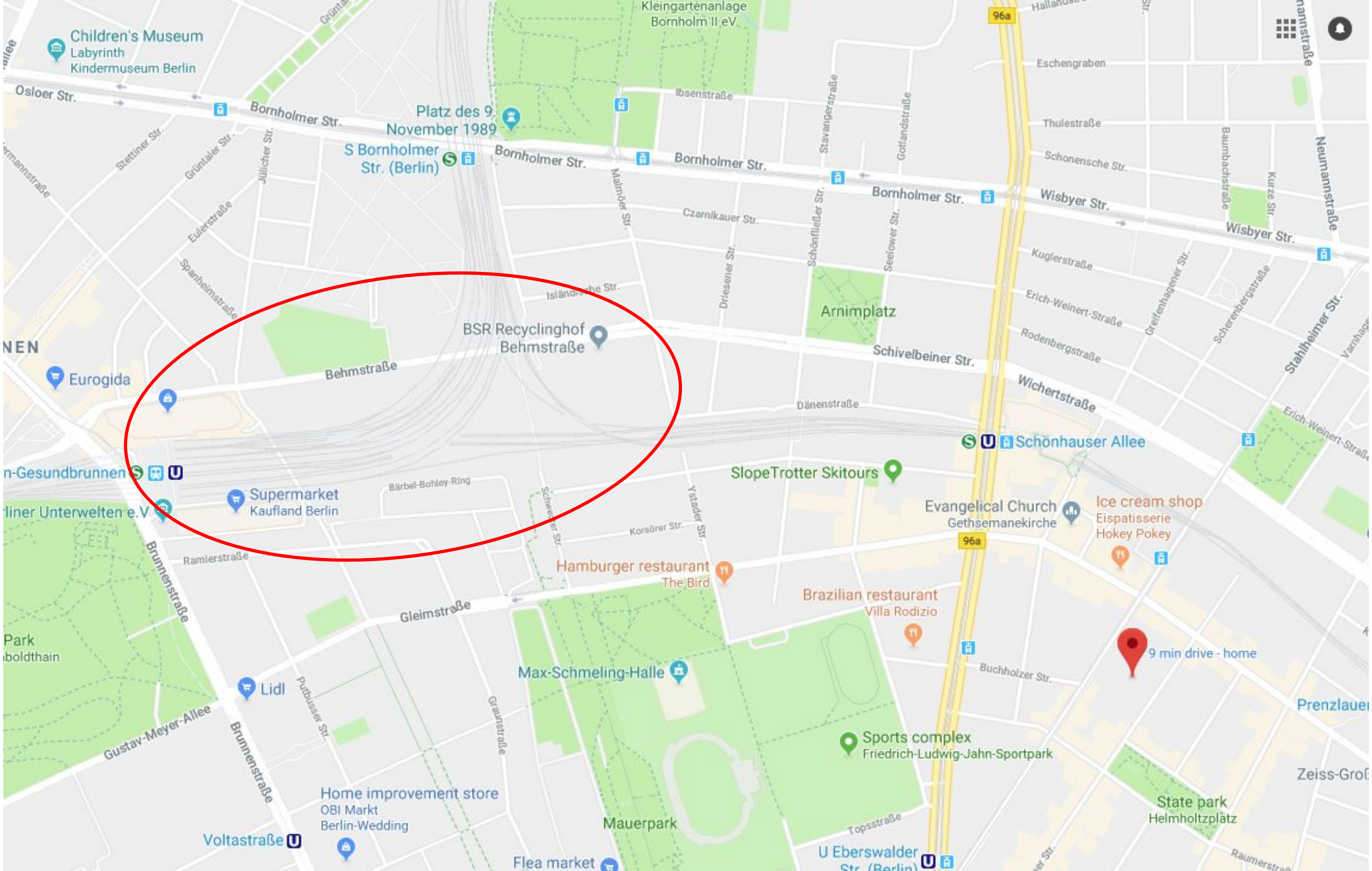


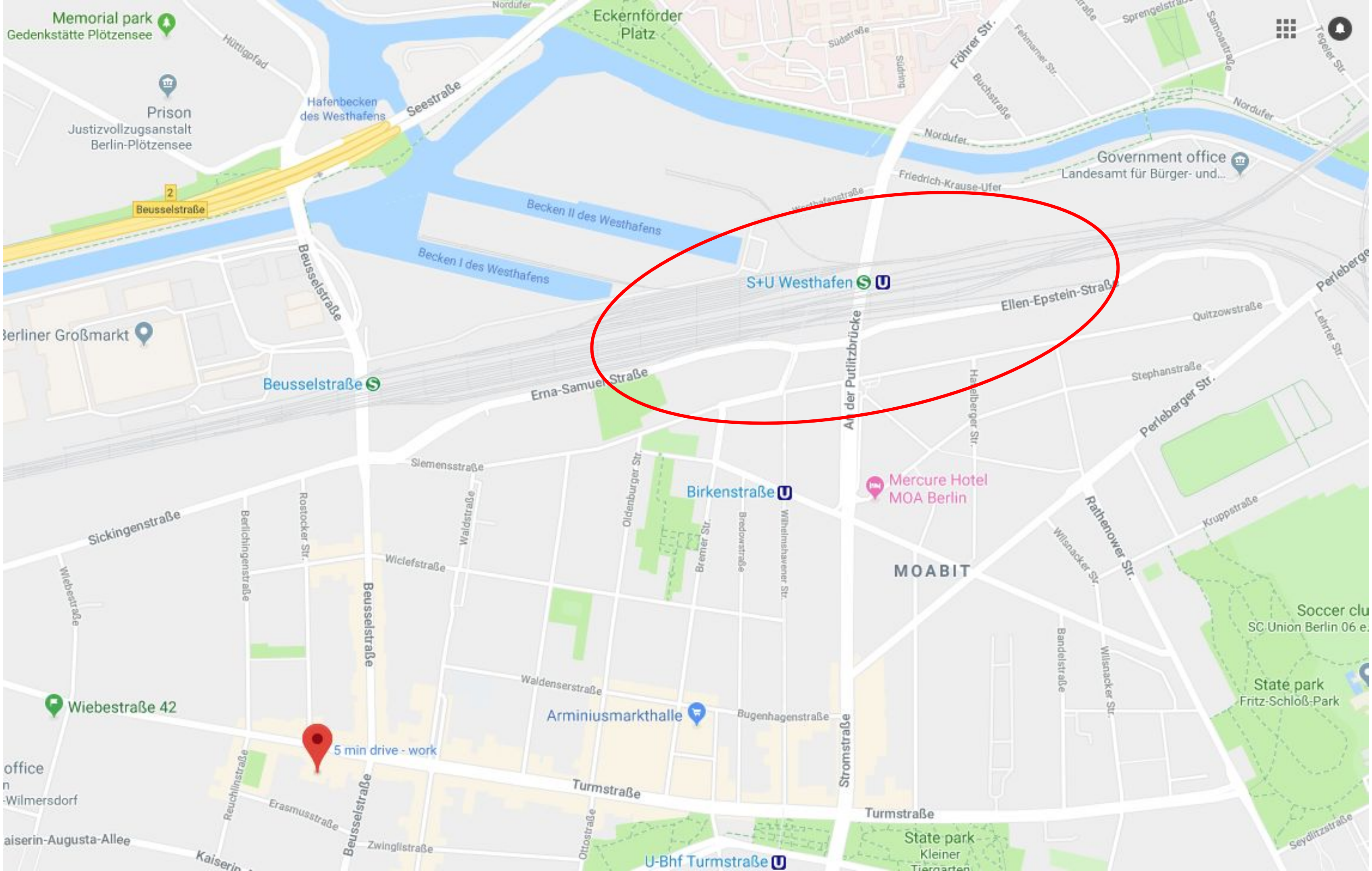
Results (OpenStreetMap)

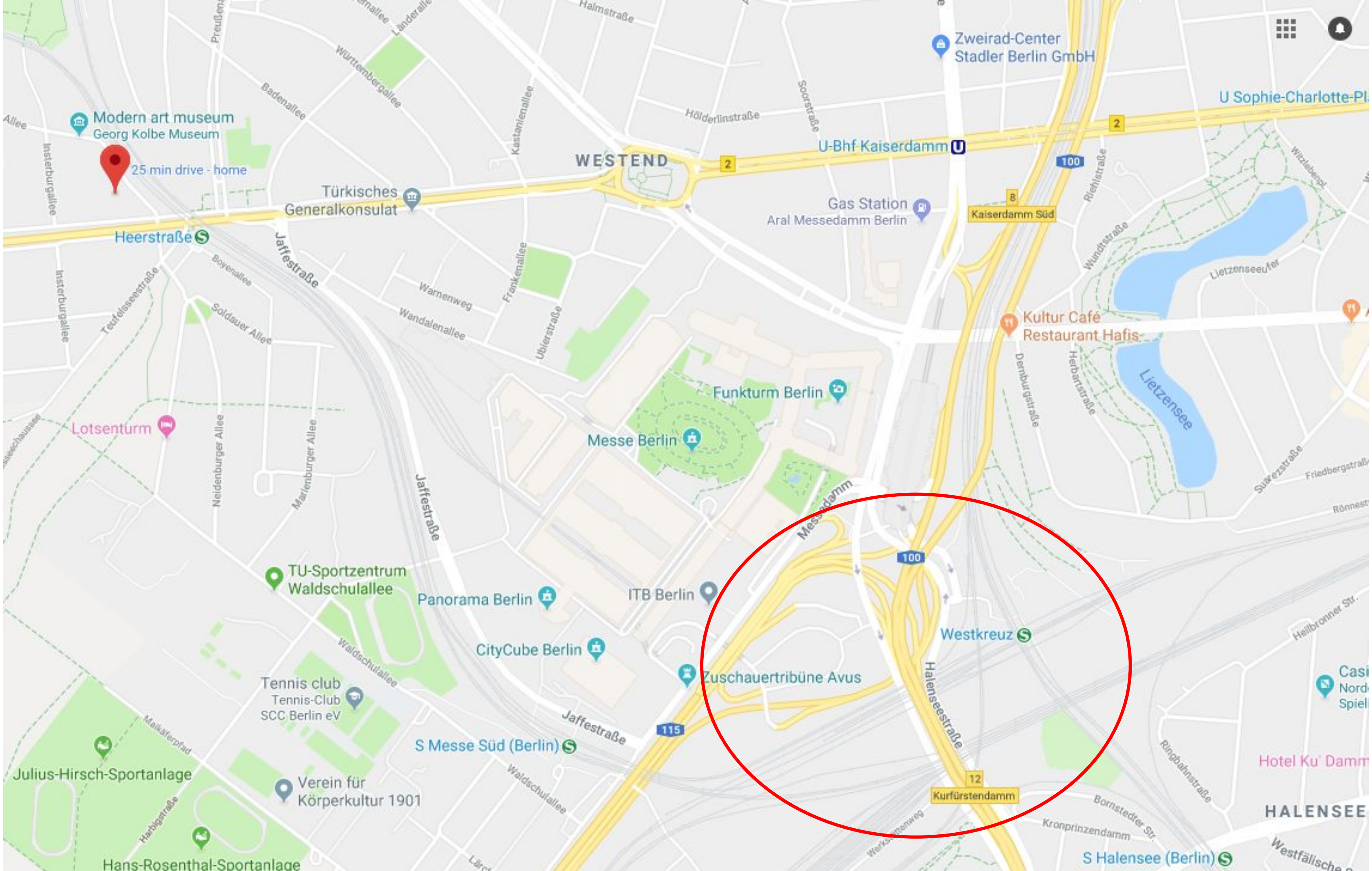
How Roads and Stations Affect Air Quality?

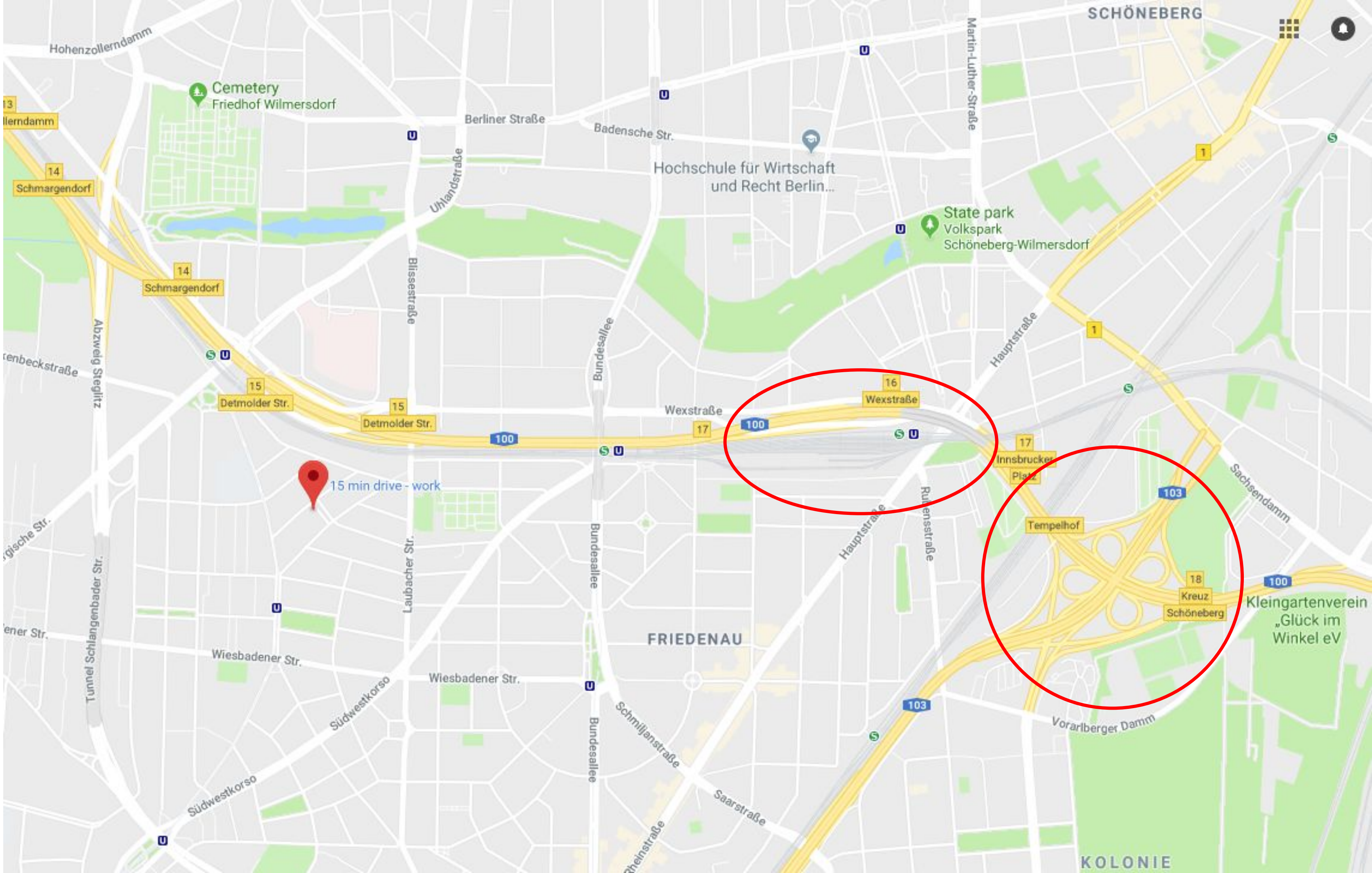
- The most polluted points are close to Ring or main S-Bahn stations in Berlin




Children's Museum
Labyrinth
Kindermuseum Berlin
Kleingartenanlage
Bornholm II eV
96a
Eschengraben
Osloer Str.
Bornholmer Str.
Platz des 9.
November 1989
Ibsenstraße
Thulestraße
S Bornholmer
Str. (Berlin)
Bornholmer Str.
Bornholmer Str.
Schonensche Str.
Bornholmer Str.
Wisbyer Str.
Wisbyer Str.
Stettiner Str.
Grüntaler Str.
Jülicher Str.
Malmöer Str.
Czarnikauer Str.
Stavangerstraße
Gotlandstraße
Baumbachstraße
Kurze Str.
Neumannstraße
Eulerstraße
Spanheimstraße
Isländische Str.
Driesener Str.
Schönfließer Str.
Seelower Str.
Kuglerstraße
Erich-Weinert-Straße
Greifenhagener Str.
Scherenbergstraße
Arnimplatz
BSR Recyclinghof
Behmstraße
Schivelbeiner Str.
Rodenbergstraße
Stahlheimer Str.
NEN
Eurogida
Behmstraße
Dänenstraße
Wichertstraße
Erich-Weinert-Straße
Schönhauser Allee
n-Gesundbrunnen
SlopeTrotter Skitours
Supermarket
Kaufland Berlin
Bärbel-Bohley-Ring
Schwedter Str.
Ystader Str.
Evangelical Church
Gethsemanekirche
Ice cream shop
Eispatisserie
Hokey Pokey
liner Unterwelten e.V
Korsörer Str.
96a
Ramlerstraße
Brunnenstraße
Hamburger restaurant
The Bird
Brazilian restaurant
Villa Rodizio
Gleimstraße
Park
boldthain
9 min drive - home
Buchholzer Str.
Max-Schmeling-Halle
Prenzlauer
Lidl
Putbusser Str.
Graunstraße
Gustav-Meyer-Allee
Sports complex
Friedrich-Ludwig-Jahn-Sportpark
Zeiss-Groß
Brunnenstraße
Home improvement store
OBI Markt
Berlin-Wedding
State park
Helmholtzplatz
Mauerpark
Topsstraße
Voltastraße
Flea market
U Eberswalder
Str. (Berlin)
Raumerstraße

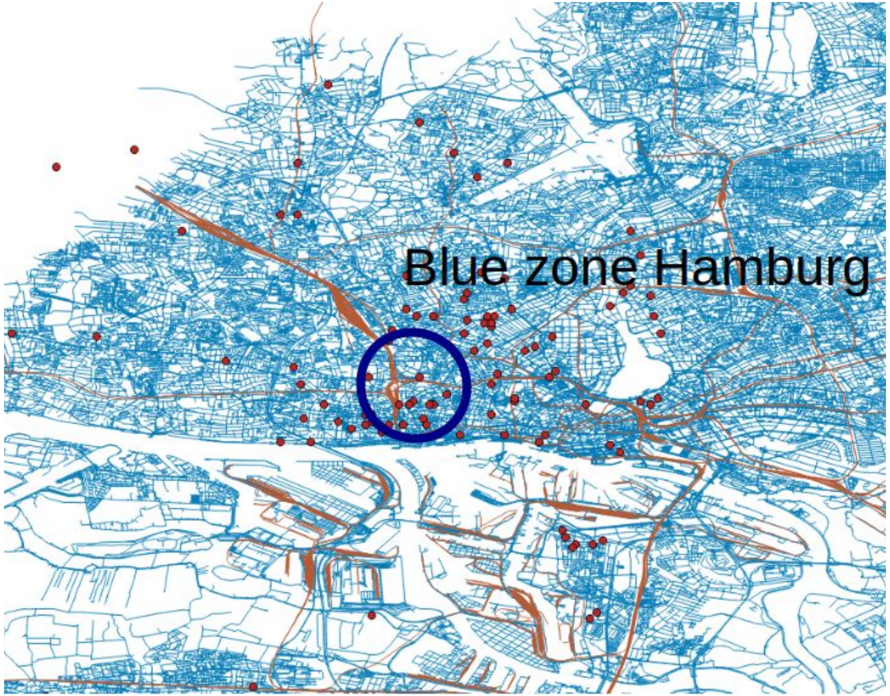

Memorial park
Gedenkstätte Plötzensee
Prison
Justizvollzugsanstalt
Berlin-Plötzensee
Hüttigpfad
Hafenbecken des Westhafens
Seestraße
Eckernförder Platz
Föhrer Str.
Beusselstraße
Becken II des Westhafens
Becken I des Westhafens
Nordufer
Friedrich-Krause-Ufer
Government office
Landesamt für Bürger- und...
Westhafenstraße
S+U Westhafen
Ellen-Epstein-Straße
Quitzowstraße
Perleberger Str.
Berliner Großmarkt
Beusselstraße
Erna-Samuel-Straße
An der Putlitzbrücke
Stephanstraße
Siemensstraße
Birkenstraße
Mercure Hotel MOA Berlin
MOABIT
Sickingenstraße
Wiclefstraße
Rathenower Str.
Kruppstraße
Soccer clu
SC Union Berlin 06 e.
Waldenserstraße
Arminiusmarkthalle
Bugenhagenstraße
Stromstraße
State park
Fritz-Schloß-Park
Wiebestraße 42
5 min drive - work
Turmstraße
office
-Wilmersdorf
Kaiserin-Augusta-Allee
Erasmusstraße
Zwinglistraße
U-Bhf Turmstraße
State park
Kleiner Tiergarten


Modern art museum
Georg Kolbe Museum
25 min drive - home
Türkisches Generalkonsulat
Heerstraße
WESTEND
U-Bhf Kaiserdamm
Zweirad-Center Stadler Berlin GmbH
U Sophie-Charlotte-Pl
Gas Station
Aral Messedamm Berlin
Kaiserdamm Süd
Kultur Café Restaurant Hafis
Lietzensee
Lietzenseeufer
Funkturm Berlin
Messe Berlin
Lotsenturm
TU-Sportzentrum Waldschulallee
Panorama Berlin
ITB Berlin
CityCube Berlin
Zuschauertribüne Avus
Westkreuz
Messedamm
Halenseestraße
Kurfürstendamm
Tennis club
Tennis-Club SCC Berlin eV
S Messe Süd (Berlin)
Julius-Hirsch-Sportanlage
Verein für Körperkultur 1901
Hans-Rosenthal-Sportanlage
Hotel Ku' Damm
HALENSEE
S Halensee (Berlin)
Jaffestraße
Waldschulallee
Württembergallee
Badenallee
Hölderlinstraße
Soorstraße
Halmstraße
Kastanienallee
Frankenallee
Ubierstraße
Warnenweg
Wandalenallee
Boyenallee
Soldauer Allee
Teufelsseestraße
Insterburgallee
Neidenburger Allee
Marienburger Allee
Dernburgstraße
Herbartstraße
Wundtstraße
Riehlstraße
Witzlebenpl
Suarezstraße
Friedbergstraße
Rönnestr
Heilbronner Str.
Ringbahnstraße
Bornstedter Str.
Kronprinzendamm
Westfälische
Malkaferpfad
Harbigstraße
Werkstättenweg
Casi Nord Spiel
2
100
8
115
12


SCHÖNEBERG
Hohenzollerndamm
Cemetery
Friedhof Wilmersdorf
Berliner Straße
Badensche Str.
Hochschule für Wirtschaft und Recht Berlin...
Martin-Luther-Straße
State park
Volkspark Schöneberg-Wilmersdorf
Uhlandstraße
Blissestraße
Bundesallee
Hauptstraße
Schmargendorf
Detmolder Str.
Wexstraße
Innsbrucker Platz
Tempelhof
Kreuz Schöneberg
Sachsendamm
Kleingartenverein „Glück im Winkel eV
15 min drive - work
Laubacher Str.
Rubensstraße
FRIEDENAU
Wiesbadener Str.
Südwestkorso
Schmiljanstraße
Saarstraße
Rheinstraße
Vorarlberger Damm
Tunnel Schlangenbader Str.
Abzweig Steglitz
KOLONIE
100
103
1
14
15
16
17
18

How Do Diesel Bans Affect Pollution?

- 10% local decrease in pollution
- No global impact
- Berlin diesel ban (1st of April 2019)
- Affected streets: e.g. Friedrichstraße
- Due to the locality, diesel bans should address the most polluted roads



Conclusion

- Luftdaten is limited by its own
- Current solutions are not effective due to the dearth of information
- Idea of enriching main dataset with external data sources
- Detected causes of pollution: e.g. public events, weather, air traffic, and etc.
- We built a general pollution explanation system that can be applied on every city

Potential Future Directions

- Exploration of pollution causes
    a. Explore more dimensions, e.g., more cities, more influencing factors,
    b. Use other ML or statistical methods
- Research direction: automated selection additional sources
    a. What are effective heuristics to choose datasets that improve explanation experience?
    b. What types of indexing mechanisms are necessary to make this process efficient?