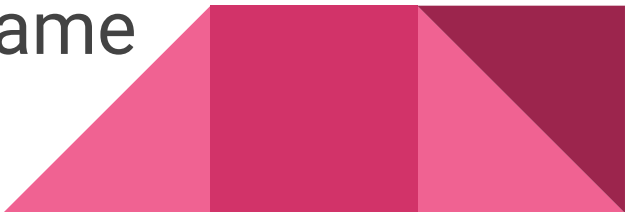




Matthew Manuel, Thien Hua, Leon Lin, Ramon Santos

Intro: Deliverable

- Dataset was taken from airsofia, a website that tracks air quality
 - Was made in response to Bulgaria for failing to implement significant anti-pollution measures
 - Gives location coordinates, air quality (pressure, temperature, humidity) and timeframe
- 

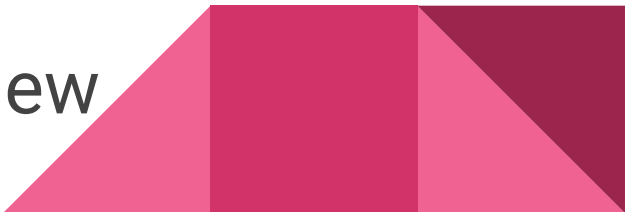
Background

- 2013 Sofia was singled out for a study by the European Environment Agency
- This was due to Sofia's presence of several air pollutants in the atmosphere
- Caused by gas and carbon monoxide from burning wood and coal.

Background (cont)

- Natural factors to air pollution
 - Geographic location
 - Stagnation of air
- Human activity is the main source
 - Transportation with the use of diesel
 - Old heating system

Background Cont.

- Two groundbreaking decisions made
 - Jnecek Case
 - ClientEarth Case
 - These decisions allowed
 - The Right of Standing
 - The Right of Legal Remedy
 - The Right to a Substantive Review
- 

Related Work

- European Environment Agency in 2013
- World Health Organization in 2016
- Both studies focus on the rampant spread of air pollution
- We used big data obtained from a data source from an agency that tracks down air quality.
- Focus on temperature of a specific region and noting changes in values over time

Specifications

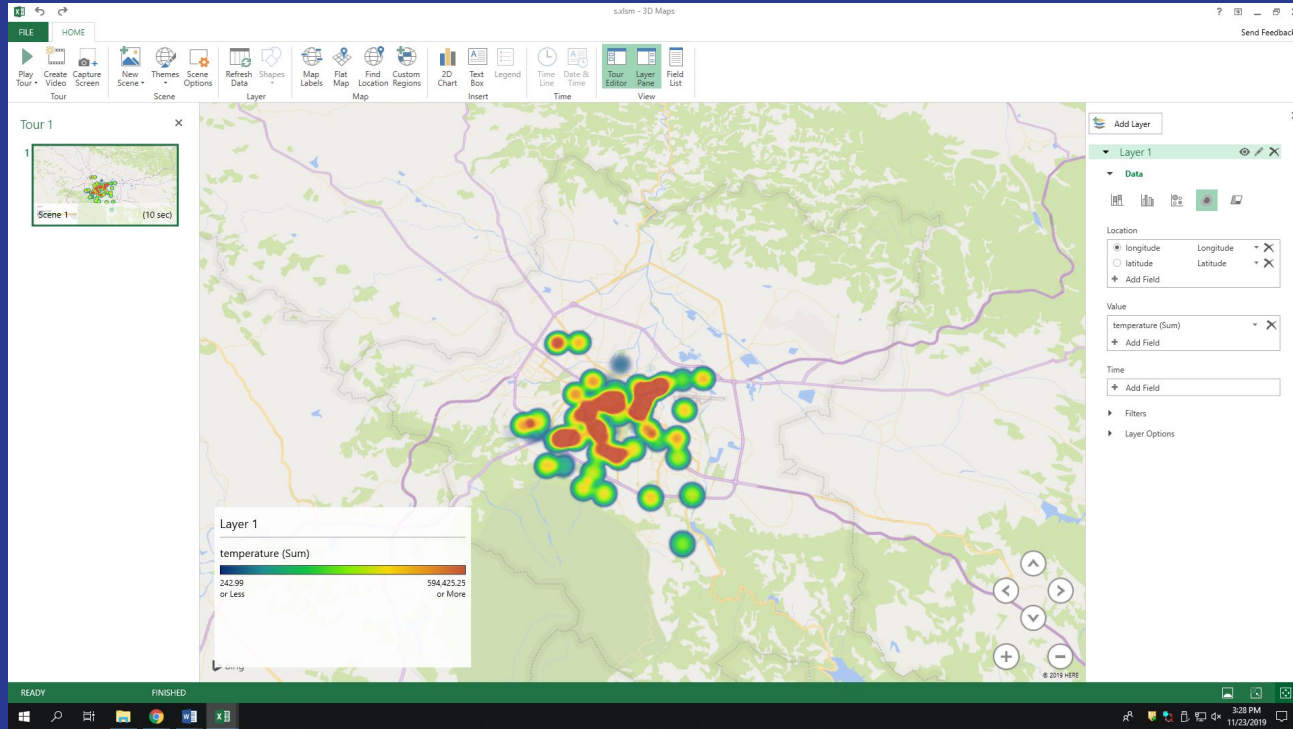
Cluster version 6.1 (includes Apache Spark 2.4.4, Scala 2.11)

Number of nodes: 1

Memory size: 2.7 GB

Cpu speed: 2.30 GHz

Our Work



Temperatures in a specific location due to air quality

Data Flow Chart

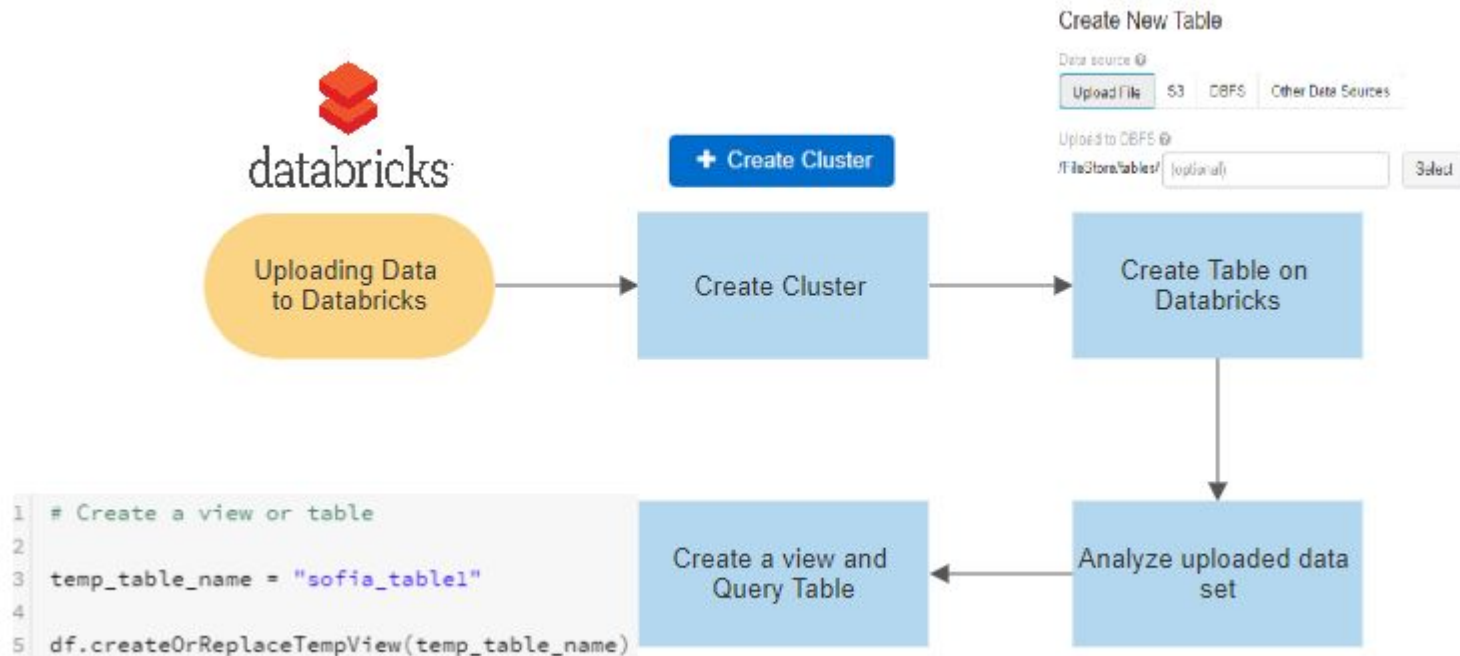


Table: sofia_table

Schema:

| col_name | data_type | comment |
|----------|-----------|---------|
| _c0 | string | null |
| _c1 | string | null |
| _c2 | string | null |
| _c3 | string | null |
| _c4 | string | null |
| _c5 | string | null |
| _c6 | string | null |
| _c7 | string | null |

Sample Data:

| _c0 | _c1 | _c2 | _c3 | _c4 | _c5 | _c6 | _c7 |
|------|-----------|----------|--------------------|--------------------|---------------------|------|------|
| null | sensor_id | location | lat | lon | timestamp | P1 | P2 |
| 56 | 7775 | 3932 | 42.669 | 23.346 | 2018-04-01T10:00:38 | 3.37 | 1.57 |
| 63 | 7917 | 4004 | 42.625 | 23.385 | 2018-04-01T10:00:38 | 4.07 | 1.4 |
| 103 | 3759 | 1895 | 42.68600000000001 | 23.250999999999998 | 2018-04-01T10:00:39 | 3.77 | 1.37 |
| 143 | 10461 | 5278 | 42.656000000000006 | 23.373 | 2018-04-01T10:00:46 | 1.67 | 1.53 |
| 194 | 5293 | 2669 | 42.693999999999996 | 23.340999999999998 | 2018-04-01T10:00:46 | 4.17 | 2.2 |
| 219 | 8673 | 4374 | 42.607 | 23.386999999999997 | 2018-04-01T10:00:46 | 5.9 | 1.3 |
| 220 | 2888 | 1453 | 42.708 | 23.283 | 2018-04-01T10:00:46 | 6.07 | 3.4 |
| 260 | 6029 | 3043 | 42.681999999999995 | 23.375999999999998 | 2018-04-01T10:00:48 | 4.43 | 1.4 |
| 273 | 11076 | 5591 | 42.626000000000005 | 23.38 | 2018-04-01T10:00:50 | 2.03 | 1.03 |
| 305 | 10945 | 5525 | 42.648 | 23.368000000000002 | 2018-04-01T10:00:51 | 2.58 | 1.05 |
| 309 | 6527 | 3297 | 42.663999999999994 | 23.255 | 2018-04-01T10:00:51 | 2.9 | 1.4 |

Data Url

Data Size: 14 GB

<https://www.kaggle.com/hmavrodiev/sofia-air-quality-dataset>