Scientific Journal of Astana IT University
ISSN (P): 2707-9031
ISSN (E): 2707-904X

DOI: 10.37943/AITU.2020.94.34.004

Zh. Sarsenova
Master of Technical Sciences, Senior lecturer of ICT Educational Program
zhibek.sarsenova@astanait.edu.kz, orcid.org/0000-0002-7780-8070
Astana IT University, Kazakhstan

Sh. Saimassayeva
Master of Technical Sciences, Senior lecturer of ICT Educational Program
sholpan.saimassayeva@astanait.edu.kz, orcid.org/0000-0003-4482-8320
Astana IT University, Kazakhstan

A. Smaiyl
Master of Technical Sciences, Senior lecturer of ICT Educational Program
assel.smaiyl@astanait.edu.kz, orcid.org/0000-0002-6215-932X
Astana IT University, Kazakhstan

DETERMINATION OF THE MOST POLLUTED ATMOSPHERIC AIR POLLUTION CATEGORY OF ALMATY CITY

Abstract: This article discusses the impact of suspended particles on human health, by providing small definitions of PM2.5, including how they appear, what particles they consist of, and how they harm the respiratory and circulatory systems. In addition, the analysis of the pollution level of the city of Almaty for the last three years from March 22, 2017, to October 6, 2020, with categorical intermediate values of suspended particles was carried out. Careful work was done with the SCV file such as data was skipped in the cells, that is, there were empty values; translation to a single data type; filling in empty cells. It also considers making a decision on the six categories provided to identify the average meeting categories.

The authors identified specific categories based on digital readings of values received from sensors, where each category has its own verbal values that are understandable for each person. The indicator displays with weights as a graph for a specific Seifullin-Dulatov sensor location with categorical and without categorical division. Then for each intersection or location of the sensors is shown in a table. It is also revealed which level or category is the rarest among the others and the most common category as well.

Keywords: air pollution, environmental monitoring, data analysis.

Introduction

Atmospheric air is one of the vital components of the environment. It performs many functions, such as biological, transport, and production. Therefore, air pollution is one of the significant health risks associated with the environment.

Problems of atmospheric protection are a vast area of research. They include not only General problems of chemical technology, mechanical engineering, and meteorology but also issues that can be addressed by narrow specialists, such as mathematicians, physicists, programmers, analysts, etc. Atmospheric pollution factors may be related to natural processes and human activities. All sources of pollution are divided into natural and artificial. Biological pollutants are those of mineral and plant origin that enter the atmosphere because of volcanic eruptions and forest fires. In addition, natural air pollutants are dust, plant pollen, and so on. Artificial factors of air pollution are divided into transport, industrial, and household.

DOI: 10.37943/AITU.2020.94.34.004

In Kazakhstan, the primary sources of air pollution are large industrial enterprises, heating systems, and vehicles [1]. According to the Ministry of ecology, geology, and natural resources, since 2017, there has been a tendency to increase the emissions claimed by nature users by 52 thousand tons, or 1.2%. At the same time, over the past five years, there has been a tendency to reduce them by 156 thousand tons or 2.2%. That is, if in 2015, the limit was 4.4 million tons of emissions; in 2019 it is already 4.3 million tons [2].

In this article [3], an info logical model is constructed that reflects the real world in a human-friendly concept, completely independent of the parameters of the semantic data model storage environment. This model solves the current problems of monitoring industrial emissions that affect air pollution. Taking into account all the factors discussed in this article and in the articles proposed by professors working in this direction, a data analysis has been carried out by us.

Ecological factors

Seven million people worldwide die from the effects of air pollution every year. Air pollution is the cause of respiratory, infectious, and heart diseases, stroke, lung cancer, and many others. Children who inhale dirty air are increasingly suffering from asthma, chronic lung failure, growth retardation, diabetes, childhood obesity, and mental retardation.

Like many other countries, Kazakhstan is firmly established in the list of countries with extremely negative environmental trends. Many residents of our country suffer from diseases of the upper respiratory tract and cancer. Furthermore, because of air pollution in Kazakhstan, the function of blood vessels in the lungs worsens, and the body's immune system is disrupted. Pollution in Kazakhstan occurs through the release of lead, benzapyrene, and formaldehyde into the air. The main source of air pollution is transport, which accounts for 70% of the total emission of harmful substances into the air basin. Moreover, it becomes more and more every year.

Therefore, according to Kazhydromet data, there are traces of active chemical compounds of hazard class 3, 4 (nitrogen oxide, carbon, sulfur) in the air, less often compounds of hazard class 1-2 [4].

The article mentions a lot about PM2.5, which is an air pollutant, that includes both solid micro particles and the smallest droplets of liquids. Both are approximately 10 nm to 2.5 microns in size. Other designations and names of PM2.5 particles: FSP (fine suspended particles), fine particles, fine particulate matter, fine suspended particles, fine dust.

All these particles and droplets smaller than 2.5 microns are suspended in the air. They are found both in the forest and on the sea, but it is in the city that they pose the greatest danger. First, there are usually a lot more of them in the city, and secondly, the chemical composition of fine aerosol in the city is more dangerous than in nature. It might be also noted, the composition of the PM2.5 aerosol and the parameters of individual particles may differ greatly in different cities [5].

Data Analysis

In the CSV file, data with 41 columns and 1291 entries and data starts from March 22, 2017, to October 6, 2020. The sensor is located in the location below at the intersection of Almaty streets. We use a Google Collaboratory to analyze this data. The Google Colaboratory is a free cloud service based on the Jupyter laptop that provides everything you need for machine learning right in the browser [6]. Fig. 1 shows the location of sensors and the default data type in the file.

Scientific Journal of Astana IT University
ISSN (P): 2707-9031
ISSN (E): 2707-904X

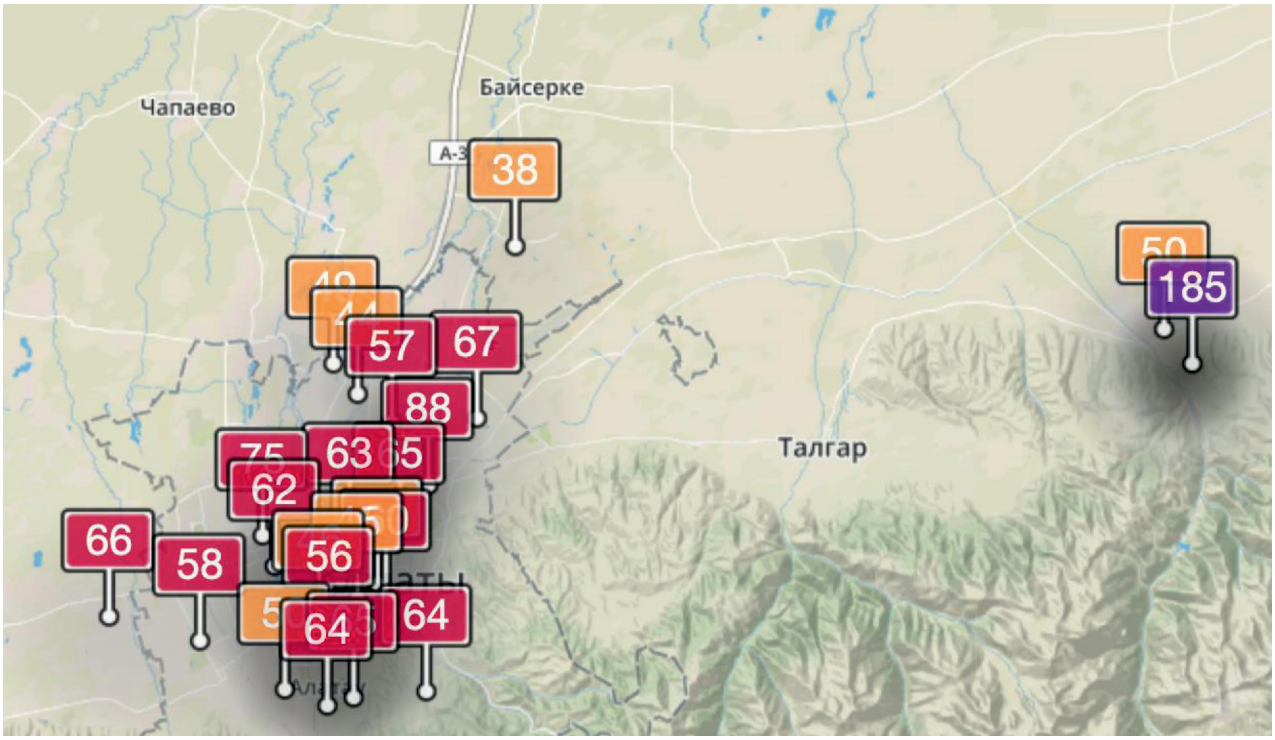

Fig.1. Locations of sensors on the map

As shown in Fig.1 above, the location of the sensors on the map is shown. In addition, the data type of each site is different and it is decided to change the data type to float type then convert it to one data type float. In our case, the last data Mukhanova_159 has an int data type. There are so many missing values. It means in those days sensors did not provide data, because of did not install sensors in that place. Table 1 shows the total count of missing values in each intersection of the street.

Table 1. Missing values

name	lat	lng	missing value
Seifullin-Dulatov	43.313	76.939	350
Al-Farabi-Markov	43.224	76.938	110
Abay-Tilendiyev	43.24	76.874	449
Gorky park	43.265	76.973	646
Tolebi-Baizakov	43.253	76.91	101
Rozybakiev - Baykadamov	43.214	76.893	72
microdistr. Kok Kainar	43.296	76.844	339
Ryskulov-Momyshuly	43.254	76.82	472
Ermensay	43.174	76.917	116
Tulebayev-Dzhambul	43.248	76.949	325
Residential complex Solnechnaya Dolina	43.189	76.868	263
Kamenskoe plateau	43.177	76.966	122
Furmanov-Tashkent	43.269	76.944	347
microdistrict Mamyr	43.216	76.848	465
Residential complex - ASYL Arman	43.231	76.754	1275
Residential complex- Zhana Kuat	43.397	77.027	244
Baganashyl	43.195	76.915	1134

DOI: 10.37943/AITU.2020.94.34.004

name	lat	lng	missing value
Kurgauldy	43.172	76.736	370
Voenniy gorodok	43.369	76.987	1022
Microdistr. Karasu	43.337	76.904	761
Satpayev-Lugansky	43.354	77.457	496
Raiymbek Rural District Karasay district	43.356	77.467	747
Shanyrak 2	43.177	76.943	1264
Energostroitel	43.241	76.96	583
Zhana Kairat	43.214	76.75	959
Almagul Microdistrict 24	43.297	76.852	1165
Akan Seri	43.299	76.792	1238
micr.Mamyr	43.312	77.001	1277
Miras 122	43.206	76.9	1124
Kurchatov 1	43.31	76.943	1289
Tatibekov	43.213	76.846	1052
KazGU Parking	43.1855	76.8716	1266
Kamyshinskaya	43.254	76.855	1172
Tulebaeva 82	43.2795	76.9677	1289
Asanbai Askarov 60	43.223	76.919	1131
3rd Ainabulak Microdistrict 107	43.3526	76.9993	1275
5th Tau Samal line	43.2576	76.9484	1280
Sagadat Nurmagambetov 330	49.628	72.986	1023
Timiryazev 17	49.628	73.046	1125
Microdistrict Karasu 2	43.178	76.87	1176
Mukanova 159	43.324	76.919	1291

After the defining count of missing values in each location to avoid, empty values and analysis show us precise values, replacing, all empty values with mean values will be a good decision.

All air pollution values are divided into 6 categories, the lower the level of air pollution in the air; the better it is for us. The Table 2 below shows the levels of PM 2.5 values.

Table 2. Levels of PM 2.5

Name of levels	Range of values PM 2.5
Low	0-12
Elevated	13-34
High	35-55
Very high	56-149
Dangerous	150-249
Poisonous	250-max value

These diagrams (Fig.3, Fig.4) show frequently occurring values of the PM 2.5 air pollution level from a sensor that is located in the Seifullin-Dulatov location.

Scientific Journal of Astana IT University
ISSN (P): 2707-9031
ISSN (E): 2707-904X

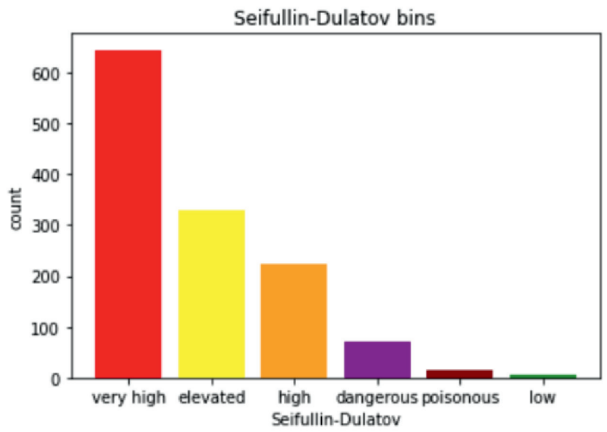

Fig. 3 The most frequent values in Seifullin-Dulatov street with category

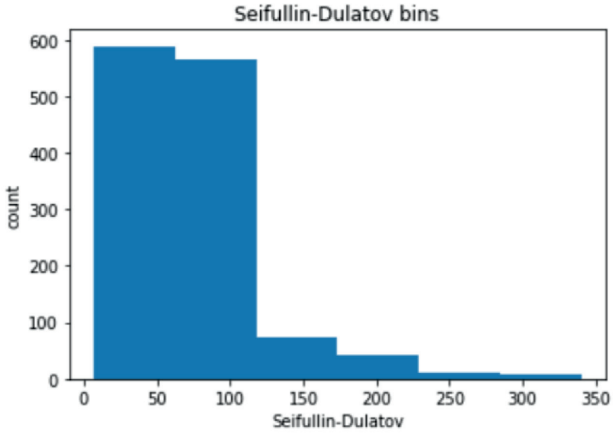

Fig. 4 The most frequent values in Seifullin-Dulatov streen without category

Table 3 shows the most frequent data values as a table in each locations.

Table 3. The most frequent values in each location

Location of sensors	Low	Elevated	High	Very high	Dangerous	Poisonous
Seifullin-Dulatov	16	330	223	5	645	72
Al-Farabi-Markov	3	599	336	54	269	30
Abay-Tilendiyev	6	350	593	96	212	34
Gorky park	2	286	752	65	157	29
Tolebi-Baizakov	13	526	262	200	250	40
Rozybakiev - Baykadamov	4	523	260	245	238	21
microdistr. Kok Kainar	21	258	199	20	699	94
Ryskulov-Momyshuly	7	253	865	25	122	19
Ermensay	0	617	194	285	192	3
Tulebayev-Dzhambul	11	128	55	11	1059	27
Residential complex Solnechnaya Dolina	0	98	1059	77	47	10
Kamenskoe plateau	0	792	188	245	65	1
Furmanov-Tashkent	4	192	999	18	71	7
microdistrict Mamyr	7	350	596	90	214	34
Residential complex - ASYL Arman	0	0	6	3	1280	2
Residential complex- Zhana Kuat	23	427	131	93	548	69
Baganashyl	0	1201	17	46	26	1
Kurgauldy	0	389	497	222	170	13
Voenniy gorodok	2	135	61	2	1076	15
Microdistr. Karasu	10	165	95	12	966	43
Satpayev-Lugansky	1	295	600	216	167	12
Raiymbek Rural District Karasay district	4	285	800	53	130	19
Shanyrak 2	0	11	1278	1	1	0
Energostroitel	13	264	94	105	771	44
Zhana Kairat	0	119	1020	68	79	5
Almagul Microdistrict 24	2	88	1196	1	4	0

DOI: 10.37943/AITU.2020.94.34.004

Location of sensors	Low	Elevated	High	Very high	Dangerous	Poisonous
Akan Seri	0	1268	3	19	1	0
micr.Mamyr	0	1279	5	5	2	0
Miras 122	0	1258	19	12	2	0
Kurchatov 1	0		1289	1	1	0
Tatibekov	0	86	1100	53	51	1
KazGU Parking	0	1288	1	2	0	0
Kamyshinskaya	0	1225	2	63	0	0
Tulebaeva 82	0	1290	0	1	0	0
Asanbai Askarov 60	0	1267	10	12	2	0
3rd Ainabulak Microdistrict 107	0	1290	1	0	0	0
5th Tau Samal line	0	1291	0	0	0	0
Sagadat Nurmagambetov 330	0	1104	28	122	35	2
Timiryazev 17	0	1249	25	6	11	0
Microdistrict Karasu 2	0	1266	1	24	0	0
Mukanova 159	1	650	194	72	337	37

We can conclude from table 3 that the average frequency of each category represented table 4.

Table 4. The average values of each category

Low	3.65
Elevated	612.3
High	367.1
Very high	64.6
Dangerous	241.4
Poisonous	16.6

As can be seen from table 4, it can be assumed that the low-level value of atmospheric air pollution appears almost 4 times in 2 years on average. The increased leveled value is 153 times higher on average than the low leveled pollution. The difference between high and dangerous levels is 126 times. Moreover, the poisonous category is less common than all other categories, with the exception of the lowest category. The rarest category is the most needed clean air in the range of 0-12 PM2. 5 is a low category.

Conclusion

This paper presents a detailed analysis based on Google Colaboratory cloud services to identify the most common categories and the most rarely encountered categories of air pollution indicators. The article [7] considered and identified the top three polluted places according to the installed sensor locations and the top 3 cleanest places analyzed for 2 years. The authors add and propose a categorical grouping for each value of the air pollution level. From six categories presented in the central part of the article, we clearly saw that the low category is infrequent and the most popular types are categories like Elevated where the value range is between 13-34, High between 35-55, and Dangerous 150-249. This means that the Almaty air is contaminated and it is necessary to take measures at the state level and fight for clean air. As the study shows that the largest part of air pollution comes from transport, the authors suggest abandoning passenger cars and switching to public transport.

Scientific Journal of Astana IT University
ISSN (P): 2707-9031
ISSN (E): 2707-904X

References

1. Internet resource: https://ru.wikipedia.org/wiki/%D0%97%D0%B0%D0%B3%D1%80%D1%8F%D0%B7%D0%BD%D0%B5%D0%BD%D0%B8%D0%B5_%D0%B0%D1%82%D0%BC%D0%BE%D1%81%D1%84%D0%B5%D1%80%D1%8B_%D0%97%D0%B5%D0%BC%D0%BB%D0%B8
2. Koch, T., & Koch, T. (2005). Cartographies of disease: maps, mapping, and medicine (p. 840). Redlands, CA: Esri Press.
3. Internet resource: https://inbusiness.kz/ru/news/top-3-zagryaznitelej-vozduha-v-kazahstane
4. Duysebekova, K., Serbin, V., Kuandykov, A., Duysebekov, T., Alimanova, M., Orazbekov, S., ... & Alimzhanova, L. (2016). The Solution of Semi-Empirical Equation of Turbulent Diffusion in Problems of Polluting Impurity Transfer by Gauss Approach. *Procedia Computer Science,* 94, 372-379.
5. https://www.who.int/ru/news/item/02-05-2018-9-out-of-10-people-worldwide-breathe-polluted-air-but-more-countries-are-taking-action
6. Internet resource: https://www.kazpravda.kz/news/obshchestvo/nechem-dishat-v-kakih-gorodah-kazahstana-samii-opasnii-vozduh
7. Internet resource: https://www.airkaz.org/pm25.php
8. Internet resource: https://informburo.kz/novosti/zagryaznyonnyy-vozduh-almatypredstavlyaet-opasnost-zdorovyu-naseleniya-issledovanie.html
9. Internet resource: https://colab.research.google.com/notebooks/welcome.ipynb?hl=ru
10. Pyagay, V.T., Sarsenova, Z.N., Duisebekova, K.S., Duzbayev, N.T., & Albanbai, N. (2020). Analysis and processing of environmental monitoring system, *Procedia Computer Science*, 170, 26-33.