ENVIRONMENTAL DATA ANALYSIS USING ORACLE ANALYTICS CLOUD

ALENKA BAGGIA,[1] ROBERT LESKOVAR,[1] MICHAL KVET,[2] ROK HARTER,[3] ALENKA BREZAVŠČEK[1]

[1] University of Maribor, Faculty of Organizational Sciences, Kranj, Slovenia
alenka.baggia@um.si, robert.leskovar@um.si, alenka.brezavscek@um.si
[2] University of Žilina, Faculty of Management Science and Informatics, Žilina, Slovakia
michal.kvet@fri.uniza.sk
[3] Municipality of Kranj, Kranj, Slovenia
rok.harter@kranj.si

The large amounts of environmental data collected over longer or shorter periods of time are worthless if we do not analyse them and gain an in-depth knowledge of past behaviour and future trends. There are many different tools on the market for data analysis, but only a few of them are offered as a cloud solution. Oracle Analytics Cloud is one of the services offered as part of Oracle Cloud Infrastructure. In the case presented, Oracle Analytics Cloud was used to test the efficiency of the tool and develop a guide for its use based on the environmental data collected by the Municipality of Kranj. The presented case shows the variety of possibilities that Oracle Analytics Cloud offers to both users with no previous experience in data analysis and experienced data analysts.

Keywords:
data analytics, environmental data, Oracle Analytics Cloud, air pollution, municipality of Kranj, case study

University of Maribor Press

DOI https://doi.org/10.18690/um.fov.3.2024.4
ISBN 978-961-286-842-0

1 Introduction

In the era of climate change, many ecosystems are gradually approaching a tipping point that would have a negative impact on the environmental services provided to humanity. As environmental degradation and crises pose a serious threat to humanity, there is an urgent need to monitor the state of ecosystems and pursue sustainable environmental management. However, environmental data is becoming increasingly complex, comprehensive and detailed. Tackling the large, multidisciplinary problems facing today's environmental scientists requires the ability to utilise the available data and information for decision-making. Successfully synthesising heterogeneous data from multiple sources to support holistic analyses and generate new insights requires the application of advanced analytical techniques and data science methods (see e.g. (Gibert et al., 2018; Gupta et al., 2021; Hristopulos et al., 2020)).

The increasing use of data analytics creates a need for a new profile of professionals who understand big data and are able to use a variety of software tools to extract useful information for decision making. To train such professionals, educational institutions need both the infrastructure and the teaching staff familiar with data analysis. These goals are being pursued with the Erasmus+ project Including EVERyone in GREEN Data Analysis (EverGreen, https://evergreen.fri.uniza.sk/)[1], which aims to develop innovative teaching materials and make them available to lecturers and students. These outputs will strengthen the digital readiness, resilience and capacity of educators and students and build their digital and sustainability competences.

To illustrate the usefulness of the EverGreen project, this paper presents the results of a case study on the analysis of environmental data on air quality of the Municipality of Kranj (MOK), analysed with Oracle Analytics Cloud (OAC), as one of the most powerful software tools for data analytics. The aim of the paper is to illustrate the main data visualization features of the OAC tool using a real-life example, thus bringing it closer to the students and the professional audience concerned.

[1] Partners of the project are: Žilinská univerzita v Žiline (Slovakia) – leading partner, Veleučilište u Šibeniku (Croatia), Univerzita Pardubice (Czech Republic), Inkubator za nove tehnologije Trokut Šibenik d.o.o. (Croatia), and Univerza v Mariboru (Slovenia).

2 Oracle Platform for Data Analytics

2.1 Oracle Cloud Infrastructure (OCI)

Oracle Cloud Infrastructure (OCI) supports all cloud computing architectures (public, private and hybrid cloud) and all types of cloud services (Infrastructure as a Service - IaaS, Platform as a Service - PaaS and Software as a Service - SaaS). It offers cloud-based solutions that were previously only available on-premise. The solution uses autonomous Oracle services, an integrated security layer, robust features and optimization techniques that provide many benefits when using the included tools. Most importantly, OCI provides an excellent location for an autonomous Oracle database that can be restored, repaired or extended in the cloud without database administrator intervention. By using machine learning to automate routine tasks, an autonomous database provides higher performance, greater security and improved operational efficiency, allowing developers to spend more time developing business applications (Oracle, 2024a). OCI offers a wide range of tools:

- **Oracle Analytics** uses machine learning and artificial intelligence (AI) to help with decision-making. It offers Oracle Analytics Cloud, Oracle Big Data Service, Oracle Big Data SQL Cloud Service, Oracle Data Science and Oracle Cloud Infrastructure Data Flow, among others.
- **Application development environment** including Tools API Gateways, Blockchain, Data Science, Digital Assistants, Oracle MySQL Database Service. Oracle Application Express (APEX) and Visual Builder also stand out among the tools that enable quick and easy development of web applications with SQL, PL/SQL and JavaScript functionalities.
- **Database** which includes autonomous transaction processing, autonomous data warehouse, JSON database, database as a cloud service (virtual computer), Exadata Cloud Service, etc.
- **Integration tools** such as API Gateway, Application Integration, Oracle GoldenGate, Oracle Data Integrator, Oracle Cloud Infrastructure Data Integration, and SOA Cloud Service.
- **Control and management tools** that enable recording, monitoring, notification, resource management, etc.

- **Connectivity and networking tools** such as DNS server, e-mail, FastConnect, health checks, load balancing, and virtual network in the cloud.
- **Storage tools** that enable archiving, block storage, data transfer, file storage, use of local NVMe SSD storage, object storage, etc.

2.2 Oracle Analytics Cloud (OAC)

Oracle Analytics Cloud (OAC) is a data analysis tool included in OCI that uses a graphical interface that allows the analyst to work intuitively and easily with a variety of features. OAC includes a wide range of tools for data visualization, custom analytics, business intelligence and advanced analytics. Connecting to different data sources is very easy, as is preparing and transforming data into the appropriate format. OAC uses machine learning methods for predictive insights (Oracle, 2023). As the name of the tool suggests, OAC is available as a cloud service, which greatly simplifies the use of the tool within the organization and makes it possible to work in the environment using only a browser. Figure 1 shows the dashboard in the OAC environment. The desktop can be personalized and adapted to the user's needs. By default, the latest workbooks, datasets, and data flows are displayed on the dashboard.

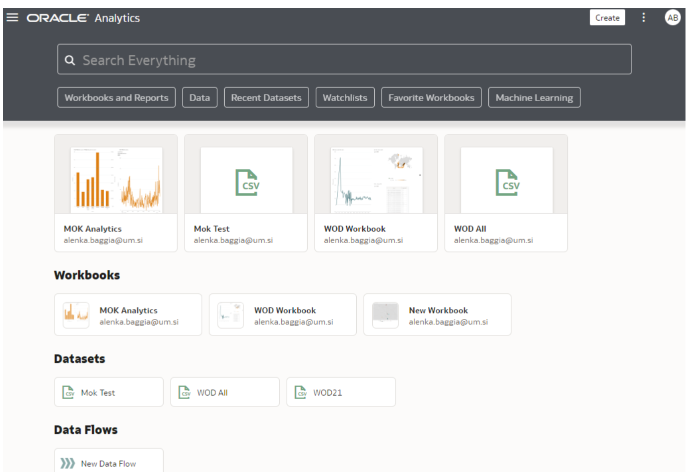

Figure 1: Oracle Analytics Cloud dashboard

Source: Own

As comprehensive cloud-based analysis platform OAC includes all the key functions required for a powerful data analysis tool:

- **Data integration and preparation**: Various data sources can be used and the process of cleansing, shaping and converting raw data into a format suitable for analysis is very intuitive. For example, a column in a dataset can be easily converted from a measure to an attribute, data formats can be changed, specific data formats can be used to visualise locations, etc.
- **Data visualisation:** Users can create interactive and insightful visualisations, dashboards, and reports without any special technical knowledge. In addition, OAC offers special automatic insights feature that uses AI to suggest different visualisations to the user.
- **Self-service versus advanced analytics:** On the one hand, OAC offers business users the ability to explore and analyse data themselves, but on the other hand, it also offers features such as predictive analytics and machine learning.

In addition, integration into the Oracle Ecosystem is one of the benefits of OAC, along with robust security measures and governance features to maintain control over the data. Since OAC is a cloud-native solution, it reduces the IT teams' infrastructure management efforts.

All these benefits led to the decision to use OAC as the primary tool for environmental data analysis in the EverGreen project. Three different cases were prepared to familiarise the user with the basic features of the OAC tool. The first case was that of the Municipality of Kranj, in which real environmental data on air quality was used. The case, the data used in this case and some examples of data analysis are presented in the following chapters.

3 Municipality of Kranj Case Study

In April 2022, Kranj was selected as part of the European Commission's project with the mission of implementation of 100 climate-neutral and smart cities by 2030. This was the result of an important step taken by the Municipality of Kranj in 2018 when it decided to implement environmental monitoring by setting up a smart city network for the Internet of Things (IoT). As part of this initiative, indicative measuring devices were installed at strategically selected locations to systematically monitor air quality, noise levels and meteorological data. The locations were selected based on the assumption of increased air pollution due to emissions from various sources such as traffic, combustion plants and industrial activities. As part of this project, citizens can access real-time environmental data via their mobile devices and the official website of the Municipality of Kranj, leading to increased awareness and engagement (Mestna občina Kranj, 2023).

Pollutants and meteorological data are measured continuously at individual measuring stations. For our case study, dust particles PM10 and PM2.5, nitrogen oxides, expressed as NO2, CO, O3, temperature, humidity, air pressure and noise were included in the analyses.

The environmental regulations in Slovenia (Decree on Ambient Air Quality (Uredba o kakovosti zunanjega zraka, 2011; Uredba o spremembah in dopolnitvah Uredbe o kakovosti zunanjega zraka, 2015; Uredba o spremembah Uredbe o kakovosti zunanjega zraka, 2018) and Environment Protection Acvt (Zakon o varstvu okolja (ZVO-2), 2022) define the limit and target values that are decisive for the protection of human health. According to this regulation, the distinction between limit and target values is defined as follows below.

A scientifically based limit value aims to prevent, avoid or reduce harmful effects on human health or the environment. It is set so that it is reached within a certain period of time and may no longer be exceeded once it has been reached. A target value for the prevention, avoidance or reduction of harmful effects on human health or the environment is ideally achieved within a certain period of time. Table 1 provides an overview of the limit and target values for the parameters measured at the monitoring stations in the municipality of Kranj. Limit values have been set for all parameters except ozone, while a target value has been set for ozone.

Table 1: Limit/target values for the parameters observed

Parameter	Unit	Measurement time period	Limit/target value
Nitrogen dioxide (NO2)	µg/m3	1 hour	200, max 18 exceedances/year
		Year	40
PM10	µg/m3	24 hours	50, max 35 exceedances/year
		Year	40
PM2.5	µg/m3	Year	20
Carbon monoxide (CO)	Mg/m3	8 hours	10
Ozon (O3)	µg/m3	8 hours	120, max 25 exceedances/year

Source: (Mestna občina Kranj, 2023)

Based on these measurements, the Air Quality Index (AQI) is used to display immediate information on current air quality so that they can adjust their activities according to air pollution. An example of an AQI meter is shown in Figure 2.

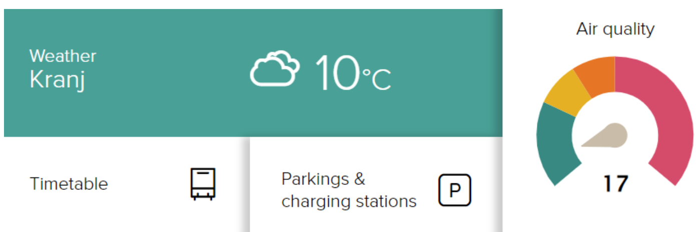

Figure 2: Excerpt from the Smart Kranj website
Source: (Mestna občina Kranj, 2024)

In addition to the environmental data, the sample data also contained noise values. According to the Decree on the Assessment and Management of Environmental Noise (Uredba o ocenjevanju in urejanju hrupa v okolju, 2004; Uredba o spremembah Uredbe o ocenjevanju in urejanju hrupa v okolju, 2019; Uredba o spremembah Uredbe o ocenjevanju in urejanju hrupa v okolju, 2022), there are four levels of noise limits, with levels II and III include residential areas that have a critical noise level of 65 and 80 dBA respectively.

3 Methodology

The input data for the analyses were provided by the MOK in .csv format. The air quality data, the air quality history and the air quality sensors for the period from November 23, 2022, to February 5, 2024, were included in the present analysis.

In order to carry out an efficient data analysis, the data contained in the dataset must be cleansed. The first challenge arose even before the data was imported into the Oracle APEX database tables (Oracle, 2024b). The values of PM10 and PM2.5, which were smaller than 2 were represented with smaller signs and the number 2: <2. This caused major problems when importing the data, as the data was initially treated as text and some values were distorted. To avoid this, we first replaced all strings »<2« with 1 before importing the data into the database.

The .csv files were first uploaded to the Oracle APEX instance in the Oracle Cloud Infrastructure to be cleaned and normalised before analysis. The three empty columns in *air_quality* and one in *air_quality_sensors* were dropped. Figure 3 shows the first three tables created in Oracle APEX.

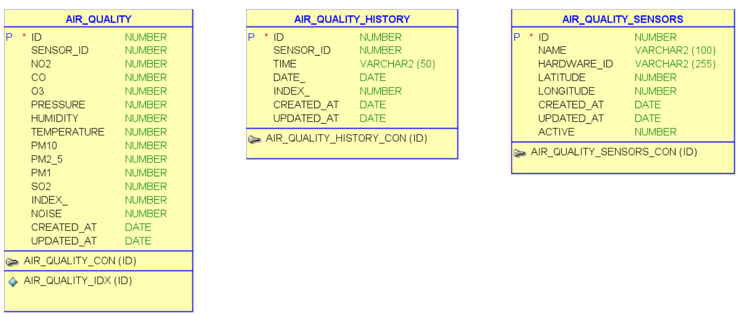

Figure 3: First tables created in Oracle APEX

Source: Own

When reviewing the data, it became clear that the data is not normalised and many values are missing in the *AIR_QUALITY* table. Not all measurements were carried out at all locations and not all measurements were carried out in the same period. In addition, some measurements (e.g. AQI) are based on moving averages of other measurements and are therefore only calculated when the required data is available. Since we did not want to lose so much valuable data, we decided to normalise the output. The AQI is available in each row for each ID. Therefore, this measure was included in the AQI table, along with the *sensor_id*, *created_at* and *updated_at* data. All other data was split into individual tables based on the ID and the specific parameter value. The tables were exported from APEX and imported into the OAC dataset. The ID was defined as a join condition for internal links between the individual data sources (Figure 4).

Figure 4: Initial dataset in Oracle Analytics Cloud
Source: Own

The process of data cleansing was continued in the OAC. First, the data was checked and the attributes defined. Based on the initial inputs, most of the imported data were identified as measures, so further steps were required to update the dataset. In addition, based on the expected analyses, the geolocation data (latitude and longitude) were converted to location settings and several conversions based on date data formats (day of the week, month of the year, time of year, time of day, etc.) were performed.

4 Data analysis and visualizations

After cleaning and formatting, the dataset was ready for analysis and visualisation. OAC offers the "Auto-Insights" feature, where visualisations and charts are actively created based on the obtained granularity of the dataset by selecting measures and dimensions and establishing relationships between the data elements. This process leads to a collection of insights from the dataset (Oracle, 2024). Apart from the automated insights, our first goal was to visualise a map of the sensor locations with the average AQI values as a mark diameter (Figure 5).

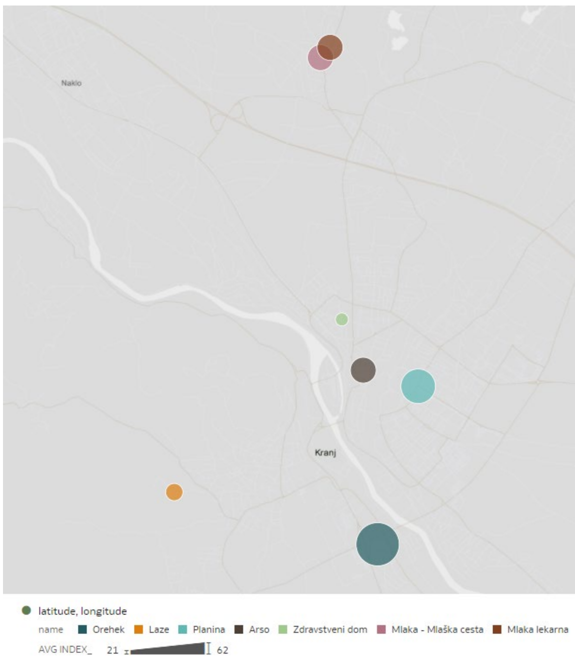

Figure 5: Locations of sensors and corresponding AQI values
Source: Own

From the variety of visualisations suggested by the Auto-Insight tool and our preferences, several visualisations were selected. First, insights into various possible presentations of the average AQI values, as shown in Figure 6.

Figure 6: Different visualizations of average AQI values
Source: Own

Further analyses focused on the noise level, the correlations between the AQI and other measured variables, and the graphical representation of PM2.5 and PM10 (Figure 7).

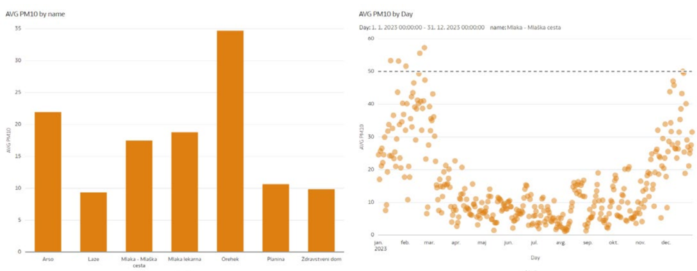

Figure 7: Visualizations of average PM10 values with limits
Source: Own

The selected visualisations presented here show only some of many different possibilities, but due to the limited length of the paper, we cannot present all of them.

5 Conclusions

Oracle Cloud Infrastructure offers a variety of functionalities, of which Oracle Analytics was used to analyse the environmental data recently collected in the Municipality of Kranj. During the development of the presented case, we experienced the flexibility of the platform. Starting with a wide range of input data and different data sources that can be used to define the dataset, to the multiple options for data cleansing and formatting. In addition, a recommendation tool in the dataset offers a variety of suggestions for improving the dataset, e.g. extracting part of the data or reformatting data. When the dataset is prepared, the visualisation tool also offers automatic generation of the most appropriate visualisation based on the type of data selected by the user. With the Auto-Insights tool, data analysis can be viewed from different perspectives and useful visualisations can be easily added to the presentation.

All in all, the OAC has proven to be an effective tool for data analysis, allowing both users with limited knowledge to take advantage of its many features and experienced users to visualise data from a variety of different perspectives. In line with the goals of the Evergreen project, further use cases are currently being developed to test the usability of the OAC, including the open data source of World Ocean data. By using different input data formats and analysis requirements, deeper insights will be gained that will later be shared with students as part of the piloting of the tutorials.

Acknowledgements

The authors would like to acknowledge the support given by the European Commission through the Action Erasmus+ Including EVERyone in GREEN Data Analysis (EverGreen), program ID 2022-1-SK01-KA220-HED-000089149, co-funded by the Erasmus+ program of the European Union. The European Commission's support for the production of this publication does not constitute endorsement of the contents which reflects the views only of the authors, and the Commission cannot be held responsible for any use which may be made of the information contained therein.

The authors would also like to thank Oracle Academy Program Manager, Mr. Darko Jureković, for his continuous support in disseminating the project results.

References

Gibert, K., Horsburgh, J. S., Athanasiadis, I. N., & Holmes, G. (2018). Environmental Data Science. Environmental Modelling & Software, 106, 4–12. https://doi.org/https://doi.org/10.1016/j.envsoft.2018.04.005

Gupta, S., Aga, D., Pruden, A., Zhang, L., & Vikesland, P. (2021). Data Analytics for Environmental Science and Engineering Research. Environmental Science & Technology, 55(16), 10895–10907. https://doi.org/10.1021/acs.est.1c01026

Hristopulos, D. T., Spagnolo, B., & Valenti, D. (2020). Open challenges in environmental data analysis and ecological complex systems(a). Europhysics Letters, 132(6), 68001. https://doi.org/10.1209/0295-5075/132/68001

Mestna občina Kranj. (2023, August 31). Kvaliteta zraka. https://www.kranj.si/kranj-moje-mesto/kakovost-zraka

Mestna občina Kranj. (2024, January 15). Pametni Kranj. https://pametni.kranj.si

Oracle. (2023, January 25). Getting Started with Oracle Analytics Cloud. https://docs.oracle.com/en/cloud/paas/analytics-cloud/acsgs/what-is-oracle-analytics-cloud.html

Oracle. (2024, April 25). Oracle APEX. https://apex.oracle.com/en/

Oracle. (2024, January 6). Oracle Cloud Infrastructure. https://www.oracle.com/cloud/

Uredba o kakovosti zunanjega zraka, Uradni list RS, 9/2011 964 (2011). https://www.uradni-list.si/glasilo-uradni-list-rs/vsebina/2011-01-0368/uredba-o-kakovosti-zunanjega-zraka

Uredba o ocenjevanju in Urejanju hrupa v okolju, Uradni list RS, 121/2004 14549 (2004). https://www.uradni-list.si/glasilo-uradni-list-rs/vsebina/2004-01-5018/uredba-o-ocenjevanju-in-urejanju-hrupa-v-okolju

Uredba o spremembah in dopolnitvah Uredbe o kakovosti zunanjega zraka, Uradni list RS, 8/2015 523 (2015). https://www.uradni-list.si/glasilo-uradni-list-rs/vsebina/2015-01-0225/uredba-o-spremembah-in-dopolnitvah-uredbe-o-kakovosti-zunanjega-zraka

Uredba o spremembah Uredbe o kakovosti zunanjega zraka, Uradni list RS, 66/2018 10140 (2018). https://www.uradni-list.si/glasilo-uradni-list-rs/vsebina/2018-01-3235/uredba-o-spremembah-uredbe-o-kakovosti-zunanjega-zraka

Uredba o spremembah Uredbe o ocenjevanju in urejanju hrupa v okolju, Uradni list RS, 59/2019 703 (2019). https://www.uradni-list.si/glasilo-uradni-list-rs/vsebina/2019-01-2668/uredba-o-spremembah-uredbe-o-ocenjevanju-in-urejanju-hrupa-v-okolju

Uredba o spremembah Uredbe o ocenjevanju in urejanju hrupa v okolju, Uradni list RS, 53/2022 3502 (2022). https://www.uradni-list.si/glasilo-uradni-list-rs/vsebina/2022-01-1162/uredba-o-spremembah-uredbe-o-ocenjevanju-in-urejanju-hrupa-v-okolju

Zakon o varstvu okolja (ZVO-2), Uradni list RS, 44/2022 2341 (2022). https://www.uradni-list.si/glasilo-uradni-list-rs/vsebina/2022-01-0873/zakon-o-varstvu-okolja-zvo-2