A week in Dexter’s Laboratory

A Wiseair Case Study

**Company value** is a key concept for us at Wiseair. A fast-growing startup is always working with few resources and little time. And yet, it needs to deliver tons of value every week. Different teams have different responsibilities: there are those who have to develop the technology, those who have to grow sales KPIs, those who have to define the long-term strategy. **Data has one responsibility: helping the entire company to deliver value, providing clients with the right answers to the right questions**.

As a Data Scientist, you will find yourself having to solve customers’ darkest doubts. Often these will be questions that you have no experience with. They will require a study of the context as well as the ability to make data talk. **Data analysis does not mean pushing buttons to run cool algos**.

In this Case Study, you'll have the opportunity to put your data science approach to the test with real insights we have been working on. **You have one week to complete it**. When you are done, share your output with us. But in the meantime and as always, we are available for you.

Good work!

Fulvio & Gabriele

# This week’s experiment

The data on which all our analyses are based is the concentration of airborne particulate matter, **measured in real time by our sensors** (in every municipality there are at least two sensors installed in different locations).

What is of particular interest to us is that the concentration of particulate matter is measured in ug/m3 and that the measurement is divided into particle size scales (PM10 means “all particles with a diameter of less than 10 um”). Keep in mind the values that administrators usually consider are the ones of PM10 an PM2.5.

The mayor of **Gazoldo degli Ippoliti** (a municipality in the province of Mantua in the north of Italy) installed our service some months ago. He is satisfied by the solution and **he would like to start taking decisions on the basis of our data.**

He called us, first of all, because he would like us **to provide him with some answers to the questions he has.** He knows that the first rule of perfect decision making is that it has to be grounded on facts, on real insights. This is an experiment for a true data scientist!

We have prepared a [Google Colab](https://colab.research.google.com/drive/14MRKBgV5gNz8acXxVjL0ywEjfJXmDP9J) you can download. It contains further information on our data. Also, we provided you with two datasets to work on (six months period of measurements in Gazoldo) **but you can decide to complement them with any public data you need** (traffic, heating, weather, etc…).

**Here’s what the mayor of Gazoldo would like to know. All the questions were taken from a real customer call.**

**Mayor of Gazoldo:** *We have been monitoring the situation with your service for quite some time now. I need to know if we have pollution problems during certain periods of the year, especially during winter. If so, could you explain to me why? I am afraid that it might be due to heating.*

Any extra insight you will be able to extract from these data will count as a plus. However, always remember the priority is, first of all, to answer the questions the client poses with the best attention possible.

The complexity of the process and the algorithm, and the numerosity of the data sources you use to produce the answers for the Mayor of Gazoldo’s sentence is not relevant: **we don’t need bazookas to kill ants**.

# Our Weekly Delivery

For this specific case **there are two outputs** to be sent in 7 days from now.

1. The first one is **a doc/pdf report (2 pages max) that answers the question and doubts expressed by the Mayor of Gazoldo.**

The Mayor of Gazoldo himself is the addressee of the document: even though we will not forward your analysis to him, keep in mind that you are writing for him. You cannot assume he has a background in data science.

**In this case, the key aspect you will be evaluated on is your capability to provide insights on the problems in a simple and clear way**. Take into consideration that your analysis has to be understood by a customer (not a technical person). Also, as you may have noticed we deliberately chose to report the questions as administrators posed them to us in the first place. Without filtering or rephrasing them. This is because **your ability to interpret what exactly the customer is asking** will be evaluated.

1. The second output we would need is **the script you wrote to solve the case. The code must be written in PYTHON. This is the only requirement.** You can decide to work using Google Collab or another environment.

The addressee of this deliverable is Wiseair team, which you can assume has a technical background and understanding of data science.

We will evaluate your ability to write algos using Python code and **the clarity of the code itself**. This means we will pay a lot of attention to the structure of the code, and to the comments you will write. **True data scientists are the ones that are able to explain why they chose an algo over another and why they selected certain variables for the analysis**. For them data is a matter of context and algos are not just black boxes.

Remember that we are a team and if you need us to clarify doubts or concerns we remain available! You can refer to Fulvio, our CTO, on Telegram to get a more immediate response.

**Fulvio’s number and mail are:** +39 3458509312 - fulvio.bambusi@wiseair.vision

We are super excited to get to know you better during the case. **See you in the lab!**