Visual Analytics Tool for Air Quality Index

Visual Analytics project 2023/24

ELISA DE BELLIS 1858927

Contents

1 Introduction

2 Related works

3 Data

4 Visualizations

5 Analytics

6 Application & Utilities

Introduction

Air pollution emissions have declined in the last decade, resulting in better air quality. Despite this improvement, air pollution remains the largest environmental health risk in Europe.

Air Quality Index Visual-Analytics tool has the aim to analyze concentration of pollutants and explore insights to capture an overview of the last 10 years about the situation of Italy.

Related works

1. The TRAFAIR air quality dashboard: The TRAFAIR dashboard monitors urban air quality in real-time, tracking pollutants like CO, NO2, NO, and O3 using low-cost sensors. It offers interactive visualizations to help decision-makers analyze trends and implement policies. (BACHECHI et al., 2020)

2. AirVis : AirVIS is a web-based tool for analyzing air quality data through visual analytics, offering GIS, Scatter Plot, and Parallel Coordinates views. It helps users track pollution trends, identify anomalies, and explore pollutant correlations with AQI. (LIAO et al., 2014)

3. Air Quality Index (AQI): The Air Quality Index (AQI) tracks real-time air quality, rating levels from 'good' to 'dangerous' based on pollutants like PM10, PM2.5, NO2, and O3. Agencies like the EPA and EEA use it to inform the public about pollution levels (EEA)

Dataset

The dataset is taken from ISPRA (Istituto Superiore per la Protezione e la Ricerca Ambientale).

It is divided into four files, one for each pollutant: PM2.5, PM10, NO2, O3

- station_eu_code
- id_regione
- id_provincia
- id_comune
- station_code
- Regione
- Provincia
- Comune
- nome_stazione
- tipo_zona
- tipo_stazione
- TIPO
- Lon
- Lat
- yy
- n
- sup25
- sup15
- media_yy
- minimo
- massimo

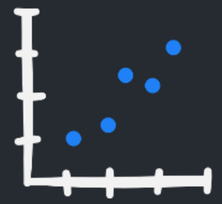
Dataset

To establish air quality on the basis of previous polutants, reference is made to the European Environment Agency (EEA)

Pollutant	Index Level (based on concentration in µg/m^3)					
	Good	Fair	Moderate	Poor	Very Poor	Extremely poor
PM2.5	0–10	10–20	20–25	25–50	50–75	75–800
PM10	0–20	20–40	40–50	50–100	100–150	150–200
NO2	0–40	40–90	90–120	120–230	230–340	340–1000
O3	0–50	50–100	100–130	130–240	240–380	380–800

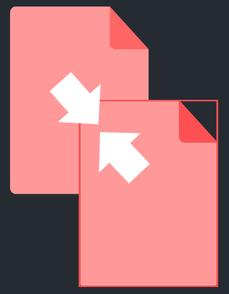
Pre-Processing

2016 Divide data by year



Select numerical value for t-SNE analysis

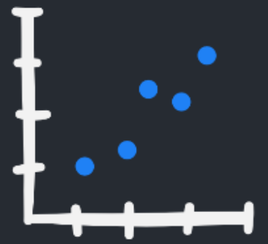
Group data by Regions



Create one file for all pollutants divided by years.

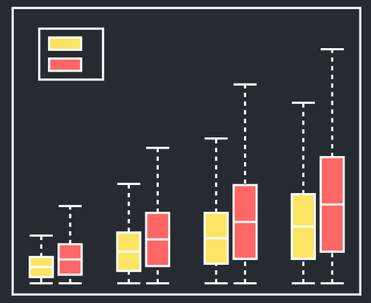
Visualizations

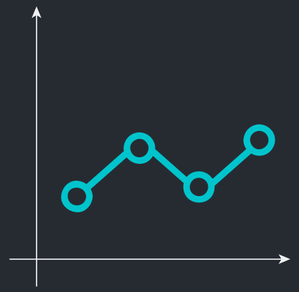
Map



Scatter-plot

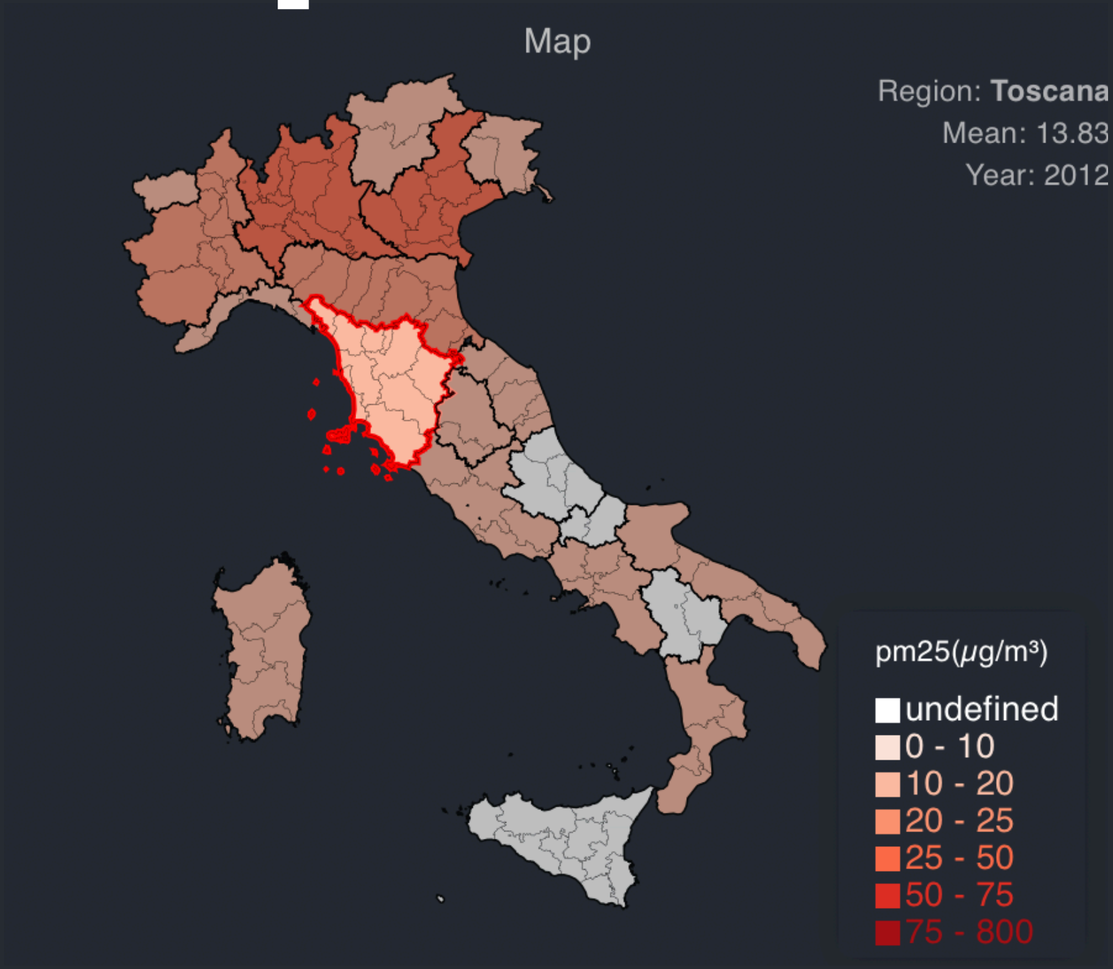
Box-Plot





Time-Series

Map



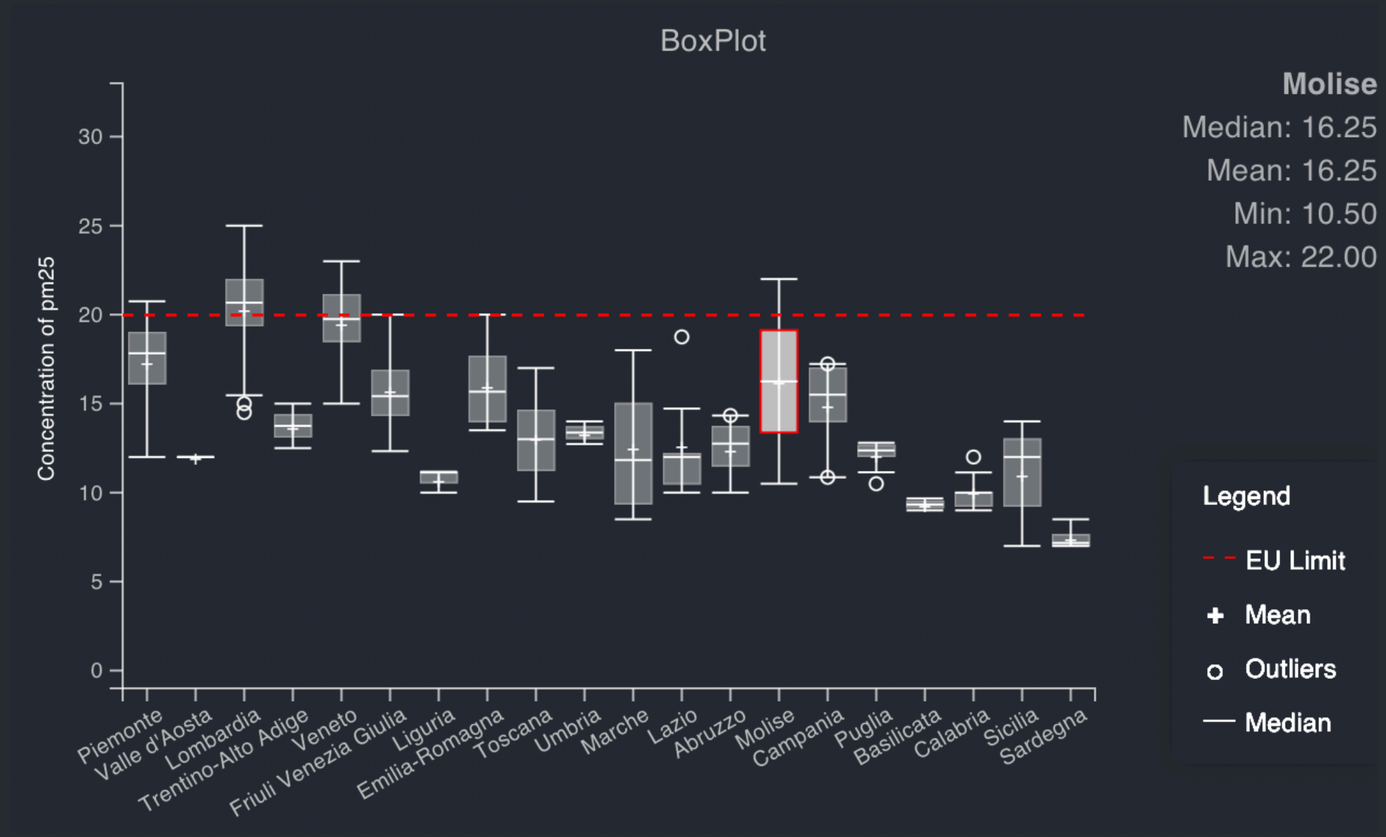
Visualization

The Map allow to show the mean concentration of the different pollutants in the Italy's regions. It consist of a map divided into 20 regions, each colour according to a shade of a red scale, according to the concentration of the selected pollutant.

Insight

- Critical regions and province
- Overview of the general situation

Box-Plot



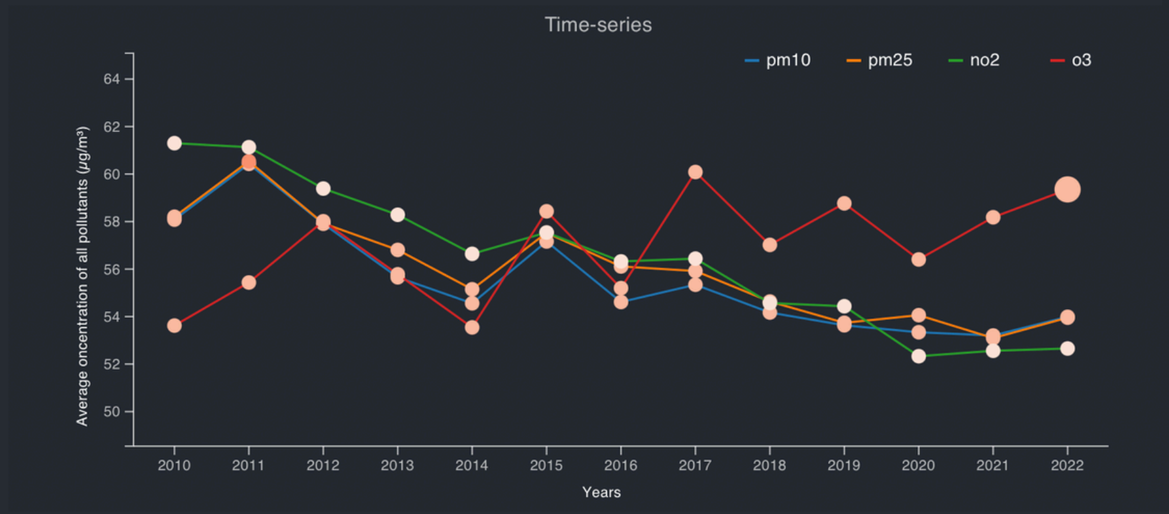
Visualization

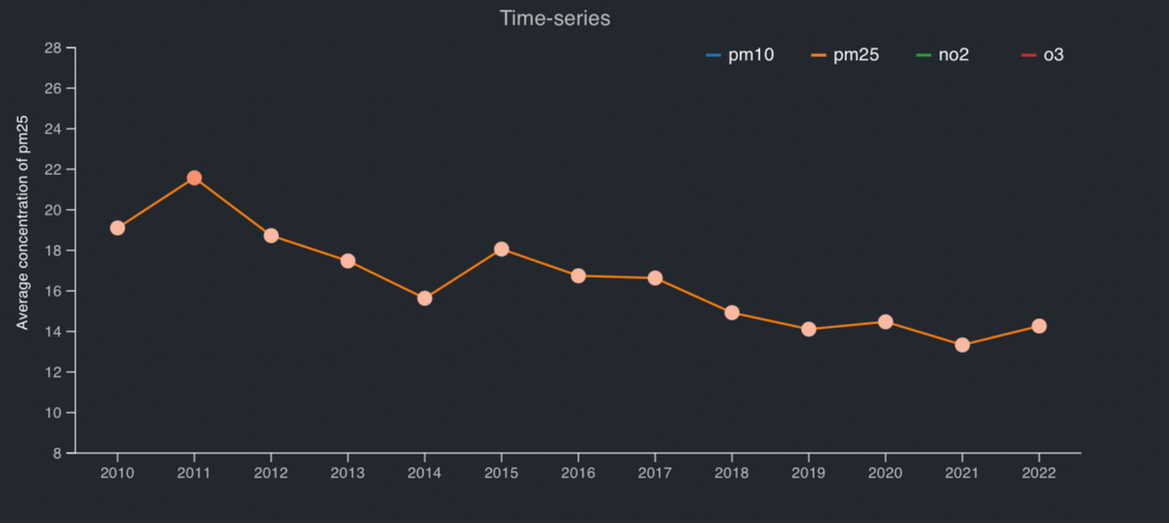
The box-plot depicts the data for each region and allows comparison of data from all regions, showing the mean, median, maximum and minimum values and outliers. The graph also shows a threshold indicating the European limit allowed for that pollutant.

Insight

- Critical regions and province
- Comparison between different regions and provinces

Time-Series





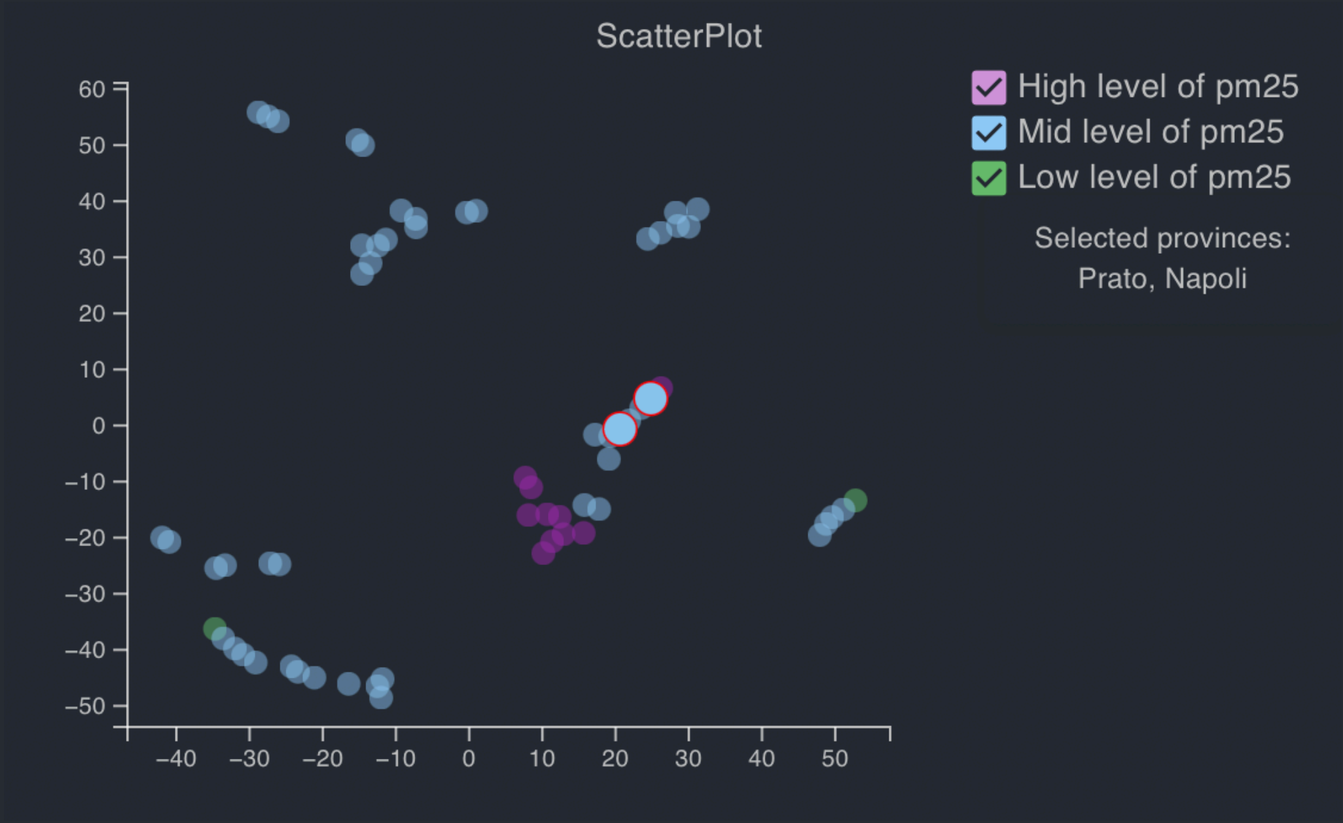
Visualization

The Time Series depicts how the average concentration in Italy in general has changed over the years. On the x-axis we have the different years taken into consideration, from 2010 to 2022, on the y-axis the average of all regions.

Insight

- Decrease in pollutants over years
- Impact of Lockdown

Scatter-Plot



Visualization

The Scatter-Plot is the result of dimensionality reduction using t-SNE. It helps to amplifies the separation between clusters of the data, that are arranged on the 2D space of the scatterplot.

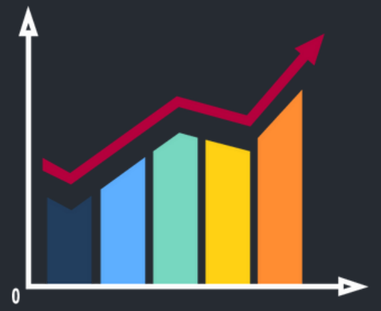
Insight

- Cluster models based on agent concentration
- Analysis of specific provinces

Analytics

The analytical process of the project focuses on applying t-distributed stochastic neighbor embedding (t-SNE) as a technique for dimensionality reduction. The following steps were undertaken to achieve the objective:

PRE-PROCESSING



STANDARDIZATION

t-SNE

Since my aim was to find clusters of the data in order to capture some insight, this solution was the one that worked the best for my purposes wrt to PCA or MDS.

Application & Utilities

Intended Users

The Visual Analytics tool for the Air Quality Index is designed to be a valuable resource for environmental analysts, policymakers.

ANALYSTS

POLICYMAKERS

DEMO