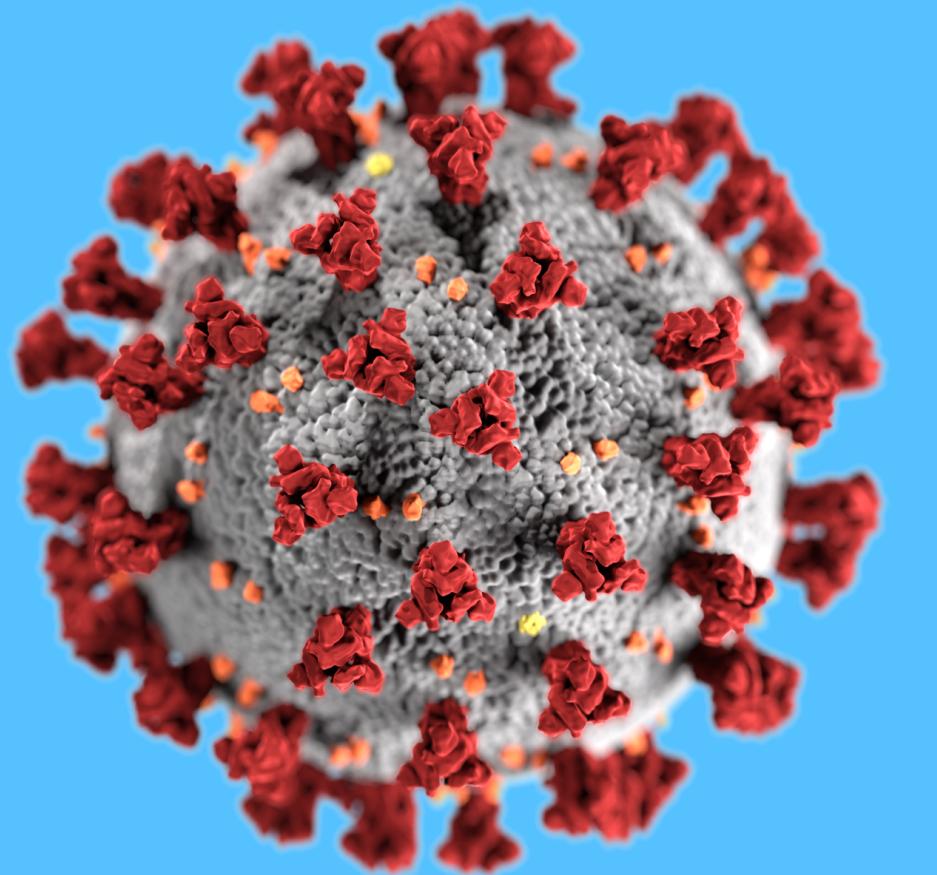


Tuscany COVID-19 Analysis

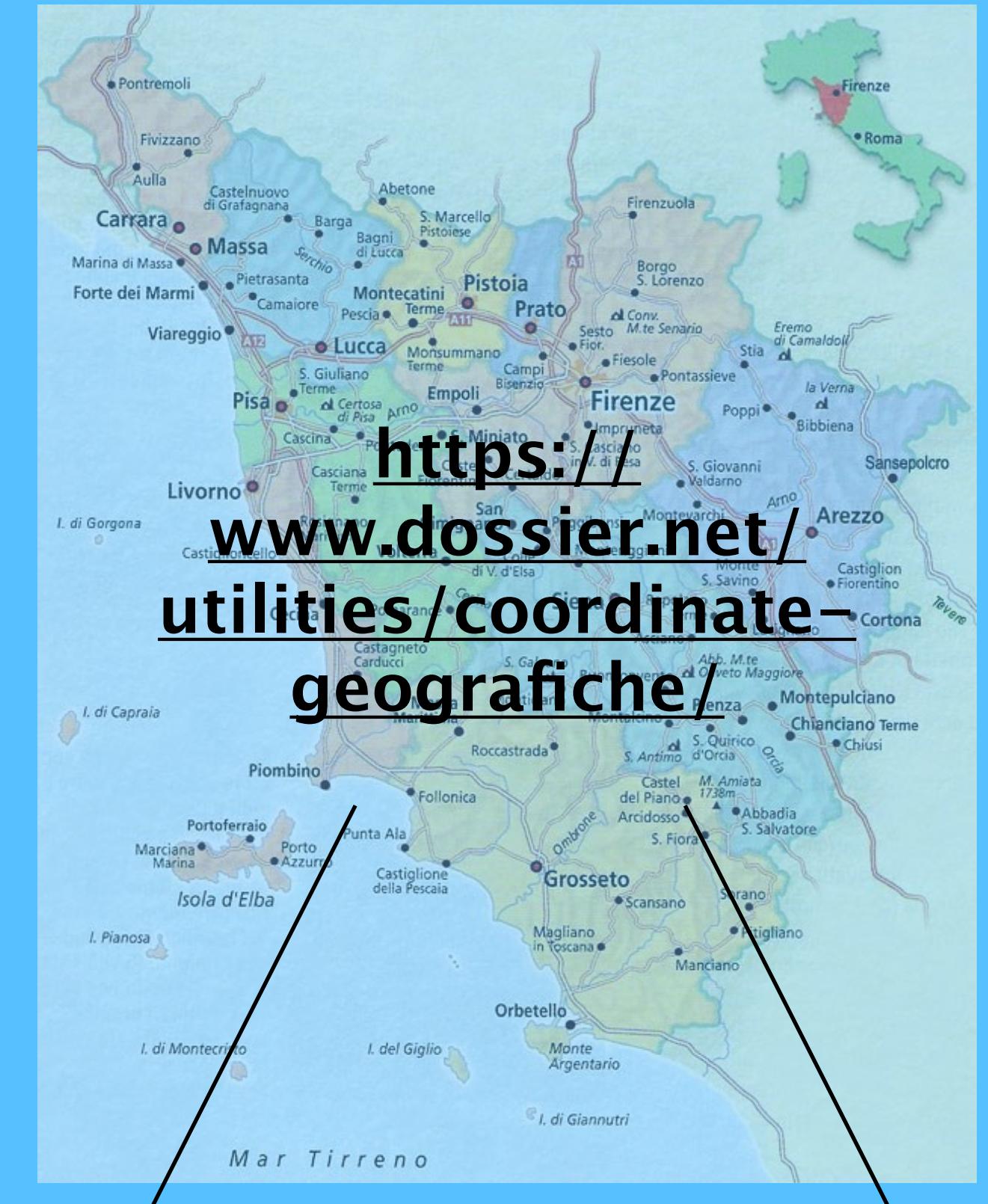


Introduction

- The interest for this analysis is due to the influence of the present virus pandemic in everyday life.
- This analysis will be based on both environmental and demographic factors.
- The aim is to assess the risk of COVID-19 infection in Tuscany Provinces and providing insights.

Data Acquisition and Cleaning

- I used the following publicly available datasets:
 - * “covidars.csv” — contains daily updated information about COVID-19 situation in Tuscany
 - * “Popolazione_indicatori_2019.xls” — including demographic factors such as old age index
 - * “Tavole_maschi_per_eta_e_classi_eta_31_12_2019.xls” — to access the percentage of male inhabitants by Province
 - * “https://it.wikipedia.org/wiki/Comuni_della_Toscana” — for information regarding area extension and population by Province and by Municipality



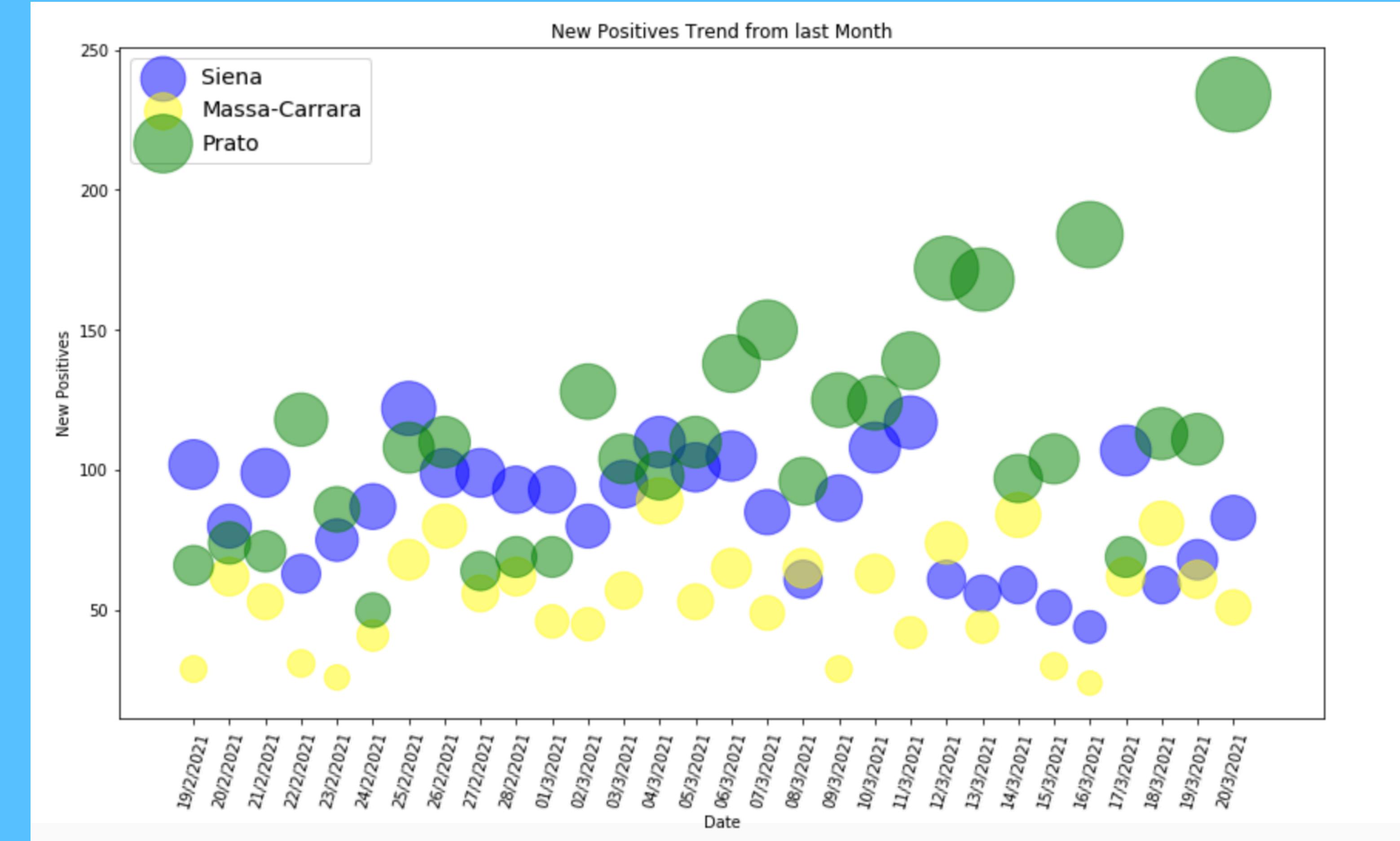
<https://www.dossier.net/utilities/coordinate-geografiche/>

Latitude Longitude

Data Exploratory Analysis

- **Siena** – my hometown
- **Massa-Carrara** – highest lethality
- **Prato** – highest percentage of infected people

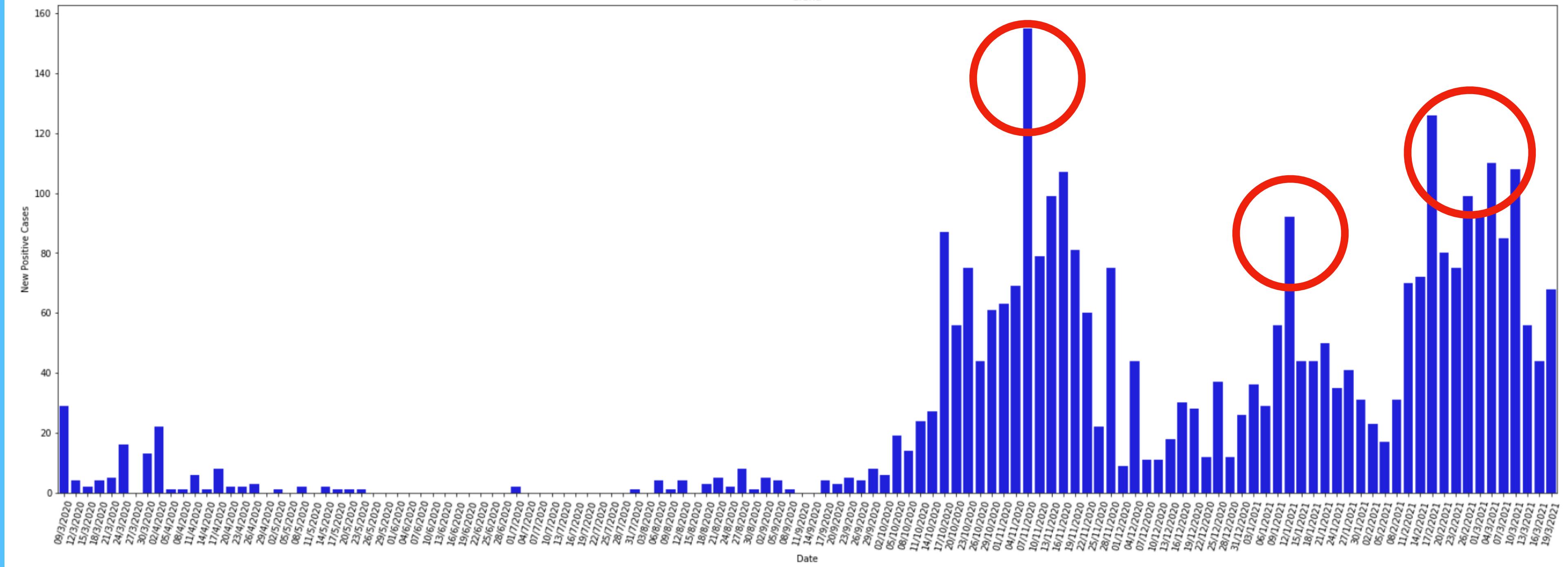
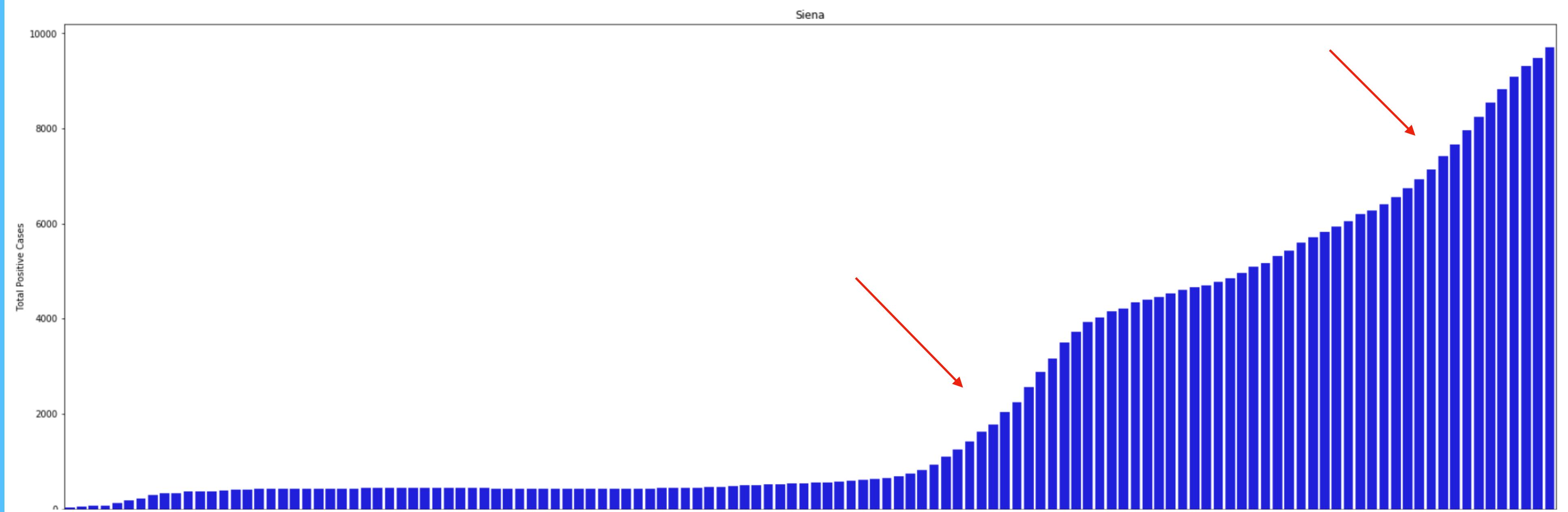
Confronting last-month situation
for these selected Provinces



Siena

March 2020

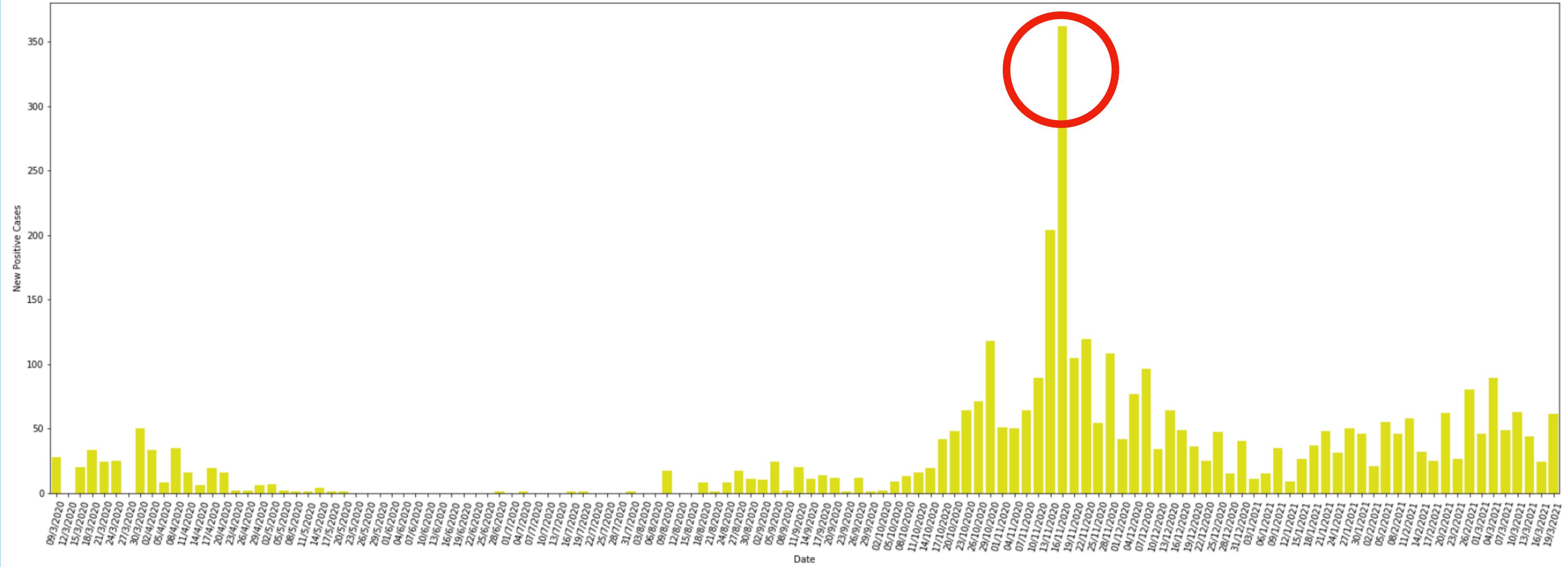
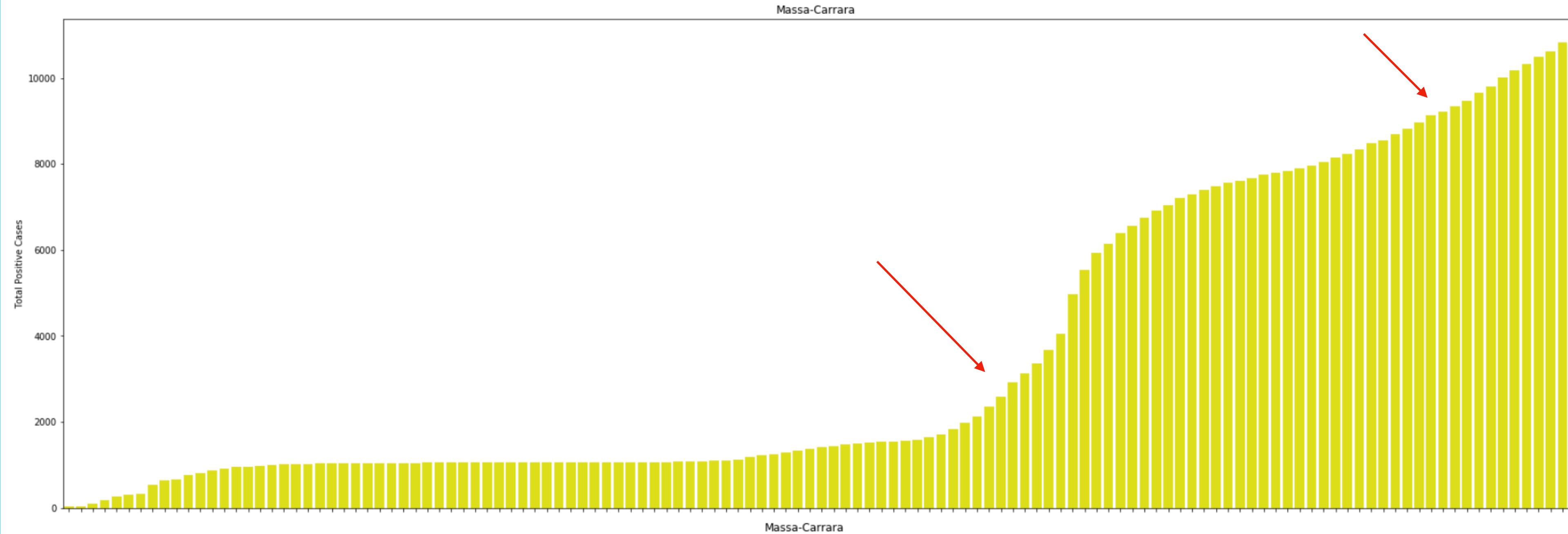
-
March 2021



Massa Carrara

March 2020

–
March 2021

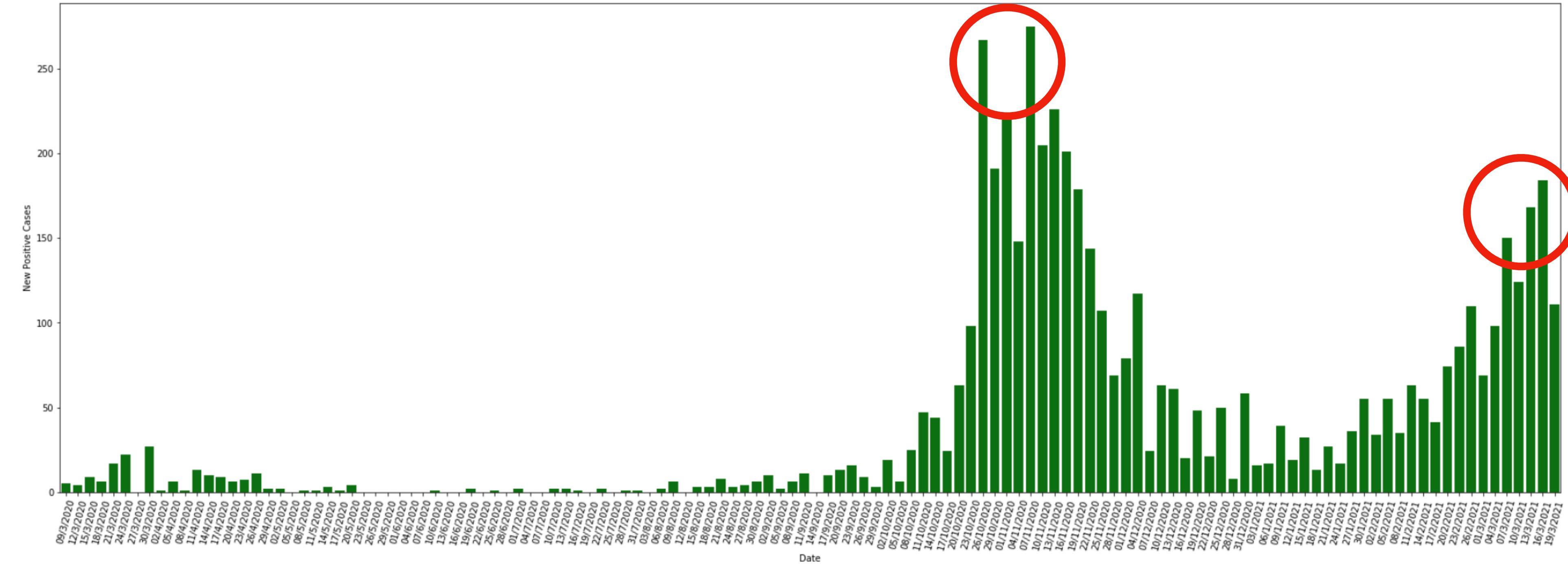
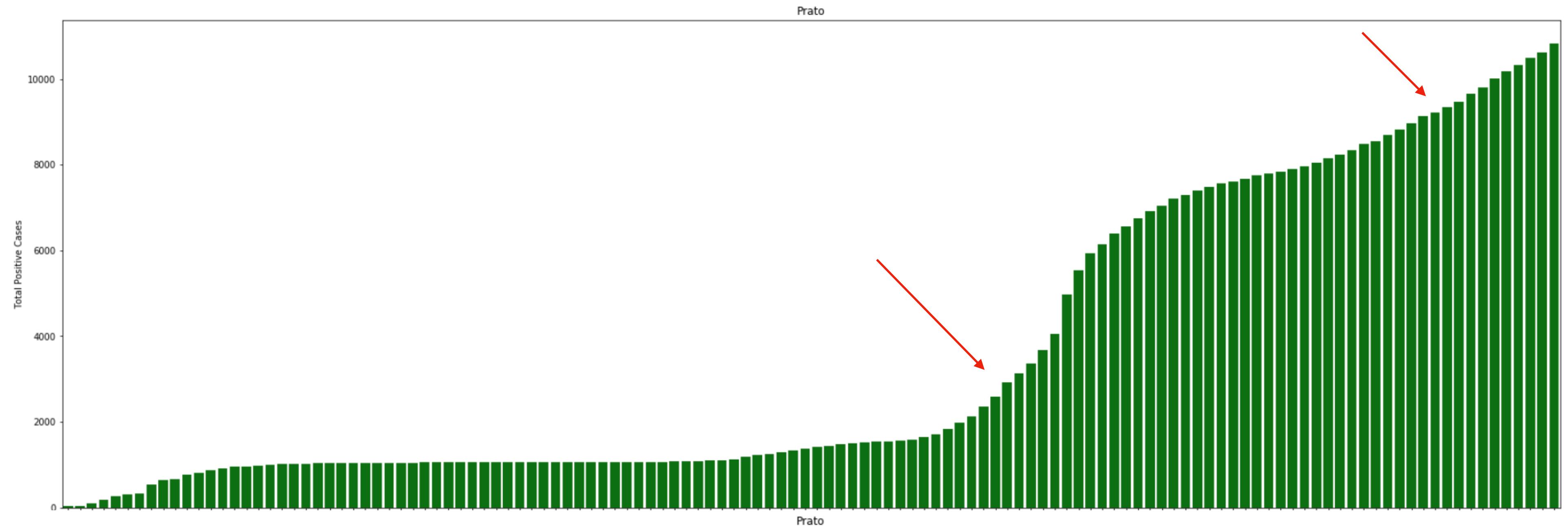


Prato

March 2020

-

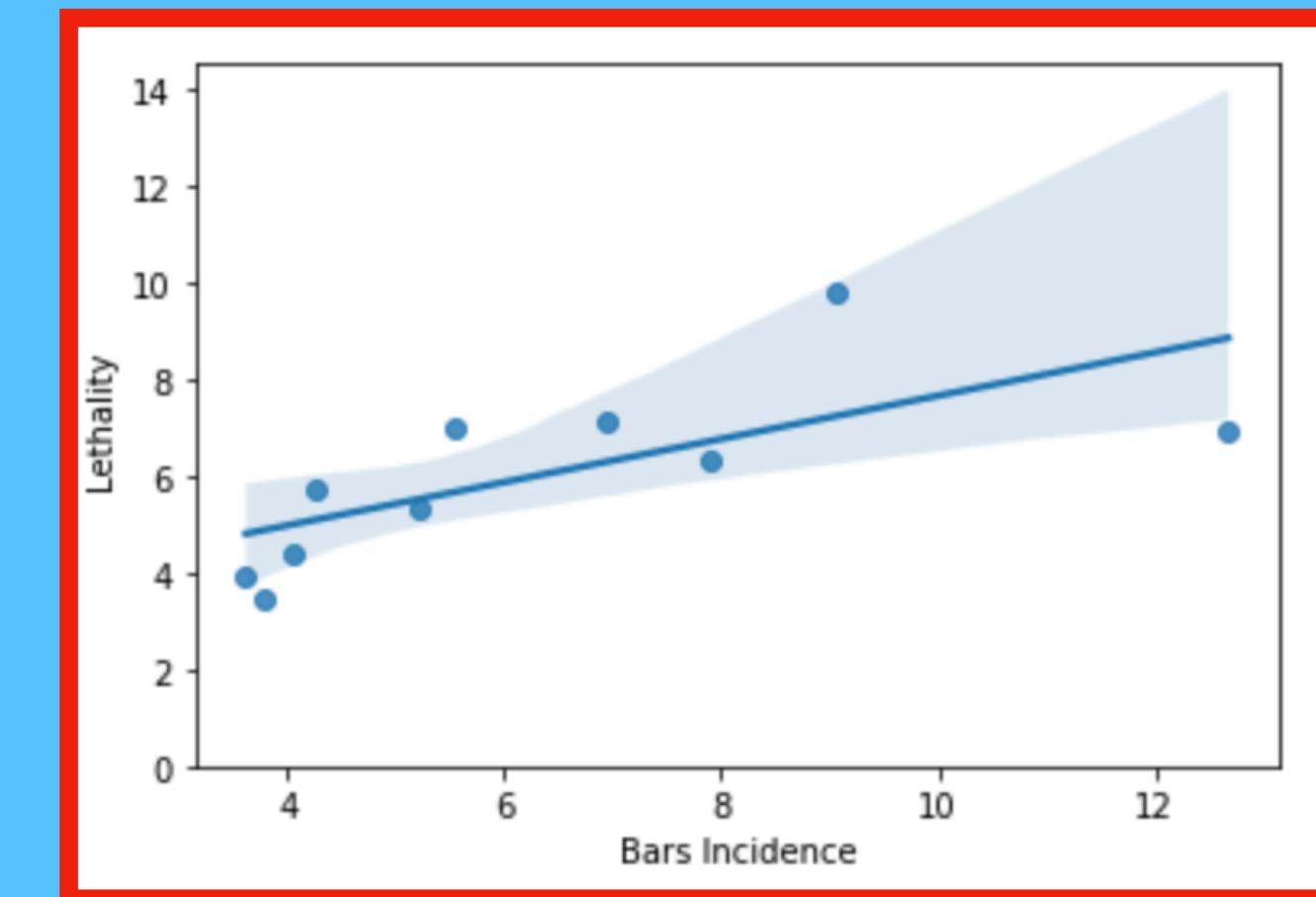
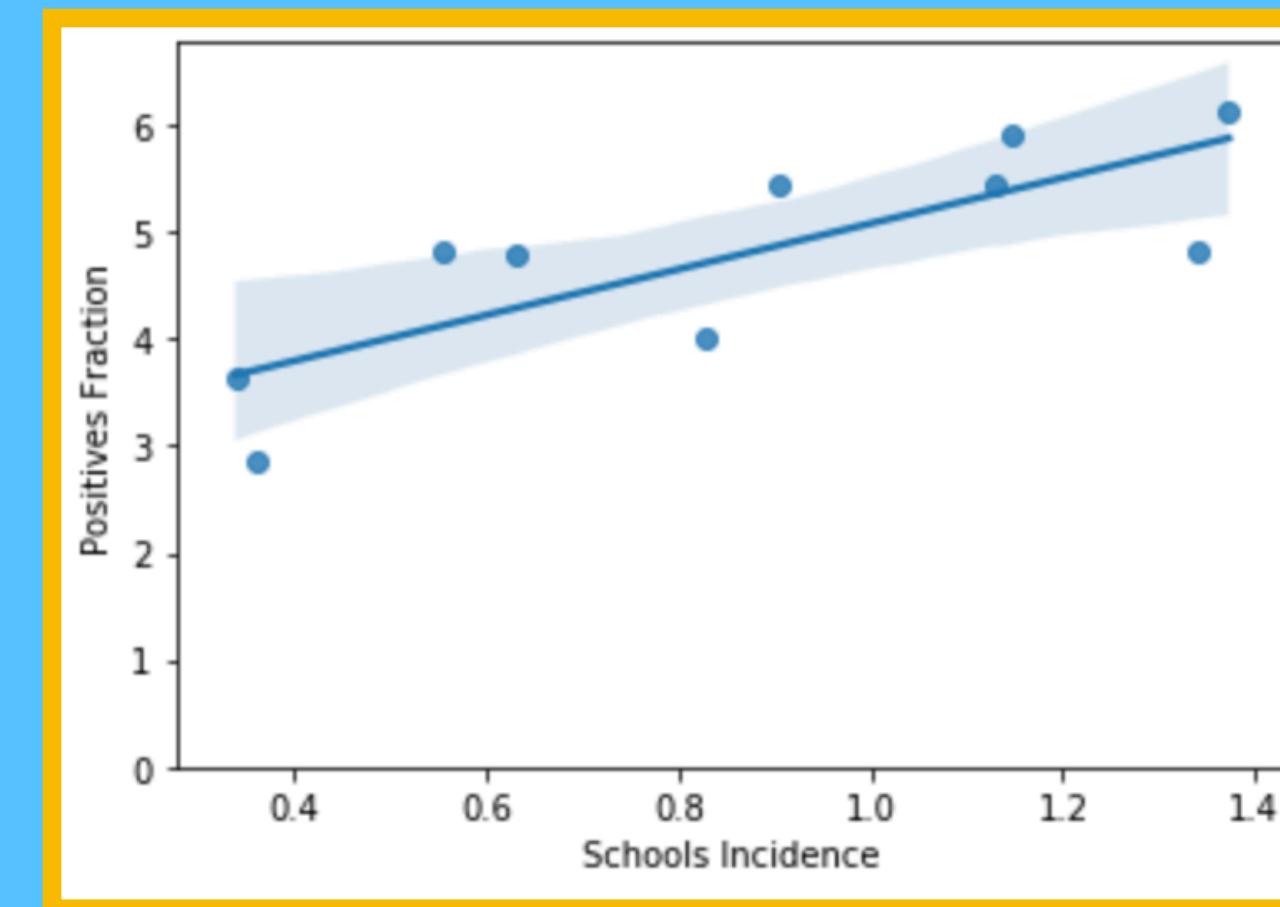
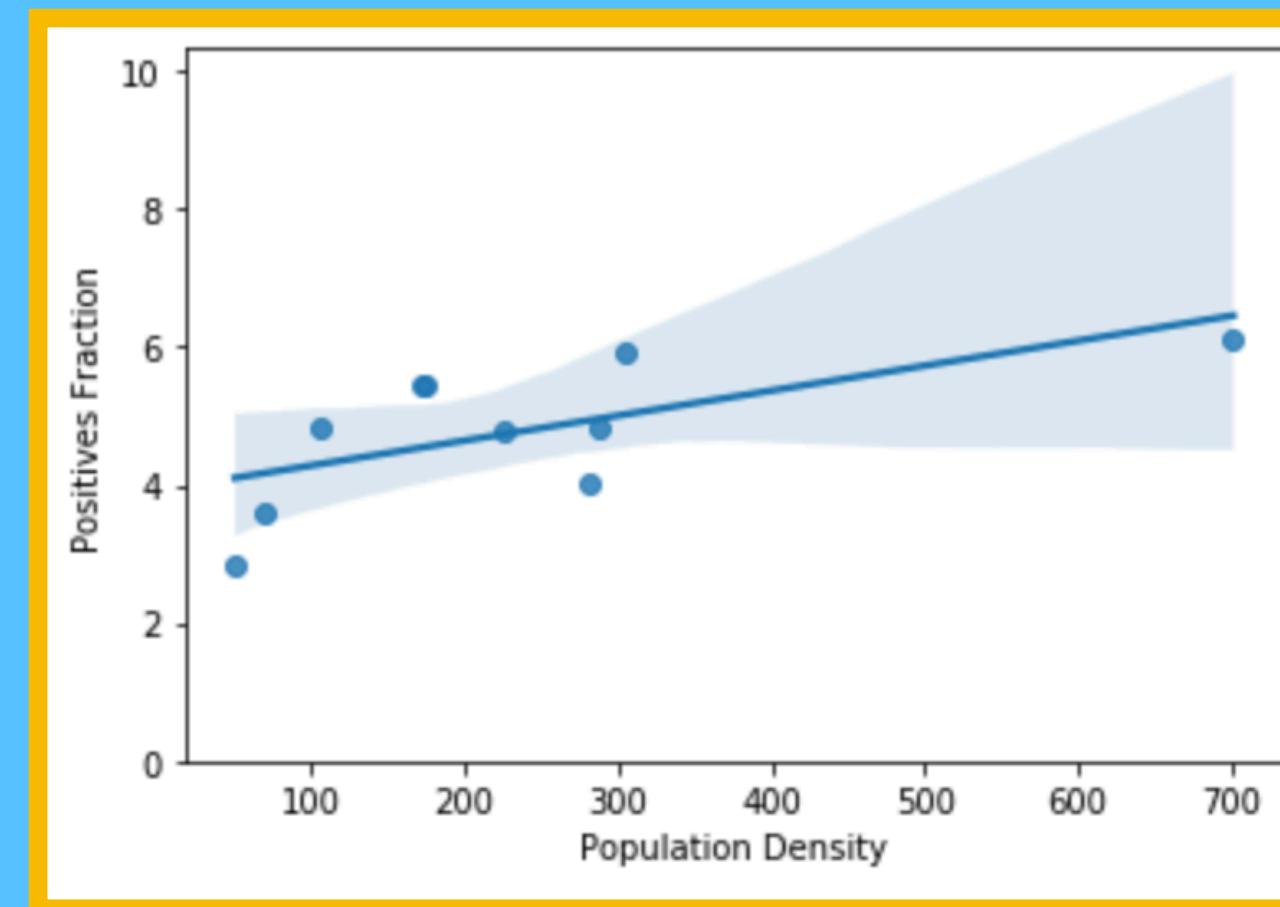
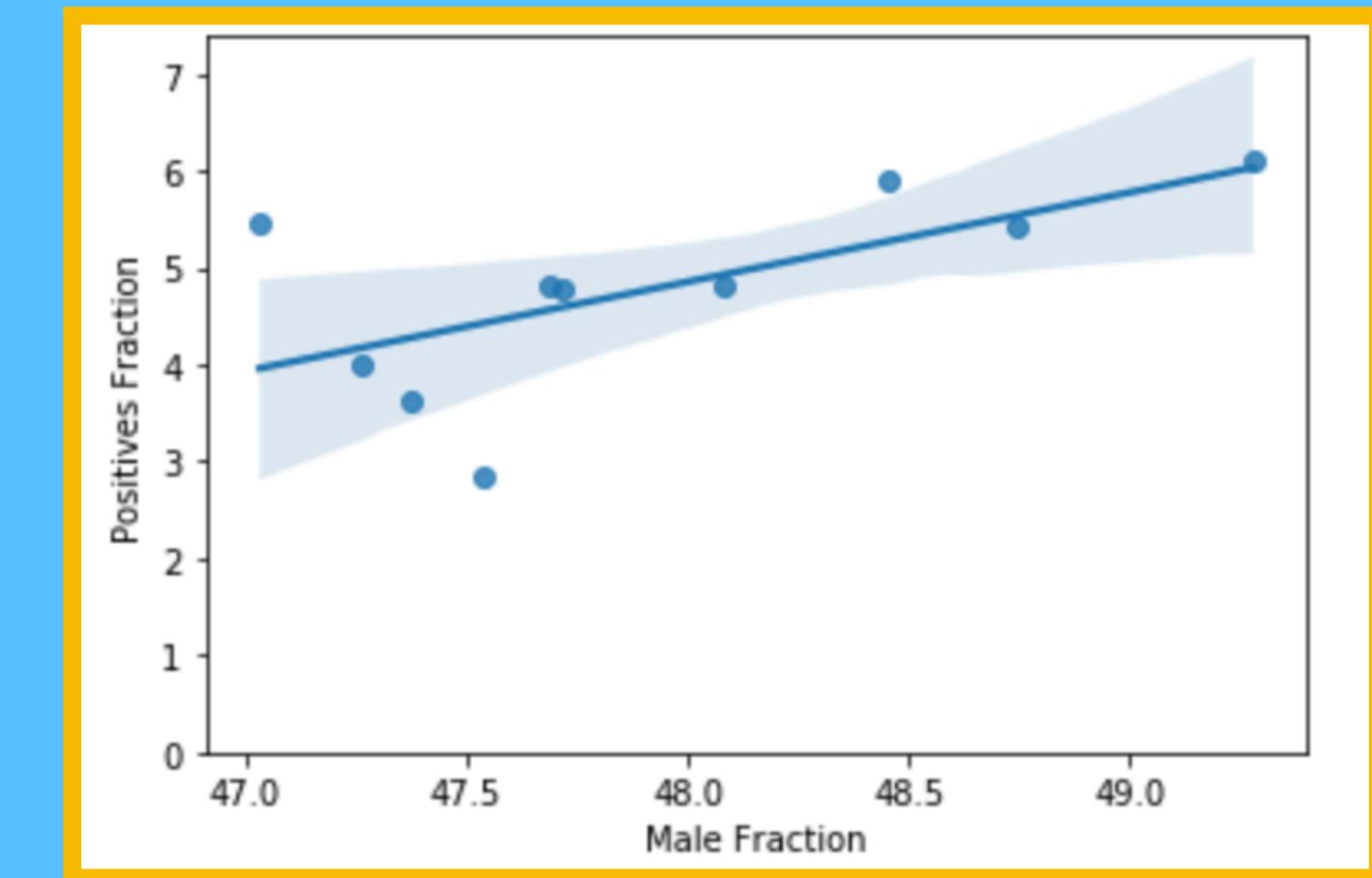
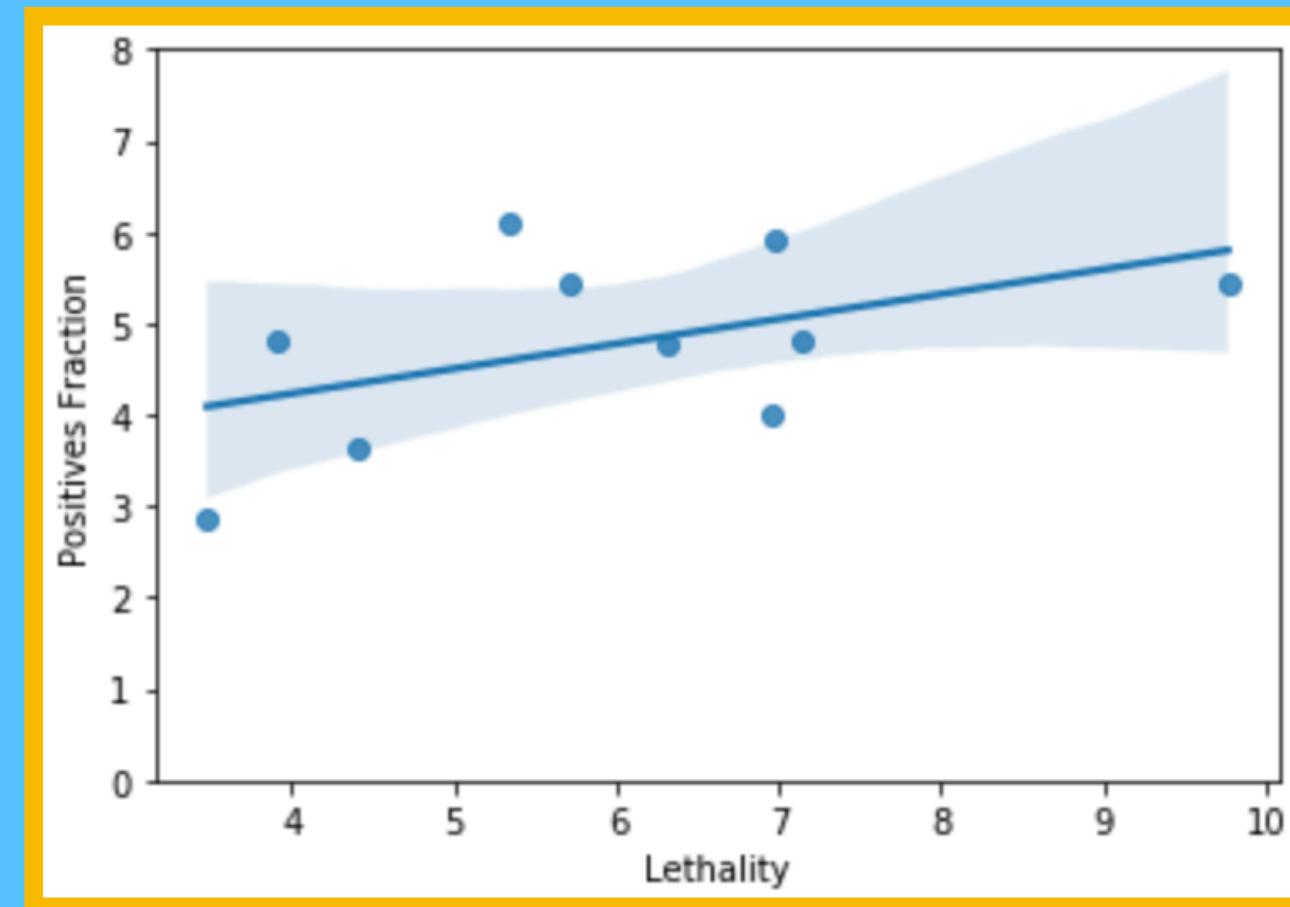
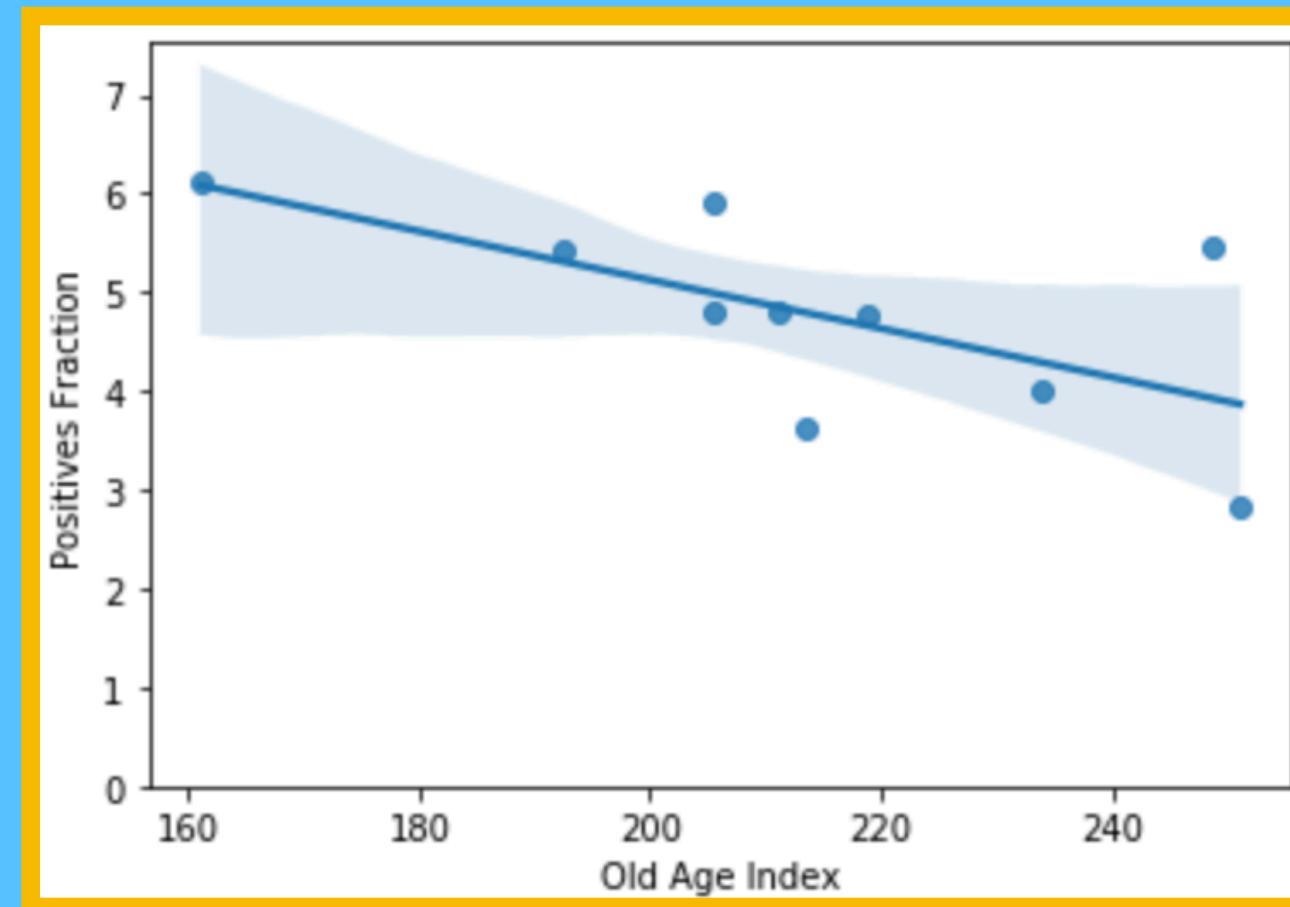
March 2021



Features Correlation

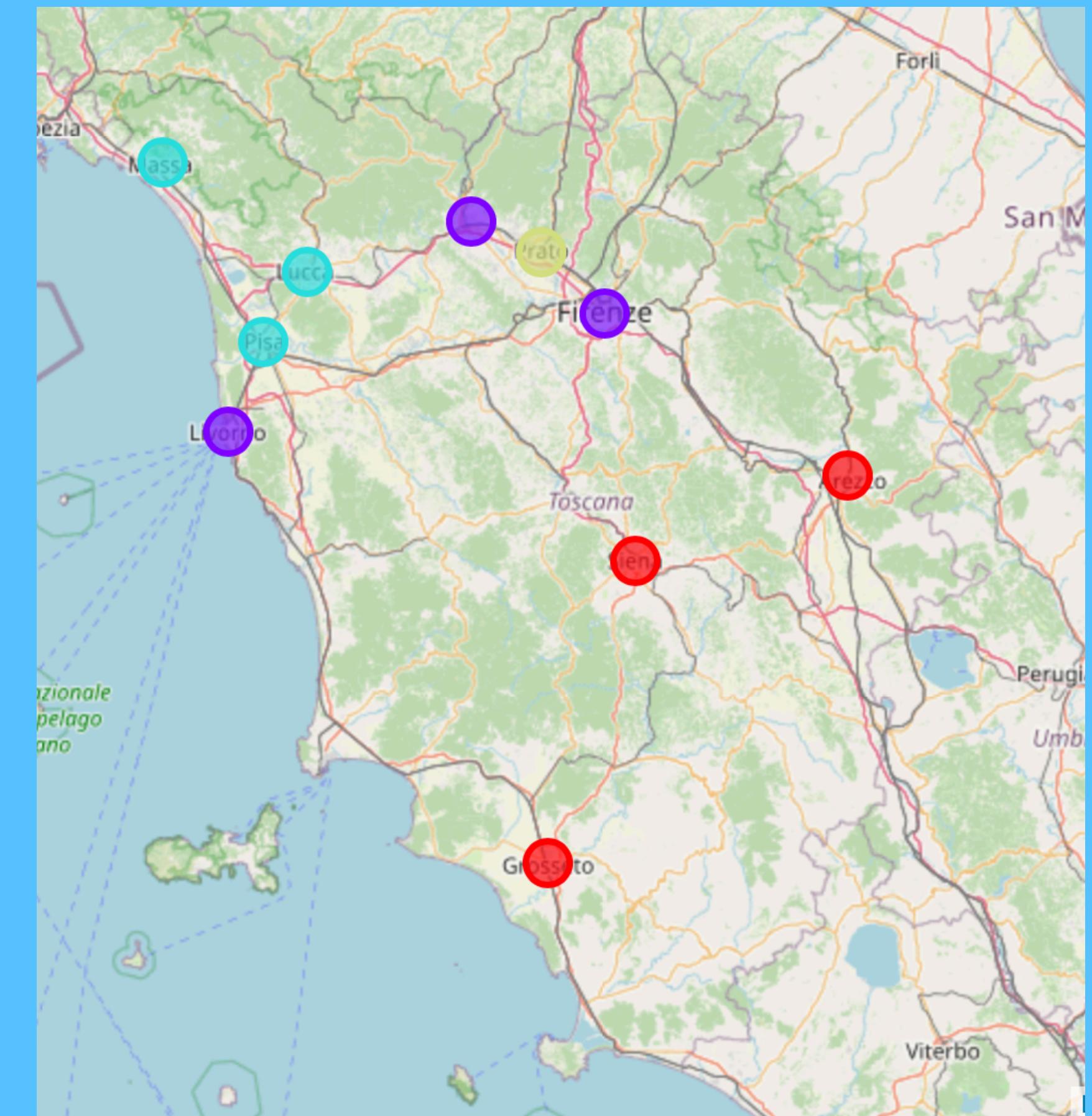
Lethality Correlation

Percentage of Infected People Correlation



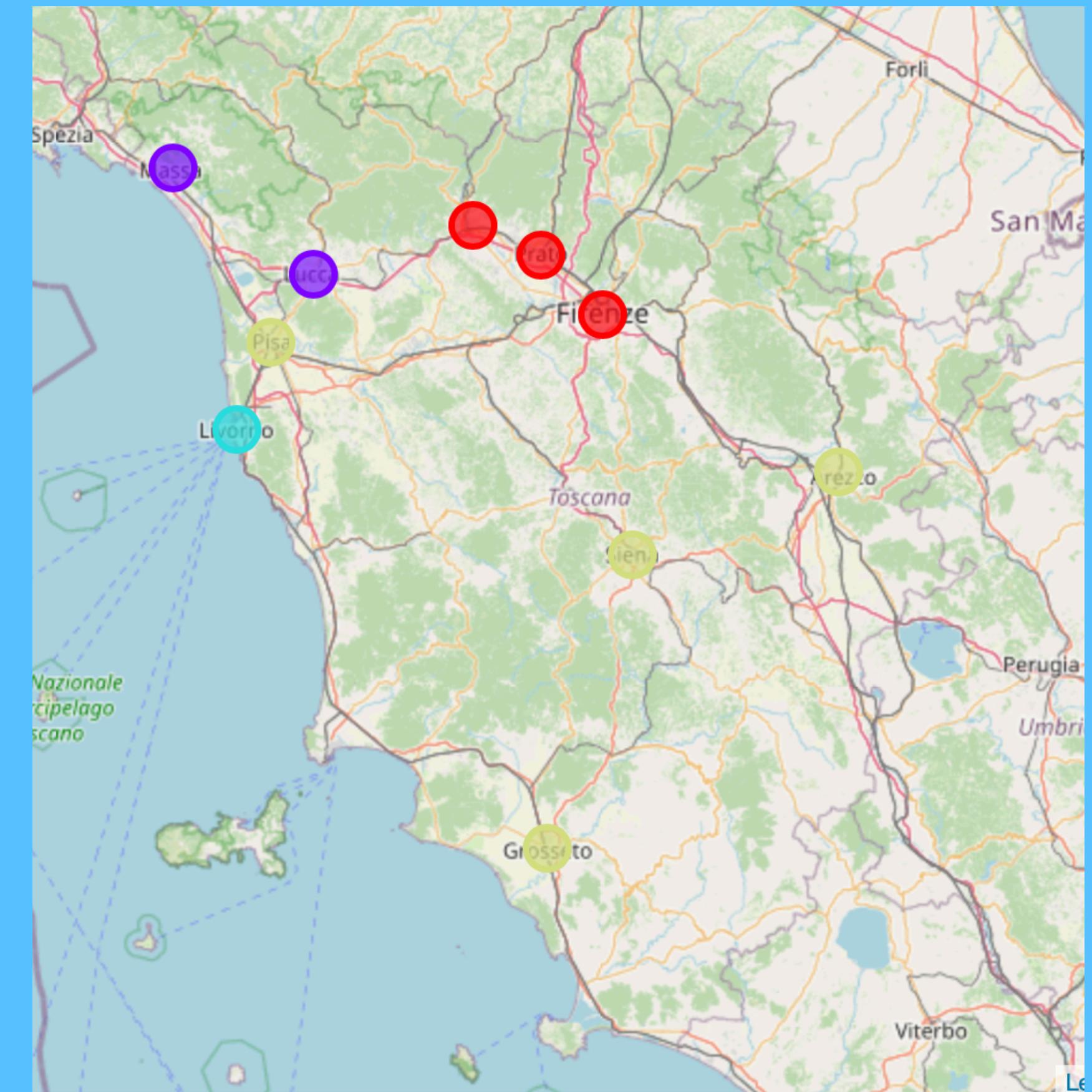
Clustering Approach based on Demographic Features

- Features Selected: Old Age Index, Lethality Index, Percentage of Male, Percentage of Infected People, Population Density
- Red : Low Risk
- Purple : High Risk
- Light Blue : Medium/Variable Risk
- Pale Green : Potential Risk



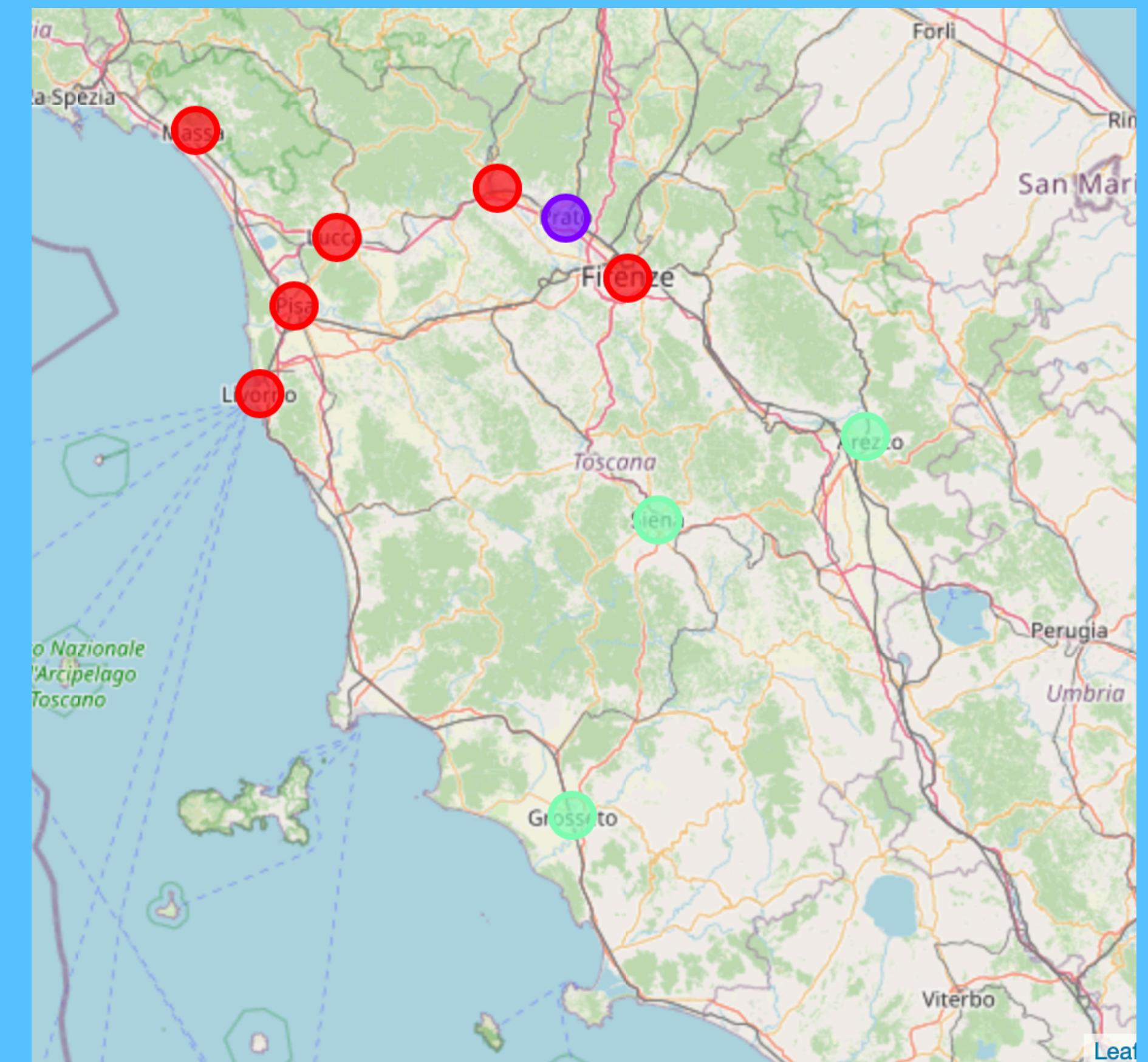
Clustering Approach based on Environmental Features

- Features Selected: Cafés Incidence, Schools Incidence
- Red : High School Density & Medium Café Density
- Purple : Medium School Density & High Café Density
- Light Blue : Medium School Density & Very High Café Density
- Pale Green : Low School Density & Low Café Density



Final Clustering Approach

- Features Selected: Old Age Index, Lethality Index, Percentage of Male, Percentage of Infected People, Population Density, Cafés Incidence, Schools Incidence
- **Red** : High Risk
- **Purple** : Potential Very High Risk
- **Green** : Low Risk



Discussion of Results & Conclusion

- Cafés incidence is not directly correlated with the total infected people, so we could think about re-opening them.
 - Schools incidence is a determinant factor in risk assessment.
 - Areas of risk are most common with near and densely populated cities, while countryside territories are safer.
 - If inhabitants are older, the percentage of infected people becomes smaller, so that younger people could unwillingly spread the virus.
- ★ The situation is rapidly evolving and it is always possible to re-apply the model when necessary, based on up-to-date information in datasets.