CRIMES IN ITALY

FROM 2006 TO 2020

ORAZIO PONTORNO 1000042940
GIUSEPPE PULINO 1000039397

AY 2021/2022
Prof. ORAZIO TOMARCHIO

01. BUSINESS QUESTIONS

Questions we want to answer through our analysis

02. DATASETS

Description of datasets used for the analysis

03. DIMENSIONAL FACT MODEL

04. ETL

Description of each step of the ETL procedure

05. DASHBOARDS

Accurate description of dashboards and their use

06. CONCLUSION

Answers to initial questions and personal opinions

01. BUSINESS QUESTIONS

1. What is the trend of crimes over time?
2. What is the most committed crime over time?
3. What is the trend of Education Rate, Economic hardship Rate and Unemployment over time?
4. How is it possible to improve the crime situation in Italy?
    1. What is the most correlated variable with the number of crimes?
    2. What is the variable which most influences the increase in the number of crimes?

02. DATASETS

POPULATION

Population19 and **Population20** tables (2304 and 128 records): these two tables provide information on the number of people for each province over time. Specifically, the first provides this data from 2004 to 2019, while the second provides the same data in 2020.

CRIMES

Crimes table (66291 records): provides information about the type of crimes declared by police from 2002 to 2020.

EDUCATION

Education table (2040 records): provides information on the number of people that have different education certifications we have information from 2004 to 2020 by a sample of 77000 households.

02. DATASETS

UNEMPLOYMENT

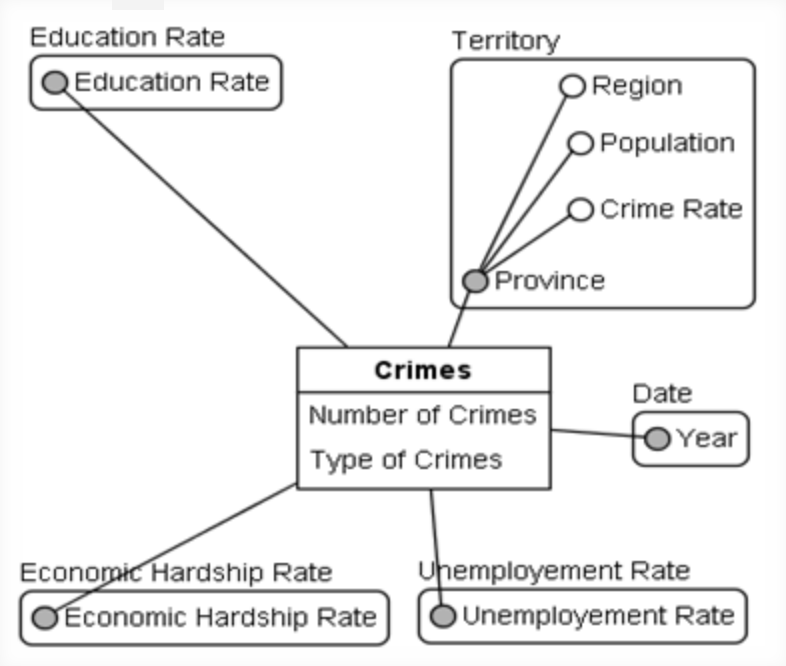
Unemployment table (3780 records): contains information regarding the percentage of people that didn't work in a specific region and year, we have information from 2002 to 2020 per each province.

ECONOMIC SITUATION

Economic situation table (2880 records): provides information regard to the way in which people can arrive at the end of year with their salary, we have information from 2002 to 2020 per each province, data are provided from a sample of 77000 households.

03. DIMENSIONAL FACT MODEL

The Dimensional Fact Model is an ad hoc and graphical formalism specifically devised to support the conceptual modelling phase.
The conceptual representation generated by the DFM consists of a set of fact schemata, which basically model facts, measures, dimensions, and hierarchies.



The **fact** is the 'CRIMES'.

There are two **measures**. Indeed, crimes can be measured by their *type* and the *number of committed*.

The **dimensions** of the Crimes are the Year in which crime where committed, the *Economic hardship rate*, the *unemployment rate* and the *Education rate*

A **dimensional attribute** is a property, with a finite domain, of a dimension. For example, the province can be described by the *region* to which it belongs and by its *population*.

04. ETL

The ETL procedure is made up of 3 steps:
Extraction, Transformation, Loading.

Our **Transformation** phase can be summarized in four steps:

- Input
- Cleaning and Computing
- Joining
- Output

We performed the Transform phase with Tableau Prep.

INPUT

In this step, we have chosen the columns to show for each table

CRIMES

- Province
- Years
- Type of crime
- Number of crimes

EDUCATION

- Region
- Years
- Education type
- Number of educated people

UNEMPLOYMENT

- Region
- Years
- Unemployment Rate

POPULATION19

- Province
- Years
- Number of people19

POPULATION20

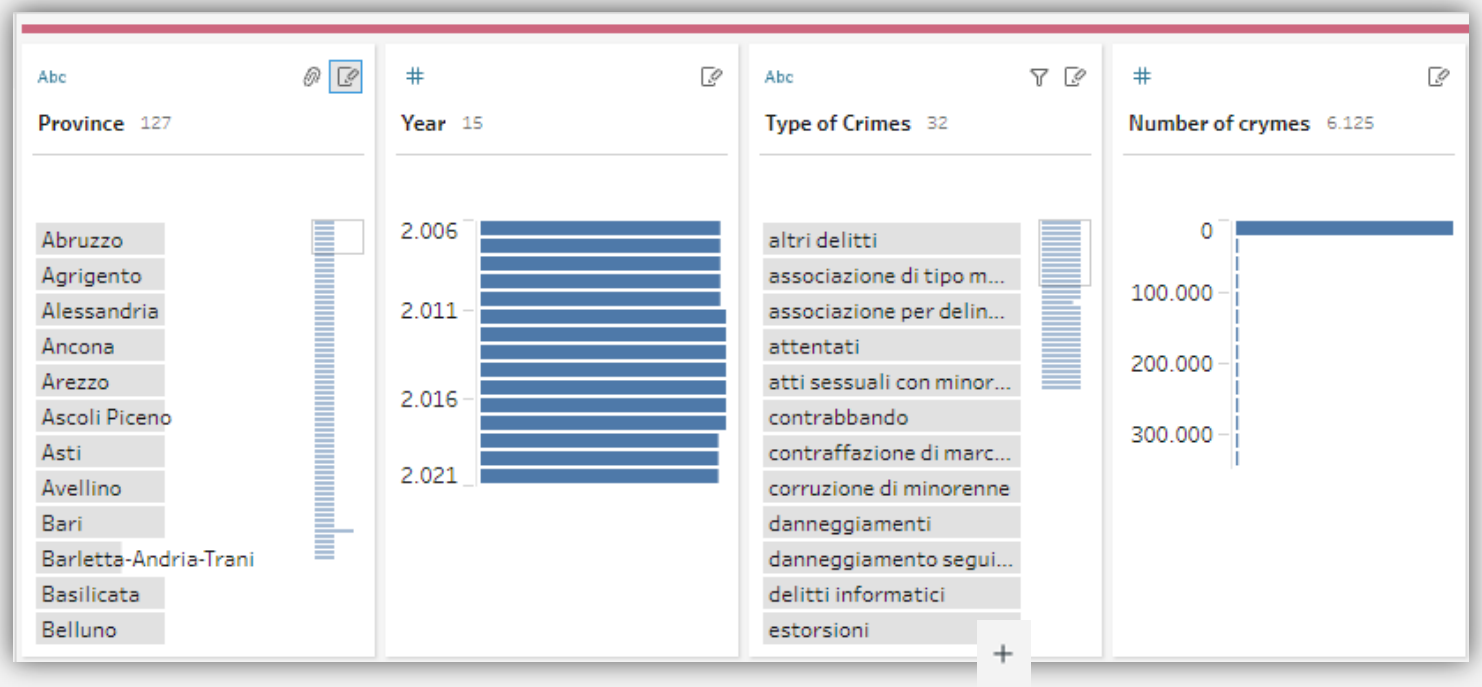
- Province
- Years
- Number of people20

ECONOMIC SITUATION

- Region
- Years
- Measure unit
- Type Data
- Economics Hardship Value

CLEANING AND COMPUTING

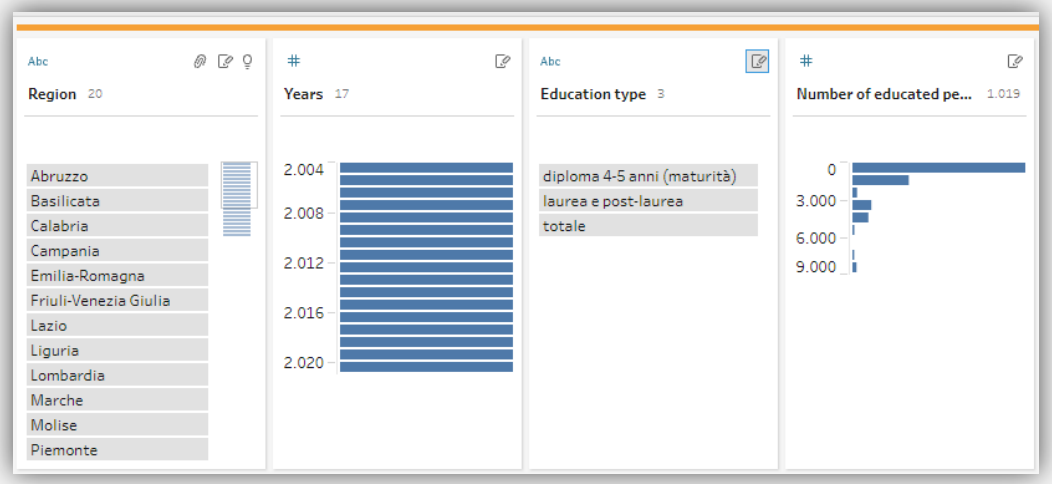
Crimes table has remained the same.

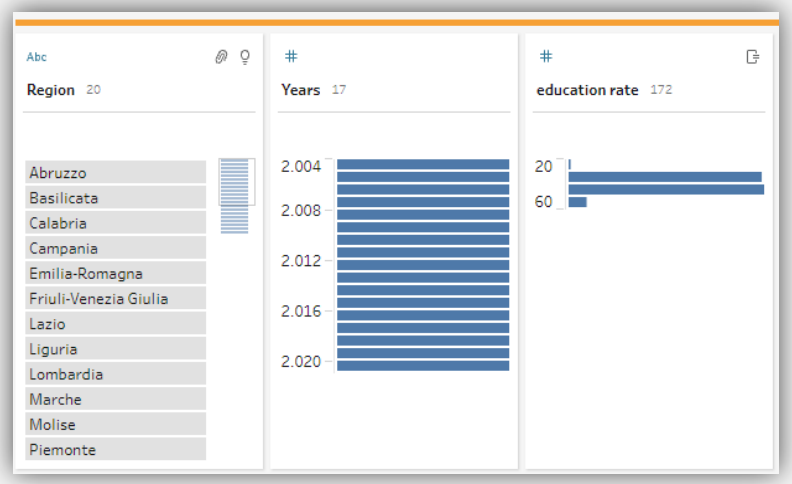


CLEANING AND COMPUTING

Education table underwent the following transformations:

- In the first step we have done a pivot from row to column of education type with number of educated people.
- In the second step we created a new column Education Rate adding up Diploma and Università and dividing by Totale.



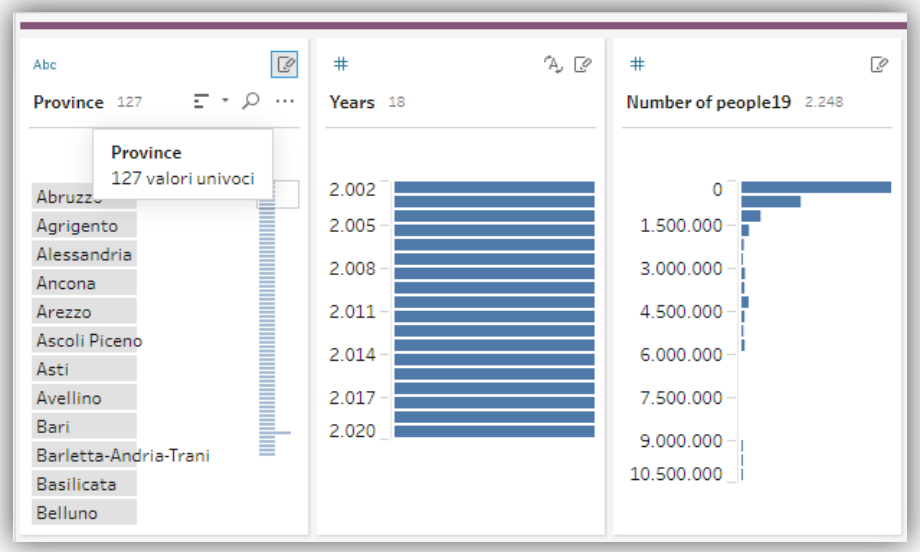


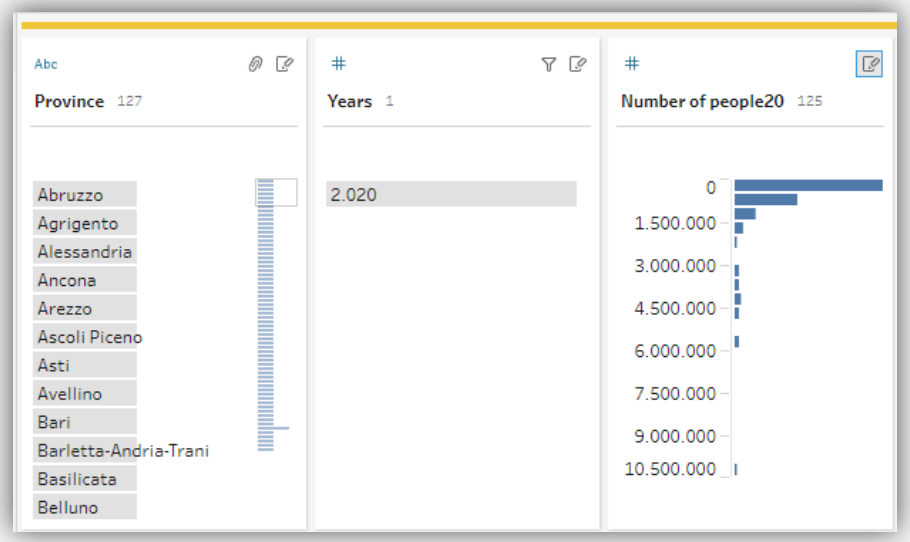
CLEANING AND COMPUTING

Unemployment table has remained the same

CLEANING AND COMPUTING

Population19 and **Population20** tables have been merged forming a new table with the same columns.

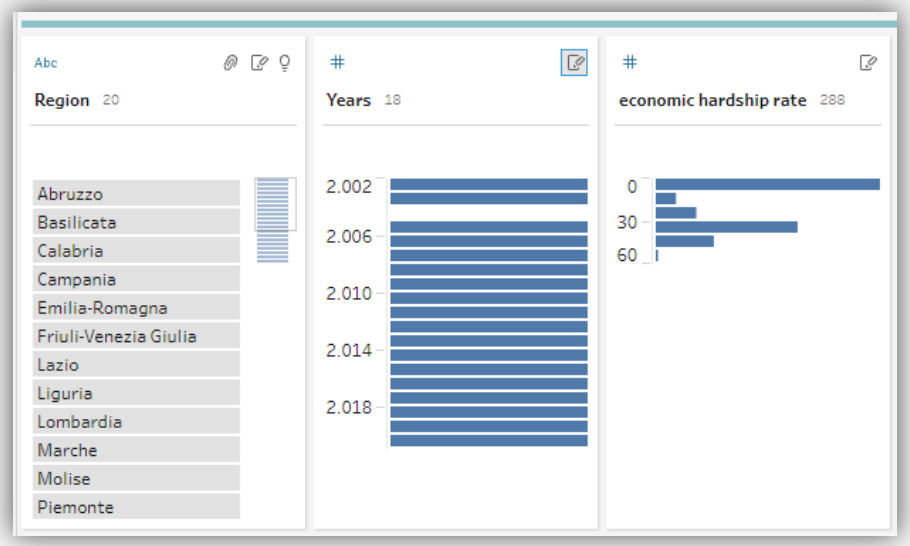




CLEANING AND COMPUTING

Economic table underwent the following transformations:

- From the field “Measure unit” has been considered only the percentage as measure unit and so the other rows have been excluded and from the field “Type Data”, the rows also have been excluded with types “Adeguate” and “Ottime”, after that the field has been removed
- In the second step we performed the aggregation per years and region related to economic hardship rate



JOINING

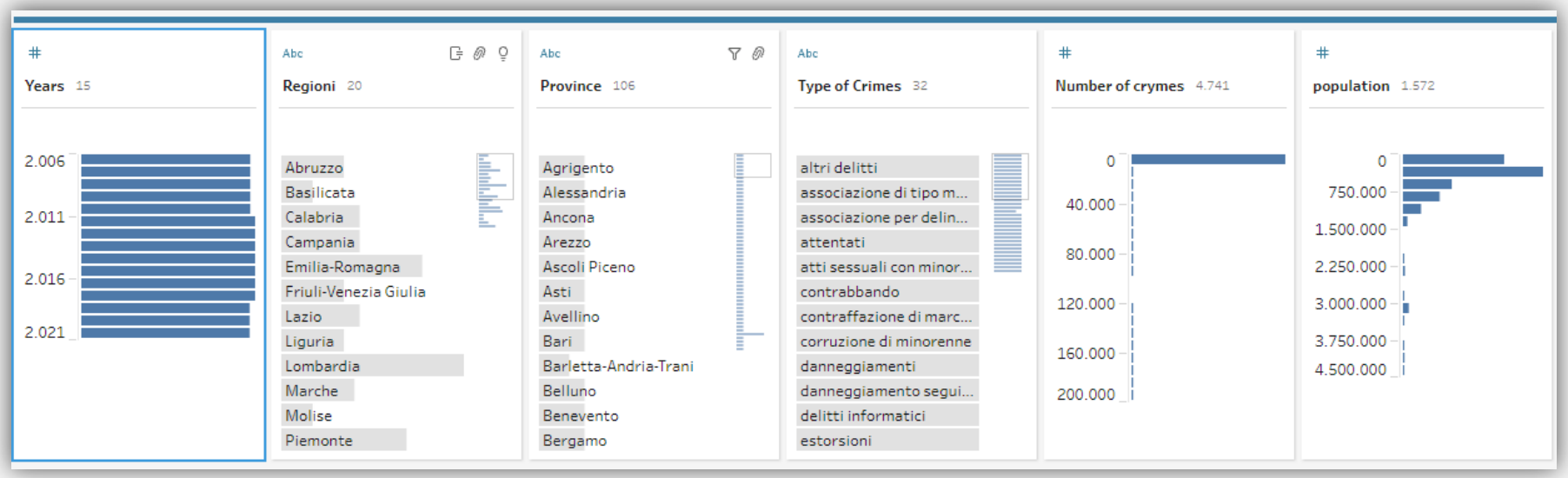
In this step, in order to obtain only two tables, the tables have been joined using inner join, in particular:

OUTPUT

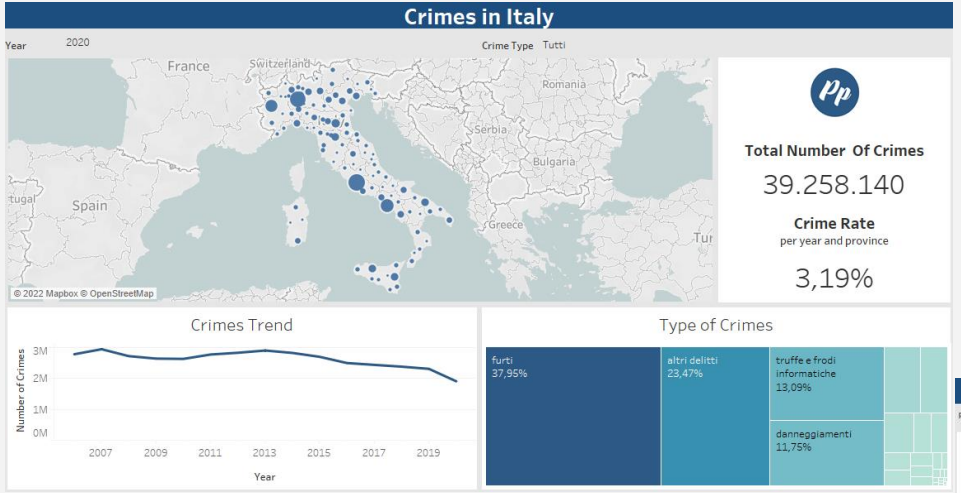
The final table **Information** has been obtained joining Economics Situation, Education and Unemployment using inner join by year and region

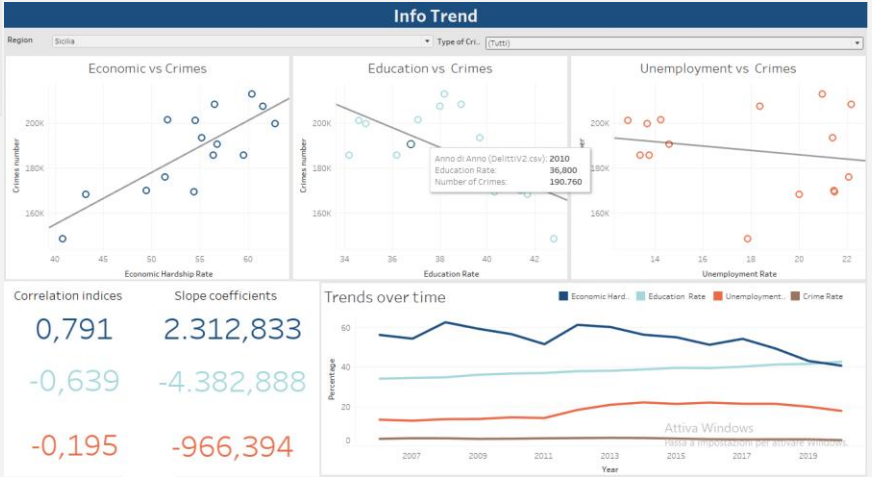
OUTPUT

The final table **Crimes** has been obtained joining Population and Crimes using inner join by year and province and adding a new computed field (**Region**) to assign to each province its own region



05. DASHBOARDS



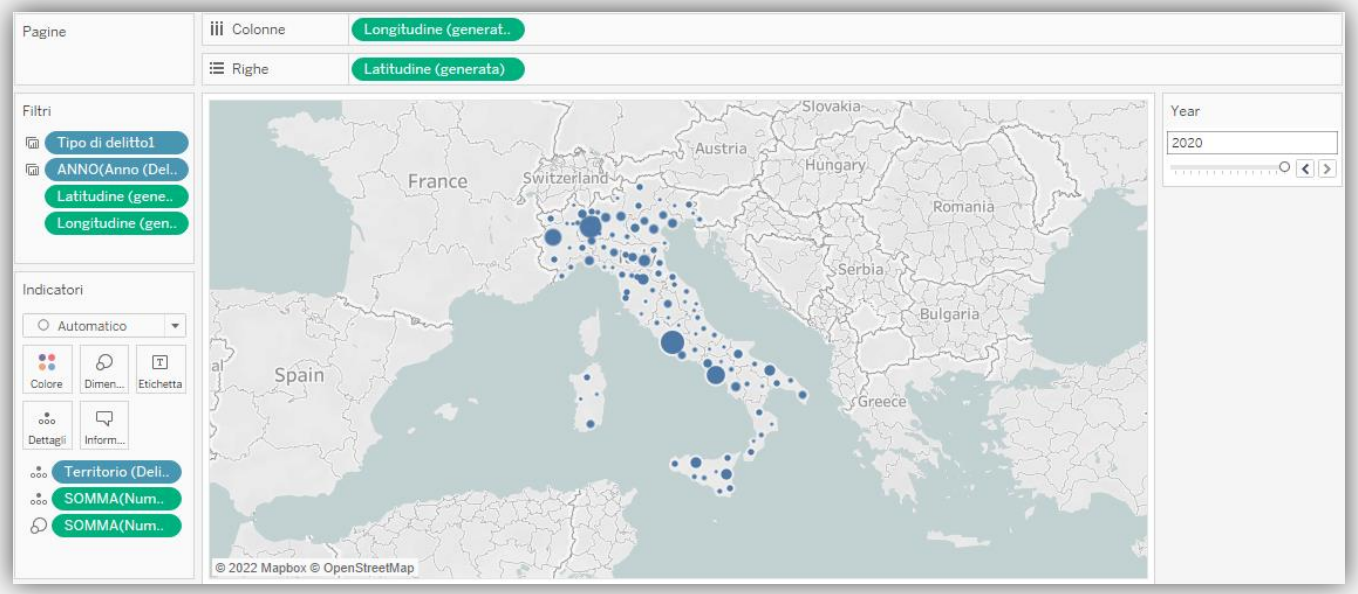


To perform the analysis, two dashboards were created. The first focuses on the distribution of crime in Italy; the second shows the relationships with the variables 'Education', 'Unemployment' and 'Economics hardship'.

Single Sheets of Dashboard 1

Crimes Map

The Crimes Map sheet is a map of Italy divided per province and it shows the total number of crimes through the dimension of the circles per province and year, selectable by the relative filter. The map is also used as filter per province for the other sheets.



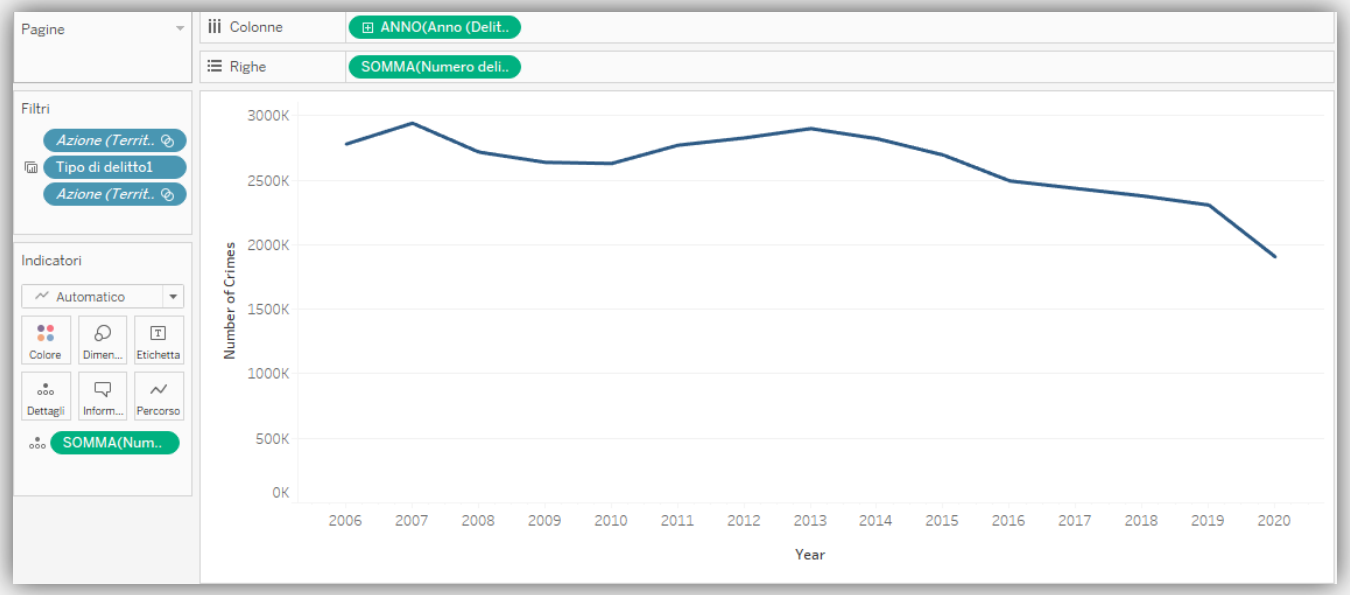
Why?

Map is the best chart when we want to show information using geocoded data for quick visualization and interpretation of data.

Single Sheets of Dashboard 1

Crimes Trend

The Crimes Trend sheet show the trend of the total number of crimes over time in Italy using a line chart. The number of crimes in a specific year is computed by summing the crimes of all provinces in the same year. Using the filters, it is possible shows only the trend of a specific crime and province.



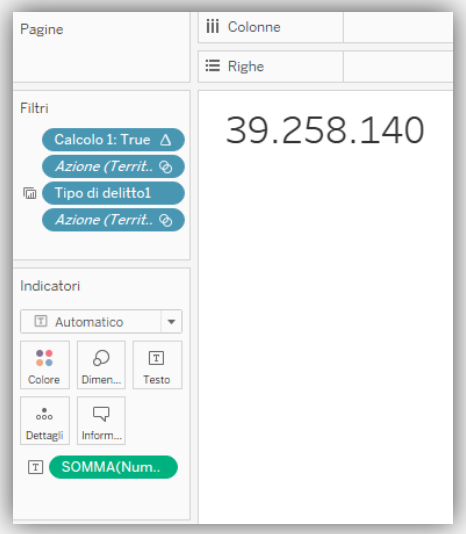
Why?

Line Chart is the most widely used chart is very useful when we want to show the trend in data over time.

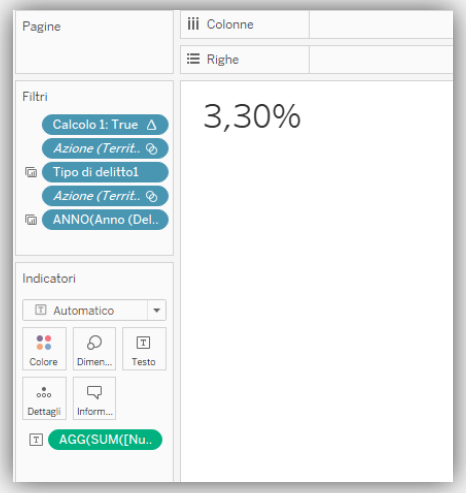
Single Sheets of Dashboard 1

Number of Crimes

The Number of crimes sheet shows the total number of crimes from 2006 to 2020 in Italy, calculated by summing all provinces' number of crimes.
Using the filters, it is possible to show only the total number of a specific crime and province.



Crime Rate



The Crime Rate sheet shows the crime rate in Italy, it has been computed by dividing the numbers of crimes by the population. Using the filters, it is possible shows the rate of a specific crime and province in a selected year.

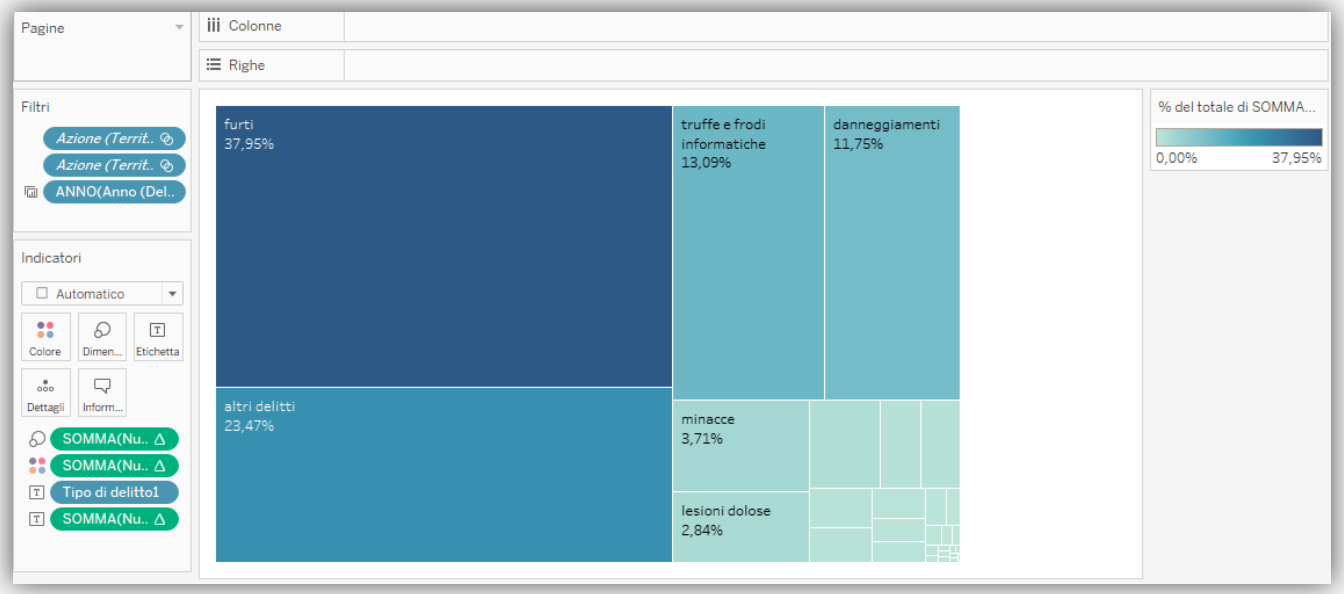
Why?

This type of sheets have been chosen to help users to have a quick visualization of the total number of crimes and crimes rate for each province.

Single Sheets of Dashboard 1

Crimes Type

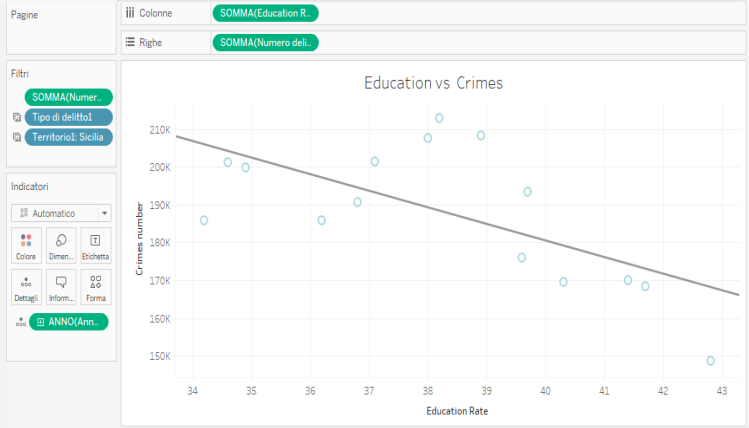
The Crime Type shows how the type of crimes are distributed as a proportion of a whole. It is possible also in this case filter data per province and year.



Why?

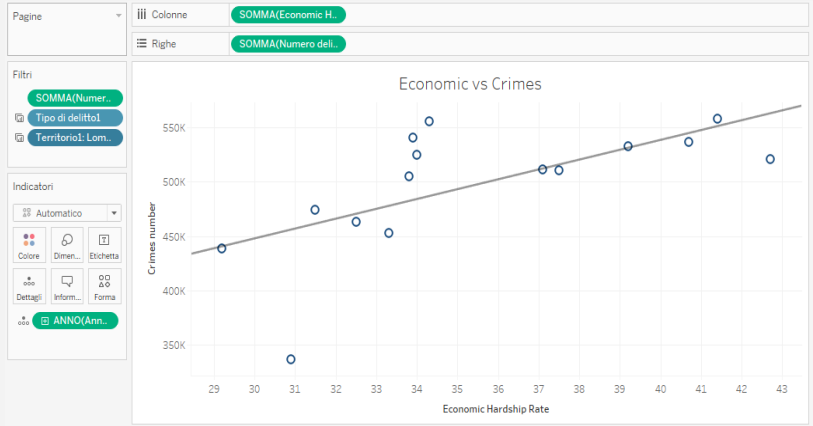
The treemap is useful when we want to show data as a proportion of a whole.

Single Sheets of Dashboard 2



Relations

The Relation sheets allows us to visualize the possible linear relation between the '*Unemployment Rate*' or '*Education Rate*' or '*Economic Hardship Rate*' and the 'Number of Crimes' of a region, who is selected by the filter. Each point is identified by a specific year; thus, in each sheet 15 points are present in the chart. Using the filter, it is possible analyse the possible presence of a linear relationship between a specific kind of crime in a specific region and the '*Unemployment Rate*' or '*Education Rate*' or '*Economic Hardship Rate*' in the same region.



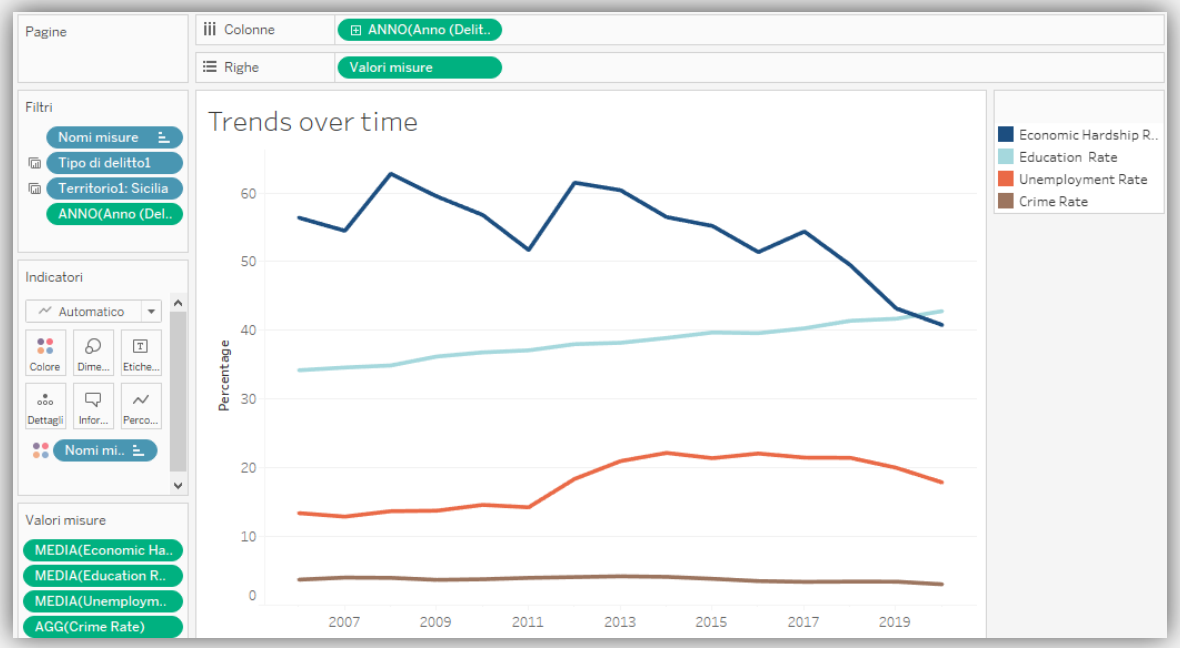
Why?

The Scatter plot is a powerful chart when we want to investigate the relationship between different variables.

Single Sheets of Dashboard 2

Trend over time

The Trend over time sheet show the trend of the variable '*Education Rate*', '*Unemployment Rate*' and '*Economic Hardship Rate*' over time in Italy using a line chart; the filters allow to show only the trend of a specific crime in a specific region.



Why?

Line Chart is the most widely used chart is very useful when we want to show the trend in data over time.

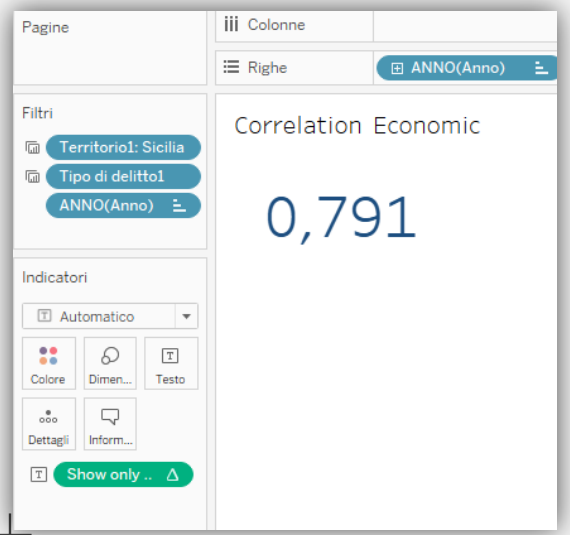
Single Sheets of Dashboard 2

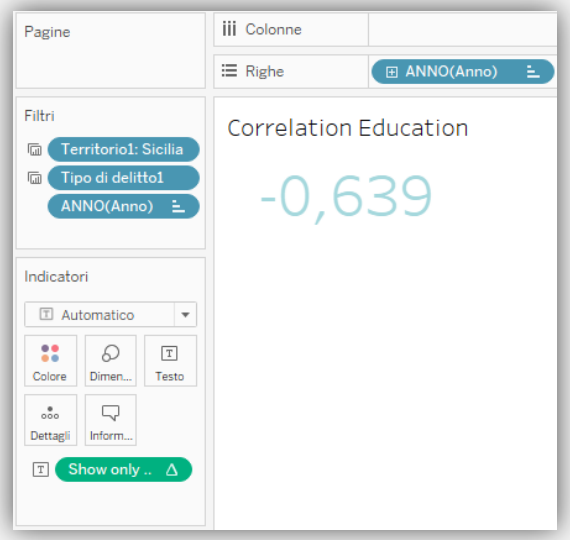
Correlation indices

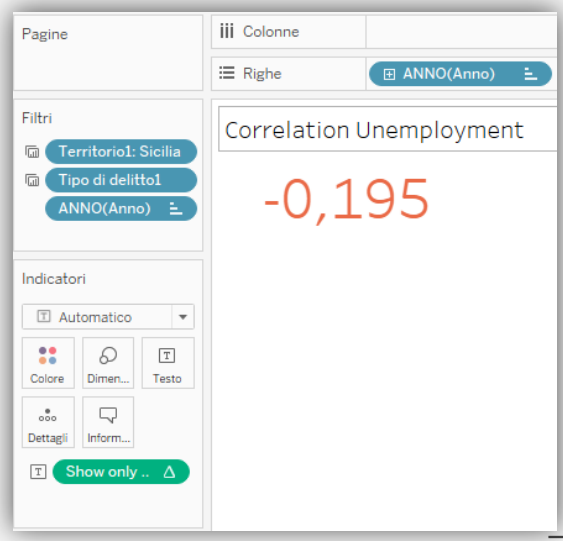
Here we have the correlation indices between each variable and the 'Number of Crimes' variable. Using the filters, it is possible to visualize these correlation indices selecting a specific kind of crime and a specific region.

The correlation index is a very powerful statistical tool that allows us to summarise the mutual interaction effect between two variables in a number.

Why?







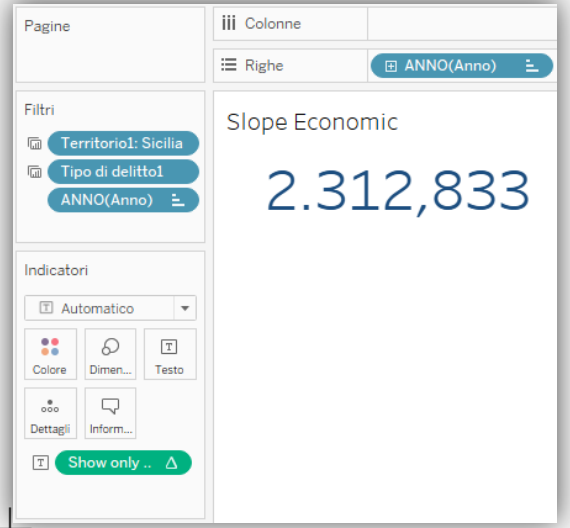
Single Sheets of Dashboard 2

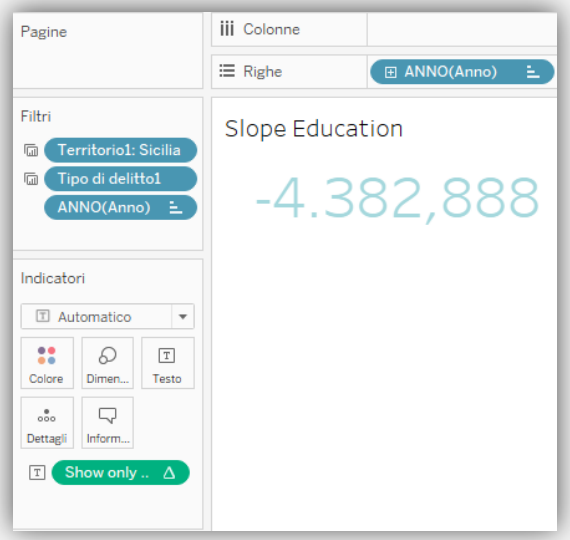
Slope coefficients

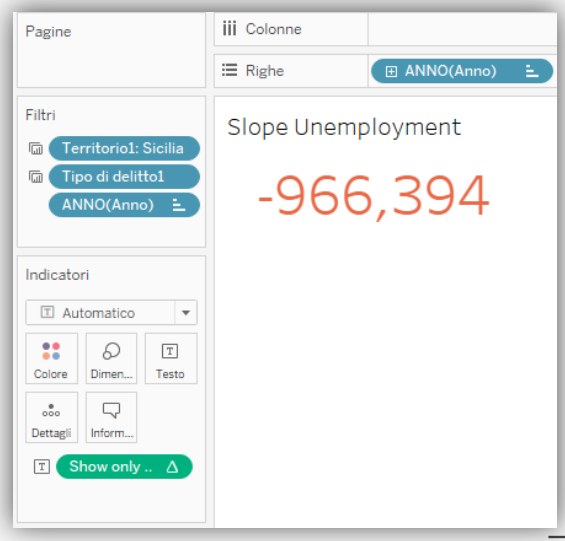
Here the slope regression coefficients calculated in the Regression sheets are shown.
As well as the scatterplots, these sheets are subjected to 'Crime type' and 'Region' filters; using them, it is indeed possible to select a specific kind of crime and a specific region.

The slope regression coefficient allows us to summarise the possible influence the independent variable has on the dependent variable in a single number.

Why?





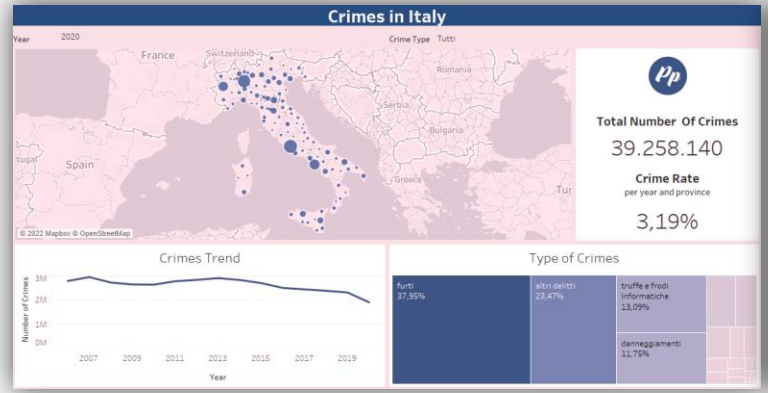


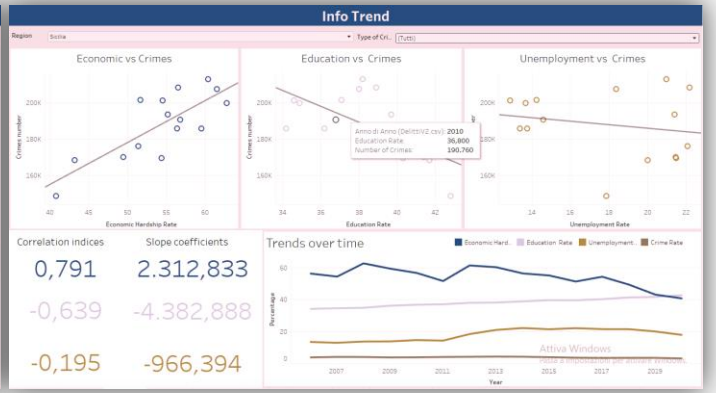
Color Blind Test

Why did we use these colors? The Color Blind Test is a useful tool for testing whether the colors chosen for the dashboards can be seen by all kinds of people, even color blind people.

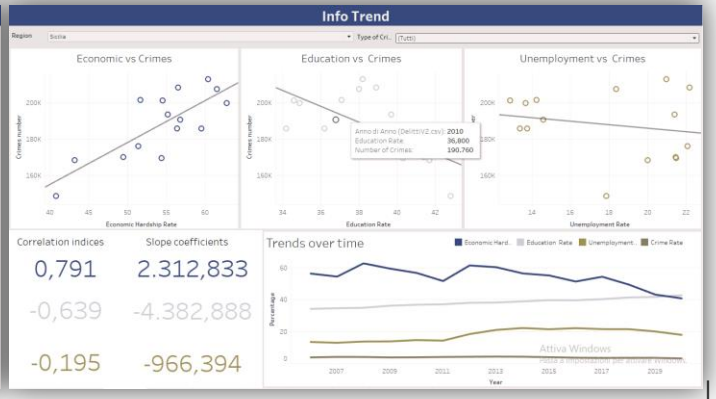
Why?

Deutan Test



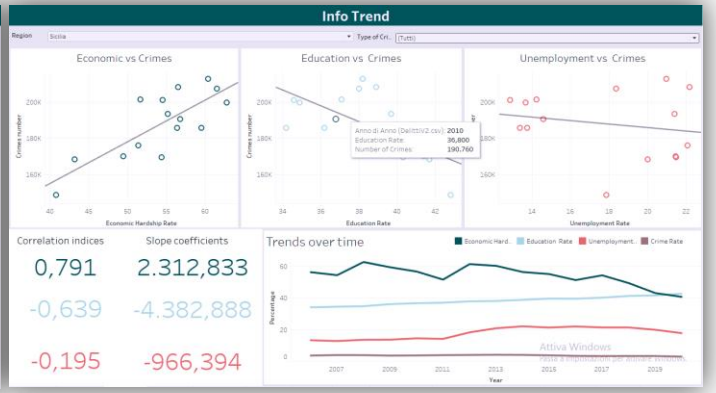


Protan Test

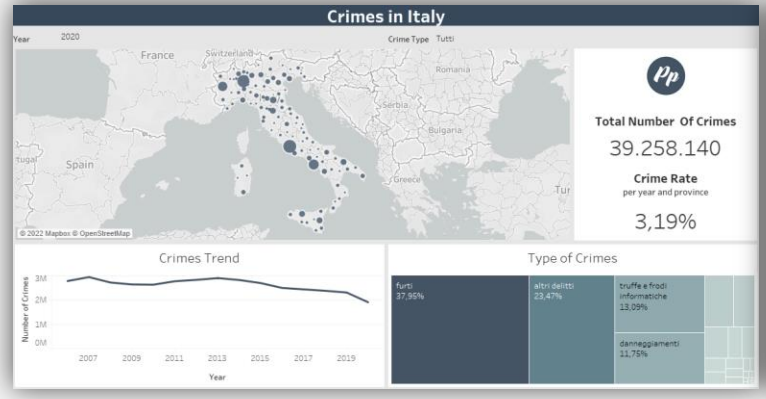


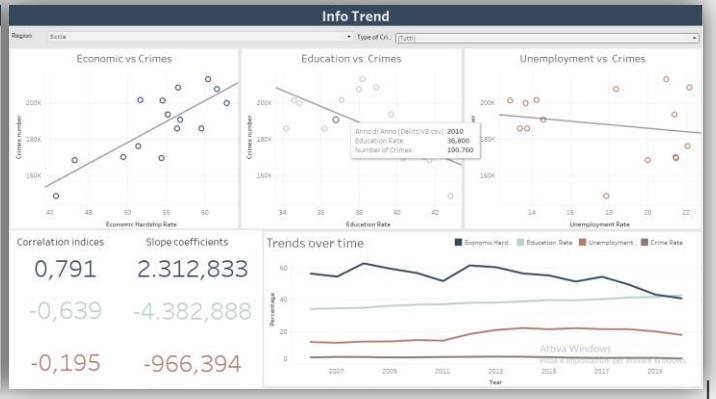
Color Blind Test

Tritan Test



Black & White Test





Why did we use these colors? The Color Blind Test is a useful tool for testing whether the colors chosen for the dashboards can be seen by all kinds of people, even color-blind people.

Why?

06. CONCLUSION

The conclusions and results of our analysis will now be exposed through the use of the dashboards

THANK YOU FOR YOUR ATTENTION AND CONSIDERATION