

THE ITALIAN LABOR MARKET: A SURVEY OF REGIONAL DYNAMICS AND AN ANALYSIS OF YOUTH EMPLOYEES AND NEETS

GIUSEPPE LOMBARDIA
MANUELA LO GIUDICE



BIG DATA ANALYTICS
PROF. ORAZIO TOMARCHIO

Index

INTRODUCTION	2
BUSINESS QUESTIONS	2
DATASETS	3
DIMENSIONAL FACT MODEL	4
ETL	5
INTRODUCTION AND OVERALL ETL STRUCTURE	5
TRANSFORMATION: INPUT STEP	6
TRANSFORMATION: CLEANING AND COMPUTING STEP	7
TRANSFORMATION: JOINING STEP	10
TRANSFORMATION: OUTPUT STEP.....	11
DASHBOARDS.....	12
DASHBOARD 1.....	12
DASHBOARD 2	15
DASHBOARD 3.....	18
DASHBOARD 4.....	20
DASHBOARD 5	21
DASHBOARD 6	21
DASHBOARD 7	22
DASHBOARD 8	23
DATA STORY.....	23
ANALYSIS REPORT.....	31

COLOR BLIND TEST35

o BLACK AND WHITE TEST35

o DEUTAN TEST35

o PROTAN TEST.....35

o TRITAN TEST..... 36

CONCLUSION.....36

INTRODUCTION

This report discusses the occupational status in Italy from 2005 to 2020. The report focuses on the employment rate and unemployment rate by gender and in each region, with a focus on Sicily. From this general analysis we move on to a more specific one dealing with Neets, that is, young people not working and not studying, and youth employment. The approach used to calculate the various employment and unemployment rates is based on the total population aged 15 and over. Some factors such as education and some economic indicators such as GDP, GDP per capita and domestic labor income are related to the two rates mentioned.

As for occupational status there are within it different categories such as the labor force which is formed by the unemployed and the employed, and finally the inactive formed by those who are not looking for and not available for work and those who are part of the gray area of inactivity.

BUSINESS QUESTIONS

- 1) How has the trend of total employment and unemployment rate varied over the years in Italy?
- a) How has the trend of female employment and unemployment rate varied over the years in Italy?

b) How has the trend of male employment and unemployment rate varied across over the years in Italy?

c) And what about Sicily?
- 2) What is the trend of other occupational conditions such as inactive, not seeking and unwilling to work and gray area of inactivity?
- 3) What is the best region in terms of youth employment and with a lower rate of Neets?
- a) What is the correlation between youth employment rate and Neets?

b) And what about Sicily?

- c) What is the picture of the Italian regions, in terms of Neets and youth employment, in 2019, 2020 and before the 2008 crisis?
- 4) What is the relationship between GDP and employment and unemployment rates indifferent regions?
- a) And what about Neets?
- b) And what about youth employment?
- 5) What is the relationship between education rate and employment and unemployment rate?
- a) And what about Neets?
- b) And what about youth employment?
- 6) How can the situation of Neets and youth employment in Italy be improved?

DATASETS

The report is based on data relating to the Occupation rate in Italy from 2005 to 2020.

From the data sources https://ec.europa.eu/eurostat/databrowser/explore/all/all_themes and <http://dati.istat.it>, **six tables** in "csv" format have been selected:

- **Gross domestic product (GDP) at current market price by NUTS 2 regions** (336 records) table: This table from the StatisticalOffice of the European Union (Eurostat) presents the gross domestic product(GDP) at market prices for NUTS Level 2 regions in Italy, expressed in millions of Euros. The basic data used in the calculation of the regional GDP estimates, which are produced by Eurostat based on a harmonized methodology, come from individual National Statistical Institutes (NSIs). The NSI statistics come from a variety of sources, including government administrative data, censuses, and business and household surveys.
- **Per-capita values:** (2688 records) table: In this table are values of GDP per capita and domestic labor income taken from the Eurostat website. Data are taken at market price for NUTS Level 2 regions in Italy, expressed in thousands of euros.
- **Population** (7056 records) table: In this table are data by occupational status of the population aged 15 years and older. They are further divided by gender and by region precisely by NUTS Level 2 so as to get a clear picture. The data are expressed in thousands.
- **Education** (6048 records) table: In this table there are data by education and by type of education in the population aged 15 years and older. They are further divided by gender and by region precisely by NUTS Level 2. The data are expressed in thousands.
- **Youth Employees** (1008 records) table: This table comes from the EUROSTAT website and contains the youth employment rate by region; it is divided by NUTS Level 2 in Italy. Again they are divided by gender and total. Values are expressed as percentages.
- **NEET** (1008 records) table: This table comes from the EUROSTAT website and contains the neet rate by region; the division is by NUTS Level 2 in Italy. Again they are divided by gender and total. Values are expressed as percentages.

DIMENSIONAL FACT MODEL

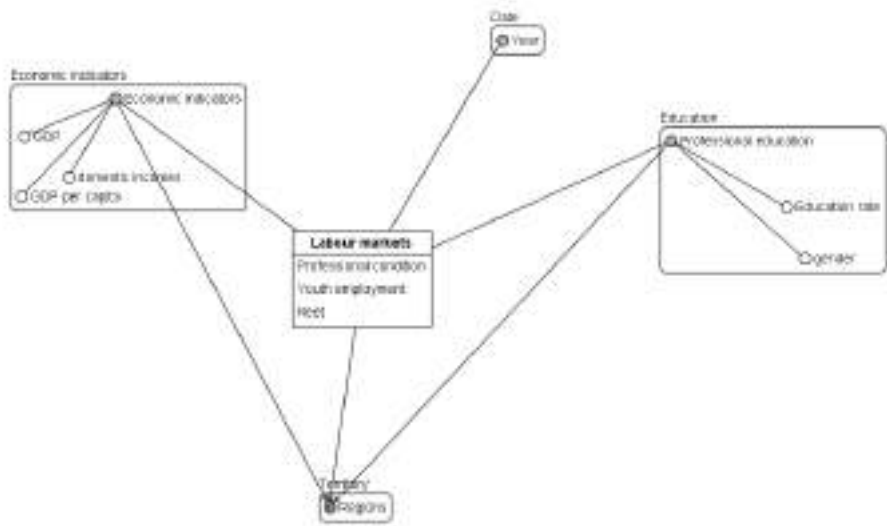
The Dimensional Fact Model is an ad hoc and graphical formalism specifically devised to support the conceptual modelling phase. The Dimensional Fact Model, or DFM, is extremely intuitive and can be used by analysts and non-technical users as well. The conceptual representation generated by the DFM consists of a set of fact schemata, which basically model facts, measures, dimensions, and hierarchies.

A **fact** is a concept relevant to decision-making processes. It typically models a set of events taking place within a company. In our case, the fact is the ‘Labour markets’.

A **measure** is a numerical property of a fact and describes a quantitative attribute that is relevant to analysis. In our case, the labor market can be measured by the types of professional conditions and their rates.

A **dimension** is a property, with a finite domain, that describes an analysis coordinate of the fact. A fact generally has multiple dimensions that define its minimum representation granularity. The dimensions of the Labour market are the Year where the analyzed rates are recorded, the Economic Indicators rate of regions where the analyzed rates are recorded and the education rate by region where the rates are analyzed.

A **dimensional attribute** is a property, with a finite domain, of a dimension. For example, economic indicators can be described by GDP, GDP per capita, etc...



In the world of data warehousing, if you need to bring data from multiple different data sources into one, centralized database.

ETL tools also make it possible to migrate data between a variety of sources, destinations, and analysis tools. As a result, the ETL process plays a critical role in producing business intelligence and executing broader data management strategies.

The ETL process is comprised of 3 steps that enable data integration from source to destination: data extraction, data transformation, and data loading.

STEP 1: Data Extraction

Most businesses manage data from a variety of data sources and use several data analysis tools to produce business intelligence.

Before data can be moved to a new destination, it must first be extracted from its source — such as a data warehouse or data lake. In this first step of the ETL process, structured and unstructured data is imported and consolidated into a single repository.

STEP 2: Data Transformation

During this phase of the ETL process, rules and regulations can be applied that ensure data quality and accessibility. You can also apply rules to help your company meet reporting requirements. The process of data transformation is comprised of several sub-processes:

Cleansing — inconsistencies and missing values in the data are resolved.

Standardization — formatting rules are applied to the dataset.

Deduplication — redundant data is excluded or discarded.

Verification — unusable data is removed, and anomalies are flagged.

Sorting — data is organized according to type.

Other tasks — any additional/optional rules can be applied to improve data quality.

Transformation is generally considered to be the most important part of the ETL process.

Data transformation improves data integrity — removing duplicates and ensuring that raw data arrives at its new destination fully compatible and ready to use.

STEP 3: Data Loading

The final step in the ETL process is to load the newly transformed data into a new destination (data lake or data warehouse). Data can be loaded all at once (full load) or at scheduled intervals (incremental load).

Full loading — In an ETL full loading scenario, everything that comes from the transformation assembly line goes into new, unique records in the data warehouse or data repository. Though there may be times this is useful for research purposes, full loading produces datasets that grow exponentially and can quickly become difficult to maintain.

Incremental loading — A less comprehensive but more manageable approach is incremental loading. Incremental loading compares incoming data with what's already on hand and only produces additional records if new and unique information is found. This architecture allows smaller, less expensive data warehouses to maintain and manage business intelligence.

Transformation: Input Step

On Tableau Prep Builder, 6 tables have been inserted, and some fields have been excluded.

- **Gross domestic product (GDP) at current market price by NUTS 2 regions** table:

the columns "DATAFLOW", "LAST UPDATE", "unit", and "OBS_FLAG" have been excluded from the flow because they are not useful for data analysis and at the end are remained the following 3 columns:

- o Regions
- o Year
- o PIL in Million

- **Per capita values** table:

the columns "ITTER107", "TIPO_DATO_CN1", "VAL", "CORREZ", "Correzione", "T_BIS", "Edizione", "TIME", "Flag Codes", "Flags" have been excluded from the flow because they are not useful for data analysis and at the end are remained the following 4 columns:

- o Regions
- o Aggregate type
- o Year
- o Value

- **Population** table:

the columns "ITTER107", "TIPO_DATO_FOL", "Tipo dato", "SEXISTAT1", "ETA1", "Classe di età", "CONDIZIONE_PROF", "TIME", "Flag Codes", "Flags" have been excluded from the flow because they are not useful for data analysis and at the end are remained the following 5 columns:

- o Regions
- o Year
- o Professional Condition
- o Sex
- o Value of Professional Condition

- **Education** table:

the columns "ITTER107", "TIPO_DATO_FOL", "Tipo dato", "SEXISTAT1", "ETA1", "TITOLO_STUDIO", "TIME", "Flag Codes", "Flags" have been excluded from the flow because they are not useful for data analysis and at the end are remained the following 5 columns:

- o Regions
- o Educational Qualification
- o Year
- o Sex
- o Value of Educational qualification

- **Youth Employees** table:

the columns "DATAFLOW", "LAST UPDATE", "freq", "age", "unit", and "OBS_FLAG" have been excluded from the flow because they are not useful for data analysis and at the end are remained the following 4 columns:

- o Regions
- o Sex
- o Year
- o Youth Employees rate

- **NEET** table:

the columns "ITTER107", "TIPO_DATO_FOL", "Tipo dato", "SEXISTAT1", "ETA1", "TIME", "Flag Codes", "Flags", "Classe di età" have been excluded from the flow because they are not useful for data analysis and at the end are remained the following 4 columns:

- o Regions
- o Year
- o Sex
- o Neet rate

Transformation: Cleaning and Computing Step

Gross domestic product (GDP) at current market price by NUTS 2 regions table underwent the following transformation:

- We have removed the not useful field.
- Change the type of "TIME_PERIOD" to "Date type".
- Group values of geo with 22 values replaced by 21 values.

- After that, we renamed the columns: “OBS_VALUE” to “PIL in Million”, “TIME_PERIOD” to “Year”, and “geo” to “regions”.



Per capita values table underwent the following transformation:

- We have removed the not useful field.
- Group values from “Territorio of Provincia autonoma Bolzano/ Bozen” replaced “Provincia autonoma Bolzano”.
- Change the type of “Seleziona period” to “Date type”.
- Filter “Valutazione” and “Tipo aggregato” to exclude “multiple” values.
- Renamed field: “Territorio” to “Regions”, “Seleziona periodo” to “Year”, “Tipo aggregato” to “Aggregate type”.
- Group values from “Regions” with 2 values replaced.



Education table underwent the following transformation:

In the first step, we made a general cleaning of our dataset as follows:

- We have removed the not useful field.
- Group values of “Territorio” and “Sex” with 3 values replaced.
- Filtered “Territorio” excluding “Territorio of Provincia autonoma Bolzano/Sudtirolo”.
- Renamed field: “Territorio” to “Regions”, “Sesso” to “Sex”, “Seleziona periodo” to “Year”, “Titolo di studio” to “Educational qualification”, “Value” to “Value of Educational qualification”.
- Change the type of “Seleziona period” to “Date type”.
- Group values of “Titolo di studio” with 7 values replaced by 6 values.
- Group values from “Territorio of Provincia autonoma Bolzano/ Bozen” replaced “Provincia autonoma Bolzano”.

In the second step, we did the pivot from columns to rows of the field Educational qualification and with the cleaning step, we transformed values from millions to percentages through the calculated field.

In the end, we did another pivot from rows to columns to create respectively the fields Education and Education rate.



Population table underwent the following transformation:

In the first step, we made a general cleaning of our dataset as follows:

- We have removed the not useful field.
- Filter “Seleziona periodo” to exclude “multiple” value.
- Renamed field: “Territorio” to “Regions”, “Sesso” to “Sex”, “Seleziona periodo” to “Year”, “Condizione professionale” to “Professional Condition”, “Value” to “Value of Professional Condition”.
- Filter “Regions” to remove fields like “Mezzogiorno”, “Nord”, “Nord-est”, “Nord-ovest”, “Centro”, “Italia” and Null values.
- Group values of “Regions” with 3 values replaced.
- Group values of “Sex” with 3 values replaced.
- Group values of “Professional Condition” with 7 values replaced.
- Change the type of “Year” to “Date type”.
- Group values from “Territorio of Provincia autonoma Bolzano/ Bozen” replaced “Provincia autonoma Bolzano”.

In the second step, we did the pivot from columns to rows of the field Professional condition and with the cleaning step, we transformed values from millions to percentages through the calculated field.

Finally, we pivoted from rows to columns to create the professional condition and value of the professional condition fields, respectively.



Youth Employees table underwent the following transformation:

- We have removed the not useful field.
- Group values of “Sex” with 3 values replaced.
- Change the type of “TIME_PERIOD” to “Date type”.
- Renamed field: “Territorio” to “Regions”, “Sesso” to “Sex”, “Seleziona periodo” to “Year”, “OBS_VALUE” to “Youth Employees rate”.

- Group values of geo with 21 values replaced.



NEET table underwent the following transformation:

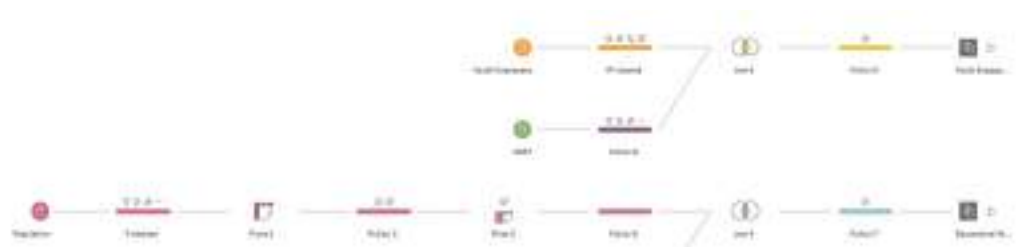
- We have removed the not useful field.
- Change the type of “Seleziona period” to “Date type”.
- Filter “Territorio” to exclude “multiple” values.
- Group values of “Territorio” with 2 values replaced.
- Filtered “Territorio” excluding “Territorio of Provincia autonoma Bolzano/Sudtirolo”.
- Group values of “Sex” with 3 values replaced.
- Filter “Regions” to exclude “Centro” value.
- Group values from “Territorio of Provincia autonoma Bolzano/ Bozen” replaced “Provincia autonoma Bolzano”.
- Group values of “Regions” with 4 values replaced by 2 values.



Transformation: Joining Step

In this step to obtain only three tables, the tables have been joined using inner joins in particular:

- In the final table, **Economic Indicators** have been obtained joining GDP and Per capita values using inner join by year and region.
- The final table **Educational Attainment and Population Trends** has been obtained joining Population and Education using inner join by year, sex, and region.
- The final table **Youth Engagement and NEET Rates** has been obtained joining Youth Employees and NEET using inner join by year, sex, and region.





Transformation: Output Step

Finally, we have obtained three tables:

- **Economic Indicators** table resulted in **2688** rows and only the following **5** columns:
 - Regions
 - Year
 - PIL in Million
 - Type of aggregates
 - Value
- **Educational Attainment and Population Trends** table resulted in **24.192** rows and only the following **7** columns:
 - Education
 - Education rate
 - Value of professional condition
 - Professional condition
 - Regions
 - Year
 - Sex
- **Youth Engagement and NEET Rates** table resulted in **1008** rows and only the following **5** columns:
 - Regions
 - Sex
 - Year
 - Youth Employees rate
 - Neet rate

That we will put in relation to Tableau.



Economic Indicators



Educational Attainment



Youth Engagement

DASHBOARDS



Dashboard 1 in Detail

- Professional condition map

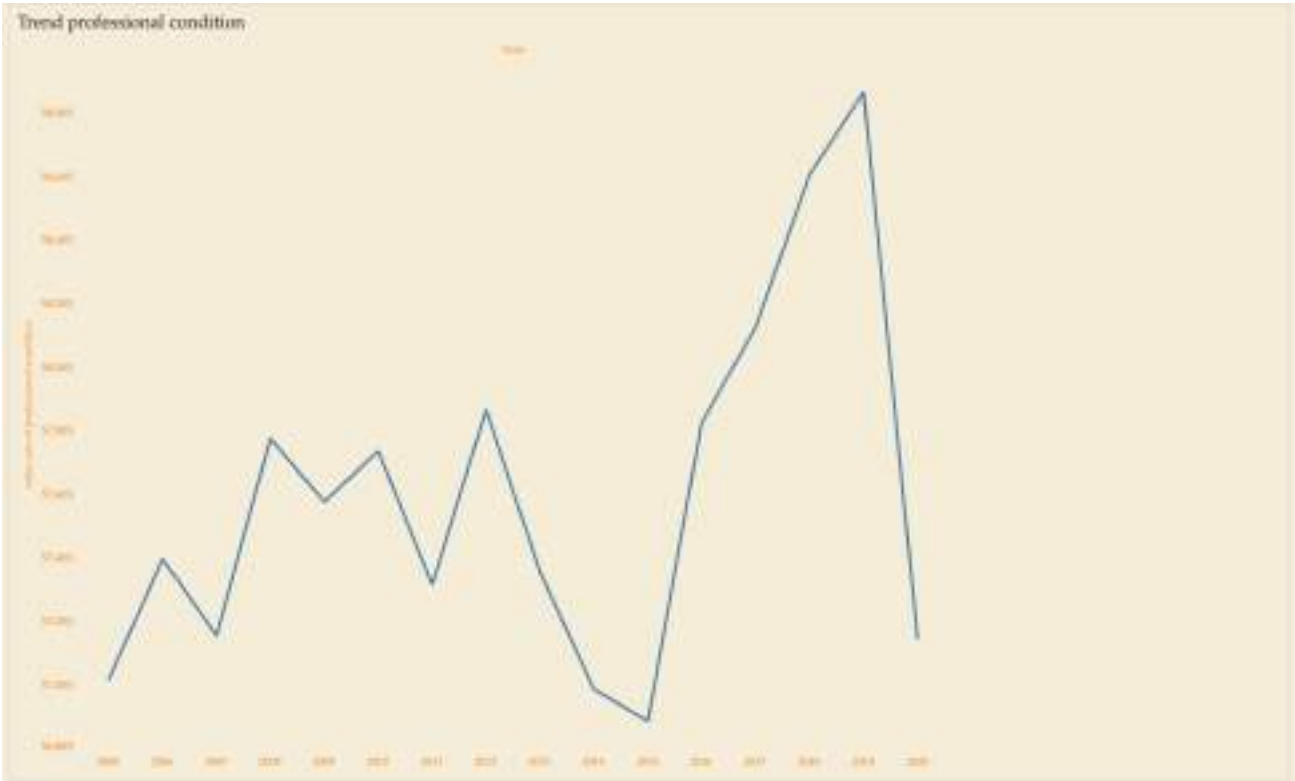
Map is the best chart when we want to show information using geocoded data for quick visualization and interpretation of data. The professional condition map sheet is a map of Italy divided per regions and autonomous provinces(Trento and Bolzano). This map shows, through the dimension of the circles, the percentage of professional condition per region, year, gender. The map is also used as

filter per region for the other sheets. For geocoding I used the NUTS 2 criteria, for the autonomous provinces I manually added the NUTS 2 codes, i.e., ITH10 and ITH20, in the edit positions.



• Professional condition trend

Line chart is the most used chart as it is very useful for showing trends over time. The occupational condition trend shows the trend of various conditions with the ability to filter by gender, region, and professional condition.





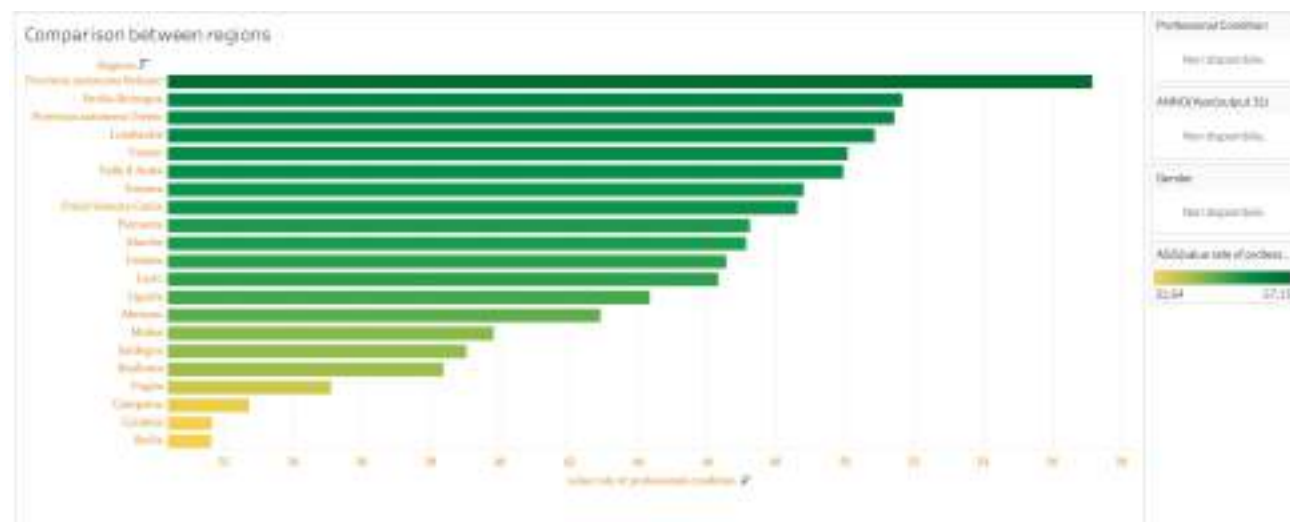
The formula **SUM ([Value of Professional Condition])/COUNT ([Value of Professional Condition])** indicates the arithmetic mean of the values present in the "Value of Professional Condition" column.

Specifically, the sum of the items in the column is divided by the total number of items in the same column. The result is an indicator of the "central tendency" of the distribution of the data in the column, that is, where the values are concentrated.

This formula could be used to analyze different aspects of the data in the column, for example, to evaluate the quality of service or the performance of the operators based on the value of the variable "Value of Professional Condition."

If the formula **SUM ([Value of Professional Condition])/COUNT ([Value of Professional Condition])** returns the exact value of the series, this means that all the values in the "Value of Professional Condition" column were taken into account and the formula was calculated correctly. In other words, the result is the arithmetic mean of the values in the column, which represents the exact numerical value of the central tendency of the data in the column. This formula is used for all chart about professional condition.

Professional condition bar chart



By using the bar chart, we can compare professional conditions across regions or autonomous provinces. The graph can be filtered by gender, year, and professional condition.

A horizontal bar graph may be preferred over other types of graphs, such as the vertical bar graph, for a few reasons. First, horizontal bars may be easier to read and compare than vertical bars because the length of the bars is horizontally aligned and comparison between bars is easier. Also, a horizontal bar graph can be useful when you have many categories or variables to display, as horizontal bars can be organized over multiple rows without losing their readability.

- **Number of professional conditions**



This sheet tells us the data not percentages but numbers of occupational conditions, filtered by year, region, and gender.

Employed + Unemployed = Labor force

Gray zone of the inactivity + people not available = Inactivity

Dashboard 2



The second dashboard analyzes the Education by regions and its compositions; it contains the following sheets: education info map, education info trend, education tree map, and finally a comparison between regions.

The following filters are available in the dashboard:

- The "Year" filter, located in the upper right corner, displays all data for the selected year.
- The "education" filter allows the user to view the data collected for each individual type of education.
- The sheet 'Education Map' is also used as filter; through it is indeed possible to view data related to the selected region, or autonomous province.
- The "Gender" filter shows data from women, men, or the total. Women's data is calculated on the total of women, same for men, however, for total is calculated on the sum of both sexes.

Dashboard 2 in Detail

- **Education info map**



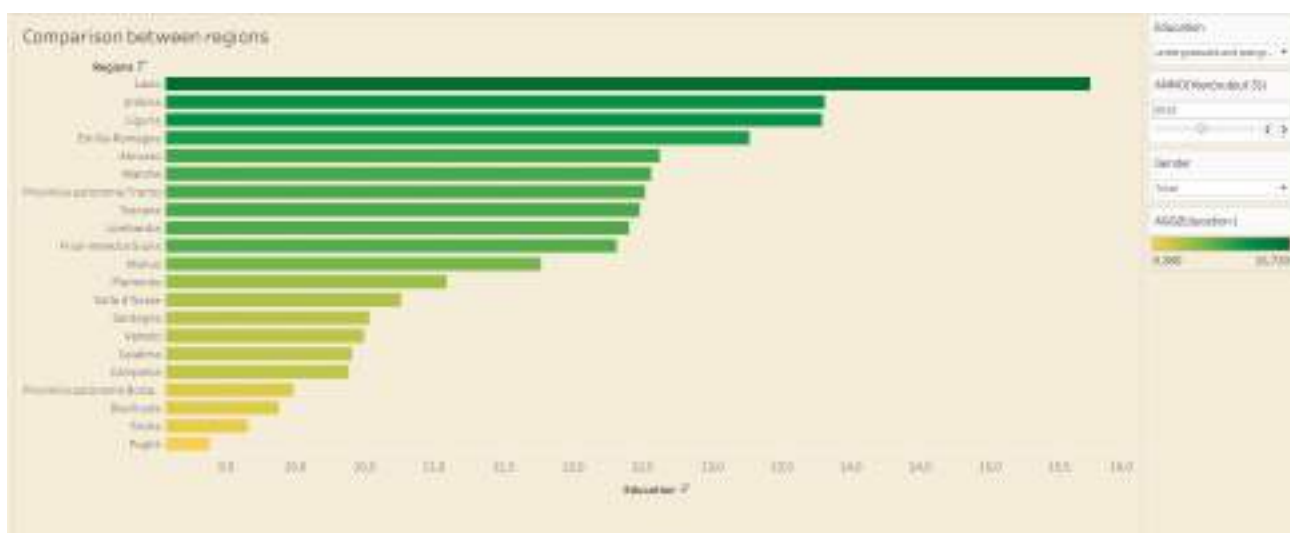
Map is the best chart when we want to show information using geocoded data for quick visualization and interpretation of data. The Education info map sheet is a map of Italy divided per regions and autonomous provinces (Trento and Bolzano). This map shows, through the dimension of the circles, the percentage of education per region, year, gender. The map is also used as filter per region for the other sheets. For geocoding I used the NUTS 2 criteria, for the autonomous provinces I manually added the NUTS 2 codes, i.e., ITH10 and ITH20, in the edit positions.

- **Education trend**



Line chart is the most used chart as it is very useful for showing trends over time. the occupational condition trend shows the trend of various conditions with the ability to filter by gender, region, and education info.

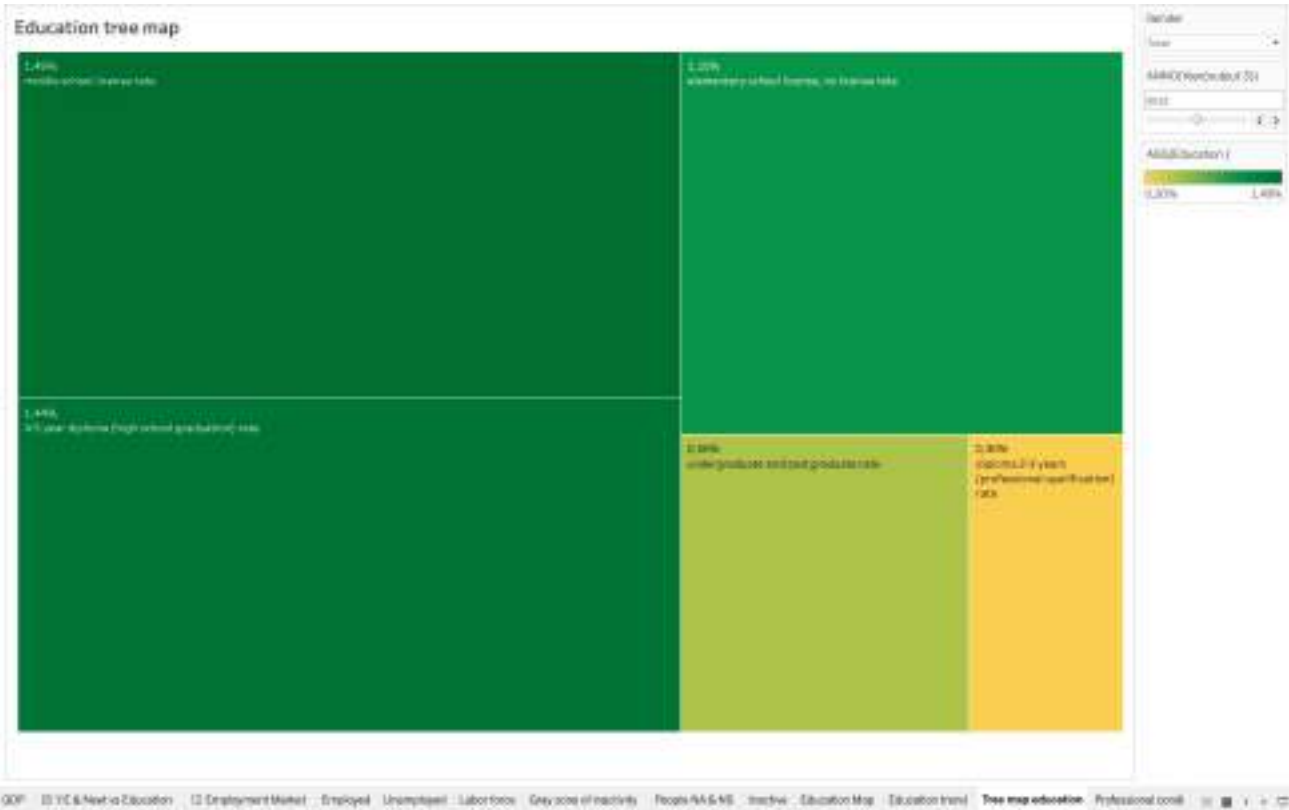
- **Comparison between regions**



By using the bar chart, we can compare professional conditions across regions or autonomous provinces. The graph can be filtered by gender, year, and education info.

A horizontal bar graph may be preferred over other types of graphs, such as the vertical bar graph, for a few reasons. First, horizontal bars may be easier to read and compare than vertical bars because the length of the bars is horizontally aligned and comparison between bars is easier. Also, a horizontal bar graph can be useful when you have many categories or variables to display, as horizontal bars can be organized over multiple rows without losing their readability.

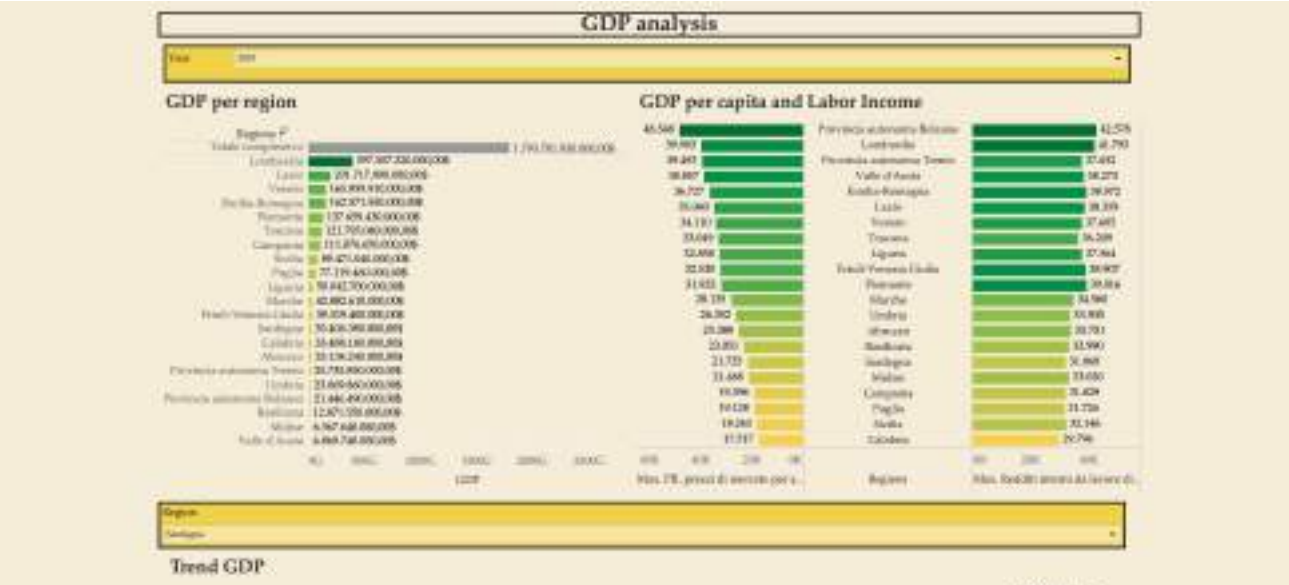
• Education tree map

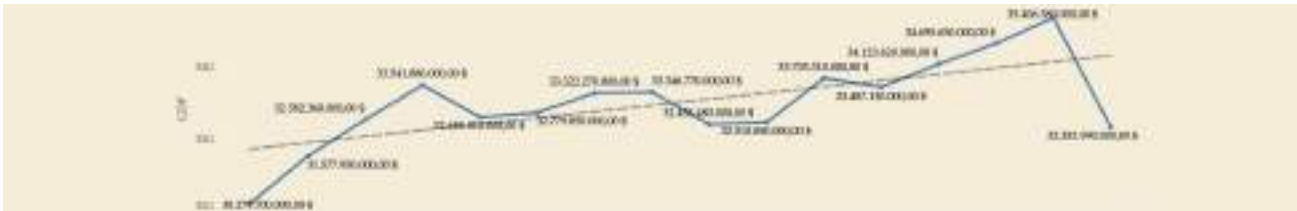


The tree-map functions as a visualization composed of nested rectangles. These rectangles represent certain categories within a selected dimension and are ordered in a hierarchy, or “tree.” Quantities and patterns can be compared and displayed in a limited chart space. Tree maps represent part to whole relationships.

In this sheet we would represent how the education is divided in one region. The sheet should be filtered by region, gender, and year.

Dashboard 3

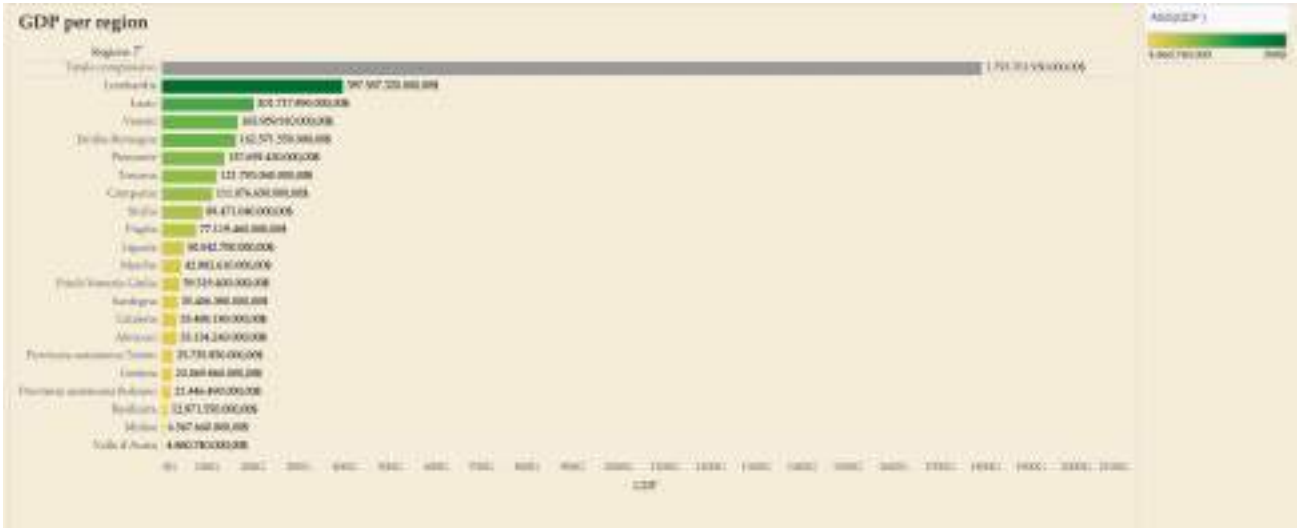




In this dashboard we analyze some important factors such as Gross Domestic Product,Gross Domestic Product per capita, and income per capita. The analysis is done by region and by type of variable considered. These variables considered are vital to our analysis of the population and labor market. We will first analyze the trends of the variables and then include them later in a more in-depth analysis.

Dashboard 3 in Detail

- GDP per region



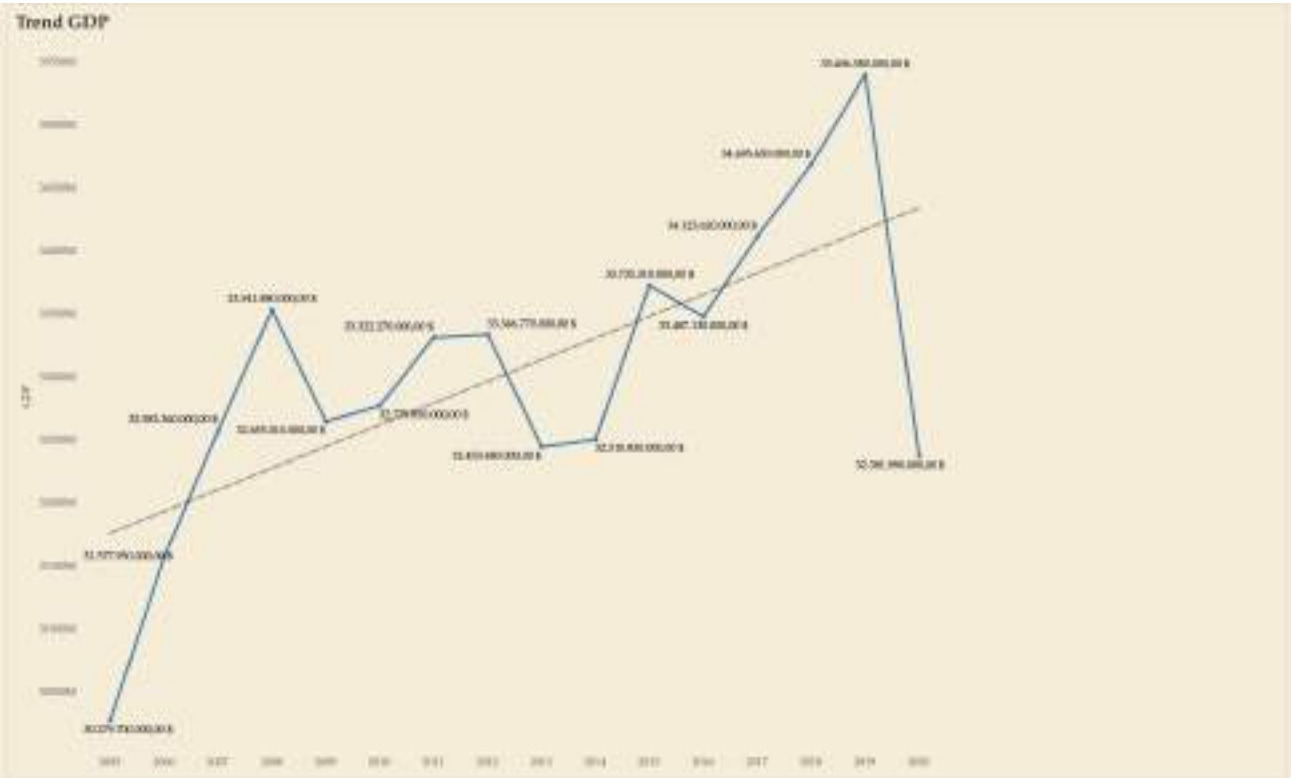
We used a horizontal bar graph to make the comparison between regions better. At the top is the grand total, which is the sum of all regions to get the GDP of Italy. The graph can only be filtered by year.

- Type of aggregates of GDP



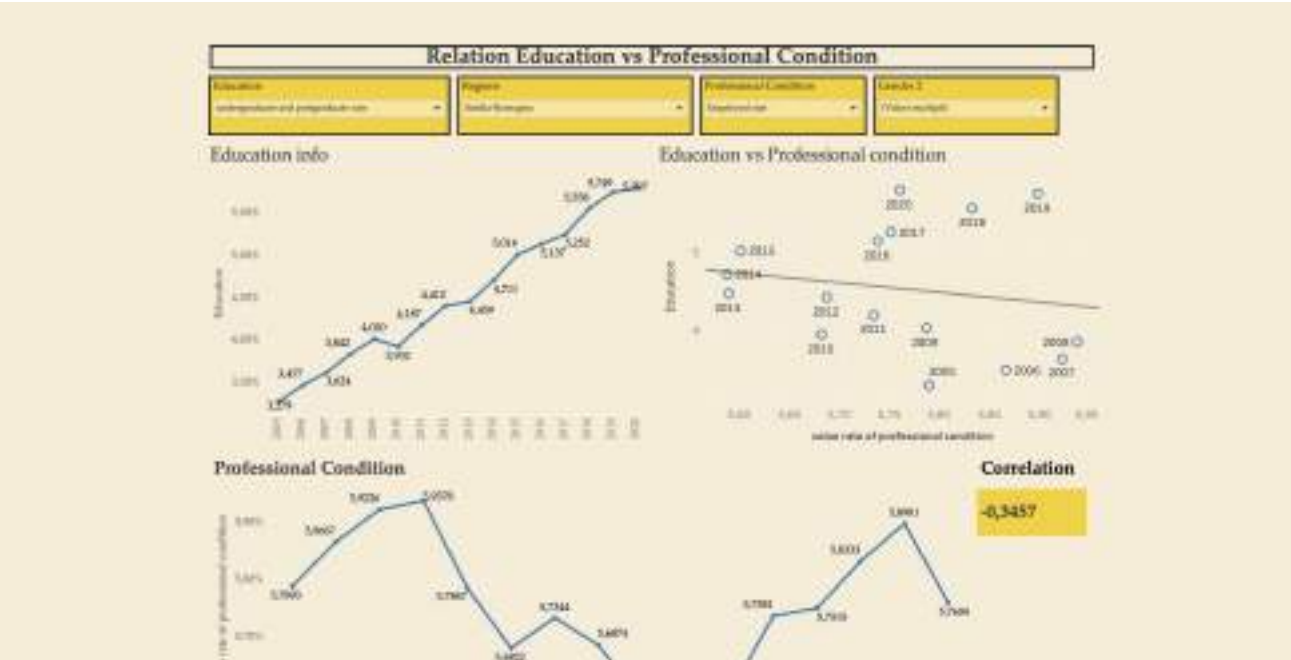
In this paper we compare two important aggregates such as per capita GDP and per capita incomes. The constructed graph compares different regions and can be filtered by year.

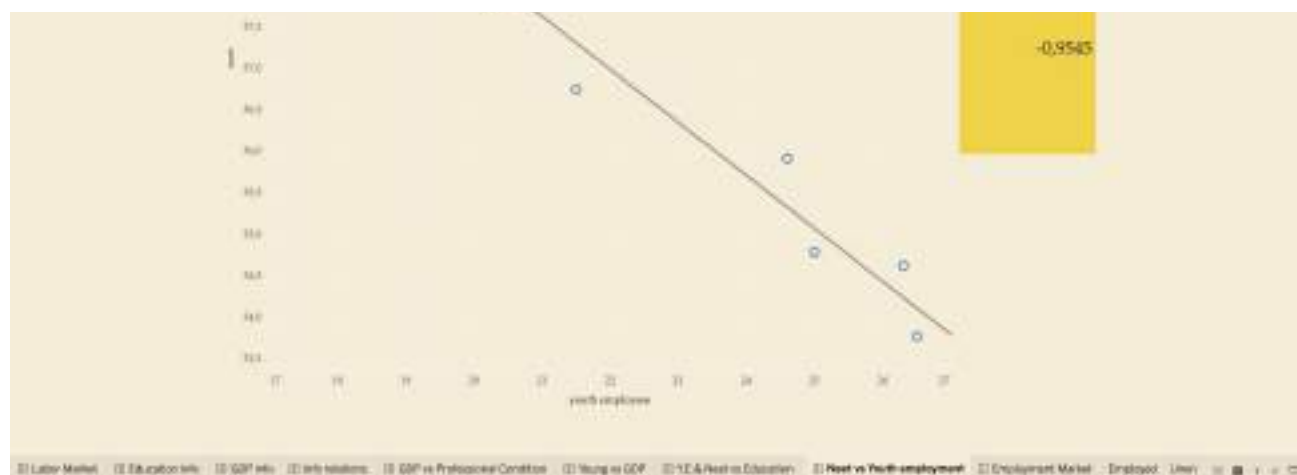
• Trend GDP



This sheet shows us how work the trend of GDP filtered by region. In the graph there is the trend line to understand in which direction the GDP is going. Despite the 2020 slump generally the trends by region are positive.

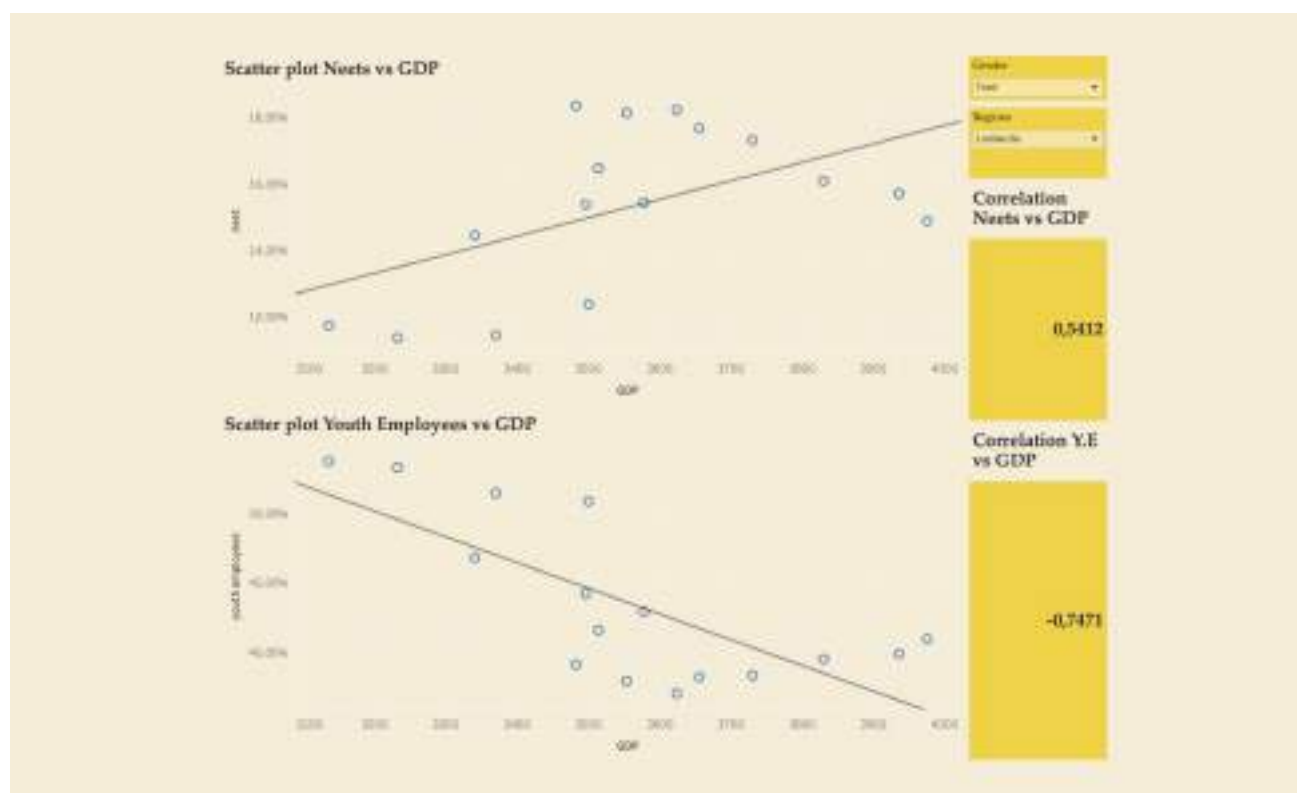
Dashboard 4





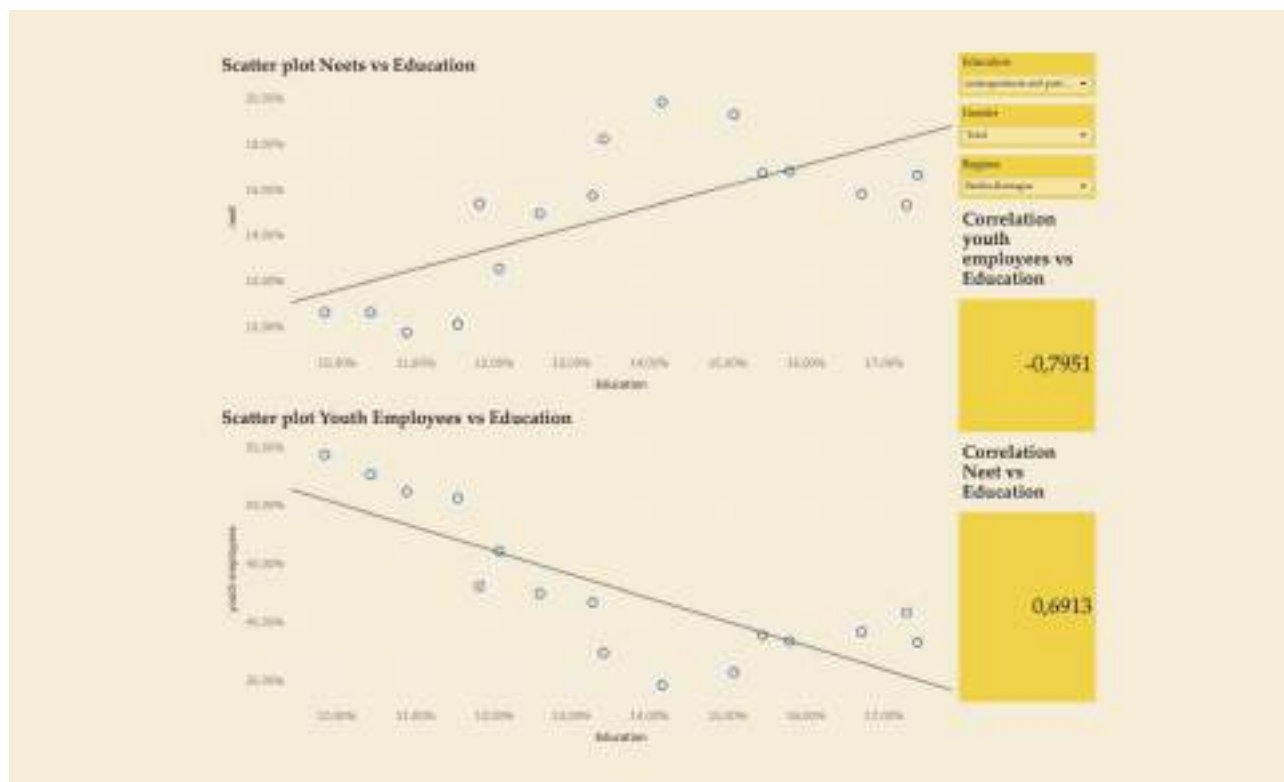
This dashboard analyzes the relationship between neets and youth employment. Clearly, we expected there to be a negative relationship. In fact there is a fairly high negative correlation almost all the time.

Dashboard 7



Scatter plots between GDP and the two categories of youth are shown in this dashboard. There are also correlation indices between the two categories of youth with GDP. These graphs can allow us to understand whether part of the phenomenon analyzed can be explained through this variable. For example, is it possible that as GDP increases, youth employment also increases? In some regions it has happened, but not in all. So, we can say that it will also depend the phenomenon on another factor, such as perhaps the study. It is certainly true that at an increase in GDP probably many young people will find jobs, thanks to investments, but it is also true that further growth in welfare might lead young people to study.

Dashboard 8

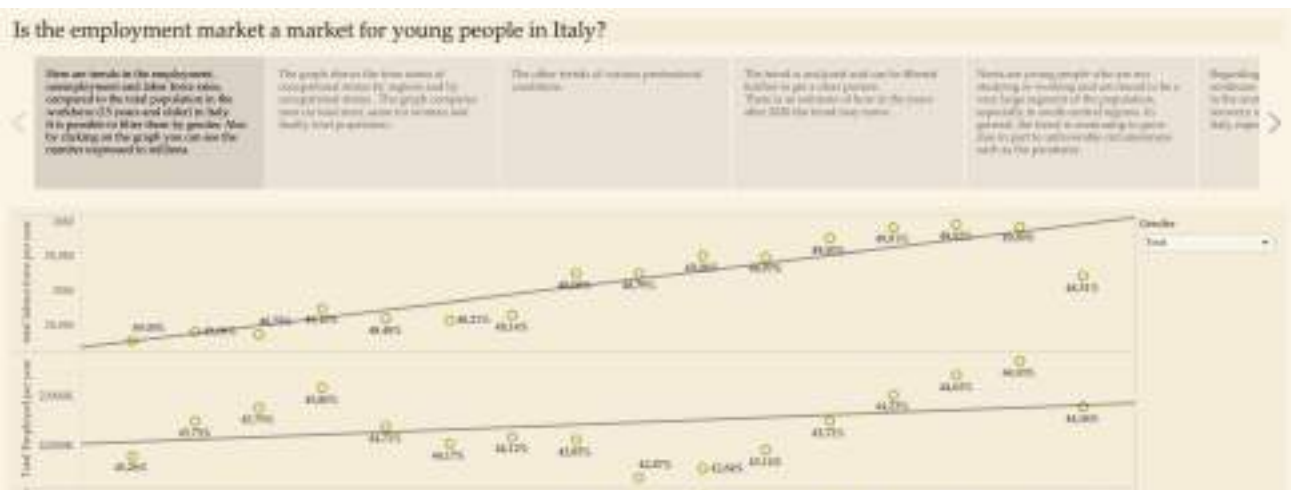


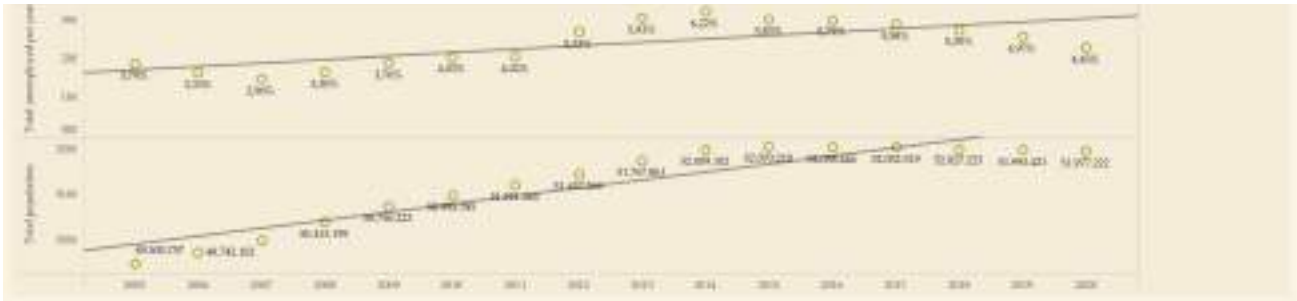
This dashboard analyzes the relationship between the two categories of youth with education. Using filters, it is possible to choose region, level of education, and gender. So that you have a special focus on each aspect.

DATA STORY

Using the data story, we want to summarize the milestones of the analysis carried out with the dashboards, step by step. In the first story, you can see trends in employment status by region and by status. The three graphs show the various genders respectively: male, female, and total.

I story

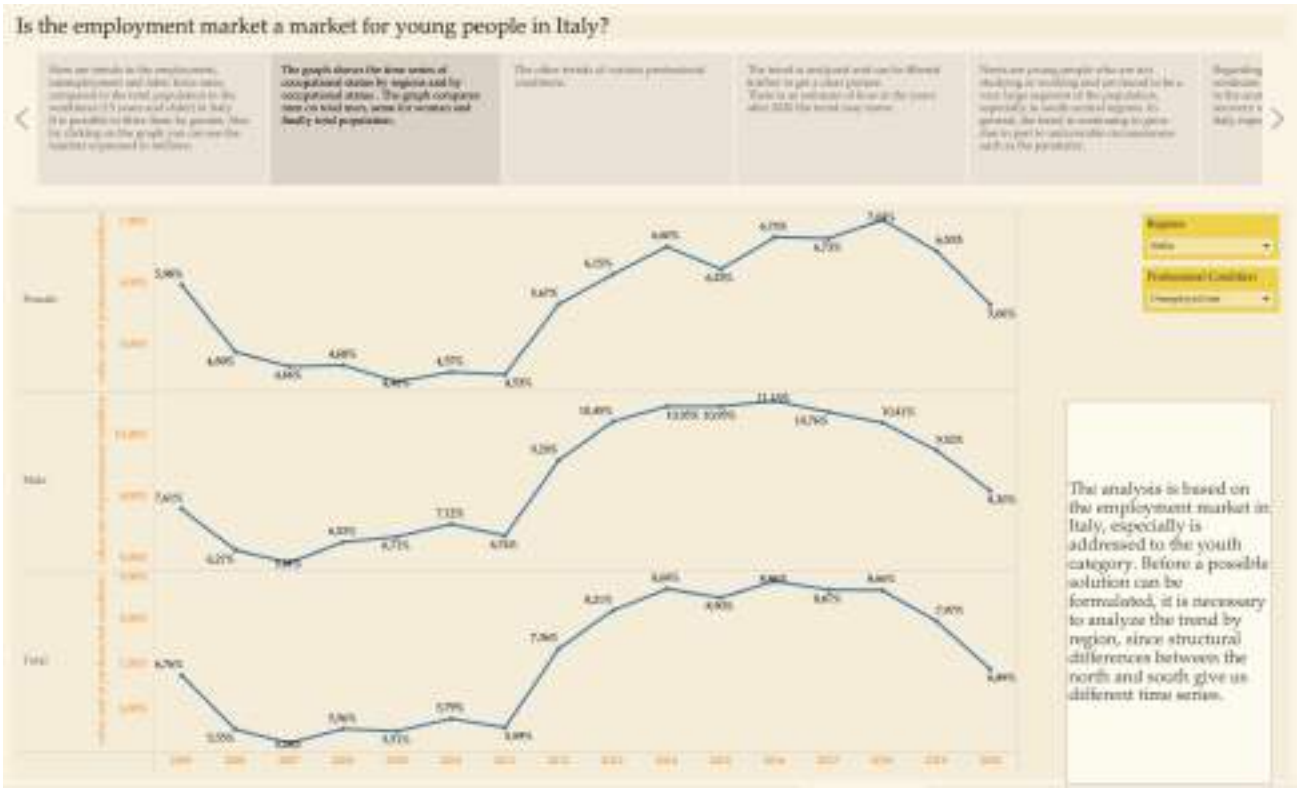




Here are trends in the employment, unemployment and labor force rates, compared to the total population in the workforce (15 years and older) in Italy.

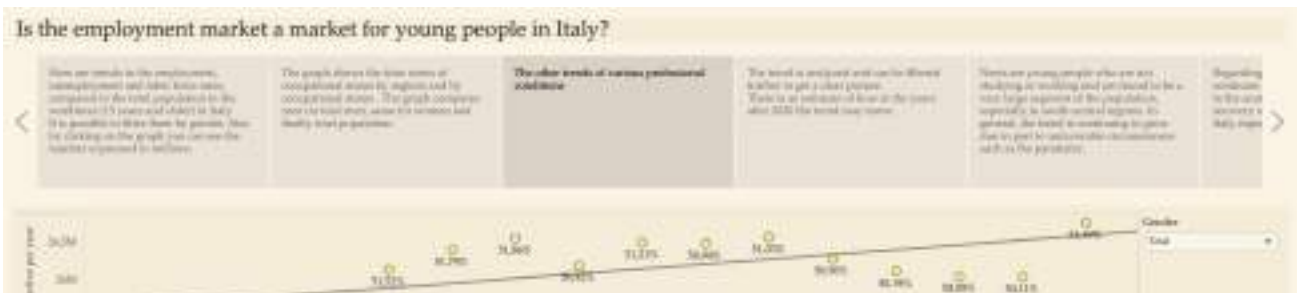
It is possible to filter them by gender. Also by clicking on the graph you can see the number expressed in millions.

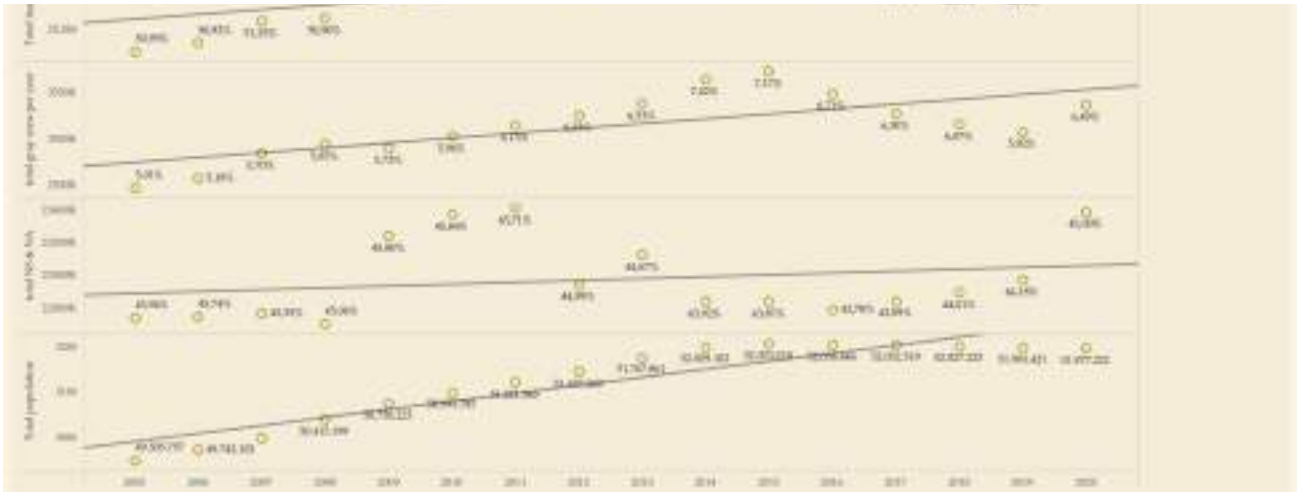
II story



This story answers the first question: what is the trend of various occupational statuses by regions? Having said that, surely you will have a clear idea of how they move over time and the critical issues there are between north vs. south-central.

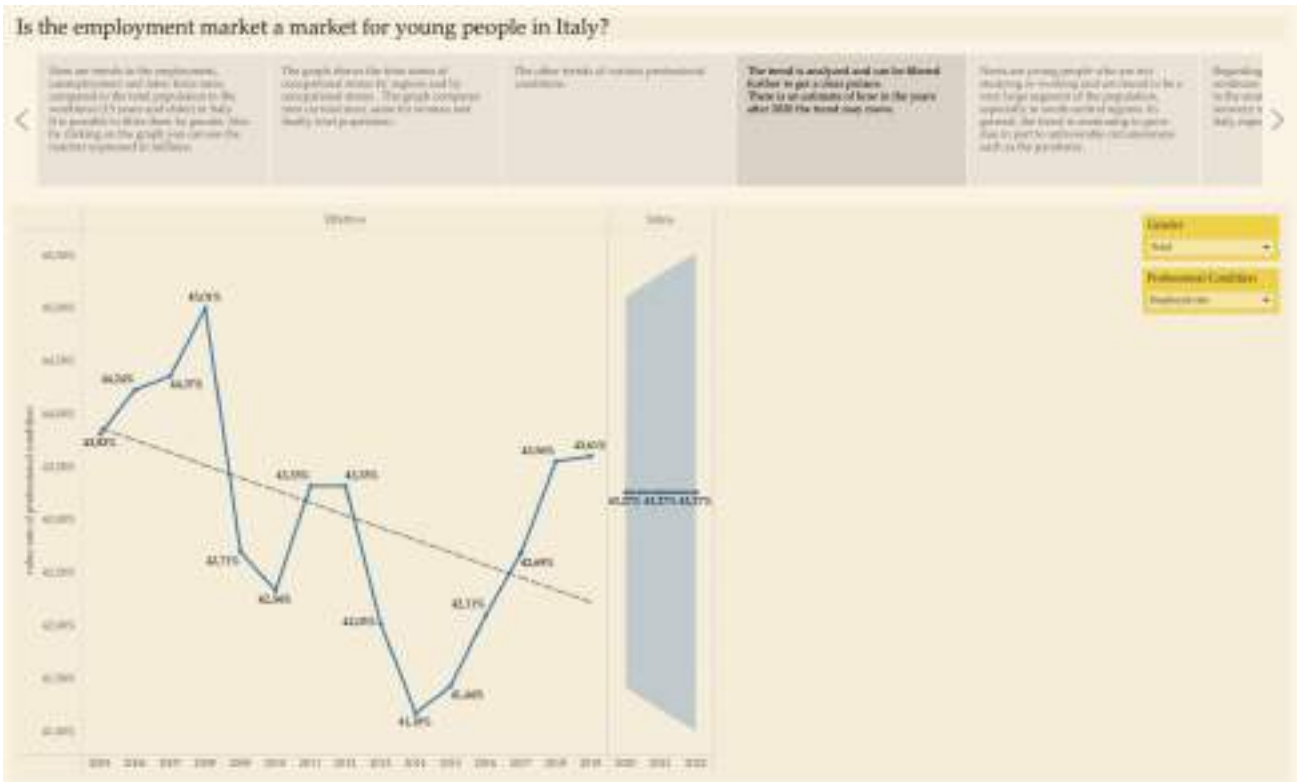
III story





The other trends of various professional conditions.

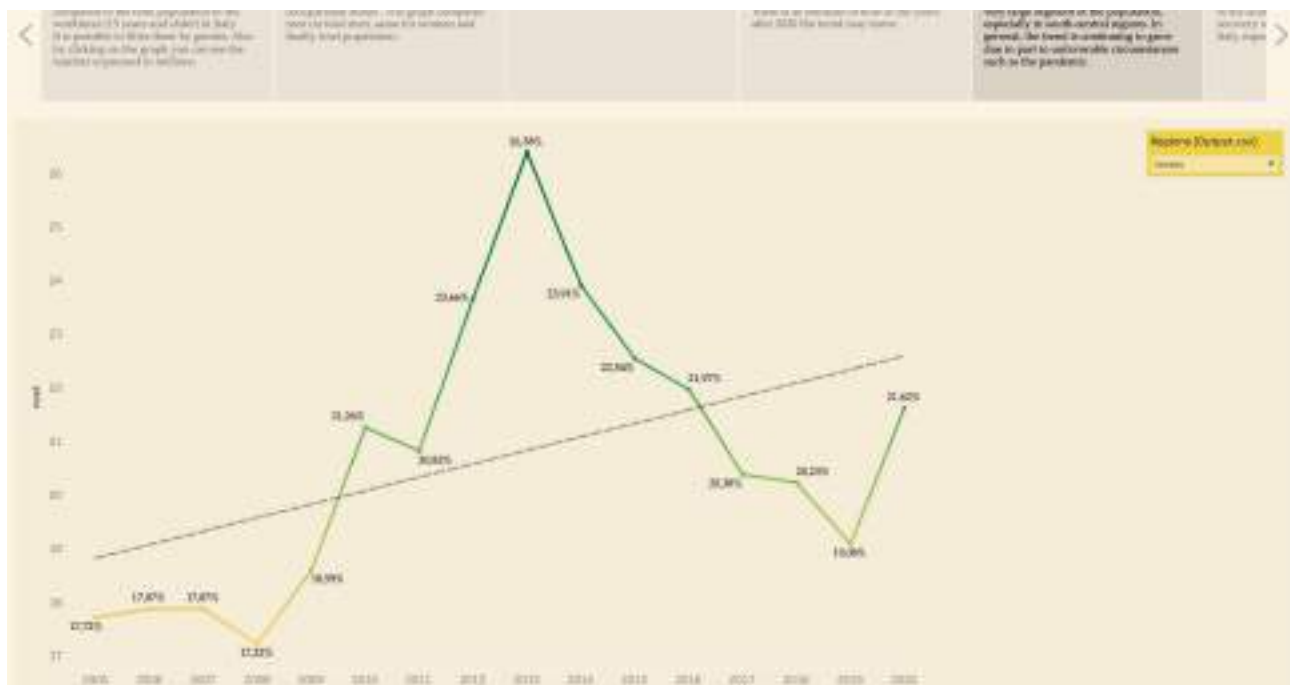
IV story



This story answers the question about the trend. It is true that there have been ups and downs in various professional conditions, but what direction is the trend moving in? Here you can see the trend by region, gender, and occupational condition. The labor force filter along with female is interesting. The trend is almost always increasing for each region. This gives us a sense of the changes taking place within the labor market. This same trend is also reflected in the female employment rate. Opposite trend is for men. Instead, the trend that includes both is often negative. The latter makes us understand how still the employment rate is mostly influenced by men than by women.

V story





This story shows the graph of Neets by gender and region. From here, you can see the trend for each individual region. Unanimously, it can be said that they are increasing in each region for both women and men. They are certainly increasing at different rates. Very negative result for southern Italy, more contained the phenomenon in the north.

VI story



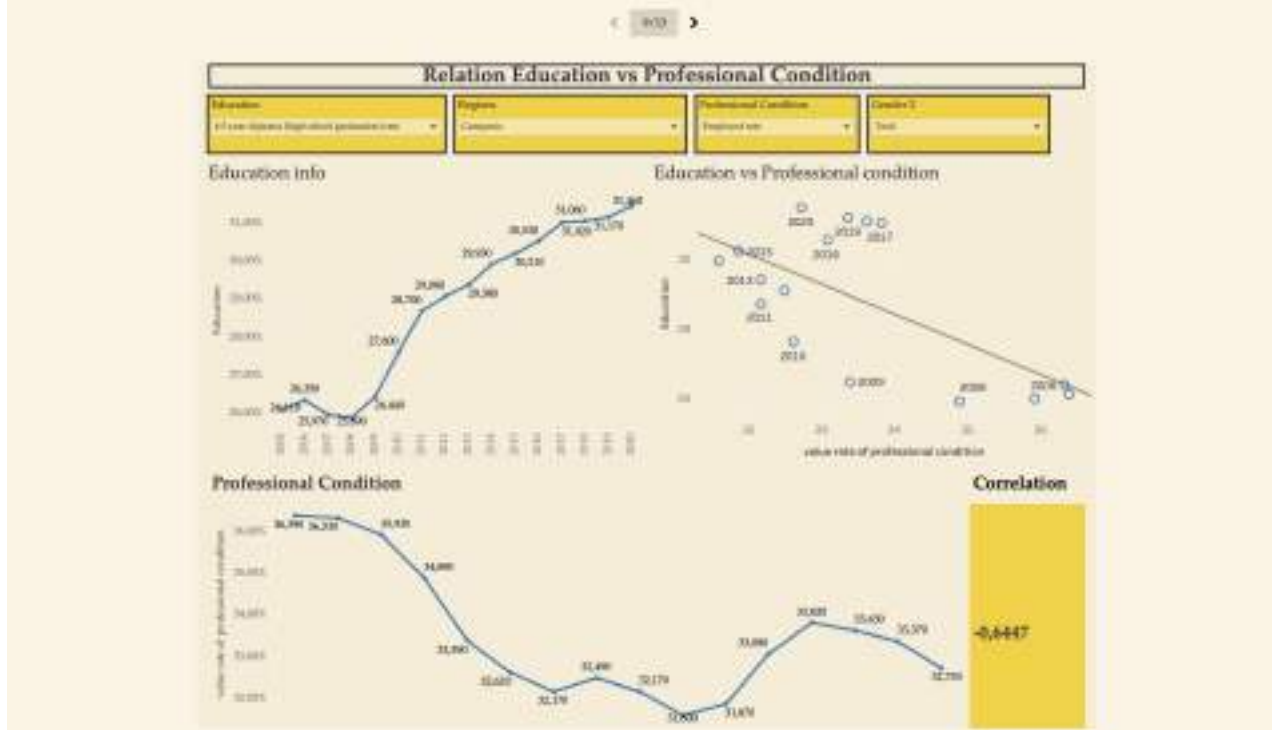
Instead, in this story we have the graph showing us the trend of youth employment (18-29 years old) by gender and region. If Neets increase as a result youth employment will drop. But what are the



Relation between Youth emoloyment vs Neet.

IX story

Is the employment market a market for young people in Italy?



In this story, dashboard 4 is taken up. This dashboard depicts the relationship between occupational status and education level. We want to understand whether there is a positive relationship between the two variables. If we look at the male employment rate, we see that there is a negative relationship between those with a bachelor's degree or more and the employment rate. Of a different trend is the female trend, which on the contrary increases along with the level of education.

X story

Is the employment market a market for young people in Italy?





In this story, GDP is analyzed using dashboard 3. Here we want to analyze GDP and other economic factors by region and year. So, you can get a clear picture of the economic situation, broadly speaking of course.

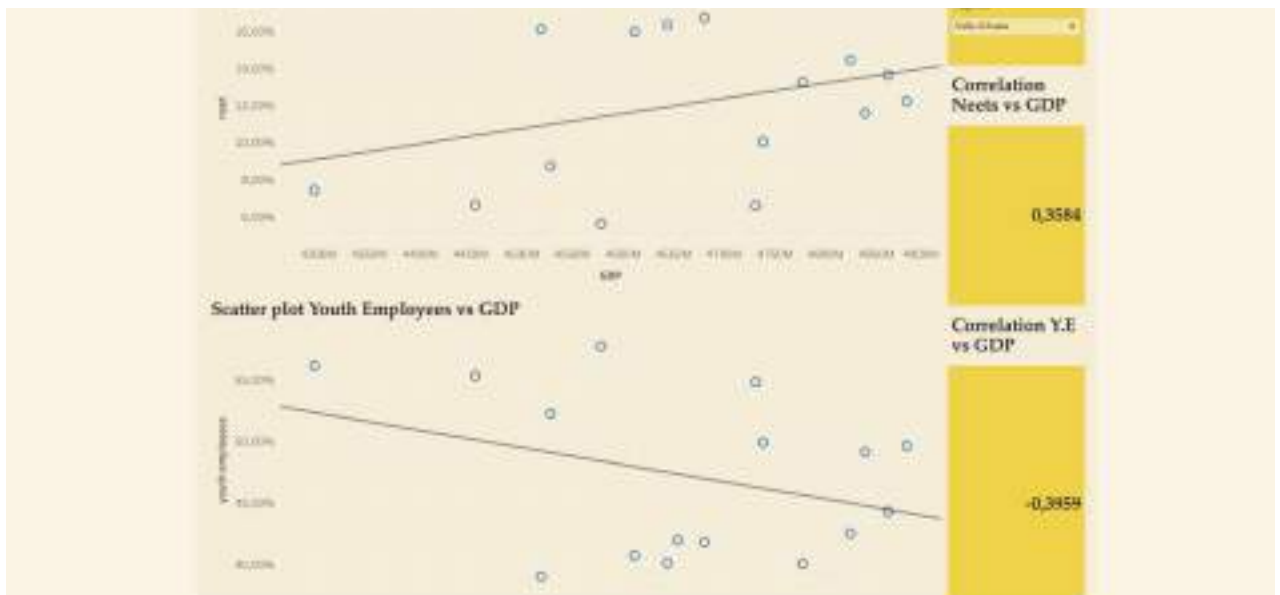
XI story



In this story, dashboard 5 is taken up, where there is the relationship between two variables such as GDP and occupational status. In this dashboard it is possible, through filters, to relate occupational status to GDP. It is possible to see how as GDP increases the female employment rate increases, for the male employment rate there is the opposite trend.

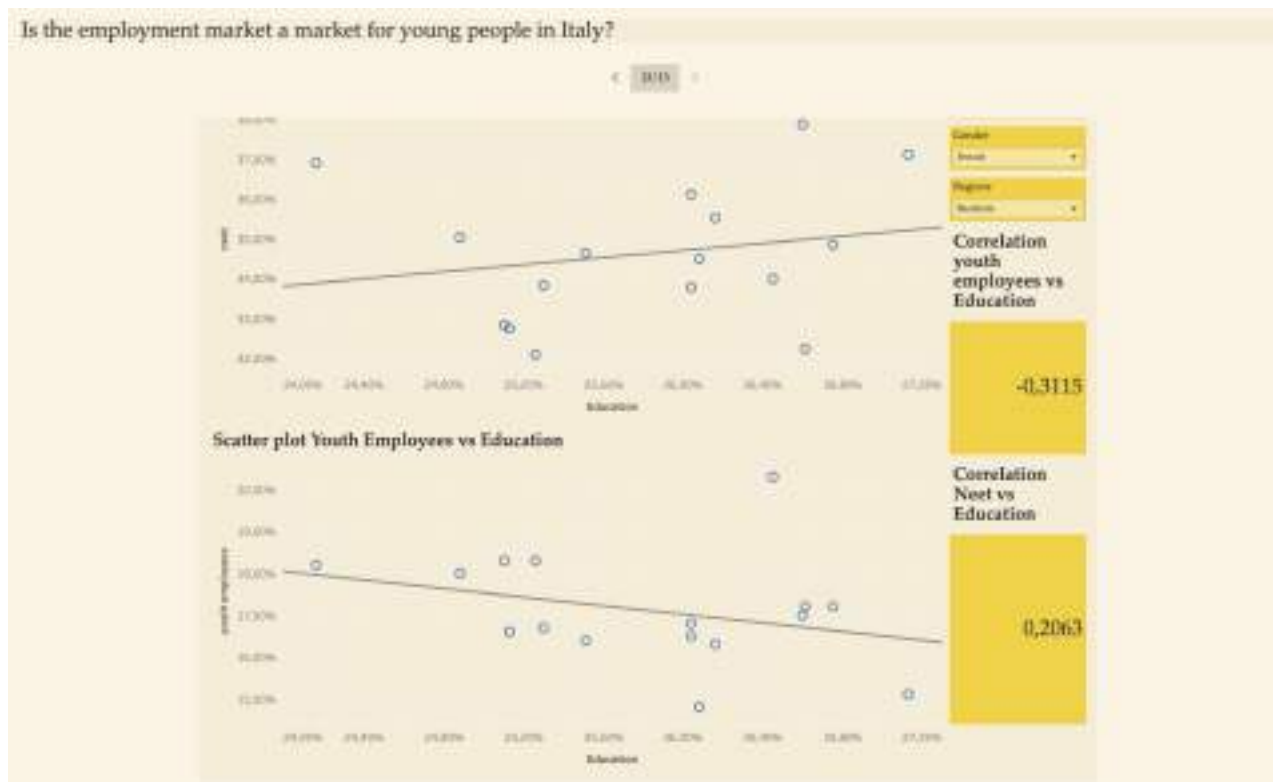
XII story





In this story we find the relationship between the two categories of youth examined and GDP. It is possible to filter by gender and by region. The relationships are quite different from gender to gender and by region. For example, in the central, southern and island regions there is an increase in youth employment and a decrease in Neets in corresponding to increase in GDP. Different story for northern regions. So, this can certainly be a growth factor for the central, southern and island regions but up to a certain point.

XIII story



The last analysis focuses on the study of education rate with neets and youth employment. The phenomenon in this case is easier to interpret. Definitely because of the increase in the education rate, thus more kids studying, there is a decrease in the youth employment rate.

ANALYSIS REPORT

Using dashboards and with the help of history, it is possible to describe a complex phenomenon such as Neets and youth employment, starting with an analysis of more general occupational status. In the first dashboard, it is possible to have a comprehensive analysis by region, by gender on occupational status. The goal of the report is to study and analyze, taking advantage of the data collected by ISTAT and EUROSTAT over time. The "rationale" is to understand why there is an increase in neets and a decrease in youth employment. Over the years the trend of neets has increased and vice versa youth employment has remained very low, mainly in the central, southern and island regions.

To answer the **first question**, that is, to describe the employment rate in Italy, mostly in the various regions, we need to start from a more general analysis and dwell on the total employment rate (men and women).

The time series further helps to develop the questions posed by this analysis.

In the first worksheet we can infer how the trend in total employment and unemployment rate over the year in Italy. We can see that the lowest unemployment rate obtained over the years was in 2007 at 2.96 percent, preceding the great crisis of 2008. After the crisis, we notice an increase year after year, peaking in 2014 with a rate of 6.22%. Following 2014, we notice a decrease until 2020 with a rate of 4.45%. The trend line predicts an increase in the succeeding years, certainly true due to the pandemic that hit the following two years. As for the employment rate, the peak occurred in 2008 equal to 45.8%, probably before the Wall Street speculative bubble burst. Thereafter we note that the employment rate declines year after year, touching a low in 2014 equal to 42.84%. After, 2014 we register a slight recovery reaching 44.93% in 2019 and decreasing to 44.06% due to the first effects of the pandemic. When it comes to the trend of the female unemployment rate in Italy, we can see that in 2007 the lowest unemployment rate of 2.97% is reached. In 2014, the highest peak of the female unemployment rate of 5.54% occurs and then gradually decreases until 2020 with a value of 4.06%. While for the female employment rate, the trend is trending upward with the highest peak in 2019 equal to 36.68 percent, decreasing to 35.78 percent in 2020. When it comes to the trend of the male unemployment rate in Italy, we can see that again in 2007 the lowest unemployment rate of 2.95 percent is reached. In 2014, the highest peak of the female unemployment rate of 6.96 percent occurs and then gradually decreases until 2020 with a value of 4.86 percent. While as for the male employment rate, the highest peak occurs in 2006 with 57.54%. Thereafter, there is a relapse until 2014 with 51.74% and a recovery in subsequent years until 2020 with 52.95%. In general, we can see how that the highest rates of employment and lowest rates of unemployment occur before the great crisis of 2008, which is followed by ever lower rates of employment and ever higher rates of unemployment, until 2017 in which there is a slight recovery until 2019, which is followed by the lapse of the pandemic. As for Sicily, the employment rate is around 30 percent, compared to other regions that hover around 50-60 percent such as the Autonomous Province of Bolzano and Lombardy. In Sicily, the peak occurred in 2006 equal to 35.84% and then there was a steep decline until 2014, when the employment rate was 30.51%. In subsequent years there is a very slight increase until 2020 with 31.64 percent. The unemployment rate in Sicily is among the highest in the Italian regions. Until 2011 it hovered around 5 percent, soaring in subsequent years reaching 8.86 percent in 2016 until declining in 2020, pre-pandemic, to 6.89 percent.

Through the **second question** we analyze trends in other employment conditions such as inactive, not seeking and unwilling to work and gray area of inactivity.

Inactives are those who have not sought work in the four weeks preceding the reference week, or are unwilling to work within the next two weeks, or both conditions. The trend from 2005 to 2020 is consistently around 50 percent with peaks in 2011 at 51.86 percent and in 2020 at 51.49 percent. In 2018 it is close to 50.08%, the lowest value in this time frame.

The "**gray zone of inactivity**" is a type of unemployment that falls outside official definitions and that Istat has identified. This includes individuals who are not considered unemployed according to statistical definitions, but who would be available for work if they had the opportunity. The official definitions of unemployment involve active job search actions and willingness to work within certain time intervals. In this trend it has been increasing until 2015 equal to 7.17% oscillating for the following years until 2020 with a value of 6.49%

Not seeking and unwilling to work are those who are not employed and not seeking employment. This trend remains around 40% with peaks in both 2011 and 2020 with values of 45.71% and 45% respectively. While the lowest peak occurs in 2015 equal to 43.87%.

To answer the **third question**, we use the seventh data story, where there are the two categories of youth, Neets and youth employment rate by region, gender, and year.

It is easy to see that the situation is better in the northern regions. The best region in terms of youth employment, albeit by a small margin, is the autonomous province of Trento, which together with Bolzano forms the region of Trentino-AltoAdige. If we consider the year 2019, one year before the crisis that COVID has identified, the autonomous province of Trento has a youth employment rate of 52% and a NEET rate of around 13.15%, considering the population of young people without gender distinction. If we filter by gender, we have a youth employment rate of 57% for the male population and a NEET rate of 8.7%. If we filter by female gender, the situation is slightly different, with an employment rate of around 47% against a NEET rate of 17.77%.

The percentages are quite high compared to the South-South Island regions.

To answer the first sub-question, we can look at the eighth story, where there is the scatterplot between Neets and youth employment and the correlation index between the two variables.

Of course, the correlation index in this case is always negative and very high, so there is a negative correlation, in fact it is intuitive to understand that as neets increase, youth employment decreases and vice versa.

To answer the second sub-question, we can use the fifth and sixth data stories. Sicily, unlike the northern regions, is in a completely different situation. Using the fifth data story, we can see the trend in the number of Neets by gender. The trend is increasing, despite some decreases after 2014, to be around 41 percent if we talk about the total youth population, 36.57 per cent for males and 45.69 per cent for females. As for the youth employment rate, it has fallen dramatically over the years, especially after the great crisis of 2008, to less than 18 per cent, 12.5 per cent for females and 22.4 per cent for males.

To answer the third sub-question, we can use the seventh data story.

If we look at the Italian situation in the years before the Great Crisis, considering 2007, we have a picture that mirrors the more recent years, but with lower percentages for Neets and higher percentages for youth employment. Looking ahead to 2019, the situation is like that in 2007, with the North doing much better than the South. It is worth noting that, 11 years after the great crisis, the percentages have never returned to the pre-crisis situation. The situation was aggravated by the COVID pandemic in 2020, which did not upset the balance, as the pre-COVID rankings remained unchanged or almost unchanged, except for a decrease in various youth employment rates and an increase in Neets.

To answer the **fourth question**, we focus on the eleventh data story, which is the relationship between GDP and working conditions. The analysis focuses on employment and unemployment rates in relation to GDP. As mentioned above, youth employment levels have not returned to pre-crisis levels as the number of neets has increased and so far, does not seem to be decreasing, especially in the South. The first factor we wanted to look at is GDP, which can be compared to a thermometer for the health of a state. It is true that GDP fell sharply after the 2008 crisis, as we can see in the tenth story. The crisis of 2008 led to an increase in unemployment and an increase in precariousness in its aftermath.

Contrary to the rates analyzed earlier, which never returned to pre-crisis levels, GDP even exceeded the 2008 level around 2015. From then on, it grew until, for obvious reasons, it suffered a setback in 2020. Note how graphically we can see that youth employment in almost all regions reversed its trend around 2015 and started to grow again, albeit slowly. Using the graph and the correlation index, we can see that the employment rate of the general population is positively correlated with GDP in the south-central regions and with some northern regions. This shows that GDP is to some extent a drag on the employment rate, but this is not always the case. In fact, the correlation index varies greatly, being high for some regions and low or even negative for others, but generally never falling below -0.3. For example, if we take Sicily, the correlation index for the employment rate for all sexes is 0.28, while if we filter for women, it is 0.57 and for men 0.20. We can see that there is a higher correlation between the female employment rate and GDP. Another example is Puglia, where the correlation index between GDP and female employment rate is 0.86, while the correlation between male employment and GDP is negative at -0.53. In almost all regions there is a higher correlation index between the female employment rate and GDP at the expense of the male correlation index. It can be said that GDP growth has a stronger effect on the female employment rate.

On the other hand, there are contradictory relationships between the unemployment rate and GDP. In some regions where the employment rate has risen, the unemployment rate has also risen. This may seem counterintuitive, but the rates are calculated based on the total population, which includes not only the economically active but also the inactive. So, there are more people entering the labour force and actively looking for work. This can be seen in Liguria, where there are both positive but low correlation indices for the employment rate and the unemployment rate. The correlation index for the active population is very high at 0.85, while the correlation index for the inactive population is -0.85. So, in essence, we can say that an increase in GDP certainly improves the situation of a region by favoring an increase in the labour force and an increase in the employment rate, but also in the unemployment rate at the expense of the inactive.

Turning our attention to NEETs and youth employment, we see that in some regions, particularly in the South, there is a low, albeit positive, correlation between the youth employment rate and GDP and, conversely, a low negative correlation between the NEETs rate and GDP. On the other hand, in the North and in the Centre, there is almost always a negative correlation between youth employment

and GDP, and a positive correlation between neets and GDP. It can therefore be said that in regions with a very low GDP, GDP growth slightly pushes up youth employment growth at the expense of Neets.

To answer the **fifth question**, we use the ninth story, where there is a relationship between occupational status, in particular unemployment and employment rates, and educational attainment. Regional structures are very different and accommodate different types of workers; for example, Sicily is a predominantly agricultural-tourist region and is unlikely to retain its graduates except in these narrow areas. Indeed, the educational attainment rate of 'graduate or above' is negatively correlated with the employment rate. There is often talk of a brain drain, at least to the northern regions, if not abroad. However, this reality, which seems to be purely a southern phenomenon, as confirmed by the various correlation indices by region, actually affects almost the whole of Italy, with small exceptions such as the autonomous province of Bolzano. There is a positive correlation between the unemployment rate and the number of graduates in all regions.

The figure, which is the result of a careful analysis, describes the structural complexity of Italy in terms of education. If we look at women, we see that in a large part of Italy there is a positive correlation between graduation and employment rate. But why is the correlation between the employment rate and the number of graduates, including both men and women, negative? It is because men are more present in the labour market and have fewer graduates than women, so they have more weight.

If we look at the Neets rate and the youth employment rate in relation to the education rate, we definitely get a negative correlation between graduates and youth employment, because young people who are studying hardly ever work.

Moving on to the NEETs rate, we see that the NEETs rate increases as the level of education increases in the different regions, with exceptions that may be dictated by some specific factors that are not examined. The opposite is true for youth employment, but this follows its own logic, as the more you study, the more time you spend outside the labour market.

On the other hand, if we go down to low levels of education, such as primary school or no driving licence, we have a positive correlation with employment.

The sixth and the last question reflects the whole analysis: a phenomenon as complex as the youth employment rate and NEETs cannot be solved in a short period of time, as the economic and social structure of entire regions would have to be changed. The analysis showed that an increase in GDP would probably succeed in raising the employment rate, but not with the desired results. Indeed, as we have seen, although GDP has exceeded the pre-crisis level of 2008, the employment rate has not returned to that level and the unemployment rate is still quite dramatic. The situation is much worse in the central and southern regions, where an increase in GDP would have a greater impact on boosting employment, especially among young people. The analysis shows a split between men and women; the trends are positive for women, who are gradually taking up a larger share of the labour market, unlike men, who are struggling to do so.

To return to the question, and as already mentioned, the only decisive factor would be an increase in GDP, through the use of public or private sector investment, which would create significant jobs for skilled workers, because let us not forget that young people are becoming increasingly qualified.

COLOR BLIND TEST

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. In the following section you can see that the colors chosen for the dashboards respond well to the color blindness test.

- **BLACK AND WHITE TEST**



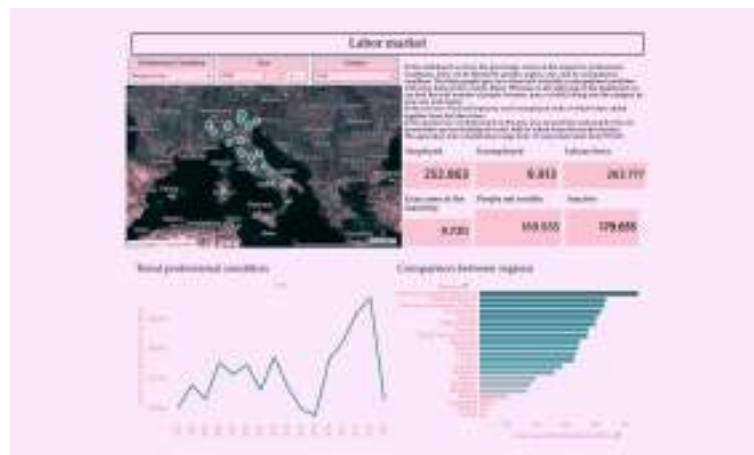
- **DEUTAN TEST**



- **PROTAN TEST**



- TRITAN TEST



Conclusion

In conclusion we can say that, the employment situation in Italy has suffered in the periods of the great crisis since the one in 2008, that despite some recoveries, the total employment rate has never regained its previous values. The lowest peak was reached around 2014 with a slight recovery in the following years before declining again before the new pandemic crisis in 2020. There is a turnaround in women's employment, which is a growing trend unlike men's employment, which grows slowly and is more affected by market conditions. But despite this, female employment is still lower in terms of absolute value.

Regarding the employment rate among regions, Sicily emerges as one of the regions with the lowest and highest employment values for unemployment.

Relative to youth employment, it continues to be a negative trend, especially in the central and southern regions. A small window opens in northern Italy, which in any case, is unable to absorb the entire youth supply of the peninsula, thus confirming that Italy tends to be a country for old people.