SMBUD 2021 - Project work 3

Aman Gabba - 10793117

Andrea Cerasani - 10680486

Giovanni Demasi - 10656704

Pasquale Dazzeo - 10562130

Vlad Marian Cimpeanu - 10606922

POLITECNICO
MILANO 1863

Contents

1 Introduction

1.1 Problem Specification

The aim of this project was to design, store and query data on a NoSQL DB supporting a data analysis scenario over data about COVID-19 vaccination statistics. The purpose is that of building a comprehensive database of vaccinations.

A vaccinations dataset has been suggested, with the purpose to pick a time interval of at least 3 months from it and, by using an ElasticSearch installation, import the data, apply the appropriate schema design choices, implement some queries aiming at exploring the data statistics and design a basic visualization dashboard of the results.

1.2 Hypothesis

The assumptions taken into account are the following:

- It is assumed that people who completed the vaccination cycle belong to the following categories:
  - have taken two vaccine doses
  - have taken a Janssen dose
  - have taken a vaccine dose after previous infection
- A person is considered vaccinated if he has already taken at least one dose or has already taken a vaccine dose after previous infection.
- A person who has been previously infected is not supposed to take any second vaccination dose.

2 Datasets

Two different datasets have been used for analysis purposes. A description of both can be found below.

2.1 Vaccine administration dataset

The Dataset used for the project is named `"somministrazioni-vaccini-latest.csv"` and it has been downloaded from the `"dati"` folder from the official Italian Government Github repository at the following link : https://github.com/italia/covid19-opendata-vaccini.

2.1.1 Schema

This dataset contains information about administered vaccines in Italy and it is made by the following fields:

Field	**Data type**	**Description**
area	string	Code of the delivery region
supplier	string	Complete name of the supplier of the vaccine
administration date	datetime	Administration date of the vaccines
age group	string	Age group to which the subjects to whom the vaccine were administered belong
male count	integer	Number of vaccinations administered to males per day, region and age group
female count	integer	Number of vaccinations administered to females per day, region and age group
first doses	integer	Number of people administered with the first dose
second doses	integer	Number of people administered with the second dose

post infection doses	integer	Number of administrations given to subjects with previous covid-19 infection in the 3-6 month period and who, therefore, conclude the vaccination cycle with a single dose
booster doses	integer	Number of people administered with an additional dose/recall
NUTS1 code	string	European classification of NUTS territorial units: NUTS level 1
NUTS2 code	string	European classification of NUTS territorial units: NUTS level 2
ISTAT region code	integer	ISTAT code of the Region
region name	string	Standard denomination of the area (where necessary bilingual denomination)

2.2 Istat population dataset

As optional point of this project, analysis has been integrated with another dataset which contains information about the Italian population, such as number of people per age range and region or number of people per gender per region. The dataset has been downloaded from a Covid-19 related Github repository at the following link: https://github.com/pcm-dpc/COVID-19, it is located in the `"dati-statistici-riferimento"` folder and it is the one named `"popolazione-istat-regione-range.csv"`.

2.2.1 Schema

This dataset contains information about population number per region and age range in Italy. It is made by the following fields:

Field	**Data type**	**Description**
ISTAT region code	integer	ISTAT code of the Region
NUTS1 code	string	European classification of NUTS territorial units: NUTS level 1
NUTS1 description	string	Cardinal direction based on the NUTS1 code
NUTS2 code	string	European classification of NUTS territorial units: NUTS level 2
region name	string	Standard denomination of the area (where necessary bilingual denomination)
area	string	Code of the region
region latitude	float	Latitude coordinate of the region
region longitude	float	Longitude coordinate of the region

age range	string	Age range to which the population numbers refers
male count	integer	Number of males in the population, per region and age range
female count	integer	Number of females in the population, per region and age range
total count	integer	Number of males plus females in the population, per region and age range

2.3 Other considerations

The data types written in the schema tables are the 'original' ones, so the ones used by the datasets creators.

The same data types have been used to implement and use the dataset in ElasticSearch because they well represent the different parameters, so no changes were needed, except for the ISTAT region code field. It is better to keep all the ISTAT region codes as keywords instead of numbers for the following reason:

they are numbers but for compatibility reasons they should be considered as keywords. In Kibana, there is the possibility to associate data to regions according to ISTAT code convention, although, the format used by Kibana is the following "01, 02, 03, 04, ...", thus, if ISTAT codes are imported as numbers they will not be compatible with that convention as Kibana will find the following codes "1, 2, 3, 4, ...".

As the original datasets do not follow the Kibana convention, they have been adapted through the script `"dataset_cleaner.py"`.

Both datasets used have Istat region code as a field, to understand what it represents, it is suggested to visit the following link:
https://en.wikipedia.org/wiki/NUTS_statistical_regions_of_Italy.

When the two datasets will be imported, both created indexes will use Istat region code to identify a region. The original dateset used for istat_population did not assign a code to Trentino region (04), instead it assigned codes for Bolzano(21) and Trento (22) provinces. For this reason codes have been corrected in order to link the two indexes.

Elastic search admits two kinds of type to handle strings: Keyword and Text.

For this project purposes, all the strings have been converted into keywords, indeed all string fields represent either an identification code (ISTAT_region_code, NUTS codes, ...) or an identification name (supplier, region name, ...) or an identification range (age group). For this reason, there is no need for text search as it is required perfect match to find certain documents.

Last but not least, in order to correctly integrate the "popolazione-istat-regione-range.csv" it was needed to fix the age ranges fields. The main dataset contains the age range '0-19', whereas the integrated one contains '0-15' and '16-19', for this reason the two age ranges have been merged through the script "merge datasets.ipynb".

After the cleaning processing, "cleaned_data.csv" is the fixed result of "somministrazioni-vaccini-latest.csv" whereas "new_isat_code.csv" is the final result of "popolazione-istat-regione-range.csv"

3 Queries and Commands

In the following chapter all the queries and commands parameters (part of the code to substitute with desired values) will be highlighted with magenta bold text.

Some parameters information can be useful for different queries or commands so they are written here to avoid writing them multiple times:

- Start date and End date are, respectevely, the starting date of a period and the ending date of a period. Dates must be in the following format YYYY-MM-DD and, obviously, End date must be subsequent to Start date.
- Date is a generic date and it must be in the following format YYYY-MM-DD.
- Supplier, for coherence w.r.t other documents, must be one of the following: Moderna, Janssen, Pfizer/BioNTech or Vaxzevria (AstraZeneca).
- Age range must follow one of this format: 12-19, x0-x9 or 90+ (where x is a number between 2 and 8)

3.1 Queries

The first eight queries refer only on the vaccination dataset. Instead, the last one refers to both vaccination and Istat population dataset.

3.1.1 Delta vaccination per area

For each region, this query returns the percentage of the difference between vaccinations of a given date and its precedent day, calculated with respect to the amount of vaccinations performed the day before. If the vaccinations have increased, the percentage will be positive, negative otherwise.

This query is thought to be used during the current day, thus the parameters "Date 1" and "Date 2" should be respectively "now" and "now-1d/d". By the way, as the database is not up to date, the query will not be correctly performed, for this reason it is suggested to use the last dates available in the dataset which are "2021-12-22" and "2021-12-21".

```
GET istat_vaccinations/_search
{
  "size" : 0,
  "aggs": {
    "group_by_area": {
      "terms": {
        "field": "nome_area"
      },
      "aggs": {
        "today_vaccinations" :{
          "filter": {
            "term" : {
              "data_somministrazione": "Date 1"
            }
          },
          "aggs": {
            "amount": {
              "sum": {
                "script": {
                  "source": "doc['sesso_maschile'].value + doc['
    sesso_femminile'].value"
                }
              }
            }
          }
        },
        "yesterday_vaccinations" : {
          "filter": {
            "term" : {
              "data_somministrazione": "Date 2"
            }
          },
          "aggs" : {
            "amount": {
              "sum" :{
                "script": {
                  "source": "doc['sesso_femminile'].value + doc['
    sesso_maschile'].value"
                }
              }
            }
          }
        },
        "delta_percentage" : {
          "bucket_script": {
            "buckets_path": {
              "today" : "today_vaccinations>amount",
              "yesterday" : "yesterday_vaccinations>amount"
            },
            "script": "(params.today - params.yesterday) / params.
    yesterday * 100"
          }
        }
      }
    }
  }
}
```

3.1.2 Percentage full covered vaccinations

The following query calculates the percentage of people which has already completed the vaccination cycle, starting from the date the vaccinations started. The percentage is calculated with respect to all the vaccinated people, so everyone that has already received at least one dose.

```
GET istat_vaccinations/_search
{
  "size" : 0,
  "aggs":{
    "group_by": {
      "date_range": {
        "field": "data_somministrazione",
        "ranges": [
          {
            "from": "2020-12-27",
            "to": "now"
          }
        ]
      },
      "aggs": {
        "sum_first_dose": {
          "sum" :{
            "field" : "prima_dose"
          }
        },
        "sum_second_dose" : {
          "sum" : {
            "field" : "seconda_dose"
          }
        },
        "sum_Janssen": {
          "filter": {
            "term" : {
              "fornitore": "Janssen"
            }
          },
          "aggs": {
            "amount" : {
              "sum" : {
                "field": "prima_dose"
              }
            }
          }
        },
        "after_infection": {
          "sum": {
            "field" : "pregressa_infezione"
```

```
     }
    },
    "full_coverage_percentage" : {
     "bucket_script": {
      "buckets_path": {
       "first_dose": "sum_first_dose",
       "second_dose": "sum_second_dose",
       "janssen_vax": "sum_Janssen>amount",
       "infection": "after_infection"
      },
      "script": "(params.second_dose + params.janssen_vax + params.
  infection)/ (params.first_dose + params.infection) * 100"
     }
    }
   }
  }
 }
}
```

3.1.3 Vaccination trend

For each day, this query returns the number of vaccinations performed.

```
GET istat_vaccinations/_search
{
 "size" : 0,
 "aggs": {
  "group_by_date": {
   "date_histogram": {
    "field": "data_somministrazione",
    "interval": "day"
   },
   "aggs": {
    "sum_vaccinations": {
     "sum": {
      "script": {
       "source": "doc['sesso_maschile'].value + doc['
  sesso_femminile'].value"
      }
     }
    }
   }
  }
 }
}
```

3.1.4 Brand administrated vaccines percentage for a given period

The following query, given a period of time, returns the percentage of the administrated vaccines per brand.

```
GET istat_vaccinations/_search
{
  "size":0,
  "aggs": {
    "group_by_date": {
      "date_range": {
        "field": "data_somministrazione",
        "ranges": [
          {
            "from": "Start date",
            "to": "End date"
          }
        ]
      },
      "aggs": {
        "total_vaccinations": {
          "sum": {
            "script": {
              "source": "doc['sesso_maschile'].value + doc['
    sesso_femminile'].value"
            }
          }
        },
        "group_by_brand": {
          "terms": {
            "field": "fornitore"
          },
          "aggs": {
            "amount": {
              "sum": {
                "script": {
                  "source": "doc['sesso_maschile'].value + doc['
    sesso_femminile'].value"
                }
              }
            }
          }
        },
        "astrazeneca_percentage": {
          "bucket_script": {
            "buckets_path": {
              "tot": "total_vaccinations",
              "astra": "group_by_brand['Vaxzevria (AstraZeneca)']>amount"
            },
            "script": "(params.astra / params.tot) * 100"
          }
        },
        "moderna_percentage": {
          "bucket_script": {
            "buckets_path": {
              "tot": "total_vaccinations",
              "moderna": "group_by_brand['Moderna']>amount"
            },
            "script": "(params.moderna / params.tot) * 100"
          }
        },
        "pfizer_percentage":{
          "bucket_script": {
            "buckets_path": {
              "tot": "total_vaccinations",
              "pfizer": "group_by_brand['Pfizer/BioNTech']>amount"
            },
            "script": "(params.pfizer / params.tot) * 100"
          }
        },
        "janssen_percentage": {
          "bucket_script": {
            "buckets_path": {
              "tot": "total_vaccinations",
              "janssen": "group_by_brand['Janssen']>amount"
            },
            "script": "(params.janssen / params.tot) * 100"
          }
        }
      }
    }
  }
}
```

3.1.5 Percentages of administrated doses, by dose number, for a given period

This query returns the percentage of first doses, second doses and boosters administrated during the given period.

```
GET istat_vaccinations/_search
{
  "size" : 0,
  "aggs": {
    "group_by_date": {
      "date_range": {
        "field": "data_somministrazione",
        "ranges": [
          {
            "from": ""Start date"",
            "to": ""End date""
          }
        ]
      },
      "aggs": {
        "first_doses": {
          "sum": {
            "script": {
              "source": "doc['prima_dose'].value + doc['
    pregressa_infezione'].value"
            }
          }
        },
        "second_doses": {
          "sum": {
            "field" : "seconda_dose"
          }
        },
        "boosters": {
          "sum": {
            "field" : "dose_addizionale_booster"
          }
        },
        "First_dose_Percentage" : {
          "bucket_script": {
            "buckets_path": {
              "First": "first_doses",
              "Second": "second_doses",
              "Booster": "boosters"
            },
            "script": "(params.First)/ (params.First + params.Second +
    params.Booster) * 100"
          }
        },
        "Second_dose_Percentage" : {
          "bucket_script": {
            "buckets_path": {
              "First": "first_doses",
              "Second": "second_doses",
              "Booster": "boosters"
            },
            "script": "(params.Second)/ (params.First + params.Second +
    params.Booster) * 100"
          }
        },
        "Booster_Percentage" : {
          "bucket_script": {
            "buckets_path": {
              "First": "first_doses",
              "Second": "second_doses",
              "Booster": "boosters"
            },
            "script": "(params.Booster)/ (params.First + params.Second +
    params.Booster) * 100"
          }
        }
      }
    }
  }
}
```

3.1.6 Vaccination percentage per age range for a given period

The following query returns the percentage of vaccinated people per age range during the given period.

```
GET istat_vaccinations/_search
{
  "size" : 0,
  "aggs": {
    "group_by_date": {
      "date_range": {
        "field": "data_somministrazione",
        "ranges": [
          {
            "from": "Start date",
            "to": "End date"
          }
        ]
      },
      "aggs": {
        "age_range": {
          "terms": {
            "field": "fascia_anagrafica"
          },
          "aggs": {
            "amount":{
              "sum": {
                "script": {
                  "source": "doc['sesso_maschile'].value + doc['
    sesso_femminile'].value"
                }
              }
            }
          }
        },
        "total": {
          "sum": {
            "script": {
              "source": "doc['sesso_maschile'].value + doc['
    sesso_femminile'].value"
            }
          }
        },
        "teen_percentage" : {
          "bucket_script": {
            "buckets_path": {
              "Teen": "age_range['12-19']>amount",
              "tot": "total"
            },
            "script": "(params.Teen) / (params.tot) * 100"
  }
},
"20s_percentage" : {
  "bucket_script": {
    "buckets_path": {
      "20": "age_range['20-29']>amount",
      "tot":"total"
    },
    "script": "(params.20)/ (params.tot) * 100"
  }
},
"30s_percentage" : {
  "bucket_script": {
    "buckets_path": {
      "30": "age_range['30-39']>amount",
      "tot": "total"
    },
    "script": "(params.30)/ (params.tot) * 100"
  }
},
"40s_percentage" : {
  "bucket_script": {
    "buckets_path": {
      "40": "age_range['40-49']>amount",
      "tot": "total"
    },
    "script": "(params.40)/ (params.tot) * 100"
  }
},
"50s_percentage" : {
  "bucket_script": {
    "buckets_path": {
      "50": "age_range['50-59']>amount",
      "tot": "total"
    },
    "script": "(params.50)/ (params.tot) * 100"
  }
},
"60s_percentage" : {
  "bucket_script": {
    "buckets_path": {
      "60": "age_range['60-69']>amount",
      "tot": "total"
    },
    "script": "(params.60)/ (params.tot) * 100"
  }
},
"70s_percentage" : {
  "bucket_script": {
    "buckets_path": {
       "70": "age_range['70-79']>amount",
       "tot": "total"
      },
      "script": "(params.70)/ (params.tot) * 100"
     }
    },
    "80s_percentage" : {
     "bucket_script": {
      "buckets_path": {
       "80": "age_range['80-89']>amount",
       "tot": "total"
      },
      "script": "(params.80)/ (params.tot) * 100"
     }
    },
    "90+_percentage" : {
     "bucket_script": {
      "buckets_path": {
       "90": "age_range['90+']>amount",
       "tot": "total"
      },
      "script": "(params.90)/ (params.tot) * 100"
     }
    }
   }
  }
 }
}
```

3.1.7 Regions ranking per number of vaccinations for a given day

The following query returns the ranking of regions per number of vaccines administrated, for a given date.

```
GET istat_vaccinations/_search
{
  "size" : 0,
  "query": {
    "bool": {
      "must":{
        "term": {
          "data_somministrazione": {
            "value": "Date"
          }
        }
      }
    }
  },
  "aggs": {
    "group_by_region": {
      "terms": {
        "field": "nome_area"
        , "size": 21
      },
      "aggs": {
        "sum_vaccinations": {
          "sum": {
            "script": {
              "source": "doc['sesso_maschile'].value + doc['
    sesso_femminile'].value"
            }
          }
        },
        "sort_by_vaccinations":{
          "bucket_sort": {
            "sort": [{ "sum_vaccinations": { "order": "desc" } }]
          }
        }
      }
    }
  }
}
```

3.1.8 Percentage of booster administrated

The following query returns percentage of people who received booster dose over those who completed the vaccination cycle at least 5 months ago and are potentially subject to booster dose.

```
GET istat_vaccinations/_search
{
  "size" : 0,
  "aggs": {
    "all_matching_docs": {
      "filters": {
        "filters": {
          "all": {
            "match_all": {}
          }
        }
      },
      "aggs":{
        "sum_booster":{
          "sum": {
            "field": "dose_addizionale_booster"
          }
        },
        "booster_candidates": {
          "filter": {
            "range": {
              "data_somministrazione": {
                "lte": "2021-12-22||-5M"
              }
            }
          },
          "aggs": {
            "sum_second_dose" : {
              "sum" : {
                "field" : "seconda_dose"
              }
            },
            "sum_janssen": {
              "filter": {
                "term" : {
                  "fornitore": "Janssen"
                }
              },
              "aggs": {
                "amount" : {
                  "sum" : {
                    "field": "prima_dose"
                  }
                }
              }
            },
            "sum_previous_infection": {
              "sum" : {
                "field" : "pregressa_infezione"
              }
            }
          }
        },
        "booster_percentage" : {
          "bucket_script": {
            "buckets_path": {
              "second_dose" : "booster_candidates>sum_second_dose",
              "janssen" : "booster_candidates>sum_janssen>amount",
              "previous_infection" : "booster_candidates>
    sum_previous_infection",
              "booster" : "sum_booster"
            },
            "script": "params.booster / (params.second_dose + params.
    janssen + params.previous_infection) * 100"
          }
        }
      }
    }
  }
}
```

3.1.9 Percentage of vaccinated people per region

This query uses both datasets and returns the percentage of vaccinated people per region over its total population.

```
GET /istat*/_search
{
  "size" : 0,
  "aggs":{
    "group_by_region":{
      "terms" : {
        "field" : "codice_regione_ISTAT"
      },
      "aggs": {
        "total_first_doses": {
          "sum": {
            "field" : "prima_dose"
          }
        },
        "total_previous_infections":{
          "sum":{
            "field": "pregressa_infezione"
          }
        },
        "total_people": {
          "sum": {
            "field" : "totale_generale"
          }
        },
        "ratio_vaccinated_people": {
          "bucket_script": {
            "buckets_path": {
              "first": "total_first_doses",
              "total" : "total_people",
              "inf": "total_previous_infections"
            },
            "script":"((params.first + params.inf)/params.total)*100"
         }
        }
      }
    }
  }
}
```

3.2 Commands

All the fields of the following commands are parameters, some will be highlighted in magenta, other won't, due to their exhaustive explanation that can be found in section 2.1.1.

3.2.1 Create new document

This command creates a new database document. Here it is reported a complete example.

```
POST /istat_vaccinations/_doc
{
  "data_somministrazione": "Date",
  "fornitore": "Supplier",
  "area": "ABR",
  "fascia_anagrafica": "Age range",
  "sesso_maschile": "1",
  "sesso_femminile": "1",
  "prima_dose": "1",
  "seconda_dose": "2",
  "pregressa_infezione": "0",
  "dose_addizionale_booster": "0",
  "codice_NUTS1": "ITF",
  "codice_NUTS2": "ITF1",
  "codice_regione_ISTAT": "13",
  "nome_area": "Abruzzo"
}
```

3.2.2 Update of number of first doses

This command, given a specific document id, updates the number of first doses. Here it is reported both the command and also the query useful to find the document to update.

Find document id
With this query it is possible to find a document id by specifing these fields: data_somministrazione, codice_regione_ISTAT, fornitore and fascia_anagrafica.

```
GET /istat_vaccinations/_search
{
  "query": {
    "bool": {
      "must": [
        {
          "match": {
            "data_somministrazione": "Date"
          }
        },
        {
          "match": {
            "codice_regione_ISTAT": "13"
          }
        },
        {
          "match": {
            "fornitore": "Supplier"
          }
        },
        {
          "match": {
            "fascia_anagrafica": "Age Range"
          }
        }
      ]
    }
  }
}
```

First doses update
Now it is possible to update the number of doses of the document found.

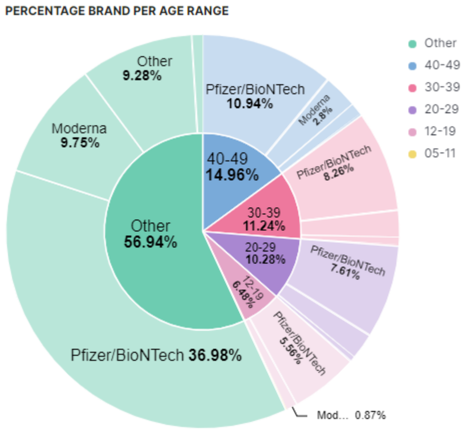
```
POST /istat_vaccinations/_update/_id
{
  "doc": {
    "prima_dose" : 3000
} }
```

4 Dashboard description

The Kibana Dashboard is made by different section, each focusing on a specific analysis. Here there is a brief description of each part is given.

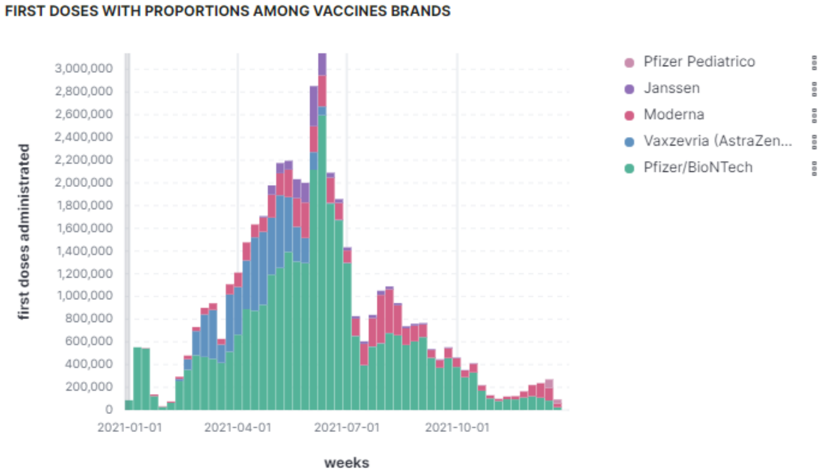
4.1 Percentage of Vaccines brand per age range

In this section a pie chart has been used to show two different things: in the inner layer there is the percentage of people vaccinated per age range, in the outer layer there is the percentage of brand vaccines used per age range.



4.2 First doses with proportion among vaccines brands

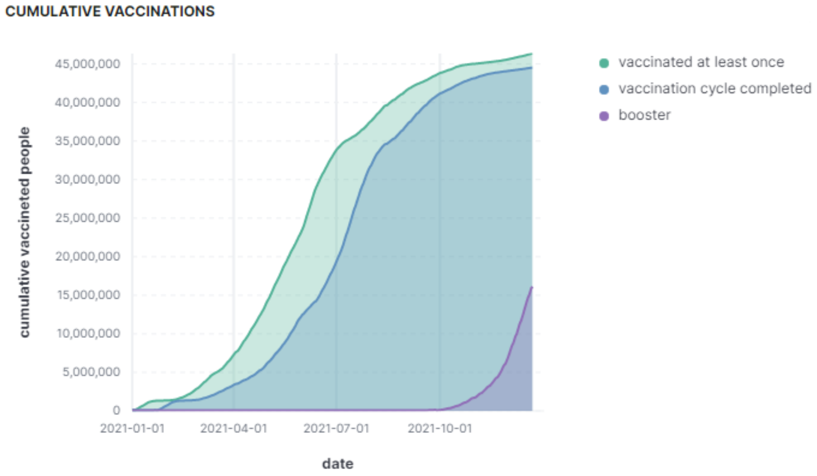
This histogram, fixed a specific interval of time, returns the amount of first doses administrated for each week. In addition the histogram shows the division per vaccine brand for each week.



4.3 Vaccinated people

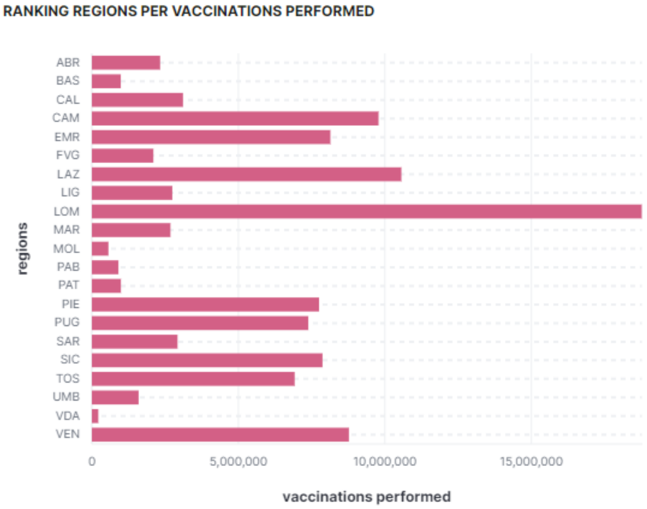
This section, given a specific period of time, returns the amount of people who have:

- vaccinated at least once
- completed the vaccination cycle
- received the booster dose



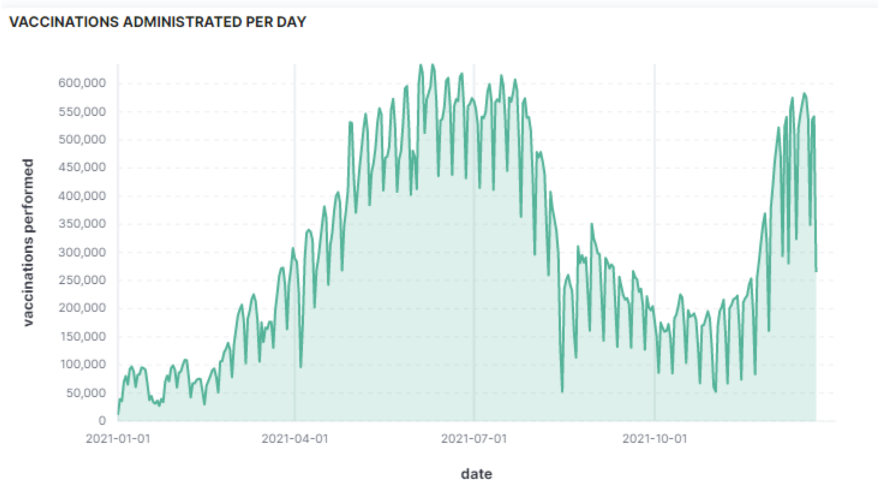
4.4 Vaccines administrated per region

Here an histogram has been used to show the amount of people who vaccinated at least once for a specific range of time per region.



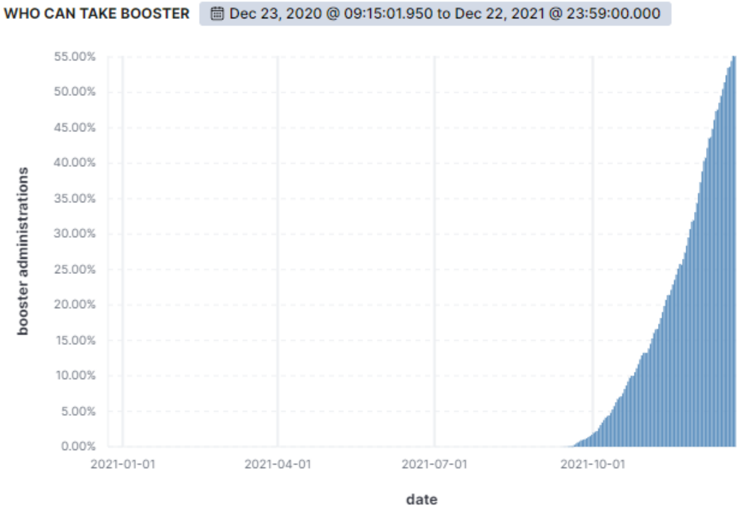
4.5 Vaccination administrated per day

This diagram shows the amount of vaccinations administrated per day during a given range of time.



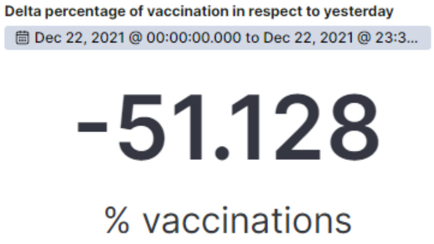
4.6 Who can take booster

This histogram, returns the percentage of people who received the booster dose for each day. The percentage has been considered with respect to all the people who are eligible to receive the booster dose, so all the people who completed the vaccination cycle at least 5 months before the analyzed date.



4.7 Delta percentage of vaccinations with respect to yesterday

It returns the percentage of the difference between vaccinations of a given date and its precedent day, calculated with respect to the amount of vaccinations performed the day before. If the vaccinations have increased, the percentage will be positive, negative otherwise. This widget is meant to be used in real time, thus it should refer to the current date. However, as the database is not up to date, the widget shows the last date available in the dataset which is "2021-12-22".



Delta percentage of vaccination in respect to yesterday

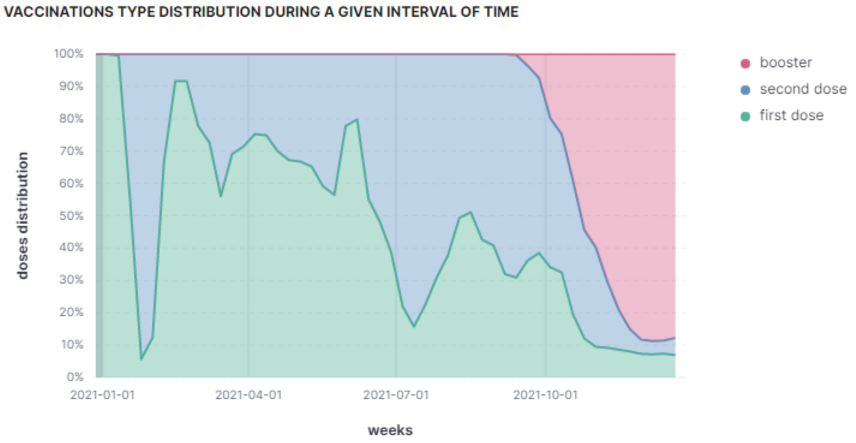
Dec 22, 2021 @ 00:00:00.000 to Dec 22, 2021 @ 23:3...

-51.128

% vaccinations

4.8 Vaccinations type distribution during a given interval of time

Given a period of time, it represents the distribution among first doses, second doses and boosters for each week.



5 User guide

5.1 Import data

5.1.1 Import vaccinations dataset

After having opened Kibana, it is necessary to click on the "upload file" button present in the home page. Now the file named "cleaned_data.csv" must be dragged and dropped in the opened page.

After that, the "import" button must be clicked. In the new page, in the advanced setting section, it is necessary to use the "istat_vaccinations" index name. In the mapping section, "codice_regione_ISTAT" type must be replaced with "keyword"(line 16).

In the Ingest pipeline section, the conversion part of "codice_regione_ISTAT" (lines from number 36 to number 42, both included) must be deleted. Then, by clicking the import button, vaccinations data will be successfully imported.

5.1.2 Import Istat population dataset

In order to import the Istat population dataset, it is necessary to repeat the same procedure described before. First, upload the "new_istat_code.csv" file and use "istat_population" as index name. Then, apply the same changes in mapping section (line 13) and ingest pipeline section (lines from number 33 to number 39, both included) as shown before.

5.2 Import dashboard

After having opened Kibana it is necessary to go in the "Stack management" section. From this page, the "Saved objects" button present in the left bar, in the Kibana section, must be clicked.

In the new page it is necessary to click the "Import data" button and then upload the "dashboard.ndjson" file.

It may appear an index conflict, in this case it is important to select as index the one referred to "istat_vaccinations". After that just click the "Import" button; the dashboard will now be visible in "Dashboard" section.

5.3 Conversion ISTAT code format from csv files

Given a csv file containing the field "codice_regione_ISTAT", it is possible to convert it to the correct with the following command:

```
"python data_cleaner.py --input original_file --output destination_file"
```

where "original_file" is the file to convert, whereas "destination_file" is the name of the file to generate with the new changes.

In order to make the script correctly work, it is necessary to have installed "pandas" package in the working enviroment.

6 Conclusion

Some interesting conclusions can be drawn from the development of this project:

Elasticsearch and Kibana are a perfect match to make different type of analysis about trends even by using a big amount of data.

Kibana can be a useful tool to display complex analysis on the available data, even without any computer science knowledge.

7 References and Sources

- Elastic Guide: https://www.elastic.co/guide/index.html
- Italian Government repository: https://github.com/italia/covid19-opendata-vaccini
- Presidenza del Consiglio dei Ministri - Dipartimento della Protezione Civile repository: https://github.com/pcm-dpc/COVID-19