AY 2021/2022

Politecnico di Milano

SMBUD 2021 - Project work delivery 3

PoliVax

Group 20

Margherita	Musumeci	(10600069)
Matteo	Nunziante	(10670132)
Piero	Rendina	(10629696)
Andrea	Sanchini	(10675541)
Enrico	Zuccolotto	(10666354)

Professor
Marco Brambilla

Sunday 9th January, 2022

Contents

1 Introduction

Since early 2020, the COVID-19 virus has represented a worldwide threat.
So far, the farmaceutical companies have been developing vaccines to tackle viral spreading. Nowadays, they represent the only way to avoid other struggling periods at home, therefore everyone must be encouraged to get vaccinated and well informed about the risks of catching the virus. Infectious-disease experts stated that vaccinating 70% to 85% of the population would enable a return to normality, nevertheless boosters may be required to prevent viral mutations from circulating.
Having said that, we need to employ an effective data-driven vaccination strategy for reaching these thresholds and targeting the anti-vaccine people.
Italy provides a well written and complete dataset[1] containing the relevant information about the vaccinations administered so far. Through the Elasticsearch engine coupled with Kibana, we query the dataset and we provide a dynamic dashboard with charts, useful to understand the trend of vaccinations and the number of fully vaccinated people.
It's worthy of attention considering the data related to our neighbouring countries. To accomplish this goal, we have included a European dataset. Moreover, a worldwide dataset has been integrated to obtain a more general picture and interesting comparisons.

[1]The dataset is available at https://github.com/italia/covid19-opendata-vaccini/dataset

2 Elasticsearch

2.1 Index

To analyze the dataset provided we have decided to exploit the automatic mapping to generate the index. It basically assigns to each integer the *long* type, to each string the *keyword* type and to the date field the correct date format and type. String-type fields use a keyword search because we need to search whole values, we have a few number of fixed terms for each field (for example the area field has only 20 values: LOM, LAZ, TOS, ...). The following table describes the meaning of each field and maps them to their data type in the index.

Field Name	Data Type	Description
Index	Long	The index of the record
Area	Keyword	Acronyms of the region of delivery
Supplier	Keyword	Complete name of the supplier of the vaccine
Administration_Date	Date (format:iso8601)	Administration date of the vaccines
Age_Group	Keyword	Age group of the people administered with the vaccines
Male_Count	Long	Number of vaccinations administered to males
Female_Count	Long	Number of vaccinations administered to females
First_Doses	Long	Number of people administered with the first dose
Second_Doses	Long	Number of people administered with the second dose
Post_Infection_Doses	Long	Number of people administered with a dose after they have been infected
Booster_Doses	Long	Number of people administered with an additional dose/recall
NUTS1_Code	Keyword	https://en.wikipedia.org/wiki/NUTS_statistical_regions_of_Italy
NUTS2_Code	Keyword	https://en.wikipedia.org/wiki/NUTS_statistical_regions_of_Italy
Region_ISTAT_Code	Long	ISTAT code of a region
Region_Name	Keyword	Name of the region (bilingual, when necessary)

2.2 Queries

1. Find the documents related to vaccinations administered in Basilicata region.

```
GET /italy_vaccine_management/_search

{
  "query":
  {
    "bool":
    {
      "filter":
      {
        "term":
        {
          "Area": "BAS"
        }
      }
    }
  }
}
```

2. Find the documents in which more than a thousand first doses of the same type of vaccine have been given to people between the ages of 80 and 89, and there is at least one person who completed his primary vaccination cycle with just one dose[2].

```
GET /italy_vaccine_management/_search
```

```
{
  "query":
  {
    "bool":
    {
      "must":
      [
        {
          "term":
          {
            "Age_Group": "80-89"
          }
        },
        {
          "range":
          {
            "First_Doses":
            {
              "gte":1000
            }
          }
        }
      ],
      "should":
      {
        "range":
        {
          "Post_Infection_Doses":
          {
            "gte": 0
          }
        }
      }
    }
  }
}
```

[2]This last assertion is checked by testing the **post infection doses** value

3. Find the vaccinations documents related to people younger than sixty years.

```
GET /italy_vaccine_management/_search
```

```
{
  "query":
  {
    "bool":
    {
      "must_not":
      [
        {
          "term":
          {
            "Age_Group": "90+"
          }
        },
        {
          "term":
          {
            "Age_Group" : "80-89"
          }
        },
        {
          "term":
          {
            "Age_Group" : "70-79"
          }
        },
        {
          "term":
          {
            "Age_Group" : "60-69"
          }
        }
      ]
    }
  }
}
```

4. Find the entries in which more than ten thousands first doses have been issued to people in Lombardia region and the vaccine type is different from Pfizer/BioNTech.

```
GET /italy_vaccine_management/_search
```

```
{
  "query":
  {
    "bool":
    {
      "must":
      {
        "range":
        {
          "First_Doses":
          {
            "gte":10000
          }
        }
      },
      "must_not":
      {
        "term":
        {
          "Supplier" : "Pfizer/BioNTech"
        }
      },

      "filter":
      {
        "term":
        {
          "Area": "LOM"
        }
      }
    }
  }
}
```

5. Find the number of vaccinations administered for each region subdivided for age group.

```
GET /italy_vaccine_management/_search
```

```
{
  "size":0,
  "aggs":
  {
    "region":
    {
      "terms":
      {
        "field": "Area",
        "size": 20
      },
      "aggs":
      {
        "age":
        {
          "terms":
          {
            "field": "Age_Group"
          },
          "aggs":
          {
            "doses":
            {
              "sum":
              {
                "script" :
                {
                  "lang": "painless",
                  "source": "doc['First_Doses'].value +
                             doc['Second_Doses'].value +
                             doc['Booster_Doses'].value +
                             doc['Post_Infection_Doses'].value"
                }
              }
            }
          }
        }
      }
    }
  }
}
```

6. Find the total number of vaccinations administered in Italy.

```
GET /italy_vaccine_management/_search
```

```
{
  "size":0,
  "aggs":
  {
    "doses":
    {
      "sum":
      {
        "script":
        {
          "lang": "painless",
          "source": "doc['First_Doses'].value +
                     doc['Second_Doses'].value +
                     doc['Booster_Doses'].value +
                     doc['Post_Infection_Doses'].value"
        }
      }
    }
  }
}
```

7. Find the top three months for number of vaccines administered.

```
GET /italy_vaccine_management/_search
```

```
{
  "size": 0,
  "aggs":
  {
    "vaccines_per_month":
    {
      "date_histogram":
      {
        "field": "Administration_Date",
        "calendar_interval": "month"
      },
      "aggs":
      {
        "doses":
        {
          "sum":
          {
            "script":
            {
              "lang": "painless",
              "source": "doc['First_Doses'].value +
                         doc['Second_Doses'].value +
                         doc['Booster_Doses'].value +
                         doc['Post_Infection_Doses'].value "
            }
          }
        },
        "top_3_months":
        {
          "bucket_sort":
          {
            "sort":
            [
              {
                "doses":
                {
                "order": "desc"
                }
              }
            ],
            "size": 3
          }
        }
      }
    }
  }}
```

8. Find the region with the lowest number of booster doses.

```
GET /italy_vaccine_management/_search
```

```
{
  "size": 0,
  "aggs":
  {
    "region":
    {
      "terms":
      {
        "field": "Region_Name"
      },
      "aggs":
      {
        "doses":
        {
          "sum":
          {
            "field": "Booster_Doses"
          }
        },
        "worst_region":
        {
          "bucket_sort":
          {
            "sort":
            [
              {
                "doses":
                {
                  "order": "asc"
                }
              }
            ],
            "size": 1
          }
        }
      }
    }
  }
}
```

9. Find the number of vaccinations administered for each type of vaccine.

```
GET /italy_vaccine_management/_search
```

```
{
  "size":0,
  "aggs": {
    "vaccine":
    {
      "terms":
      {
        "field": "Supplier"
      },
      "aggs":
      {
        "doses":
        {
          "sum":
          {
            "script" :
            {
              "lang": "painless",
              "source": "doc['First_Doses'].value +
                         doc['Second_Doses'].value +
                         doc['Booster_Doses'].value +
                         doc['Post_Infection_Doses'].value"
            }
          }
        }
      }
    }
  }
}
```

10. Retrieve the distribution of the vaccinations through the different age groups.

```
GET /italy_vaccine_management/_search
```

```
{
  "size": 0,
  "aggs":
  {
    "age":
    {
      "terms":
      {
        "field": "Age_Group"
      },
      "aggs":
      {
        "doses":
        {
          "sum":
          {
            "script" :
            {
              "lang": "painless",
              "source": "doc['First_Doses'].value +
                         doc['Second_Doses'].value +
                         doc['Booster_Doses'].value +
                         doc['Post_Infection_Doses'].value"
            }
          }
        },
        "percent_of_total_doses":
        {
          "normalize":
          {
            "buckets_path": "doses",
            "method": "percent_of_sum",
            "format": "00.00%"
          }
        }
      }
    }
  }
}
```

11. Retrieve the percentage of people who were issued with Moderna vaccine as first dose.

```
GET /italy_vaccine_management/_search
```

```
{
  "size": 0,
  "aggs": {
    "all": {
      "terms": {
        "script": "1"
      },
      "aggs": {
        "total_doses": {
          "sum": {
            "field": "First_Doses"
          }
        },
        "Moderna": {
          "filter": {
            "term": {
              "Supplier": "Moderna"
            }
          },
          "aggs": {
            "doses": {
              "sum": {
                "field": "First_Doses"
              }
            }
          }
        },
        "percentage": {
          "bucket_script": {
            "buckets_path": {
              "Mod": "Moderna>doses",
              "Tot": "total_doses"
            },
            "script": "params.Mod / params.Tot * 100"
          }
        }
      }
    }
  }
}
```

12. Retrieve the ratio between people older than 90 that received the booster dose and the people, older than 90, that received the second dose, grouped by region.

```
GET /italy_vaccine_management/_search
```

```
{
  "query":
  {
    "bool":
    {
      "filter":
      {
        "term":
        {
          "Age_Group": "90+"
        }
      }
    }
  },
  "size": 0,
  "aggs":
  {
    "region":
    {
      "terms":
      {
        "field": "Region_Name",
        "size":  20
      },
      "aggs":
      {
        "2nd_doses":
        {
          "sum":
          {
             "field": "Second_Doses"
          }
        },
        "3rd_doses":
        {
         "sum":
         {
            "field": "Booster_Doses"
          }
        },
        "rate":
        {
          "bucket_script":
          {
```

```
            "buckets_path":
            {
              "third": "3rd_doses",
              "second": "2nd_doses"
            },
            "script": "params.third / params.second *100"
          }
        }
      }
    }
  }
}
```

13. Retrieve the percentage of people vaccinated with previous infections w.r.t. the total number of vaccinations for each region.

```
GET /italy_vaccine_management/_search
```

```
{
  "size": 0,
  "aggs":
  {
    "region":
    {
      "terms":
      {
        "field": "Region_Name",
        "size":  20
      },
      "aggs":
      {
        "tot_doses":
        {
          "sum":
          {
             "script" :
            {
              "lang": "painless",
              "source": "doc['First_Doses'].value +
                         doc['Second_Doses'].value +
                         doc['Booster_Doses'].value +
                         doc['Post_Infection_Doses].value"
            }
          }
        },
        "inf":
        {
         "sum":
         {
            "field": "Post_Infection_Doses"
          }
        },
        "rate":
        {
          "bucket_script":
          {
            "buckets_path":
            {
              "infect": "inf",
              "tot": "tot_doses"
            },
            "script": "params.infect / params.tot*100"
          }}}}}}
```

14. Find all the days in which more than 400000 first doses have been administered.

```
GET /italy_vaccine_management/_search
```

```
{
  "size": 0,
  "aggs":
  {
    "first_doses_per_day":
    {
      "date_histogram":
      {
        "field": "Administration_Date",
        "calendar_interval": "day"
      },
      "aggs":
      {
        "doses":
        {
          "sum":
          {
            "field":"First_Doses"
          }
        },
        "gte400000":
        {
          "bucket_selector":
          {
            "buckets_path":
            {
              "numberOfDoses": "doses"
            },
            "script": " params.numberOfDoses>400000"
          }
        }
      }
    }
  }
}
```

15. Retrieve the total number of male and female vaccinations.

```
GET /italy_vaccine_management/_search
```

```
{
  "size": 0,
  "aggs":
  {
    "male_doses":
    {
      "sum":
      {
        "field": "Male_Count"
      }
    },
    "female_doses":
    {
      "sum":
      {
        "field": "Female_Count"
      }
    }
  }
}
```

16. Compute the amount of the first doses issued between April 10th 2021 and April 18th 2021.

```
GET /italy_vaccine_management/_search
```

```
{
  "query":{
    "bool": {
      "must": {
        "range": {
          "Administration_Date": {
            "gte": "2021-04-10T00:00:00",
            "lte": "2021-04-18T00:00:00"
          }
        }
      }
    }
  },
  "size": 0,
  "aggs":
  {
    "first_doses":
    {
      "sum":
      {
        "field": "First_Doses"
      }
    }
  }
}
```

17. Find the documents related to AstraZeneca vaccines administered in Lombardia region to people aged between 12 and 19.

```
GET /italy_vaccine_management/_search
```

```
{
  "query": {
    "bool": {
      "must": [
        {
          "term":
          {
            "Supplier": "Vaxzevria (AstraZeneca)"
          }
        },
        {
          "term":
          {
            "Area": "LOM"
          }
        },
        {
          "term":
          {
            "Age_Group": "12-19"
          }
        }
      ]
    }
  }
}
```

18. Find the amount of Pfizer/BioNTech vaccinations, grouped by sex, administered in Lazio on October 31th to people aged between 80 and 89.

```
GET /italy_vaccine_management/_search
```

```
{
   "query": {
     "bool": {
       "filter": [
         {
           "term":
           {
             "Area": "LAZ"
           }
         },
         {
           "term":
           {
             "Administration_Date": "2021-10-31T00:00:00"
           }
         },
         {
           "term":
           {
             "Age_Group": "80-89"
           }
         },
         {
           "term":
           {
             "Supplier": "Pfizer/BioNTech"
           }
         }
       ]
     }
   },
   "aggs":
   {
     "male_doses":
     {
       "sum":
       {
         "field": "Male_Count"
       }
     },

     "female_doses":
     {
       "sum":
       {
```

```
        "field": "Female_Count"
      }
    }
  }
}
```

2.3 Commands

1. Command to add 10 **booster doses** from Pfizer/BioNTech supplier, issued today to males aged between 20 and 29 in Friuli Venezia Giulia region.

   ```
   POST /italy_vaccine_management/_update_by_query
   ```

   ```
   {
     "query": {
       "bool": {
         "must": [
           {
             "term": {
               "Administration_Date": "now"
               }
           },
           {
             "term": {
               "Supplier": "Pfizer/BioNTech"
             }
           },
           {
             "term": {
               "Age_Group": "20-29"
             }
           },
           {
             "term": {
               "Region_Name": "Friuli-Venezia Giulia"
             }
           }
         ]
       }
     },
     "script": {
       "source": "ctx._source.Booster_Doses+=10;
                  ctx._source.Male_Count+=10;",
       "lang": "painless"
     }
   }
   ```

2. Command to add a new field inside all the documents to show the total number of vaccines.

```
POST /italy_vaccine_management/_update_by_query
```

```
{
  "script": {
    "source": "ctx._source.Total_Doses =
               ctx._source.First_Doses +
               ctx._source.Second_Doses +
               ctx._source.Booster_Doses +
               ctx._source.Post_Infection_Doses",
    "lang": "painless"
  }
}
```

3. The following command triggers the creation of a new field named *post-second_dose_infections*. It comes in handy to analyze the number of people who caught the virus within the last vaccination and the booster dose. Moreover, its purpose is also to provide an overview of the vaccines' efficacy.

```
POST /italy_vaccine_management/_update_by_query
```

```
{
  "query": {
    "bool": {
      "filter": [
        {"term": {
          "Administration_Date": "now"
        }}
      ],
      "must": [
        {
          "term": {
          "Region_Name": "Lombardia"
          }
        },
        {
          "term": {
            "Age_Group":"30-39"
          }
        },
        {
          "term": {
            "Supplier": "Pfizer/BioNTech"
          }
        }
      ]
    }
  },
```

```
  "script": {
    "params": {"number_of_infections": 1000},
    "source": "ctx._source.Post_Second_Dose_infections =
    params.number_of_infections",
    "lang": "painless"
  }
}
```

4. Generic update of all the numerical fields in a document.

   `POST /italy_vaccine_management/_update_by_query`

```
{
  "query": {
    "bool": {
      "must": [
        {
          "term": {
            "Administration_Date": "now-3d/d"
            }
        },
        {
          "term": {
            "Supplier": "Moderna"
          }
        },
        {
          "term": {
            "Age_Group": "20-29"
          }
        },
        {
          "term": {
            "Region_Name": "Lombardia"
          }
        }
      ]
    }
  },
  "script": {
    "params": {
                "booster_doses": 10,
                "first_doses": 10,
                "second_doses": 10,
                "male_count": 15,
                "female_count": 15,
                "post_infection_doses": 10
            },
    "source": "ctx._source.Booster_Doses = params.booster_doses;
               ctx._source.First_Doses = params.first_doses;
               ctx._source.Second_Doses = params.second_doses;
```

```
                ctx._source.Male_Count = params.male_count;
                ctx._source.Female_Count = params.female_count;
                ctx._source.Post_Infection_Doses = params.post_infection_doses;
                ctx._source.Total_Doses =
                params.first_doses + params.second_doses + params.booster_doses+
                params.post_infection_doses",
    "lang": "painless"
  }
}
```

3 Dashboard

In order to visualize, analyze and interact with data sets, it's been created a user interface using Kibana.

The dashboard is made of different components:

1. Three panels showing the trend of vaccinations in Italy

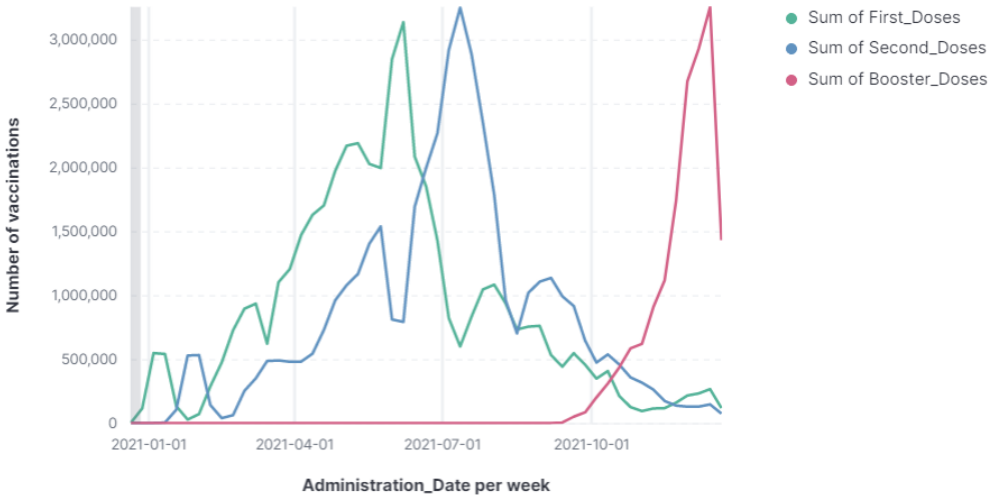


Figure 1: *Vaccinations divided by first,second and third doses*

Figure 2: *All vaccinations divided by male and female*

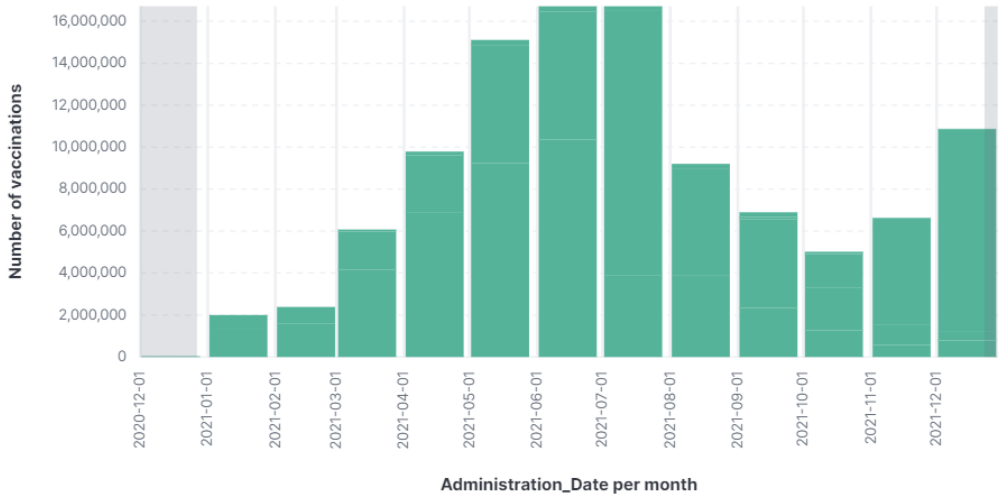


Figure 3: *All vaccinations did group by month*
This panels is linked to query 7

2. A panel describing the trend of vaccinations in Lombardia region. This is possible thanks to a filter applied on the data:

```
{
  "query":
  {
    "match":
    {
      "Region_Name": "Lombardia"
    }
  }
}
```

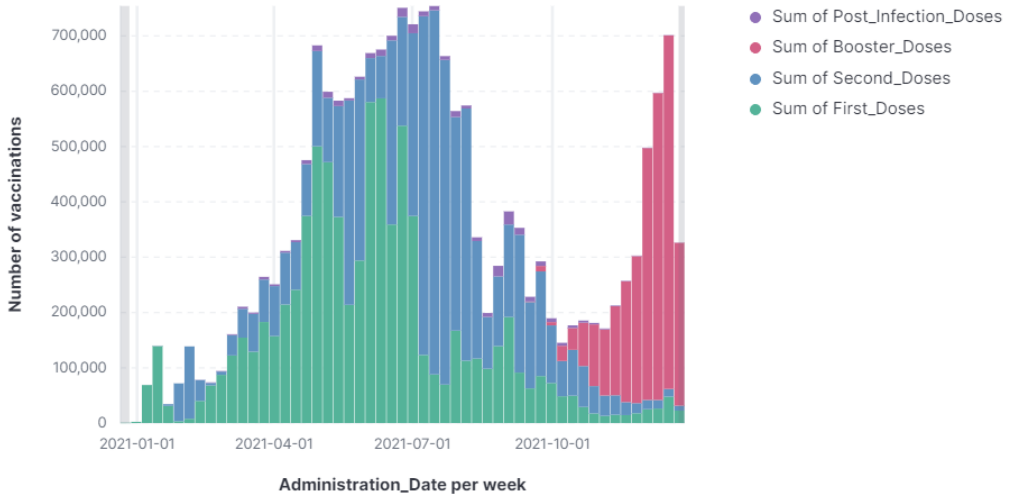


Figure 4: *All vaccinations in Lombardia*

3. Three panels describing the trends of first, second and booster doses respectively. In each of them is been added a second layer in order to show the linear trend

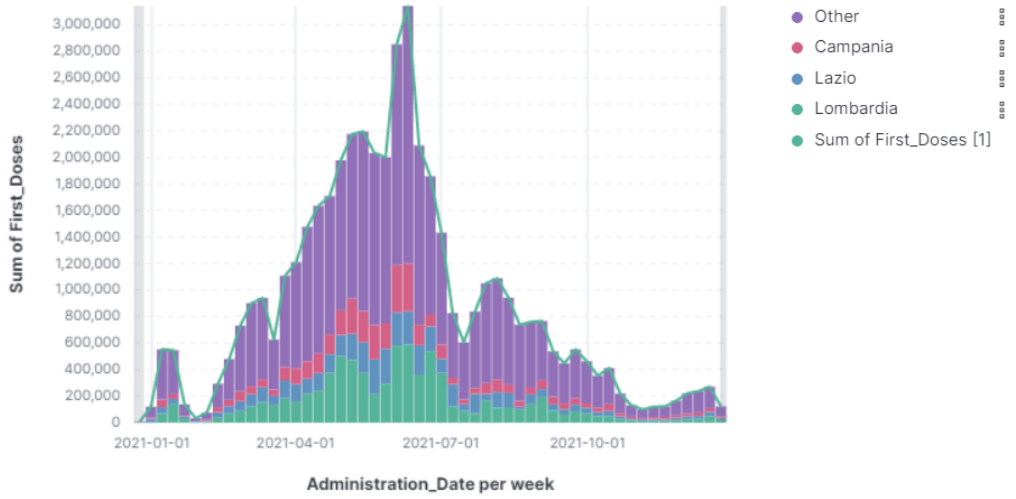


Figure 5: *Trend of first doses*
This panel is linked to query 14

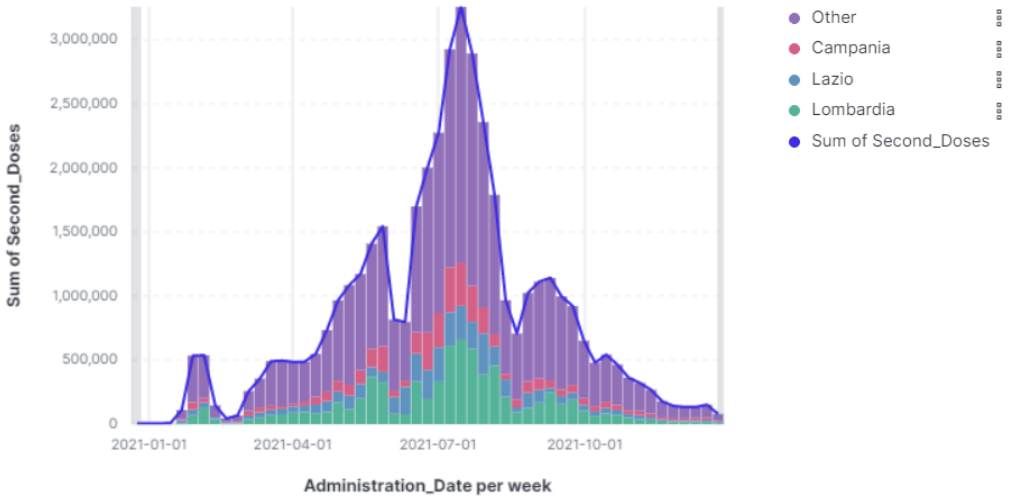


Figure 6: *Trend of second doses*

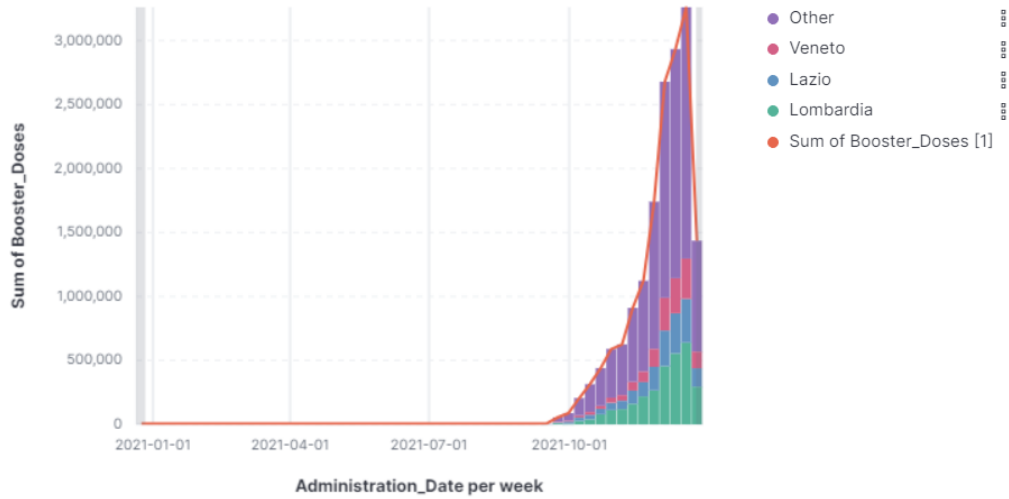


Figure 7: *Trend of booster doses*

4. Three panels describing the density of vaccinations in each region

Figure 8: *Density of first doses*

Figure 9: *Density of second doses*

Figure 10: *Density of booster doses*

5. Four panels that show directly the total amount of first, second, booster doses and the total amount of *Post Infection Doses*

Sum of First_Doses

46,403,899

Sum of Second_Doses

43,033,634

Sum of Booster_Doses

16,404,968

Sum of Post_Infection_Doses

1,639,057

6. Three panels describing the distribution of vaccine's type over vaccinations

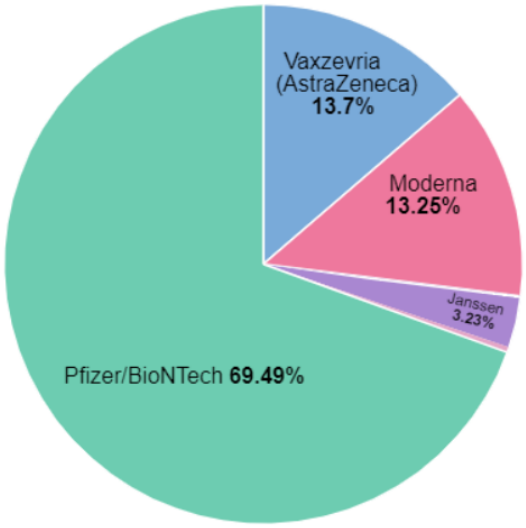


Figure 11: *Distribution in first doses*
This panel is linked to query 11

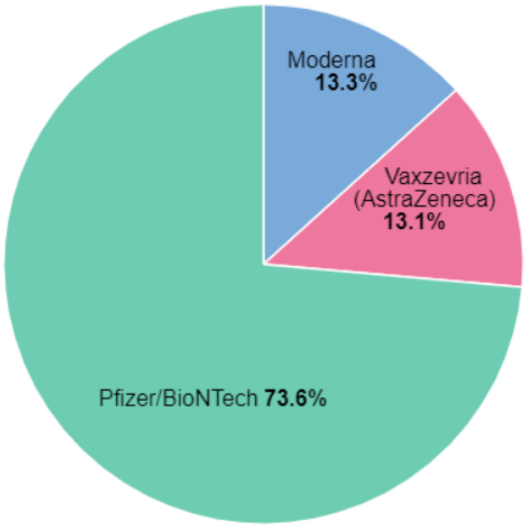


Figure 12: *Distribution in second doses*

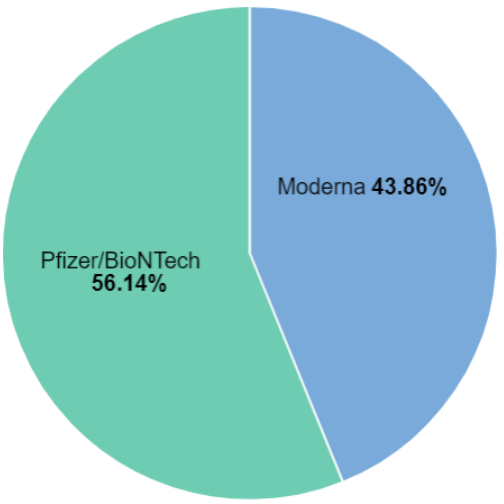


Figure 13: *Distribution in booster doses*

7. A panel that show the maximum number of second doses done in one day in one region

Max of Second_Doses

37,640

8. Three panels describing the distribution of the doses over the age and the distribution of *Post Infection Doses* over the age

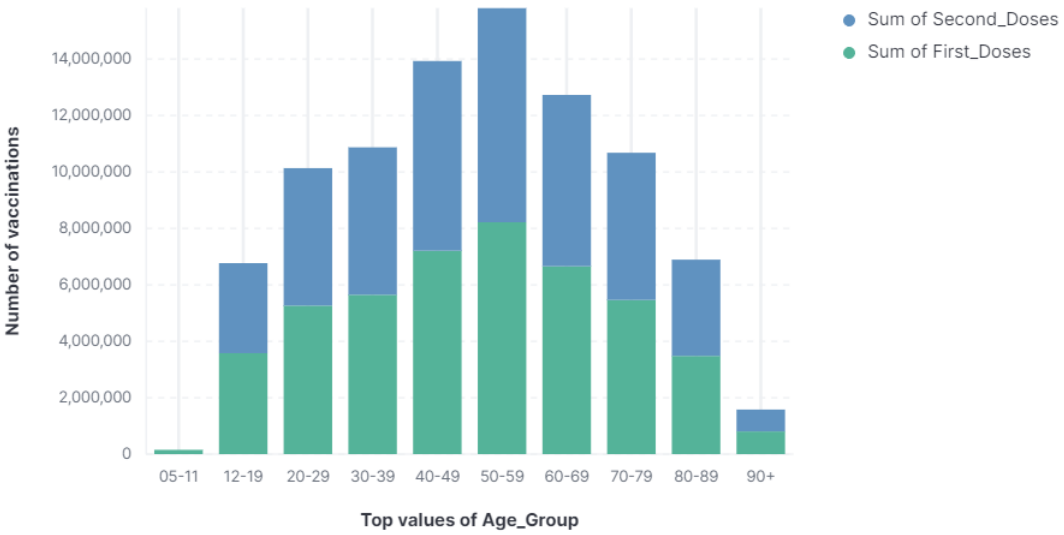


Figure 14: *Distribution of the first/second doses over the age*

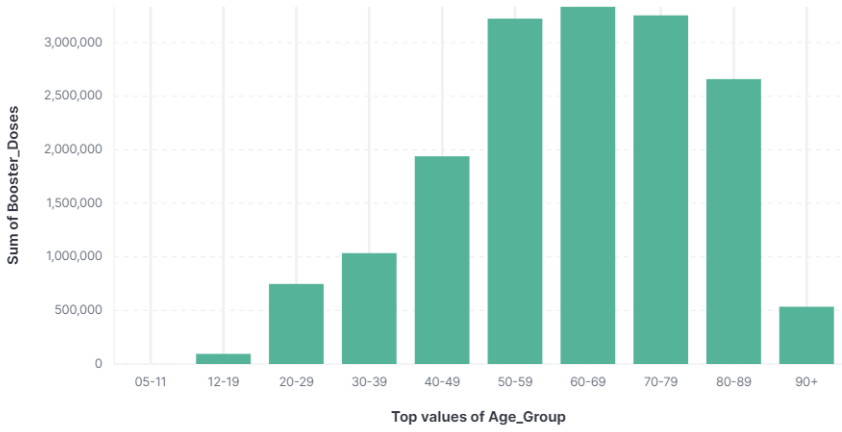


Figure 15: *Distribution of booster doses over the age*

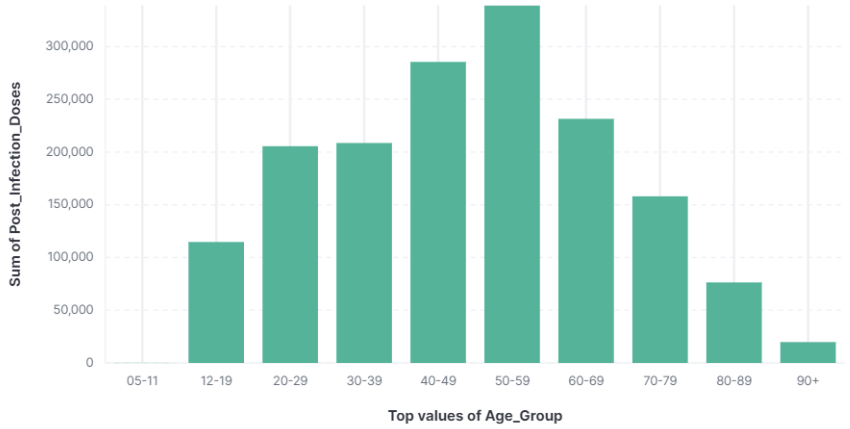


Figure 16: *Distribution of post infection doses over the age*

9. A panel describing the distribution of the vaccine's types over the total amount of vaccinations

Figure 17: *Distribution of vaccine's producers over the total vaccinations*
This panel is linked to query 9

10. A heat map that shows the relation between some regions and the last two months of vaccinations. The value is the number of second vaccinations. The constraints on the dates it's possible thank to the following filter applied on the data:

```
{
  "query": {
    "bool": {
      "filter": {
        "range": {
          "Administration_Date": {
            "gte": "2021-11-01"
          }
        }
      }
    }
  }
}
```

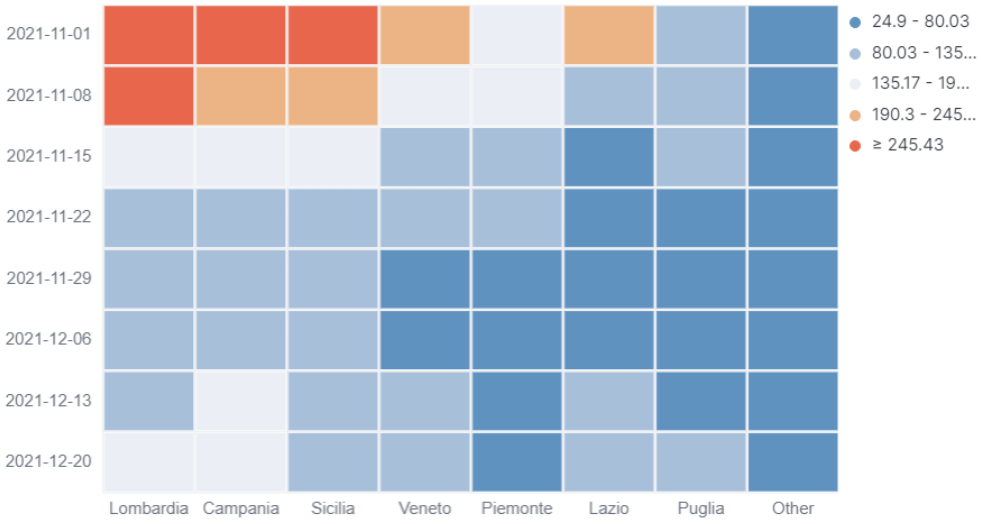


Figure 18: *Dates/regions relations*

11. A panel that shows the trend of post infection doses.

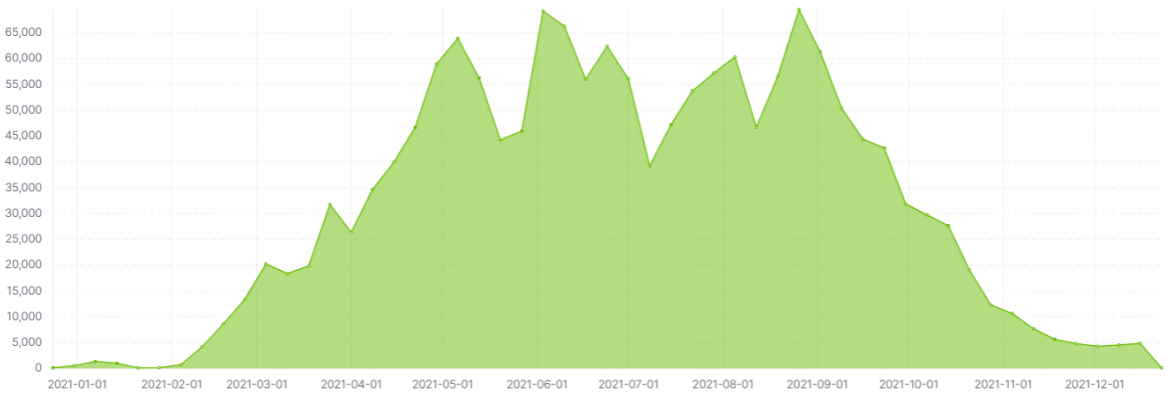


Figure 19: *Number of post infection doses over the time*

12. A panel describing the distribution of doses over the NUT1 areas

Figure 20: *Distribution of doses in NUT1 areas*

NUT1 codes correspond to the following areas:

- ITC: northwest Italy
- ITF: south Italy
- ITG: insular Italy
- ITH: northeast Italy
- ITI: central Italy

13. Three tables describing numerically the amount of vaccines used as first, second and third dose

Top values of NUTS1_Code	Pfizer/BioNTech › Sum of First_Doses	Vaxzevria (AstraZeneca) › Sum of First_Doses	Moderna › Sum of First_Doses
ITC	8,654,593	1,818,421	1,754,498
ITF	7,263,950	1,560,990	1,333,494
ITI	6,443,289	1,303,874	1,232,181
ITH	6,346,330	1,165,913	1,182,894
ITG	3,567,700	512,756	653,121
	Sum: 32,275,862	**Sum: 6,361,954**	**Sum: 6,156,188**

Figure 21: *Numerical description of vaccine's types used as first dose*

Top values of NUTS1_Code	Pfizer/BioNTech › Sum of Second_Doses	Moderna › Sum of Second_Doses	Vaxzevria (AstraZeneca) › Sum of Second_Doses
ITC	8,563,220	1,645,126	1,578,806
ITF	7,132,499	1,238,306	1,367,784
ITI	6,302,390	1,186,853	1,175,600
ITH	6,209,616	1,073,881	1,065,330
ITG	3,483,771	581,155	453,419
	Sum: 31,691,496	**Sum: 5,725,321**	**Sum: 5,640,939**

Figure 22: *Numerical description of vaccine's types used as second dose*

Top values of NUTS1_Code	Pfizer/BioNTech › Sum of Booster_Doses	Moderna › Sum of Booster_Doses	Vaxzevria (AstraZeneca) › Sum of Booster_Doses
ITI	2,401,867	1,130,229	0
ITF	2,386,735	1,338,077	0
ITC	2,167,934	2,632,111	0
ITH	1,769,120	1,629,049	0
ITG	717,261	649,004	0
	Sum: 9,442,917	**Sum: 7,378,470**	**Sum: 0**

Figure 23: *Numerical description of vaccine's types used as third dose*

14. A table showing numerically the distribution of booster doses over ages in the top five regions (top five by number of booster doses)

Top values of Age_Group	Lombardia › Sum of Booster_Dos	Lazio › Sum of Booster_Doses	Veneto › Sum of Booster_Doses	Campania › Sum of Booster_Dos	Emilia-Romagna › Sum of Booste
05-11	0	0	0	0	0
12-19	14,646	10,580	5,300	15,576	5,041
20-29	149,123	62,583	59,721	83,699	54,684
30-39	225,874	101,612	71,011	108,254	71,315
40-49	357,727	220,042	169,513	206,063	126,905
50-59	614,698	356,957	284,270	327,118	218,776
60-69	594,136	364,081	304,598	279,601	256,006
70-79	540,451	348,209	300,960	249,591	249,903
80-89	487,125	252,146	225,073	158,269	240,369
90+	95,429	49,126	45,982	23,560	54,719

Figure 24: *Numerical description of the booster vaccinations distributed over the ages for some regions. This panel is linked to the query 5*

In addition to panels for visualizing and checking data, in the dashboard are also present four control panels that allow the user to interact with data by filtering them according to his interests. The filters are applied to all the components in the dashboard. Below they are presented with some practical examples.

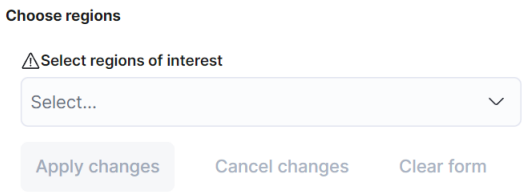


Figure 25: *Control for filtering data according to the region*

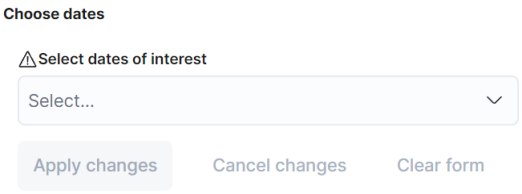


Figure 26: *Control for filtering data according to the date*

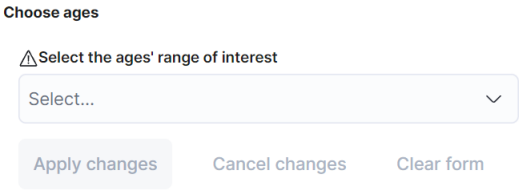


Figure 27: *Control for filtering data according to the age group*

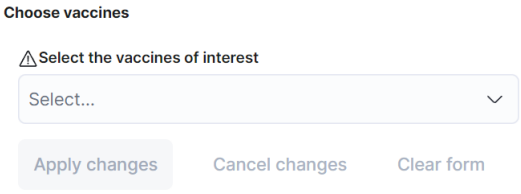


Figure 28: *Control for filtering data according to the vaccine's type*

Examples:

- Filtering on region

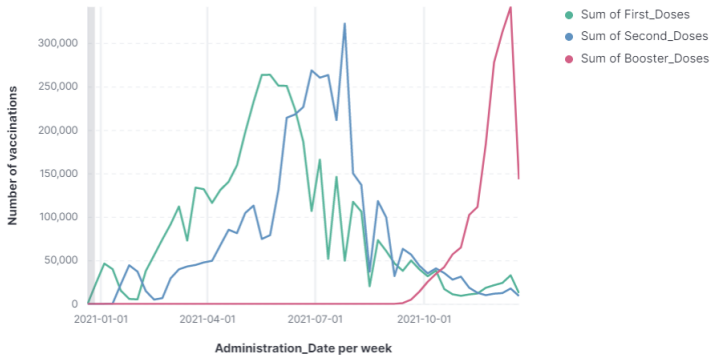


Figure 29: *All the vaccinations done in Lazio*

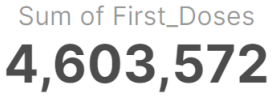


Figure 30: *Number of first doses done in Lazio*

Figure 31: *Density of the first doses done in Lazio and in Campania*

- Filtering on dates

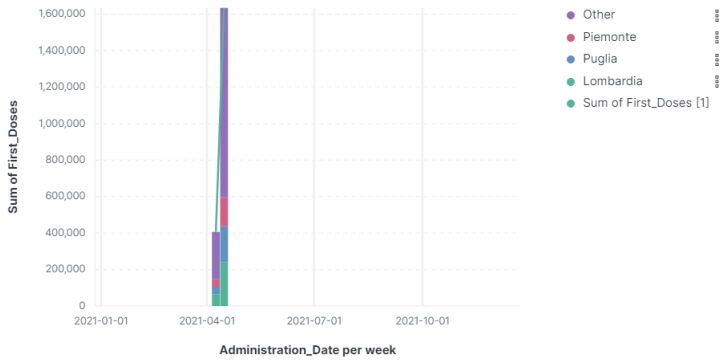


Figure 32: *Total of the first doses done between April 10th and April 18th This panel is linked to the query 16*

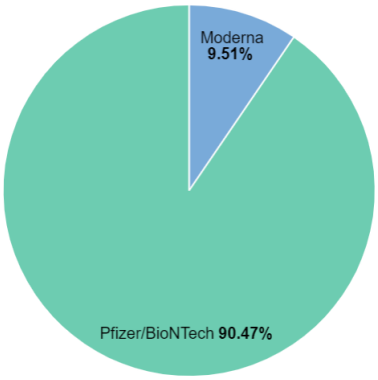


Figure 33: *Distribution of vaccine's types for the second dose between April 10th and April 18th*

- Filtering on ages

Figure 34: *Density of booster doses for people aged between 40 and 49*

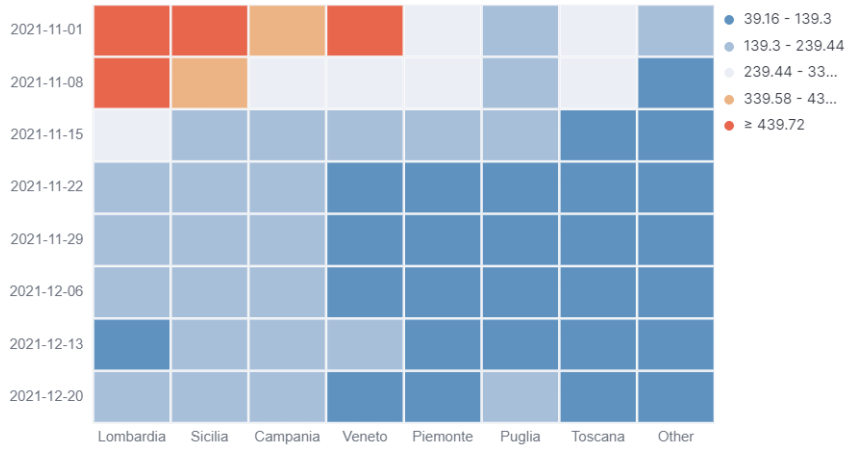


Figure 35: *Heat map date/region for person in the range of 40-49 age with values the number of second doses*

- Filtering on vaccine's type

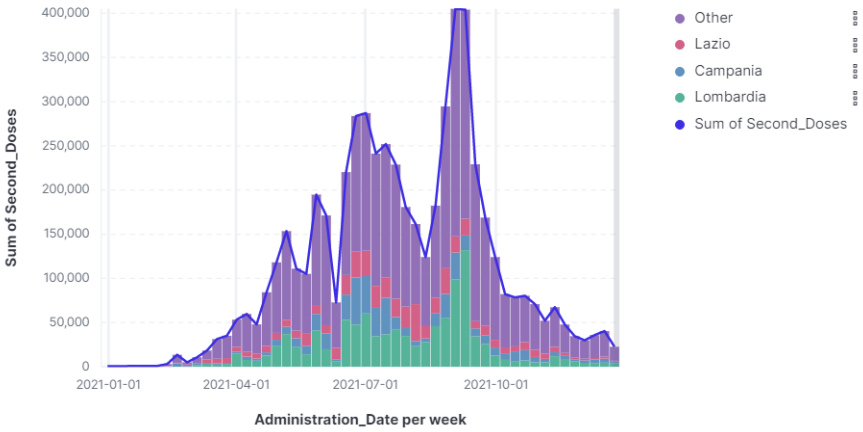


Figure 36: *Trend of second doses done with Moderna*

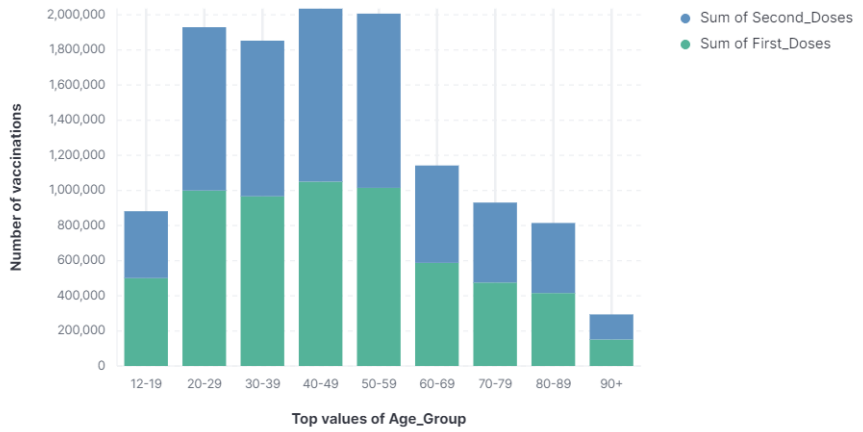


Figure 37: *Distribution of first/second doses over the age for vaccinations done with Moderna*

- Combination of the previous filters

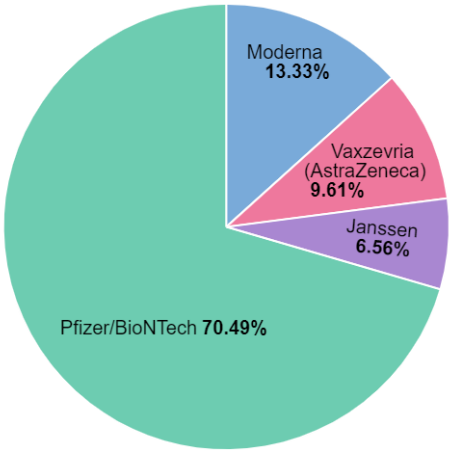


Figure 38: *Distribution of vaccine's type of the first doses done in Lombardia to people aged between 40 and 49 and between 50 and 59*

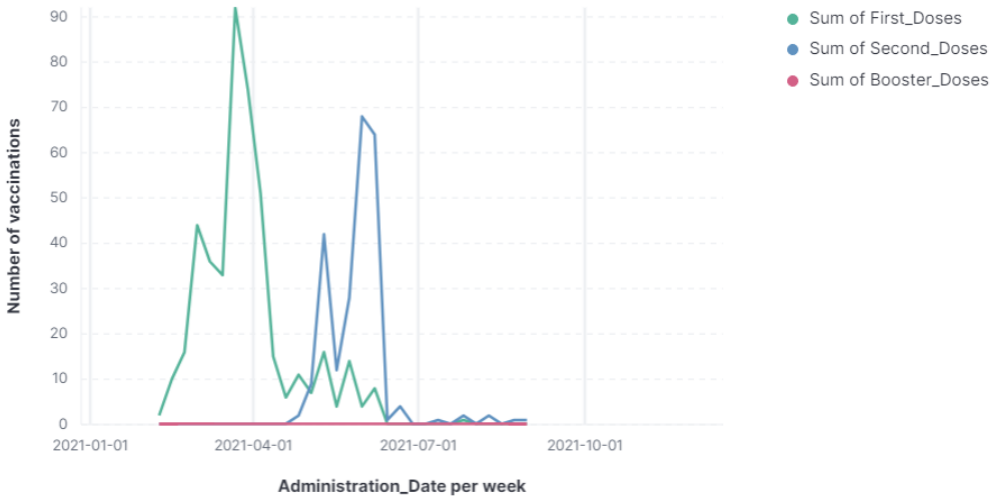


Figure 39: *All the vaccinations of AstraZeneca done in Lombardia to person with age between 12 and 19. This panel is linked to query 17*

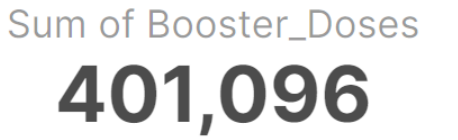


Figure 40: *Total number of booster doses done in Veneto between December 10 and December 18*

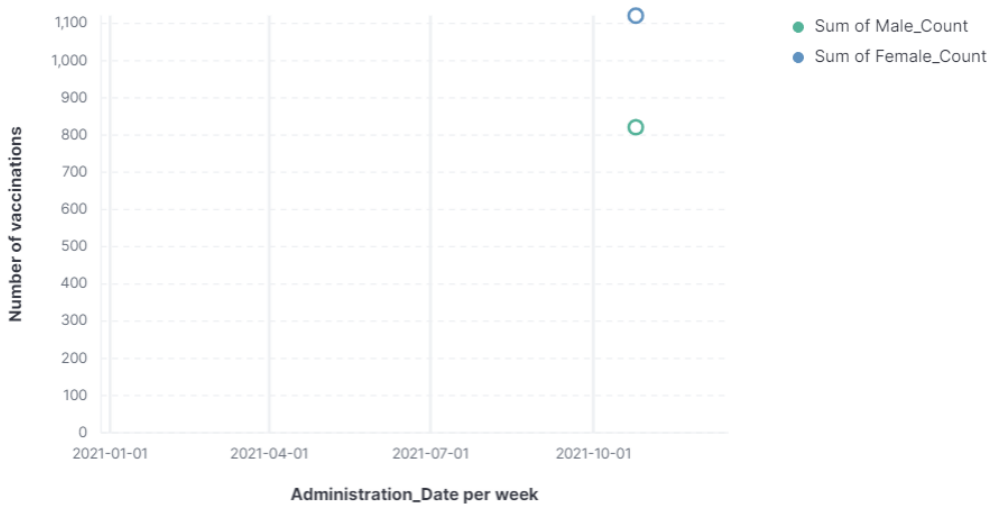


Figure 41: *Amount of Pfizer/BioNTech vaccinations divided by sex done on Lazio in October 31st to person in the range of 80-89 ages. This panel is linked to query 18*

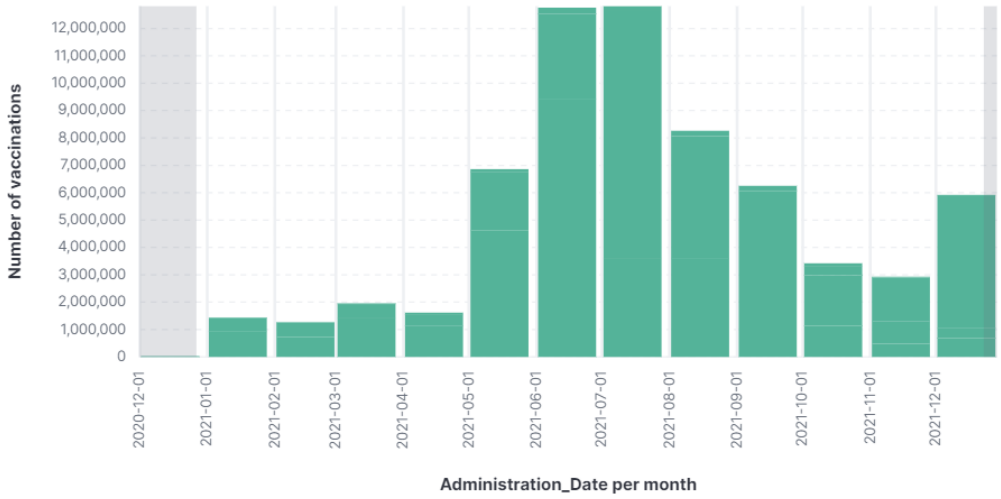


Figure 42: *All vaccinations done to persons younger than 60 years old This panels is linked to query 3*

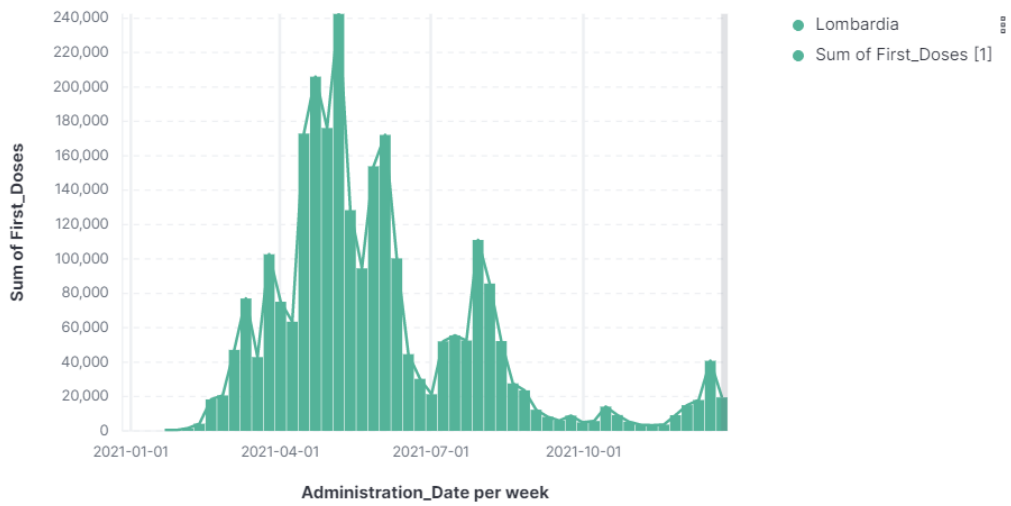


Figure 43: *All the first doses done in Lombardia with vaccine's type different from Pfizer/BioNTech. This panel is linked to query 4*

4 Optional 1: European and World vaccinations

4.1 Datasets description

To give a broader view on the progress of the vaccination campaign in different countries, **two additional datasets** have been integrated and analyzed:

- The first one, named ***owid_ covid_ data***[3], is a very large dataset that contains several information related to the SARS-COVID19 pandemic, concerning almost all countries in the world. Some attributes of the dataset, especially those of lesser importance with respect to the delivery topic and those for which many rows had null values, have been eliminated.
  To generate the elasticsearch index, it was sufficient to use the automatic mapping. Most of the fields in fact contain numeric values and have been correctly mapped to double, while the few string fields concerning geographic information have been mapped to keywords as text analysis is not required.

Field Name	Data Type	Description
index	Long	The index of the record
continent	Keyword	Continent of the reporting country
date	Date (format: iso8601)	Date at the time of reporting
gdp_per_capita	Double	GDP per capita of the reporting country
human_development_index	Double	HDI of the reporting country
iso_code	Keyword	ISO code of the reporting country
location	Keyword	The reporting country
median_age	Double	Median Age in the reporting country
new_cases	Double	New covid cases
new_cases_per_million	Double	New covid cases per million
new_deaths	Double	New deaths due to covid
new_deaths_per_million	Double	New deaths per million due to covid
new_tests	Double	New tests performed
new_tests_per_thousand	Double	New tests performed per thousand
new_vaccinations	Double	New vaccinations
people_fully_vaccinated	Double	Number of fully vaccinated people
people_fully_vaccinated_per_hundred	Double	Number of fully vaccinated people per hundred
people_vaccinated_per_hundred	Double	Number of vaccinated people per hundred
people_vaccinated	Double	Number of partially vaccinated people
population	Double	Population of the reporting country
population_density	Double	Population density of the reporting country
positive_rate	Double	Positive rate of the reporting country
total_boosters	Double	Total administered booster doses
total_boosters_per_hundred	Double	Total administered booster doses per hundred
total_cases	Double	Total covid cases
total_cases_per_million	Double	Total covid cases per million
total_deaths	Double	Total deaths due to covid
total_deaths_per_million	Double	Total deaths due to covid per million
total_tests	Double	Total performed tests
total_tests_per_thousand	Double	Total performed tests per thousand
total_vaccinations	Double	Total vaccinations
total_vaccinations_per_hundred	Double	Total vaccinations_per_hundred

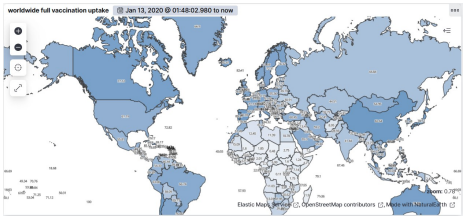
- The second dataset, named ***vaccinations_ europe***[4], contains details on vaccine administrations at european level, collected weekly and grouped by age group, vaccine type and region. For simplicity, the rows concerning sub-national data have been eliminated, keeping only national level information.
  As for the first dataset, automatic mapping has been used to generate the elasticsearch index, with the only exception of the *YearWeekISO* field which has been manually specified because it was not automatically recognized.

[3]https://ourworldindata.org/covid-vaccinations
[4]https://www.ecdc.europa.eu/en/publications-data/data-covid-19-vaccination-eu-eea

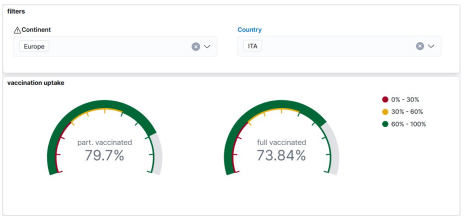
Field Name	Data Type	Description
RecordNumber	Long	The index of the record
YearWeekISO	Date (format: weekyear_week)	Date when the vaccine was received/administered
ReportingCountry	Keyword	ISO code of the reporting country
Denominator	Double	Population denominators for target groups
NumberDosesReceived	Double	Number of vaccine doses distributed in the reporting week
NumberDosesExported	Double	Number of vaccine doses donated or sold in the reporting week
FirstDose	Long	First doses administered in the reporting week
SecondDose	Long	Second doses administered in the reporting week
DoseAdditional1	Long	Booster doses administered in the reporting week
UnknownDose	Long	Administered doses in the reporting week whose type is unknown
Region	Keyword	The reporting country
TargetGroup	Keyword	Target group for vaccination
Vaccine	Keyword	Name of the vaccine
Population	Long	Population of the country

4.2 Dashboard

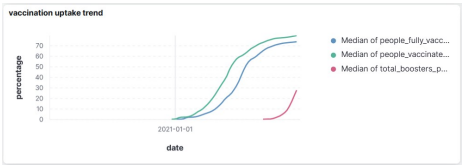
In order to better visualize the data described above, an additional dashboard has been created, split into two main sections.
The first focuses on the global dataset and shows macro-level information such as the percentage of vaccinated, the number of deaths caused by COVID-19, and the vaccination uptake over time. The different panels can be customized using drop-down menus, which allow you to filter data at the continent or individual country level.



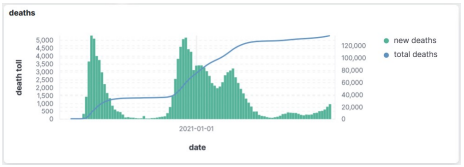
(a) *World map showing the full vaccination uptake in every recorded country.*



(b) *Drop-down menus used to customize the different panels.*



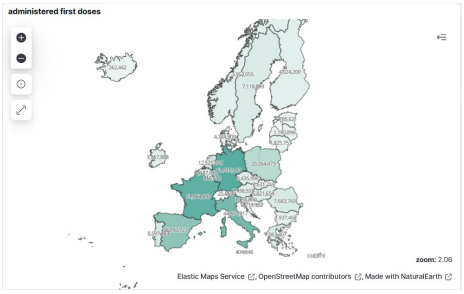
(c) *Vaccination uptake over time.*



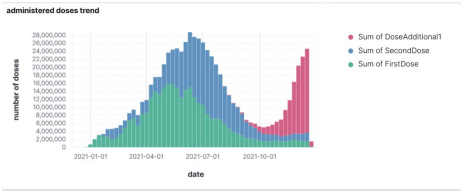
(d) *Death toll over time.*

Figure 44: The first section of the dashboard.

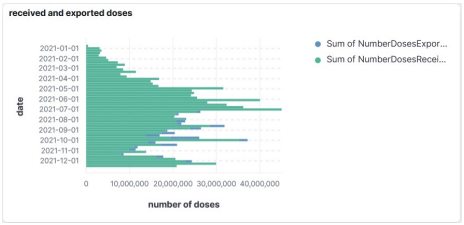
The second one utilizes the information contained in the second dataset to show some more detailed infographics regarding the status of vaccinations in the European continent. Once again, the different panels can be customized through a drop-down menu that allows to choose a single country.



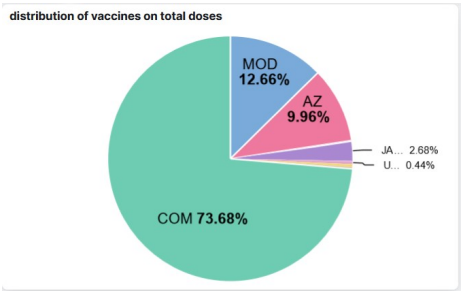
(a) *European map showing the number of first doses administered in each country.*



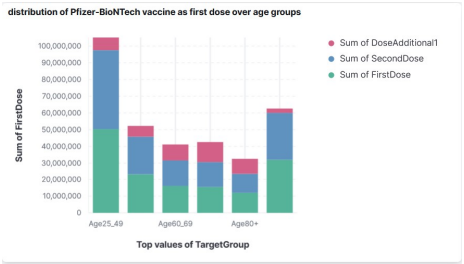
(b) *Trend of administered doses, subdivided by dose type.*



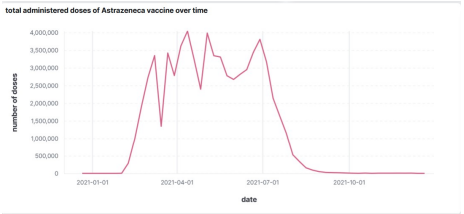
(c) *Received and exported doses over time.*



(d) *Vaccines distribution over total administered doses.*



(e) *Distribution of Pfizer-BioNTech vaccine over age groups.*



(f) *Administered doses of Astrazeneca vaccine over time.*

Figure 45: Some panels included in the second section of the dashboard.

5 Optional 2: Cassandra

The first step for interacting with Cassandra is to create a new **keyspace**. It involves the following commands:

```
CREATE KEYSPACE vaccines_management
WITH replication = {'class': 'SimpleStrategy', 'replication_factor': 3};

DESCRIBE KEYSPACES

USE project
```

The second step consists in the creation of a table containing the data related to the vaccinations divided by region, age group. We use as primary key the tuple **[administration_date, age_group, area, supplier]**

```
CREATE TABLE vaccinations (
administration_date date,
supplier varchar,
area varchar,
age_group varchar,
male_count varint,
female_count varint,
first_doses varint,
second_doses varint,
post_infection_doses varint,
booster_doses varint,
NUTS1_code varchar,
NUTS2_code varchar,
region_ISTAT_code varint,
region_name varchar,
PRIMARY KEY(administration_date, age_group, area, supplier);
```

To integrate data from an external sources (.csv file) we use the command COPY.

```
'
FROM 'YourPath/somministrazioni-vaccini-latest.csv'
WITH HEADER = TRUE AND DELIMITER = ';';
```

5.1 Queries

We will use toDate(now()) to build the current date.

1. Recover all data related to vaccination in Basilicata region.

   ```
   CREATE INDEX region ON vaccinations (region_name);

   SELECT *
   FROM vaccinations
   WHERE region_name = 'Basilicata' ;
   ```

2. Count how many times have been carried out more than 1000 vaccines of the same type to people from 20 to 29 years old.

```
SELECT COUNT(*)
FROM vaccinations
WHERE first_doses >= 1000 and age_group = '20-29'
ALLOW FILTERING;
```

3. Count how many first doses of vaccine were carried out in Lombardia region.

```
SELECT SUM(first_doses)
FROM vaccinations
WHERE area = 'LOM'
ALLOW FILTERING;
```

4. Check in the last 10 days, how many people have been vaccinated despite having already had COVID-19.

```
SELECT administration_date, area, age_group, post_infection_doses
FROM vaccinations
WHERE administration_date > toDate(now()) - 10d
AND post_infection_doses > 0
ALLOW FILTERING;
```

5. Calculate the average number of males vaccinated daily in Tuscany from 30 to 39 years old

```
SELECT AVG(male_count)
FROM vaccinations
WHERE area = 'TOS' and age_group = '30-39'
ALLOW FILTERING;
```

6. Display ten *NUTS1_code* codes

```
CREATE INDEX codeN1 ON vaccinations(NUTS1_code);

SELECT NUTS1_code
FROM vaccinations
LIMIT 10;
```

7. Show when, where, what vaccine and at what age range doses of vaccine were administered to a male-only population

```
SELECT administration_date, area, supplier, age_group,
        male_count, female_count
FROM vaccinations
WHERE female_count = 0
ALLOW FILTERING;
```

8. See if *Janssen* was given to people over 80 who had faced COVID-19

```
SELECT *
FROM vaccinations
WHERE supplier IN ('Janssen')
    AND age_group >= '80-89'
    AND post_infection_doses > 0
ALLOW FILTERING;
```

5.2 Commands

Before delving into the details of Cassandra commands, it's helpful to describe the idea behind them. Since Cassandra is a technology more suitable for write-intensive applications, there is no particular need for complex updates or deletions.

1. Creation of a new row inside the *vaccination* table.

```
INSERT INTO vaccinations
(administration_date,
supplier,area,age_group,
male_count,female_count,
first_doses,second_doses,
post_infection_doses,booster_doses,
NUTS1_code,NUTS2_code,
region_ISTAT_code,region_name)
VALUES
('2021-12-31',
'Pfizer/BioNTech','LOM','40-49',
100,250,50,50,250,0,
'ITC','ITC4',
3,'Lombardia');
```

2. The following command updates the previously added row. It best fits whenever we need to change some values after a recalculation. The attributes' ordering in the **where** clause is fundamental for retrieving the right row. The declaration[5] must comply with the *primary key* definition in the table creation.

```
UPDATE vaccinations
SET female_count = 300,
post_infection_doses = 0,
booster_doses = 300
WHERE
administration_date = '2021-12-31' and age_group = '40-49'
and area = 'LOM' and supplier = 'Pfizer/BioNTech';
```

[5]This rule applies also to the queries.

6 Resources

All the resources covered in this report are available at the following page https://github.com/AndreaSanchini/PoliVax.
Refer to it for guidelines on how to import the datasets and Kibana dashboards.