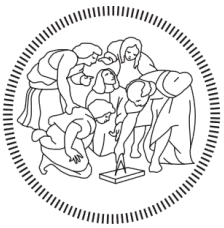


POLITECNICO
MILANO 1863

System and Methods for Big and Unstructured Data

Course project 2021 - Third part

Vaccinations Data Analysis

Authors

Marcello Bavaro - 10597421

Francesco Puddu - 10583315

Filippo Rescalli - 10609906

Elena Righini - 10796621

Simone Sarti - 10580595

Table of contents

1. Overview

The goal of this project is to build an information system for the management of data related to a pandemic (like the COVID-19 one) in a given country.

In this document we describe the third part, which consists in the analysis of COVID-19 vaccinations information, stored in an Elasticsearch database. The system shall be able to perform queries and to visualize reports and summary statistics through dashboards realized in Kibana, with the objective of giving insights on the current pandemic situation.

2. Datasets

All the updated and unmodified datasets our system relies on can be found at this link. These datasets are open. They are provided and maintained by the italian "Struttura Commissariale per l'Emergenza Covid-19".

The .csv files we effectively used can be found in the delivery folder, together with a customized dataset.

2.1 Original Dataset

somministrazioni-vaccini-latest (index name: *somministrazioni-vaccini-latest*)

Field	Data types	Description
index	integer	The index of the record
area	string	Acronyms of the region of delivery
fornitore	string	Complete name of the supplier of the vaccine
data_somministrazione	datetime	Administration date of the vaccines
fascia_anagrafica	string	Age group of the people administered with the vaccines
sesso_maschile	integer	Number of vaccinations administered to males
sesso_femminile	integer	Number of vaccinations administered to females
prima_dose	integer	Number of people administered with the first dose
seconda_dose	integer	Number of people administered with the second dose

pregressa_infezione	integer	Number of people administered with a dose after they have been infected
dose_addizionale_booster	integer	Number of people administered with an additional dose/recall
codice_NUTS1	string	https://en.wikipedia.org/wiki/NUTS_statistical_regions_of_Italy
codice_NUTS2	string	https://en.wikipedia.org/wiki/NUTS_statistical_regions_of_Italy
codice_regione_ISTAT	integer	ISTAT code of a region
nome_regione	string	Name of the region (bilingual, when necessary)

We uploaded the dataset into Elasticsearch through the interface offered by Kibana. Elasticsearch tries to automatically assign data types to the fields of datasets being ingested, in our case the only field that needed a manual correction was the datetime, for which we defined a custom date mapping (*yyyy-MM-dd)*. All the String types were mapped to Keyword, we accepted it as for our task we are only going to perform exact matches.

2.2 Additional Datasets

To perform more advanced queries we took advantage of some additional datasets. Their description is reported here:

2.2.1 platea (index name: *platea*)

Field	**Data types**	**Description**
index	integer	The index of the record
area	string	Acronyms of the region
nome_area	string	Name of the region (bilingual, when necessary)
fascia_anagrafica	string	Age group to which vaccinable people belong
totale_popolazione	integer	Total amount of the population present in platea for a certain age group

The Elasticsearch automatic mapping on this dataset was correct, no need for intervention.

2.2.2 consegne-vaccini-latest (index name: *consegne*)

Field	Data types	Description
index	integer	The index of the record
area	string	Acronyms of the region
fornitore	string	Complete name of the supplier of the vaccine
data_consegna	datetime	Delivery date of the vaccine
numero_dosi	integer	Number of doses delivered on that date per region
codice_NUTS1	string	https://en.wikipedia.org/wiki/NUTS_statistical_regions_of_Italy
codice_NUTS2	string	https://en.wikipedia.org/wiki/NUTS_statistical_regions_of_Italy
codice_regione_ISTAT	integer	ISTAT code of a region
nome_regione	string	Name of the region (bilingual, when necessary)

Just like for "somministrazioni-vaccini-latest", the automatic mapping was fine, except for the datetime that had to be transformed into a custom (yyyy-MM-dd) date.

2.2.3 result_geo (index name: *geo_test4*)

Being this dataset derived from "somministrazioni-vaccini-latest", the structure is the same for most of the columns, the only exceptions being:

- "data_somministrazione" : this was transformed into an ISO date to avoid the need to create a custom mapping on import.
- "region_code" : this field stores an identification code for each region. This was needed to use the choropleth layer for the fourth query (Q4).
- "latitude" and "longitude" : were meant to be needed for another query

Thanks to the new format of the date field, the Elasticsearch automatic mapping on this dataset was correct, no need for intervention.

3. Queries and Commands

3.1 Queries

What follows is our set of queries, enriched where possible by the relative visualization. All the graphs displayed in this section are included in the delivery folder. Some of the reported graphs represent weekly results unlike the corresponding query (always daily), but that's just for better readability.

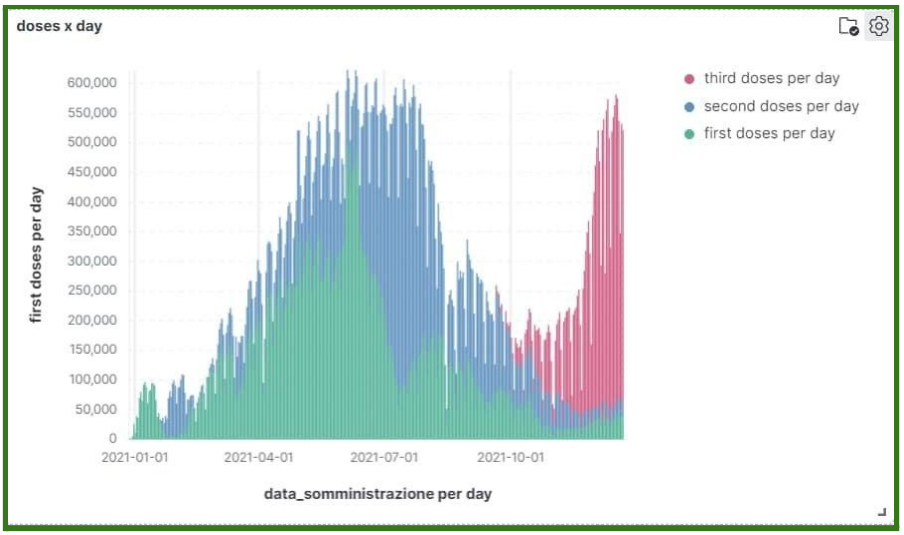
Q1 - Total number of first, second and third doses administered

```
GET /somministrazioni-vaccini-latest/_search
{
   "size": 0,
   "aggs": {
      "first_doses_tot": {
         "sum": {
            "field": "prima_dose"
         }
      },
      "second_doses_tot": {
         "sum": {
            "field": "seconda_dose"
         }
      },
      "additional doses tot": {
        "sum": {
          "field": "dose_addizionale_booster"
        }
      }
   }
}
```

Description:
The query uses an aggregation command to define three fields: first_doses_tot, second_doses_tot and additional_doses_tot, representing the sum of the values in fields prima_dose, seconda_dose, dose_addizionale_booster of the documents, respectively . The results are displayed through a bar plot. As expected, the number of first and second doses is much higher than that of third doses.

Q2.1 - **Total number of vaccinations per day** (split by number of dose)

```
GET /somministrazioni-vaccini-latest/_search
{
  "size":0,
  "aggs": {
    "single_days":{
      "terms": {
        "field": "data_somministrazione"
      },
      "aggs": {
        "first_doses": {
          "sum": {
            "field": "prima_dose"
          }
        },
        "second_doses": {
          "sum": {
            "field": "seconda_dose"
          }
        },
        "third_doses": {
          "sum": {
            "field": "dose_addizionale_booster"
          }
        }
      }
    }
  }
}
```



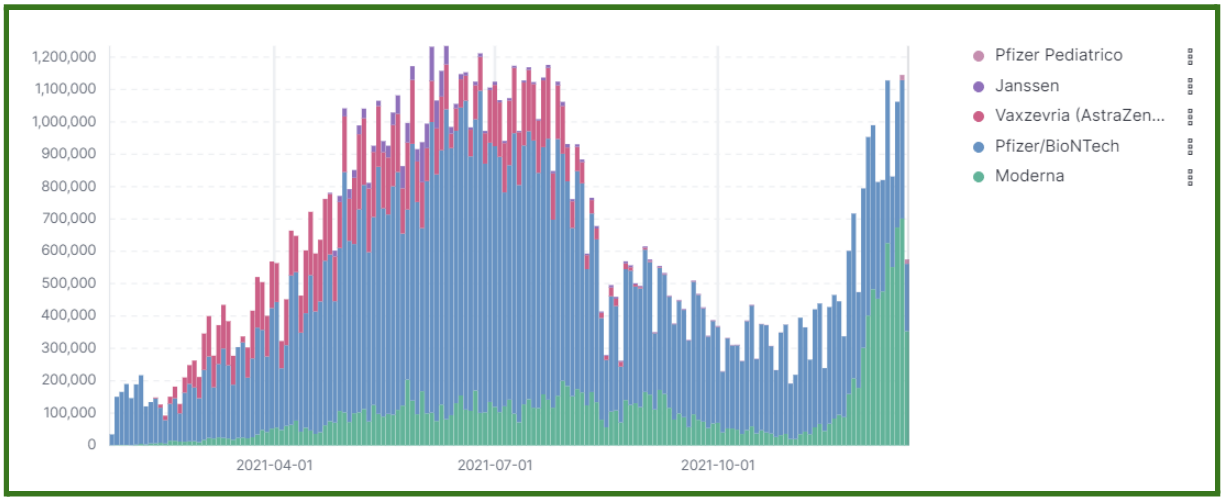
Description:

Firstly, the query aggregates all the documents by the date in which the vaccination was performed. Then, for each date, it uses another aggregation command to create three fields: first_doses, second_doses and third_doses. They contain the sum of the values of the fields prima_dose, seconda_dose, dose_addizionale_booster appearing in the vaccination documents for that day. Finally, the results are displayed in an histogram which has a bar for every day in the period analyzed. Each bar represents the

number of doses administered on that specific day, and it's given by the sum of three bars representing the number of first/second/third doses. As expected, the numbers of first and second doses increase together at the beginning, to then diminish, followed by an increase in the number of third doses.

Q2.2 - **Total number of vaccinations per day** (split by vaccine supplier)

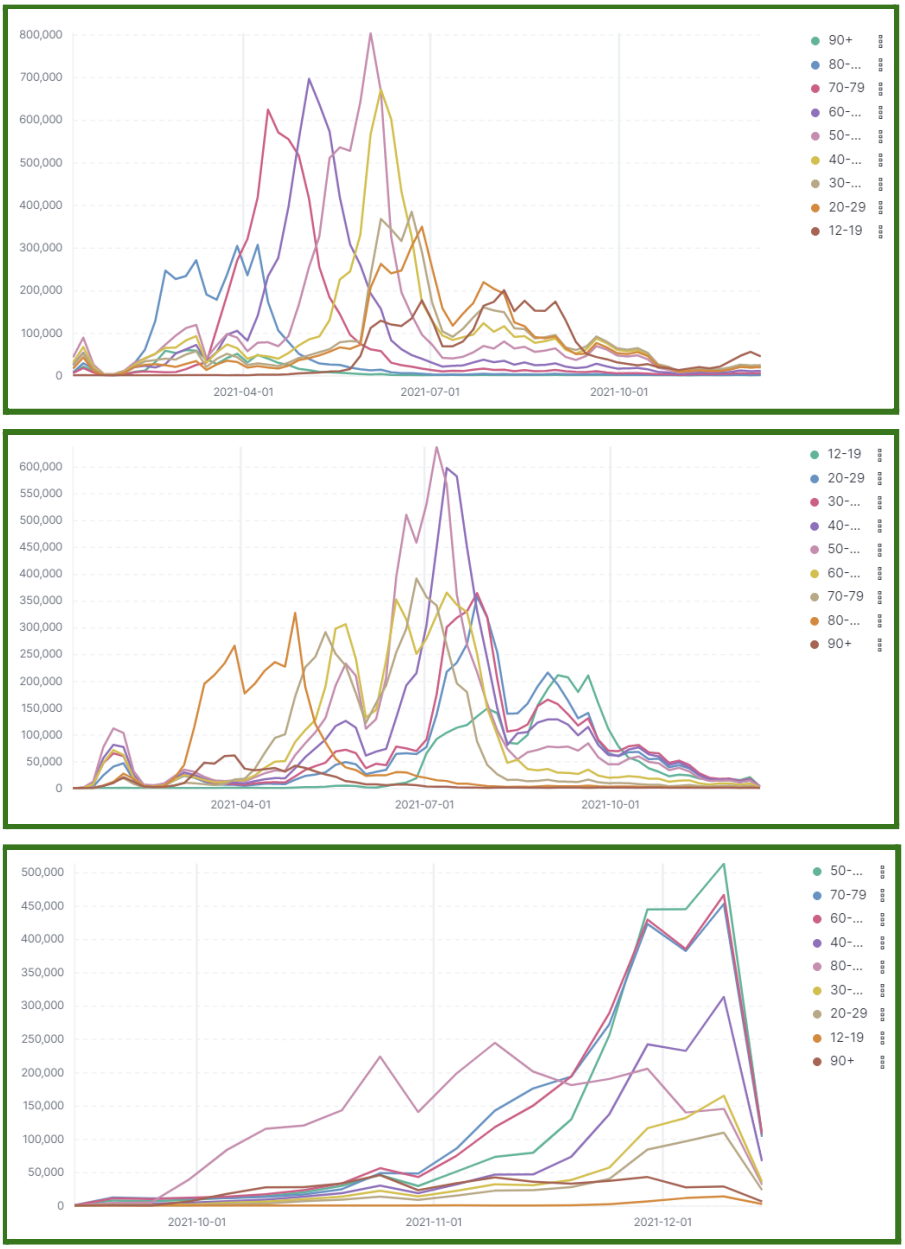
```
GET /somministrazioni-vaccini-latest/_search
{
  "size":0,
  "aggs": {
    "single_days":{

      "multi_terms": {
        "terms": [
          {
            "field": "data_somministrazione"
          },
          {
            "field": "fornitore"
          }
        ]
      },
      "aggs": {
        "total_vaccinations": {
          "sum" : {
            "script" : "doc['prima_dose'].value + doc['seconda dose'].value + doc['dose addizionale booster'].value"
          }
        }
      }
    }
  }
}
```



Description:
Firstly, the query aggregates all the documents by the date in which the vaccination was performed. Then, for each date, it uses another aggregation command to count, for each supplier, the total number of doses used. The script command is used to sum together the values of first,second and third doses appearing in the same document, as the aggregation would just sum them separately. Finally, the results are displayed in an histogram. Each bar represents the number of doses administered on that specific day, and it's given by the sum of multiple bars representing the number of doses of different vaccine types. It's Interesting to see how the administration of different vaccine types varies over time, even if Pfizer/BioNTech remains the most used vaccine over the analyzed time period.

Q3 - Daily number of vaccinations per age group and dose

```
GET /somministrazioni-vaccini-latest/_search
{
    "size":0,
    "aggs": {
        "doses_per_day_per_age group":{
            "multi_terms": {
                "terms": [
                    {"field": "data somministrazione"},
                    {"field": "fascia anagrafica"}
                ]
            },
            "aggs":{
                "first_doses":{
                    "sum":{
                        "field":"prima_dose"
                        }
                    },
                "second_doses":{
                    "sum":{
                       "field":"seconda_dose"
                       }
                    },
                "third_doses":{
                    "sum":{
                       "field":"dose_addizionale_booster"
                    }
                }
            }
        }
    }
}
```

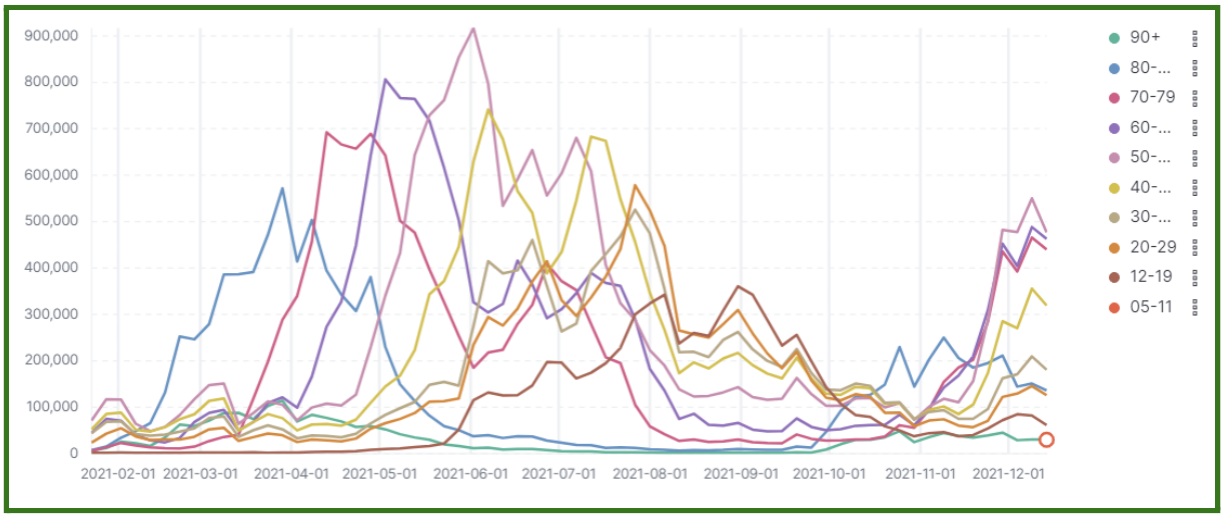


Description:

The query aggregates together documents that represent vaccinations performed on the same day and on the same age groups. Then, for each group, it uses an aggregation command to create three fields whose value is given by the sum of values of prima_dose, seconda_dose and dose_addizionale_booster for that group, respectively. The results are shown in 3 line graphs, each one displaying when the different age groups received respectively the first, second and third dose. As expected, the first graph (which refers to first dose) presents higher values during the first months, the second graph (referring to the second dose) has higher peaks in the middle of the time span (even if some high values can also be find also in the first part, probably due to vaccinations of weaker subjects or health workers) and finally the third graph (referring to the third dose) has higher values in the last part of the analyzed time window.

Q3bis - Daily total number of vaccinations per age group

```
GET /somministrazioni-vaccini-latest/_search
{
    "size":0,
    "aggs": {
        "doses_per_day_per_age group":{
            "multi_terms": {
                "terms": [
                    {"field": "data somministrazione"},
                    {"field": "fascia anagrafica"}
                ]
            },
            "aggs": {
                "number_vaccinations": {
                    "sum" : {
                        "script" : "doc['prima_dose'].value + doc['seconda dose'].value +
doc['dose addizionale booster'].value"
                    }
                }
            }
        }
    }
}
```

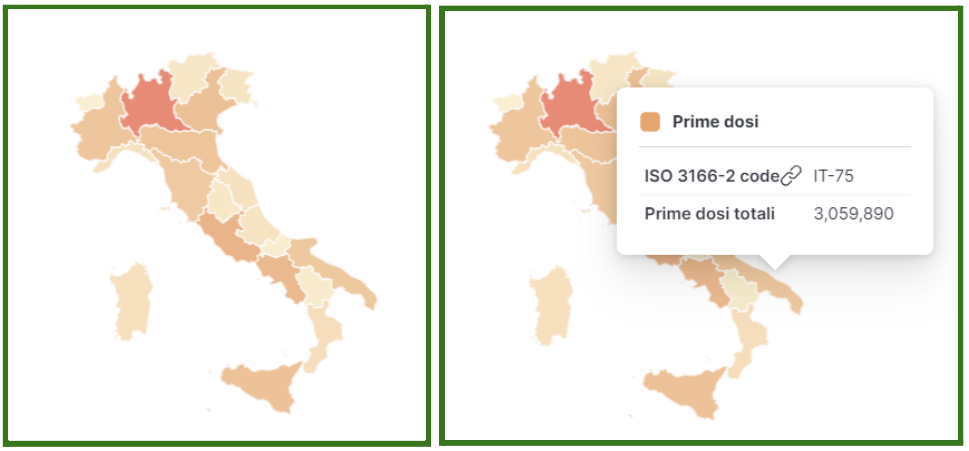


Description:

The query aggregates documents that represent vaccinations performed on the same day and on the same age groups. Then, for each group, it uses an aggregation command to create a single field whose value is given by the sum of all the values of prima_dose, seconda_dose and dose_addizionale_booster, summed together through the script command. The results are shown in a line graph displaying when the different age groups received their vaccinations. Each line, representing one age group, has a slightly different behavior w.r.t. the others. We can note how higher values in the graph are registered during the first months.

Q4 - Total number of first/second/third doses per region

We selected this query with the aim of exploring the map-based visualizations of Kibana. In particular, we wanted to replicate the heatmaps that usually appear in graphic reports about the pandemic. As we decided to leverage the choropleth layers offered by Kibana, we realized that the dataset needed to be augmented to include also the identification codes (in the ISO 3166-2 format) of regions, and not just the names. We retrieved the codes from Elasticsearch's documentation and wrote a small python script to add the new field to each record. Given the size of the dataset, this operation has been parallelized with a library (implemented by us) that is included in the delivery.



In the choropleth layers (one for the first dose, and one for the second), we assign to each region a shade of red that is directly proportional to the absolute number of vaccinations administered. The actual number can be retrieved by simply positioning the cursor over a region area.

Q4.1 - Total number of first/second/third doses per region (first dose)

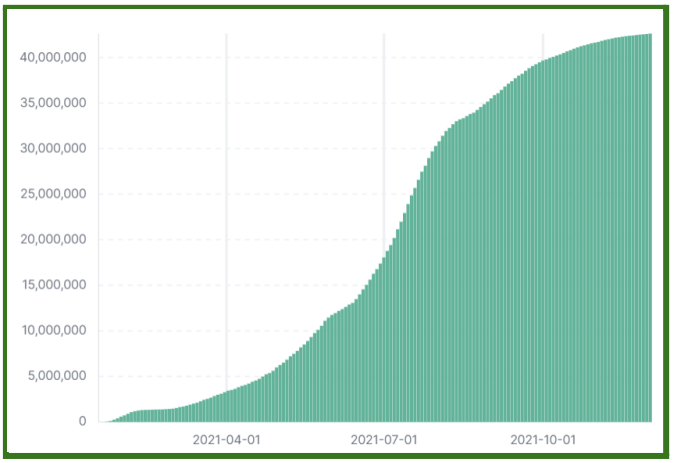
The query that follows is an attempt at extracting more or less the same information we can get from the choropleth layer, that is the number of first dose vaccinations administered in each region.

```
GET /somministrazioni-vaccini-latest/_search
{
   "size": 0,
   "aggs": {

      "regione": {
         "terms": {
            "field": "nome area"
         },

         "aggs": {
            "first_doses": {
              "sum": {
                "field": "prima_dose"
              }
            }
         }
      }
   }
}
```

Description:
Firstly, the query aggregates together documents that represent vaccinations administered in the same region. Then, using another aggregation command, the query computes the number of first dose vaccinations registered. The same query and analysis can be done for the second and third doses of the vaccine (we avoid reporting them).

Q5 - Incremental number of people who have completed the vaccination cycle

```
GET /somministrazioni-vaccini-latest/_search
{
  "size":0,
  "aggs":{
    "vaccini_al_giorno":{
      "date_histogram" : {
                "field" : "data somministrazione",
                "interval" : "day"
            },
            "aggs": {
                "vaccinati completi": {
                    "sum": {
                        "field": "seconda_dose"
                    }
                },
                "vaccinati completi cumulativo": {
                    "cumulative_sum": {
                        "buckets path": "vaccinati completi"
                    }
                }
      }
    }
  }
}
```

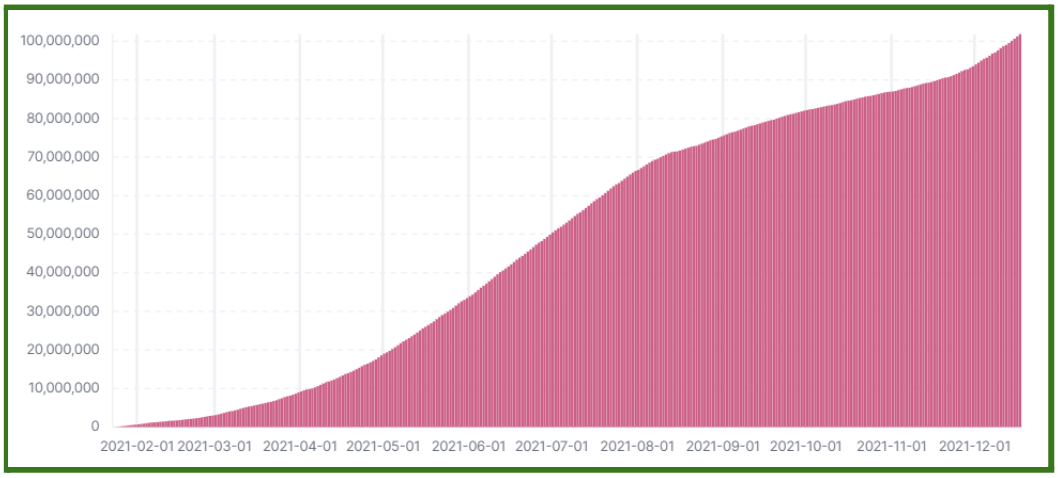


Description:
The query aggregates together documents that represent vaccinations performed on the same day. Then, for each day, using another aggregation command, the query computes the number of second doses. At last, a cumulative sum is applied. The result is displayed in a histogram, and shows a more rapid growth in the middle months of the timeframe.

Note: By “people who have completed the vaccination cycle” we mean those people who have received both the first and the second doses of vaccine and not those who have recovered from Covid and were subject to only one vaccination.

Q5bis - Incremental number of administered vaccinations

```
GET /somministrazioni-vaccini-latest/_search
{
    "size":0,
    "aggs":{
        "vaccini_al_giorno":{
            "date_histogram" : {
                "field" : "data somministrazione",
                "interval" : "day"
            },
            "aggs": {
                "somma prime dosi": {
                    "sum": {
                        "script" : "doc['prima_dose'].value + doc['seconda dose'].value +
doc['dose addizionale booster'].value"
                    }
                },
                "prime dosi cumulative": {
                    "cumulative_sum": {
                        "buckets path": "somma prime dosi"
                    }
                }
            }
        }
    }
}
```



Description:
Firstly, the query aggregates together documents that represent vaccinations performed on the same day. Then, for each day, using another aggregation command, the query computes the number of vaccinations performed on that day (by summing first,second and third doses). At last, a cumulative sum is applied. After a regular start, we can first notice a slight slowdown in correspondence with the decline of Covid cases during summer season, then a sudden surge in the final part of the year due to a new increase in the number of

cases, the possibility of getting an additional dose of vaccine for adults and the green light of the AIFA to children's vaccinations.

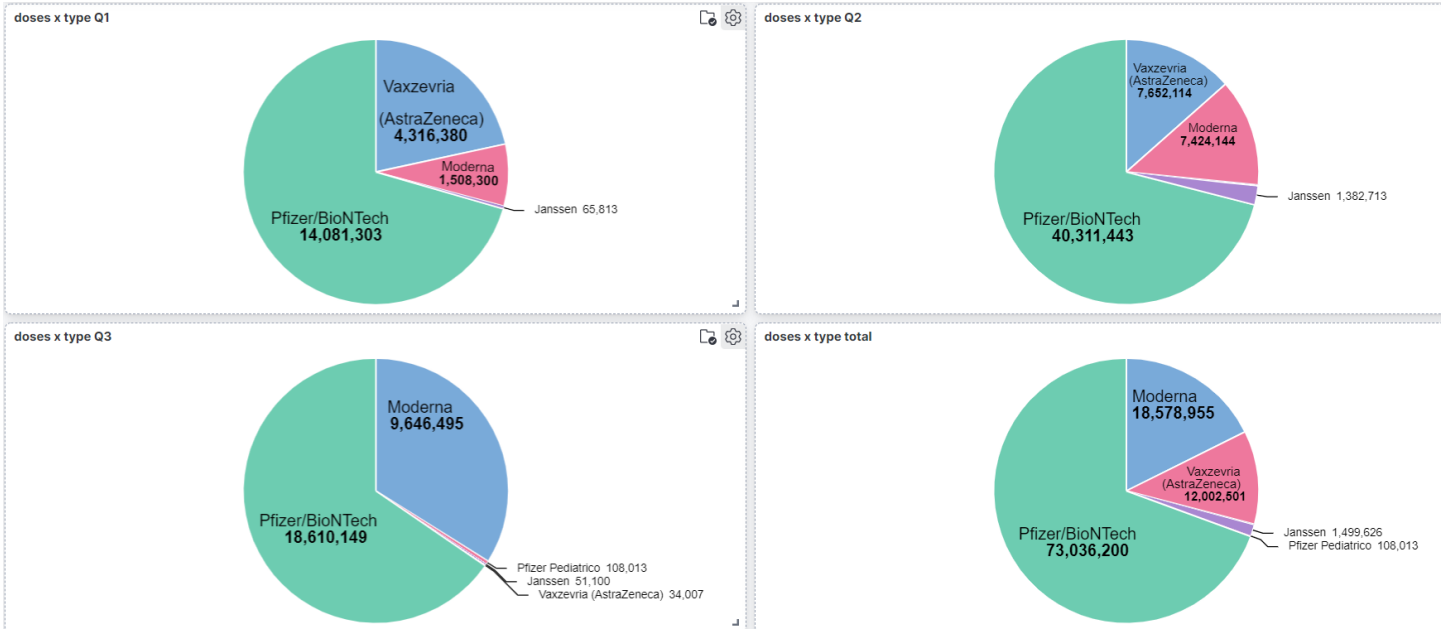
Q6: Amount of doses administered per supplier each quarter and in total

```
GET /somministrazioni-vaccini-latest/_search
{
  "size":0,
  "query": {
    "range": {
      "data somministrazione": {
        "from": "2021-01-01",
        "to": "2021-04-30"
      }
    }
  },
  "aggs": {
    "supplier": {
      "terms": {
        "field": "fornitore"
      },
      "aggs": {
        "doses": {
          "sum": {
          "script" : "doc['prima_dose'].value + doc['seconda dose'].value + doc['dose addizionale booster'].value"

          }
        }
      }
    }
  }
}
```

```
GET /somministrazioni-vaccini-latest/_search
{
  "size":0,
  "query": {
    "range": {
      "data somministrazione": {
        "from": "2021-05-01",
        "to": "2021-08-31"
      }
    }
  },
  "aggs": {
    "supplier": {
      "terms": {
        "field": "fornitore"
      },
      "aggs": {
        "doses": {
          "sum": {
          "script" : "doc['prima_dose'].value + doc['seconda dose'].value + doc['dose addizionale booster'].value"

          }
        }
      }
    }
  }
}
```

```
GET /somministrazioni-vaccini-latest/_search
{
  "size":0,
  "query": {
    "range": {
      "data somministrazione": {
        "from": "2021-09-01",
        "to": "2021-12-31"
      }
    }
  },
  "aggs": {
    "supplier": {
      "terms": {
        "field": "fornitore"
      },
      "aggs": {
        "doses": {
          "sum": {
          "script" : "doc['prima_dose'].value + doc['seconda dose'].value + doc['dose addizionale booster'].value"

          }
        }
      }
    }
  }
}
```

```
GET /somministrazioni-vaccini-latest/_search
{
  "size":0,
  "aggs": {
    "supplier": {
      "terms": {
        "field": "fornitore"
      },
      "aggs": {
        "doses": {
          "sum": {
          "script" : "doc['prima_dose'].value + doc['seconda dose'].value + doc['dose addizionale booster'].value"

          }
        }
      }
    }
  }
}
```

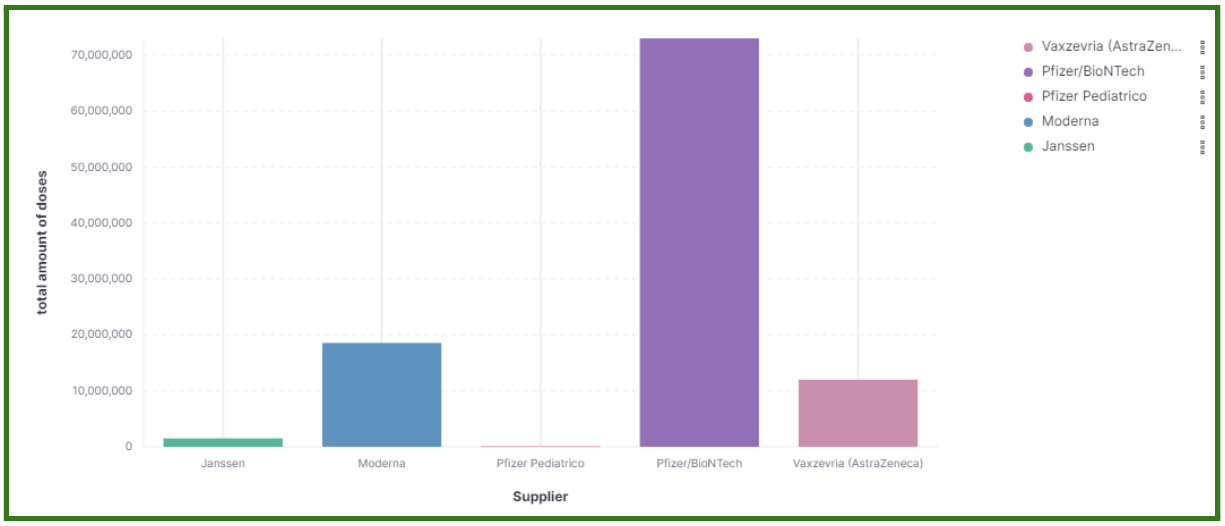


Description

This query is split in 4 parts, the first 3 consider only documents belonging to the first/second/third quarter of the 2021 year respectively, they then use an aggregation command to count, for each vaccine supplier, the total amount of doses administered in the considered time frame. The fourth part instead takes into consideration the whole year. The results are displayed in the pie charts above, it's evident that the Pfizer/BioNTech vaccine was by far the most used, followed by the Moderna one. It can also be observed that, starting from the third quarter, the use of the J&J and AstraZeneca vaccines was abandoned, more Moderna doses were used to compensate.

Q7 - Total number of vaccinations per vaccine supplier

```
GET /somministrazioni-vaccini-latest/_search
{
    "size":0,
    "aggs":{
        "type_of_vaccine":{
            "terms":{
                "field":"fornitore"
            },
            "aggs":{
                "number_of_doses":{
                    "sum":{
                        "script" : "doc['prima_dose'].value + doc['seconda dose'].value + doc['dose addizionale booster'].value"
                    }
                }
            }
        }
    }
}
```



Description

The query aggregates together documents representing vaccinations performed using vaccines from the same supplier. For each group, it defines a field containing the sum of the values for prima_dose, seconda_dose and dose_addizionale_booster. Finally, the result is displayed in a bar plot, with a vertical bar for each vaccine supplier. The PfizerBioNTech vaccine is by far the most used.

3.2 Commands

C1. Insertion of a new document

```
POST /somministrazioni-vaccini-latest/_doc
{
  "area": "CAM",
  "codice regione ISTAT": 15,
  "nome area": "Campania",
  "data somministrazione": "2021-12-20",
  "dose addizionale booster": 0,
  "codice NUTS": "ITF",
  "fascia anagrafica": "50-59",
  "prima dose": 1,
  "pregressa infezione": 0,
  "fornitore": "Janssen",
  "@timestamp": "2021-12-22T00:00:00.000+01:00",
  "seconda dose": 0,
  "sesso_maschile": 1,
  "codice NUTS": "ITF3",
  "sesso_femminile": 0
}
```

Description:
This command inserts a new document into the *somministrazioni-vaccini-latest* index

C2.1: Deletion of a document with a specific id

```
DELETE /somministrazioni-vaccini-latest/_doc/<_id>
```

Description:
This command deletes a document based on its _id.

C2.2: Deletion of a document using a query

```
POST /somministrazioni-vaccini-latest/_delete_by_query
{
 "query": {
  "match": {
      "<field>": "<value>"
        }
    }
}
```

Description:
This command deletes from the index all the documents satisfying the specified query. It enables the user to perform more refined delete operations, as he/she can define the field-value pair present in the match clause.

C3: Update of a document

```
POST /somministrazioni-vaccinali-latest/update/<_id>
{
  "script" : {
    "source": "ctx._source.seconda dose = params.count",
    "lang": "painless",
    "params" : {
      "count" : 4
    }
 }
}
```

Description:
This command updates the value of the attribute "seconda_dose" of a document identified by the specified _id. The attribute is set to the value specified as parameter in "params".

4. Optional part: Integrations with other datasets

Note on the graphical representation: Unfortunately it seems that Kibana doesn't allow to create representations showing results obtained from mathematical operations performed over data coming from different indexes, therefore the results of some of these queries (Qx1 & Qx2) won't be represented graphically

Qx1. Percentage of first, second and third dose vaccinations per age group

```
GET /platea,somministrazioni-vaccini-latest/_search
{
    "size":0,
    "query":{
        "bool":{
            "must_not":[
                {"term":{"fascia anagrafica":"80-89"}},
                {"term":{"fascia anagrafica":"90+"}},
                {"term":{"fascia anagrafica":"80+"}}
            ]
        }


    },
    "aggs":{
        "doses_for_agegroup":{
            "terms":{
                "field":"fascia anagrafica"
            },
            "aggs":{
                "first_doses":{
                    "sum":{
                        "field":"prima dose"
                        }
```

```
                },
            "second_doses":{
                "sum":{
                   "field":"seconda_dose"
                   }
                },
            "third_doses":{
                "sum":{
                   "field":"dose addizionale booster"
                }
            },
            "persons":{
                "sum":{
                    "field":"totale popolazione"
                }
            },

            "%_first_dose": {
                "bucket_script": {
                    "buckets_path": {
                             "v1": "first_doses",
                             "v2": "persons"
                            },
            "script": "(Math.round(params.v1/params.v2*10000))/100.0"
                }
            },
            "%second dose": {
                "bucket_script": {
                    "buckets_path": {
                             "v1": "second_doses",
                             "v2": "persons"
                            },
            "script": "(Math.round(params.v1/params.v2*10000))/100.0"
                }
            },
            "%_third_dose": {
                "bucket_script": {
                    "buckets_path": {
                             "v1": "third_doses",
                             "v2": "persons"
                            },
            "script": "(Math.round(params.v1/params.v2*10000))/100.0"
                }
            }
        }

    }
  }
}
```

Description

This query uses two indexes: `somministrazioni-vaccini-latest` and `platea`.
Firstly, the query excludes all documents having values for "fascia_anagrafica" equal to 80+, 90+ or 80-89 (that's because the italian ministry uses the ranges 80-89 & 90+ in the datasets of vaccinations, while it uses just 80+ for the dataset describing the population in each region, making the two incompatible). Then, an aggregation over the field "fascia_anagrafica" is performed, which leads to documents in both indexes being aggregated on that field at the same time. For each "fascia_anagrafica", an aggregation command is used to compute the total number of first,second and third doses, as well as to the total population belonging to that age group. Successively, bucket script commands are used to compute the fractions, and script commands are used to compute the percentages with two decimal places.

Qx2. Percentage of first/second/third dose vaccinations per region

```
GET /platea,somministrazioni-vaccini-latest/_search
{
    "size":0,
    "aggs":{
        "vaccinations_per_region":{
            "terms":{
                "field":"nome area"
            },
            "aggs":{
                "first_doses":{
                    "sum":{
                        "field":"prima dose"
                        }
                    },
                "second_doses":{
                    "sum":{
                       "field":"seconda_dose"
                       }
                    },
                "third_doses":{
                    "sum":{
                       "field":"dose addizionale booster"
                    }
                },
                "persons":{
                    "sum":{
                        "field":"totale popolazione"
                    }
                },

                "%_first_dose": {
                    "bucket_script": {
                        "buckets_path": {
                                 "v1": "first_doses",
                                 "v2": "persons"
                                },
```

```
                "script": "(Math.round(params.v1/params.v2*10000))/100.0"
                    }
                },
                "%second dose": {
                    "bucket_script": {
                        "buckets_path": {
                                "v1": "second_doses",
                                "v2": "persons"
                               },
                "script": "(Math.round(params.v1/params.v2*10000))/100.0"
                    }
                },
                "%_third_dose": {
                    "bucket_script": {
                        "buckets_path": {
                                "v1": "third_doses",
                                "v2": "persons"
                               },
                "script": "(Math.round(params.v1/params.v2*10000))/100.0"
                    }
                }
            }
        }
    }
}
```

Description

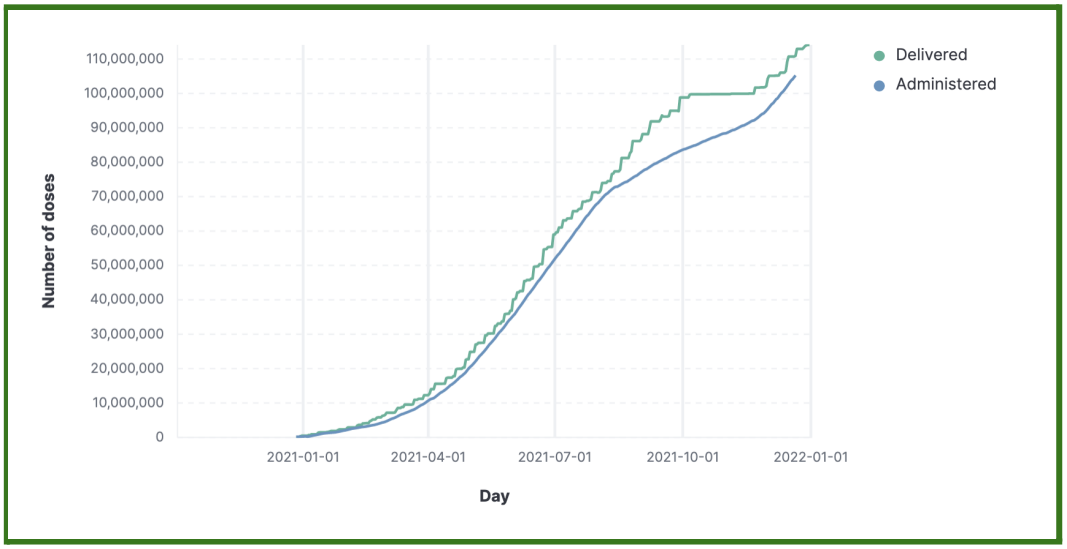
This query uses two indexes: `somministrazioni-vaccini-latest` and `platea`.

The query aggregates all documents with the same value for the "region" field. The aggregation is performed simultaneously on documents belonging to box indexes, because the "region" field is shared. For each region, the total number of first, second and third doses, plus the total population of the region are computed. Bucket script commands are then used to compute the fractions, Script commands compute the percentages using two decimal places.

Qx3: cumulative sum of delivered and administered doses

```
GET /consegne,somministrazioni-vaccini-latest/_search
{
  "size": 0,
  "aggs": {
    "vaccinations_per_day": {
      "date_histogram": {
        "field": "data somministrazione",
        "interval": "day"
      },
      "aggs": {
        "administered":{
          "sum": {
            "script": "doc['prima_dose'].value + doc['seconda dose'].value + doc['dose
addizionale booster'].value"
```

```
          }
        },
        "prime dosi cumulative": {
          "cumulative_sum": {
            "buckets_path": "administered"
            }
          }
      }
    },
    "deliveries_per_day":{
      "date_histogram": {
        "field": "data consegna",
        "interval": "day"
      },
      "aggs": {
        "delivered": {
          "sum": {
            "field": "numero dosi"
          }
        },
        "consegne cumulative":{
          "cumulative_sum": {
            "buckets_path": "delivered"
          }
        }
      }
    }
  }
}
```



Description

The query uses two indexes: `somministrazioni-vaccini-latest` and `consegne`.

The query aggregates together documents representing vaccinations administered on the same day, and for each day computes the total number of doses administered . A cumulative sum is then applied to compute the running total for each day. The same process is applied to the delivery data. The final result is the

cumulated number of doses delivered and administered up to each day. From the graph, we can clearly see that almost all the doses delivered to vaccination centers are administered within a short time.
More in depth, we can notice the graph outlining an initial correspondence between the supply of doses and their relative administration, then it moves on to a period in which the range between the two curves increases to reach the maximum distance in October, and finally those curves tend to get closer at the end of the year. We could assume that the gap between the two curves during autumn represents the phase in which the Italian government had stocked up to perform a strong vaccination campaign during the winter.

5. Importing the Datasets and visualizing the Dashboards

All the material needed to visualize the dashboard is provided into the delivery folder.

5.1 Importing the Datasets

For each dataset, do the following:

1. open kibana (Home)
2. click the “upload a file” button
3. select/drag&drop the .csv file containing the dataset
4. if needed, under the “summary” section, click “override settings” and adjust the automatic type mapping as described in section 3 of this document, then save
5. click the “import” button
6. in the “index name” text area, insert the name of the index as indicated below, and press “import”

Dataset Name	**Index Name**
somministrazioni-vaccini-latest	somministrazioni-vaccini-latest
platea	platea
consegne-vaccini-latest	consegne
result_geo	geo_test4

5.2 Importing the Dashboard

1. return to the Kibana Home
2. open the drop-down menu on the upper left
3. under the “management” section, select “stack management”
4. in the left panel, under the “Kibana” section, select “Saved Objects”
5. on the top right part of the screen press the “import” button
6. select the provided .ndjson file provided, and import it, press “done” when the process is complete
7. in the “Saved Objects” panel, click on the “Final dashboard” element that should have appeared
8. adjust the time frame on the upper right, and apply the changes

5.3 Solving Duplicate Conflicts

While testing the import and export of the dashboard we came across a Duplication error a couple of times, we solved by following these steps:

1 Removing mappings

1. reach the “Saved Objects” section as described in 5.2 (points 1 to 4)
2. select the duplicates/wrong objects that might be the cause of conflicts, and delete them

2 Removing Indexes

1. go to the Kibana Home
2. open the drop-down menu on the upper left
3. under the “management” section, select “Dev Tools”
4. in the console write `DELETE /<index_name>` and execute, this deletes the index that might be causing the conflict

3 Repeating the imports

1. retry 5.1
2. retry 5.2