Boreholes Data Visual Analysis

Autore:

Chiara Firmani

Professore:

Salvatore Rinzivillo

Ottobre 2020

Abstract

This project aims to create a visual dashboard to explore and browse geographic and geological boreholes' data. This is a joint project with the IGG institute of the CNR.

Specifically, we propose an application containing several visual widgets, each of which provides insights into a selection of original data dimensions. The different visual widgets interact with each other, creating a series of connected displays to browse data on multiple dimensions.

Interactivity was implemented by providing toolbars, selections and data filters. Moreover, we used *dc.js* library, which is a javascript charting library with native crossfilter support, allowing highly efficient exploration on large multi-dimensional datasets. It leverages d3 to render charts in CSS-friendly SVG format. Charts rendered in this project are data driven and reactive and therefore provide instant feedback to user interaction. The need to create such a dashboard arises to ensure the interaction between the three data tables, object of our analysis:

- The *Pozzi*4326 dataset
- The *Lithostratigraphic* dataset
- The *Temperature* dataset

Note that the last two datasets are in one-to-many relationship, thanks to the "key" attribute that identifies each borehole.

In order to give consistency to the analysis conducted, various tools and software were used, based on the phases of the analysis process. In particular:

- Python, Jupyter Notebook, as an open source web application for the preliminary phase of cleaning and preprocessing of the data (originally provided in csv format, then converted in *json* files).
- Vue.js, as an open-source JavaScript framework in model–view–view-model configuration, for creating user interfaces and single-page applications.
- Dc.js as the main javascript charting library with native crossfilter support, in order to provide interactive charts.
- Both Webstorm and Visual Studio Code. The first as integrated development environment for javascript, and the latter, as a modern easy handling javascript editor.

In this way we have provided a completely interactive structure, spread over two pages, Map and Graphs and Statistics, which allows the user to easily investigate the most relevant properties of the boreholes' datasets.

1 Cleaning and Preprocessing

As a first step of the analysis, a phase of exploration and cleaning of the data was carried out, specifically aimed at their subsequent use.

The original dataset, $3811 * 24$ in size, contains a total of 9170 missing values. In particular, some columns contain a disproportionately high number of NaN, so at first we thought of dropping the columns with more than 500 missing values, ie: 'entitam', 'camploc', 'locgeo' and 'tr'. However, for future purposes, we have decided to keep the first three columns and dropped only the last one. Obviously we use these variables only for background information, since they were still relevant, when present, to boreholes' researched scientists and did not burden the graphics or analysis in any way. Finally we have kept only the rows with at least 62% of correct, non missing, values and we dropped the others. This makes the resulting dataset 3805 x 23 in size.

In the section below, we can see many histograms, useful for carrying out a preliminary and holistic visual investigation of the objects under analysis.

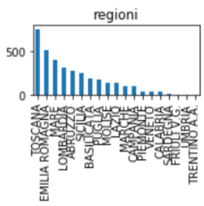
Note that the region with the highest density of boreholes is Tuscany, followed immediately by Emilia-Romagna. From the second histogram it is clear that the main purpose for which boreholes are planted in Italy is the search for hydrocarbons or for the production of geothermal energy. Similarly, the two most common types of boreholes are the exploration ones and the production ones. The difference between the two typologies is substantial: the exploration borehole is drilled in an unexplored area, in order to ascertain, for example, the presence of hydrocarbon accumulations. If the drilling leads to the discovery of economically extractable hydrocarbons, then that borehole becomes a production one.

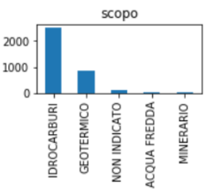
The 'state' attribute of boreholes is not specified for most of them, as seen by running a groupby.count on the distinct entries of this feature. Furthermore, many of them are temporarily closed, closed or abandoned. There are 331 boreholes drilled currently in production in Italy and 2 in reinjection.

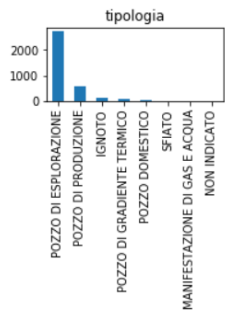
The other two datasets have many more rows, however each row has at least 2 missing values. We could have eliminated a few columns, but in light of the sheer number of rows, it wouldn't have made any difference, so we preserved the original datasets, converting them to jason format.

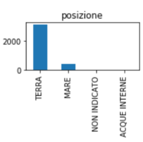
1.1 Histograms

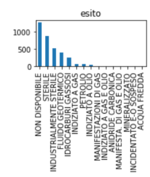
Analysis of preprocessing and visualization through histograms of the features that we then really used to build the visualization structures in VUE.js.







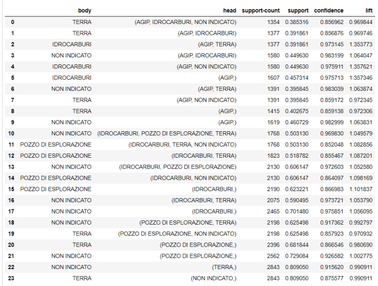




stato	
ABBANDONATO	83
CEMENTATO	162
CHIUSO MINERARIAMENTE	82
CHIUSO TEMPORANEAMENTE	407
COMPLETAMENTO DOPPIO	1
ESAURITO	8
IN PRODUZIONE	331
IN PROVA	2
IN REINIEZIONE	2
NON INDICATO	2291
OSTRUITO	1
TAPPATO E ABBANDONATO	144

1.2 Association rules mining

As a last preprocessing operation, we performed a search for associations between frequent itemsets, using the Apriori algorithm, and showing, for example, that an exploration borehole is, with 87% confidence, for hydrocarbon purposes.



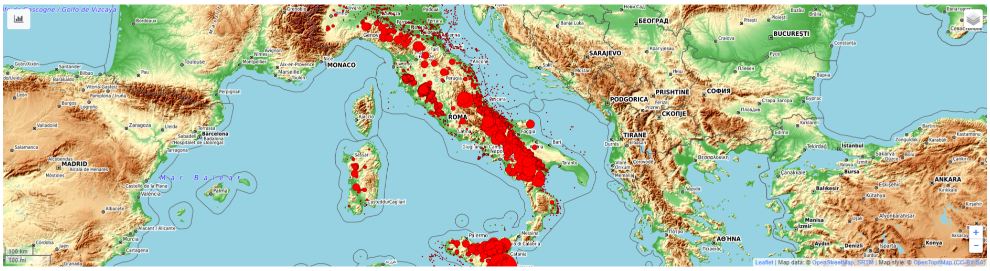
	body	head	support-count	support	confidence	lift
0	TERRA	(AGIP, IDROCARBURI, NON INDICATO)	1354	0.385316	0.856962	0.969844
1	TERRA	(AGIP, IDROCARBURI)	1377	0.391861	0.856876	0.969746
2	IDROCARBURI	(AGIP, TERRA)	1377	0.391861	0.973145	1.353773
3	NON INDICATO	(AGIP, IDROCARBURI)	1580	0.449630	0.983199	1.064047
4	IDROCARBURI	(AGIP, NON INDICATO)	1580	0.449630	0.975911	1.357621
5	IDROCARBURI	(AGIP,)	1607	0.457314	0.975713	1.357346
6	NON INDICATO	(AGIP, TERRA)	1391	0.395845	0.983039	1.063874
7	TERRA	(AGIP, NON INDICATO)	1391	0.395845	0.859172	0.972345
8	TERRA	(AGIP,)	1415	0.402675	0.859138	0.972306
9	NON INDICATO	(AGIP,)	1619	0.460729	0.982999	1.063831
10	NON INDICATO	(IDROCARBURI, POZZO DI ESPLORAZIONE, TERRA)	1768	0.503130	0.969830	1.049579
11	POZZO DI ESPLORAZIONE	(IDROCARBURI, TERRA, NON INDICATO)	1768	0.503130	0.852048	1.082856
12	POZZO DI ESPLORAZIONE	(IDROCARBURI, TERRA)	1823	0.518782	0.855467	1.087201
13	NON INDICATO	(IDROCARBURI, POZZO DI ESPLORAZIONE)	2130	0.606147	0.972603	1.052580
14	POZZO DI ESPLORAZIONE	(IDROCARBURI, NON INDICATO)	2130	0.606147	0.864097	1.098169
15	POZZO DI ESPLORAZIONE	(IDROCARBURI,)	2190	0.623221	0.866983	1.101837
16	NON INDICATO	(IDROCARBURI, TERRA)	2075	0.590495	0.973721	1.053790
17	NON INDICATO	(IDROCARBURI,)	2465	0.701480	0.975851	1.056095
18	NON INDICATO	(POZZO DI ESPLORAZIONE, TERRA)	2198	0.625498	0.917362	0.992797
19	TERRA	(POZZO DI ESPLORAZIONE, NON INDICATO)	2198	0.625498	0.857923	0.970932
20	TERRA	(POZZO DI ESPLORAZIONE,)	2396	0.681844	0.866546	0.980690
21	NON INDICATO	(POZZO DI ESPLORAZIONE,)	2562	0.729084	0.926582	1.002775
22	NON INDICATO	(TERRA,)	2843	0.809050	0.915620	0.990911
23	TERRA	(NON INDICATO,)	2843	0.809050	0.875577	0.990911

Figure 1 – Association rules with confidence > 85%.

2 Visualization tools: Home page

The home page consists of a blank page with a brief presentation of the project and two buttons: 'Maps' and 'Graphs and Statistics'. By clicking on 'Maps' you are directed to a page that allows the interactive location of boreholes on a geographical map and a selection filter that allows you to search for them by name, obtaining a focus on their main information.

2.1 Maps



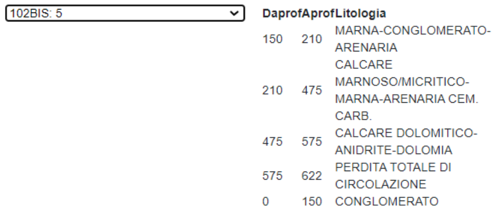
We realized the map above, using *Street Open Maps*. This interactive map is a tool that allows the user to easily interact with the data, in particular with the visualization of their geographical location. The red circles represent the borehols and, by clicking on each dot, it is possible to display a window on the right side of the map that shows the main information of that object (extrapolated from dataset *pozzi*4326).The top right button allows the user to change the type of geographical map, choosing from the following: Toner map, POIs map, Topography map (in the picture) and Terrain map. Zoom in and zoom out are provided by the down right bottons.

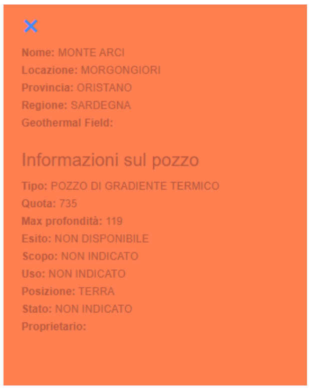
The top left button allows the user to open a selection window (with a filter by regions), which shows the relative Bar Chart, whose legend shows the colors referring to wells of sea, land, inland waters or unspecified. It also provides the regional statistics of maximum altitude, maximum depth and total number of boreholes in that specific region. If no region is selected as a filter, the Bar Chart shows the global Italian situation (all regions). By clicking on the top right drop-down menu, the user can download the resulting barchart in three different formats (SVG, PNG, CSV). The window can be closed only by clicking on the top right cross.

2.2 Interactive drop-down menu with the map

The lower part of the *Map's* page is occupied by a drop down menu, which contains the names of the data points, in the form of a dimensional variable obtained through crossfilter with the second dataset (that of lithologies)

In this way the user can select the name of the borehole he is interested in and, by clicking on it, the corresponding red dot will be highlighted on the map and an information box will appear on the right side of the map. At the same time, a three-column data table will be shown, containing the names of the lithologies found in that borehole, with their respective depth ranges. This is one of the places where a crossed interaction between the first two datasets takes place.





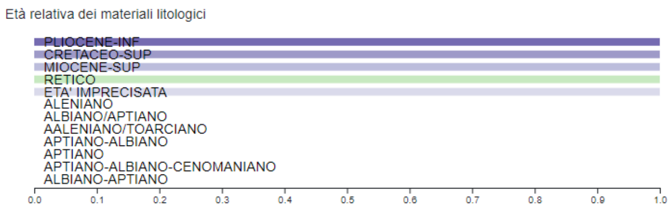
3 Graphs and Statistics

In this second section the user can find an interactive set of charts, specially designed to allow an interested user to travel through the properties of these data. The selection of the attributes explored was made on the basis of what could be the specific interests of a geologist or a borehole engineer.

3.1 Horizontal Bar Chart

Here there is a visualization tool which shows the relative ages of the rocks and lithological materials found in a well. The well in question can be easily selected from the drop-down menu next to the pie chart below. Hovering the cursor over each bar, it displays the total number material to which the age is referred to.

In this case the data on which the visualization is built come from the third dataset, that is the one relating to the lithology of the boreholes.



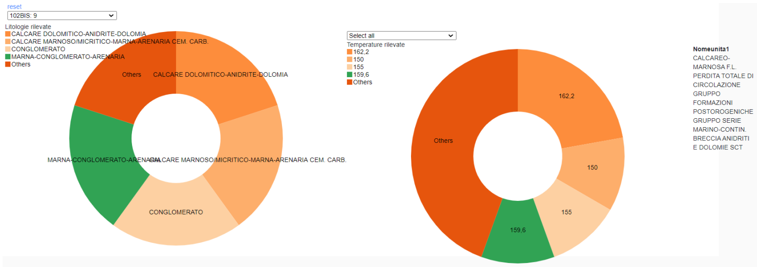
3.2 Pie Charts and data Table

The choice to use pie charts is inherent in the very nature of these structures, they are circles that intuitively represent the proportions between the objects of a set. Pie charts provide the user with an effective and qualitative way to visualize some of the most essential attributes of boreholes. In particular, by selecting the name of a well from the drop-down menu, in the pie chart on the left we will see the lithological composition detected in that well at different depth levels. In the pie chart on the right we will see the temperatures detected in the same well, at different depths. And finally in the data table we will display the names of the main units of the selected borehole.

3.3 Descriptive Pie Chart and Bar Plot

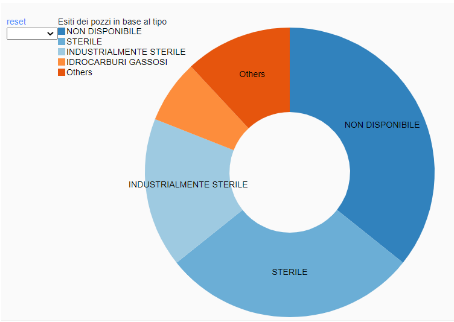
Finally we find two mutually interactive charts, which deepen the exploration of the first data set:

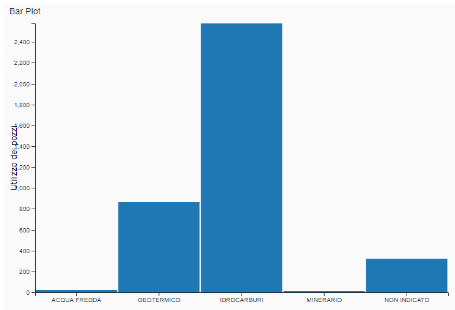
- A descriptive pie chart, which does not interact with others, but provides valuable information on the data. In fact, from the drop-down menu it is possible to select a type of borehole, and the 'esito' (scope) for which that type of well is intended will be shown (in proportion) on the pie chart.



- A bar plot showing the actual purpose for which each borehole is intended. By clicking on a particular bar of the graph, the pie chart on the right will also change, showing the results and types of the selected boreholes category.

In practice, the interaction between the two tools acts as a filter, which can be applied in three different ways: from the drop-down menu by selecting a type, or by selecting an outcome directly from the pie chart, or finally by clicking on a bar and selecting a purpose.





4 Technical specifications

1. We decided to use Vuejs since it is a JavaScript open-source configuration Model-View-viewmodel framework, for creating user interfaces and single page applications. It is also use friendly and it combines functional and OO programming together. Vuejs is a progressive framework, build from the ground and incrementally adoptable. It focuses on the view layer and it makes the development easy to pick up and integrate with other libraries.

2. OpenStreetMap is an open source map to create a free editable map of the world. Currently, we need to pay the license of Google maps. Using OpenStreetMap can be an efficient alternative for free maps visualization development.

3. We used dc.js library, because it is a javascript charting library with native crossfilter support, allowing highly efficient exploration on large multi-dimensional datasets. Charts rendered using dc.js are data driven and reactive and therefore provide instant feedback to user interaction. The dc.js is an easy yet powerful javascript library for data visualization and analysis in the browser and on mobile devices.

5 User case

In this section we propose a case, by way of example, of how a user could use our framework.

We recommend that the user first choose the 'Map' page. Suppose a user browses our page and is interested in viewing the geo-location of the boreholes, he can do it very easily simply by looking at the map. Suppose then that the user is intrigued by a particular borehole found on the map, by clicking on the dot that represents it, the user will see a small box with name, province, altitude and type of that borehole. If, on the other hand, the user wants to view a specific borehole whose name he knows, he can just select it in the drop-down menu under the map. At this point the red dot indicating its geolocation will be highlighted and an information box will appear on the right side.

Then, suppose that user is focused on Abruzzo and he wants to know what the general situation is like in that region. By clicking on the button at the top left of the map, a filter will open, selecting 'Abruzzo', he will see the general statistics and the bar chart of the wells in that region.

Now suppose our user wants to know what types of materials were actually detected in that borehole. As a first action, he can select the name of the borehole on the drop-down menu on the same page and view the lithologies with depth layers.

However, for a more in-depth analysis he should use the second page of our website, Graphs and Statistics, by clicking on the button at the top right. There he can select the name of the well in the drop-down menu next to the first pie chart and he will see the structure of the pie chart changing, each slice, of a different color, represents a lithology found at different levels of depth, in that same borehole.

At the same time, without the user doing anything, the pie chart relating to temperatures, the horizontal bar chart of the relative ages and the data table with the names of the structural units of the rocks, will also change, showing the different proportions in that borehole.

Basically, with a single click (i.e. by selecting the name of a borehole), the user will be able to know, qualitatively and quantitatively, different information on the object selected, through interactive visualization structures.

Now suppose that our user is interested in understanding the outcome ('esito') of the borehole, grouping them by their type. For example, he wants to know what the outcomes are for exploration boreholes. Then it will use the descriptive pie chart on the bottom left, selecting the 'exploration' type from the drop-down menu and see the result in the pie chart. At the same time, the bar plot alongside will also change in appearance, showing the actual purpose of that type of well selected.

This was an example of using the web page, obviously there are many variants of user cases, based on different needs; we have reported one of the most interesting and complete usage.