



UNIVERSITÀ DI PISA

Visual Analytics
VAST Challenge 2014: Mini-Challenge 1

Ferri Lorenzo (607828)
email l.ferri11@studenti.unipi.it

Visual Analytics, Academic Year 2020/2021
Prof. Salvatore Rinzivillo

Contents

1	Introduction	1
2	Data Preparation	1
3	Visualizations	1
3.1	Layout	2
3.1.1	Position	2
3.1.2	Style	2
3.2	Dashboard	2
3.3	TheForm	3
3.4	BarChart	4
3.5	Network	4
4	Answers to Questions	4
4.1	Provide a visual representation of the structure of the Protectors of Kronos network	4
4.1.1	Who are the leaders? How has the group structure and organization changed over time?	4
4.1.2	Who is part of the extended network? Where are the potential connections between the POK and GASTech?	5
4.2	Describe the events of January 20-21, 2014	5
4.3	Identify at least two possible explanations why the GASTech employees may be missing	5

1 Introduction

Lo scenario è un'isola, Kronos, dove sono presenti un'azienda che estrae gas naturale, la GASTech, ed un'associazione, i Protectors of Kronos (POK). L'azienda provoca l'inquinamento delle acque nell'isola e l'associazione protesta contro l'azienda. Siamo nel Gennaio 2014, mese della IPO di GASTech, quando diversi dipendenti dell'azienda scompaiono. Lo scopo della Mini-Challenge 1 è proprio fornire delle visualizzazioni che permettano di indagare sulla scomparsa. I primi sospettati sono i membri della POK.

2 Data Preparation

La Challenge fornisce principalmente due dataset strutturati: "EmployeeRecords.xlsx", con i dati sui dipendenti di GASTech e "email headers.csv", con i dati sullo scambio di email tra i dipendenti stessi nelle date dal 6 gennaio al 17 gennaio 2014.

I due dataset sono stati processati, uniti ed integrati con altri dati, in modo da renderli utili per produrre visualizzazioni che permettano di individuare i legami tra GASTech e POK e che descrivano le caratteristiche dei dipendenti scomparsi. Le categorie di visualizzazioni prodotte sono: Lista di Nomi, Cards con statistiche, Bar Charts, Network.

Il network in particolare consiste in un grafo di Nodi, labelizzati con i Nomi, che vengono uniti in modo Undirected. Esso deriva dal dataset delle Mails, legando due dipendenti se si sono scambiati più di 2 mails (almeno 3), ed estendendolo con il grafo dei membri della POK, legandoli sulla base del loro rapporto gerarchico.

Oltre agli attributi già presenti sono stati aggiunti i seguenti: "Id", che è un intero crescente che identifica ogni persona (ogni riga del dataset) e serve per il network; "FullName", che contiene il nome completo; "Kidnapped", che è un attributo binario che definisce chi è stato rapito (anche se non è stata fornita una lista precisa); "Membership", che definisce chi appartiene alla POK, alla GASTech o ad entrambe; "CssClass" e "LabelClass", che forniscono la stessa suddivisione di "Membership" e servono per il network; "TargetNodes", che, per ogni persona, definisce la lista di identificativi delle persone a cui essa è collegata nel network (i nodi target). Inoltre, le date sono state tutte trasformate in differenze di anni rispetto alla data di riferimento presente nella challenge (il 2014-01-21). Infine, l'età e gli anni trascorsi dall'inizio del lavoro dei dipendenti sono stati discretizzati raccogliendoli in bin di 3 range di valori.

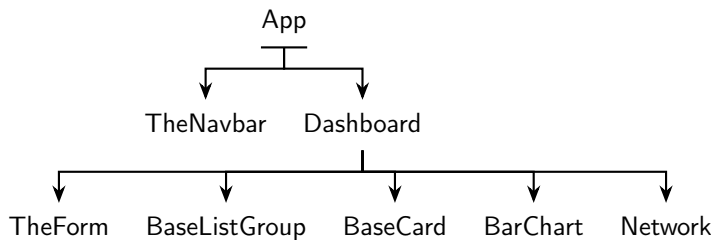
3 Visualizations

Le visualizzazioni sono state prodotte usando lo stack Front End (HTML, CSS, JS), Vue.js, Bootstrap, Plotly.js e vue-d3-network.js.

Il progetto è una single-page web app con una route, "/dashboard", che mostra tutte le visualizzazioni.

Tutto il progetto è stato strutturato dividendo coerentemente ogni componente utilizzato, in modo da renderlo più modulare possibile. La struttura gerarchica dei

componenti è la seguente:



App - contiene tutta la pagina.

TheNavbar - contiene la barra con gli headers in cima alla pagina. Questo è l'unico elemento sempre presente in tutta la app.

Dashboard - Contiene la logica per caricare i dati e trattarli sulla base dei filtri selezionati mediante Crossfilter. Contiene anche tutti i componenti delle visualizzazioni, che si basano tutte su quei dati filtrati.

TheForm - contiene i filtri selezionabili dall'utente.

BaseCard - contiene la struttura di base per le Cards, dove sono presenti le statistiche.

BarChart - contiene la struttura di base ed il trattamento dei dati per la visualizzazione dei Bar Charts.

Network - contiene la struttura di base ed il trattamento dei dati per la visualizzazione del network.

3.1 Layout

3.1.1 Position

La disposizione degli elementi nella pagina è stata effettuata tramite il sistema delle Grid di Bootstrap: a ciascun elemento è assegnata una posizione in termini di righe e colonne, dal componente che contiene tutta l'app in giù nella gerarchia dei componenti. La maggior parte sono posizionati giustificando il contenuto delle righe al centro e facendo adattare le colonne in maniera automatica, mentre alcuni sono posizionati sulla base del sistema a 12 colonne. Bootstrap stesso è stato utilizzato anche per trattare margini e padding.

3.1.2 Style

Sono state utilizzate le classi base di Bootstrap anche per lo stile della maggior parte dei componenti: Navbar, Filtri e Cards. Per i Bar Charts e il Network una parte dello stile viene assegnato direttamente dalla librerie. Per il resto gli stili utilizzati si trovano nei file CSS in "src/assets/styles", che vengono importati nel componente "App".

3.2 Dashboard

In questo componente sono presenti tre categorie di variabili di stato: quelle per i dati fetchati, quelle per le opzioni selezionabili dall'utente nei filtri e quelle per i dati da inserire nella pagina web.

Tutti i dati necessari, riferiti a tutte le persone, vengono letti nella fase di Mounted tramite fetch dal file "persons.json" presente in public/static/data. La Promise restituita che contiene i dati viene poi trattata per eseguire il parsing dei dati, istanziare Crossfilter su quei dati e creare le dimensioni sugli attributi riferiti a: Membership, rapimento, sesso, tipo di lavoro, range di età e range di anni di lavoro. I primi 4 di questi attributi vengono usati come opzioni selezionabili dall'utente (per cui settiamo in automatico i capi "value" e "options" delle apposite variabili di stato), mentre gli ultimi 3 vengono usati nelle visualizzazioni dei Bar Charts (si veda la pagina web per maggiori dettagli).

Un metodo, "refreshDashboard", è stato predisposto per aggiornare tutti i dati che vengono utilizzati nelle visualizzazioni della dashboard. Questo metodo viene richiamato ogni qual volta un filtro viene selezionato dall'utente. E' infatti presente uno Watcher per ogni variabile di stato che contiene i filtri selezionabili dall'utente: quando l'utente seleziona una nuova opzione il capo "value" della variabile riferita a quell'opzione si modifica e quindi lo Watcher triggera una funzione che filtra i dati nella relativa dimensione di crossfilter secondo l'opzione selezionata e richiama la funzione che refresha la Dashboard.

Oltre a tutto ciò, in questo stesso componente vengono inseriti tutti i componenti figli riferiti alla Form, alla Lista, alle Cards, ai Bar Charts e al network, passandogli come props tutti i dati che questi necessitano per funzionare. Cioè in quelle componenti non avremo il fetch di nessun tipo di dato, perché tutti i dati necessari si trovano nello stesso file (persons.json), che viene fetchato e trattato una sola volta nel componente "Dashboard".

3.3 TheForm

La form viene usata come componente figlio di Dashboard, passandogli come props le variabili di stato riferite alle opzioni selezionabili dall'utente, cioè quelle degli attributi sopra elencati, che hanno come campi il "value" corrente selezionato e le possibili "options" della variabile (che non sono altro che i valori distinti dell'attributo). Sono state prodotte due tipi di opzioni selezionabili: quelle a bottone ("Radio") e quelle a tendina.

Le opzioni a bottone sono costituite da tanti tag "input" quante sono le opzioni selezionabili, tutti raccolti in un unico div con class="input-group". Per automatizzare la creazione delle opzioni è stato necessario incapsulare un "input" in un div che, mediante un "v-for", itera sulle opzioni presenti nell'apposita variabile di stato. Per quanto riguarda le opzioni a tendina invece, esse sono costituite da un unico tag "select" con all'interno un numero di tag "option" pari al numero di opzioni selezionabili. Per automatizzarne la creazione quindi è stato usato un "v-for" nel tag "option", che itera sulle opzioni presenti nell'apposita variabile di stato.

In entrambe le categorie di filtri è stato usato un "v-model" per legare il valore selezionato dall'utente al campo "value" dell'apposita variabile di stato.

3.4 BarChart

Anche i Bar Charts sono componenti figli di Dashboard, a cui vengono passati come props i dati da visualizzare, già filtrati e aggregati nel componente padre. Questi dati vengono stored in una variabile di stato e poi passati alla proprietà "data" del componente "Plotly", importato da "vue-plotly". Al fine di rendere dinamici i grafici, nel componente è presente uno watcher sulla props che contiene i dati aggregati. Quando un utente seleziona un filtro, grazie agli watcher nel componente Dashboard viene triggerata la funzione "refreshDashoard", che aggiorna, tra le altre, anche le variabili di stato riferite ai dati aggregati che i Bar Charts devono mostrare, le quali vengono poi passate come props al componente figlio. Successivamente, poiché queste props sono ascoltate dallo watcher del componente figlio, i dati per il Bar Chart vengono aggiornati in automatico.

3.5 Network

Per il network infine viene applicato lo stesso meccanismo applicato per i Bar Charts. In questo caso sono due le variabili passate come props che il componente deve ascoltare: quelle relative ai dati per i nodi e quelle relative ai dati per gli archi. Queste variabili sono inizializzate nel componente Dashboard e passate già processate al componente Network.

4 Answers to Questions

Le visualizzazioni permettono di portare in evidenza fatti importanti per le indagini. Si nota che i soggetti appartenenti alla POK che sono anche dipendenti di GASTech sono 7 persone che fanno parte (per la maggioranza) del dipartimento "Security" ed hanno un'età media e anni di lavoro molo bassi. Di contro coloro che sono stati presumibilmente rapiti fanno parte dei dipartimenti "Administration" ed "Executive" ed hanno un età media e anni di lavoro molto maggiori. Tramite queste ed altre evidenze possiamo passare a rispondere ad alcune delle domande poste.

4.1 Provide a visual representation of the structure of the Protectors of Kronos network

4.1.1 Who are the leaders? How has the group structure and organization changed over time?

- 1997: viene fondata la POK, il cui leader era Henk Bodrogi. Inizialmente l'organizzazione era un piccolo gruppo con sette persone, poi Carmine Osvaldo ha aiutato Bodrogi a connettersi con altro supporto esterno. Il reclutatore era Jeroen Karel.
- 2001: i fondatori iniziali si ritirano dal gruppo e Bodrogi affida la leadership a Elian Karel (giovane attivista con metodi più violenti), che inizia a reclutare civili.

- 2004: Mandor Vann si unisce alla POK e ne diviene il più importante stratega politico.
- 2005: un'organizzazione locale di nome "Save our Wildlands" si unisce con POK. Lucio Jakab (ex membro di quell'associazione) diventa responsabile dei mezzi mediatici di POK (banner, slogan, cartelli, ecc...).
- 2009: Elian muore in prigione per un attacco di cuore e Silvia Marek diventa il leader. A quel punto POK ha 200-300 membri, dei quali i più importanti sono: Mandor Vann, Isia Vann, Lucio Jakab e Lorenzo Di Stefano.

4.1.2 Who is part of the extended network? Where are the potential connections between the POK and GASTech?

Elenchiamo tutti i membri passati e attuali della POK, comprendendo anche coloro che sono solo sospettati di farne parte o che abbiano avuto solo qualche legame. Specifichiamo inoltre il loro ruolo all'interno dell'organizzazione e quali di essi sono dipendenti della GASTech:

Carmine Osvaldo (fondatore), Ale L. Hanne (fondatore), Jeroen Karel (fondatore), Valentine Mies (fondatore), Yanick Cato (fondatore), Joreto Katell (fondatore), Henk Bodrogi (fondatore ed ex leader), Elian Karel (ex leader), Silvia Marek (leader attuale); Mandor Vann (zio di Juliana e Isia Vann); Petrus Gerhard (giornalista), Maha Salo (giornalista); Joclyn Reynolds (WFA project manager); Cesare Nespoli (Government Minister of Health); Lucio Jakab (Ex membro di "Save Our Wildlands"); Lorenzo Di Stefano (professore di scienze ambientali all'Università di Abila); Michale Kraft (ex avvocato di Elian Karel); Jon L. (presente ai comizi); Kriste Bronte (presente ai comizi); Isia Vann (fratello di Juliana Vann e dipendente di GASTech), Hank Fluss (dipendente di GASTech), Inga Ferro (dipendente di GASTech), Loreto Bodrogi (dipendente di GASTech), Hennie Osvaldo (dipendente di GASTech), Minke Mies (dipendente di GASTech), Ruscella Mies Haber (dipendente di GASTech), Rachel Pantanal (dipendente di GASTech, che potrebbe avere una relazione con Isia Vann).

4.2 Describe the events of January 20-21, 2014

La descrizione di questi eventi era interamente in formato testuale non strutturato e quindi difficilmente visualizzabile. Cournque si lascia a disposizione il file "events of January 20-21, 2014", in "research/data", che elenca tutti gli eventi.

4.3 Identify at least two possible explanations why the GASTech employees may be missing

1. Una possibile ragione è la vendetta della famiglia Vann nei confronti di GASTech e del governo di Kronos. L'inchiesta sulla morte di Juliana Vann ha mostrato che la causa fu l'inquinamento causato da GASTech nel 1998 ed il governo locale non ha preso misure forti per prevenire i danni o migliorare la situazione.

2. Un'altra possibile motivazione è il denaro. A causa del minore sostegno finanziario pervenuto alla POK, per colpa del restringimento della rete, i membri potrebbero aver deciso di rapire i dipendenti e richiedere un riscatto di 20 milioni di dollari.
3. Gli altri motivi potrebbero essere legati alla volontà della POK di danneggiare la GASTech, rea di aver inquinato l'ambiente dell'isola, richiamando così l'attenzione della popolazione sulla questione dell'inquinamento.
4. Oppure dovuti ad una rivalsa dei dipendenti della GASTech di basso livello nei confronti di quelli di alto livello.