

Aula 4 - Análise textual de uma amostra do Twitter

José Devezas <<u>ild@fe.up.pt</u>> Faculdade de Engenharia da Universidade do Porto

Gestão da Informação em Redes Sociais Mestrado em Ciência da Informação

Conteúdos

- Introdução
- Conceitos básicos sobre análise de texto
- Tutorial básico de KNIME
- 2^ª avaliação

Introdução

- Nesta aula vamos esquecer por algum tempo a rede e focar apenas na análise textual (text mining).
- Em particular, vamos identificar:
 - Entidades mencionadas em grupos de documentos (*named entity recognition*);
 - Bem como o sentimento dominante em grupos de documentos (sentiment analysis).
- Para isso, vamos usar o KNIME, software open source para análise de dados, baseado em programação visual
 - Os "Workflows" serão fornecidos, apenas requerendo pequenas alterações.
- Segundo trabalho individual focado na análise de texto.

Conceitos básicos sobre análise de texto

Pré-processamento

- Ler ficheiro e remover ruído.
 - Ruído é tudo aquilo que é irrelevante para o nosso objetivo de análise.
 - Para a análise textual do Twitter, removemos:
 - i. URLs;
 - ii. Menções (p.e., @elonmusk);
 - iii. Inícios de tweet com "RT";
 - iv. Caracteres que representam símbolos ou emoji (p.e., 😳 🕸 오).
 - Os elementos de i. a iii. são melhor analisados com recurso ao grafo (p.e., i.) link analysis para URLs, ii.) grafo de menções, iii.) grafo de retweets).
- Converter o texto de cada tweet para uma lista de palavras (tokenization).

Conceitos básicos sobre análise de texto

Reconhecimento de entidades mencionadas

- Entidades são sequências de palavras que se referem a pessoas, organizações, localizações, etc.
- A deteção de entidades depende geralmente da criação de modelos (por exemplo aprendidos automaticamente com *machine learning*, ou criados com recurso a gramáticas), que dependem da língua e de um conjunto de dados previamente anotado.
 - As anotações identificam as fronteiras das entidades no texto, identificando a sua classe. Por exemplo:
 "<person>José Devezas</person> trabalha na <organization>FEUP</organization>, que fica na cidade do <location>Porto</location>."
- Uma forma rápida de caracterizar o conjunto de dados é identificar as palavras-chave ou entidades mais frequentes, visualizando-as sobre a forma de uma *tag cloud*.

Conceitos básicos sobre análise de texto

Análise de sentimento

- Uma das formas mais simples de análise de sentimento é a classificação da polaridade de um texto:
 - Positivo;
 - Neutro;
 - Negativo.
- Para isso é geralmente necessário aprender modelos com técnicas de *machine learning* e com base num conjunto de textos previamente anotados com a classe de polaridade.
- Durante a aprendizagem, o modelo vai sendo melhorado conforme vão sendo fornecidos mais textos anotados com a polaridade.
 - Intuição: quanto mais textos positivos, melhor sabemos se determinada palavra é usada em textos com sentimento positivo; palavras mais positivas implicam textos mais positivos (e vice-versa).

Tutorial básico de KNIME

Visão geral

O KNIME tem vários nós que executam tarefas específicas. Os nós podem ter entradas e saídas. Existem nós, como o "File Reader", que apenas têm uma saída e servem para ler ficheiros do sistema.

Workflows-exemplo (EXAMPLES) Descrição detalhada e workflows criados pelo sobre o nó selecionado. utilizador (LOCAL) 📑 🗕 🔛 🔛 😓 😂 100% -C O O 🦻 A KNIME Explorer 23 📋 🔥 0: GIRS - Tag Cloud 🕱 🛕 2: GIRS - Polarity Classificatio 👍 Node Descri... 🔉 📟 🗖 BEE File Reader Loading and Filtering ▶ A EXAMPLES (knime-guest@http:// File Reader String Replacer Strings To Document Column Filter ▼ ALOCAL (Local Workspace) This node can be used to read D+ + ► 📆 I ► +🗎 I ► ±± ► A GIRS - Polarity Classification data from an SCII file or URL location. It can be configured to read various formats. Load Twitter data Remove URLs Create Document Extract only the from text' column Document' column When you open the node's configuration dialog and provide a filename, it tries to 🕼 Workflow Coach 😫 🛛 🚯 - -Named Entity Recognition guess the reader's settings by analyzing the content of the OpenNLP NE tagger OpenNLP NE tagger OpenNLP NE tagger file. Check the results of these ► 🛱 ► ► 📸 ► settings in the preview table. If BA the data shown is not correct or an error is reported, you can Identify persons Identify organizations Identify locations - 0 A Node Repository adjust the settings manually (see below) 2 The file analysis runs in the Named Entity Frequency Calculation) L IO background and can be cut short by clicking the "Quick 🕨 🎤 Mar Bag of Words Tag Filter Creator Tags to String TE GroupBy scan", which shows if the Q Views → 闘 analysis takes longer. In this Analytics ▶ 融 • <u>-</u>-case the file is not analyzed ▶ ■ Database completely but only the first ▶ Cother Data Types BE Outline 23 Console 🕅 R. AT RO -----> () Structured Data . (Scripting NIME Conse ERROR Split Collection Column 3:298 Tool Integration Temp folder */tmp/knime 08 Sentiment Cl79 10 10 10 10 10 ERROR Random Forest Predictor 3:302 Temp folder "/tmp/knime 08 Sentiment Cl79-🕨 🍻 Community Nades ERROR Scorer 3:280 Temp folder */tmp/knime 08 Sentiment Cl79431 KNIME Labs -Temp folder '/tmp/knime_08_Sentiment_Cl79431' ERROR Scorer 3:280 No row filter specified Vorkflow Control WARN Row Filter 2:307 WARN Joiner 2:305:295 Memory is low. I have no chance to free memo > 🔥 Social Media * * * * * EFROR Vocabulary Extractor (deprecated) 2:306:293 Unable to clone input data at pr ▶ ➡ Reporting WARN KNIMEApplication\$3 Potential deadlock in SWT Display thread dete WARN Joiner 2:306:299 Please define at least one joining column pa: ► a Chemistry * * WARN loine 2:305:299 Memory is low. I have no chance to free memor ChemAxon / Infocom ERROR Vocabulary Extractor (deprecated) 2:306:293 Unable to clone input data at pr

Nós disponíveis

Configurar um nó

Cada nó pode ser ligado a outros nós (clicando numa entrada ou saída e arrastando), devendo ser previamente configurado (selecionando "Configure" com o botão direito do rato ou utilizando duplo clique).



Configurar o "File Reader"

O nó "File Reader" permite carregar dados a partir de ficheiros de texto. No nosso caso, vamos carregar um CSV com o conjunto de tweets recolhidos a partir das contas do @realDonaldTrump e do @elonmusk.

Enter AS	[Flow Variables Job Mana III data file location: (press 'El	ger Selection Memory Policy nter' to update preview)
file:/hom	e/jldevezas/Teaching/GIRS/Le	ture4/datasets/elonmusk 🔻 Browse
	🔲 Preserve user settir	ngs for new location Rescan
-Basic Set	tings	
read r	ow IDs Column d	elimiter: , 🔹 🚺 Advanced)—
ile Read	n: Advanced Settings	
Limit Row	S Character decoding Mis	sing Value Pattern
Quote su	port Decimal Separator	Ignore spaces Short Lines unique Rowll
Quotes ca Escape ch	n be multi-character patterns nracter (if checked) is always	(for example: <quote>). the backslash ('\').</quote>
Enter a ne	w quote character:	currently set quotes
		lesc. ()
📃 su	pport esc character ('\')	< bba
		< Romova
\backslash		< Remove
1		
1		
Allow mult	Line quoted strings:	
Allow mult	i line quoted strings: strings can extend over multi	ple lines
Allow mul	i line quoted strings: strings can extend over multi	ple lines
Al low mul	Lline quoted strings: strings can extend over multi	ple lines
Allow mul	Lline quoted strings: strings can extend over multi	ple lines
Allow mul	i line quoted strings: strings can extend over multi Read from	ple lines
Allow mul	i line quoted strings: strings can extend over multi Read fror @elonmusk Well done, love #Te BT @elonmusk. Officially verified	ple lines NML file Ok Cancel sla https://t.co/qql2ftm8/v as the first perduction electric car to exceed 1000
Allow mul	Line quoted strings: strings can extend over multi Read fror @elonmusk Well done, love #Te RT @elonmusk. Officially verified RT @elonmusk. Officially verified	ple lines Dk Cancel sla https://t.co/qql2ftm8vv as the first production electric car to exceed 1000 as the first production electric car to exceed 1000 as the first production electric car to exceed 1000
Allow mul	Line quoted strings: strings can extend over multi Read fror @elonmusk Well done, love #TE RT @elonmusk: Officially verified RT @elonmusk: Dfficially verified RT @elonmusk: Dfficially verified	ple lines NML file Ok Cancel sla https://t.co/qql2ftm8/v as the first production electric car to exceed 1000 as the first production electric car to exceed 1000 laining. That has remained constant throughout hin
Allow mul	Line quoted strings: strings can extend over multi Read fror @elonmusk Well done, love #7E RT @elonmusk: Officially verified RT @elonmusk: Except for comp RT @elonmusk: Except for comp	ple lines NKML file Ok Cancel sla https://t.co/qql2ftm8Vv as the first production electric car to exceed 1000 as the first production electric car to exceed 1000 laining. That has remained constant throughout hi laining. That has remained constant throughout hi baining. That has remained constant throughout hi
Allow mul	Etine quoted strings: strings can extend over multi Read fror @elonmusk Well done, love #Te RT @elonmusk: Officially verified RT @elonmusk: Except for comp RT @elonmusk: Except for comp RT @elonmusk: Except for comp	ple lines NXML file Ok Cancel sla https://t.co/qql2ftm8vv sta first production electric car to exceed 1000 as the first production electric car to exceed 1000 as the first production electric car to exceed 1000 as the first production electric car to exceed 1000 Jaining. That has remained constant throughout hi laining. That has remained constant throughout hi laining. That has remained constant throughout hi laining.
Allow mol Quoted Row16 Row17 Row18 Row19 Row20 Row21	Read from @elonmusk: Well done, love #TE RT @elonmusk: Officially verified RT @elonmusk: Stept for comp RT @elonmusk: Except for comp @william_sw @MrGeeDeeArr @jc @Kabeerisgod @elonmusk Fuck	ple lines NML file Ok Cancel sla https://t.co/qql2ftm8Vv as the first production electric car to exceed 1000 as the first production electric car to exceed 1000 laining. That has remained constant throughout hi laining. That has remained constant throughout hi htm_Gardi @TeslaOwnersIT @elonmusk @marcodali off
Allow mel Quoted Row16 Row17 Row18 Row20 Row20 Row21 Row22 Row22	Read from @elonmusk Well done, love #Te RT @elonmusk: Officially verified RT @elonmusk: Officially verified RT @elonmusk: Except for comp @william, sw @MrGeeDeeArr @jc @Kabeerisgad @elonmusk Fuck Donmed a comp true @elonemusk Fuck Donmed a comp true @elonemusk Fuck	ple lines NML file NKML f
Allow mult	Read from @elonmusk Well done, love #Te RT @elonmusk: Officially verified RT @elonmusk: Except for comp RT @elonmusk: Except for comp RT @elonmusk: Except for comp RT @elonmusk: Except for comp @william sw @MrGeeDearr @jo @william su @MrGeeDearr @jo @william su @more and a comp to a compared william su @MrGeeDearr @jo	ple lines NML file Ok Cancel Isla https://t.co/gql2tfm8Vv as the first production electric car to exceed 1000 as the first production electric car to exceed 1000 laining. That has remained constant throughout hi laining. That has remained constant throughout hi laining. That has remained constant throughout hi laining. Gardi @TeslaOwnerstT @elonmusk @marcodali off wk bttps://t.codbull.Pi4pa5b
Allow mult	Read from @elonmusk: Well done, love #Te RT @elonmusk: Officially verified RT @elonmusk: Strept for comp RT @elonmusk: Except for comp @william_sw @MrGeeDeeArr @jc @kabeerisgod @elonmusk Fuck Dreame do comp two. @elopmusk	ple lines n XML file OK Cancel sla https://t.co/qql2ftm8W as the first production electric car to exceed 1000 as the first production electric car to exceed 1000 laining. That has remained constant throughout hi laining. That has remained constant hi laining. That has remained constant hi laining. Th
Allow mol	Ine quoted strings: strings can extend over multi Read fror @elonmusk Well done, love #Te RT @elonmusk: Officially verified RT @elonmusk: Except for comp RT @elonmusk: Except for comp @william_sw @MTGeeDeeArr @jc @kabeerisgod @elonmusk Fuck Dromme do como true. @elonmusk	ple lines NXML file OK Cancel sla https://t.co/qql2ftm8Vv as the first production electric car to exceed 1000 as the first production electric car to exceed 1000 laining. That has remained constant throughout hi hph. Gardi @Tesla0wnersIT @elonmusk @marcodali off code bittos://t.co/bULB/4045b I
Altow musi 2 Quoted Row16 Row17 Row18 Row19 Row20 Row20 Row22 Pow23	Eline quoted strings: strings can extend over multi Read fror @elonmusk. Well done, love #1e RT @elonmusk. Officially verified RT @elonmusk. Except for comp RT @elonmusk. Except for comp @william sw @MrGeeDearr @jo @william sw @jo @william sw @jo @william sw @MrGeeDearr @jo @william sw @jo	ple lines n XML file Ok Cancel sla https://t.co/qqlZftm8Vv as the first production electric car to exceed 1000 iaining. That has remained constant throughout hi laining. That has remained constant throughout hi hybn. Gardi @TeslaOwnersIT @elonmusk @marcodali off usk bttps://t.co/bull.BidpASb.
Allow musi Quoted Row16 Row17 Row18 Row20 Row20 Row20 Row22 Dow22	Line quoted strings: strings can extend over multi Read fror @elonmusk Well done, love #Te RT @elonmusk: Officially verified RT @elonmusk: Except for comp RT @elonmusk: Based and the second for a finite RT @elonmusk	ple lines NML file Ok Cancel sla https://t.co/qql2ftm8/v as the first production electric car to exceed 1000 as the first production electric car to exceed 1000 laining. That has remained constant throughout hi laining. That has remained constant throughout hi htm, Gardi @TeslaOwnersIT @elonmusk @marcodali off uck https://t.co/blil.DidaASb.
Allow muł Quoted Row16 Row17 Row18 Row19 Row20 Row20 Row20 Row22 Dow22	Ine quoted strings: strings can extend over multi Read from @elonmusk Well done, love #TE RT @elonmusk: Officially verified RT @elonmusk: Except for comp RT @elonmusk: Except for comp @william_sw @MrGeeDeeArr @jc @kabeerisgod @elonmusk Fuck Drome do como true_@elopmusk	ple lines n XML file Ok Cancel sla https://t.co/qql2ftm8W as the first production electric car to exceed 1000 as the first production electric car to exceed 1000 laining. That has remained constant throughout hi laining. That has remained con

Importante!

Executar o Workflow

Para executar o Workflow, deve clicar-se com o botão direito no nó para o qual queremos obter um resultado e clicar em "Execute". O nó deve passar para o estado verde ou então devolver um aviso de erro.



Visualizar os resultados

Para visualizar os resultados de um nó, deve clicar-se com o botão direito e selecionar uma das opções disponíveis, geralmente com um ícone de lupa+tabela. As opções disponíveis dependem do nó em particular.



Visualizar os resultados para o "String Replacer"

O "String Replacer" remove ruído do texto (URLs, menções a contas do Twitter, expressões de retweet, e caracteres não-alfanuméricos). Ao clicar em "Input with replaced values" podemos ver o resultado deste processamento.

Row ID S te RowO Arg Row1 Arg Row2 Offit Row3 Have Row6 Elon Row7 Offit Row6 elon Row10 Offit Row8 Offit Row10 Offit Row11 Offit Row12 Exc Row13 Exc Row14 Offit Row15 Offit Row16 Well Row17 Offit Row18 Offit Row19 Exc Row10 Diffit Row11 Offit Row12 Exc Row13 Exc Row14 Offit Row21 Fut Row22 Fut Row24 Agre Row25 Offit Row25 Offit	xt iments followed of course by complaints ECHOCHAMBER OF ially verified as the first production ele you considered entering LeMans New you considered entering LeMans New you considered entering that has remained 0 kmh vask My dyo coloured this for my birthd ially verified as the first production ele ially verified as the first production ele	S favorit FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE	favorit 0	S replyT NA Google NA elonmusk NA NA NA NA NA NA NA NA NA NA	Sur Sur Sur Sur Sur Sur Sur Sur Sur Sur
Row0 Arg Row1 Particle Row2 Offic Row3 Have Row4 Exc Row5 4 Row6 elonr Row7 Offic Row8 Offic Row9 Exc Row10 Offic Row11 Offic Row13 Exc Row14 Offic Row15 Offic Row16 Well Row19 Exc Row10 Offic Row12 Offic Row13 Exc Row14 Offic Row19 Exc Row20 Exc Row21 Fuc Row22 Fuc Row23 Drea Row24 Agre Row25 Offic Row26 Exc	Imments followed of course by complaints ECHOCHAMBER OF ially verified as the first production ele you considered entering LeMans New spt for complaining That has remained of kmh usk My 4yo coloured this for my birthd. ially verified as the first production ele ially verified as the first production ele	FALSE FALSE	0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0	NA Google NA elonmusk NA NA NA NA NA NA NA NA NA	Sur Sur Sur Sur Sur Sur Sur Sur Sur Sur
Row1 Row2 Offiti Row3 Have Row3 Row4 Exc. Row5 4 Row5 4 Row6 elont Row7 Offiti Row9 Exc. Row10 Offiti Row11 Offiti Row11 Offiti Row12 Exc. Row13 Exc. Row14 Offiti Row14 Offiti Row15 Offiti Row15 Offiti Row17 Offiti Row19 Exc. Row20 Exc. Row20 Exc. Row21 Fuc Row21 Fit Row22 Fuc Row22 Fuc Row23 Offet Row24 Agre Row25 Offet	ECHOCHAMBER OF ially verified as the first production ele you considered entering LeMans New pt for complaining That has remained 0 kmh nusk My 4yo coloured this for my birthd ially verified as the first production ele ially verified as the first production ele ially verified as the first production ele ially verified as the first production ele pt for complaining That has remained pt for complaining That has remained ially verified as the first production ele ially verified as the first production ele	FALSE	0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0	Google NA elonmusk NA NA NA NA NA NA NA NA NA	Sur Sur Sur Sur Sur Sur Sur Sur Sur Sur
Row2 Offin Row3 Have Row4 Exc. Row5 4 Row6 elonr Row7 Offin Row8 Offin Row10 Offin Row12 Exc. Row13 Exc. Row16 Well Row17 Offin Row18 Offin Row19 Exc. Row18 Offin Row19 Exc. Row10 Offin Row11 Offin Row12 Exc. Row13 Exc. Row14 Offin Row15 Offin Row20 Exc. Row21 F Row22 Fuc. Row23 Drea Row23 Drea Row24 Agre Row25 Offin	ially verified as the first production ele you considered entering LeMans New py for complaining That has remained 0 kmh usk My dyo coloured this for my birthd ially verified as the first production ele ially verified as the first production ele	FALSE	0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0	NA elonmusk NA NA NA NA NA NA NA NA NA	Sui Sui Sui Sui Sui Sui Sui Sui Sui
Row3 Have Row4 Exc: Row5 4 Row6 elonr Row7 Offit Row8 Offit Row9 Exc: Row10 Offit Row11 Offit Row12 Exc: Row13 Exc: Row14 Offit Row15 Offit Row16 Well Row17 Offit Row19 Exc: Row20 Exc: Row21 f Row22 Fuc Row23 Drea Row24 Agre Row25 Offit Row25 Offit	you considered entering LeMans New ppt for complaining That has remained to kmh usk My dyo coloured this for my birthd ially verified as the first production ele ially verified as the first production ele ially verified as the first production ele ially verified as the first production ele pt for complaining That has remained pt for complaining That has remained ially verified as the first production ele ially verified as the first production ele	 FALSE 	0 0 0 0 0 0 0 0 0 0 0 0 0 0 0	elonmusk NA NA NA NA NA NA NA NA NA	Sui Sui Sui Sui Sui Sui Sui Sui
Row4 Exc. Row5 4 Row6 elonr Row7 Offitt Row8 Offitt Row10 Offitt Row11 Offitt Row12 Exc. Row13 Exc. Row14 Offitt Row15 Offitt Row16 Well Row17 Offitt Row18 Offitt Row21 F Row22 Fuc Row23 Drea Row24 Agre Row25 Offitt Row25 Offitt	ppt for complaining That has remained 0 kmh uak My 4yo coloured this for my birthd ially verified as the first production ele ially verified as the first production ele ally verified as the first production ele ially verified as the first production ele done love Tesla	 FALSE 	0 0 0 0 0 0 0 0 0 0 0 0 0 0 0	NA NA NA NA NA NA NA NA NA	Sui Sui Sui Sui Sui Sui Sui Sui
Row5 4 Row6 elon Row7 Offit Row8 Offit Row10 Offit Row11 Offit Row12 Exc: Row13 Exc: Row14 Offit Row15 Offit Row16 Well Row17 Offit Row218 Offit Row219 Exc: Row22 Fuc Row23 Drea Row24 Agre Row25 Offit Row26 Exc:	0 kmh usk My 4yo coloured this for my birthd ially verified as the first production ele pt for complaining That has renduction ele ially verified as the first production ele ially verified as the first production ele pt for complaining That has remained ially verified as the first production ele ially verified as the first production ele	FALSE	0 0 0 0 0 0 0 0 0 0 0 0 0 0	NA NA NA NA NA NA NA NA	Sui Sui Sui Sui Sui Sui Sui
Row6 elonr Row7 Offit Row8 Offit Row8 Offit Row10 Offit Row11 Offit Row12 Exc. Row13 Offit Row14 Offit Row15 Offit Row17 Offit Row18 Exc. Row20 Exc. Row20 Exc. Row21 Fit Row22 Fuc Row23 Drea Row24 Agre Row25 Offit	usk My dyo coloured this for my birthd ially verified as the first production ele ially verified as the first production ele pt for complaining That has remained ially verified as the first production ele pt for complaining That has remained pt for complaining That has remained ially verified as the first production ele ially verified as the first production ele ially verified as the first production ele ially verified as the first production ele to ne love Tesla	 FALSE 	0 0 0 0 0 0 0 0 0 0 0	NA NA NA NA NA NA NA	Sui Sui Sui Sui Sui Sui
Row7 Offit Row8 Offit Row9 Row10 Offit Row11 Offit Row12 Exc. Row13 Exc. Row13 Exc. Row13 Exc. Row13 Exc. Row14 Offit Row15 Offit Row17 Offit Row16 Well Row17 Offit Row18 Offit Row18 Offit Row20 Exc. Row21 f Row22 Fuc. Row22 Fuc. Row22 Prea Row23 Drea Row24 Agre Row25 Offit Row25 Offit Row26 Exc. Row26 Exc.	ially verified as the first production ele ially verified as the first production ele ially verified as the first production ele ially verified as the first production ele ept for complaining That has remained pt for complaining That has remained ially verified as the first production ele ially verified as the first production ele ially verified as the first production ele ially verified as the first production ele	FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE	0 0 0 0 0 0 0 0 0 0	NA NA NA NA NA NA	Sui Sui Sui Sui Sui
Row8 Offin Row9 Exc: Row10 Offin Row11 Offin Row13 Exc: Row14 Offin Row15 Offin Row16 Well Row19 Exc: Row19 Exc: Row20 Exc: Row21 If Row22 Fuc Row23 Drea Row24 Agre Row25 Offin	ially verified as the first production ele ppt for complaining That has remained ially verified as the first production ele pt for complaining That has remained pt for complaining That has remained ially verified as the first production ele ially verified as the first production ele ially verified as the first production ele tone love Tesla	FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE	0 0 0 0 0 0 0 0	NA NA NA NA NA	Sui Sui Sui Sui Sui
Row9 Exc. Row10 Offit Row11 Offit Row12 Exc. Row13 Exc. Row14 Offit Row15 Offit Row16 Well Row17 Offit Row18 Exc. Row21 Ft Row22 Fuc Row23 Drea Row24 Agre Row25 Offit	pt for complaining That has remained ially verified as the first production ele ially verified as the first production ele pt for complaining That has remained pt for complaining That has remained ially verified as the first production ele ially verified as the first production ele done love Tesla	FALSE FALSE FALSE FALSE FALSE FALSE FALSE	0 0 0 0 0 0	NA NA NA NA NA	Sui Sui Sui Sui
Row10 Offitt Row11 Offitt Row12 Exc. Row13 Exc. Row14 Offitt Row15 Offitt Row16 Well Row18 Offitt Row19 Exc. Row20 Exc. Row21 F Row22 Fut Row23 Drea Row24 Agret Row25 Offit Row26 Exc.	ially verified as the first production ele ially verified as the first production ele spt for complaining That has remained ially verified as the first production ele ially verified as the first production ele ially verified as the first production ele	FALSE FALSE FALSE FALSE FALSE FALSE	0 0 0 0 0	NA NA NA	Su Su Su
Row11 Offit Row12 Exc: Row13 Exc: Row14 Offit Row15 Offit Row16 Well Row17 Offit Row19 Exc: Row20 Exc: Row21 Fit Row22 Fuc Row23 Drea Row24 Agre Row25 Offit Row26 Exc:	ially verified as the first production ele ppt for complaining That has remained ppt for complaining That has remained ially verified as the first production ele ially verified as the first production ele done love Tesla	FALSE FALSE FALSE FALSE FALSE	0 0 0 0	NA NA NA	Su
Row12 Exc. Row13 Exc. Row14 Offil Row15 Offil Row16 Well Row17 Offil Row18 Offil Row19 Exc. Row20 Exc. Row21 Fl Row23 Drea Row24 Agre Row25 Offil Row26 Exc.	ept for complaining That has remained opt for complaining That has remained ially verified as the first production ele ially verified as the first production ele done love Tesla	FALSE FALSE FALSE FALSE	0 0 0	NA NA	Su
Row13 Exc. Row14 Offit Row15 Offit Row16 Well Row18 Offit Row19 Exc. Row21 Flat Row22 Fuc. Row23 Drea Row24 Agre Row25 Offit Row26 Exc.	pt for complaining That has remained ially verified as the first production ele ially verified as the first production ele done love Tesla	FALSE FALSE FALSE	0 0	NA	SLU
Row14 Offit Row15 Offit Row16 Well Row17 Offit Row18 Offit Row20 Exc. Row21 F Row22 Fuc Row23 Drea Row24 Agre Row25 Offit Row26 Exc.	ially verified as the first production ele ially verified as the first production ele done love Tesla	FALSE FALSE	0		1 U UI
Row15 Offit Row16 Well Row17 Offit Row18 Offit Row20 Exc. Row21 Ft Row22 Fuc Row23 Drea Row24 Agre Row25 Offit Row26 Exc.	ially verified as the first production ele done love Tesla	FALSE		INA	Su
Row16 Well Row17 Offnit Row18 Offnit Row20 Exc. Row21 F Row22 Fuc Row23 Drea Row23 Drea Row24 Agre Row25 Offnit Row26 Exc.	done love Tesla		0	NA	Su
Row17 Offic Row18 Offic Row19 Exc: Row20 Exc: Row21 F Row22 Fuc Row23 Drea Row24 Agree Row25 Offic Row26 Exc: Row26 Exc:		FALSE	0	elonmusk	Su
Row18 Offic Row19 Exci Row20 Exci Row21 Fit Row23 Drea Row24 Agre Row25 Offic Row26 Exci	ially verified as the first production ele	FALSE	0	NA	Su
Row19 Exci Row20 Exci Row21 F Row22 Fuc Row23 Drea Row24 Agre Row25 Offrit Row26 Exci	ially verified as the first production ele	FALSE	0	NA	Su
Row20 Exc. Row21 F Row22 Fuc Row23 Drea Row24 Agre Row25 Offic Row26 Exc.	ept for complaining That has remained	. FALSE	0	NA	Su
Row21FRow22FucRow23DreaRow24AgreRow25OfficRow26Exco	ept for complaining That has remained	. FALSE	0	NA	Su
Row22FucRow23DreaRow24AgreeRow25OfficeRow26Exce	rom their website	FALSE	0	william sw	Su
Row23 Drea Row24 Agre Row25 Offic Row26 Excr	< off	FALSE	0	Kabeeris	Su
Row24 Agre Row25 Offic Row26 Exce	ns do come true	FALSE	0	NA	Su
Row25 Offic Row26 Exc	ed Complaining is a universal constant .	. FALSE	0	elonmusk	Su
Row26 Exc	ially verified as the first production ele	FALSE	0	NA	Su
	ept for complaining That has remained	. FALSE	0	NA	Su
Row27 True	although the changemyview subreddit.	. FALSE	0	NA	Su
Row28 Exci	ept for complaining That has remained	. FALSE	0	NA	Su
Row29 Exc	ept for complaining That has remained	. FALSE	0	NA	Su
Row30 Exc	ept for complaining That has remained	. FALSE	0	NA	Su
Row31 Offic		FALSE	0	NA	Sur
Row32 Offic	ially verified as the first production ele	FALSE	0	NA	Su
Row33 Th	ially verified as the first production ele ially verified as the first production ele			planetary	Sur

2ª avaliação

- Individualmente (à vez, caso não haja computadores para todos):
 - Vamos selecionar dois atributos (colunas no CSV) e decidir valores para filtrar essas colunas (p.e., retweetCount > 10000).
 - Alunos no mesmo computador não podem escolher as mesmas colunas.
 - Vamos explorar a tag cloud de entidades mencionadas em grupos de tweets restringidos por um e pelo ou filtro.
 - Vamos explorar a polaridade nos mesmos grupos de tweets, aplicando alternadamente ambos os filtros.
- Individualmente:
 - Cada aluno deve escrever um mini relatório, analisando manualmente de forma crítica os resultados obt (p.e., citando alguns tweets identificados correta ou incorretamente como positivos ou negativos).
- Entrega:
 - Pode ser feita até á próxima aula.