Hands-on: Introdução a Python na Análise de Dados

...

Rodrigo Ramos

Ciência de Dados

- Motivação
- Produto de Dados
- Ciência de Dados
- Processo de Construção

[Loucos] Tempos Modernos

Produto de Dados

“Um produto de dados é um produto que usa dados / informações para facilitar o fornecimento do valor esperado.”

Exemplos de empresas que fornecem data products:

NETFLIX

waze

amazon

Os problemas não estão relacionados “apenas” ao grande volumes de dados!

os Vs de Big Data

- Volume
- Velocidade
- Variedade
- Veracidade

....

adicione o seu próprio V aqui :)

Assim como o petróleo, dados geram riquezas ...

... porém, precisam ser extraídos, armazenados, e refinados.

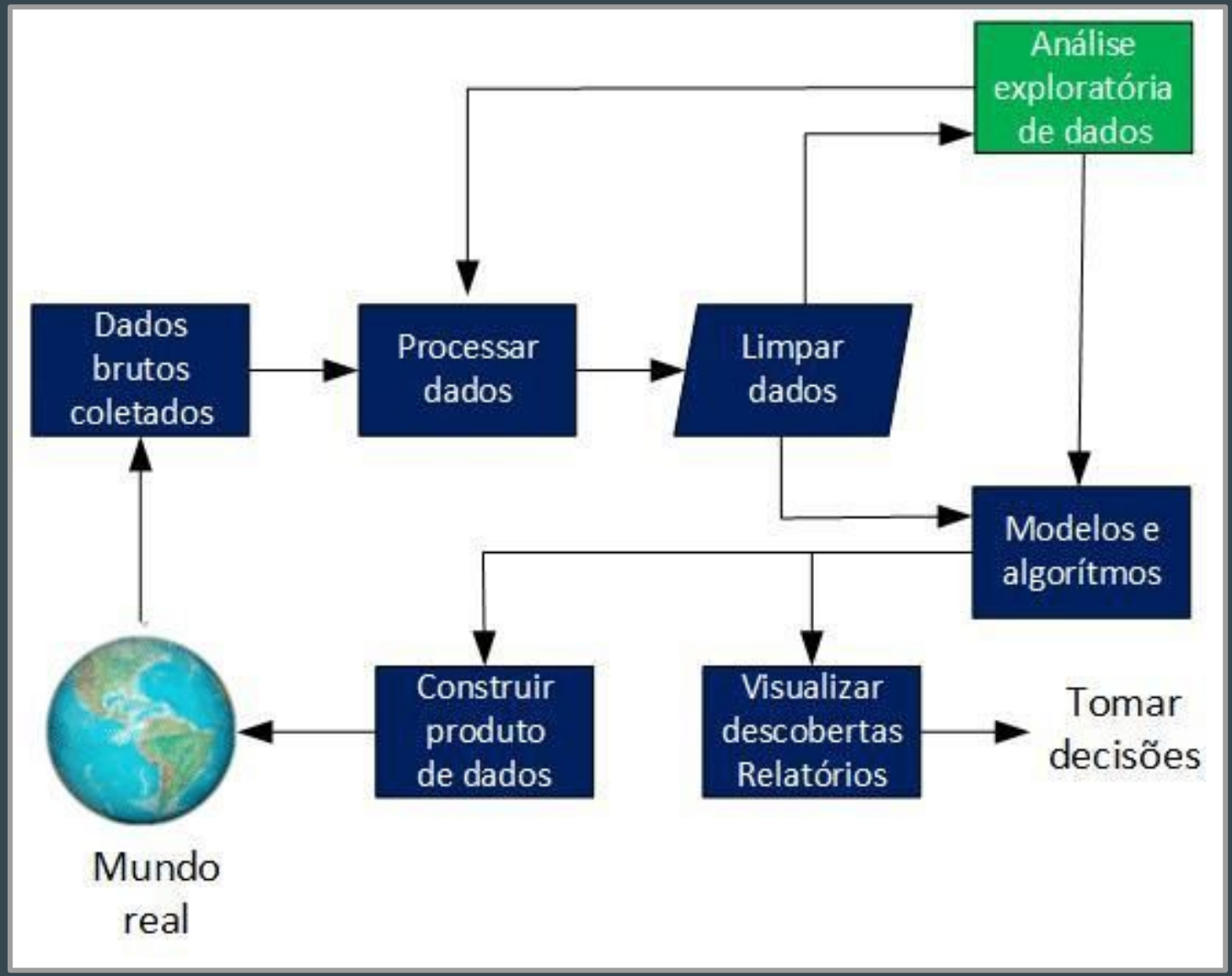
antes de serem úteis.

Ciência de Dados:
A ciência [ou arte] de ...

- Descobrir o que não conhecemos sobre os dados;
- Compreender a estrutura e relação dos dados;
- Predizer informações sobre os dados;
- Criar *produto de dados* de impacto no mercado;
- Comunicar estórias relevantes ao mercado a partir dos dados;
- Apoiar com confiança decisões guiadas pelos dados;

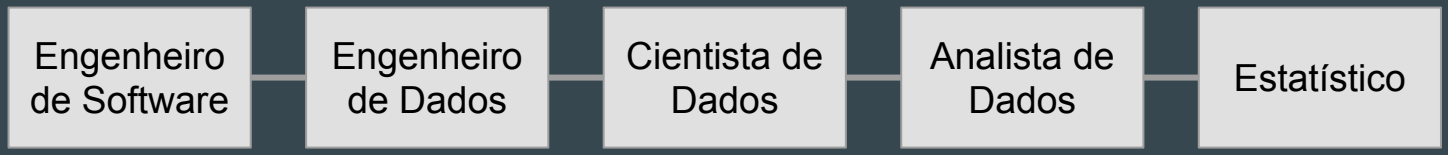
Diagrama de Venn da Ciência de Dados

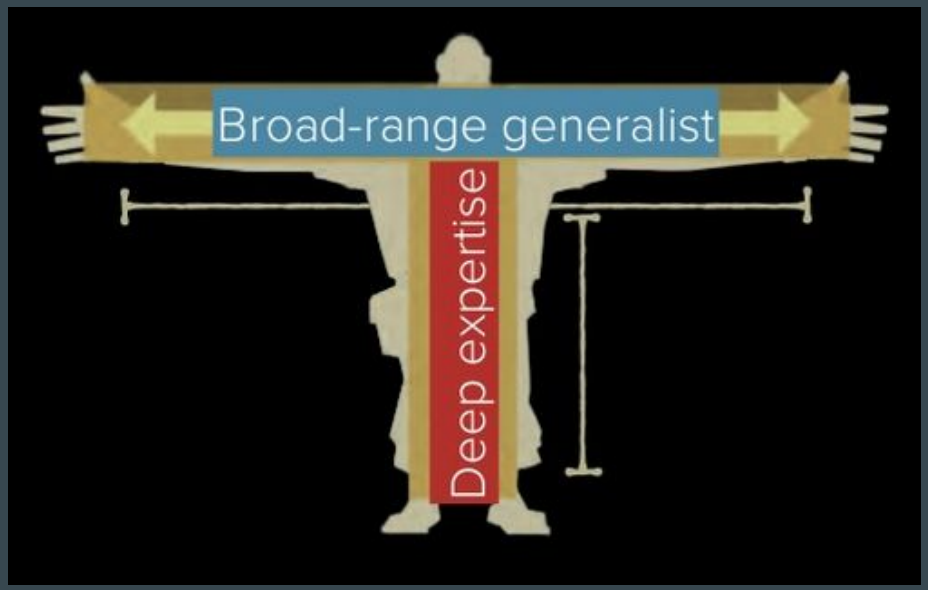
Processo da Ciência de Dados



Equipes de desenvolvimento

- Data Products são criados por equipes multidisciplinares
  - Igual aos esportes, especialização é importante

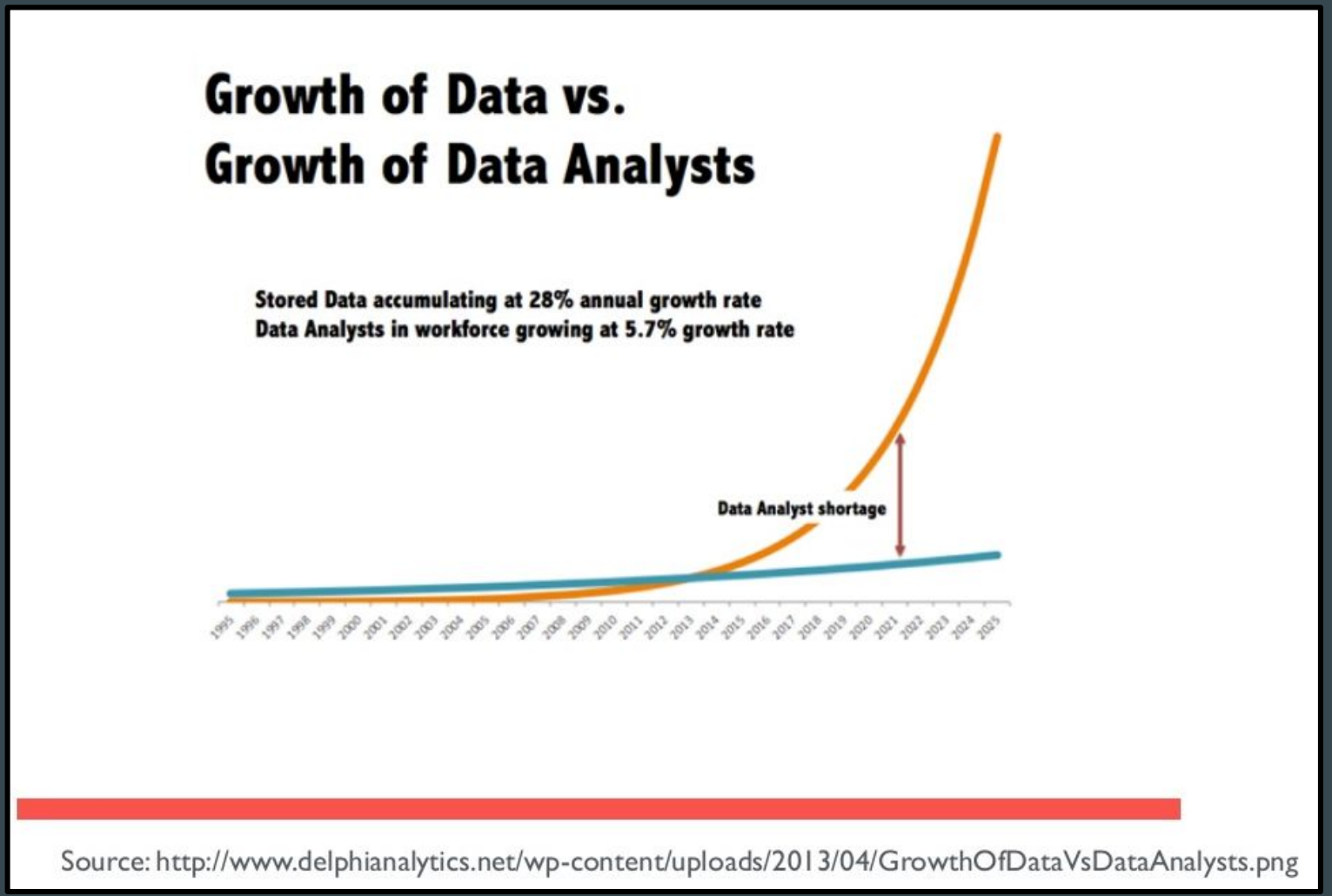




Fonte: http://www.exploringdatascience.com/the-data-science-clock/

Ciência de Dados: Uma área em ascensão

... porém com uma grande demanda de profissionais.



Introdução a Python

Por que Python?

- Fácil de aprender
- Linguagem de uso geral, porém também atrativa para computação científica
- Conjunto vasto e poderoso de bibliotecas

Algumas Bibliotecas importantes para análise de dados

- NumPy
- Pandas
- Matplotlib
- SciPy
- Statsmodels
- Natural Language Toolkit (NLTK)

Mão na massa - Python Básico

- Instale o Python
  - disponível https://www.python.org/
  - versão 3.x
- Instale os pacotes numpy , pandas, matplotlib

- Acompanhe o notebook em: https://github.com/rodrigoteixeira/demo_python_data_analysis

1o Exemplo: Consumo de Álcool Mundial

Dados oriundos de

FiveThirtyEight

Who Drinks The Most Beer, Spirits And Wine?

Top 10 countries by servings consumed per person, 2010

	BEER		SPIRITS		WINE	
1	Namibia	376	Grenada	438	France	370
2	Czech Rep.	361	Belarus	373	Portugal	339
3	Gabon	347	Russia	326	Andorra	312
4	Germany	346	Haiti	326	Switzerland	280
5	Lithuania	343	Saint Lucia	315	Denmark	278
6	Poland	343	Guyana	302	Slovenia	276
7	Venezuela	333	Slovakia	293	Luxembourg	271
8	Ireland	313	Dominica	286	Croatia	254
9	Palau	306	Thailand	258	Italy	237
10	Romania	297	Cook Islands	254	Eqtr. Guinea	233

FIVETHIRTYEIGHT SOURCE: WORLD HEALTH ORGANIZATION

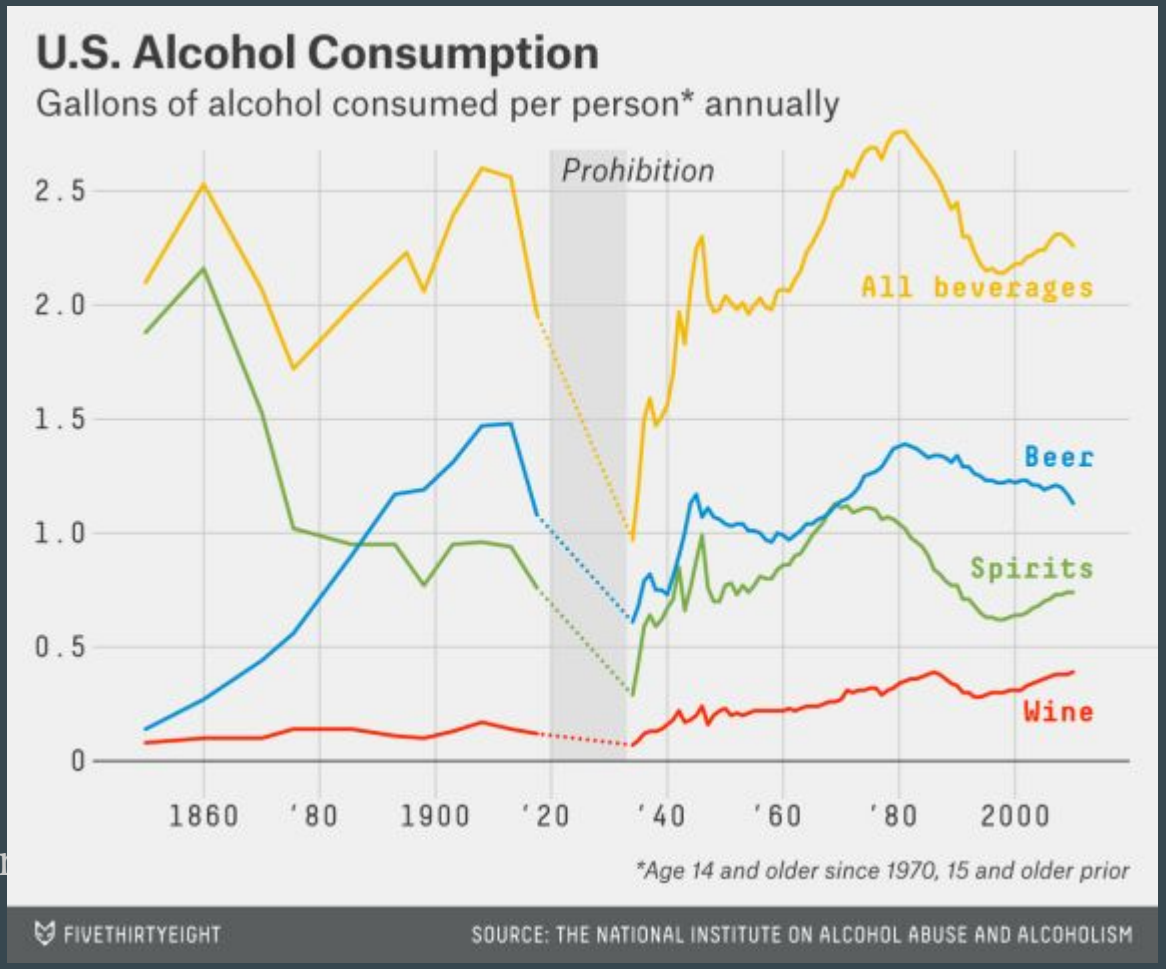
https://fivethirtyeight.com/features/dear-mona-followup-where-do-people-drink-the-most-beer-wine-and-spirits/

1o Exemplo: Consumo de Álcool Mundial (cont)

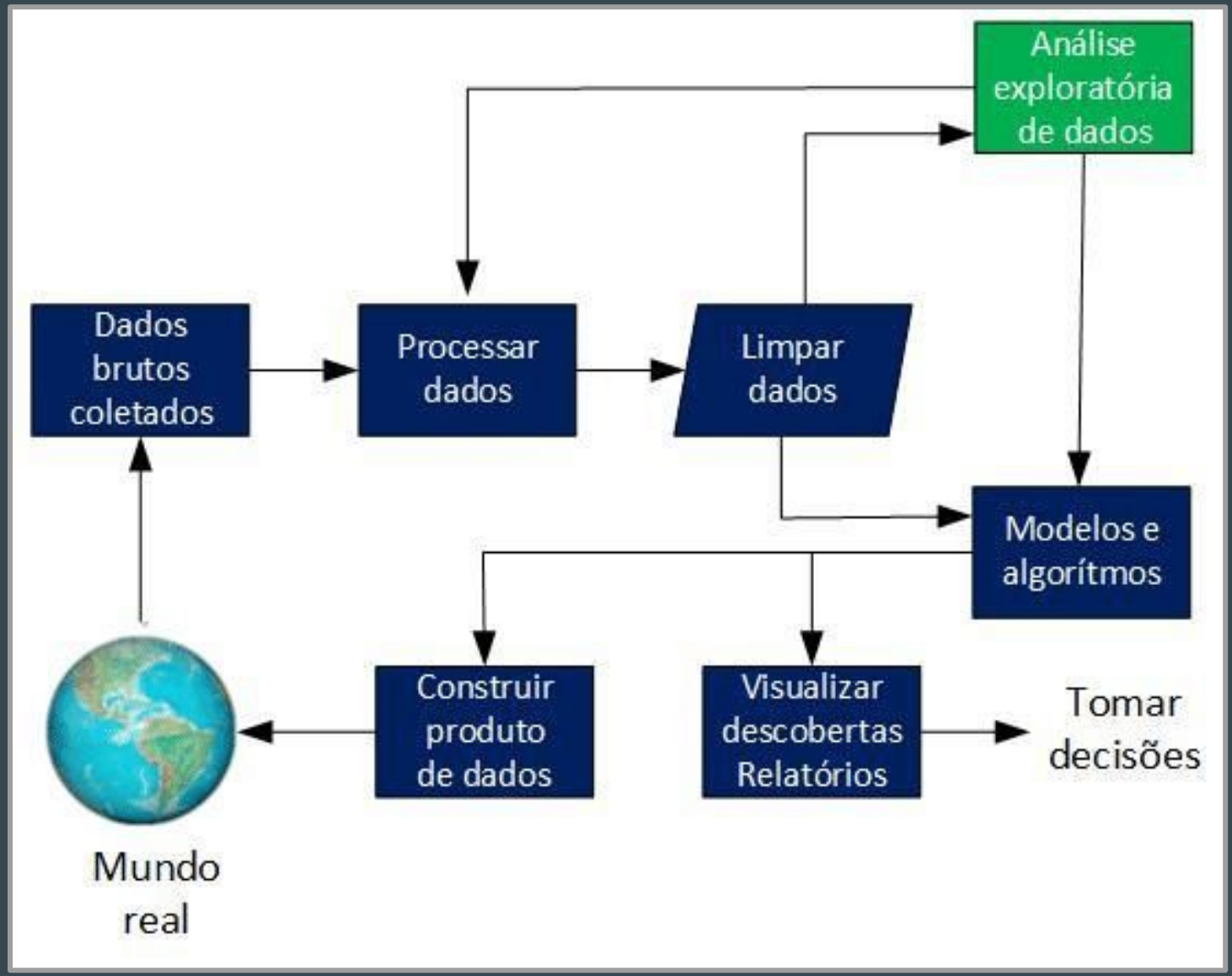
Dados oriundos de

FiveThirtyEight

https://fivethirtyeight.com/featur
ost-beer-wine-and-spirits/

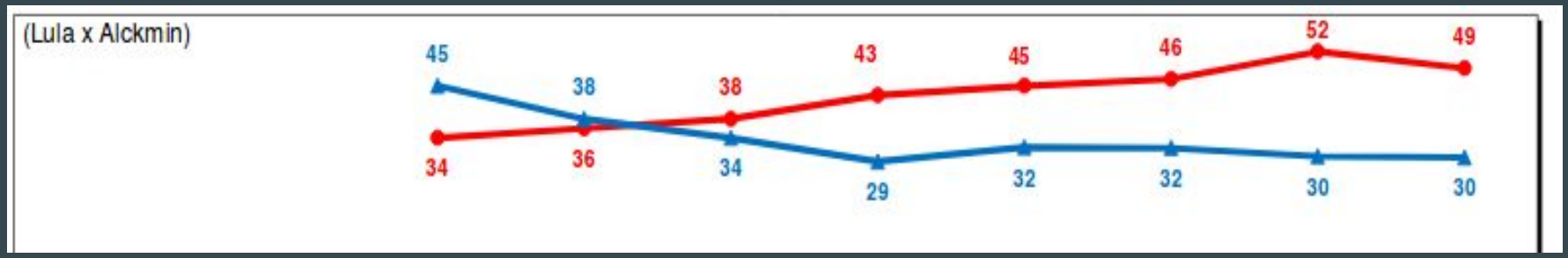


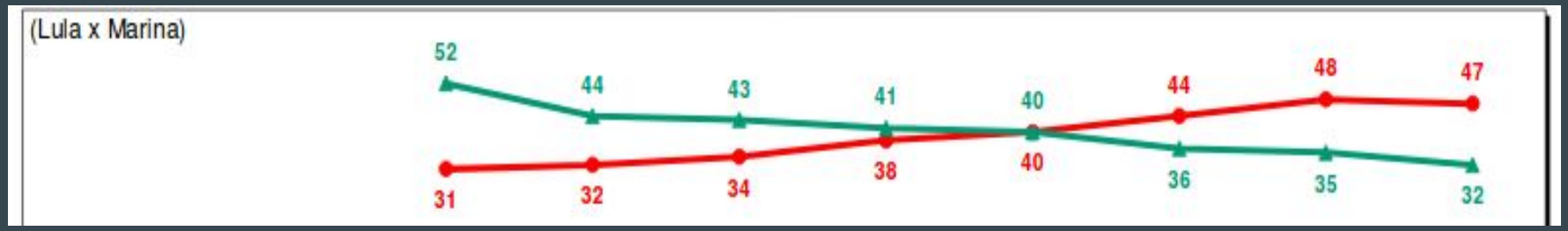
Processo da Ciência de Dados

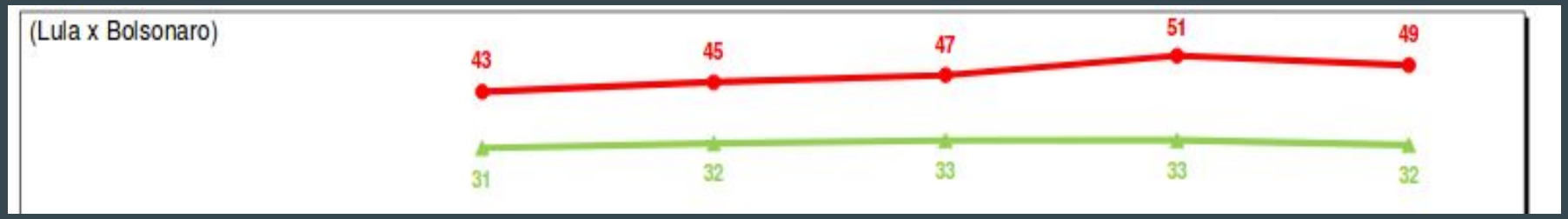


2o Exemplo - Popularidade dos pré-candidatos a presidente

Fonte Datafolha - Intenção de voto para presidente 2018 - 2º turno, período de 04/2017 a 01/2018

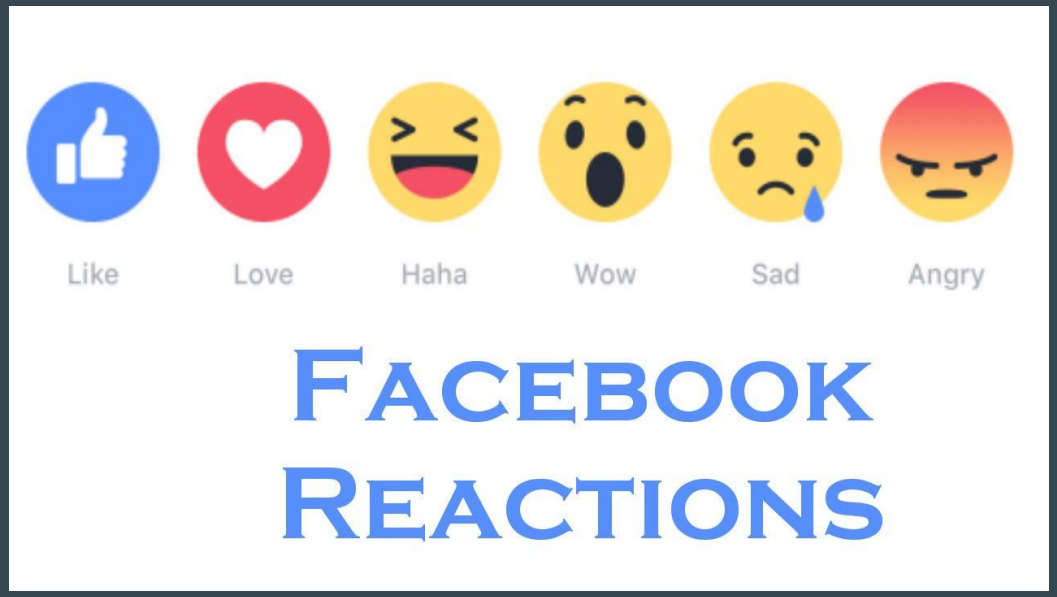






2o Exemplo - Popularidade dos pré-candidatos a presidente no Facebook

Vamos buscar as reações dos internautas aos posts do facebook dos pré-candidatos a presidente no último ano.



2o Exemplo - Popularidade dos pré-candidatos a presidente no Facebook (Extração de dados)

Recuperando dados do facebook:

API do facebook

https://developers.facebook.com/tools/explorer

IDs dos candidatos no facebook:

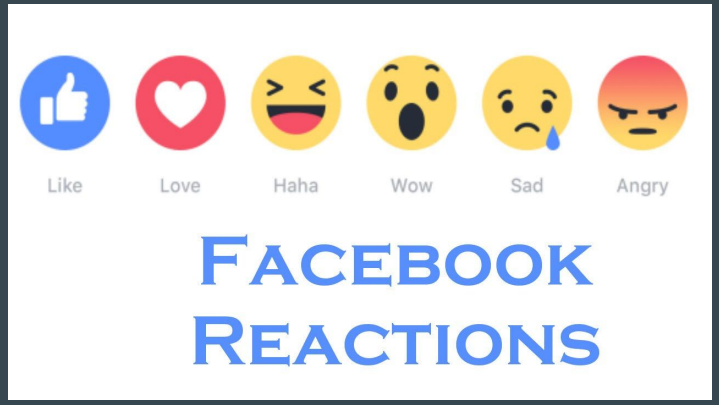
- Jair Bolsonaro : jairmessias.bolsonaro
- Luiz Inácio Lula : Lula
- João Doria : jdoriajr
- Marina Silva : marinasilva.oficial
- Ciro Gomes : cirogomesoficial
- Geraldo Alckmin : geraldoalckmin

2o Exemplo - Popularidade dos pré-candidatos a presidente no Facebook (Modelo)

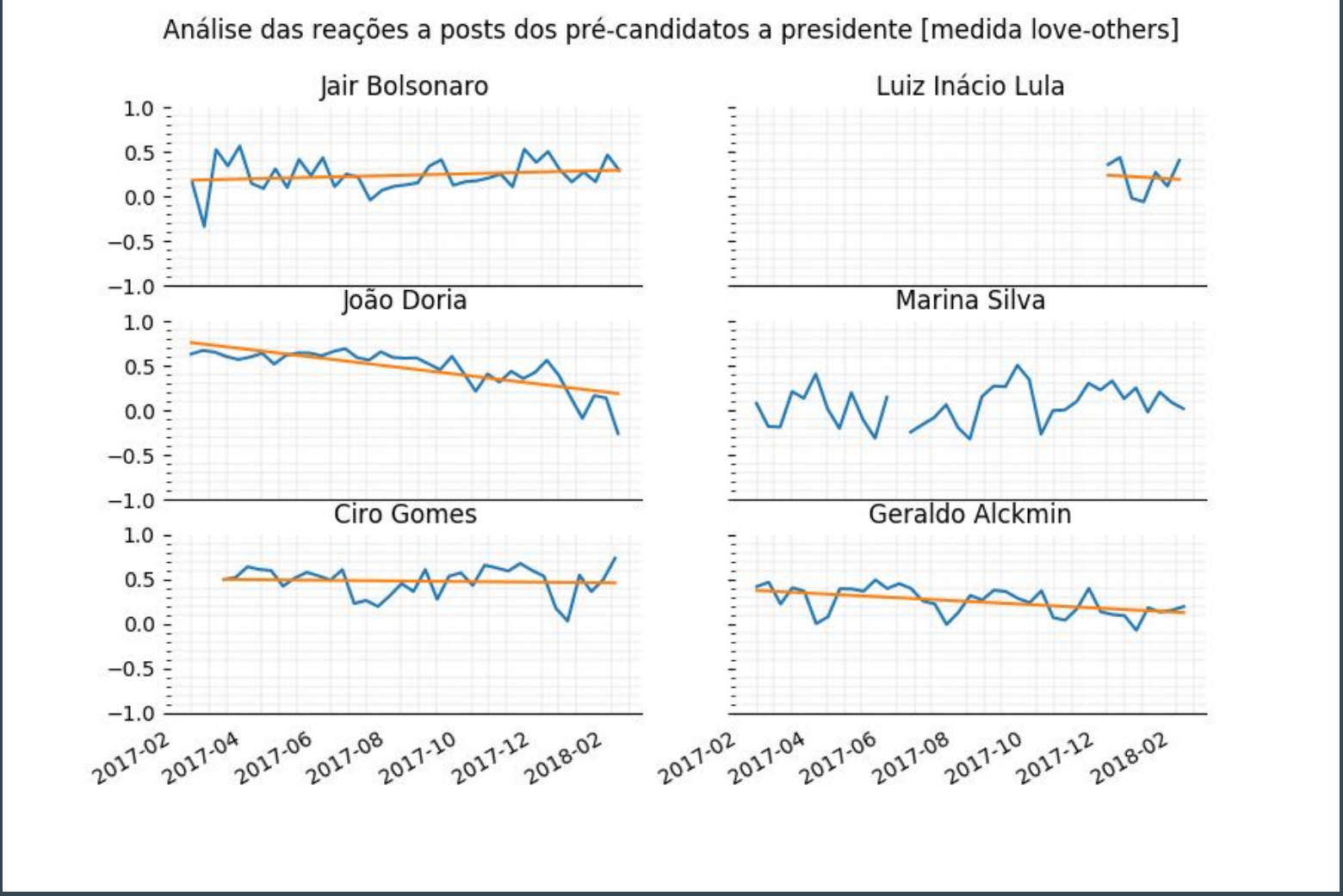
love-angry = (love-angry)/(love+angry)

love-others = (love-others)/(love+others)

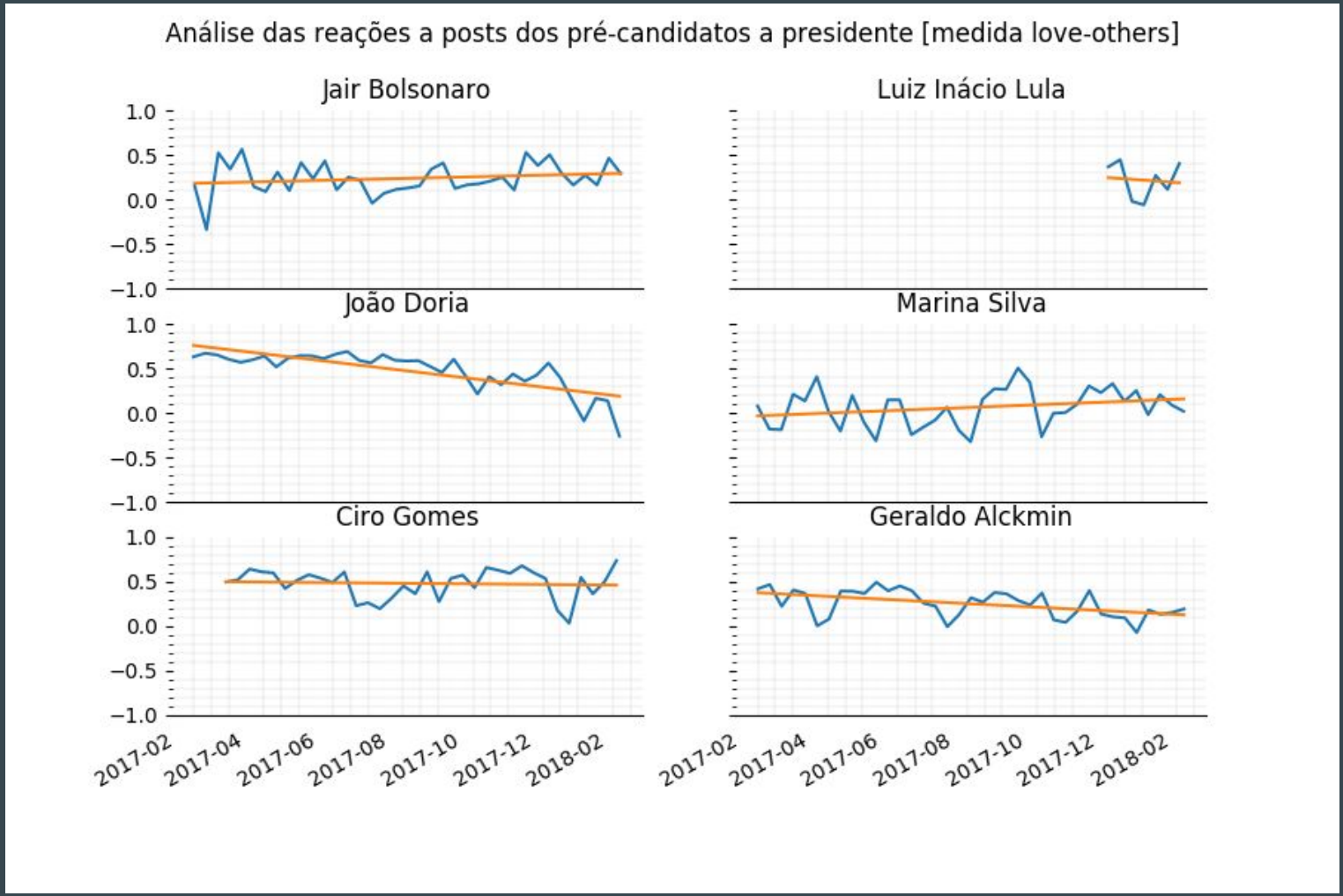
others : Haha, Wow, Sad, Angry



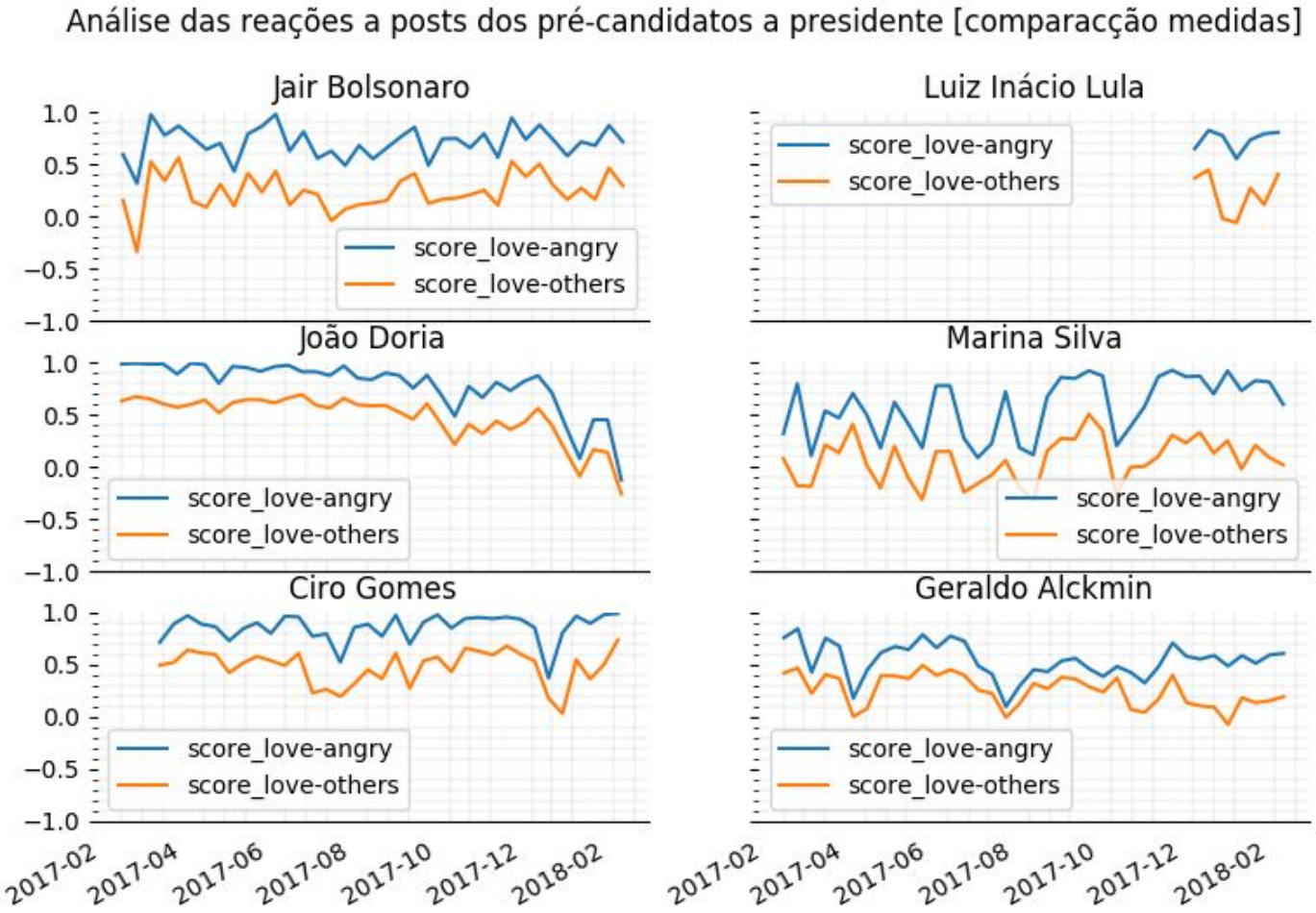
2o Exemplo - Popularidade dos pré-candidatos a presidente no Facebook (Apresentação tendência)



2o Exemplo - Popularidade dos pré-candidatos a presidente no Facebook (Apresentação tendência - corrigida)



2o Exemplo - Popularidade dos pré-candidatos a presidente no Facebook (Apresentação comparação índices)



Auto aprendizado

- Básico Python: Codecademy, Google's Python Class, Python Tutor, Python para zumbis
- Bibliotecas: pandas e numpy
- Ciência de Dados: cursos no udacity e cursera