



PUC Minas

IEC - Instituto de Educação Continuada
Pós-Graduação em Ciência dos Dados e Big Data

**Recuperação da Informação na Web
e em Redes Sociais**

**Análise exploratória de dados
coletados da rede social Twitter com
citações ao Bancoltau - Um dos
maiores banco do Brasil.**

Aluno: Rodolfo Sato Junior e Akilles Yuri Nakabori.

Professor: Zilton Cordeiro Jr.



PUC Minas

IEC - Instituto de Educação Continuada
Pós-Graduação em Ciência dos Dados e Big Data

Projeto Final

Análise exploratória de dados coletados da rede social Twitter com citações ao Bancoltau - Um dos maiores banco do Brasil.

Trabalho apresentado ao Instituto de Educação Continuada (IEC) da pós-graduação em Ciência dos Dados e Big Data da PUC Minas, como requisito parcial para a obtenção de créditos na disciplina de Recuperação da Informação na Web e em Redes Sociais.

Aluno: Rodolfo Sato Junior e Akilles Yuri Nakabori.

Professor: Zilton Cordeiro Jr.

Conteúdo

1	Resumo	1
2	Introdução	2
3	Descrição das Atividades	3
3.1	Coleta	3
3.2	Armazenamento	3
3.3	Processamento e Análise	4
4	Análise dos Resultados	6
4.1	Estado dos usuarios que mais falam sobre o Itau	6
4.2	Dias com maior frequencia de Tweets	7
4.3	Frequência das palavras com polaridade Negativa	10
4.4	Frequência das palavras com polaridade Positiva	11

1 Resumo

Este projeto foi feito utilizando dados retirado de uma das maiores rede social usadas hoje mundialmente, o “Twitter” que hoje representa uma grande fonte de dados por muitos dev e cientistas de dados, iremos pesquisar utilizando a palavra chave “Itau” e filtrar somente dados em Portugues do Brasil (pt-br), iremos filtrar dados somente do dia 10/07/2021 até o dia 18/07/2021 ou seja uma amostra de 9 (nove) dias, contendo um total de 7670 Tweets, e também com esses dados iremos retirar palavras-chaves positivas ou negativas, que vista por cima podemos perceber diversas coisas com essas palavras-chaves sendo possivel ver elogios o criticas se vistas por um especialista, também veremos de qual cidade e estado são os usuarios da rede social que comentam falando sobre o banco Itau para podermos ter uma noção em quais cidades e estados essas pessoas pertencem, sendo assim podemos também fazer um calculo para podermos descobrir aproximadamente aonde o Banco Itau foi mais falado nesses 9 dias de Tweets.

2 Introdução

O Itaú Unibanco foi fundado em 1924 por João Moreira Salles e Alfredo Egydio de Souza Aranha, em Poços de Caldas, Minas Gerais. Na época, o banco Itaú se chamava Banco Central de Crédito, enquanto o Unibanco era conhecido como Casa Moreira Salles. Nos anos 1960, as duas instituições financeiras se uniram.

Entre os segmentos de operação do Itaú Unibanco estão o bancário, investimentos, seguros, previdência, capitalização e consórcio. Para dar conta da demanda, a instituição conta com cerca de 5 mil agências no Brasil e no exterior e 26 mil caixas eletrônicos e pontos de atendimento. Além disso, a empresa está presente em 21 países e é responsável por 60 milhões de contas-correntes.

O Twitter é uma rede social e um serviço de micro blog para comunicação em tempo real usado por milhões de pessoas e organizações. Usuários do Twitter ficam interconectados ao publicar atualizações ao site, conhecidas como “Tweets”, para compartilhar, trocar e descobrir informações. Tweets consistem em 140 ou menos caracteres e podem contar ideias e vários tipos de informação, como fotos, vídeos e links para artigos. Usuários podem acessar essas mensagens no Twitter pela Web ou por um dispositivo móvel com conexão a Internet.

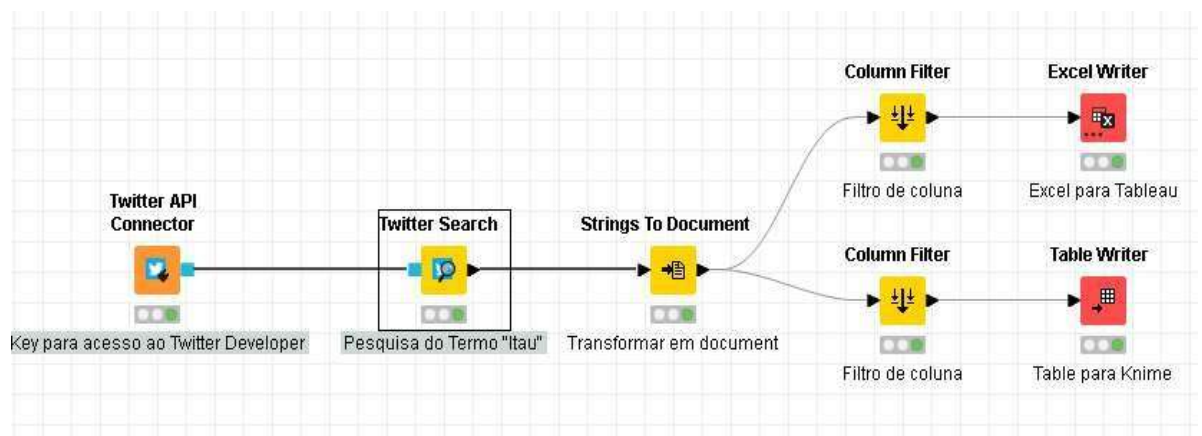
As ferramentas que iremos utilizar é a plataforma KNIME Analytics que é um software de código aberto para todas as suas necessidades de dados que contém todas as principais técnicas de data wrangling e aprendizado de máquina baseados em programação visual e também iremos usar um pouco de Tableau que é um software destinado ao segmento de analytics (análises de dados)

3 Descrição das Atividades

Nesta seção são descritas as etapas de trabalho desempenhadas para alcance dos resultados, sendo elas: coleta de dados, armazenamento, processamento e análise.

3.1 Coleta

Inicialmente iremos pegar informações do Twitter, que contenham o termo “Itau” da data de 10/07/2021 até o dia 18/07/2021 para podermos ter uma amostra de 9 (nove) dias, sendo assim ter bastante informações para podermos utilizar para análise:



3.2 Armazenamento

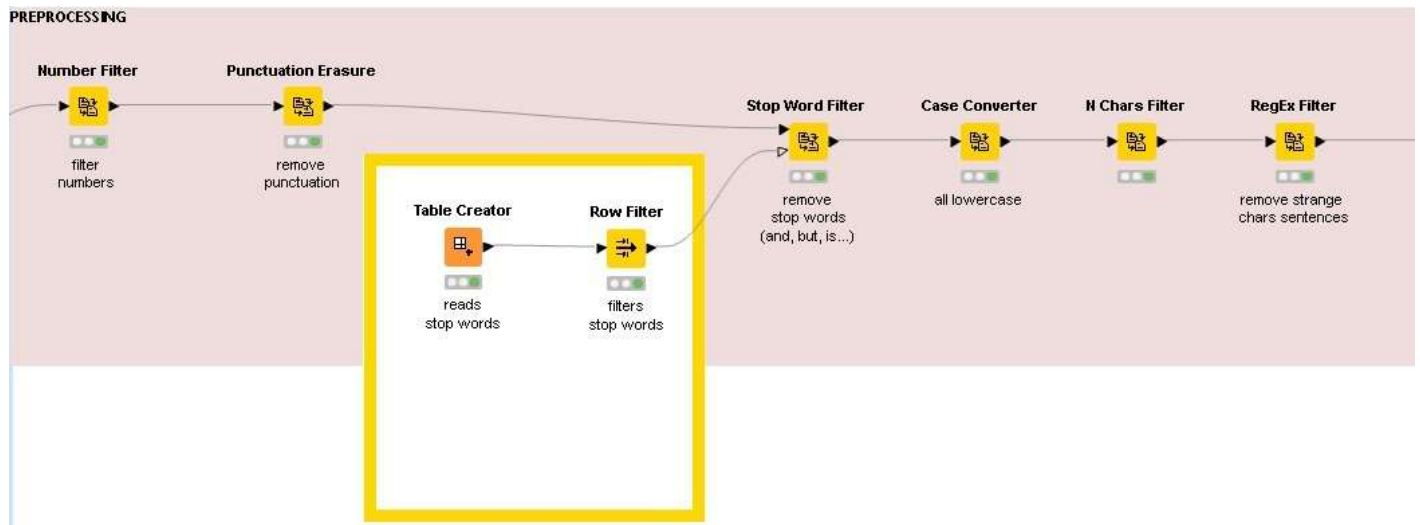
Para armazenamento dos dados coletados foi utilizado a propria maquina local, utilizei dois tipos de arquivos, sendo .table para gerar uma tabela para uso do fluxo de text mining no knime, e arquivo excel para se utilizar na manipulação dos dados no tableau.

3.3 Processamento e Análise

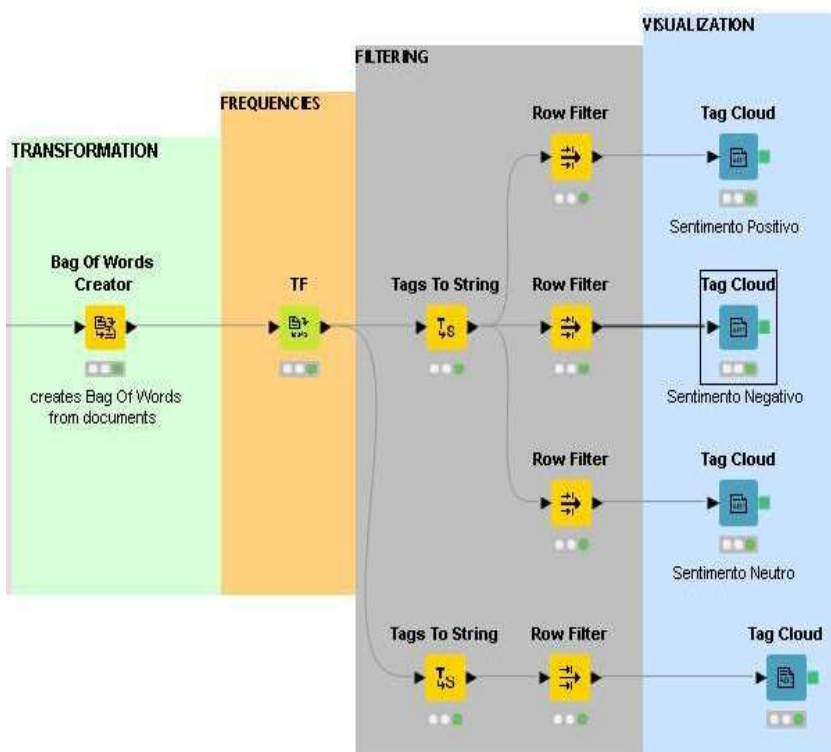
Essa etapa teve como objetivo processar e analisar todos os tweets coletados e armazenados conforme descrição das etapas anteriores, a fim de responder às seguintes questões:



Nesse processo o fluxo de Text mining começa utilizando a base antes gerada pelo twitter no passo 3.1, após termos esses dados, passamos ele por esse fluxo para no fim gerar os resultados desejados que nesse caso são as palavras-chaves positivas ou negativas, o fluxo segue inicialmente recebendo o arquivo da máquina local, de um diretório, inicialmente passa pela leitura e criação de documentos, depois ele passa pela parte de enriquecimento, que tem a função de enriquecer a base inicial com informações que iremos usar no fluxo mais pra frente.



Nessa parte do fluxo fazemos o pré processamento que como o nome diz é um passo antes de começar a processar os resultados dos dados.



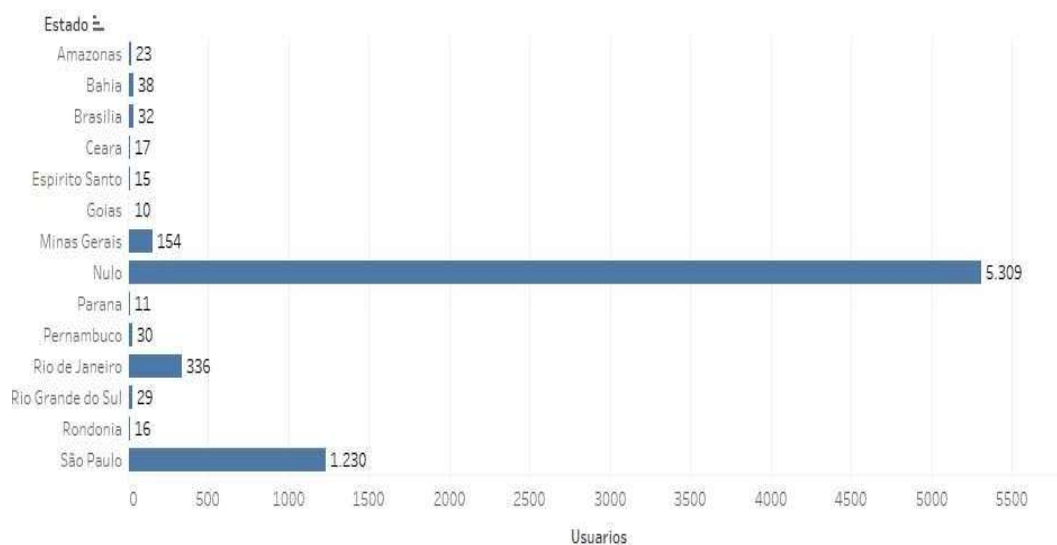
Nessa parte do fluxo temos os ultimos 4 processos, sendo eles a transformação, que tem a função de transformar algumas palavras-chaves em outra para dizer se essas palavras são palavras positivas, negativas ou neutras, apos isso temos a frequencia, que juntas todos os temos usados nos tweets e faz a função de frequencias para poder descobrir as palavras mais usadas, apos isso temos o filtro que tem a função de separar se o termo é positivo, negativo, neutro, ou até mesmo uma palavra que corresponde a uma determinada empresa, que no caso é o ultimo filtro, por fim temos a parte de visualização, que tem a função de visualizar os resultados.

4 Análise dos Resultados

A figura 4 apresenta os dados utilizados para análise e produção de resultados, após conclusão das etapas de coleta, armazenamento e processamento, sendo:

4.1 Estado dos usuarios que mais falam sobre o Itau

O objetivo desta etapa foi identificar qual a distribuição entre os estados brasileiros para descobrir quais estados comentam mais sobre o Itau:



Podemos ver que a maior parte do publico que falaram sobre o Itau coloca seu endereço como nulo, pois a rede social não necessariamente te obriga a colocar o local de vivencia ou nascimento, apos isso podemos ver que a maioria do publico que comentou sobre o Itau nesses 9 dias são de São Paulo e seguidamente pelo Rio de Janeiro...

4.2 Dias com maior frequencia de Tweets

O objetivo desta etapa foi identificar quais dia da semana houve a maior quantidade de Tweets e por qual motivo aparente:

Time	
2021-07-10 (Sab)	341
2021-07-11 (Dom)	821
2021-07-12 (Seg)	4.486
2021-07-13 (Ter)	1.298
2021-07-14 (Qua)	625
2021-07-17 (Qui)	2
2021-07-18 (Sex)	97

Podemos ver que no dia 12/07/2021 houve um pico de tweets citando o Itau, indo a fundo descobrimos que nesse dia houve um “protesto” online por conta da reclamação de alguns funcionarios que aparentemente se sentem sobrecarregados pela que parece metas abusivas, e varios outros problemas, alguns exemplos:



Carlos Damarindo @CDamarindo · 12 de jul

#QueVergonhaltaú na fogueira da vaidade e pela disputa interna dos executivos do bco fica o bancário com a insegurança medo da demissão metas desumanas que adoecem e as publicidades enganosas.



5

60

217





Osmar Carvalho @osmarpaulinode1 · 12 de jul

...

O Itaú continua cobrando metas impossíveis de se atingir, totalmente fora da realidade.

#QueVergonhaItaú



8

70

353



Como se não fosse somente isso no mesmo dia, também houve uma outra Hashtag que subiram contra o apresentador de TV Sikeira Jr, pelo fato de ele ter falado coisas contrárias a comunidade LGBT, pelo fato do Itaú patrocinar o mesmo, o público pedia pela desmonetização do mesmo:



Sleeping Giants Brasil @slpng_giants_pt · 13 de jul

...

🚀 O @itau É A MAIS EMPRESA QUE RETIRA O PATROCÍNIO DO SIKÉRA JR!

Agradecemos a todos os seguidores que conscientizaram a empresa e também parabenizamos a Itaú pelo posicionamento! 🏳️🌈

#DesmonetizaSikera



Itaú (em 🏠) @itau · 12 de jul

Em resposta a @slpng_giants_pt

Olá, @slpng_giants_pt. Estamos atentos por aqui! Não toleramos discursos de ódio e conteúdos que geram desinformação, por isso, reforçamos ainda mais nossos controles de exposição de marca e seguimos acompanhando de perto este movimento.

67

236

2,5 mil





Sleeping Giants Brasil @slpng_giants_pt · 12 de jul

...

Olá @itau, tudo bem? Um anúncio #FeitoCom Sikera Jr não é nada legal. Afinal, tem lugares melhores para investir sua publicidade do que um canal no @YouTubeBrasil que propaga a lgbtfobia e preconceito, né? Pfv, #DesmonetizaSikera 🏳️🌈

...contas em dia,
sua vida financeira tranquila.

Simule agora

Pular anúncios ▶

Anúncio · 1 de 2 · 0:03 ⓘ

 **Banco Itaú**
Anúncio Instalado **ABRIR APLICATIVO** ▼

#GALOSIMPORTAM ABSURDO BRASIL! ▼

10 mil visualizações · há 1 hora

 2,3 mil  Não gostei  Compartilhar  Download  Salvar

Tirando isso de base, podemos ver que a media dos tweets serem divergentes se da pelo fato desses fatores externos que mudam a média de tweets do Itaú.

4.3 Frequência das palavras com polaridade Negativa

O objetivo desta etapa foi identificar quais as palavras de polaridade negativas, vendo essas palavras, podemos entender alguns problemas na empresa, vendo apenas por cima dessas palavras:



Vimos que a palavra “Vergonha” esta em maior destaque e se dá pelo fato da Hashtag que subiram #QueVergonha!taú, mostrando que o publico em geral estao insastifeitos com algo relacionado a empresa, podemos ver a palavra “propaganda” se da ao fato da retirada das propagandas do apresentador Sikeira Jr que é pelo fato da Hashtag #DesmonetizaSikera.

4.4 Frequência das palavras com polaridade Positiva

O objetivo desta etapa foi identificar quais as palavras de polaridade positivas, vendo essas palavras, podemos entender alguns pontos fortes ou elogios na empresa pelos usuarios, vendo apenas por cima dessas palavras:

