



Documentação & Manual de Uso

Equipe Proconfie
<http://proconfie.vod.dcc.ufmg.br>

Sumário

1	Motivação	4
2	Processamento do Arquivo de Reclamações	5
3	Descritores de Risco das Empresas	6
4	Modelo de Risco para Empresas	8
5	Sobre o Aplicativo e Requisitos Mínimos	9
6	Como Usar	11
7	Equipe Proconfie	18

1 Motivação

A principal motivação para o desenvolvimento do aplicativo *Proconfie* é prover o cidadão brasileiro com uma ferramenta capaz de acessar, de forma rápida e inteligente, a base de dados do Procon. A ideia é, a partir de poucos cliques, informar o quão comprometida é uma empresa com o consumidor. Essa informação é mostrada de forma clara e intuitiva através de um valor que chamamos de *risco*, que indica o quão propensa é essa empresa a causar transtornos para o consumidor. O *risco* da empresa varia de 1, que representa risco baixo, a 10, que representa risco altíssimo.

Outra motivação para o desenvolvimento do *Proconfie* é nossa crença que somente a visualização simples das informações contidas na base de dados não é suficiente para que o cidadão tenha conhecimento sobre o comportamento e comprometimento das empresas para com a sociedade brasileira. Mostraremos neste documento que o comportamento das empresas depende de diversos fatores, como a categoria que ela se encontra e o ano de atuação. Assim, o *risco* que calculamos para cada empresa não somente considera o número de reclamações que essa recebeu, mas também o percentual dentro da sua categoria, o número de reclamações que foram resolvidas, o desempenho ao longo do tempo, entre outros fatores.

Dentre as funcionalidades do *Proconfie*, está um ranking das empresas dentro das suas respectivas categorias. Esse ranking pode ajudar a sociedade brasileira de duas maneiras fundamentais. Primeiro, ele auxilia o consumidor na tomada de decisão quando esse for contratar um serviço ou comprar um produto, informando quais são as empresas que mais podem lhe causar transtornos. Assim, a partir do *Proconfie*, o consumidor pode evitar aquelas empresas que menos comprometimento têm para com ele. Segundo, o *Proconfie* pode auxiliar as próprias empresas, que terão um *feedback* imediato das suas situações perante outras empresas da mesma categoria. Caso uma empresa esteja entre as empresas de maior risco dentro de uma categoria, ela deverá, a fim de preservar a sua clientela e sua fatia de mercado, tomar medidas urgentes e significativas na direção de melhorar os serviços prestados. Quando as empresas tomam tais atitudes, quem sai ganhando é o consumidor.

2 Processamento do Arquivo de Reclamações

Para o desenvolvimento do aplicativo, foram fornecidos pelos organizadores três arquivos contendo informações sobre as reclamações registradas no Procon até o ano de 2011, são eles

- `reclamacoes-fundamentadas-sindec-2009.csv`,
- `reclamacoes-fundamentadas-sindec-2010.csv` e
- `reclamacoes-fundamentadas-sindec-2011.csv`.

Assim, o primeiro passo que tomamos foi a criação de um programa processador desses arquivos a fim de popular uma base de dados relacional, que julgamos ser mais apropriada para o desenvolvimento do aplicativo. Esse programa, que foi desenvolvido na linguagem *Python*, está disponível publicamente juntamente com o código do aplicativo, no arquivo `ProconCSVParser.py`.

Ao todo foram criadas 8 tabelas a partir dos arquivos de reclamações:

- `assuntos`,
- `idades`,
- `estados`,
- `categorias`,
- `problemas`,
- `radicais_cnpj`,
- `empresas` e
- a própria tabela de `reclamações`.

Esta parte do processamento não envolve nenhuma inteligência em particular, servindo apenas para separar as informações contidas nos arquivos e para colocar as mesmas em suas devidas tabelas, fazendo o uso de chaves estrangeiras quando for o caso. Em suma, a tabela `reclamações` contém o mesmo número de linhas que os três arquivos de reclamações fornecidos, mas sem qualquer informações redundante, que estão armazenadas nas outras tabelas.

Vale ressaltar que foram encontrados muitos campos nos arquivos sem o devido preenchimento. Além disso, muitas informações estão no formato puramente descritivo, como os campos `idade` e `estado`, por exemplo. Assim, uma porção significativa dos dados registrados nas tabelas contém valores criados por nós, sejam eles códigos para referenciar os campos descritivos dos arquivos, ou `-1`, quando um campo não foi preenchido.

3 Descritores de Risco das Empresas

A fim de analisar quais são as empresas que oferecem um maior risco para o consumidor, extraímos da tabela de reclamações uma série de descritores que indicam o potencial que uma empresa tem para gerar uma reclamação no Procon. Tais descritores de risco de uma empresa são:

- $F0$: número de reclamações anteriores à 2009;
- $F1$: número de reclamações no ano de 2009;
- $F2$: número de reclamações no ano de 2010;
- $F3$: número de reclamações no primeiro semestre de 2011;
- $F4$: número de reclamações no segundo semestre de 2011;
- $F5$: percentual de reclamações considerando somente as reclamações dentro da categoria da empresa;
- $F6$: percentual de reclamações no ano de 2009 considerando somente as reclamações dentro da categoria da empresa;
- $F7$: percentual de reclamações no ano de 2010 considerando somente as reclamações dentro da categoria da empresa;
- $F8$: percentual de reclamações no primeiro semestre de 2011 considerando somente as reclamações dentro da categoria da empresa;
- $F9$: percentual de reclamações no segundo semestre de 2011 considerando somente as reclamações dentro da categoria da empresa;
- $F10$: inclinação da reta da regressão linear considerando os pontos $(1, F5)$, $(2, F6)$, $(3, F7)$, $(4, 2 \times F8)$, $(5, 2 \times F9)$;
- $F11$: número normalizado¹ de reclamações de uma empresa na sua categoria;
- $F12$: mediana dos tempos que demoram para uma reclamação ser resolvida;
- $F13$: média dos tempos que demoram para uma reclamação ser resolvida;
- $F14$: *skewness* dos tempos que demoram para uma reclamação ser resolvida;
- $F15$: data da primeira reclamação;

¹dividido pelo maior número de reclamações dentro da sua categoria.

- $F16$: número total de reclamações resolvidas;
- $F17$: proporção de reclamações do segundo semestre de 2011 que foram resolvidas;
- $F18$: proporção de reclamações do primeiro semestre de 2011 que foram resolvidas;
- $F19$: proporção de reclamações do ano de 2009 que foram resolvidas;
- $F20$: proporção de reclamações do ano de 2010 que foram resolvidas;
- $F10$: inclinação da reta da regressão linear considerando os pontos $(1, F17)$, $(2, F18)$, $(3, F19)$, $(4, F20)$;

Os descritores $F0$ a $F4$ servem como uma medida simples para verificar a evolução do número de reclamações contra a empresa ao longo do tempo. Os descritores $F5$ a $F9$ também servem para analisar a evolução do número de reclamações, mas consideram o universo da categoria da empresa. Isso é importante porque, como mostrado na Figura 1, há várias categorias de baixo risco, como a categoria “COMÉRCIO VAREJISTA DE OBJETOS DE ARTE”, que possui somente uma reclamação, e outras de alto risco, como “BANCOS MÚLTIPLOS, COM CARTEIRA COMERCIAL”, com 24648 reclamações. Assim, é importante não comparar empresas que fazem parte de negócios intrinsecamente diferentes, que atingem mercados que atuam sobre regras diferentes. O descritor $F10$ é a inclinação da reta formada pelos pontos dos descritores anteriores, ou seja, $(1, F5)$, $(2, F6)$, $(3, F7)$, $(4, 2 \times F8)$, $(5, 2 \times F9)$. Multiplicamos os dois últimos valores por 2 porque eles consideram um intervalo de 6 meses, não de 1 ano como os outros. Quando $F10$ é positivo (negativo), então o número de reclamações contra essa empresa está crescendo (diminuindo) dentro da própria categoria, indicando que o risco dessa empresa também está aumentando (diminuindo).

Por fim, é interessante explicar os descritores de risco $F17$ a $F21$, que como as descritas anteriormente, também alimentarão o nosso modelo de risco para as empresas. Os descritores $F17$ a $F20$ indicam a evolução da proporção de reclamações que foram resolvidas contra aquela empresa. Assim, quando $F21$ é positivo (negativo), então a proporção de reclamações resolvidas contra a empresa está crescendo (diminuindo), indicando que a empresa está (não está) tomando medidas necessárias para resolver os problemas dos seus clientes e melhorando (piorando) os seus serviços.

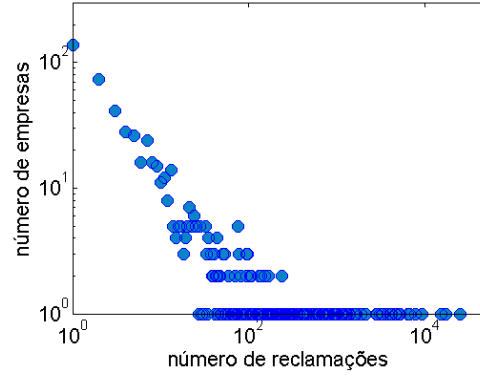


Figura 1: Número de reclamações registradas em cada categoria de empresa.

4 Modelo de Risco para Empresas

Dados os descritores de risco listados na seção anterior, propomos agora um modelo que agrega alguns desses descritores a fim de indicar qual é o risco, em uma escala simples de 1 (risco mínimo) a 10 (risco máximo), de ter problemas com uma empresa ao usufruir de seus serviços. A escala inicia-se com 1 porque todas as empresas cadastradas têm, pelo menos, uma reclamação registrada no Procon. Empresas que não foram registradas na base do Procon foram “premiadas” com risco 0.

O risco da empresa é calculado em algumas etapas. Primeiro, calcula-se a média ponderada dos percentuais de reclamações dentro da categoria (descritores $F6$ a $F9$), que chamamos de M_1 , e a média ponderada das proporções de reclamações resolvidas (descritores $F17$ a $F20$), que chamamos de M_2 . Faz-se a média ponderada para valorizar mais o comportamento mais recente da empresa. Os valores M_1 e M_2 são calculados da seguinte maneira:

$$M_1 = \frac{4 \times F9 + 3 \times F8 + 2 \times F7 + F6}{10}$$

$$M_2 = 1 - \frac{4 \times F17 + 3 \times F18 + 2 \times F19 + F20}{10}$$

Note que a média calculada em M_2 é subtraída de 1 para que M_2 tenha uma correlação positiva com o risco da empresa, assim como M_1 . Assim, calculamos uma terceira média ponderada M_3 , considerando os valores de M_1 e M_2 , a fim de dar mais peso para o percentual de reclamações recebidas dentro da categoria. Isso é feito da seguinte maneira:

$$M_3 = \frac{3 \times M_1 + M_2}{4}$$

Depois, considerando o valor de M_3 e $F10$, calcula-se um valor r entre 0 a 1 para o risco da empresa, segundo as seguintes equações:

$$\begin{aligned} r_1 &= 1 - e^{-0.1 \times M_3} \\ r_2 &= \frac{1}{1 + e^{-F10}} \\ r &= r_1 \times r_2 \end{aligned}$$

As transformações feitas em M_3 e $F10$ são feitas para mudar a escala dos mesmos. Observe na Figura 2 como esses valores variam.

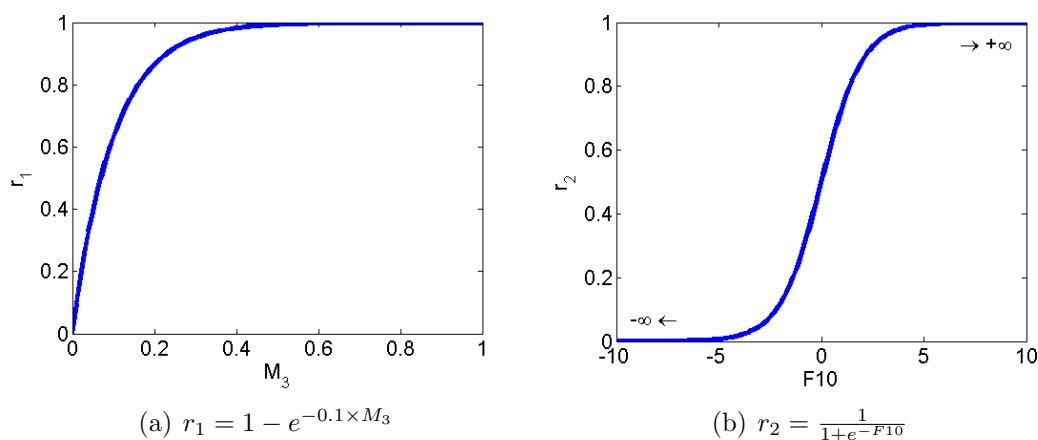


Figura 2: Mudança de escala dos valores M_3 e $F10$.

Por fim, o risco R da empresa é calculado através de uma simples mudança de escala:

$$R = 1 + 9 \times r$$

Observe na Figura 3 o histograma do número de empresas com determinado risco R . Note que a grande maioria das empresas apresenta um risco baixo, entre 1 e 2.

5 Sobre o Aplicativo e Requisitos Mínimos

O PROCONFIE é um aplicativo web open source de avaliação de empresas. Sendo um aplicativo web, seu uso é muito simples e pode ser realizado de diversos dispositivos. Basta apontar o seu navegador preferido para <http://proconfie.vod.dcc.ufmg.br> e pronto! O aplicativo foi desenvolvido usando as seguintes ferramentas:

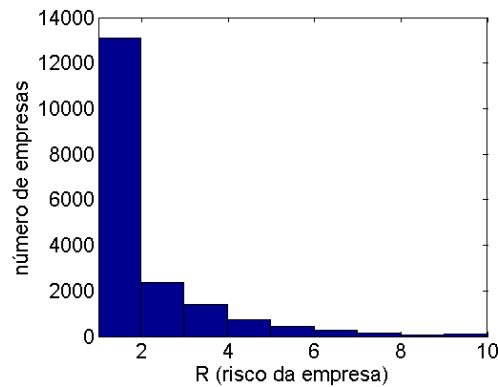


Figura 3: Histograma do número de empresas com determinado risco R .

- Ruby - <http://ruby.org/>
- Rails - <http://rubyonrails.org/>
- Passenger - <http://www.phusionpassenger.com/>
- Nginx - <http://nginx.org/>
- Foundation Front-end framework 3 - <http://foundation.zurb.com/>
- Morris JS - <http://www.oesmith.co.uk/morris.js/>
- JQuery - <http://jquery.com>
- MySQL - <http://mysql.org>

O desenvolvimento do mesmo como um aplicativo web usando tais ferramentas foi feito com os principais objetivos de: (1) facilidade de manutenção do código; e (2) portabilidade. Por ser um aplicativo web o mesmo pode ser acessado através de um extenso leque de dispositivos. Não só isso, como o usuário tem uma interface unificada independente do dispositivo sendo utilizado. Em especial, testamos o funcionamento do PROCONFIE nos seguintes dispositivos e sistemas operacionais:

- Firefox nos Sistemas Operacionais Windows e Linux
- Google Chrome nos Sistemas Operacionais Windows e Linux
- Internet Explorer no Sistema Operacional Windows

Além dos dispositivos acima para PCs, o sistema também foi testado nos seguintes sistemas operacionais de smartphones e tablets:

- Google Chrome para Android Ice Cream Sandwich
- Opera para Android Gingerbread e Android Ice Cream Sandwich
- Safari para dispositivos com Sistema Operacional IOS (iPhone/iPod/iPad)

6 Como Usar

Para iniciar o uso do sistema basta abrir seu navegador preferido² e ir para o seguinte website: <http://proconfie.vod.dcc.ufmg.br>. Após a carga, o usuário se encontrará com a interface representada na Figura 4 a sua frente. Tal figura, assim como as outras dessa seção, apresenta a interface usando um navegador desktop (esquerda) e um navegador de dispositivo móvel (direita).



Figura 4: Tela Inicial do PROCONFIE

Nesse momento o usuário é apresentado com uma pequena descrição do aplicativo, além de duas possíveis formas de interação com o aplicativo. Inicialmente, o usuário pode escolher em qual estado se localiza a empresa qual ele deseja saber mais informação. Após isso o usuário pode digitar o nome completo ou parte do nome da empresa na caixa de busca. A ação de selecionar estados está representada na Figura 5.

O PROCONFIE contém um avançado sistema de busca que faz uso de auto completação para ajudar o usuário. Caso o usuário selecionou um estado no passo anterior, a busca será realizada apenas para empresas daquele estado. Caso contrário

²Na Seção 5 apresentamos uma lista de navegadores testados

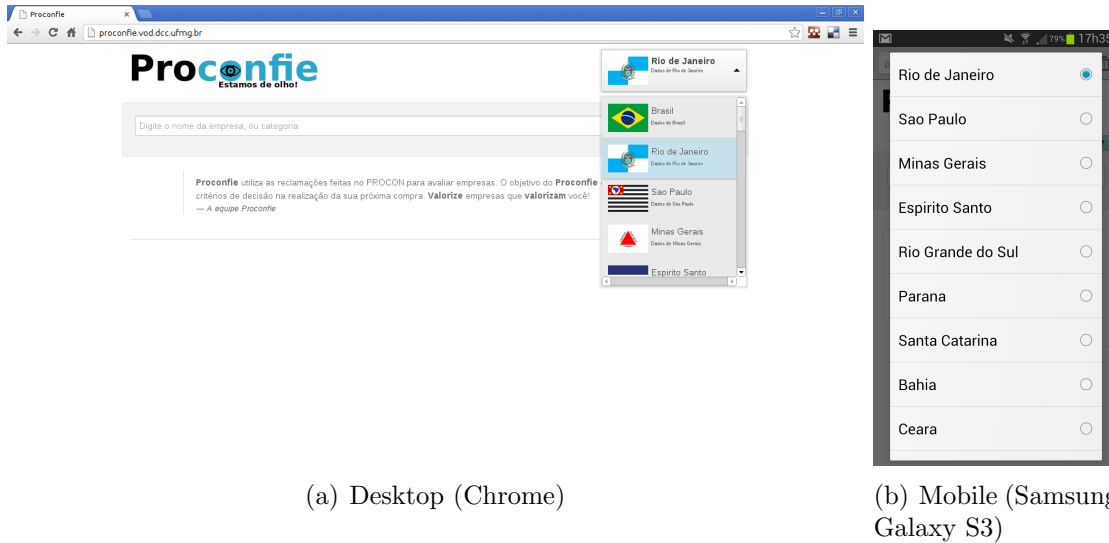


Figura 5: Seleção de Estados

será realizada em empresa do todo o Brasil. Nesse momento o usuário pode digitar sua busca parcial e a funcionalidade de auto completar entrará em ação assim como mostrada na Figura 6. Ao ser apresentado com a lista de possíveis resultados o usuário deve então clicar (desktop) ou tocar no nome (mobile) da empresa que deseja ver mais informações.

Ao escolher uma empresa o usuário será direcionado para a página da mesma. Um exemplo de tal página é apresentado na Figura 7. Nesse momento uma caixa descritiva de algumas informações sobre a empresa é fornecida. **É importante a caixa onde contém o risco de uma empresa.** Esta caixa contém uma das informações que é um diferencial da nossa aplicação. Através do uso de algoritmos de mineração de dados e aprendizado de máquina, desenvolvemos um ranking inteligente de empresas em uma mesma categoria. Neste ranking, indicamos para o usuário o **quão ruim** é uma empresa. Note que isto é diferente de indicar empresas boas. Os dados do Procon são de reclamações e não elogios. A ferramenta PROCONFIE serve então o propósito de prover mais informação sobre a ineficiência de empresas em relação ao Procon. Assim, o usuário pode escolher quais empresas evitar!

Ainda na página apresentada na Figura 7, o usuário pode ver a nota da empresa definida pelo nosso ranking inteligente. Essa nota é chamada de *risco*. Ao colocar o mouse sobre a palavra *Risco* (desktop) ou realizar a *gesture* de segurar o dedo em cima da palavra (móvel), uma *tooltip*, apresentada na forma de caixa de texto, vai apresentar para o usuário detalhes adicionais sobre como é computado o risco de uma empresa. Tal ação está representada na Figura 8.



Figura 6: Funcionalidade de Busca e Auto Completar

Tal tela inicial também mostra o gráfico de pizza com o percentual de reclamações não resolvidas pela empresa. Novamente movendo o mouse por cima do gráfico ou usando o dedo, o usuário pode extrair mais informações do PROCONFIE. Essas ações estão representadas na Figura 9. As reclamações não resolvidas contribuem para aumentar o risco de fazer negócios com uma determinada empresa. Ou seja, quanto maior a percentagem de reclamações não resolvidas contra uma empresa, maior também pode ser o risco de fazer negócios com ela.

Seguindo o link *Saiba Mais* o usuário será apresentado com quatro gráficos sobre o desempenho da empresa no Procon ao longo do tempo. Tais gráficos estão apresentados na Figura 10. Assim como o gráfico pizza, a exploração desses gráficos é feita com o cursos do mouse ou o dedo do usuário, dependendo do dispositivo sendo utilizado para acessar o PROCONFIE.

O gráfico da Figura 10 (a) apresenta o número de reclamações obtidas por uma empresa ao longo dos anos. Um elevado número de reclamações não necessariamente significa que a empresa possui “alto risco”, ou seja, maior chance de gerar reclamações por usuários. Essa é uma informação importante na avaliação de empresas, mas para avaliar o risco de uma empresa com maior justiça é preciso considerar também outras informações, como as mostradas nos outros gráficos desta página.

O gráfico da Figura 10 (b) mostra a percentagem do número de reclamações de uma empresa em relação a todas as reclamações da sua categoria. Essa informação é interessante, pois possibilita avaliar, de forma geral, o desempenho de uma determinada empresa em determinada categoria.



Figura 7: Dados Iniciais Sobre uma Empresa

Enquanto isso, o gráfico da Figura 10 (c) mostra a percentagem de reclamações resolvidas ao longo dos anos. O crescimento desse valor ao longo do tempo indica que a empresa está apresentando bom desempenho na resolução de problemas. Observe que esse valor percentual não é necessariamente proporcional ao número de reclamações. Isso significa que uma empresa pode apresentar um número de reclamações menor ao longo dos anos, mas diminuir o número de reclamações resolvidas. Essa é uma informação fundamental na criação do ranking de piores empresas em uma determinada categoria.

Por fim, o gráfico da Figura 10 (d) mostra os cinco tipos de problemas mais comuns para uma determinada empresa, possibilitando uma visão geral dos mesmos. Essa informação também pode ser útil no processo de decisão de escolha de uma determinada empresa. Por exemplo, sendo considerada como um critério de desempate.

Por fim, o usuário pode conhecer outras empresas da mesma categoria daquele que ele se encontra atualmente. Para isto, basta clicar no nome da categoria na página da empresa atual. Após esta ação o usuário será apresentado com a lista de outras empresas da mesma categoria. Além disso, é mostrado o risco de cada uma dessas empresas. Assim, o usuário pode explorar tais outras empresas e escolher qual empresa vai fazer negócios. A Figura 11 mostra as empresas de uma mesma categoria que o usuário pode explorar.



Figura 8: *Tooltip* do Risco de uma Empresa

Por fim, resumimos na Figura 12 as possíveis ações que usuários podem fazer no sistema. Lembre-se que as teclas do navegador como Avançar e Voltar podem ser usadas para navegar entre ações já visitadas.



Figura 9: Gráfico Pizza de Percentual de Reclamações Resolvidas/Não-Resolvidas

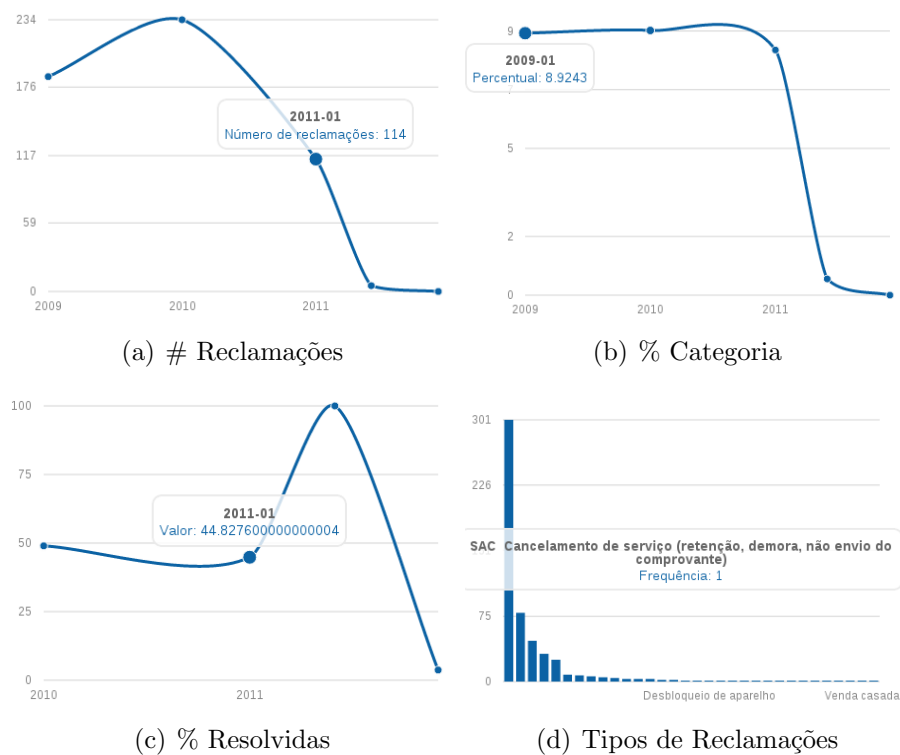


Figura 10: Gráficos de mais Informações sobre Empresas

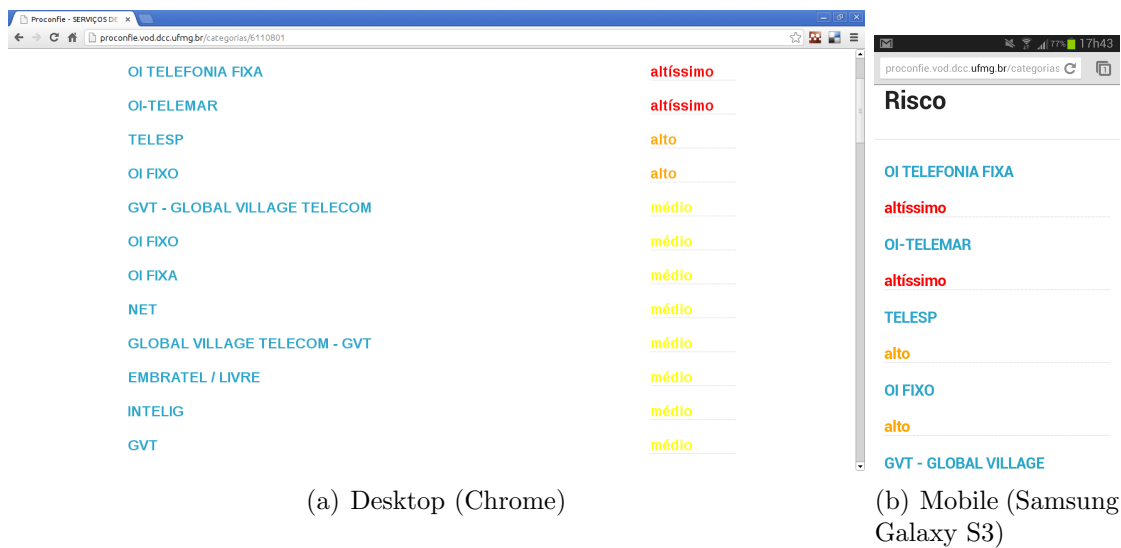


Figura 11: Empresas de uma mesma Categoria

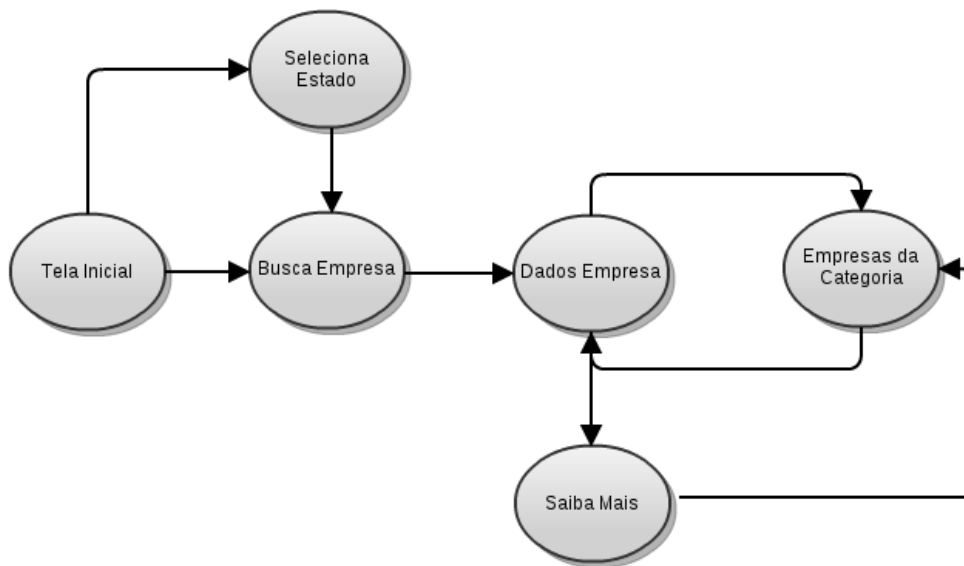


Figura 12: Ações Possíveis Proconfe

7 Equipe Proconfie

A equipe de desenvolvimento do aplicativo Proconfie é composta por membros da Universidade Federal de Minas Gerais, bem como da École Polytechnique de Paris/França. Biografia resumida dos integrantes da equipe:

- **Eduardo Mucelli Rezende Oliveira:** Atualmente é estudante de doutorado na École polytechnique em Paris/França. Eduardo também possui mestrado em Ciência da Computação pela Universidade Federal de Minas Gerais e graduação em Ciência da Computação pela Pontifícia Universidade Católica de Minas Gerais, 2008. Bolsista por dois anos em projetos de iniciação científica na área de Redes de Sensores sem Fio com foco em aplicação para detecção de incêndios. Eduardo também foi pesquisador visitante em 2011 na Telecom Itália (Future Centre) em Veneza/Itália durante seis meses. Currículo Lattes: [http : //lattes.cnpq.br/7979903064399755](http://lattes.cnpq.br/7979903064399755)
- **Thiago Henrique Silva:** Thiago é doutorando em Ciência da Computação na Universidade Federal de Minas Gerais (UFMG). Além disso é mestre em Ciência da Computação pela UFMG (2009) e graduado em Ciência da Computação pela Universidade de Itáúna (2004). Ele realizou o seu doutorado sanduíche em 2012 na University of Birmingham na Inglaterra. Thiago também foi pesquisador visitante em 2011 na Telecom Itália (Future Centre) em Veneza/Itália durante seis meses. Possui experiência profissional e de pesquisa em Redes de Computadores, Sistemas Distribuídos, Computação Móvel e Ubíqua. Outros tópicos de interesse incluem: Computação urbana, Data-science e Interação Homem Computador. Currículo Lattes: [http : //lattes.cnpq.br/9568338246493587](http://lattes.cnpq.br/9568338246493587)
- **Flavio Vinicius Diniz de Figueiredo:** Atualmente Flávio é estudante de doutorado na Universidade Federal de Minas Gerais. Além disso, possui graduação em Ciência da Computação pela Universidade Federal de Campina Grande (2007) e mestrado em Ciência da Computação pela Universidade Federal de Minas Gerais (2010). Currículo Lattes: [http : //lattes.cnpq.br/9481210393304645](http://lattes.cnpq.br/9481210393304645)
- **Pedro Olmo Stancioli Vaz de Melo:** Pedro é professor Adjunto no departamento de Ciência da Computação da Universidade Federal de Minas Gerais. Ele possui graduação em Ciência da Computação pela Pontifícia Universidade Católica de Minas Gerais (2003) e mestrado em Informatica pela Pontifícia Universidade Católica de Minas Gerais (2007). Concluiu seu doutorado pela Universidade Federal de Minas, com doutorado sanduíche na Carnegie Mellon University sob orientação do professor Christos Faloutsos. Ele também foi um pesquisador visitante por cinco meses no INRIA

em Lyon/França. Seus interesses de pesquisa são redes complexas, redes sociais, teoria dos jogos e redes de sensores sem fio. Currículo Lattes: <http://lattes.cnpq.br/3262926164579789>