

Bruno Carvalho Logerfo
Isabele Romanini Carvasan

**ANÁLISE DE DADOS EM ORGANIZAÇÕES NÃO GOVERNAMENTAIS PARA APOIO À
TOMADA DE DECISÃO**

Bauru
2019

UNIVERSIDADE ESTADUAL PAULISTA “JÚLIO DE MESQUITA FILHO”
FACULDADE DE CIÊNCIAS
BACHARELADO EM SISTEMAS DE INFORMAÇÃO

Bruno Carvalho Logerfo
Isabele Romanini Carvasan

**ANÁLISE DE DADOS EM ORGANIZAÇÕES NÃO GOVERNAMENTAIS PARA APOIO À
TOMADA DE DECISÃO**

Trabalho de conclusão de curso apresentado
ao curso de Bacharelado em Sistemas de
Informação da Universidade Estadual Paulista
“Júlio de Mesquita Filho” como requisito para
obtenção do título de Bacharel em Sistemas de
Informação.

Orientador: Prof. Dr. João Pedro Albino

Bauru
2019

Resumo

Ciência de dados é uma área interdisciplinar abrangente, voltada para a análise de dados previamente coletados, com o objetivo de extrair conhecimento que auxilie na tomada de decisões. Esses tomadores de decisão podem ser empresas, organizações não governamentais (ONGs), figuras públicas, dentre outros. Para cumprir seu propósito, essa ciência faz uso de técnicas voltadas à estatística, matemática, economia, banco de dados, programação, além de outras. Seu estudo já existe há mais de 30 anos, mas se popularizou recentemente devido ao desenvolvimento de ferramentas emergentes como big data – termo que trata sobre grandes conjuntos de dados – e aprendizagem de máquina. Isto posto, o objetivo deste trabalho é, em conjunto com a ONG de nome Centro Voluntariado Universitário – CVU (sedes de Bauru/SP e Ribeirão Preto/SP), estudar, avaliar e minerar dados através da ciência de dados e de técnicas coadjuvantes a fim de encontrar padrões que auxiliem a ONG na construção de relatórios, previsões estatísticas, tomada de decisões, ampliação de áreas atingidas pela ONG, dentre outras finalidades.

Palavras-chave: Ciência de dados, Estatística, Centro Voluntariado Universitário, Trabalho voluntário, Tomada de decisão.

Abstract

Data science is a comprehensive interdisciplinary area, focused on the analysis of previously collected data, with the objective of extracting knowledge that assists in making decisions. These decision makers can be companies, non-governmental organizations (NGOs), public figures, among others. In order to fulfill its purpose, this science makes use of techniques focused on statistics, mathematics, economics, databases, programming, and others. Its study has been in existence for more than 30 years, but has recently become popular due to the development of emerging tools as big data – a term that deals with large data sets – and machine learning. Therefore, the objective this work is, together with the NGO named University Volunteer Center – CVU (headquarters of Bauru/SP and Ribeirão Preto/SP), to study, evaluate and mine data through data science and supporting techniques in order to find patterns that help the NGO in the construction of reports, statistical forecasts, decision making, expansion of areas reached by the NGO, among other purposes.

Keywords: Data science, Statistics, University Volunteer Center, Volunteer work, Decision-making.

Lista de ilustrações

Figura 1 – Painel do CoCo Analytics.	11
Figura 2 – Diagrama de caso de uso.	13
Figura 3 – Diagrama de atividades.	15
Figura 4 – Exemplo planilha de dados (.xls).	21
Figura 5 – Exemplo arquivo de dados (.csv).	22
Figura 6 – Fórmula de retenção.	24
Figura 7 – Fórmula de evasão.	25
Figura 8 – Gráfico de setores por gênero do CVU de Bauru sobre a ação de natal em 2018.	30
Figura 9 – Gráfico de setores por gênero do CVU de Ribeirão Preto sobre a ação de “bem estar” em 2017.	30
Figura 10 – Gráfico de setores por curso do CVU de Ribeirão Preto da ação de trote social em 2019.	31
Figura 11 – Gráfico descritivo anual de setores por curso do CVU de Bauru do ano 2018.	32
Figura 12 – Gráfico comparativo proveniente dos dados relacionados à análise de retenção envolvendo o CVU de Bauru e de Ribeirão Preto perante os anos analisados.	33
Figura 13 – Gráfico comparativo proveniente dos dados relacionados à análise de evasão envolvendo o CVU de Bauru e de Ribeirão Preto perante os anos analisados.	34
Figura 14 – Gráfico de área por gênero do CVU de Bauru.	35
Figura 15 – Gráfico de área por gênero do CVU de Ribeirão Preto.	35
Figura 16 – Gráfico de ranking por curso do CVU de Bauru.	36
Figura 17 – Gráfico de ranking por curso do CVU de Ribeirão Preto.	37

Lista de abreviaturas e siglas

CVU	Centro Voluntariado Universitário
ONG	Organização Não Governamental
UNESP	Universidade Estadual Paulista “Júlio de Mesquita Filho”
API	Application Public Interface
CSV	Comma Separated Values
CPF	Cadastro de Pessoa Física
RG	Registro de pessoa física
UML	Unified Modeling Language

Sumário

	Lista de ilustrações	4
	Lista de abreviaturas e siglas	5
	Sumário	6
1	INTRODUÇÃO	8
2	SOLUÇÕES EXISTENTES E OBJETIVOS DO SISTEMA	10
3	ESPECIFICAÇÃO DO SISTEMA DESENVOLVIDO	12
3.1	Análise de recorrência	12
3.2	Computação De Gênero	12
3.3	Computação De Curso	12
3.4	Diagrama De Caso De Uso	13
3.5	Diagrama De Atividades	14
4	TECNOLOGIAS UTILIZADAS	17
4.1	Python	17
4.2	GitHub	17
4.3	Google Sheets	17
4.4	Visual Studio Code	18
4.5	Genderize	18
4.6	Expressões regulares	18
4.7	Python String Similarity	18
4.8	CSV	19
5	IMPLEMENTAÇÃO DO SISTEMA DESENVOLVIDO	20
5.1	Entrada de dados	20
5.2	Processamento de dados	23
6	RESULTADOS	26
6.1	Sugestões	26
6.2	Gráficos	29
7	DIFICULDADES ENCONTRADAS	38
7.1	CVU Bauru	38
7.2	CVU Ribeirão Preto	39
7.3	Refatoração de código	41
8	TRABALHOS FUTUROS	42
8.1	Interface de usuário	42
8.2	Gráficos de ranking por área de conhecimento	42
8.3	Gráficos de densidade de retenção e evasão	42

8.4	Projeções	42
8.5	Correlações aprofundadas	43
8.6	Feedback	43
9	CONCLUSÃO	44
	REFERÊNCIAS	46

1 INTRODUÇÃO

Organizações não governamentais (ONGs) são organizações autônomas sem fins lucrativos, criadas com o objetivo de promover ações solidárias, apoiando populações excluídas das condições de cidadania. A ascensão desse tipo de instituição está ligada tanto à crise do Estado quanto ao desenvolvimento da população no sentido de uma sociedade colaborativa e ativa.

Atualmente, é difícil para instituições voluntárias medirem seus índices de eficiência. Isso acontece porque elas operam sem fins lucrativos, e o lucro é um sinal de eficiência na economia. Sem indicadores confiáveis, as instituições ficam sem saber onde focar, o que fazer de melhor, como promover ações e assim por diante. É bem provável que, por causa disso, seus resultados não sejam os melhores possíveis. Através da correlação de dados, é possível que uma análise de previsão mude esse panorama.

Vale a pena proferir que organizações não governamentais funcionam, principalmente, através de doações financeiras e voluntariado – conjunto de ações sociais e comunitárias sem recebimento de qualquer remuneração ou lucro, contribuindo para a melhoria de vida de populações marginalizadas.

De acordo com Petenate (2016), a correlação entre variáveis é uma importante fonte para o entendimento de um problema e uma maneira de encontrar possíveis soluções. Nesse trabalho, entre outros aspectos importantes, busca-se avaliar o trabalho produzido pela ONG parceira do trabalho – Centro de Voluntariado Universitário (CVU, 2019) – e a saída é o serviço realizado, porém este serviço pode ser avaliado sobre diferentes pontos de vista ou pela correlação de variáveis importantes para a instituição. Ainda segundo Petenate (2016), o objetivo da correlação é o de encontrar relação entre variáveis e, encontrando esta relação pode-se prever uma forma de melhorar o serviço prestado às comunidades carentes atendidas pela ONG.

Um estudo realizado pela Everyaction (2016) indica que 90% das organizações sem fins lucrativos coletam dados. Entretanto, apenas metade estão cientes das maneiras que os dados podem impactar seus trabalhos. O que faz sentido, já que dois terços dessas organizações não possuem um membro específico dedicado ao gerenciamento de dados. O que chama atenção nesse estudo é que 5% das organizações tomam decisões baseadas em dados, um número que pode ser considerado baixo.

Dessa forma, o objetivo do trabalho apresentado é, em parceria com a ONG denominada Centro de Voluntariado Universitário (CVU, 2019), otimizar e qualificar processos de voluntariado a fim de melhorar a qualidade de vida dos atingidos e dos próprios voluntários.

Segundo a descrição no site da própria instituição, o CVU é:

[...] uma associação sem fins lucrativos, de caráter não-religioso e não-político, fundada em Abril de 2011, em Ribeirão Preto (SP), composta por estudantes e professores universitários. A finalidade do CVU é fomentar o trabalho voluntário no meio acadêmico a partir de eventos; cursos de capacitação; projetos sociais; parcerias; informação e acompanhamento para o voluntário. (CVU, 2019)

Atualmente, o CVU possui oito núcleos em todo o Brasil: Bauru (SP), Limeira (SP), Maringá (PR), Ribeirão Preto (SP), São Carlos (SP), Uberaba (MG), Uberlândia (MG) e Franca (SP).

Para atingir os propósitos delineados neste trabalho, os desenvolvedores utilizaram os dados de dois núcleos do CVU: um deles localizado na cidade de Bauru, mais especificamente no campus da Universidade Estadual Paulista “Júlio de Mesquita Filho” (UNESP, 1976) e o outro localizado na cidade de Ribeirão Preto, no campus da Universidade de São Paulo (USP, 1934).

O método utilizado foi o desenvolvimento de uma ferramenta que analisa dados disponibilizados pela ONG parceira em busca de correlações e padrões, para que seja possível criar projeções e previsões – sociais, econômicas, estatísticas, dentre outras – com o intuito de facilitar processos, organizar melhor os dados, auxiliar na tomada de decisões e aumentar a quantidade de pessoas atingidas.

Como desenvolvimento futuro e a partir da estrutura desenvolvida neste projeto, pode-se dar continuidade ao trabalho por meio do desenvolvimento de uma interface para automatizar o processamento.

Nesta fase inicial de desenvolvimento, foram identificadas dificuldades, tais como o não domínio das ferramentas envolvidas. Outro obstáculo que também poderia ser enfrentado é em relação aos dados disponibilizados pela instituição parceira; isso porque, se não houvesse dados significativos e em grande quantidade, o processo de análise dos mesmos seria menos preciso e, conseqüentemente, pode trazer uma falta de exatidão nas análises desenvolvidas.

2 SOLUÇÕES EXISTENTES E OBJETIVOS DO SISTEMA

Após realizada uma pesquisa em busca de soluções similares e já existentes para o problema apresentado – ou parte dele –, foi encontrado o programa CoCo Analytics (Connect Online Connect Offline), da Digital Green. Se trata de um sistema proprietário de coleta de dados com um painel de cadastro e visualização em um *website*. A ideia é que ele seja usado por diferentes instituições e que seus dados estejam disponíveis publicamente, permitindo que os usuários coletem e visualizem informações cruciais a qualquer momento (DIGITAL GREEN, 2019).

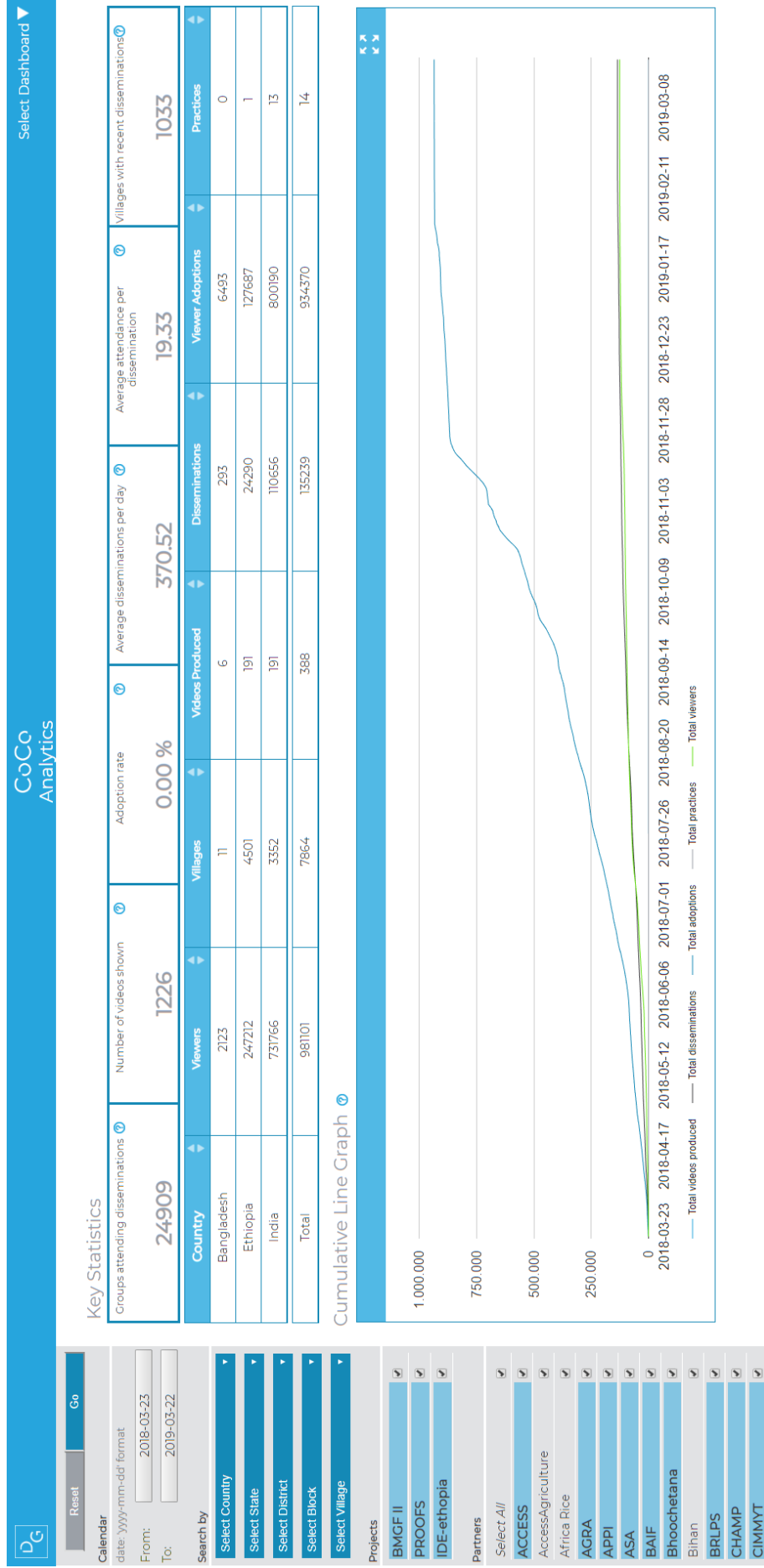
Assim como a Digital Green, o objetivo deste trabalho é acelerar, através da tecnologia, os processos – principalmente em relação à ações de voluntariado – da ONG analisada, auxiliando a propagação de qualidade de vida para pessoas mais carentes.

Pode-se fazer um comparativo entre o sistema desenvolvido no decorrer deste trabalho e o próprio CoCo Analytics. Dentre as vantagens que a plataforma CoCo Analytics apresenta está seu alto nível de interação possível entre usuário e sistema, ou seja, enquanto possui um painel de cadastro e visualização – permitindo que o usuário cadastre manualmente suas informações, filtre os dados como desejar e obtenha informações praticamente instantâneas sobre determinada instituição, ação, período ou área de atuação desejada –, o sistema deste trabalho não possui tais funcionalidades, sendo não tão amigável. Além disso, usuários do CoCo Analytics podem acessar a plataforma de qualquer dispositivo – desde que tenham acesso estável à internet –, o que não se pode afirmar do sistema desenvolvido neste trabalho, que torna-se acessível somente através de navegadores em computadores.

Por outro lado, também há notáveis desvantagens do CoCo Analytics em relação ao sistema deste trabalho. Podemos citar a questão da acessibilidade: o sistema desenvolvido é completamente acessível por todo potencial usuário, já que possui seu código totalmente aberto e nenhuma restrição de uso. Enquanto isso, o CoCo Analytics mantém seu código privado, além de estar disponível apenas para instituições cadastradas e asiáticas.

A Figura 1 exemplifica como é organizado e estruturado o painel do CoCo Analytics:

Figura 1 – Painel do CoCo Analytics.



Fonte: (DIGITAL GREEN, 2019)

3 ESPECIFICAÇÃO DO SISTEMA DESENVOLVIDO

3.1 Análise de recorrência

A análise de recorrência se dá ao identificar que o mesmo voluntário participante das ações do CVU – seja em relação ao CVU de Bauru ou em relação ao CVU de Ribeirão Preto – esteve presente em duas ou mais ações durante o período dos dados de entrada do sistema – de 2015 a 2019 no caso do CVU de Bauru, e de 2016 a 2019 no caso do CVU de Ribeirão Preto. Para que esse processo aconteça, é necessário fazer uso de um identificador do voluntário.

3.1.1 Análise De Retenção

Utilizando os resultados da análise de recorrência, a análise de retenção visa identificar o índice de retenção de novos voluntários em cada ação, ou seja, quantos voluntários determinada ação trouxe e reteve para as próximas ações a serem desenvolvidas.

3.1.2 Análise De Evasão

A análise de evasão tem como objetivo realizar o trabalho inverso da análise de retenção. Enquanto a análise de retenção busca por participações retidas, a análise de evasão busca por participações cessadas.

3.2 Computação De Gênero

Devido à falta de alguns dados relevantes sobre o voluntariado nas planilhas analisadas – tanto em relação ao CVU de Bauru quanto em relação ao CVU de Ribeirão Preto – surgiu a necessidade de inferir o gênero de cada voluntário participante das ações.

Isso se deu por conta da importância de disponibilizar um balanceamento dessas informações aos encarregados pelo CVU, para possíveis análises futuras sobre recorrência e participação de seus voluntários divididos por gênero. Sendo assim, nessa etapa de especificação do sistema, os desenvolvedores passaram a pesquisar maneiras de contornar a falta dessa informação nos dados a serem computados.

3.3 Computação De Curso

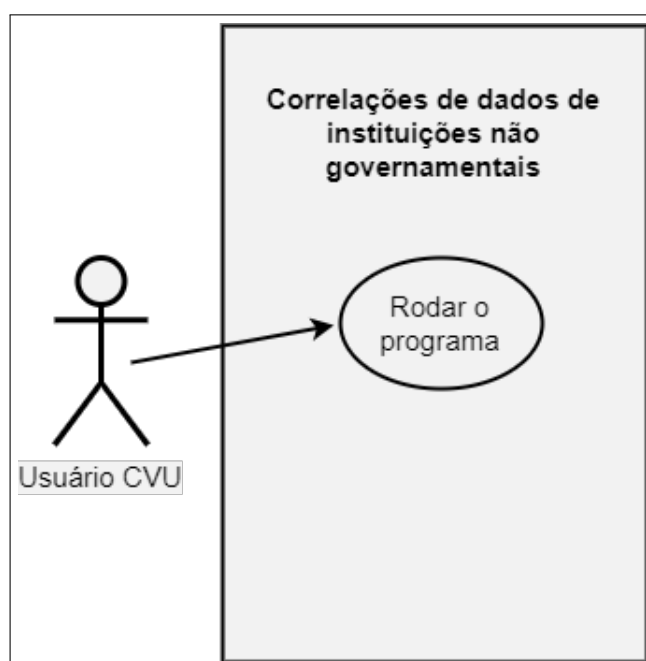
A computação de curso se dá de forma semelhante à computação de gênero, com uma maior variedade de classificações. Assim como para cada voluntário computa-se um gênero, também computa-se um curso. A diferença está na quantidade de classificações: enquanto trabalha-se com três classificações fixas de gênero (masculino, feminino e não informado), o número de classificações de curso é variável e tende a aumentar conforme o número de voluntários.

3.4 Diagrama De Caso De Uso

O diagrama de caso de uso – ferramenta para o levantamento dos requisitos funcionais – descreve a funcionalidade proposta para um novo sistema que será projetado. É uma unidade discreta de um trabalho significativamente maior e representa uma interação entre um ator – sendo esse ator um humano ou entidade máquina que interage com o sistema para executar um trabalho ou atividade – e o sistema.

Cada caso de uso tem uma descrição da funcionalidade que será construída no sistema proposto. Exemplos de caso de uso são: login do sistema, registrar no sistema, criar pedidos, dentre outros. Sendo assim, parte do processo de especificação desse sistema envolve o desenvolvimento de um diagrama de casos de uso.

Figura 2 – Diagrama de caso de uso.



Fonte: elaborada pelos autores.

Como é possível observar através da Figura 2, há apenas uma ação disponível a ser realizada por um possível usuário que possuir o código fonte do sistema desenvolvido. A proposta para esse trabalho consiste em encontrar padrões entre dados internos da instituição envolvida e externos – de assuntos no geral que possam ser relevantes ao estudo. Os resultados são relatados de forma que classifiquem correlações por suas forças e desclassifiquem dados sem correlação alguma. Uma vez obtidas essas informações, é possível aplicá-las para otimizar os processos relacionados à instituição parceira, CVU.

Sendo assim, o sistema desenvolvido não abrange questões envolvendo o usuário, ou seja, não há uma interface gráfica implementada que permita o usuário realizar ações dentro do sistema. Em outras palavras, isso representa que o programa não é amigável a nível de usuário: os parâmetros de entrada estão diretos no código fonte, a saída não é parametrizável, é necessário

alterar o código fonte para adicionar novos cursos dos voluntários, não há documentação fora do código fonte para explicar ao usuário o que é feito, dentre outras características.

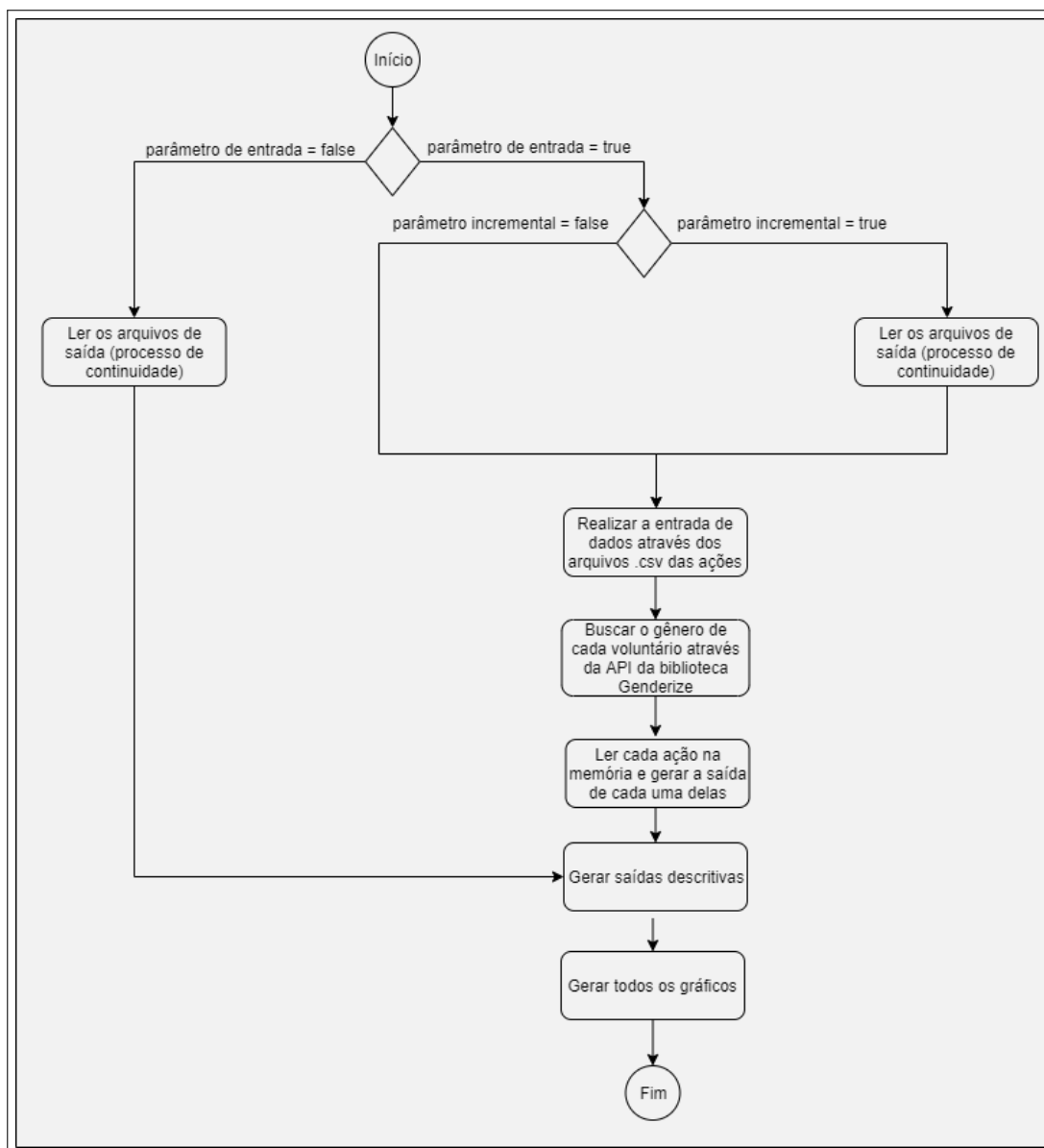
Por outro lado, o sistema é escalável também a nível de usuário, ou seja, é realizável implementar também sistemas de entrada – sejam cadastros ou APIs – e saída – simples relatórios ou painéis de análise – de dados, utilizando as correlações encontradas para transformar tudo isso num fluxo de atividades.

Entretanto, o tempo hábil desse trabalho não permitiu uma implementação tão profunda – a nível de usuário – quanto os desenvolvedores desejavam. Assim, é importante dar ênfase de que o sistema desenvolvido foi apenas uma maneira de atingir resultados para informar e auxiliar o CVU em relação às melhores formas de organizar e processar seus próprios dados originados das ações executadas. Sendo assim, fica aberto o convite a outros acadêmicos ou curiosos para, ao final desse trabalho, dar continuidade ao mesmo, em qualquer estado que seja concluído.

3.5 Diagrama De Atividades

Um diagrama de atividade, definido pela Linguagem de Modelagem Unificada (UML), é utilizado para representar os fluxos conduzidos por processamentos, ou seja, envolve a modelagem das etapas sequenciais de um processo computacional, mostrando o fluxo de controle de uma atividade para outra. Sendo assim, parte do processo de especificação desse sistema também envolve o desenvolvimento de um diagrama de atividades, como representado na Figura 3.

Figura 3 – Diagrama de atividades.



Fonte: elaborada pelos autores.

Vale a pena também descrever com mais detalhes o que representam os parâmetros de entrada, o incremental e o processo de continuidade, criados com o intuito de acelerar processos, evitando que toda execução do sistema repita atividades já feitas na execução anterior e consideradas, portanto, desnecessárias.

O parâmetro de entrada – que pode ser determinado e alterado programaticamente pelos desenvolvedores do sistema ou todo e qualquer indivíduo que tenha acesso ao código fonte – delibera, quando seu valor é `false`, a possibilidade do sistema executar apenas a parte da geração dos gráficos com base nos resultados processados pela execução anterior. Se o valor for `true`, o fluxo de execução do sistema é deliberado também pelo parâmetro incremental.

Por sua vez, o parâmetro incremental – também necessário de ser determinado ou pos-

sivelmente alterado no código fonte do sistema desenvolvido nesse trabalho – cuja relevância depende do valor determinado primeiramente pelo parâmetro de entrada, delimita a necessidade, após o início do processamento, de realizar ou não o processo de continuidade – que envolve a leitura dos arquivos de saída dos processamentos feitos anteriormente – antes de passar para a entrada de novos dados provenientes das ações dispostas nos arquivos .csv.

4 TECNOLOGIAS UTILIZADAS

Neste capítulo será disponibilizada uma pequena introdução sobre as tecnologias escolhidas pelos desenvolvedores do sistema.

4.1 Python

A linguagem de programação escolhida para o tratamento de dados, bem como geração de gráficos e todo e qualquer tipo de codificação necessária durante o decorrer do desenvolvimento do sistema foi a Python.

Criada por Guido Van Rossum em 1991, a linguagem de programação Python foi desenvolvida sob uma licença de código aberto aprovada pela OSI (OPENSOURCE.ORG, 1998), o que a torna livremente utilizável e distribuível – seja para uso pessoal ou comercial.

Caracterizada por ser uma linguagem de programação de alto nível, orientada a objetos e funcional, é principalmente utilizada para processamento de textos e de dados científicos, aprendizagem de máquina, inteligência artificial, criação de CGIs para páginas dinâmicas da web, dentre outras funcionalidades. A licença do Python é administrada pela Python Software Foundation (PYTHON SOFTWARE FOUNDATION, 2019).

4.2 GitHub

A nível de versionamento de código, a tecnologia escolhida foi o GitHub: plataforma de hospedagem de código-fonte com controle de versão. Ele permite que os desenvolvedores do sistema contribuam com o mesmo de forma fácil e clara, além de promover a comunicação através de recursos que relatam problemas ou mesclam repositórios remotos. Dessa forma, tudo o que é desenvolvido pelos programadores fica registrado e versionado, fazendo com que nenhum código seja perdido ou esquecido (GITHUB INC., 2019).

Por sua vez, o GitHub desempenha seu papel através da utilização do Git: software livre de controle de versões distribuído, usado principalmente no desenvolvimento de software para registrar o histórico de edições de qualquer tipo de arquivo (LINUS TORVALDS, 2019).

4.3 Google Sheets

Outra tecnologia também utilizada na formatação, bem como durante o processo de levantamento do volume e vastidão dos dados disponíveis pela instituição parceira foi o editor de planilhas do Google, o *Google Sheets* (GOOGLE, 2019b).

Em conjunto com essa ferramenta, os desenvolvedores puderam ter maior organização e visão macro de quais seriam os dados que dariam entrada ao estudo prático proposto. Além disso, também foi utilizada para conversão de planilhas no formato *.xls* que continham dados do voluntariado em arquivos de formato *.csv* devido a maior facilidade de processamento – conforme mencionado anteriormente.

4.4 Visual Studio Code

Lançado em 2015 pela Microsoft como um software livre e de código aberto, o Visual Studio Code foi escolhido como editor do código-fonte do sistema desenvolvido. Dentre suas inúmeras funcionalidades, ele inclui suporte para depuração, controle de Git incorporado, realce de sintaxe, complementação inteligente de código, *snippets* e refatoração de código que auxiliam os desenvolvedores no processo de criação do sistema, características que foram de grande peso no momento da escolha dessa ferramenta (MICROSOFT, 2019).

4.5 Genderize

Como dito anteriormente, a informação específica do gênero de cada voluntário participante das ações analisadas não estava contida dentre os dados disponibilizados pela instituição. Sendo assim, para atingir tal objetivo de inferir essa informação foi utilizada uma biblioteca chamada *Genderize* (DEMOGRAFIX APS, 2014).

4.6 Expressões regulares

Expressões regulares, ou *RegEx*, são cadeias de caracteres utilizadas para análise de outras cadeias de caracteres. São muito utilizadas para classificação predeterminada e validação de dados, o que condiz com a necessidade demonstrada para a computação de curso por voluntário.

4.7 Python String Similarity

Para o cumprimento do objetivo dos programadores de realizar no sistema a análise de recorrência do voluntariado – partindo do pressuposto que nem todos os voluntários disponibilizaram seus respectivos números de documento de identidade (RG) e/ou de documento de pessoa física (CPF) – foi necessário implementar essa função através do nome de cada voluntário participante das ações do CVU, tanto de Bauru quanto de Ribeirão Preto.

Para isso, os desenvolvedores contaram com o auxílio de uma biblioteca de similaridade de cadeia chamada de *Python String Similarity* (ZHOUYANG, 2018) disponibilizada através da plataforma de compartilhamento de códigos GitHub (GITHUB INC., 2019).

O funcionamento dessa biblioteca acontece através de diferentes medidas de similaridade e distância de cadeias. Para isso, conta com a implementação de diversos algoritmos como, por exemplo, Levenshtein, Jaro-Winkler, Longest Common Subsequence, semelhança de cosseno, dentre outros.

Utilizando apenas o método conhecido como Jaro-Winkler, os programadores desse sistema estipularam uma porcentagem de 96% de similaridade entre os nomes, ou seja, nomes 96% ou mais similares são considerados iguais. Tal porcentagem foi escolhida por ser a menor onde não aconteceram conflitos entre nomes diferentes ao ser executada a função.

A documentação da biblioteca utilizada é bem ampla e rica em detalhes, o que auxiliou no entendimento da mesma pelos programadores desse sistema.

4.8 CSV

CSV, ou em sua nomenclatura extensa *comma-separated-values*, é um tipo de formato de arquivo de texto utilizado principalmente para armazenar dados e possibilitar transferências de informações entre plataformas diferentes. A formatação desses dados se dá através da separação e delimitação por uma vírgula dos campos que compõem o arquivo, possibilitando a leitura dos mesmos por meio de diversos aplicativos diferentes.

Nesse trabalho, os desenvolvedores utilizaram essa tecnologia durante a entrada de dados no sistema: como o CVU, tanto de Bauru quanto de Ribeirão Preto, armazenavam os dados dos seus respectivos voluntários em planilhas *.xls*, foi necessário transformar esses arquivos em *.csv* a fim de utilizá-los em conjunto com a linguagem de programação Python (PYTHON SOFTWARE FOUNDATION, 2019), além de também garantir um desempenho mais rápido de processamento do software.

5 IMPLEMENTAÇÃO DO SISTEMA DESENVOLVIDO

5.1 Entrada de dados

Com o objetivo de iniciar o sistema proposto, foi necessário que a instituição parceira disponibilizasse para os desenvolvedores os dados previamente armazenados: fossem eles relacionados à área social e de voluntariado, ou à parte financeira da ONG, ou qualquer outro tipo de dado que pudesse ser útil para o cruzamento de informações.

Sendo assim, foram coletadas, no total relacionado ao CVU de Bauru, 21 planilhas contendo informações do voluntariado participante das ações entre os anos de 2015 a 2019, 16 planilhas com dados relacionados aos vídeos publicados na página do Facebook da instituição e 29 planilhas com dados relacionados à página do Facebook propriamente dita – informações como, por exemplo, sobre posts publicados, engajamento da página, dentre outras.

Já em relação ao CVU de Ribeirão Preto, foram coletadas 23 planilhas contendo informações apenas sobre os voluntários participantes das ações desenvolvidas durante os anos de 2016 a 2019.

Para possibilitar o processamento dos dados contidos nessas planilhas, os desenvolvedores tiveram que, primeiramente, promover a entrada desses dados no software desenvolvido. Para isso, foi necessário transformar esses arquivos de formato `.xls` (Figura 4) em arquivos de formato `.csv` (Figura 5), devido à facilidade de leitura desse tipo de arquivo pelas bibliotecas de Python utilizadas pelos desenvolvedores.

Em segundo plano, foi necessário dar manutenção a esses dados – agora em arquivos `.csv` – e limpá-los: foram removidas colunas em branco ou com informações incompletas, além de desenvolvidos padrões de texto e organização para aqueles dados recorrentes e significativos.

Por fim, após esse processo de mineração e limpeza de dados, as informações sobre o voluntariado consideradas relevantes e utilizadas pelos desenvolvedores foram, principalmente: nome completo, CPF e/ou RG, curso universitário e gênero.

Figura 4 – Exemplo planilha de dados (.xls).

Programa Semente_ Páscoa.xlsx * x +									
Heloísa Helena Figueiredo Paes de Oliveira									
A	B	C	D	E	F	G	H	I	
1	Carimbo de data/hora	Nome completo	Telefone para contato	CPF	Link do seu perfil do facebook	Você pode oferecer car	Se sim, quantas	Você faz algum curso?	Como você soube da ação? Você po
2	4/10/2019 17:38:03	Júlia Corrêa	411 111 1111	411 111 1111	Não, preciso de carona	Não, preciso de carona		Relações Públicas	Evento no facebook
3	4/10/2019 18:34:50	Júlia Corrêa	411 111 1111	411 111 1111	Não, preciso de carona	Não, preciso de carona		Arquitetura	Evento no facebook
4	4/11/2019 12:26:43	Júlia Corrêa	411 111 1111	411 111 1111	Não, preciso de carona	Não, preciso de carona		Psicologia	Indicação
5	4/11/2019 15:50:17	Júlia Corrêa	411 111 1111	411 111 1111	Sim	Sim	4	Pós em Marketing Digital	Evento no facebook
6									
7	4/11/2019 20:21:21	Júlia Corrêa	411 111 1111	411 111 1111	Não, preciso de carona	Não, preciso de carona		Psicologia	
8	4/12/2019 20:59:14	Júlia Corrêa	411 111 1111	411 111 1111	Sim	Sim	4, mas acho q v	BCC	ALO CULTURA
9	4/13/2019 12:07:46	Júlia Corrêa	411 111 1111	411 111 1111	Não, preciso de carona	Não, preciso de carona		Engenharia Elétrica	Sou do CVU
10	4/13/2019 12:35:57	Júlia Corrêa	411 111 1111	411 111 1111	Não, preciso de carona	Não, preciso de carona		Relações públicas	Sou de Cultura
11	4/13/2019 15:15:20	Júlia Corrêa	411 111 1111	411 111 1111	Não, preciso de carona	Não, preciso de carona	Não posso ofere	Sim. Arquitetura e Urbanismo	Divulgação durante o processo
12	4/14/2019 11:26:19	Júlia Corrêa	411 111 1111	411 111 1111	Não, preciso de carona	Não, preciso de carona		Sim. Arquitetura e Urbanismo	Divulgação durante o processo
13	4/14/2019 12:09:23	Júlia Corrêa	411 111 1111	411 111 1111	Não, preciso de carona	Não, preciso de carona		Ciência da Computação	Estou no CVU kkkkk
14	4/14/2019 13:38:57	Júlia Corrêa	411 111 1111	411 111 1111	Não, preciso de carona	Não, preciso de carona		Artes Visuais Unesp	Status do WhatsApp
15	4/15/2019 0:26:16	Júlia Corrêa	411 111 1111	411 111 1111	Não, preciso de carona	Não, preciso de carona		Sim. Jornalismo	Evento no facebook
16	4/15/2019 0:53:29	Júlia Corrêa	411 111 1111	411 111 1111	Sim	Sim	3	RP	
17	4/15/2019 12:05:34	Júlia Corrêa	411 111 1111	411 111 1111	Não, preciso de carona	Não, preciso de carona		Engenharia elétrica	Evento no facebook
18	4/15/2019 13:24:01	Júlia Corrêa	411 111 1111	411 111 1111	Não, preciso de carona	Não, preciso de carona		Engenharia elétrica UNESP	Exclusividade de organização
19	4/15/2019 23:41:13	Júlia Corrêa	411 111 1111	411 111 1111	Não, preciso de carona	Não, preciso de carona		Jornalismo	Sou membro do CVU

Fonte: elaborada pelos autores.

Figura 5 – Exemplo arquivo de dados (. csv).

```

1 Carimbo de data/hora,Nome completo,"Telefone para contato (preferencialmente, whatsapp)","CPF,Link do seu perfil do facebook,Você pode oferecer carona?,"Se sim, quantas vagas?"
2 4/10/2019 17:38:03,Julia Cristina Rodrigues,1199114-0000,400.111-111-111,"Não, preciso de carona",,Relações Públicas,Evento no facebook,Sim,Eu concordo com o termo de volunt
3 4/10/2019 18:34:50,Artur Henrique Figueiredo Neto de Oliveira,000.000-000-000,"Não, preciso de carona",,Arquitetura,Evento no facebook,Sim,Eu concordo com o
4 4/11/2019 12:26:43,Letícia Aguiar Sousa Galvães,1199114-0000,400.111-111-111,"Não, preciso de carona",,Psicologia,Indicação,Sim,Eu concordo
5 4/11/2019 15:50:17,Letícia Aguiar Sousa Galvães,1199114-0000,400.111-111-111,"Não, preciso de carona",,Marketing Digital,Evento no facebook,Sim,Eu concordo com c
6 ,,,,,,
7 4/11/2019 20:21:21,Letícia Aguiar Sousa Galvães,1199114-0000,400.111-111-111,"Não, preciso de carona",,Psicologia,,Sim,Eu concordo com o termo de voluntariado,Concordo com o te
8 4/12/2019 20:59:14,Apelidos dos Rodrigues,1199114-0000,400.111-111-111,"Não, preciso de carona",,BCC,ALO CULTURA,,Sim,Eu conc
9 4/13/2019 12:07:46,Roberto Carlos Mendes,1199114-0000,400.111-111-111,"Não, preciso de carona",,Engenharia Elétrica,Sou do CVU,Talvez,Eu concordo com o termo de voluntariad
10 4/13/2019 12:35:57,Letícia Aguiar Sousa Galvães,1199114-0000,400.111-111-111,"Não, preciso de carona",,Relações Públicas,Sou de Cultura,Talvez,Eu concordo com o termo de voluntariado,(
11 4/13/2019 15:15:20,Letícia Aguiar Sousa Galvães,1199114-0000,400.111-111-111,"Não, preciso de carona",,Não posso oferecer carona,,Sim, Arqu
12 4/14/2019 11:26:19,Letícia Aguiar Sousa Galvães,1199114-0000,400.111-111-111,"Não, preciso de carona",,Talvez,Eu concordo com o termo de voluntariado,Concordo com o termo de
13 4/14/2019 12:09:23,Letícia Aguiar Sousa Galvães,1199114-0000,400.111-111-111,"Não, preciso de carona",,Ciência da Computação,Estou no CVU kkkkk
14 4/14/2019 13:38:57,Letícia Aguiar Sousa Galvães,1199114-0000,400.111-111-111,"Não, preciso de carona",,Não, preciso de carona",,Sim, Jornalismo,Evento no facebook,
15 4/15/2019 0:26:16,Letícia Aguiar Sousa Galvães,1199114-0000,400.111-111-111,"Não, preciso de carona",,Sim, Jornalismo,Evento no facebook,
16 4/15/2019 0:53:29,Letícia Aguiar Sousa Galvães,1199114-0000,400.111-111-111,"Não, preciso de carona",,Sim, Jornalismo,Evento no facebook,
17 4/15/2019 12:05:34,Letícia Aguiar Sousa Galvães,1199114-0000,400.111-111-111,"Não, preciso de carona",,Sim, Jornalismo,Evento no facebook,
18 4/15/2019 13:24:01,Letícia Aguiar Sousa Galvães,1199114-0000,400.111-111-111,"Não, preciso de carona",,Sim, Jornalismo,Evento no facebook,
19 4/15/2019 23:41:13,Letícia Aguiar Sousa Galvães,1199114-0000,400.111-111-111,"Não, preciso de carona",,Sim, Jornalismo,Evento no facebook,
20 ,,,,,,

```

Fonte: elaborada pelos autores.

5.2 Processamento de dados

Sobre o processamento dos dados, foram desenvolvidas diversas funções utilizando a linguagem de programação Python, onde as informações previamente coletadas da instituição parceira sobre o voluntariado foram padronizadas: aquelas consideradas relevantes mantiveram-se no estudo, enquanto as incompletas ou desnecessárias ao mesmo foram descartadas.

Além disso, outras funcionalidades foram desenvolvidas como, por exemplo, aquelas relacionadas à validação do número de documento – seja ele o documento de identidade (RG) ou o registro de pessoa física (CPF) –, ou então relacionada à análise dos gêneros e cursos presentes nas ações, além daquelas relacionadas à criação e formação de gráficos apresentando resultados (que será melhor descrito na seção 6.2), dentre outras. A seguir estão detalhadas as funções mais impactantes para o cumprimento do objetivo do sistema como um todo.

5.2.1 Determinação do gênero

A biblioteca *Genderize* (DEMOGRAFIX APS, 2014) foi utilizada com o objetivo de trazer ao sistema desenvolvido a funcionalidade de, através de requisições em uma API, determinar o gênero de um indivíduo por meio da verificação do seu nome fornecido.

Seu exercício acontece da seguinte forma: aos programadores interessados, é possível inferir o gênero de até 10 nomes por requisição solicitada apenas enviando um vetor com tais informações contidas em um parâmetro “nome”. Após o processamento dos mesmos, em um cenário perfeito, a biblioteca retorna também um array com o gênero inferido para cada voluntário. A utilização dessa API é gratuita em até 1000 requisições por dia; após esse número, é necessário obter uma API key no site da biblioteca.

No decorrer do desenvolvimento do sistema, alguns desafios provenientes do uso da biblioteca surgiram. Certos nomes de voluntários não eram processados pela API, talvez pela alta complexidade dos mesmos ou pelo fato de serem pouco convencionais. Para contornar esse problema, foi necessário criar um dicionário contendo esses nomes específicos com seus gêneros já determinados pelos próprios programadores, com o objetivo de possuir essa informação mesmo sem o retorno da biblioteca.

Outro ponto a ser mencionado é a análise de dados levando em consideração a cisgeneridade: os programadores do sistema não possuem a informação se o nome apresentado do voluntário é social ou de registro em cartório. De qualquer forma, identificar o gênero de um indivíduo através do seu nome apresentado exclui possíveis casos de pessoas transexuais, não-binárias, dentre outros. Dessa forma, a utilização dessa biblioteca restringe o voluntariado em determinadas questões, apresentando, então, a possibilidade de uma margem de erro na análise do gênero do mesmo.

5.2.2 Determinação de curso

Diferentemente da determinação de gênero, não existe uma API pronta para lidar com isso. Como alternativa, uma análise de voluntário a voluntário foi feita, de forma a categorizar cada entrada no campo livre de curso como um curso predeterminado através de expressões regulares. O sistema conta com uma lista de aproximadamente 60 cursos predeterminados, garantindo que a análise de todos os voluntários testados seja exaustiva.

Caso seja feita entrada de uma nova ação contendo um voluntário cujo curso não está especificado nessa lista, esta precisa ser atualizada para contemplar esse curso até então desconhecido pelo sistema de classificação.

5.2.3 Análise de recorrência

Para o desenvolvimento da análise de recorrência, foram utilizados os números dos documentos pessoais, tanto RG quanto CPF, dependendo de qual foi apresentado pelo voluntário. Entretanto, nem todos os registros de voluntários acompanham algum número de documento. Além disso, não é possível identificar recorrência pelos números dos documentos entre registros com tipos de documentos diferentes.

Por esses motivos, também foi feita uma análise de similaridade de nome. Nomes 96% ou mais similares são considerados iguais. Essa porcentagem foi escolhida em função de ser a menor onde não houve conflito de nomes diferentes, porém 96% similares. O algoritmo utilizado para o cálculo de similaridade foi o *Jaro-Winkler*, através da biblioteca *Python String Similarity* (ZHOUYANG, 2018).

5.2.4 Análise de retenção

É considerada retentora a primeira ação que determinado voluntário participou, apenas caso esse mesmo voluntário venha a participar pelo menos uma segunda vez. O índice de retenção é representado em porcentagem por ação e seu cálculo é dado pela fórmula descrita na Figura 6.

Figura 6 – Fórmula de retenção.

$$\frac{M}{N} * 100$$

M quantidade de voluntários novos que vieram a participar pelo menos uma outra vez
 N quantidade de voluntários que vieram a participar pelo menos uma outra vez

Fonte: elaborada pelos autores.

5.2.5 Análise de evasão

É considerada evasora a última ação que determinado voluntário participou, apenas caso esse mesmo voluntário tenha participado pelo menos uma outra vez. O índice de evasão é representado em porcentagem por ação e seu cálculo é dado pela fórmula descrita na Figura 7.

Figura 7 – Fórmula de evasão.

$$\frac{M}{N} * 100$$

M	quantidade de voluntários recorrentes que não vieram a participar uma outra vez
N	quantidade de voluntários recorrentes

Fonte: elaborada pelos autores.

6 RESULTADOS

Após o desenvolvimento de todo o sistema, como resposta à todos os dados coletados e toda implementação realizada, se fez necessário apresentar resultados – estatísticos ou não – tanto ao CVU quanto a quem acompanhou toda a criação do software. Sendo assim, os desenvolvedores encontraram duas formas de mostrar esses resultados, uma delas quantitativa e a outra qualitativa.

A forma qualitativa diz respeito somente ao CVU: durante o desenvolvimento do sistema, os programadores tiveram algumas dificuldades, como mencionadas anteriormente, em relação à coleta e padronização dos dados do voluntariado disponibilizados pela ONG. Sendo assim, conforme isso ocorria, foram levantadas possíveis soluções para esses desafios, a fim de mostrar aos responsáveis pelo CVU algumas formas de melhor coleta e armazenamento dos dados em um contexto futuro. Tal assunto pode ser lido mais detalhadamente na seção 6.1.

A medida quantitativa foi através da elaboração de diversos gráficos utilizando os dados dos voluntários participantes das ações do CVU durante o período analisado – de 2015 à 2019 no caso do CVU de Bauru, e de 2016 à 2019 no caso do CVU de Ribeirão Preto – além de um arquivo .csv descritivo – ou seja, observando todo o conjunto amostral, envolvendo o CVU de Bauru e o de Ribeirão Preto – contendo o resultado da análise de recorrência desenvolvida. Dentre os gráficos estão: de setores para a análise por gênero e por curso de cada ação desenvolvida e de cada grupo por análise combinatória – ou seja, para cada profundidade das pastas lógicas onde as planilhas foram armazenadas, sendo elas organizadas em uma combinação de ano, mês e cidade – além do gráfico descritivo de retenção, descritivo de evasão, ranking de cursos – dividido por cidade ou também descritivo – e gráfico de área por gênero. Tal assunto pode ser lido mais detalhadamente na seção 6.2.

6.1 Sugestões

Dentre as sugestões elaboradas pelos programadores para melhor coleta e armazenamento dos dados do CVU – de Bauru e de Ribeirão Preto –, é possível citar:

6.1.1 Documento de identificação

Durante o desenvolvimento do sistema, a não padronização e inconsistência em relação à solicitação do documento de identificação do voluntário – por vezes o documento de identidade (RG) era solicitado, em outras o documento de cadastro de pessoa física (CPF) e em outras nada era especificado – trouxe a impossibilidade de utilizar em todos os registros uma função de análise de recorrência de voluntariado por meio de algum desses documentos.

Sendo assim, uma das medidas resolutivas propostas pelos programadores aos responsáveis pelo CVU é a implementação de um padrão acerca dessa informação do voluntariado: é de extrema importância que haja um documento que identifique cada um dos participantes das ações do CVU e que tal documento seja sempre padronizado, ou seja, que sempre

seja solicitado o mesmo documento do voluntário no cadastro deste perante determinada ação a ser desenvolvida.

Por outro lado, vale a pena mencionar o fato de que seria mais vantajoso à pesquisa desenvolvida através do sistema que fosse solicitado o documento de cadastro de pessoa física (CPF) ao voluntário, pois esse documento é emitido pela Receita Federal do Brasil, ou seja, válido e padronizado em todo território nacional; diferentemente do documento de identidade (RG), que é emitido pelos órgãos de segurança dos Estados da Federação e pelo Distrito Federal, podendo, assim, sofrer alterações e despadronização dependendo do estado de emissão, além de permitir repetição de números entre os órgãos emissores.

Outro ponto importante acerca da solicitação dessa informação é a possibilidade de implementar uma investigação da idade com base no documento fornecido, ou seja, através do número do CPF do voluntário é factível inferir sua idade e, portanto, realizar novas análises e elaborar novos gráficos focando criticamente nessa informação como um todo – ou seja, descritivamente – ou por ação elaborada pelo CVU. Em contrapartida, se a data de nascimento do voluntário for solicitada – como melhor informado no item 6.1.4 –, não há mais vantagem do documento para inferir a idade de cada participante.

Por meio do número do RG essa mesma funcionalidade é praticável apenas se também for disponibilizado aos programadores o estado de emissão do documento. Sendo assim, a implementação dessa funcionalidade fica também como sugestão à possíveis projetos baseados na continuação desse trabalho.

6.1.2 Solicitar gênero

A fim de garantir maior assertividade em relação ao gênero de cada voluntário para as análises feitas pelo software a partir dessa informação, seria interessante solicitar, durante o cadastro do mesmo na ação, o gênero com o qual se identifica.

Tal informação não deve ser de preenchimento livre pelo voluntário para evitar despadronização das respostas fornecidas, ou seja, deve ser apresentada uma lista previamente elaborada pelos responsáveis pelo CVU com as possíveis escolhas do voluntário para responder essa questão.

Tal atitude enriqueceria os dados coletados, além de tornar obsoleta a utilização da biblioteca *Genderize* (DEMOGRAFIX APS, 2014) – descrita na seção 4.5 – nesse sistema e minar qualquer margem de erro em relação à cisgeneridade ou também causada pelo eventual mau funcionamento da biblioteca citada.

6.1.3 Solicitar nome

Por diversos motivos, sendo um deles a implementação da análise de recorrência, se faz necessário que sempre haja a informação relacionada ao nome completo de cada voluntário participante das ações do CVU.

Dessa forma, outra medida resolutiva proposta pelos desenvolvedores do sistema aos responsáveis pelo CVU é a padronização também envolvendo os nomes dos voluntários: sempre solicitar o nome original completo nos cadastros das ações. Em casos de apresentação de nome social, tal informação deve também ser apresentada.

Dados relacionados ao gênero e a algum documento de identificação, seja ele o CPF ou RG, de cada voluntário participante das ações desenvolvidas também se fazem importantes e, se presentes dentre os dados fornecidos pela instituição parceira, torna desnecessária a solicitação do nome do voluntário – uma vez que o nome é utilizado apenas para inferir seu gênero e também tornar possível a análise de recorrência.

6.1.4 Solicitar data de nascimento

Devido à possibilidade de implementações futuras no software – sejam elas por meio dos desenvolvedores desse trabalho ou demais desenvolvedores que se interessarem em dar continuidade a este trabalho – seria interessante também que os responsáveis pelo CVU solicitassem, durante o cadastro dos voluntários nas ações a serem promovidas, a data de nascimento dos mesmos.

Se tal informação estivesse contida dentre os dados coletados, enriqueceria a pesquisa de modo que se tornaria possível implementar novas análises por meio da idade do voluntariado como, por exemplo: análise da idade média do voluntariado por ação, maior recorrência de idade dentre os voluntários, gráficos elaborados sobre a idade dos voluntários participantes, dentre outras.

6.1.5 Pesquisa de satisfação

Mais uma medida resolutiva importante a ser proposta aos responsáveis pelo CVU é a elaboração de um formulário de pesquisa de satisfação, ou seja, assim como é feito o cadastro dos voluntários participantes de cada ação através do preenchimento de formulários no *Google Forms* (GOOGLE, 2019a), os desenvolvedores desse sistema propõem que seja elaborado também um formulário para pesquisa de satisfação dos voluntários em relação a cada ação promovida.

O preenchimento de tal formulário deve acontecer posteriormente à realização da ação, pessoalmente – se for de interesse dos responsáveis imprimir uma cópia do questionário presente no formulário por voluntário participante – ou através do próprio *Google Forms* (GOOGLE, 2019a). Tais registros de feedback darão maior direcionamento aos responsáveis pela ONG para promover ações mais assertivas ao gosto do voluntariado e eliminar possíveis recorrências de problemas.

Foi elaborado pelos desenvolvedores desse trabalho um exemplo de como seria tal formulário e quais informações deveriam ser requisitadas para a pesquisa, além daquelas necessárias e escolhidas pelos responsáveis do CVU e suas estratégias. Dentre os campos importantes para os desenvolvedores estão: CPF, idade e gênero. Por outro lado, se tais

informações forem fornecidas no formulário de cadastro dos voluntários, não se faz necessário as mesmas serem requisitadas no questionário de feedback. Da mesma forma, algumas questões interessantes como meio de feedback ao CVU são:

- “Como você qualificaria sua experiência participando da ação?”;
- “Essa foi a sua primeira vez participando de uma ação do CVU?”;
- “Na sua opinião, qual seria a chance de você participar de outra ação do CVU no futuro?”;
- “O que te motivou a participar da ação?”.

6.1.6 Formatação padrão dos dados

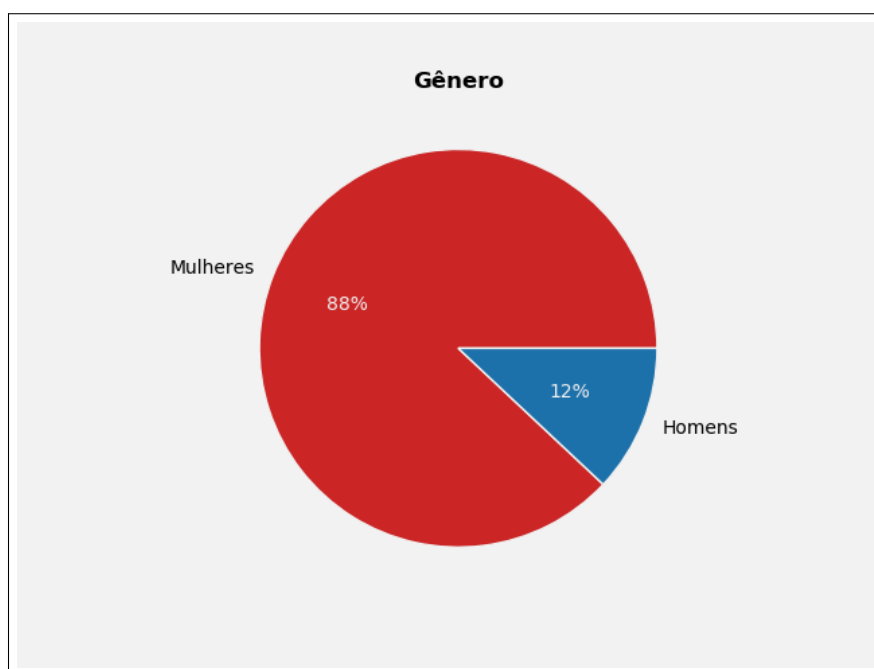
Por fim, tal medida resolutiva diz respeito à padronização da solicitação de informações pelos responsáveis pelo CVU acerca dos seus voluntários. Os desenvolvedores sugerem a continuidade da obtenção dos dados através de formulários preenchidos no *Google Forms* (GOOGLE, 2019a) e colhidos através do *Google Sheets* (GOOGLE, 2019b), porém com a ressalva de que esses dados devem seguir sempre uma mesma padronização em relação à sua ordem e disposição nos formulários, a fim de evitar problemas como os encontrados durante o desenvolvimento desse sistema, por exemplo, além de possíveis problemas futuros envolvendo os dados coletados. Essa medida resolutiva também se enquadra no contexto dos formulários envolvendo pesquisas de satisfação.

6.2 Gráficos

6.2.1 Setores por gênero

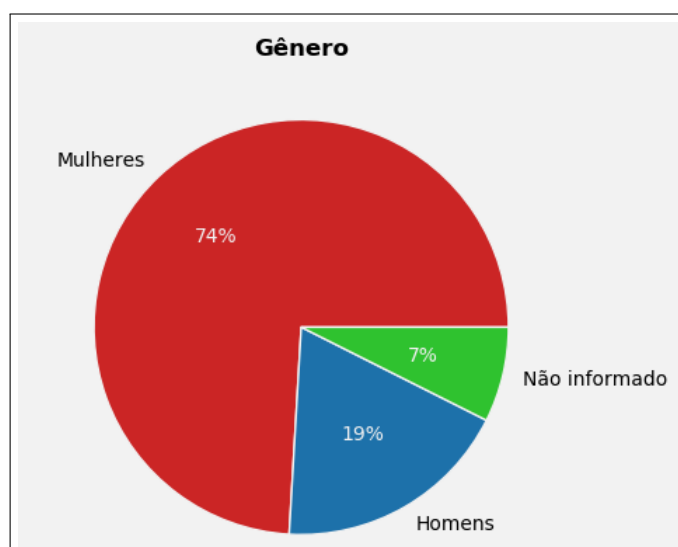
A seguir, como podemos observar através das figuras 8 e 9, estão exemplos de gráficos provenientes dos dados relacionados aos gêneros dos voluntários do CVU – de Bauru e de Ribeirão Preto –, respectivamente.

Figura 8 – Gráfico de setores por gênero do CVU de Bauru sobre a ação de natal em 2018.



Fonte: elaborada pelos autores.

Figura 9 – Gráfico de setores por gênero do CVU de Ribeirão Preto sobre a ação de “bem estar” em 2017.



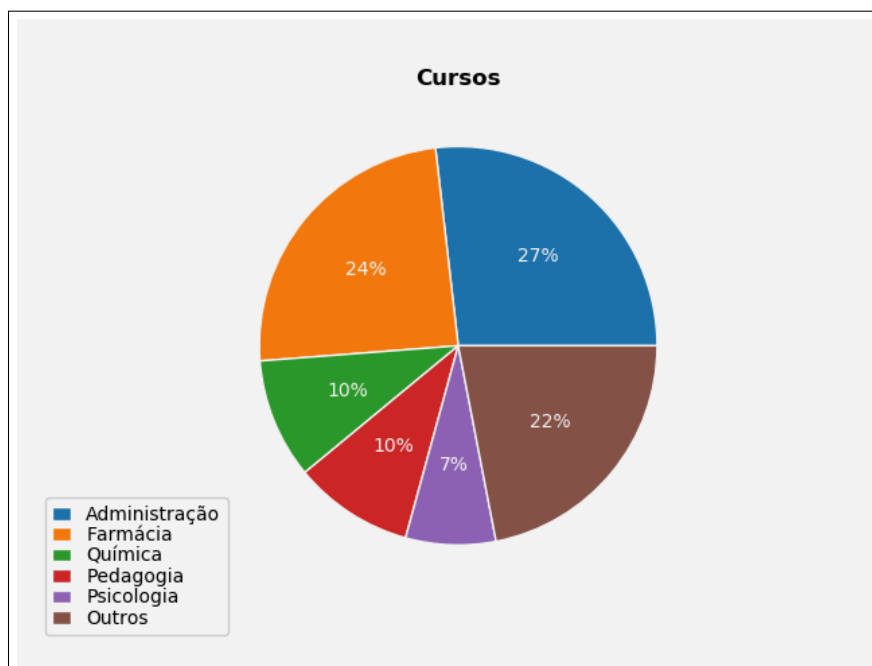
Fonte: elaborada pelos autores.

Sobre os gráficos de cursos, é possível notar a existência de um número fixo e pequeno de classificações – masculino, feminino e não informado. Dentre as ações analisadas, nem todas apresentam o gênero “não informado”, sendo a ocorrência desses casos justificada pela inexistência dessa informação dentre os dados coletados e/ou pela dificuldade da biblioteca *Genderize* em inferir o gênero com base no nome do voluntário.

6.2.2 Setores por curso

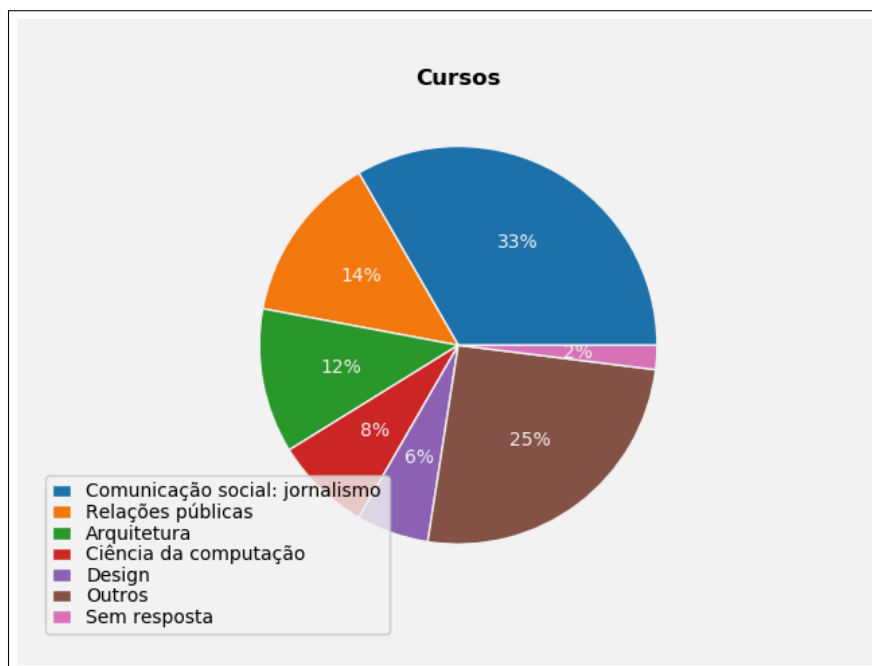
A seguir, como podemos observar através das figuras 10 e 11, estão exemplos de gráficos provenientes dos dados relacionados aos cursos dos voluntários do CVU – de Ribeirão Preto e de Bauru –, respectivamente.

Figura 10 – Gráfico de setores por curso do CVU de Ribeirão Preto da ação de trote social em 2019.



Fonte: elaborada pelos autores.

Figura 11 – Gráfico descritivo anual de setores por curso do CVU de Bauru do ano 2018.



Fonte: elaborada pelos autores.

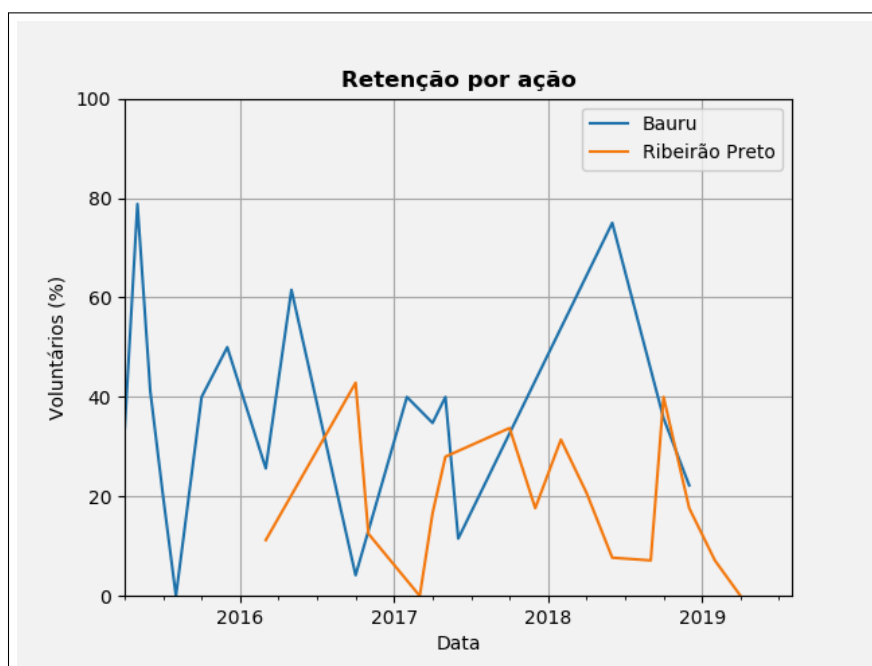
É possível observar a presença de setores que não são cursos propriamente ditos. Isso foi necessário em função da vasta quantidade de voluntários com essas informações diferentes. Por isso, o gráfico se limita a apresentar apenas cinco cursos, agrupando o restante num grupo denominado “outros”.

Além disso, também foi necessário o uso de outros setores como “sem resposta” – para planilhas onde o curso não era de preenchimento obrigatório – e “não universitário” – para ações que permitem a participação do público em geral.

6.2.3 Retenção

A Figura 12 apresenta o gráfico de linhas gerado à partir da análise de retenção envolvendo os voluntários do CVU de Ribeirão Preto e de Bauru.

Figura 12 – Gráfico comparativo proveniente dos dados relacionados à análise de retenção envolvendo o CVU de Bauru e de Ribeirão Preto perante os anos analisados.



Fonte: elaborada pelos autores.

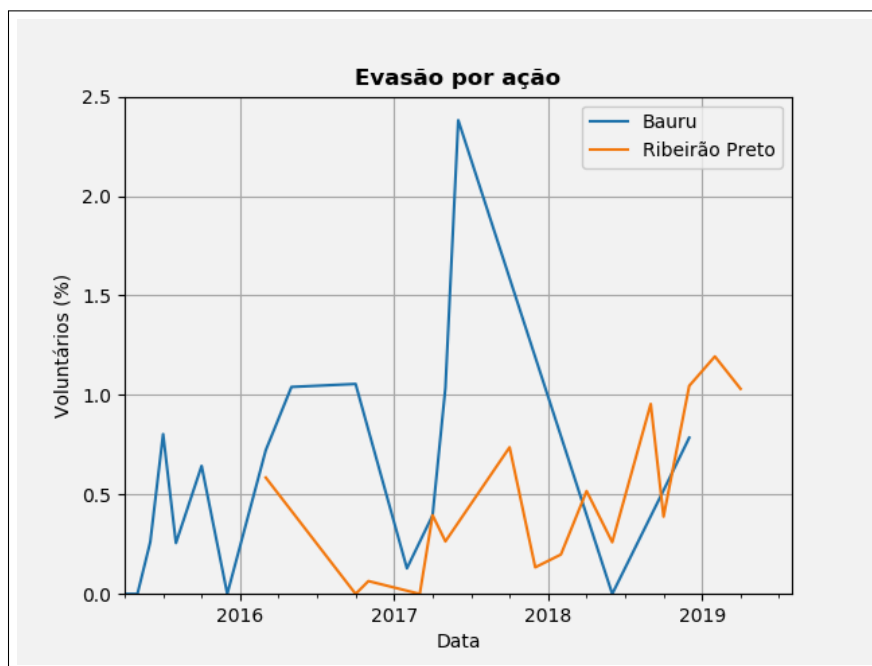
Nesse gráfico, podemos observar alguns picos. Entretanto, eles não são tão relevantes quanto parecem. Isso porque o primeiro pico se dá logo no início do período analisado, enquanto a instituição já trabalhava antes de começar a armazenar dados, representando então um acúmulo de retenções anteriores.

Os demais picos apresentados estão ligados a ações com números baixos de voluntários totais, aumentando a variação de retenção, justamente por ser apresentada em porcentagem. Para evitar esse problema, foi identificado que talvez seja melhor organizar essas mesmas informações através de um gráfico de densidade.

6.2.4 Evasão

A Figura 13 mostra o gráfico de linhas gerado à partir da análise de evasão envolvendo os voluntários do CVU de Ribeirão Preto e de Bauru.

Figura 13 – Gráfico comparativo proveniente dos dados relacionados à análise de evasão envolvendo o CVU de Bauru e de Ribeirão Preto perante os anos analisados.



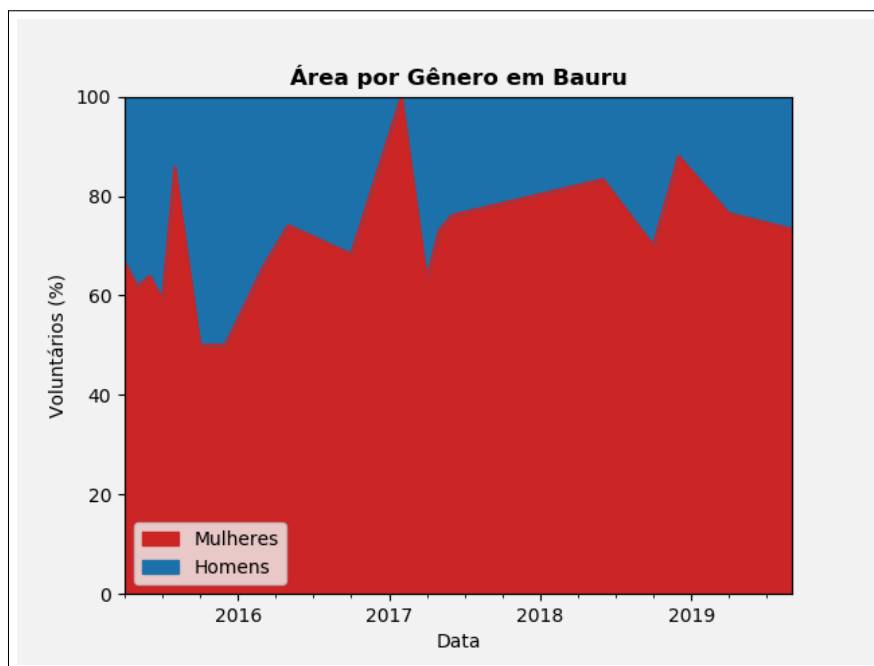
Fonte: elaborada pelos autores.

Considerando os dados apresentados pelos dois gráficos – o de retenção e o de evasão –, é possível observar que, em geral, o CVU tem um alto índice de recorrência dentre seus voluntários.

6.2.5 Área por gênero

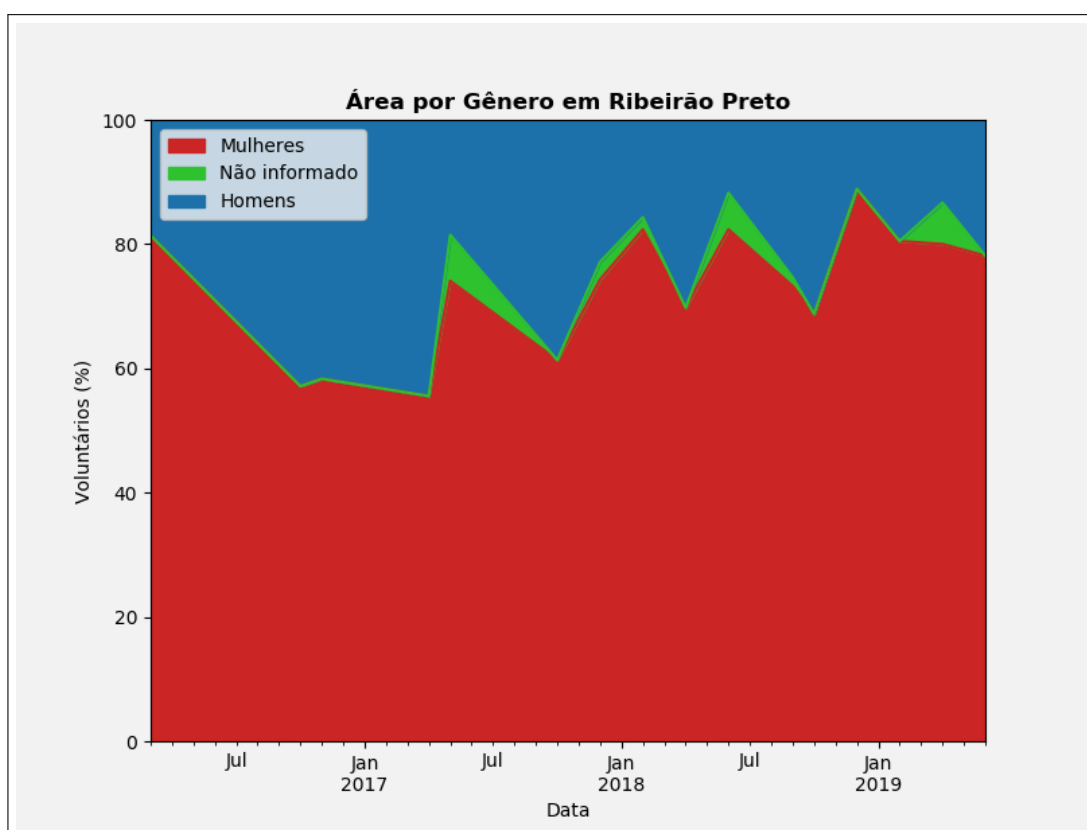
A seguir, as Figuras 14 e 15 mostram os gráficos de área gerados à partir da análise de gêneros envolvendo os voluntários do CVU de Bauru e de Ribeirão Preto, respectivamente, contabilizando todos os dados do período analisado pelo sistema.

Figura 14 – Gráfico de área por gênero do CVU de Bauru.



Fonte: elaborada pelos autores.

Figura 15 – Gráfico de área por gênero do CVU de Ribeirão Preto.



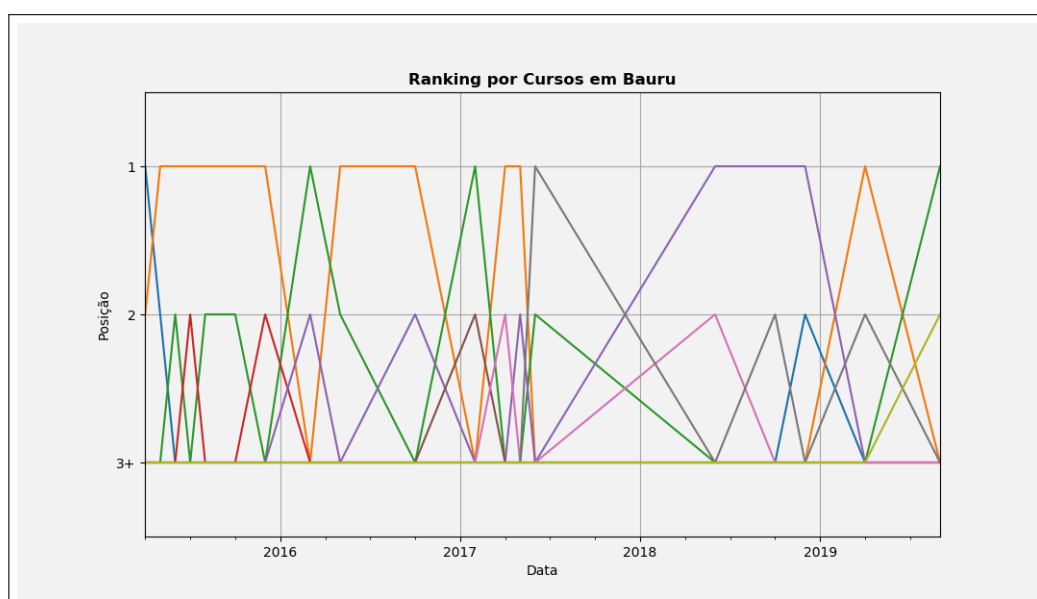
Fonte: elaborada pelos autores.

Através da análise dos gráficos de setores e de área gerados, é possível concluir que o CVU – tanto de Bauru quanto de Ribeirão Preto – tem maior incidência de voluntários do gênero feminino.

6.2.6 Ranking por curso

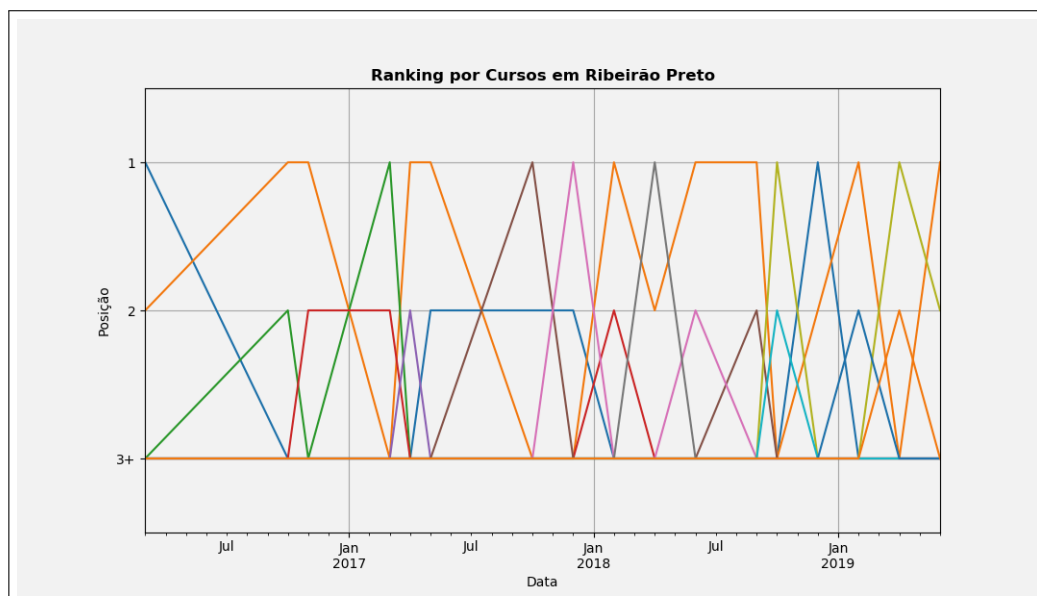
Nas Figuras 16 e 17 estão os gráficos de ranking gerados à partir da análise de cursos envolvendo os voluntários do CVU de Bauru e de Ribeirão Preto, respectivamente, contabilizando todos os dados do período analisado pelo sistema.

Figura 16 – Gráfico de ranking por curso do CVU de Bauru.



Fonte: elaborada pelos autores.

Figura 17 – Gráfico de ranking por curso do CVU de Ribeirão Preto.



Fonte: elaborada pelos autores.

Devido à grande rotatividade de voluntários nas ações desenvolvidas pelo CVU, há também uma abrangente amostra em relação aos cursos presentes nas ações; tal característica faz com que o gráfico de ranking por cursos seja desproporcional se comparado com os demais e, portanto, irrelevante à pesquisa.

7 DIFICULDADES ENCONTRADAS

No decorrer do trabalho, surgiram diversas dificuldades com as quais os desenvolvedores tiveram que lidar prontamente. Durante o estabelecimento de soluções para esses problemas, foram elaboradas medidas resolutivas como uma forma de sugestão aos responsáveis pelo CVU, tanto de Bauru quanto de Ribeirão Preto, em relação a como melhor documentar e organizar seus dados.

Um dos desafios principais aconteceu envolvendo a coleta de dados necessários para a realização da pesquisa: a ONG parceira – tanto a sede situada em Bauru quanto a sede em Ribeirão Preto – é composta por estudantes universitários que se voluntariam para colaborar com a causa que acreditam; sendo assim, o tempo destinado à assuntos voltados à organização por muitas vezes é corrido e atribuído a janelas de horários de aulas.

Dessa forma, houve empecilhos para obter os dados relacionados às ações de forma rápida – em algumas situações os desenvolvedores ficaram por dias aguardando retorno da organização –, o que tomou um certo tempo e preocupou os envolvidos em relação ao cronograma.

7.1 CVU Bauru

Outro ponto relevante a ser comentado ocorreu durante o período após a coleta de dados do CVU de Bauru: vinte e oito planilhas foram disponibilizadas para análise e nelas estavam contidos os dados relacionados aos 281 voluntários válidos que participaram de cada ação desenvolvida pela instituição em questão, desde o ano de 2015 até o ano atual, 2019.

Porém, apesar do número relativamente grande de documentos, não foram todos os dados aproveitados nas análises; isso porque alguns estavam apenas parcialmente preenchidos (com algumas informações relevantes ignoradas pelo voluntário que preencheu o documento), ou fora do padrão dos demais, ou apenas continham informações que não interessavam aos cientistas de dados. Sendo assim, tal situação provocou um afunilamento das informações do voluntariado disponibilizadas pela instituição parceira.

Durante a fase de descoberta de possíveis dados a serem minerados, surgiu uma contestação pelos desenvolvedores do trabalho em torno da possibilidade de estimar a idade aproximada de cada voluntário a partir do seu número de documento, fosse ele o documento de identidade (RG) ou o cadastro de pessoa física (CPF). Entretanto, não foi dada continuidade nessa tratativa em virtude de dois problemas:

1. O número do documento de identidade (RG) não é único nacionalmente, mas sim por estado. Apesar das sedes da instituição estarem localizadas no estado de São Paulo, não é interessante inserir essa limitação no software. Ademais, esse fato não garante que os voluntários emitiram seus documentos nesse estado específico, apesar da alta probabilidade.

2. O número do documento do cadastro de pessoa física (CPF) é pouco recorrente nos dados colhidos, o que faria com que o custo-benefício de tal implementação fosse irrelevante.

Além do que já foi mencionado em relação às dificuldades em estabelecer padrões com os dados disponíveis, os responsáveis pelo sistema notaram uma particularidade relevante durante a análise dos mesmos: dos anos de 2015 a 2017 a identificação do voluntário era feita por meio do documento de registro de identidade (RG); posteriormente a essa data, nos anos de 2018 e 2019, essa mesma identificação passou a ser feita através do documento de cadastro de pessoa física (CPF). Essa especificidade dificultou o trabalho dos desenvolvedores, principalmente em relação a determinação da recorrência entre os voluntários quanto à participação nas ações solidárias do CVU de Bauru.

Sendo assim, alguns dos riscos previamente identificados foram comprovados: a dificuldade em relação à dependência e disponibilidade dos indivíduos que compõe a ONG parceira, bem como a insegurança sobre o formato, consistência, tamanho e relevância dos dados apresentados para análise.

7.2 CVU Ribeirão Preto

O processamento dos dados da instituição parceira CVU localizada na cidade de Ribeirão Preto foi iniciado posteriormente ao desenvolvimento de algumas análises referentes ao CVU localizado em Bauru. Sendo assim, dificuldades citadas anteriormente já haviam sido ultrapassadas quando os desenvolvedores começaram a analisar e organizar os dados do CVU de Ribeirão Preto. Porém, novos desafios surgiram.

Assim como aconteceu com os dados referentes ao CVU de Bauru, apesar do número relativamente grande de ações desenvolvidas pelo CVU de Ribeirão Preto, não eram todas elas que possuíam os voluntários participantes organizados em algum documento.

De qualquer forma, após a coleta dos documentos válidos para a análise de dados, foram incorporadas no trabalho vinte e três novas planilhas que continham informações referentes ao voluntariado participante das ações do CVU de Ribeirão Preto, desde o ano de 2016 até o ano atual, 2019. Sendo assim, apesar da falta de documentação relacionada à algumas ações, após a compilação de todos esses novos documentos, os desenvolvedores adicionaram mais 494 voluntários na análise, totalizando 775 voluntários, se somados aos 281 já existentes nos dados do CVU de Bauru.

Uma particularidade referente ao CVU de Ribeirão Preto observada foi a presença de voluntários não universitários em suas ações desenvolvidas. De acordo com o próprio nome da instituição, Centro de Voluntariado Universitário (CVU, 2019), a ONG tem o objetivo principal de incentivar e promover ações de voluntariado em universidades. Porém, no caso do CVU de Ribeirão Preto, voluntários da comunidade da cidade também são bem-vindos a participarem das ações desenvolvidas. O mesmo, por outro lado, não acontece no CVU de Bauru, onde todos os

voluntários são, de fato, universitários. Tal peculiaridade do CVU de Ribeirão Preto trouxe aos desenvolvedores uma dificuldade em acrescentar esses voluntários na geração de gráficos por cursos, já que essa informação não condizia com o voluntário a ser analisado. Para contornar essa circunstância, foi criada uma nova categoria no gráfico de setores agrupados por cursos chamada de “Não universitário”, onde todos esses casos foram contabilizados.

Também houve casos onde, apesar de universitário, a informação sobre o curso do voluntário não era totalmente especificada: em alguns registros era esclarecida apenas sua faculdade, em outros nada era informado. Dessa forma, foi criada mais uma categoria no gráfico de setores agrupados por curso chamada de ‘Universitário: curso não informado’ para abranger todos os casos caracterizados pelas situações mencionadas.

Sobre as dificuldades em estabelecer padrões com os dados disponíveis envolvendo o CVU de Ribeirão Preto, os desenvolvedores se depararam com uma particularidade específica em relação à ausência total de documento de identidade (RG) ou de documento de cadastro de pessoa física (CPF) dos voluntários.

A princípio, existe a questão ética: os responsáveis pelo CVU de Ribeirão Preto, dessa forma, não possuem qualquer tipo de identificação pessoal legal da cidadã ou cidadão que se voluntaria em suas ações que traga algum tipo de legitimidade à sua participação e, logo, nenhum tipo de garantia civil à adesão e bom cumprimento do trabalho voluntário conforme previsto pela lei de número 9.608 inicialmente vigorada em 18 de Fevereiro de 1998 onde é dito pelo Art. 2º: “O serviço voluntário será exercido mediante a celebração de termo de adesão entre a entidade, pública ou privada, e o prestador do serviço voluntário, dele devendo constar o objeto e as condições de seu exercício” (CASA CIVIL, 2016).

Além disso, houve a limitação técnica aos desenvolvedores: esse caso trouxe a impossibilidade de utilizar em todos os dados coletados uma função de análise de recorrência de voluntariado através de algum dos documentos citados anteriormente (RG e/ou CPF), o que traz um impacto maior tendo em vista a complexidade de se estabelecer padrões em nomes escritos por extenso de forma livre (nome completo do voluntário) em comparação ao uso de qualquer documento pessoal cuja probabilidade de sofrer alterações é menor, trazendo maior facilidade e credibilidade à análise de dados. Para contornar esse problema, realizou-se a análise de recorrência em duas etapas (por identidade de RG ou CPF e, em caso de inexistência desses dados do voluntário, por similaridade de nome) que foram apresentadas mais detalhadamente no item 5.2.3 desse texto.

A partir do desenvolvimento da análise de recorrência por nome, outro desafio foi a existência de registros de voluntários onde, ao invés do nome completo, havia disponível apenas o apelido do universitário. Além disso, houve casos onde, em apenas um registro, dois nomes eram descritos juntos. Tal acontecimento, além de dificultar a entrada desses voluntários na análise de recorrência, também impossibilitou a dedução do gênero dos mesmos através da biblioteca *Genderize* (DEMOGRAFIX APS, 2014). Sendo assim, devido à baixa taxa de assertividade desses casos, voluntários que se encaixavam nessas situações foram excluídos da

análise de dados.

7.3 Refatoração de código

Durante o mês de setembro realizou-se uma refatoração no código do programa, o que consumiu certo tempo. O objetivo dessa refatoração foi ampliar o índice de aproveitamento dos dados de entrada.

Até então, a ordem de processamento de dados acontecia de forma que a análise de retenção era feita por último. Após a refatoração, essa ordem foi alterada e a análise de retenção passou a ser feita primeiro.

O resultado prático dessa refatoração se dá pela possibilidade de armazenar uma estrutura de dados de voluntários, além da estrutura de dados de ações. Assim, é possível reaproveitar dados de voluntários recorrentes em registros onde há dados vagos.

Por exemplo: determinado voluntário participou da ação X, onde preencheu no formulário que cursa psicologia. Ao participar de uma segunda ação (Y), o curso não estava na lista de dados do voluntário a serem preenchidos no formulário. A partir dessa estrutura de dados de voluntários, é possível reaproveitar o curso preenchido no formulário da ação X como dado para a ação Y. Isso aumenta a taxa de descoberta do trabalho ao preencher lacunas dos dados de entrada.

Além disso, também realizou-se uma refatoração na parte de requisições da API *Genderize*. Como a API tem um limite de 1000 requisições por dia na versão gratuita e esse trabalho lida com um rol de quase 800 voluntários, foi necessário fazer um uso mais consciente das requisições. Isso ocorreu de três maneiras:

- Uso do lote de até 10 nomes por requisição, que a API oferece;
- Reaproveitamento do resultado para pessoas com o mesmo nome;
- Uso de um arquivo local que salva o resultado das requisições para um próximo uso.

Essas alterações não só solucionaram o problema do limite de requisições, como também aumentaram a performance do programa como um todo, visto que as requisições da biblioteca compunham a maior parte do tempo de processamento, justamente por serem feitas via internet.

8 TRABALHOS FUTUROS

Como recomendações aos desenvolvedores que tenham a intenção de elaborar possíveis projetos futuros e dar continuidade neste trabalho, algumas sugestões foram reunidas:

8.1 Interface de usuário

Como vantagem identificada da solução mencionada no Capítulo 2, a interface de usuário seria um grande passo no sentido de continuidade desse trabalho. O intuito é permitir que o usuário faça a entrada dos dados – tanto em relação a novas ações desenvolvidas, quanto de novos voluntários participantes ou todo e qualquer tipo de dado disponibilizado pela instituição parceira CVU –, além de poder gerar saídas de dados – como gráficos em tempo real com filtros customizáveis e painéis de visualização dos gráficos e análises geradas.

8.2 Gráficos de ranking por área de conhecimento

Como apresentado no item 6.2.6, devido à grande rotatividade de voluntários nas ações desenvolvidas pelos CVU's, os cursos analisados pelo sistema desenvolvido apresentaram uma vasta amostra. Dessa forma, o gráfico de ranking por cursos tornou-se irrelevante à pesquisa e impossível de inferir algum resultado significativo à análise de dados.

Portanto, como sugestão para desenvolvimentos futuros está também a criação de um gráfico de ranking de cursos, porém por área de conhecimento, ou seja, propõe-se a divisão dos cursos já cadastrados em áreas de ciências humanas, ciências da natureza, linguagens e códigos ou ciências exatas e, posteriormente, a elaboração de gráficos de linha sob tais divisões. Assim, espera-se resultados mais significativos e que possam dar um direcionamento aos responsáveis pelo CVU acerca da área que mais participa de suas ações.

Para desenvolver esse gráfico, será necessário categorizar os cursos presentes no dicionário do código fonte dentre as áreas de conhecimento já mencionadas.

8.3 Gráficos de densidade de retenção e evasão

Os gráficos de recorrência foram insatisfatórios em consequência dos seus picos em função de ações com baixo número de voluntários. Para mitigar esse problema, seria interessante que esses gráficos de linha fossem substituídos por gráficos de densidade, assim adicionando o número de voluntários como uma variável a mais para visualização.

8.4 Projeções

Todos os gráficos desenvolvidos nesse trabalho representam dados já existentes. Como sugestão de continuidade, seria interessante o desenvolvimento de gráficos de projeção, ou seja, previsão de indicadores futuros baseados nos indicadores passados.

8.5 Correlações aprofundadas

As análises desse trabalho não fazem correlações profundas entre os dados disponíveis, apenas focam em cada dado individualmente, como prova de conceito. Para trabalhos futuros, a sugestão é que esses dados sejam correlacionados a ponto de garantir novos indicadores a partir dos já existentes.

8.6 Feedback

Através da pesquisa de satisfação que foi desenvolvida para o CVU como parte desse trabalho, é possível utilizar as respostas como novos dados de entrada para o sistema desenvolvido, assim criando novas oportunidades para o desenvolvimento de indicadores a partir desses dados.

9 CONCLUSÃO

Esse trabalho concentra-se na área de ciência de dados – importante extensão multidisciplinar que promove a coletânea e processamento de dados a fim de auxiliar na tomada de decisões de instituições, planejamentos econômicos, estatísticos e de mercado, dentre outras finalidades.

Sendo assim, em parceria com a ONG Centro de Voluntariado Universitário (CVU, 2019) – mais especificamente com as sedes de Bauru e Ribeirão Preto –, esse sistema traz como objetivo principal gerar insumos – através do estabelecimento de padrões e correlações entre os dados analisados – que tragam benefícios, aos responsáveis pela ONG, relacionados à melhor organização e processamento de suas informações sobre o voluntariado. Dessa forma, através de gráficos e análises descritivas, o intuito é que, com o auxílio desse sistema, o CVU consiga aperfeiçoar suas ações a fim de conseguir cada vez mais aumentar a participação popular e atingir pessoas em situações carentes, além de melhorar seus processos internos.

Feita pesquisa em busca de soluções cujos objetivos podem ser considerados similares aos propostos pelo sistema deste trabalho, foi encontrado o programa CoCo Analytics (Connect Online Connect Offline), da Digital Green. Tal plataforma realiza a coleta e análise de dados através de um painel de cadastro e visualização *online*. Em comparação com o sistema desenvolvido neste trabalho, o CoCo Analytics possui vantagens e desvantagens em relação à sua utilização pelos usuários. Dentre as vantagens que o CoCo Analytics fornece perante ao sistema deste trabalho, pode-se citar o alto nível de interação possível do usuário com o sistema – devido ao painel de manipulação de dados – e a maior acessibilidade – por estar hospedado em um *website*. Em relação às desvantagens, tem-se o código fechado do CoCo Analytics – ou seja, o fato do código fonte da plataforma não estar disponível a todos que possuam interesse em conhecê-lo –, a necessidade de cadastro na plataforma para possibilitar sua utilização e a restrição dessa utilização apenas para ONGs asiáticas.

Vale a pena mencionar que, para a linha de pesquisa utilizada nesse trabalho, quanto maior a qualidade das informações analisadas, bem como seu volume e relevância, mais assertivos são os resultados. Tal verdade tornou-se um risco para o bom desenvolvimento do sistema como um todo. Portanto, dentre os diversos desafios encontrados pelos desenvolvedores no decorrer da implementação, a maioria foi acerca dessa qualidade de informações: os dados obtidos através do CVU eram bidimensionalmente escassos, além de despadronizados e flutuantes (algumas informações contidas em certas planilhas não estavam contidas também em outras); houve também a falta de uma informação identificadora, ou seja, um dado que identificasse cada voluntário individualmente nos registros das ações com o passar dos anos; dentre outras questões. Além das adversidades já comentadas, existiram dificuldades técnicas enfrentadas pelos desenvolvedores do sistema, como o não domínio das linguagens, ferramentas e tecnologias que foram utilizadas, além do tempo relativamente curto para o desenvolvimento de um trabalho complexo, dentre outras.

Após o desenvolvimento e implantação de todo o sistema, para de garantir boa execução

e estabilidade, os desenvolvedores do projeto necessitaram de uma maneira de validar seu funcionamento. Dessa forma, tal atividade foi exercida através das planilhas contendo os dados do voluntariado disponibilizadas pela instituição parceira CVU – Bauru e Ribeirão Preto – ao utilizarem esses dados como informação de entrada no sistema.

A fim de melhorar a organização e armazenamento dos dados coletados pelo CVU, foram elaboradas algumas sugestões como, por exemplo: realizar pequenas alterações no questionário de cadastro do voluntário na ação a ser desenvolvida como, por exemplo, sempre solicitar documento de identificação do mesmo – seja ele o documento de identidade (RG) ou o cadastro de pessoa física (CPF) –, além de informações envolvendo o gênero, o nome e a data de nascimento de cada voluntário; elaborar um questionário de pesquisa de satisfação que deverá ser preenchido pelos participantes pós realização da ação; formatação padrão em qualquer forma de entrada de dados do sistema; elaboração de simples relatórios ou painéis de análise de informações, utilizando correlações encontradas e transformando tudo isso num fluxo de atividade; dentre outras funcionalidades.

Para os próximos anos, os responsáveis por esse sistema deixam em aberto a possibilidade de novas implementações através de outros desenvolvedores ou quaisquer interessados em continuar o trabalho feito até então. Algumas sugestões de novas funcionalidades aos possíveis trabalhos futuros são: criação de uma interface de usuário, elaboração do gráfico de ranking por área de conhecimento, modificação dos gráficos de retenção e evasão para o estilo de gráfico de densidade, elaboração de correlações mais aprofundadas e projeções futuras e utilização da pesquisa de satisfação como feedback à instituição e como forma de entrada de dados no sistema.

Isto posto, tem-se por conclusão desse trabalho uma prova de conceito: foram elaboradas mais sugestões do que indicadores propriamente ditos aos responsáveis pela ONG parceira e aos desenvolvedores de possíveis projetos futuros que tenham este como base, além de feita a padronização e utilização dos dados disponibilizados pela instituição parceira.

Referências

- CASA CIVIL. *Lei Nº 9.608, de 18 de Fevereiro de 1998*. 2016. Disponível em: http://www.planalto.gov.br/ccivil_03/leis/l9608.htm. Acesso em: 03 out. 2019.
- CVU. *Centro Voluntariado Universitário*. 2019. Disponível em: <http://www.cvu.org.br/>. Acesso em: 22 mar. 2019.
- DEMOGRAFIX APS. *Genderize*. 2014. Disponível em: <https://genderize.io/>. Acesso em: 06 out. 2019.
- DIGITAL GREEN. *CoCo Analytics*. 2019. Disponível em: <https://solutions.digitalgreen.org/coco/analytics/>. Acesso em: 22 mar. 2019.
- EVERYACTION. *5 Things You Need to Know about Nonprofits + Big Data*. 2016. Disponível em: <https://blog.everyaction.com/5-things-you-need-to-know-about-nonprofits-big-data/>. Acesso em: 20 jun. 2019.
- GITHUB INC. *GitHub*. 2019. Disponível em: <https://github.com/>. Acesso em: 16 jun. 2019.
- GOOGLE. *Google Forms*. 2019. Disponível em: <https://www.google.com/forms/about/>. Acesso em: 27 out. 2019.
- GOOGLE. *Google Sheets*. 2019. Disponível em: <https://www.google.com/sheets/about/>. Acesso em: 16 jun. 2019.
- LINUS TORVALDS. *Git*. 2019. Disponível em: <https://git-scm.com/>. Acesso em: 16 jun. 2019.
- MICROSOFT. *Visual Studio Code*. 2019. Disponível em: <https://code.visualstudio.com/>. Acesso em: 16 jun. 2019.
- OPENSOURCE.ORG. *Open Source Initiative*. 1998. Disponível em: <https://opensource.org/>. Acesso em: 07 out. 2019.
- PETENATE, M. *Por que é importante entender correlação entre variáveis*. 2016. Disponível em: <https://www.escolaedti.com.br/entender-correlacao-entre-variaveis/>. Acesso em: 26 jun. 2019.
- PYTHON SOFTWARE FOUNDATION. *Python*. 2019. Disponível em: <https://www.python.org/about/>. Acesso em: 05 jun. 2019.
- UNESP. *Universidade Estadual Paulista Júlio de Mesquita Filho*. 1976. Disponível em: <http://www2.unesp.br/>. Acesso em: 12 out. 2019.
- USP. *Universidade de São Paulo*. 1934. Disponível em: <http://www5.usp.br/>. Acesso em: 12 out. 2019.
- ZHOUYANG, L. *Python String Similarity*. 2018. Disponível em: <https://github.com/luozhouyang/python-string-similarity>. Acesso em: 19 out. 2019.