



# **Entrega da Solução Coronathon 2020**

## **Equipe Sinergia (Número 20)**

## Sumário

<b>1</b>	<b>INTRODUÇÃO</b>	<b>5</b>
<b>2</b>	<b>ENTENDIMENTO DO PROBLEMA E ANÁLISE DOS DADOS</b>	<b>8</b>
	<b>2.1 Nível de Escolaridade</b>	<b>8</b>
	2.1.1 Região Sudeste	8
	2.1.2 Região Sul	10
	2.1.3 Região Centro-Oeste	11
	2.1.4 Região Norte	13
	2.1.5 Região Nordeste	14
	<b>2.2 Dados faltantes</b>	<b>16</b>
<b>3</b>	<b>SOLUÇÃO PROPOSTA</b>	<b>18</b>
<b>4</b>	<b>MVP</b>	<b>21</b>
<b>5</b>	<b>CONCLUSÃO</b>	<b>29</b>
<b>6</b>	<b>APÊNDICE A – CÓDIGO DESENVOLVIDO EM PYTHON PARA ANÁLISE DOS</b>	
<b>DADOS</b>	<b>31</b>	
<b>7</b>	<b>BIBLIOGRAFIA, SITES E BASES DE DADOS</b>	<b>34</b>

**LISTA DE FIGURAS**

Figura 1 – Dados de desemprego no Brasil	6
Figura 2 – Perfil de escolaridade dos trabalhadores cadastrados no SINE: Região Sudeste	9
Figura 3 – Estudantes cadastrados no SINE: Região Sudeste	9
Figura 4 – Perfil de escolaridade dos trabalhadores cadastrados no SINE: Região Sul	10
Figura 5 – Estudantes cadastrados no SINE: Região Sul	11
Figura 6 – Perfil de escolaridade dos trabalhadores cadastrados no SINE: Região Centro-Oeste	12
Figura 7 – Estudantes cadastrados no SINE: Região Centro-Oeste	12
Figura 8 – Perfil de escolaridade dos trabalhadores cadastrados no SINE: Região Norte	13
Figura 9 – Estudantes cadastrados no SINE: Região Norte	14
Figura 10 – Perfil de escolaridade dos trabalhadores cadastrados no SINE: Região Nordeste	15
Figura 11 – Estudantes cadastrados no SINE: Região Nordeste	15
Figura 12 – Esquemático da solução	20
Figura 13 – Tela inicial	21
Figura 13 – Tela de dados	22
Figura 13 – Tela de opções de entrevista	23
Figura 13 – Tela de criação	24
Figura 13 – Tela de perguntas por vídeo	25
Figura 13 – Tela de perguntas por chat	26
Figura 13 – Tela de vagas para perfil vermelho	27
Figura 13 – Tela de vagas para perfil amarelo	28
Figura 13 – Tela de vagas para perfil verde	29



**LISTA DE TABELAS**

Tabela 1 – Porcentagem de cadastrados de cada região	8
Tabela 2 – Análise de dados faltantes	17

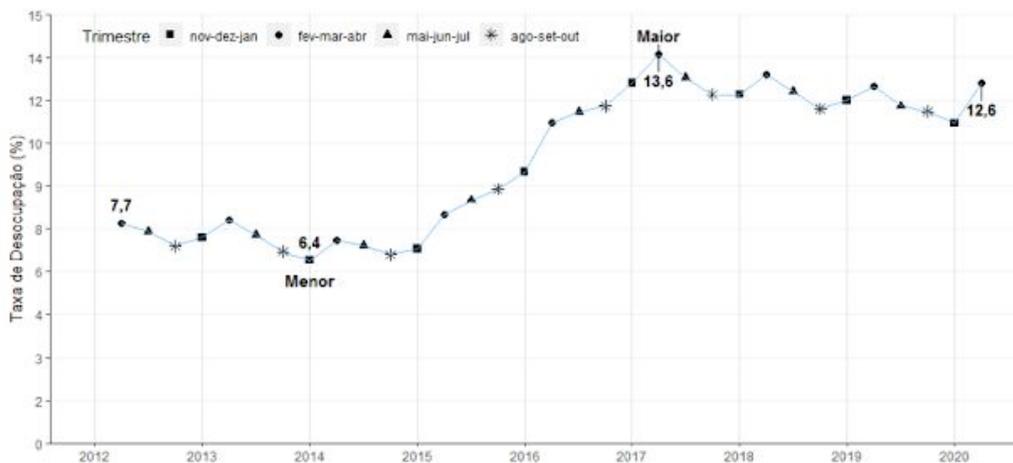
## 1 INTRODUÇÃO

O desemprego, de forma simplificada, se refere às pessoas com idade para trabalhar (acima de 14 anos) que não estão trabalhando, mas estão disponíveis e tentam encontrar trabalho. A PNAD Contínua é a pesquisa que mostra quantos desempregados há no Brasil. Nela, o que é conhecido popularmente como “desemprego” aparece no conceito de “desocupação” (IBGE, 2020).

Frequentemente divulgada na mídia, A taxa de desemprego ou desocupação, é obtida com base na PNAD Contínua e corresponde a porcentagem de pessoas na força de trabalho que estão desempregadas.

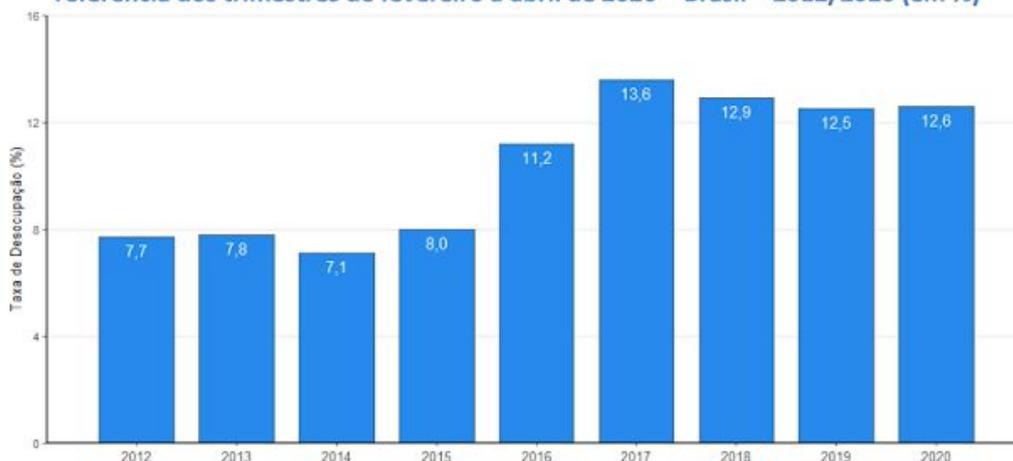
A série histórica da taxa de desocupação apresenta uma tendência de aumento a partir do ano de 2012, como pode ser observado na Figura 1.

**Gráfico 1 - Taxa de desocupação das pessoas de 14 anos ou mais de idade, na semana de referência, de todos os trimestres comparáveis - Brasil - 2012/2020 (em %)**



Fonte: IBGE, Diretoria de Pesquisas, Coordenação de Trabalho e Rendimento, Pesquisa Nacional por Amostra de Domicílios Contínua.

**Gráfico 2 - Taxa de desocupação das pessoas de 14 anos ou mais de idade, na semana de referência dos trimestres de fevereiro a abril de 2020 – Brasil – 2012/2020 (em %)**



Fonte: IBGE, Diretoria de Pesquisas, Coordenação de Trabalho e Rendimento, Pesquisa Nacional por Amostra de Domicílios Contínua.

**Figura 1 – Dados de desemprego no Brasil**

O advento da covid-19 no Brasil tem provocado um efeito devastador no mercado de trabalho. A taxa de desocupação foi estimada em 12,6% no trimestre móvel referente aos meses de fevereiro a abril de 2020, registrando variação de 1,3 ponto percentual em relação ao trimestre de novembro de 2019 a janeiro de 2020 (11,2%).

As medidas de distanciamento social impostas para se tentar limitar o avanço da doença também têm reduzido o nível de procura por trabalho, fazendo com que o desemprego pareça menor do que o apresentado nas estatísticas oficiais.

O quadro sintético do PNAD (anexo) aponta que os efeitos da pandemia foram mais sentidos no setor de Indústria (5,6%, ou menos 685 mil pessoas), Construção (13,1%, ou menos 885 mil pessoas), Comércio, reparação de veículos automotores e motocicletas (6,8%, ou menos 1,2 milhão de pessoas), Transporte, armazenagem e correio (4,9%, ou menos 242 mil pessoas), Alojamento e alimentação (12,4%, ou menos 699 mil pessoas), Outros serviços (7,2%, ou menos 366 mil pessoas) e Serviços domésticos (11,6%, ou menos 727 mil pessoas).



pnadc\_202004\_quad  
roSintetico.pdf

A partir do exposto acima, a equipe se reuniu e optou por trabalhar na solução do desafio de habilidades e competências: *“Como identificar e disponibilizar aos trabalhadores cadastrados no SINE informações sobre as qualificações exigidas pelas ocupações em alta?”*

## **2 ENTENDIMENTO DO PROBLEMA E ANÁLISE DOS DADOS**

A fim de entender melhor o problema, analisou-se a base de dados de trabalhadores do SINE. Para tanto, foi desenvolvido um código em Python, em que diversas características dos trabalhadores foram avaliadas, utilizando uma divisão por regiões do Brasil: Sudeste, Sul, Centro-Oeste, Norte e Nordeste.

OBS: Todos os códigos aqui utilizados podem ser encontrados no

APÊNDICE A – Código desenvolvido em Python para análise dos dados.

- Além disso, os arquivos de imagem e excel estão no seguinte link: [https://github.com/joabemsouza/Coronathon\\_BR](https://github.com/joabemsouza/Coronathon_BR).
- Contato para maiores dados: <https://www.linkedin.com/in/joabemsouza/>

**Tabela 1 – Porcentagem de cadastrados de cada região**

<b>Região</b>	<b>Porcentagem de trabalhadores cadastrados</b>
Sudeste	47.7 %
Sul	18.6 %
Centro-Oeste	10.2 %
Norte	6.1 %
Nordeste	17.4 %

## **2.1 Nível de Escolaridade**

### **2.1.1 Região Sudeste**

Os gráficos a seguir mostram que a maioria dos trabalhadores da região Sudeste cadastrados no SINE possuem ensino médio completo, ensino médio incompleto e fundamental completo. Um dado que chama a atenção é que uma porcentagem relativa dos trabalhadores não souberam ou não informaram qual a escolaridade. Em outras palavras, fica evidente que a procura maior pelo SINE é de um público com um nível de escolaridade menor, cada vez menos especializados e que não está estudando, conforme Figura 3.

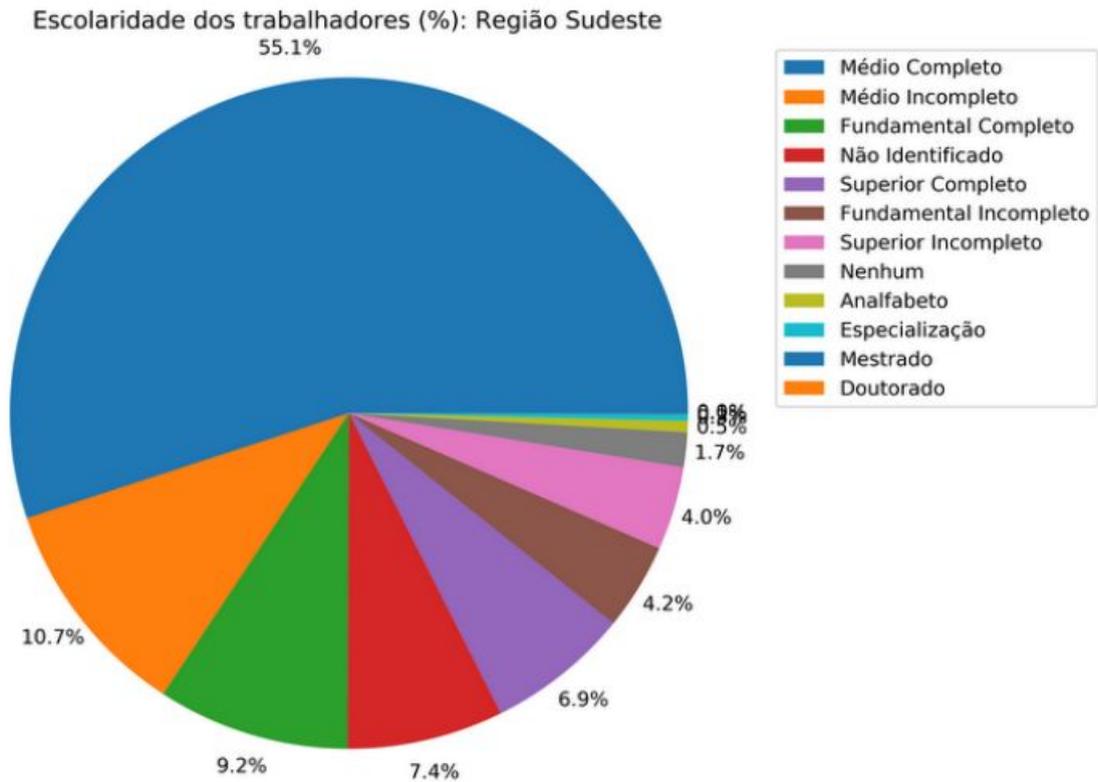
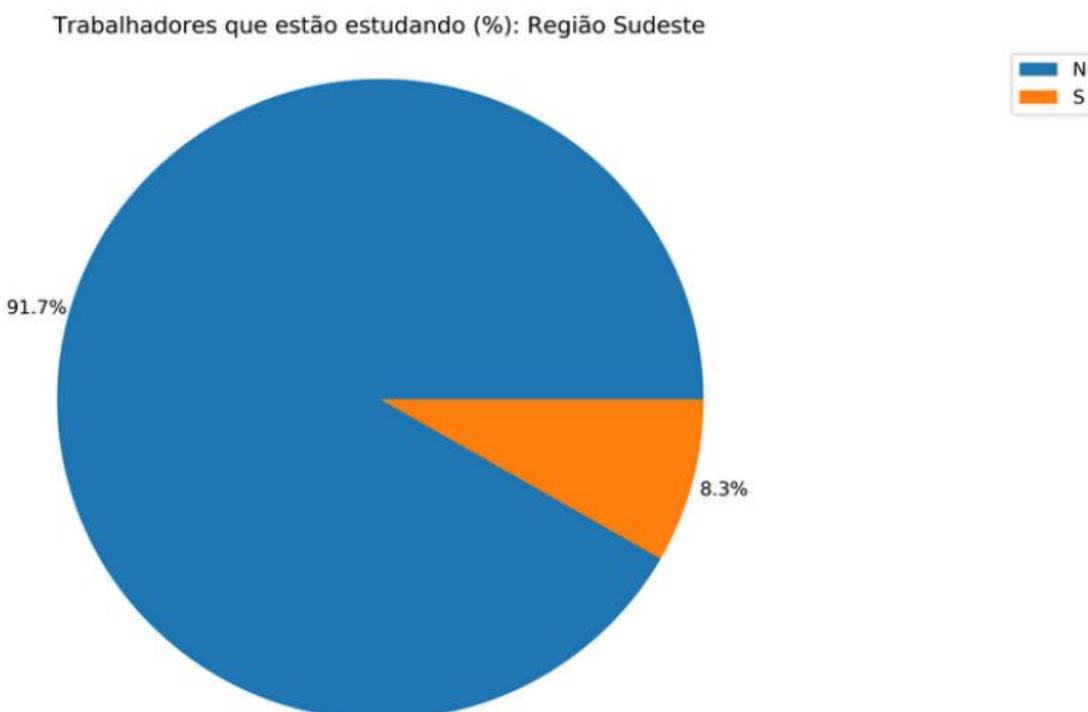


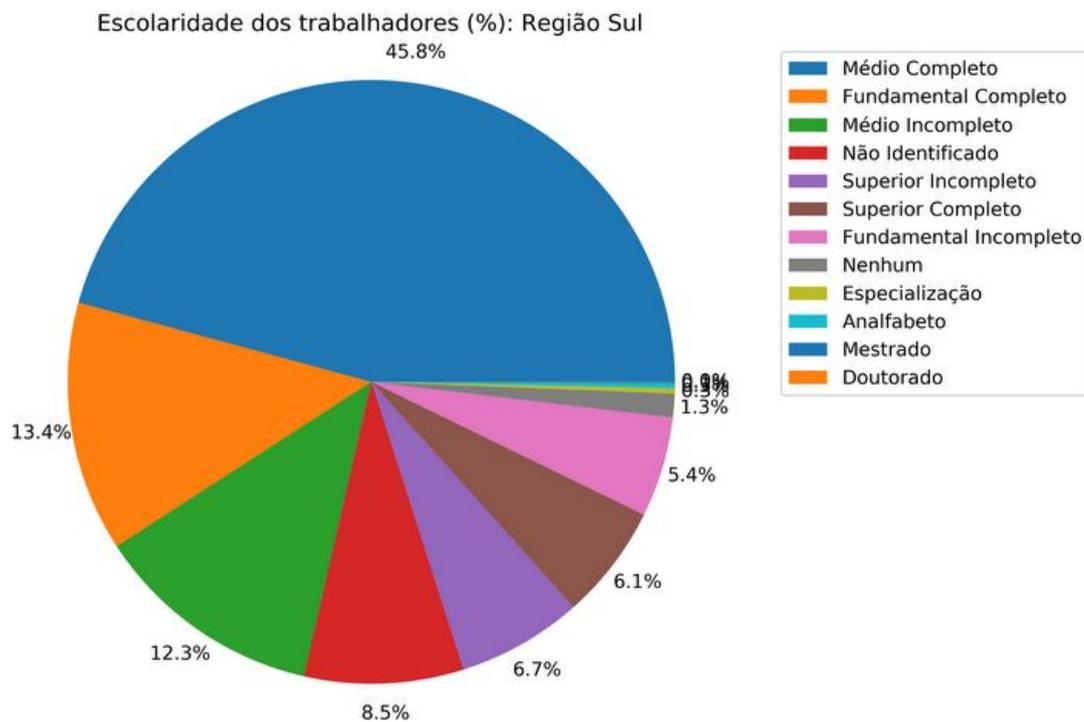
Figura 2 – Perfil de escolaridade dos trabalhadores cadastrados no SINE: Região Sudeste



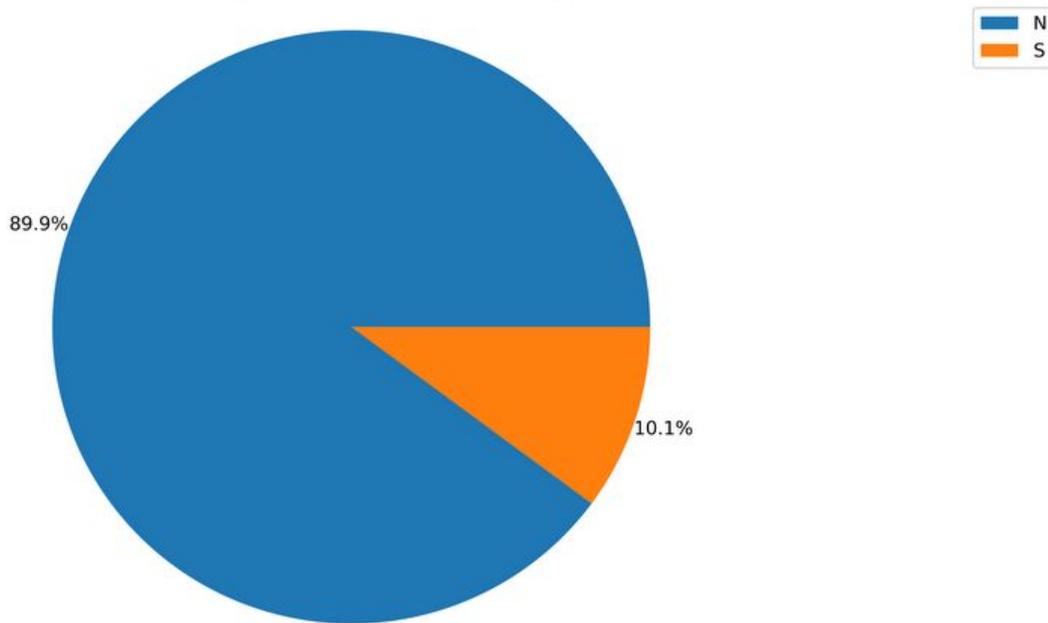
**Figura 3 – Estudantes cadastrados no SINE: Região Sudeste**

### 2.1.2 Região Sul

Para o Sul, o mesmo padrão do Sudeste se repete, porém a maioria dos trabalhadores da região Sul cadastrados no SINE possuem ensino médio completo, fundamental completo e ensino médio incompleto. Além disso, repete-se a tendência de que uma porcentagem relativa dos trabalhadores não souberam ou não informaram qual a escolaridade.

**Figura 4 – Perfil de escolaridade dos trabalhadores cadastrados no SINE: Região Sul**

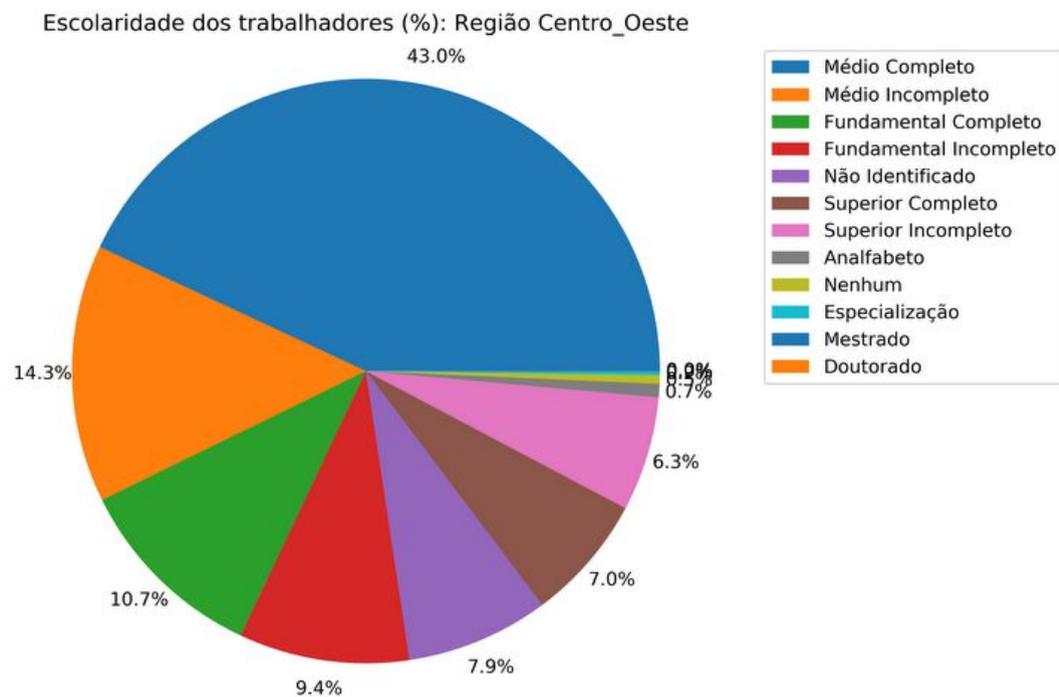
Trabalhadores que estão estudando (%): Região Sul



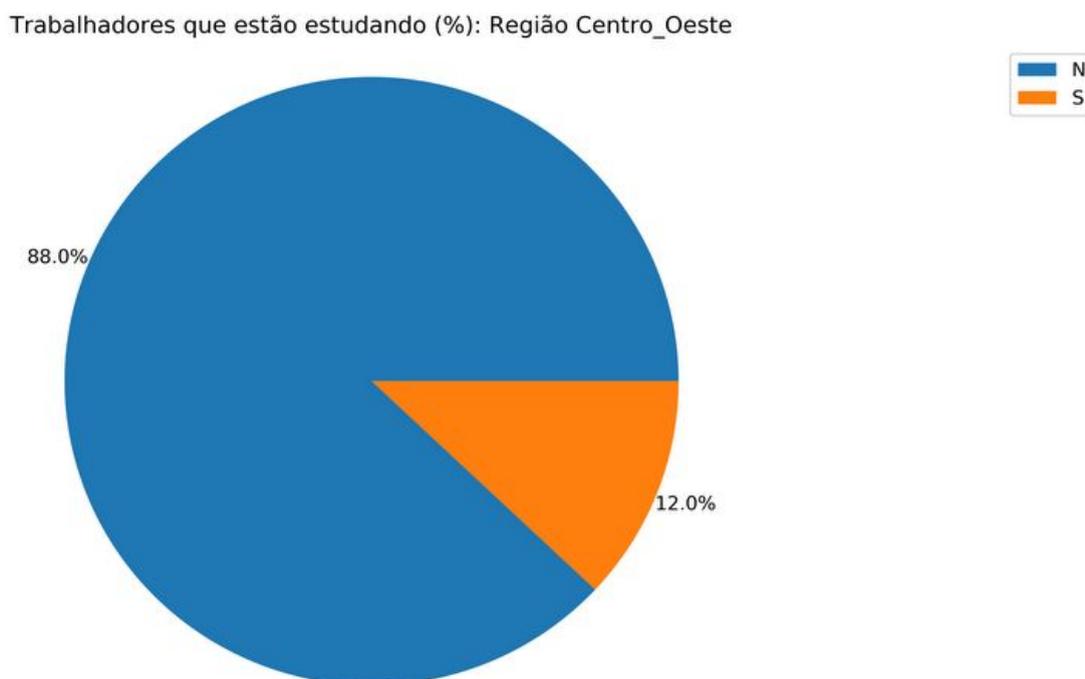
**Figura 5 – Estudantes cadastrados no SINE: Região Sul**

### 2.1.3 Região Centro-Oeste

Para a região centro-oeste, as mesmas tendências já verificadas na região sul e sudeste se repetiram.



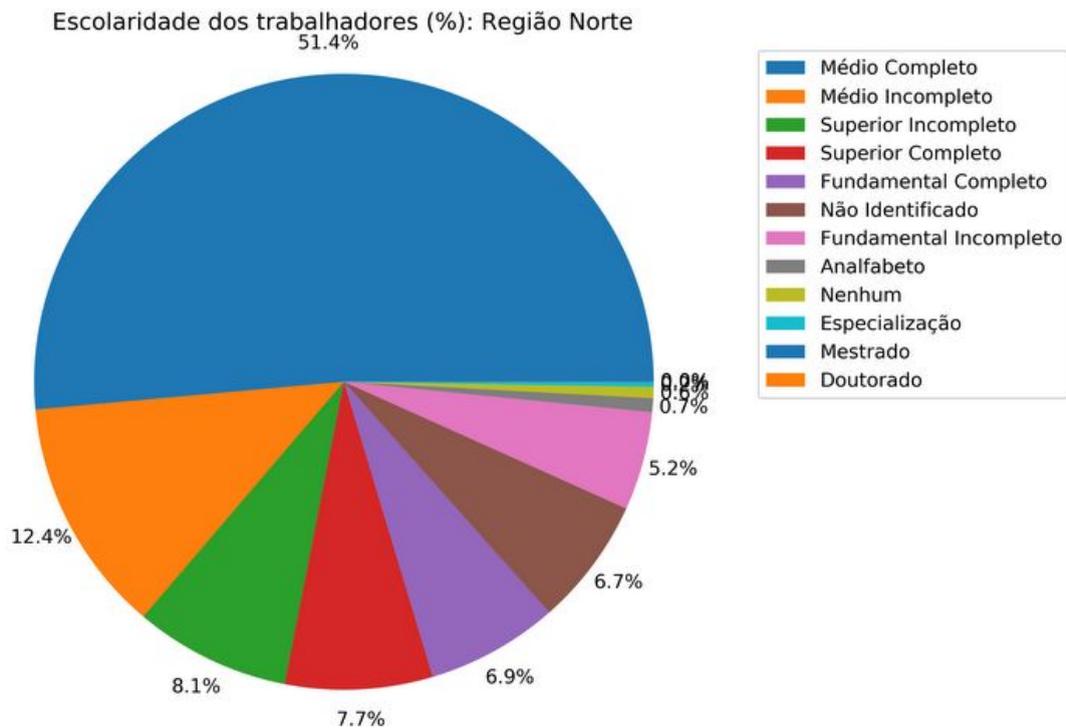
**Figura 6 – Perfil de escolaridade dos trabalhadores cadastrados no SINE: Região Centro-Oeste**



**Figura 7 – Estudantes cadastrados no SINE: Região Centro-Oeste**

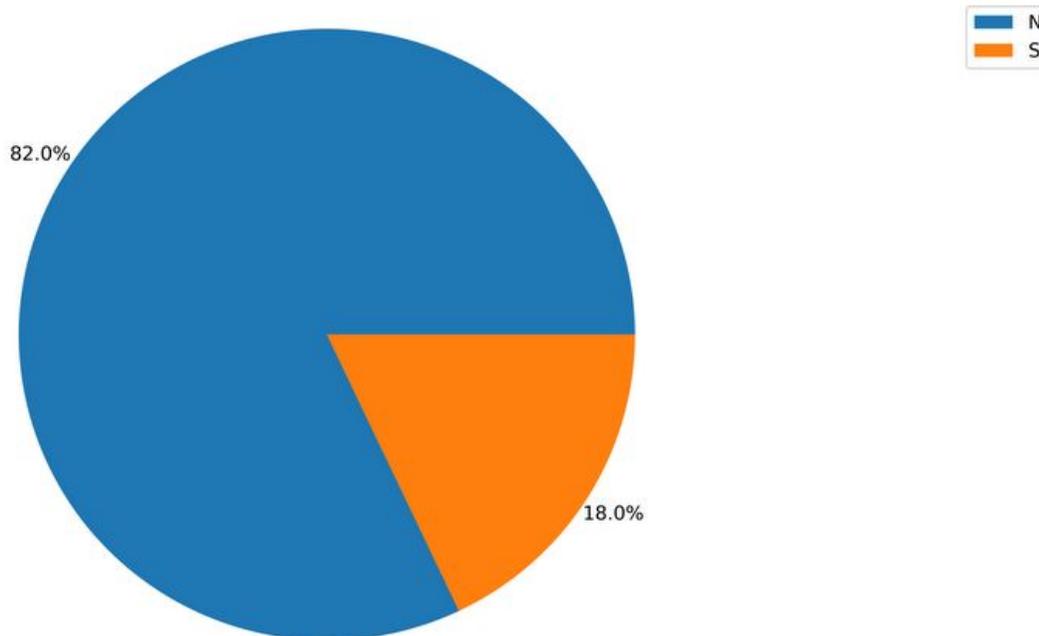
### 2.1.4 Região Norte

Para a região norte, observa-se uma diferença no perfil do trabalhador cadastrado no SINE. Nota-se que os trabalhadores com ensino superior incompleto e completo representam um valor significativo dos que mais procuram emprego. Além disso, nota-se um aumento, em relação as regiões sul e sudeste, no cadastro de trabalhadores que estão estudando.



**Figura 8 – Perfil de escolaridade dos trabalhadores cadastrados no SINE: Região Norte**

Trabalhadores que estão estudando (%): Região Norte

**Figura 9 – Estudantes cadastrados no SINE: Região Norte**

### 2.1.5 Região Nordeste

Para a região nordeste, observa-se que a maioria dos estudantes cadastrados possuem ensino médio completo, porém a segunda categoria mais cadastrada não soube informar a escolaridade. Além disso, nota-se que essa é a região em que um menor número percentual de estudantes estão cadastrados.

Escolaridade dos trabalhadores (%): Região Nordeste

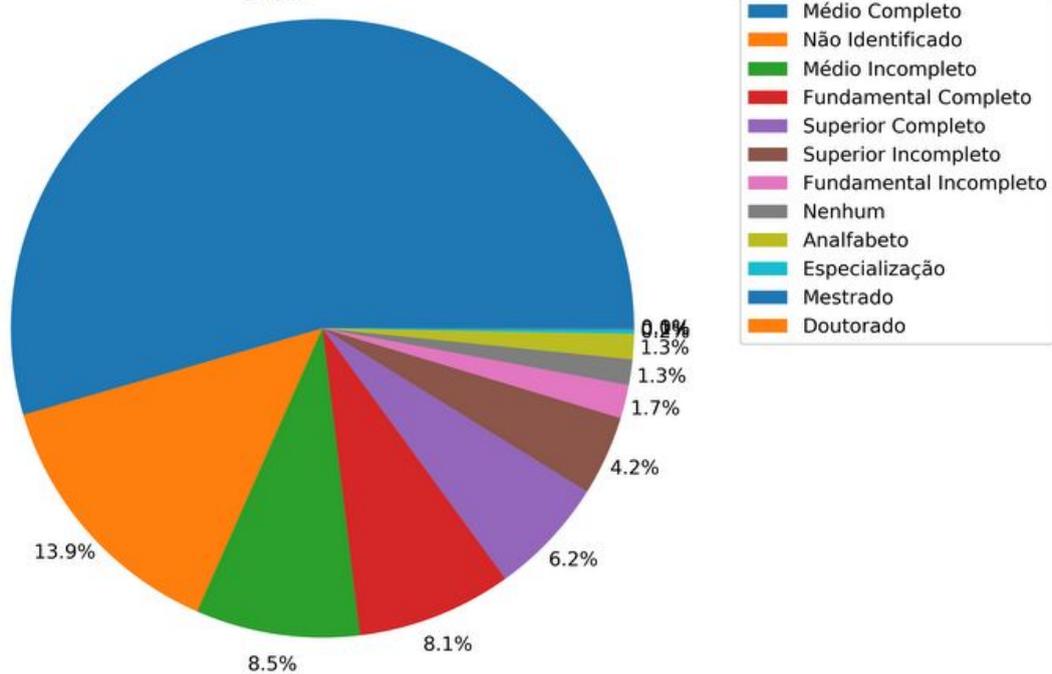


Figura 10 – Perfil de escolaridade dos trabalhadores cadastrados no SINE: Região Nordeste

Trabalhadores que estão estudando (%): Região Nordeste

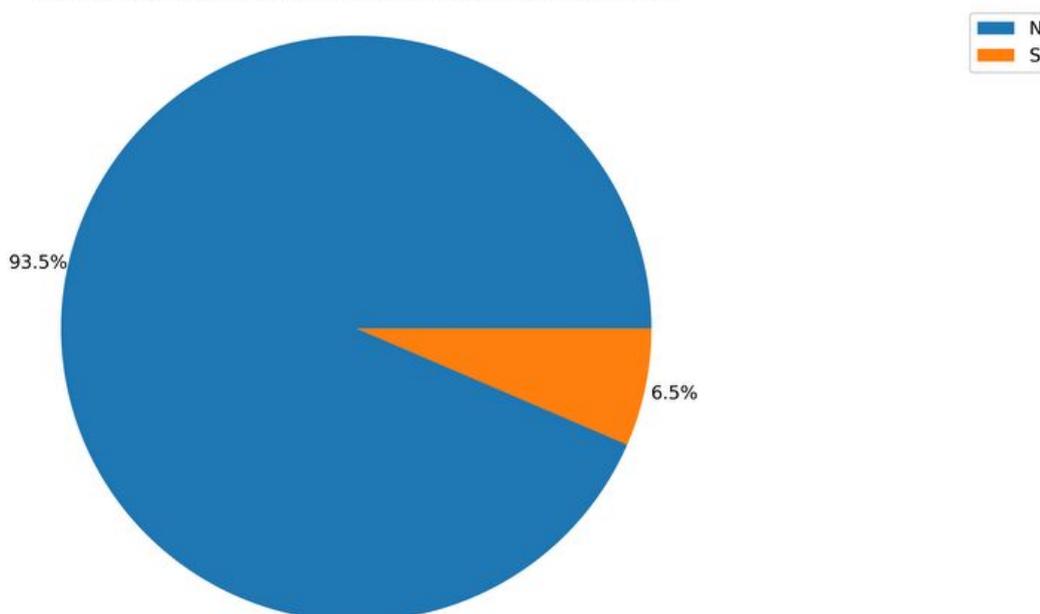


Figura 11 – Estudantes cadastrados no SINE: Região Nordeste

## **2.2 Dados faltantes**

Após a análise da escolaridade dos trabalhadores cadastrados, fez uma análise dos dados faltando por região e categoria a fim de entender melhor a base de dados do SINE (Tabela 2). Seguem algumas conclusões obtidas a partir desse estudo:

- Acima de 75% dos trabalhadores não preencheram ou não possuem: cursos profissionalizantes, cursos superiores e cursos de idiomas
- A maioria (acima de 90%) não colocou os municípios de interesse
- A maioria (acima de 90%) não possui ou não colocou deficiência

Tabela 2 – Análise de dados faltantes

	Valores percentuais de NaN						
	Sudeste	Sul	Centro-Oeste	Norte	Nordeste		
NACIONALIDADE	0.00	0.00	0.00	0.00	0.00		
DEFICIENCIAS	99.22	99.19	99.31	99.02	99.08		
BAIRRO	25.78	26.74	29.06	25.98	23.24		
CEP	0.02	0.01	0.06	0.14	0.08		
CODIGO_MUNICIPIO_IBGE	0.00	0.00	0.00	0.00	0.00		
NOME_MUNICIPIO	0.00	0.00	0.00	0.00	0.00		
UF	0.00	0.00	0.00	0.00	0.00		
ESCOLARIDADE	0.00	0.00	0.00	0.00	0.00		
ESTUDANTE	0.00	0.00	0.00	0.00	0.00		
CURSOS_PROFISSIONALIZANTES	78.63	84.86	83.08	75.20	75.66		
GRADUACOES	88.26	87.42	87.07	89.05	89.39		
POS_GRADUACOES	99.70	99.58	99.78	99.73	99.65		
IDIOMAS	96.88	97.59	98.27	98.02	97.05		
HABILITACAO	0.00	0.00	0.00	0.00	0.00		
VEICULOS	0.00	0.00	0.00	0.00	0.00		
DISP_VIAJAR	0.00	0.00	0.00	0.00	0.00		
DISP_DORMIR_EMP	0.00	0.00	0.00	0.00	0.00		
DISP_AUSENTAR_DOMIC	0.00	0.00	0.00	0.00	0.00		
PRETENSOES	1.67	1.29	1.17	2.07	1.53		
MUNICIPIOS_INTERESSE	95.42	96.44	97.51	98.92	91.97		

### 3 SOLUÇÃO PROPOSTA

A partir da análise do problema os seguintes pontos ficaram evidentes:

- O público alvo é o trabalhador que possui um nível de escolaridade menor, sem possuir até mesmo os cursos de qualificações para as vagas disponíveis.
- Esse público não possui muito conhecimento técnico para preenchimento do perfil, deixando vários campos sem preencher, não apresentando em seu currículo as habilidades e competências que realmente possui. Isso se evidenciou pelo grande número de trabalhadores que nem souberam cadastrar seu nível de escolaridade
- O perfil de quem procura o SINE não apresenta cursos de idiomas, nem de cursos profissionalizantes.
- O trabalhador não sabe informar as cidades que pretende trabalhar

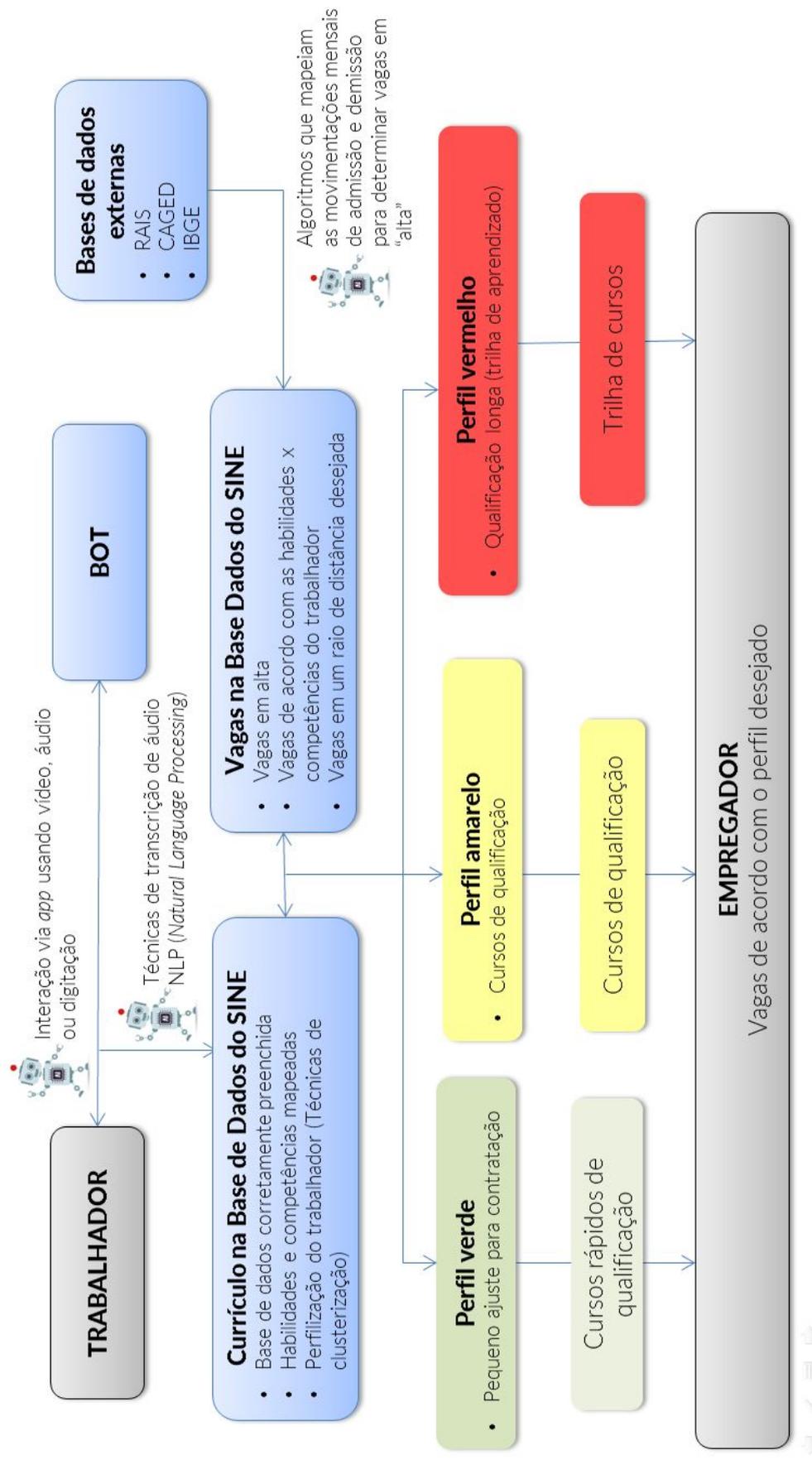
Diante do exposto, procurou-se desenvolver uma plataforma via *app* que da forma mais fácil possível é capaz de extrair as informações pessoais, técnicas, habilidades, competências e preencher um currículo de forma completa em uma base de dados no SINE, além de indicar as profissões que estão em alta, bem como os cursos necessários para conquistar tal posição. Para atingir esse objetivo, os seguintes passos acontecem:

1. O trabalhador irá utilizar o *app* desenvolvido e iniciar uma conversa via áudio, vídeo ou chat com um BOT que irá conduzir a conversa de forma a preencher corretamente uma base de dados no SINE, incluído técnicas para mapeamento do perfil emocional, de habilidades e competências.
2. O áudio obtido será transcrito e através das técnicas de NLP (*Natural Language Processing*) os dados serão automaticamente salvos na base de dados do SINE.
3. Um script de clusterização irá identificar o perfil do trabalhador para as profissões que ele tem afinidade, tem disponibilidade para morar em um raio determinado por ele, bem como para aquelas que são de outro setor, mas estão em alta.

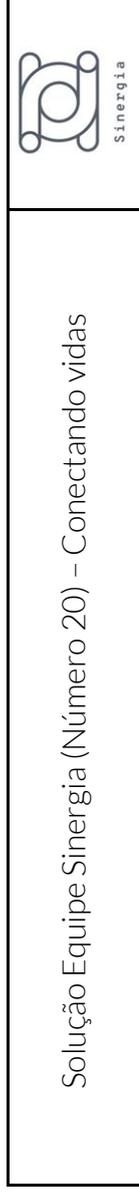


4. A análise das profissões que estão em alta será feita utilizando as métricas de movimentação de admitidos e demitidos nas bases de dados: RAIS, CAGED e IBGE.
5. Com as vagas em alta mais as vagas disponíveis, será apresentado ao trabalhador seu perfil e quais os cursos necessários para que ele possa concorrer àquelas vagas.
6. Por fim, o matching é realizado, em que o trabalhador pode então aplicar às vagas para as quais ele se qualificou, aumentando as probabilidades de contratação.

Apresenta-se assim o esquema do funcionamento da solução (Figura 12):



Como identificar e disponibilizar aos trabalhadores cadastrados no SINE informações sobre as qualificações exigidas pelas ocupações em alta?



**Figura 12 – Esquemático da solução**

Como identificar e disponibilizar aos trabalhadores cadastrados no SINE informações sobre as qualificações exigidas pelas ocupações em alta?



Ao longo do processo de ideação, desenvolveu-se também o CANVAS do projeto para que a equipe pudesse ter uma visão completa do modelo de negócio.

Canvas\_Equipe\_20.d  
ocx

#### 4 MVP

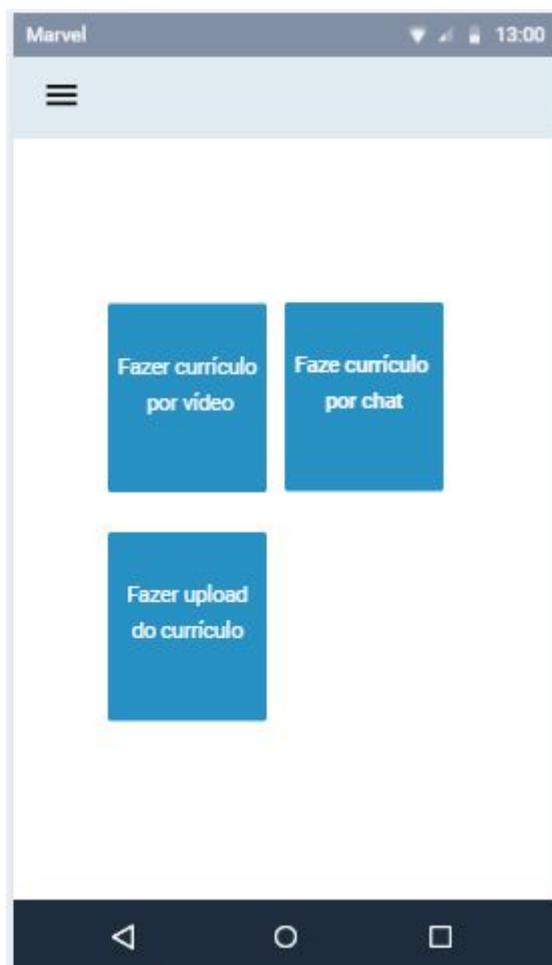
A fim de apresentar um conceito do *app* a ser desenvolvido, apresenta-se a seguir as telas de interface da solução proposta:



**Figura 13 – Tela inicial**



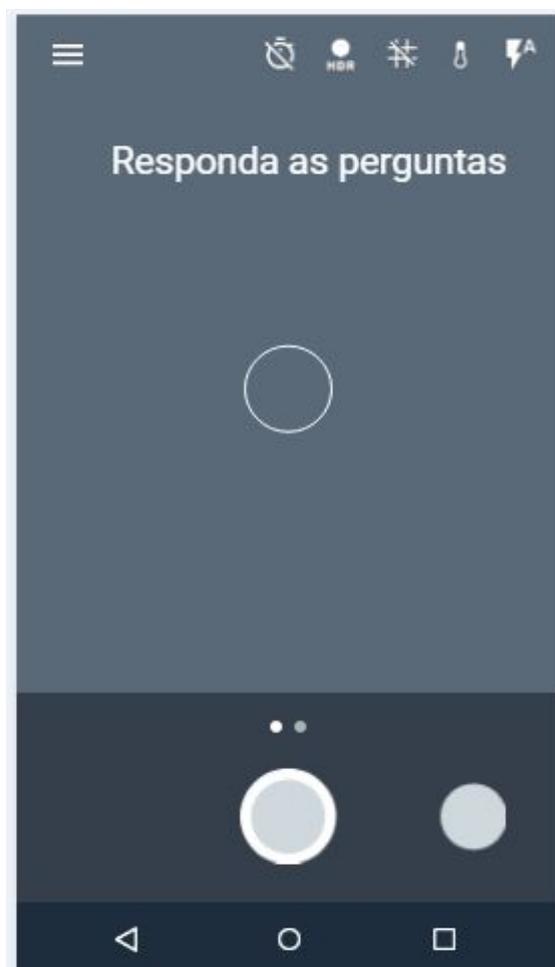
**Figura 14 – Tela de dados**



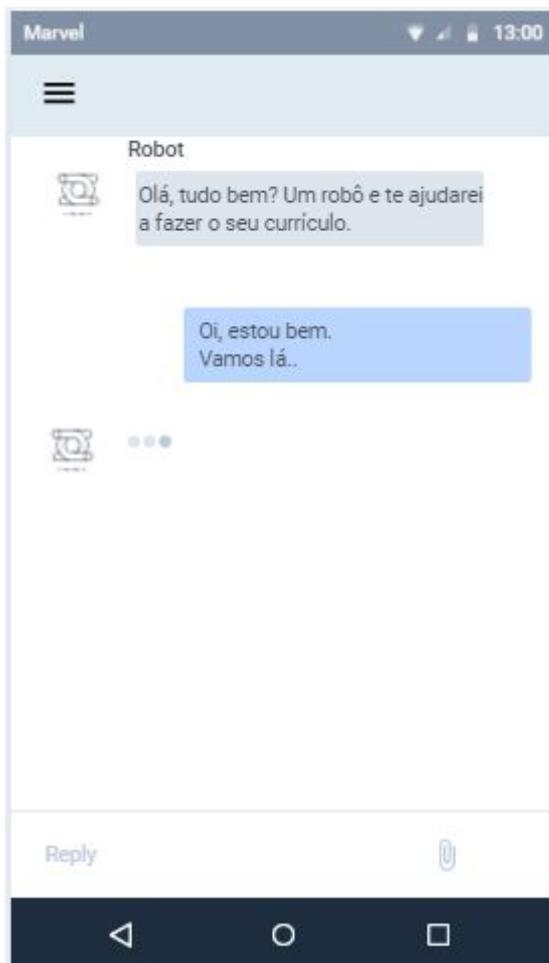
**Figura 15 – Tela de opções de entrevista**



Figura 16 – Tela de criação



**Figura 17 – Tela de perguntas por vídeo**



**Figura 18 – Tela de perguntas por chat**



Figura 19 – Tela de vagas para perfil vermelho



Figura 20 – Tela de vagas para perfil amarelo



Figura 21 – Tela de vagas para perfil verde

## 5 CONCLUSÃO

A partir do desafio proposto, analisou-se a base de dados do SINE para entender melhor os perfis dos trabalhadores cadastrados. A partir desse entendimento propôs-se uma solução usando as técnicas disponibilizadas ao longo do evento. A análise dos dados mostrou um grande déficit de informações no cadastro de habilidades e competências dos trabalhadores, além disso, evidenciou-se que o público alvo é o trabalhador que possui um nível de escolaridade menor, sem possuir cursos profissionalizantes para as vagas disponíveis. Sendo assim, procurou-se desenvolver uma plataforma via *app* que da forma mais fácil possível é capaz de extrair as informações pessoais, técnicas, habilidades, competências e preencher um currículo de forma completa em uma base de dados no SINE, além de indicar as profissões que estão em alta, bem como os cursos necessários para conquistar tal posição. Por fim, o trabalhador pode se candidatar às vagas pretendidas, uma vez que já possui as qualificações necessárias, aumentando assim sua probabilidade de conquistar a tão sonhada recolocação.

## 6 APÊNDICE A – CÓDIGO DESENVOLVIDO EM PYTHON PARA ANÁLISE DOS DADOS

```
# -*- coding: utf-8 -*-
'''
Created on Sat Jun 13 12:10:23 2020

@author: Joabe
'''

import pandas as pd
import matplotlib.pyplot as plt

Sudeste = ['SP', 'RJ', 'MG', 'ES']
Sul = ['PR', 'SC', 'RS']
Centro_Oeste = ['DF', 'GO', 'MT', 'MS']
Norte = ['AC', 'AM', 'AP', 'PA', 'RO', 'RR', 'TO']
Nordeste = ['AL', 'BA', 'CE', 'MA', 'PB', 'PE', 'PI', 'RN', 'SE']
regions = [Sudeste, Sul, Centro_Oeste, Norte, Nordeste]
regions_name = ['Sudeste', 'Sul', 'Centro_Oeste', 'Norte', 'Nordeste']

for index, regions_var in enumerate(regions):

    df_trabalhadores = pd.DataFrame()

    for state in regions_var:
        print(10*'- ' + 'CALCULANDO DADOS PARA O ESTADO: '+ state +
              10*'-)

        name_csv =
'D_ETL_IMO_EXTRACAO_SINE_ABERTO_TRABALHADORES_' + state + '.csv'
        df_temp = pd.read_csv(name_csv, sep=';', encoding='iso-8859-1')
        df_trabalhadores = pd.concat([df_trabalhadores, df_temp])
```

```
# Analise da escolaridade

series_categ = df_temp['ESCOLARIDADE'].value_counts()
series_categ_porcentagem = df_temp['ESCOLARIDADE'].value_counts(normalize =
True) * 100

labels = series_categ_porcentagem.index.tolist()
sizes = series_categ_porcentagem.to_numpy()

fig1, ax1 = plt.subplots()
name1 = 'escolaridade_trabalhadores_regiao_' + regions_name[index]
plt.xticks(rotation=45)
series_categ.plot.bar(x = 'Escolaridade dos trabalhadores: Região ' +
str(regions_var), fontsize=8, rot=20)
ax1.set_title('Escolaridade dos trabalhadores: Região ' +
regions_name[index])
figManager = plt.get_current_fig_manager()
figManager.window.showMaximized()
plt.show()
plt.pause(0.1)
fig1.savefig(name1 + '.png', dpi = 600) # save the figure to file
plt.close(fig1) # close the figure window

name2 = 'escolaridade_trabalhadores_porcentagem_regiao_' +
regions_name[index]
fig2, ax2 = plt.subplots()
pie = ax2.pie(sizes, autopct='%1.1f%%', pctdistance = 1.1, labeldistance =
1.1)
ax2.axis('equal') # Equal aspect ratio ensures that pie is drawn as a circle.
ax2.legend(pie[0], labels, loc='upper right', bbox_to_anchor = (1,1))
```

```
ax2.set_title('Escolaridade dos trabalhadores (%): Região ' +
regions_name[index])
figManager = plt.get_current_fig_manager()
figManager.window.showMaximized()
plt.show()
plt.pause(0.1)
fig2.savefig(name2 + '.png', dpi = 600) # save the figure to file
plt.close(fig2) # close the figure window

# Analise sobre estudo

series_categ = df_temp['ESTUDANTE'].value_counts()
series_categ_porcent = df_temp['ESTUDANTE'].value_counts(normalize =
True) * 100

labels = series_categ_porcent.index.tolist()
sizes = series_categ_porcent.to_numpy()

fig1, ax1 = plt.subplots()
name1 = 'estudante_trabalhadores_regiao_' + regions_name[index]
plt.xticks(rotation=0)
series_categ.plot.bar(x = 'Trabalhador estudando? Região ' +
str(regions_var), fontsize=8, rot=0)
ax1.set_title("Trabalhadores que estão estudando: Região ' +
regions_name[index])
figManager = plt.get_current_fig_manager()
figManager.window.showMaximized()
plt.show()
plt.pause(0.1)
fig1.savefig(name1 + '.png', dpi = 600) # save the figure to file
plt.close(fig1) # close the figure window
```

```

        name2 = 'estudante_trabalhadores_porcentagem_regiao_' +
regions_name[index]
        fig2, ax2 = plt.subplots()
        pie = ax2.pie(sizes, autopct='%1.1f%%', pctdistance = 1.1, labeldistance =
1.1)
        ax2.axis('equal') # Equal aspect ratio ensures that pie is drawn as a circle.
        ax2.legend(pie[0], labels, loc='upper right', bbox_to_anchor = (1,1))
        ax2.set_title('Trabalhadores que estão estudando (%): Região ' +
regions_name[index])
        figManager = plt.get_current_fig_manager()
        figManager.window.showMaximized()
        plt.show()
        plt.pause(0.1)
        fig2.savefig(name2 + '.png', dpi = 600) # save the figure to file
        plt.close(fig2) # close the figure window

# Análise de preenchimento - Qntdade de informações não preenchidas
series_nan = df_trabalhadores.isnull().sum(axis = 0)
series_nan_porc = series_nan / len(df_trabalhadores) * 100
df_nan = pd.concat([series_nan, series_nan_porc], axis=1)
        df_nan.rename(columns={0: "Valores absolutos de NaN", 1: "Valores
percentuais de NaN"}, inplace = True)
        df_nan.to_excel('nan_data_excel_'+regions_name[index]+'.xlsx')
        df_nan.to_csv('nan_data_csv_'+regions_name[index]+'csv')

```

## 7 BIBLIOGRAFIA, SITES E BASES DE DADOS

1.

[ftp://ftp.ibge.gov.br/Trabalho\\_e\\_Rendimento/Pesquisa\\_Nacional\\_por\\_Amostra\\_de\\_Domicilios\\_continua/Mensal/Quadro\\_Sintetico/2020/pnadc\\_202004\\_quadroSintetico.pdf](ftp://ftp.ibge.gov.br/Trabalho_e_Rendimento/Pesquisa_Nacional_por_Amostra_de_Domicilios_continua/Mensal/Quadro_Sintetico/2020/pnadc_202004_quadroSintetico.pdf) .  
[Online]

2. <https://www.ibge.gov.br/explica/desemprego.php>. [Online]

3.

<https://economia.uol.com.br/noticias/estadao-conteudo/2020/06/09/covid-19-mascara-dados-de-desemprego.htm> . [Online]

4.

<https://www.ibge.gov.br/estatisticas/sociais/trabalho/9171-pesquisa-nacional-por-amostra-de-domicilios-continua-mensal.html?edicao=27774&t=destaques>. [Online]

5. <https://sineaberto.economia.gov.br/dadosAbertos.html>. [Online]

6. <https://sineaberto.economia.gov.br/vagas.html>. [Online]

7. <https://datasebrae.com.br/emprego/#saldoempregoMPE>. [Online]

8. <http://pdet.mte.gov.br/microdados-rai-e-caged>. [Online]

9. <http://www8.receita.fazenda.gov.br/SimplesNacional/Aplicacoes/ATBHE/>. [Online]

10.

<https://www.ibge.gov.br/estatisticas/sociais/populacao/9127-pesquisanacional-por-amostra-de-domicilios.html?=&t=microdados>. [Online]