



Data Science Academy

# Mentoria

Data Science Academy

Time Data ree

## INSPIRAÇÃO



**O Problema de Negócio**  
Dificuldades de encontrar pessoas com competências ideais para vagas de emprego ofertadas no mercado

## SOLUÇÃO



**A Solução - SkillNet**  
Solução tecnológica que por meio de um modelo preditivo e técnicas de machine learning identificam o candidato ideal para vaga de emprego ideal.

## PROJETO



**O Projeto- SkillNet**  
**Etapa 1 - Realizada na mentoria**  
 ✓ Preparação da Base de Dados  
 ✓ Enriquecimento de Dados  
 ✓ Análise Exploratória dos Dados  
 ✓ Testes de Hipótese  
**Etapa 2 - A realizar pos mentoria**  
 Otimização do Modelo Preditivo  
 Desenvolvimento dos Modelos  
 Desenvolvimento da Aplicação (Front)  
 Piloto (Validação)

## ETAPA 1

**Etapa 1 - Roadmap**



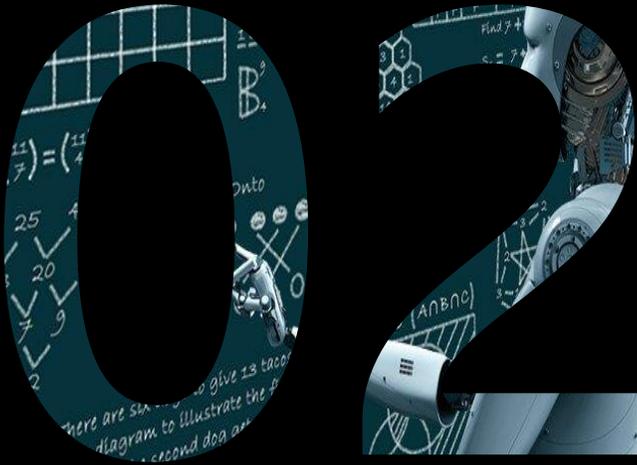
TRATAMENTO    DADOS    TIME    TESTE    ANÁLISE

Formação do Time pela Mentoria



# O Problema de Negócio

Dificuldades de encontrar pessoas com competências ideais para vagas de emprego ofertadas no mercado



## A Solução - SkillNet

Solução tecnológica que por meio de um modelo preditivo e técnicas de machine learning identificam o candidato ideal para vaga de emprego ideal.



## O Projeto- SkillNet

### Etapa 1 - Realizada na mentoria

- ✓ Preparação da Base de Dados
- ✓ Enriquecimento de Dados
- ✓ Análise Exploratória dos Dados
- ✓ Testes de Hipótese

### Etapa 2 – A realizar pos mentoria

- Otimização do Modelo Preditivo
- Desenvolvimento dos Modelos
- Desenvolvimento da Aplicação (Front)
- Piloto (Validação)

# Etapa 1 - Roadmap

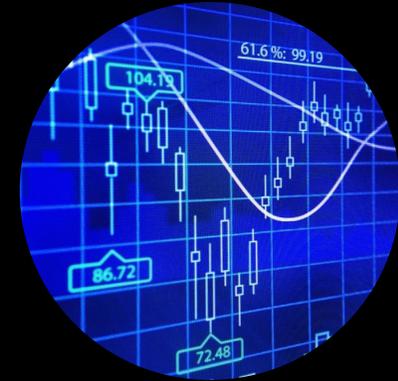
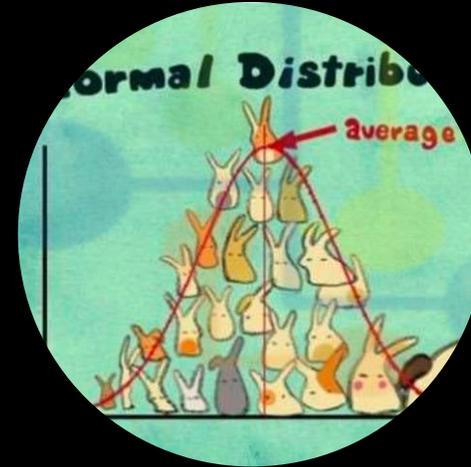
TRATAMENTO

DADOS

TIME

TESTE

ANÁLISE



Formação do Time pela Mentoria



## Felipe Reis Camargo Pacheco (Mentorado)

Brasiliense, bancário, formado em Ciências da Computação com especialização em Governança de TI, Gestão de Projetos e Processos. Atualmente trabalha na Diretoria de Tecnologia do Banco do Brasil na área Estratégica de Inteligência Competitiva de TI e está se especializando na carreira de Data Science.



## Hugo Paulo Hübner (Mentorado)

Formado em Processamento de dados com especialização em análise de sistemas e redes. Funcionário público federal.



## Rodolfo Risetto Terra (Mentor)

Economista, pós graduado em direito processual do trabalho e tecnólogo em cientista de dados. Atuei em diversos projetos com a utilização de inteligência Artificial, em Startup Americana, nos diversos nicho de mercado, desde imobiliário, saúde, financeiro, logística e contábeis. Atuai em diversas área dentro de projeto, como modelador de dados, analista BI, web scraping, cientista de dados e atualmente trabalho como Engenheiro de Dados na Ambev.

# Etapa 1 - Roadmap

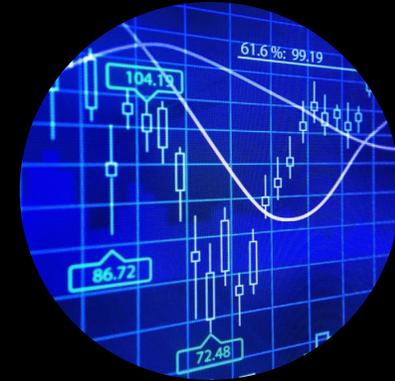
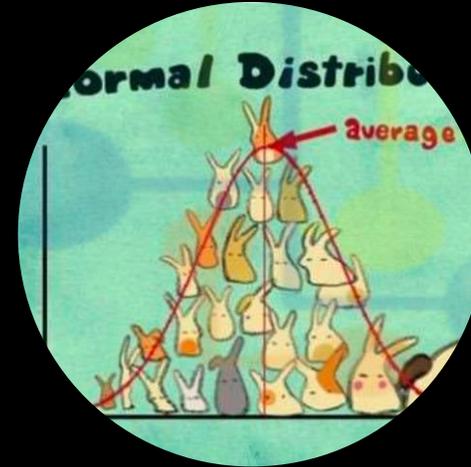
TRATAMENTO

DADOS

TIME

TESTE

ANÁLISE



Formação do Time pela Mentoria

# Etapa 1 - Roadmap

ANÁLISE

TRATAMENTO

DADOS

TIME

TESTE



Definição da Base de Dados

# SINE aberto

Ministério da Economia Trabalho



- O QUE É
- INTERMEDIÇÃO
- DADOS ABERTOS - TRABALHADORES
- DADOS ABERTOS - VAGAS
- INFORMAÇÕES API



O SINE Aberto é uma plataforma que visa aumentar a efetividade na colocação do trabalhador no mercado de trabalho, por meio do compartilhamento dos dados dos trabalhadores cadastrados no Sistema Nacional de Emprego (SINE).

Para isso, o governo federal propõe realizar parcerias com empresas, ampliando os atores que fazem intermediação de mão de obra e potencializando a geração de empregos.

- O QUE É
- INTERMEDIÇÃO
- DADOS ABERTOS - TRABALHADORES
- DADOS ABERTOS - VAGAS
- INFORMAÇÕES API



O acesso aos dados profissionais dos usuários cadastrados no SINE, por parte de empresas privadas de intermediação de mão de obra, amplia a oferta de vagas disponíveis para os trabalhadores inscritos no SINE. Com isso, aumentam as chances dos trabalhadores inscritos encontrarem emprego.

As empresas privadas de intermediação de mão de obra, por sua vez, poderão acessar o cadastro dos inscritos no SINE para prover serviços de intermediação de mão de obra, podendo efetuar a cobrança desses serviços junto aos empregadores. O acesso à base de dados deve ser feito mediante autenticação com certificado digital.

Vale lembrar que os serviços de intermediação de mão de obra realizados no âmbito do SINE Aberto devem ser totalmente gratuitos para o trabalhador.

Clique aqui para acessar o SINE Aberto [Acessar](#)

Em caso de dúvidas, sugestões ou problemas, contate-nos por meio do endereço [sineaberto@economia.gov.br](mailto:sineaberto@economia.gov.br).

- O QUE É
- INTERMEDIÇÃO
- DADOS ABERTOS - TRABALHADORES
- DADOS ABERTOS - VAGAS
- INFORMAÇÕES API



Arquivos com informações sobre trabalhadores cadastrados.

- Dicionário de Dados (.pdf)
- Baixar todos os arquivos (.zip)

### CENTRO-OESTE

- Distrito Federal (.zip)
- Goiás (.zip)
- Mato Grosso (.zip)
- Mato Grosso do Sul (.zip)

### NORTE

- Acre (.zip)
- Amapá (.zip)
- Amazonas (.zip)
- Pará (.zip)
- Rondônia (.zip)
- Roraima (.zip)
- Tocantins (.zip)

### NORDESTE

- Alagoas (.zip)
- Bahia (.zip)
- Ceará (.zip)
- Maranhão (.zip)
- Paraíba (.zip)
- Pernambuco (.zip)
- Piauí (.zip)
- Rio Grande do Norte (.zip)
- Sergipe (.zip)

### SUDESTE

- Espírito Santo (.zip)
- Minas Gerais (.zip)
- Rio de Janeiro (.zip)
- São Paulo (.zip)

### SUL

- Paraná (.zip)
- Rio Grande do Sul (.zip)
- Santa Catarina (.zip)

# Etapa 1 - Roadmap

ANÁLISE

TRATAMENTO

DADOS

TIME

TESTE



Escolha e análise da Base de Dados

# Etapa 1 - Roadmap

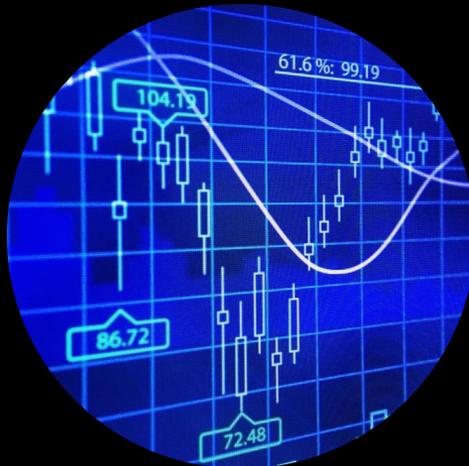
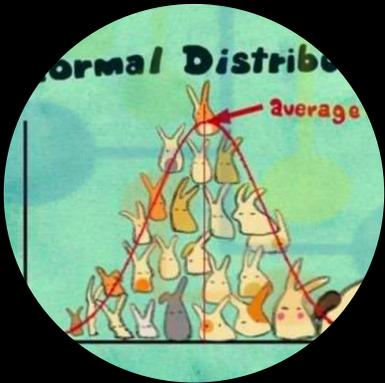
TESTE

ANÁLISE

TRATAMENTO

DADOS

TIME



Data Cleaning  
Data Wrangling  
Enriquecimento da base

# DATA CLEANING – DATA WRANGLING – ENRIQUECIMENTO DA BASE

## 1. Carregando conjunto de dados



Santa Catarina



Dados candidatos a empregos

**195.522** registros

## 2. Limpeza dos Dados

Dados removidos  
classificados como:

Não identificados

**6,43%**

Ignorados

**11,76%**

## 3. Enriquecimento de Base

Criação de Variáveis:

Regiões Municípios de SC

PIB por Cidade

PIB por Regiões M

PIB Médio

## 4. Otimização de Variáveis

Estágio

Horário de Trabalho

Pretensão Salarial

# Etapa 1 - Roadmap

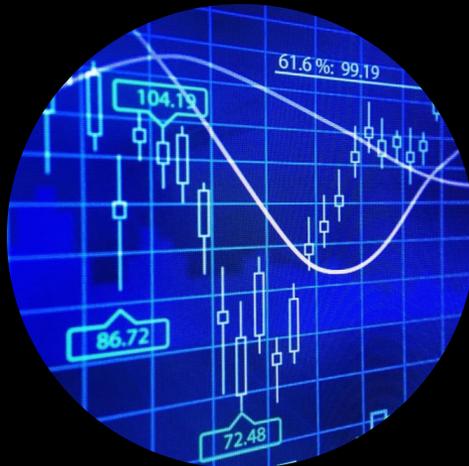
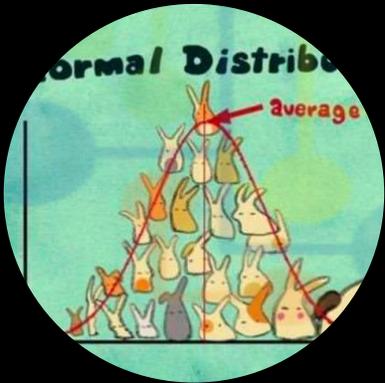
TESTE

ANÁLISE

TRATAMENTO

DADOS

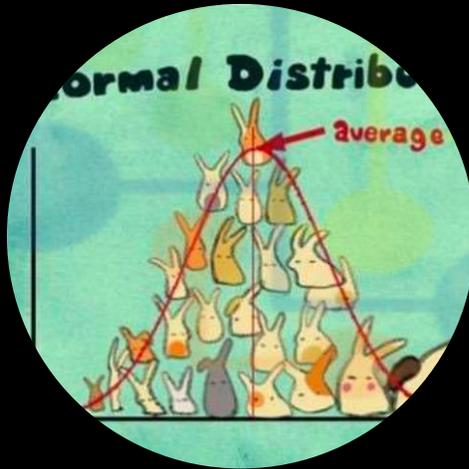
TIME



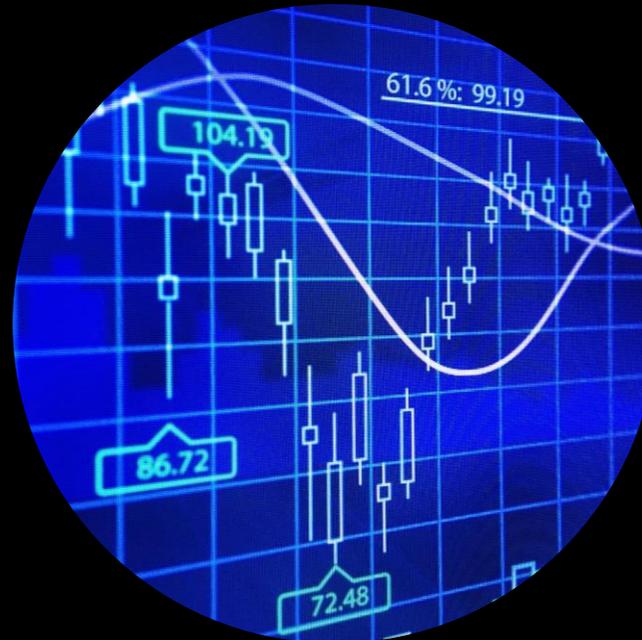
Data Cleaning  
Data Wrangling  
Enriquecimento da base

# Etapa 1 - Roadmap

TESTE



ANÁLISE



TRATAMENTO



DADOS



Análise Exploratória dos Dados

# ANÁLISE EXPLORATÓRIA

## 1. Carregando conjunto de dados



Santa Catarina



Dados candidatos a empregos

**195.522** registros

## 2. Explorando as Variáveis

Quantidade de Candidatos por Nacionalidade

Quantidade de Candidatos por Município

Quantidade de Candidatos por Escolaridade

Percentual de Candidatos que são Estudantes

Quantidade de Candidatos por Quantidade de Cursos Profissionalizantes

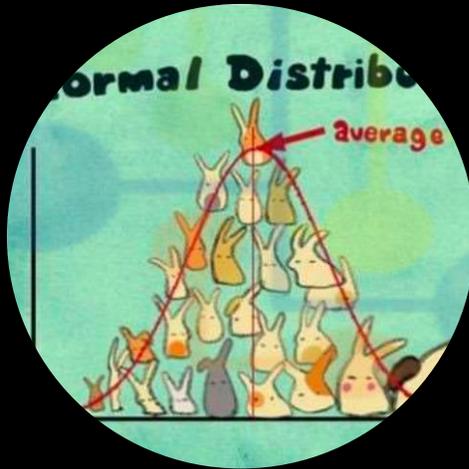
Graduações e Quantidade de Candidatos

Tipo de Habilitação

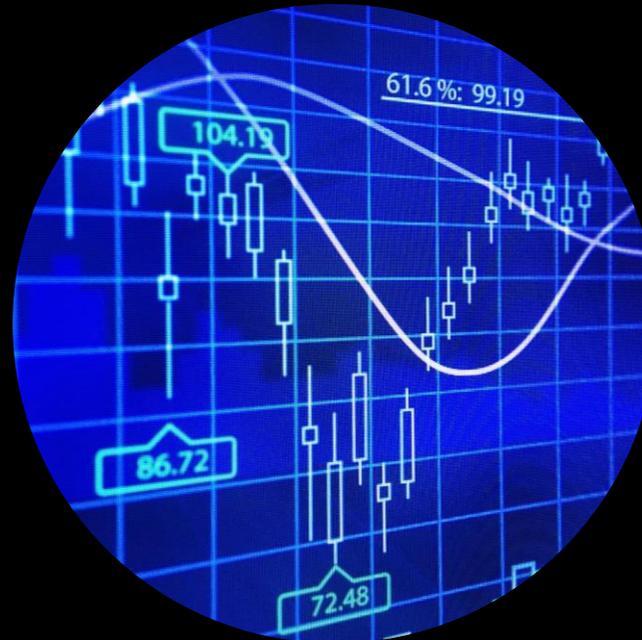
Em relação ao horário pretendido para trabalhar, existe diferença entre escolaridade?

# Etapa 1 - Roadmap

TESTE



ANÁLISE



TRATAMENTO



DADOS



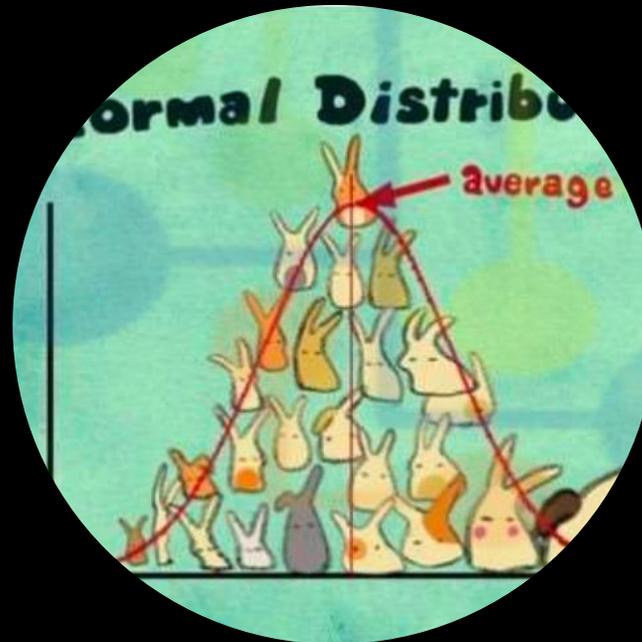
Análise Exploratória dos Dados

# Etapa 1 - Roadmap

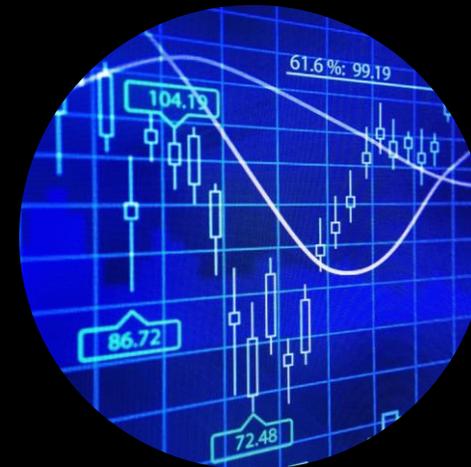
TESTE

ANÁLISE

TRATAMENTO



Testes de Hipótese





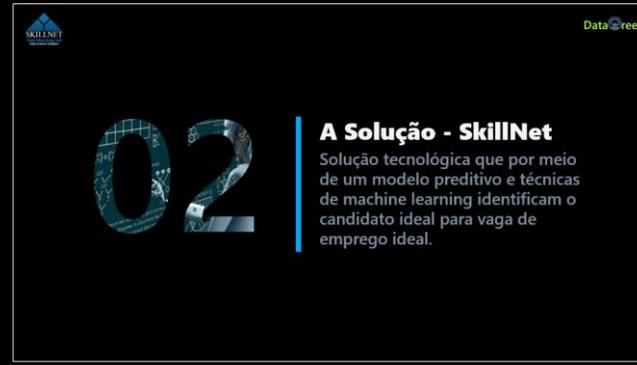
# OBRIGADO

## INSPIRAÇÃO



**01** | **O Problema de Negócio**  
Dificuldades de encontrar pessoas com competências ideais para vagas de emprego ofertadas no mercado

## SOLUÇÃO



**02** | **A Solução - SkillNet**  
Solução tecnológica que por meio de um modelo preditivo e técnicas de machine learning identificam o candidato ideal para vaga de emprego ideal.

## PROJETO



**03** | **O Projeto - SkillNet**

- Etapa 1 - Realizada na mentoria
  - ✓ Preparação da Base de Dados
  - ✓ Enriquecimento de Dados
  - ✓ Análise Exploratória dos Dados
  - ✓ Testes de Hipótese
- Etapa 2 - A realizar pos mentoria
  - Otimização do Modelo Preditivo
  - Desenvolvimento dos Modelos
  - Desenvolvimento da Aplicação (Front)
  - Piloto (Validação)

## ETAPA 1



**Etapa 1 - Roadmap**

TRATAMENTO   DADOS   TIME   TESTE   ANÁLISE

Formação do Time pela Mentoria