



ICDA6 - Introdução à
Ciência de Dados

Ciclo de Vida dos Dados

Uma análise sobre a saúde mental na
indústria de software

João Vitor Silva Bispo

SP3052672

Viviane de Santana Queiroz

SP3053601



Tópicos

- Introdução ao tema
- Resumo das etapas das análises e resultados obtidos
- Aplicabilidade do modelo
- Referências



Introdução ao tema

Datasets utilizados e origem dos dados

Health Expenditure and Suicide Rates [2000-2019]

Relation between health expenditure and suicide rates, in 2000/2019, WHO stats.



Mental Health in the Tech Industry

Survey on Mental Health in the Tech Workplace in 2014, 2016, 2017, 2018, 2019

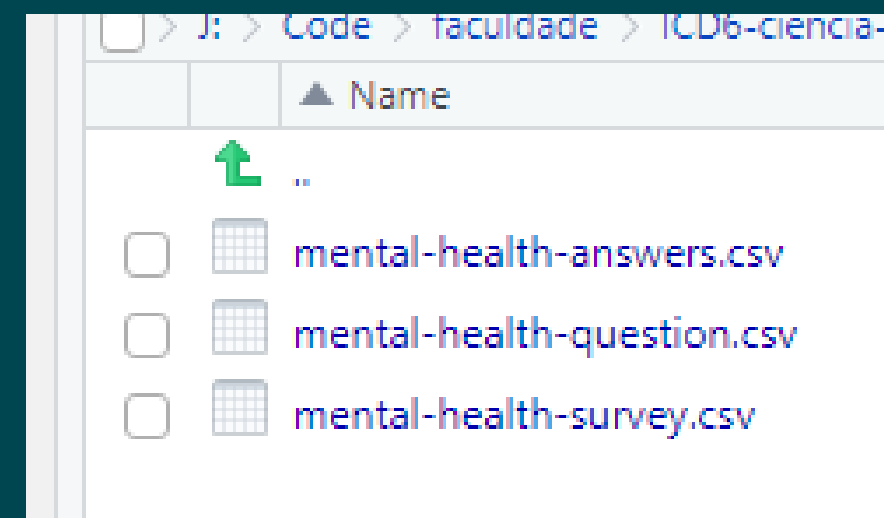


- Mais de 280 milhões de pessoas no mundo sofrem de depressão
- A discussão sobre saúde mental muitas vezes é negligenciada por ignorância do assunto
- A área de tecnologia prevê um déficit de 85 milhões de profissionais em 2030 (de acordo International Monetary Fund - IMF) em todo o mundo
- É estimada a criação de 797 mil vagas no Brasil nos próximos 5 anos (de acordo com a Brasscom)



Análise explícita

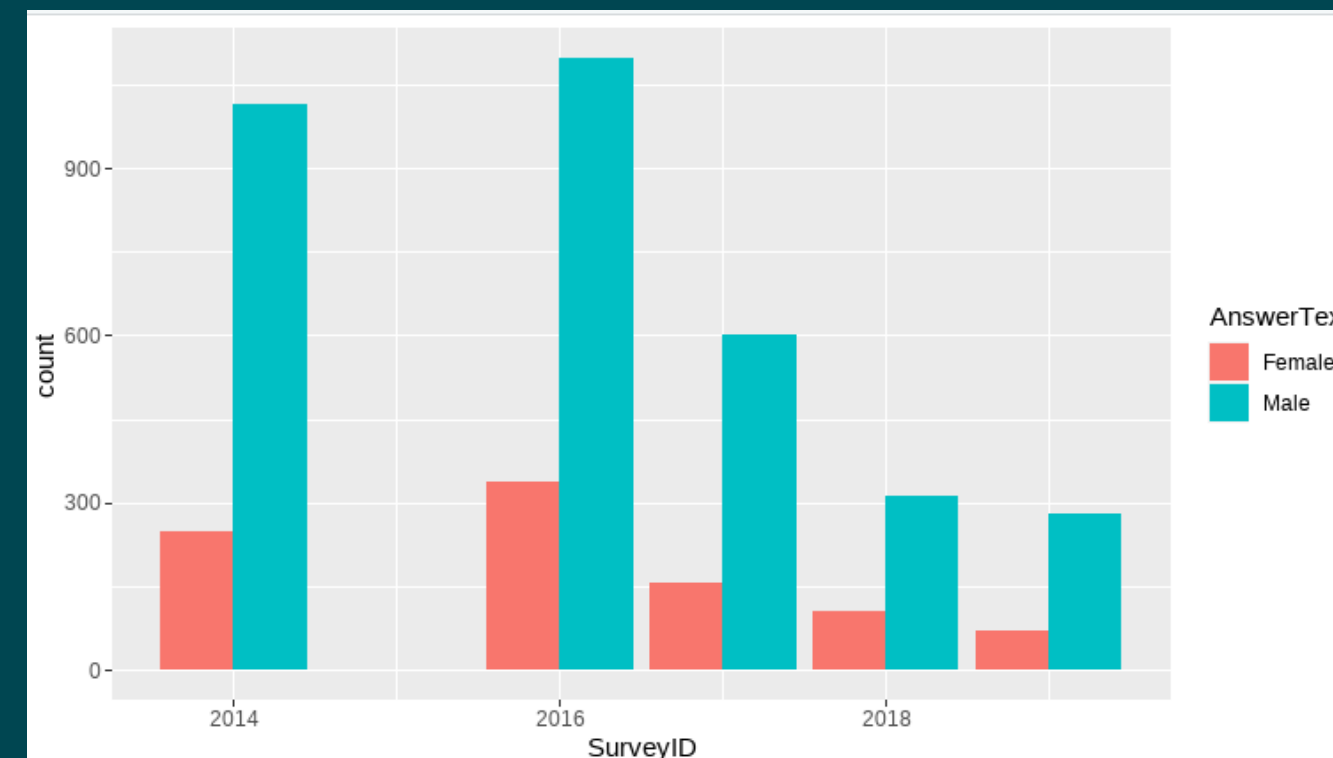
- CRISP-DM: Entendimento do negócio
- CRISP-DM: Entendimento dos dados
- Junção e filtro tabelas
- Sumarizações
- Análise do conjunto de dados
- Estabelecimento de critérios para seleção das variáveis
 - Maior quantidade de dados válidos
 - Maior relevância para negócio
- Objetivo de negócio: Um modelo que indique se o profissional da área de tecnologia já procurou ajuda profissional para tratar questões mentais



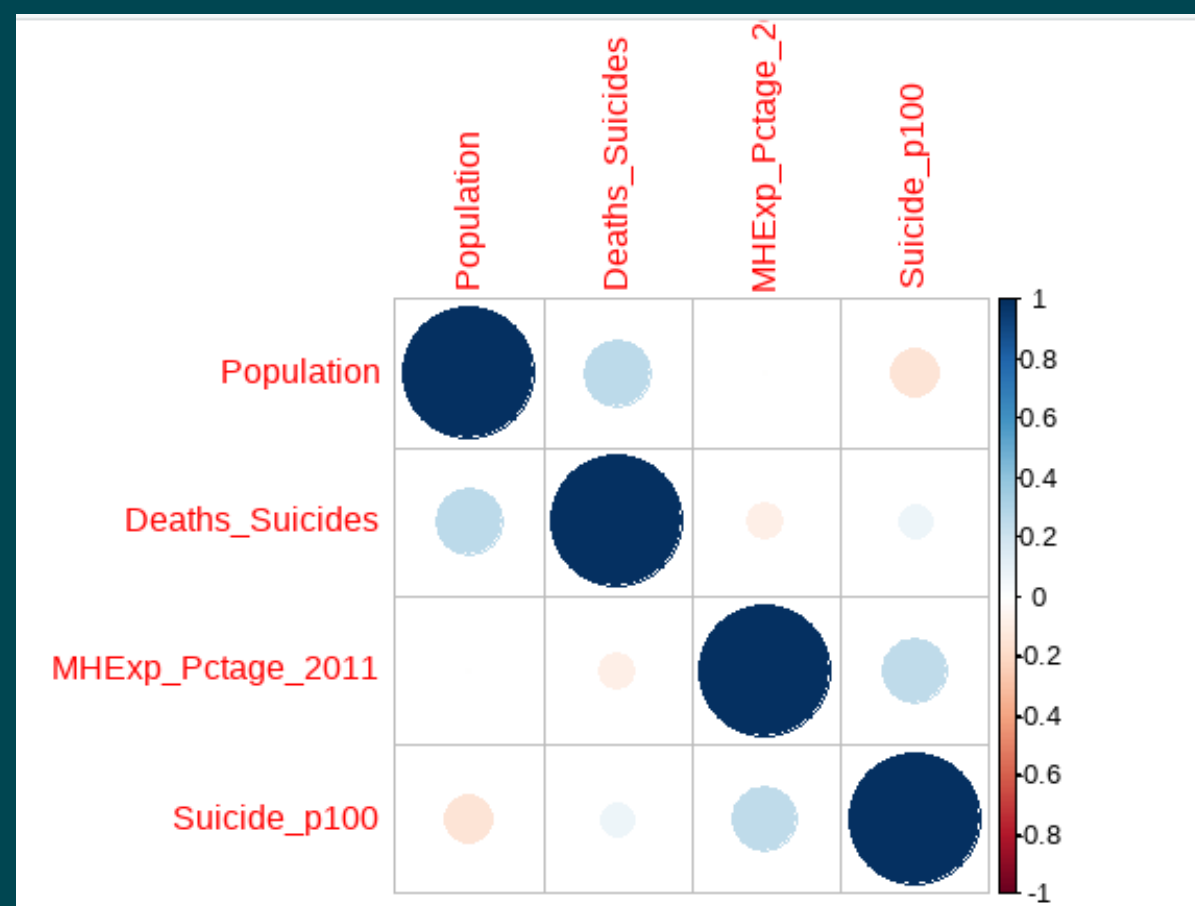
```
> head(answers)
  AnswerText SurveyID UserID QuestionID
1         37    2014     1           1
2         44    2014     2           1
3         32    2014     3           1
4         31    2014     4           1
5         31    2014     5           1
6         33    2014     6           1
> head(questions)
  questiontext questionid
1 what is your age?      1
2 what is your gender?  2
3 what country do you live in? 3
4 If you live in the United States, which state or territory do you live in? 4
5 Are you self-employed? 5
6 Do you have a family history of mental illness? 6
> head(survey)
  SurveyID Description
1    2014 mental health survey for 2014
2    2016 mental health survey for 2016
3    2017 mental health survey for 2017
4    2018 mental health survey for 2018
5    2019 mental health survey for 2019
> head(answers)
  AnswerText SurveyID UserID QuestionID
1         37    2014     1           1
2         44    2014     2           1
3         32    2014     3           1
4         31    2014     4           1
5         31    2014     5           1
6         33    2014     6           1
> head(questions)
```

Análise exploratória

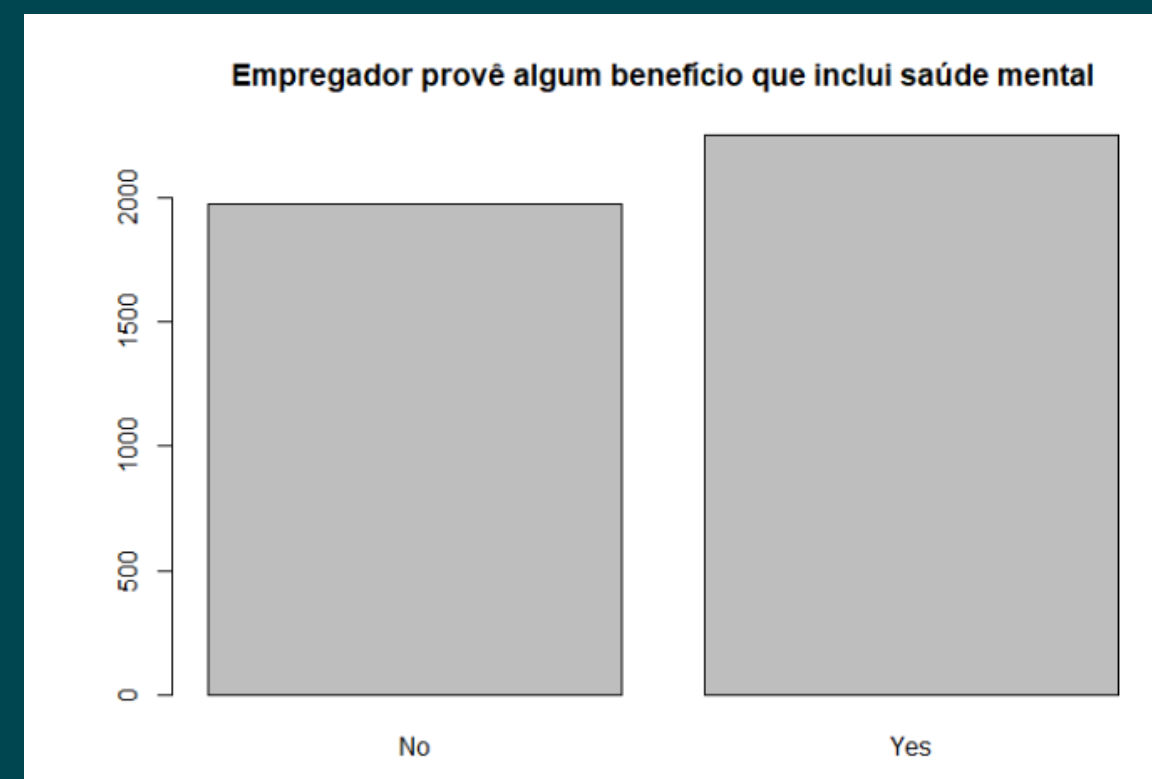
- CRISP-DM: Preparação dos Dados
- Limpeza e tratamento de outliers e dados faltantes
- Visualizações para dimensionar o comportamento das variáveis escolhidas



Dataset 1: Mental Health in the Tech Industry



Dataset 2: Health Expenditure and Suicide Rates [2000-2019]

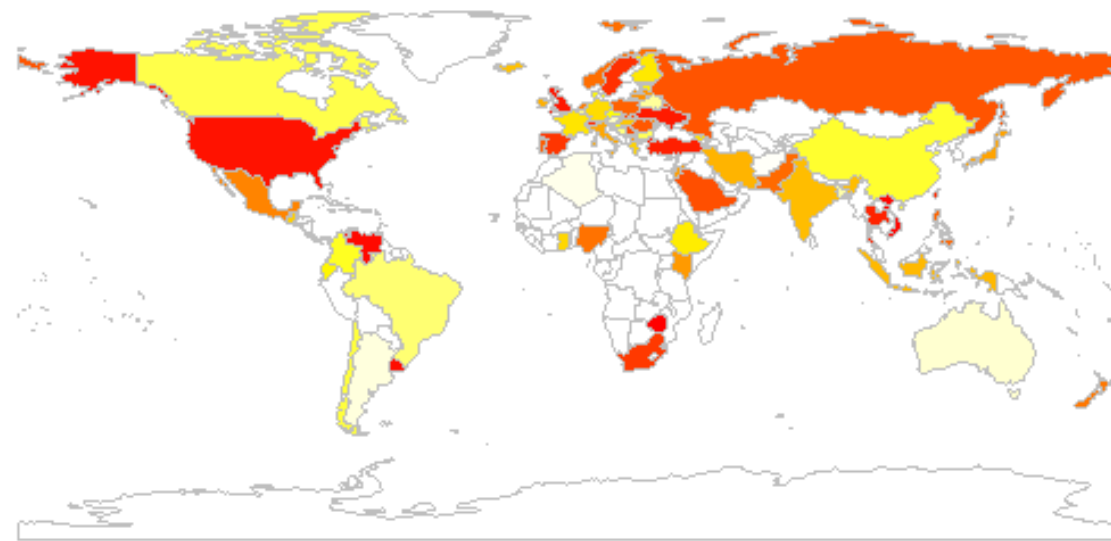


Dataset 2: Health Expenditure and Suicide Rates [2000-2019]

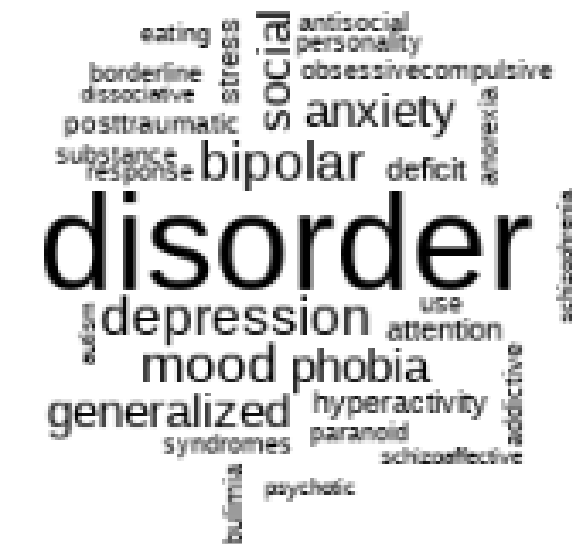


Análise exploratória

Países participantes das pesquisas

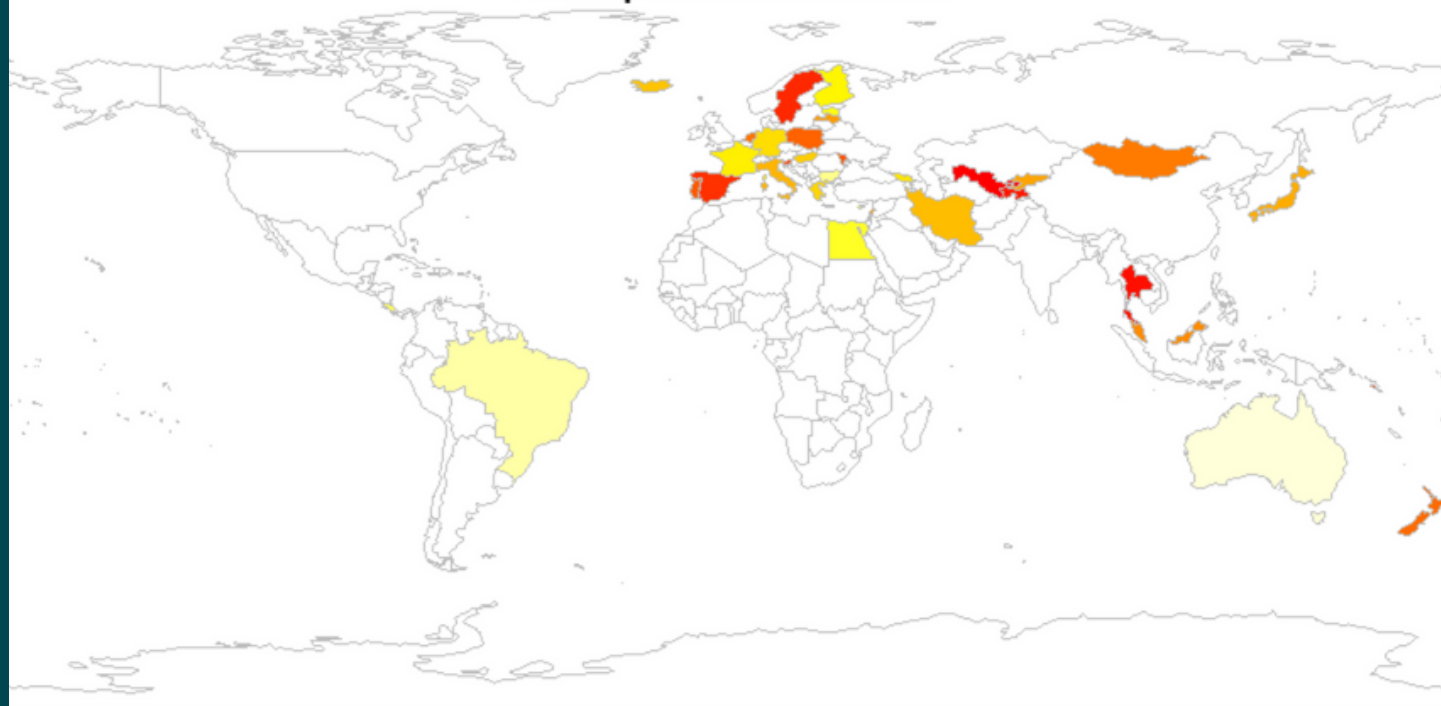


Dataset 1: Mental Health in the Tech Industry



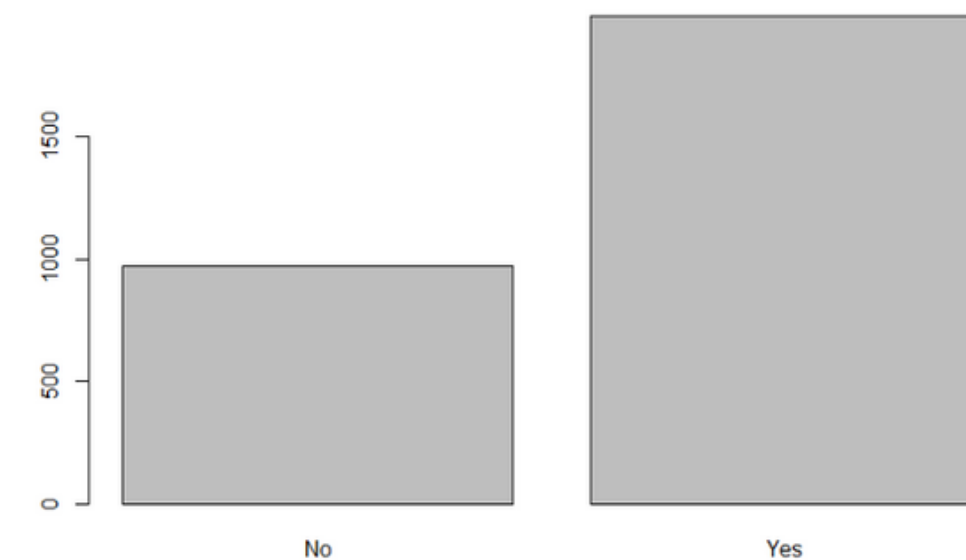
Dataset 1: Mental Health in the Tech Industry

Países presentes na análise



Dataset 2: Health Expenditure and Suicide Rates [2000-2019]

Possui algum distúrbio mental

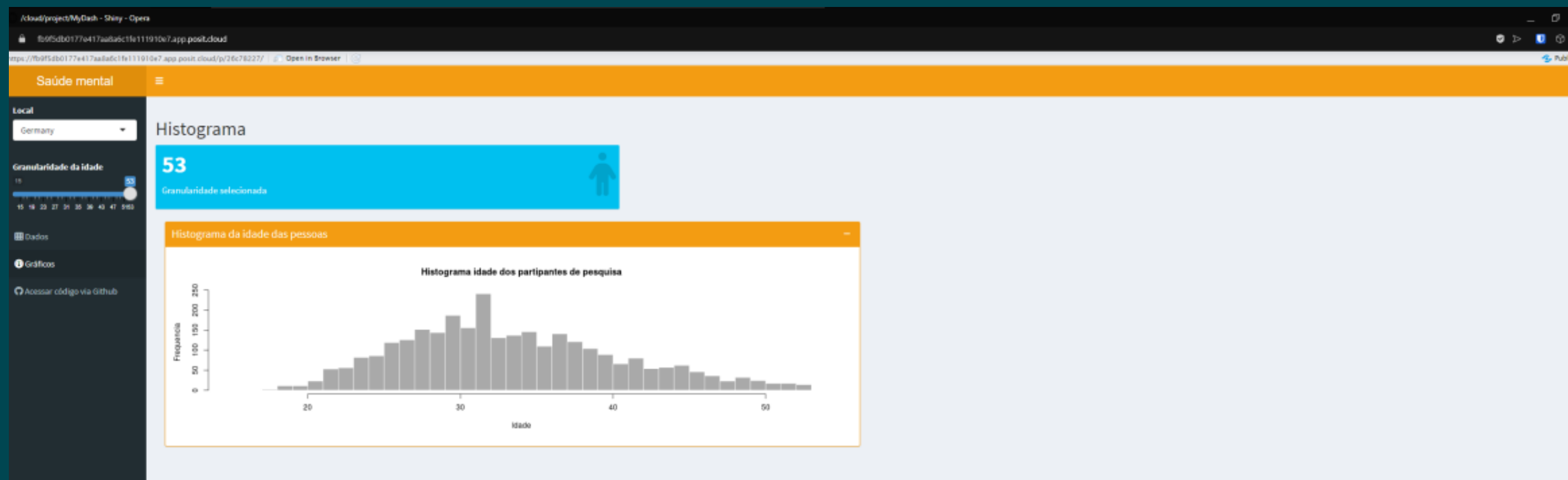
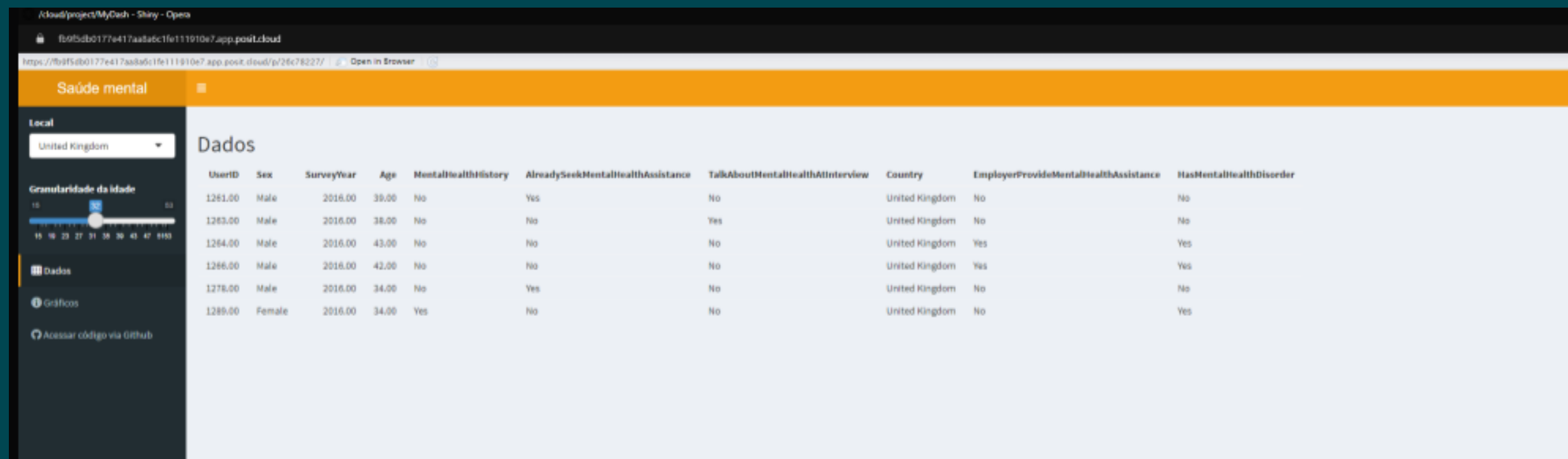


Dataset 1: Mental Health in the Tech Industry



Análise exploratória

Dashboard interativo Shiny



Análise exploratória

Algumas observações interessantes

- + de **100** perguntas no dataset 1
- Majoritariamente **homens**
- **60%** possui histórico familiar de distúrbio mental
- **53%** têm algum auxílio à saúde mental do empregador
- **95%** não abordaria saúde mental em uma entrevista de emprego



Análise implícita

- CRISP-DM: Modelagem
- CRISP-DM: Aplicação
- Regressão Logística Múltipla
- Hold out (70 / 30)
- Variável dependente

AlreadySeekMentalHealthAssistance

- Variáveis independentes

Sex

Age

MentalHealthHistory

TalkAboutMentalHealthAtInterview

HasMentalHealthDisorder

Country

Matriz de confusão

```
> confusionMatrixTable
      AlreadySeekMentalHealthAssistance
glm.pred.test  No  Yes
No      459  113
Yes     61   254
> |
```

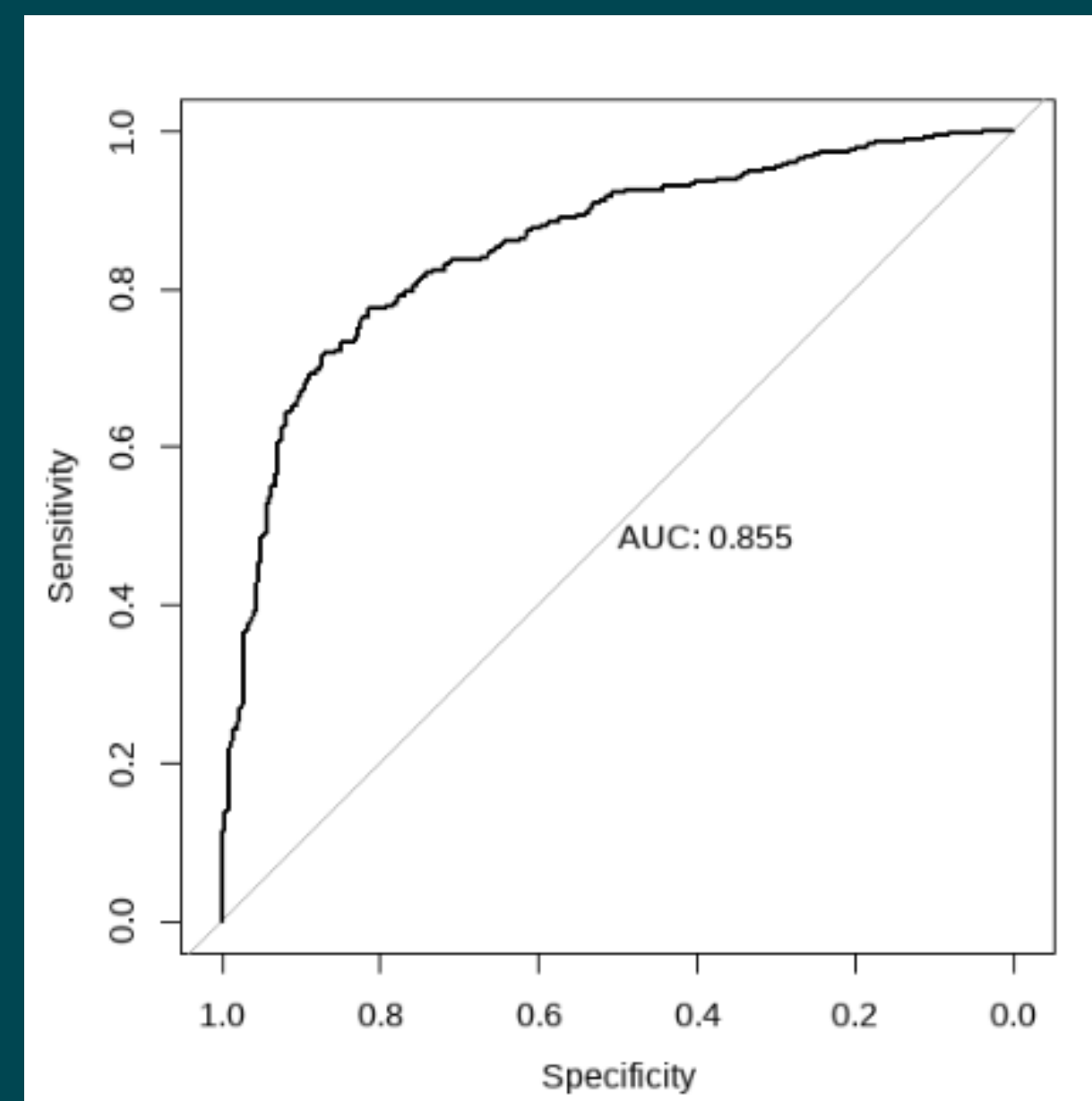
Acurácia alcançada:

80%

Valor de AUC:

80%

Curva ROC AUC





Aplicabilidade do modelo

Setor de pessoas em empresas de tecnologia

- Criação de benefícios voltados à saúde mental de acordo com o perfil dos colaboradores
- Ações preventivas

Compor outros estudos que auxiliam na divulgação sobre saúde mental na tecnologia

- Fomentar a discussão e criação de novos estudos acerca da saúde mental na indústria



<https://osmihelp.org/>

<https://www.imf.org/en/Publications/fandd/issues/2019/03/global-competition-for-technology-workers-costa>

<https://www.who.int/news-room/fact-sheets/detail/depression>

<https://brasscom.org.br/estudo-da-brasscom-aponta-demanda-de-797-mil-profissionais-de-tecnologia-ate-2025/>