

**Instituto Federal de Educação, Ciência e Tecnologia  
de São Paulo**

**Introdução a Ciência de Dados - ICDA6**

**Profa. Dra. Josceli M. Tenorio**

**Viviane de Santana Queiroz**

**SP3053601**

**João Vitor Silva Bispo**

**SP3052672**

**Projeto - Ciclo de Vida dos Dados**

*Uma análise sobre a saúde mental na indústria de  
software*

São Paulo

2022

## INTRODUÇÃO

Cerca de 280 milhões de pessoas ao redor do mundo sofrem de depressão, de acordo com o artigo “Depression” publicado no site da World Health Organization (WHO) em 13 de Setembro de 2021. Especialistas da área, como os psiquiatras brasileiros Ivo Querino e Ana Beatriz, vive-se uma pandemia silenciosa, em que muitos sofrem de transtornos mentais que levam até mesmo ao suicídio.

Diante desse fato e da importância de se discutir sobre o tema, este trabalho se propõe a analisar dois datasets extraídos do site Kaggle.com que abordam a questão de saúde mental. O objetivo é expandir o entendimento de como os profissionais da indústria de software têm sido impactados mentalmente, quais fatores podem influenciar positiva ou negativamente para que se previna o crescimento de distúrbios psicológicos nesse meio e assim levantar possíveis estratégias e insights.

## DATASETS ESCOLHIDOS

### Dataset 1: **Mental Health in the Tech Industry**

Fonte: <https://www.kaggle.com/datasets/anth7310/mental-health-in-the-tech-industry>

Origem dos dados:

Nesse dataset, foram compilados dados de pesquisas feitas com pessoas que trabalham na indústria de software pela empresa estrangeira sem fins lucrativos Open Sourcing Mental Health (OSMI). Anualmente, são disponibilizados questionários que contém mais de 100 questões que se relacionam com a saúde mental dos colaboradores das mais diversas empresas e de todo o mundo. Os integrantes atuam como embaixadores que visam difundir conhecimento e ajuda sobre o tema no mercado de tecnologia.

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

Fonte: <https://www.kaggle.com/datasets/fernandoretamales/health-expenditure-and-suicide-rates>

Origem dos dados:

Nesse segundo dataset usado, foram compilados dados sobre os gastos com saúde e níveis de depressão e suicídio ao redor do mundo. É um trabalho que se baseia nas informações oficiais disponibilizadas pelo WHO que agregam os períodos de 2000 a 2019.

## VISÃO GERAL DO CICLO DE VIDA DOS DADOS

### 1. Coleta

Para esse trabalho, a coleta se deu por meio do download dos arquivos disponibilizados na plataforma Kaggle para os 2 datasets especificados no índice anterior.

### 2. Processamento

Após o download, todos os arquivos foram convertidos para o formato CSV a fim de padronizar a manipulação desses dentro da ferramenta RStudio.

### 3. Análise

Foram efetuadas as análises explícita, exploratória e implícita, todas por meio da linguagem R, na ferramenta RStudio, e as quais serão detalhadas no próximo capítulo.

### 4. Publicação

Alguns resultados obtidos foram adicionados em um dashboard Shiny.

### 5. Armazenamento

Os dados dos dois datasets escolhidos somados aos gerados ao longo da criação desse trabalho se manterão armazenados em repositório open source na plataforma Github. Não havia e não foi gerado nenhum dado sensível que comprometesse a LGPD.

### 6. Exclusão

O repositório no Github pode sofrer exclusão em caso de exclusão da conta dos dois integrantes do grupo ou por alguma limitação futura da plataforma em relação ao armazenamento.

### 7. Reutilização

Por estar em uma plataforma open source tal como o Kaggle, o resultado do trabalho fica disponível a outros usuários que podem reutilizá-lo livremente.

## ANÁLISE EXPLÍCITA

### CRISP-DM: Entendimento do negócio

Ainda na etapa de análise explícita, a análise ampla do que ambos os datasets ofertavam de dados levou ao estabelecimento de um objetivo de negócio mais específico com base no que foi analisado:

Buscar um modelo que indique se o profissional da área de tecnologia já procurou ajuda profissional para tratar questões mentais

Dessa forma, é possível que indivíduos e instituições possam desenvolver estratégias para abordar o assunto e tratá-lo de forma eficiente.

### CRISP-DM: Entendimento dos dados

Estabelecido o objetivo de negócio, os dados foram *coletados, transformados e carregados* na ferramenta RStudio a fim de identificar seus outliers e itens faltantes. Eles foram conforme necessário por moda, em caso de dados do tipo categóricos, e por mediana, no caso de dados quantitativos.

Também foram exploradas as escalas e tipos específicos de cada dado e, em paralelo, ocorreu um levantamento preliminar de possíveis modelos de Machine Learning que pudessem ser utilizados na etapa de Análise Implícita de acordo com o que estava sendo analisado nesse primeiro momento.

Contudo, o foco permaneceu em validar se os dados escolhidos apoiavam de forma consistente o objetivo de negócio, e chegou-se à conclusão que sim.

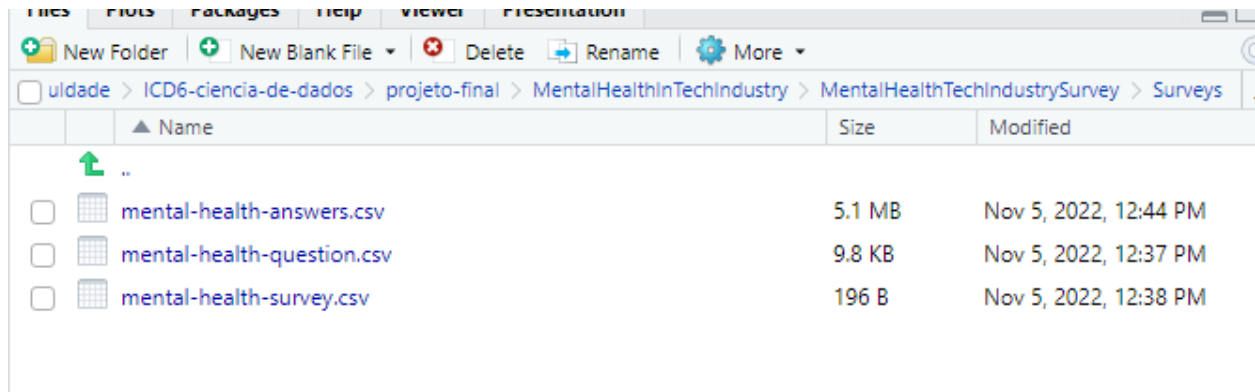
De forma mais específica, seguem as etapas mais importantes seguidas nessa análise para cada um dos datasets:

#### Dataset 1: **Mental Health in the Tech Industry**

Objetivos:

- Associar perguntas, respostas e ano de pesquisa

Após transformar os arquivos de sqlite para csv, foram gerados arquivos apartados conforme a imagem seguir:



Screenshot RStudio - árvore de arquivos Dataset 1

- Analisar as perguntas de forma quantitativa e qualitativa

Após a associação dos três arquivos do Dataset 1 feita, foram analisadas as 118 perguntas presentes a fim de escolher as mais relevantes para o objetivo de análise da saúde mental no mercado de tecnologia. Para isso, foram estabelecidos dois critérios:

a. Perguntas mais respondidas (após a junção foi constatado que nem todos os participantes responderam a todas perguntas)

b. Perguntas com alto valor para o negócio e ao mesmo tempo genéricas, que pudessem abarcar o maior número possível de indivíduos sem restrição a grupos específicos

A partir disso, foram escolhidas as seguintes perguntas contidas no arquivo original mental-health-question.csv:

1. *What is your age?*
2. *What is your gender?*
3. *What country do you live in?*
4. *Do you have a family history of mental illness?*
5. *Have you ever sought treatment for a mental health disorder from a mental health professional?*
6. *Does your employer provide mental health benefits as part of healthcare coverage?*
7. *Would you bring up a mental health issue with a potential employer in an interview?*
8. *Do you currently have a mental health disorder?*

Dessa forma, o Dataset 1 iniciou com 4 classes de interesse: ID do usuário, ID do ano da pesquisa e ID da questão e ID da resposta.

9. *Do you currently have a mental health disorder?*
10. *What condition(s) have you been diagnosed with?*

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

Objetivos:

- Obter os dados dos anos de interesse que fossem os mesmos presentes no Dataset 1  
Sendo o Dataset 1 mais significativo quantitativamente - com 4218 observações contras as 531 observações do Dataset 2, ele foi tomado como referência para a decisão dos anos que seriam considerados entre 2000 e 2019, sendo esses:

2014, 2016, 2017, 2018 e 2019

- Entendimento e validação da pertinência das colunas presentes no Dataset 2

Foram escolhidas as seguintes colunas para prosseguir na análise seguidas da descrição adicionada pelos próprios responsáveis pelo Dataset 2 no Kaggle:

Deaths\_Suicides:Deaths by suicides according to ICD10

HExp\_Pctage\_Y - Health Expenditure as % of GDP

MHExp\_Pctage\_2011 - Mental Health Expenditure index in 2011

Dep\_Num\_2015 - Depression estimate index in 2015

Suicide\_p100 - Number of suicides per 100,000 population

Tais colunas estão alinhadas ao objetivo de negócio estabelecido ao passo que possibilitam uma visão macro da questão da saúde mental que pode servir como comparativo ou referencial dentro da indústria de software.

## **ANÁLISE EXPLORATÓRIA (EDA)**

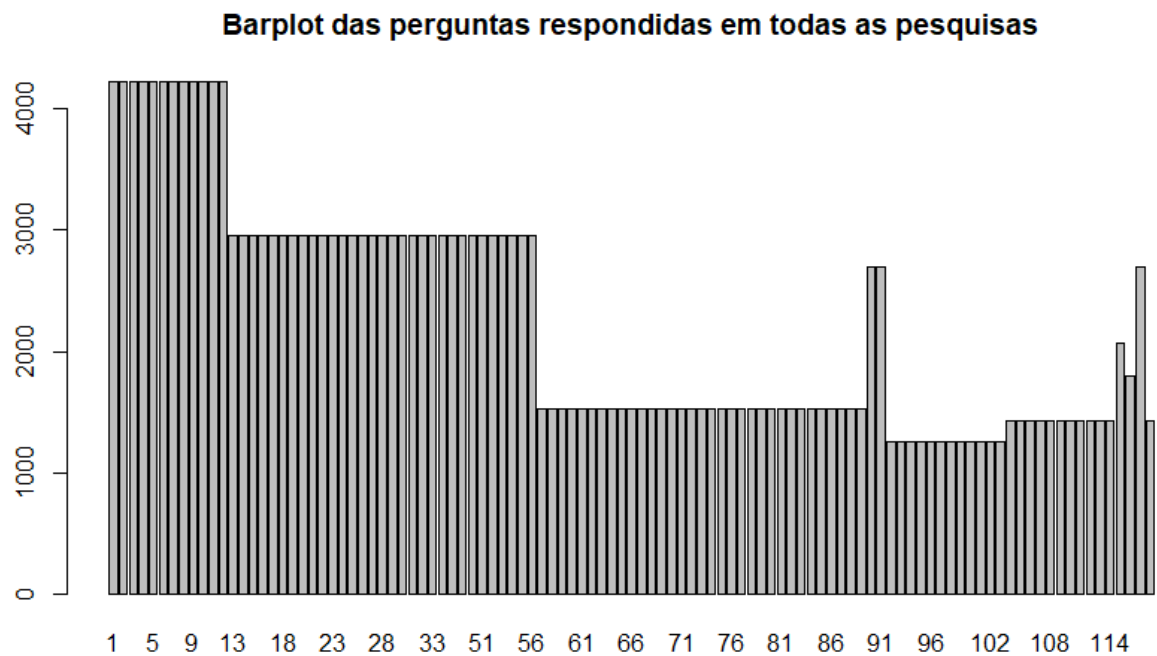
### **CRISP-DM: Preparação dos Dados**

Após o entendimento inicial dos dados, limpeza e tratamento daqueles a serem explorados, nessa etapa de preparação de dados foram construídas visualizações e retornos que pudessem sustentar, de fato, a pertinência de continuar usando as variáveis escolhidas inicialmente.

Nessa etapa, foi construído também uma nova tabela que compilasse todos os valores de interesse em ambos os datasets.

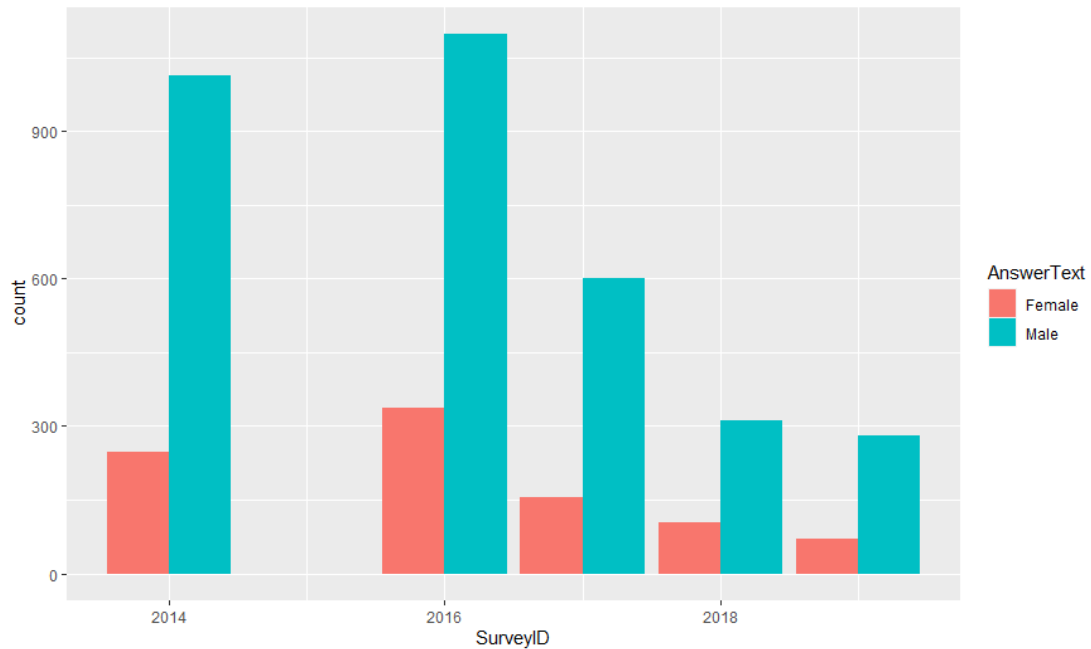
## Dataset 1: Mental Health in the Tech Industry

- **Visualização das perguntas mais respondidas:**



Apesar do alto número de questões, é possível confirmar que as mais respondidas concentram-se no início e tendem a decair. Em conta disso, permaneceu-se com as perguntas previamente escolhidas, pois continuavam a atender os critérios estabelecidos na Análise Explícita.

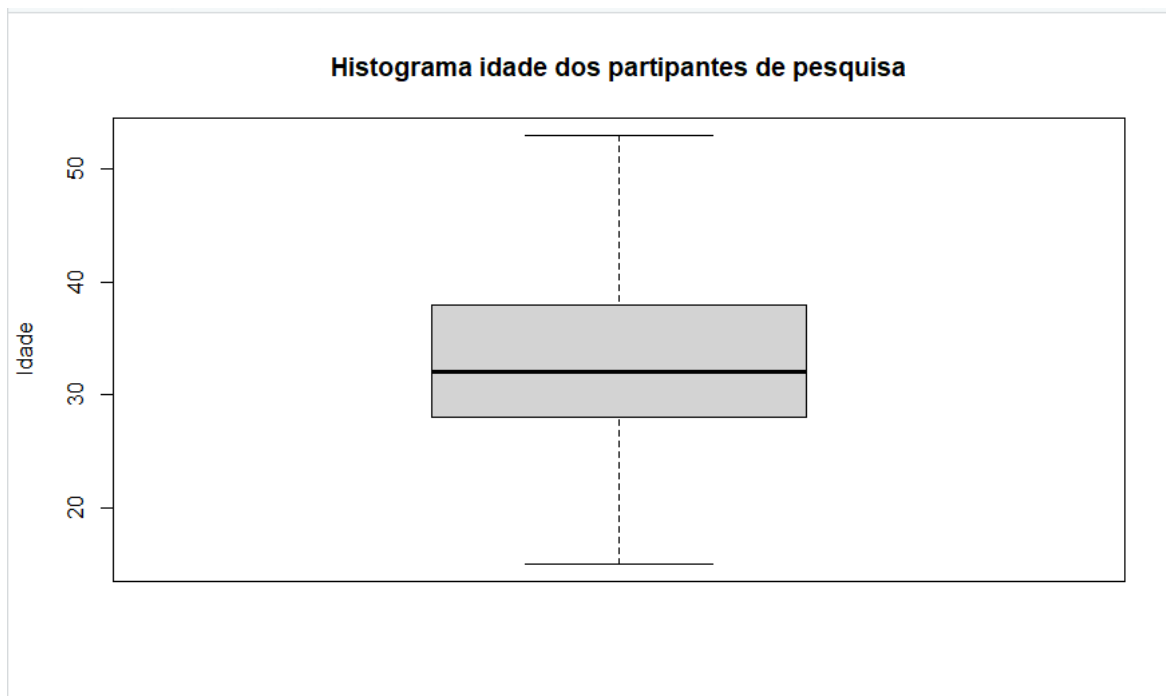
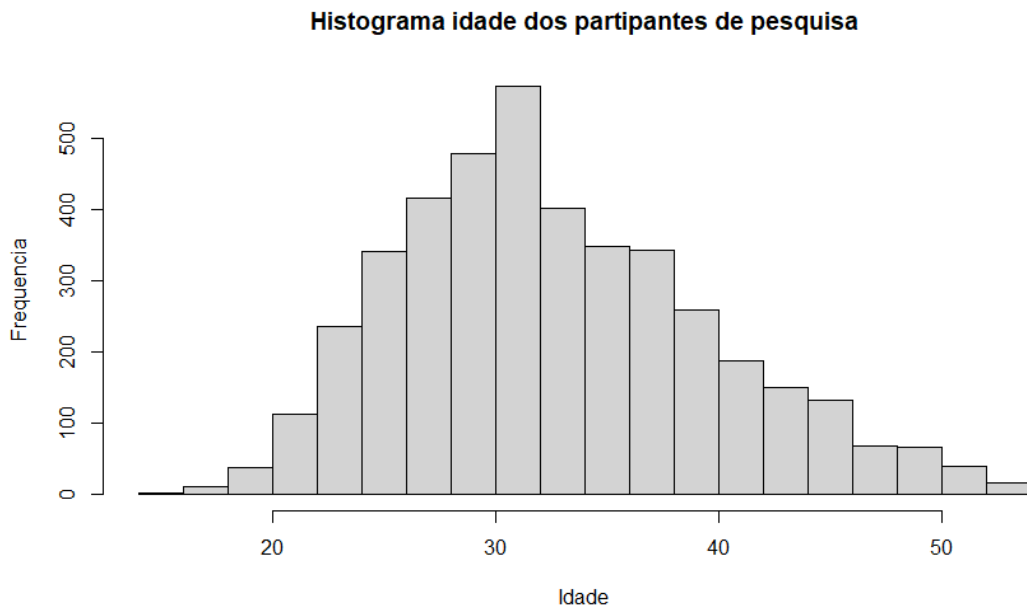
- Total de respostas a cada ano por gênero:



Observa-se uma predominância do sexo masculino em todos os anos de pesquisa realizada. De fato, o mercado de tecnologia ainda é composto em diversos setores majoritariamente por pessoas do sexo masculino. Também se mantém um dado de interesse por isso permaneceu na análise.

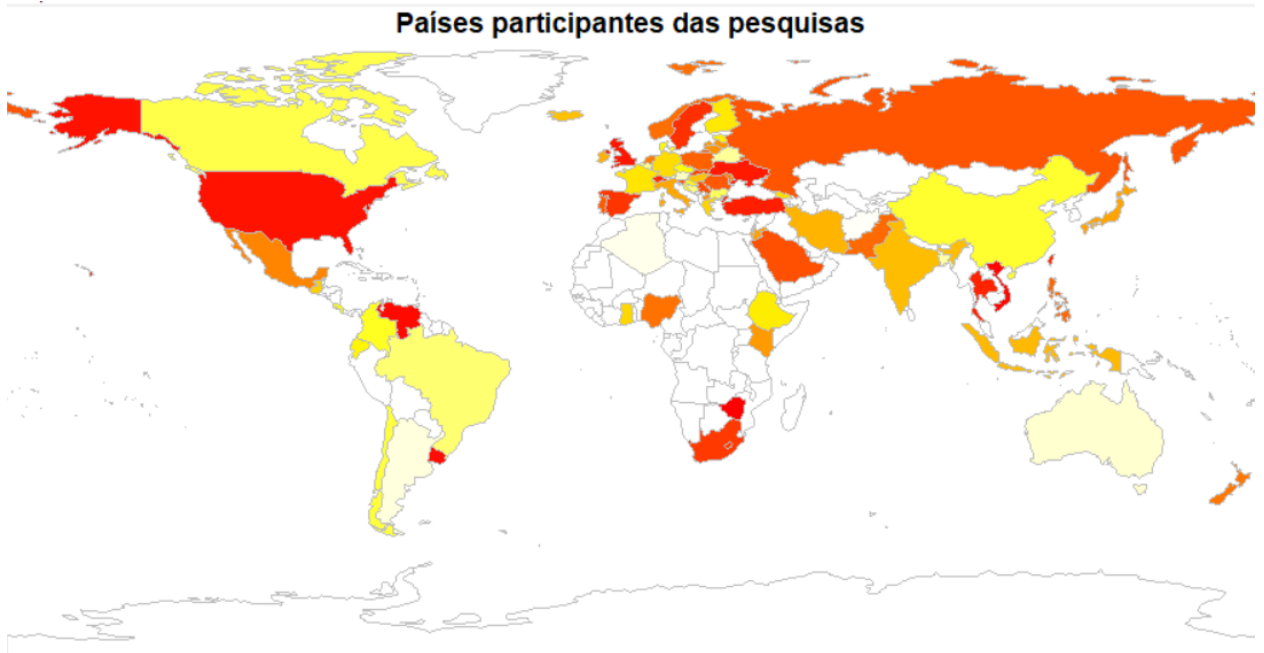


- **Distribuição da idade a cada ano de pesquisa**



Observa-se uma concentração entre os 30 anos e a idade mais recorrente foi a de participantes com 32 anos.

- **Países participantes das pesquisas**

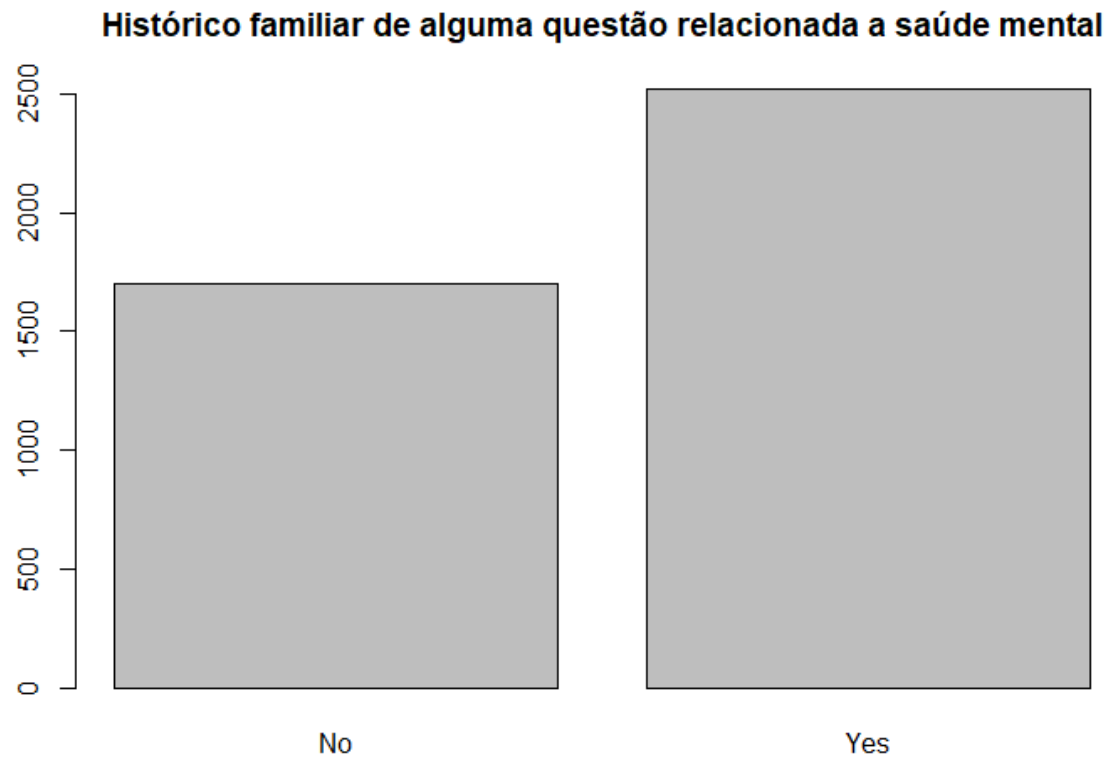


A pesquisa foi respondida por uma diversidade de países e com os filtros realizados foram analisadas 2958 respostas. Os países com mais respostas foram:

1. Estados Unidos - 2608 respostas - representa 62% das respostas totais
2. Reino Unido - 482 respostas - representa 11% das respostas totais
3. Canadá - 199 respostas - representa 5% das respostas totais
4. Germany - 136 respostas - representa 3% das respostas totais

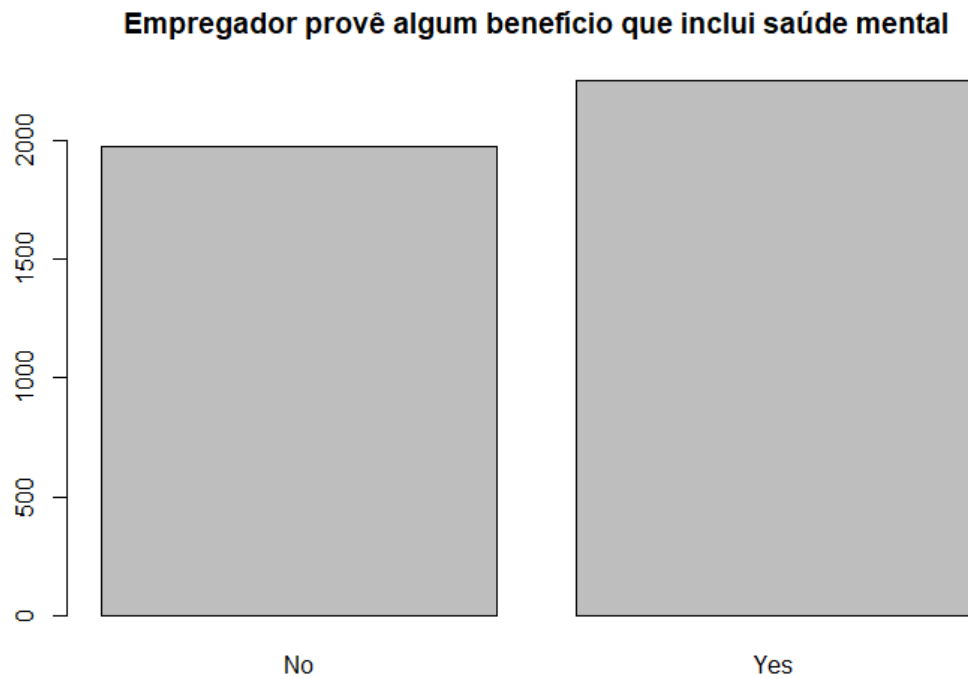
- **Histórico familiar de distúrbio mental**

---



Nessa visualização se observa um número expressivo de respostas positivas para um histórico familiar de distúrbio mental, representando 60% da amostra. Dado o contexto, continua sendo uma variável relevante para a análise.

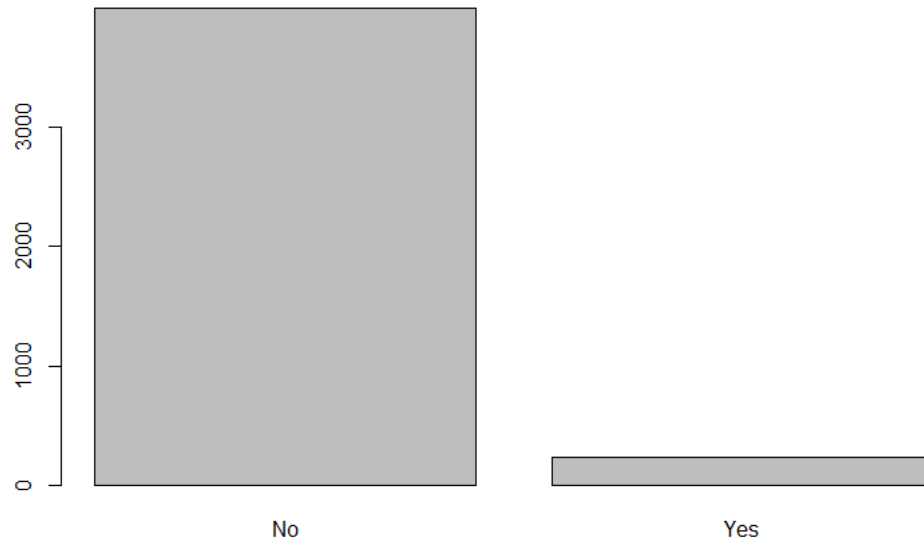
- **Empregador provê benefícios que abrangem saúde mental**



Nessa visualização, já se encontra um equilíbrio entre os participantes. Para 53% deles é oferecido algum tipo de auxílio à saúde mental, já para 47% não.

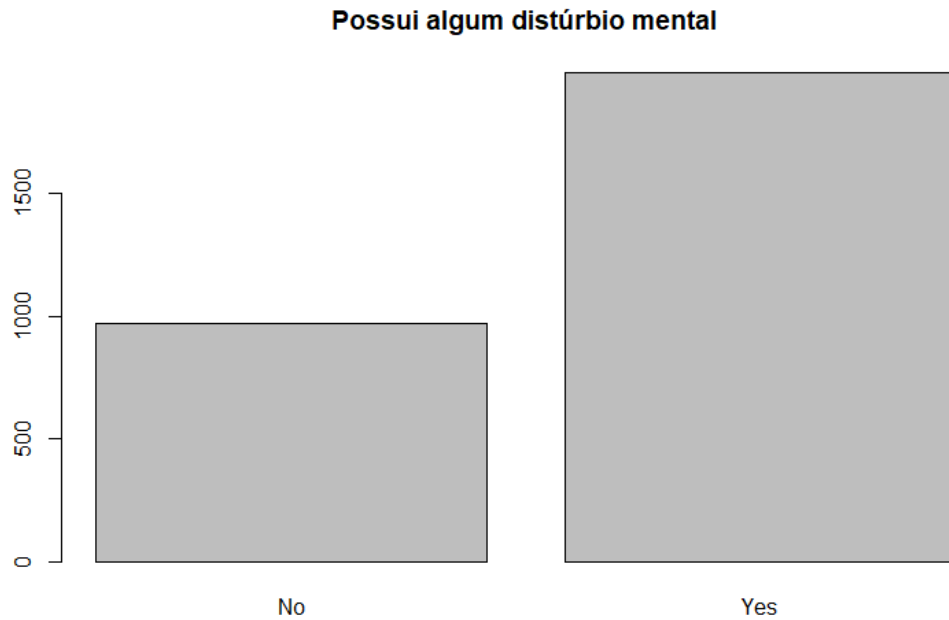
- **Abordar saúde mental em entrevista de emprego**

**Abordaria questão de saúde mental pessoal em uma entrevista de emprego**



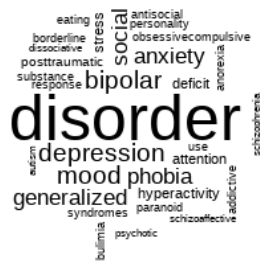
Nessa visualização, percebe-se até então a maior divergência entre as variáveis binomiais até o momento exploradas. 95% dos participantes não abordariam alguma questão de saúde mental em uma entrevista de emprego. Esse número já pode gerar insights sobre o quão confortáveis e seguras as pessoas se sentem para abordar o tópico.

- Possui algum distúrbio mental



67% dos participantes alegaram possuir algum distúrbio mental, número também expressivo e que confirma fazer sentido continuar observando essa variável.

- Diagnósticos mais comuns - nuvem de palavras



O transtorno mais recorrente na análise foi o classificado como "Mood Disorder (Depression, Bipolar Disorder, etc)", aparecendo em 34% das respostas, seguido de "Anxiety Disorder (Generalized, Social, Phobia, etc)"

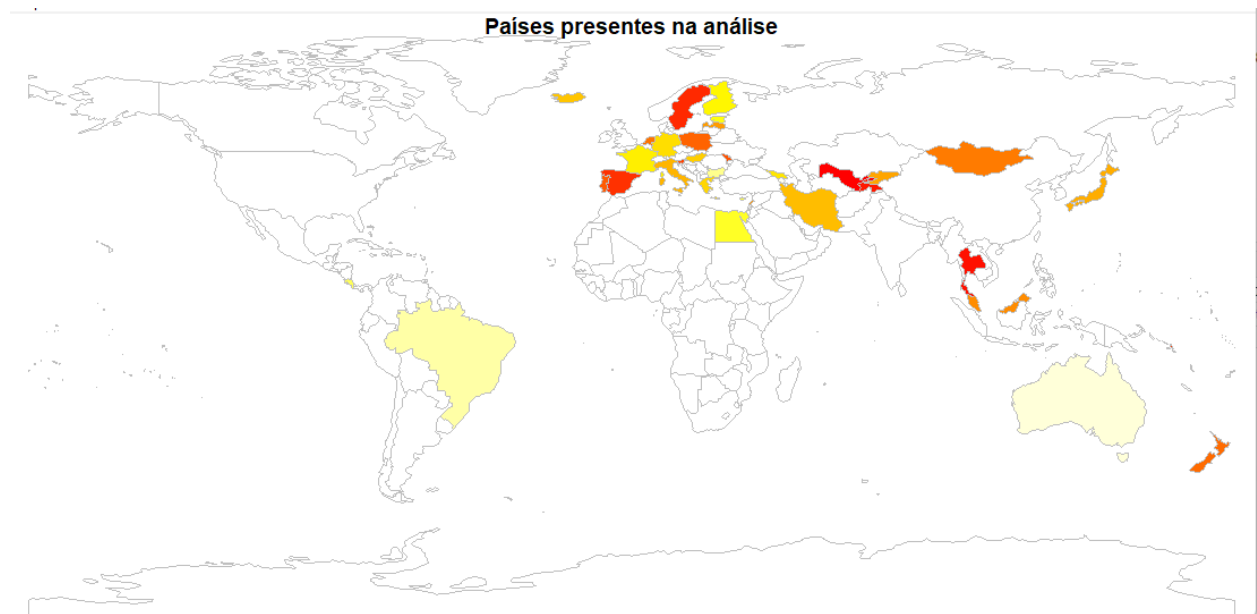
com 29% e “Attention Deficit Hyperactivity Disorder” com 10%.

Após a confirmação das variáveis pertinentes para prosseguir à análise, foi gerado um novo arquivo CSV `mental-health-in-tech-overview.csv` que as agrupa de forma a facilitar a construção do modelo.

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

Nesse segundo dataset, buscou-seo entender a correlação entre os gastos com saúde mental e o impacto em índices como diagnósticos de depressão e o total de suicídios a cada 100 mil habitantes.

Para ter uma visualização do alcance desses dados, foi gerado também uma visualização do mapa mundi com a frequência dos países participantes. A partir dele, constatou-se que frente ao Dataset 1, o número de países é significativamente menor (531), bem como o total de observações dentre os anos de 2014 a 2019, excluindo-se o ano de 2015, como aponta a imagem a seguir:



Aqui a configuração dos países com mais respostas também difere e cada um representa

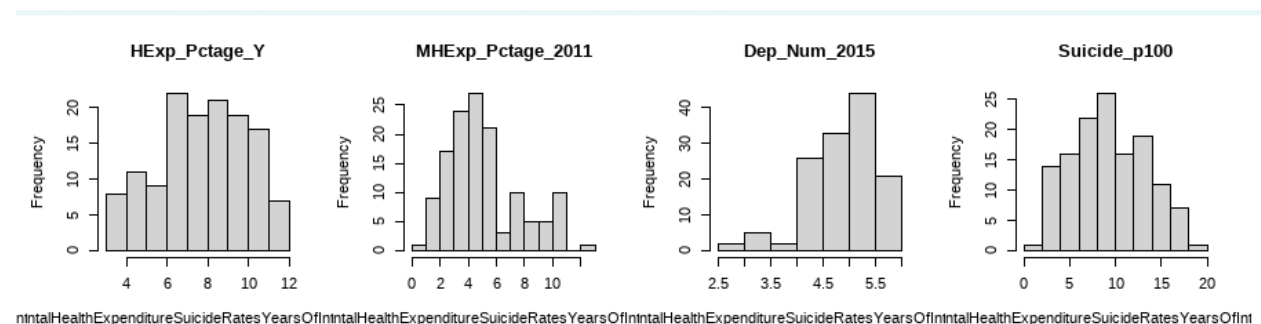
4% dos dados totais, sendo 7 deles:

Armênia, Austrália, Bulgária, Egito, Geórgia, Alemanha e Grécia.

Considerando o objetivo de entender os possíveis impactos dos investimentos em saúde mental nesses países, foi gerada uma tabela de correlação:

```
> correlationTable # print
# A tibble: 6 x 3
  var1          var2      coef_corr
  <fct>        <fct>      <dbl>
1 MHExp_Pctage_2011 Deaths_Suicides -0.140
2 Suicide_p100 Deaths_Suicides 0.0376
3 Deaths_Suicides MHExp_Pctage_2011 -0.140
4 Suicide_p100 MHExp_Pctage_2011 0.238
5 Deaths_Suicides Suicide_p100 0.0376
6 MHExp_Pctage_2011 Suicide_p100 0.238
> |
```

Tem-se uma correlação negativa entre os índices de suicídio e investimento em saúde mental. Dada a natureza desse tipo de análise, o menor investimento em saúde mental não pode ser apontado como causa raiz, pois correlação não é causa, mas torna-se um evento de interesse e que pode ser analisado com mais profundidade.



Nesse gráfico, pode-se observar a frequência para as variáveis de maior interesse e assim constatar que existe uma menor quantidade de recursos destinados à saúde mental, contudo, é esperado pois esse valor advém do total investido em saúde como um todo. Além disso, o índice de depressão analisado tem frequência alta em suas maiores incidências.



Ao realizar o plot do correlograma, pode-se observar a correlação mediana e positiva entre o número da população e índice de suicídio.

Proporcionalmente, como é contabilizado a cada 100 mil habitantes, os índices de suicídio diminuem ao passo que a população total também diminui.



## ANÁLISE IMPLÍCITA

### CRISP-DM: Modelagem

Com os dados devidamente tratados e com maior entendimento de como eles se relacionam, foi determinado o modelo de Machine Learning a ser utilizado a fim de atender o objetivo de se ter uma predição que informe a probabilidade de o profissional da área de tecnologia já ter procurado ajuda profissional para tratar questões mentais. Com as informações e relações extraídas do Dataset 2, entendeu-se que identificar a causa raiz de distúrbios mentais é uma tarefa que requer um olhar multifatorial. Por conta disso, o modelo se aplicou ao Dataset 1 que traz de forma mais abrangente o contexto do grupo de indivíduos a ser analisado

## **Técnica de Machine Learning escolhida:**

### **Regressão Logística Múltipla**

O modelo foi aplicado na base dados do arquivo mental-health-in-tech-overview.csv gerado a partir das análises anteriormente expostas.

Ele foi escolhido pois permite que seja realizada a predição de valores tomados por uma variável categórica, neste caso binária, a partir de uma série de variáveis explicativas contínuas e/ou binárias, o que se encaixa na necessidade de negócio.

Além disso, ao buscar o contexto em que geralmente é aplicada essa técnica, como por exemplo na área da saúde para previsão de diagnósticos de doenças, o modelo também se caracterizou como favorável de ser utilizado.

Variável dependente e para a qual busca-se a predição:

*AlreadySeekMentalHealthAssistance* - indica se o indivíduo já buscou ajuda para saúde mental anteriormente

Variáveis independentes e as quais são relevantes para ajudar na acurácia da predição:

*Sex*

*Age*

*MentalHealthHistory*

*TalkAboutMentalHealthAtInterview*

*HasMentalHealthDisorder*

*Country*

### **Técnica para validação interna - Hold out**

A base foi dividida em 70% de dados para treinamento e 30% para teste do modelo.

Além disso, foi estabelecido como sucesso as probabilidades com valores maiores que 0.5.

## **Resultados**

### **CRISP-DM: Avaliação**

A fim de se analisar os resultados obtidos com o modelo, foram usadas algumas técnicas e cálculos para identificar sua acurácia.

Matriz de confusão em treinamento:

```
> table(glm.pred.traning, AlreadySeekMentalHealthAssistance)
      AlreadySeekMentalHealthAssistance
glm.pred.traning  No  Yes
      No  1093  278
      Yes  162  538
>
```

Acurácia do modelo em treinamento de 79%:

```
> print(paste('Acurácia do modelo em treinamento: ', paste0(round(100 * accuracyvalue, 0), '%')))
[1] "Acurácia do modelo em treinamento: 79%"
>
```

Matriz de confusão em teste:

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

Acurácia do modelo em teste de 80%:

```
> print(paste('Acurácia do modelo em teste: ', paste0(round(100 * accuracyvalue, 0), '%')))
[1] "Acurácia do modelo em teste: 80%"
> |
```

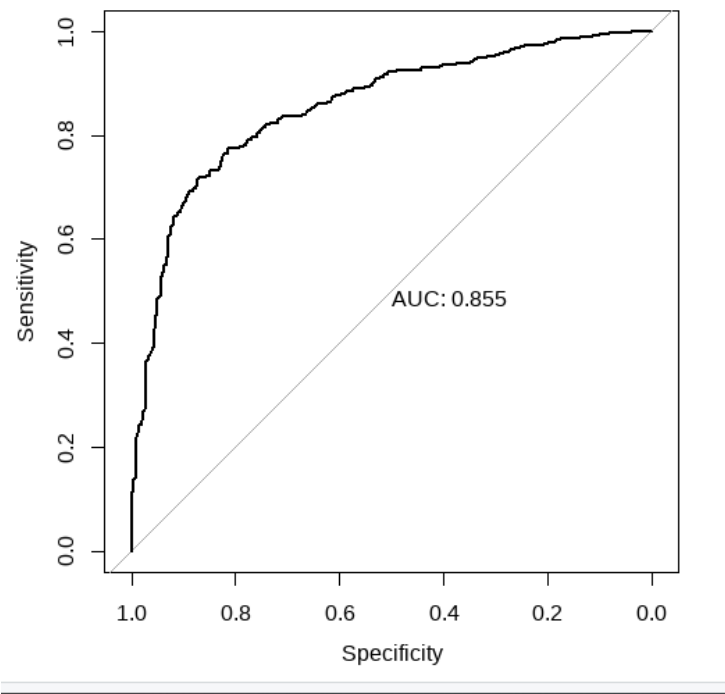
Cheragem de multicolinearidade:

```
> vif(model)
      GVIF Df GVIF^(1/(2*Df))
Sex      1.066109 1      1.032525
Age      1.084699 1      1.041489
MentalHealthHistory 1.129362 2      1.030880
TalkAboutMentalHealthAtInterview 1.024498 1      1.012175
HasMentalHealthDisorder 1.116761 1      1.056769
Country  1.219328 50      1.001985
~ |
```

Após testar o modelo com outras variáveis chegou-se a esse modelo final que não aponta valores que indicam multicolinearidade.

### Curva ROC e AUC - Sensibilidade e especificidade do modelo:

A partir do resultado obtido com a modelagem, obteve-se um valor de AUC de 0.85, o que está dentro do limiar esperado como aceitável para a precisão das classificações geradas pelo modelo de regressão logística múltipla criado.



## APLICABILIDADE DO MODELO

### CRISP-DM: Aplicação

Diante do resultado obtido com o modelo, podem ser consideradas algumas das seguintes aplicabilidades:

1. Uso pelo RH ou setor de Pessoas de empresas, além de Consultorias:

O bem estar físico e emocional dos colaboradores têm sido abordado como parte da cultura de algumas empresas hoje, como por exemplo a IBM e o Itaú Unibanco, que promovem ações internas para tratar especificamente desse tema.

Com o modelo desenvolvido, seria possível à área de pessoas dessas instituições promoverem pesquisas que pudessem fomentar o modelo e assim entender se faria sentido ofertar algum outro benefício no plano de saúde especificamente voltado à saúde mental, ou se algum setor tem uma maior tendência de possuir colaboradores que busquem por esse tipo de ajuda.

Dessa forma, a empresa pode até mesmo melhorar seu planejamento estratégico e financeiro, reter mais funcionários e aumentar a sua produtividade, pois eles se sentirão mais seguros e amparados psicologicamente no ambiente de trabalho. Além, é claro, de poder ter ações preventivas, sendo possível implementar o modelo nas áreas de contratação de pessoas.

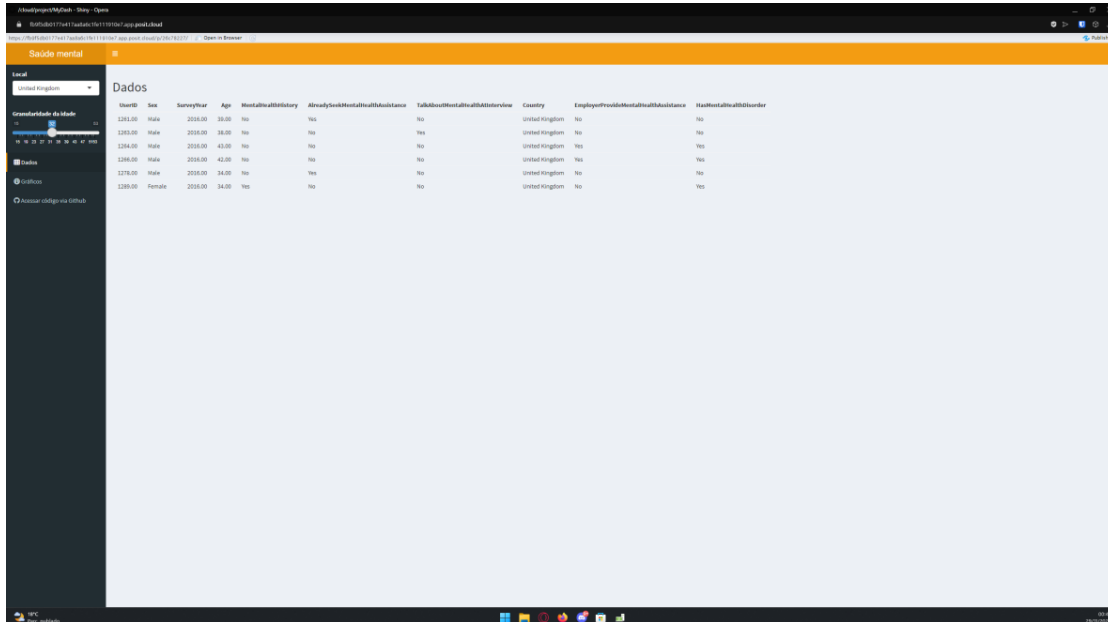
## 2. Uso por instituições de ensino e pesquisa

Além do mercado de trabalho, o meio acadêmico pode se beneficiar criando outras análises a partir do modelo ou até mesmo integrando-o a análises mais complexas. Com os dados sendo fomentados no meio acadêmico, outras áreas podem se beneficiar além do próprio mercado de trabalho: as estatísticas geradas podem ser referenciadas por grandes meios de veiculação midiática, como TV ou internet, de forma a democratizar o conhecimento sobre o assunto com a população em geral.

Distúrbios mentais afetam, como visto na introdução, milhões de pessoas por todo o mundo. Os números podem influenciar pessoas e autoridades pública a direcionarem mais atenção ao tema, o que pode contribuir para a redução de índices vistos durante o trabalho como o de suicídio e depressão.

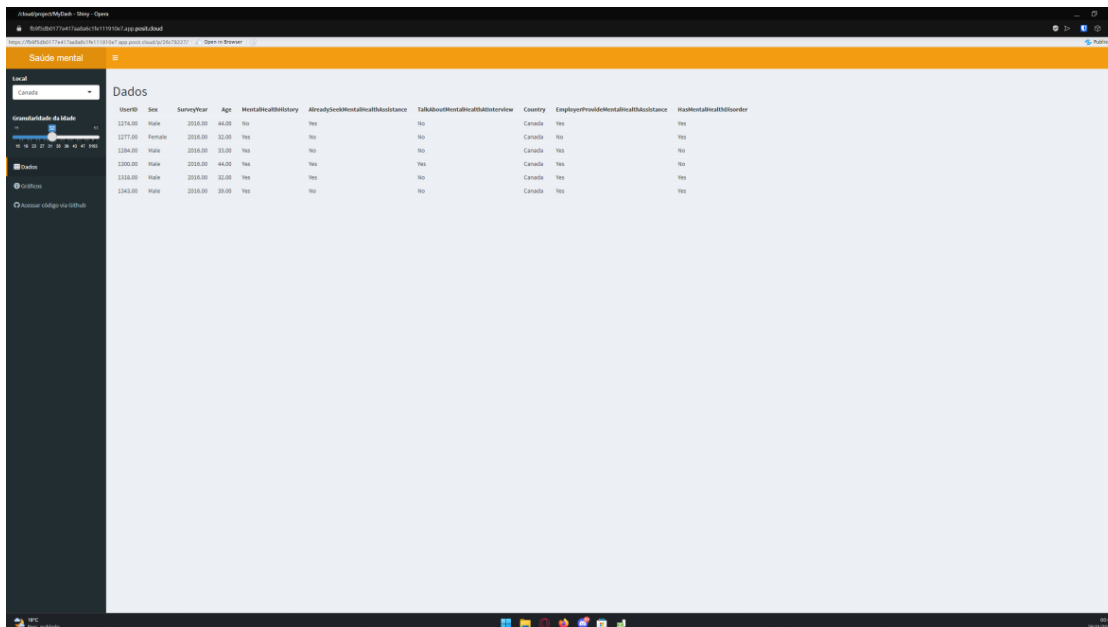
## DASHBOARD SHINY

Visualização geral da base mental-health-in-tech-overview.csv gerada nas análises, nele, é possível realizar o filtro por país. Os screenshots demonstram o filtro feito nos 4 países com maior incidência.



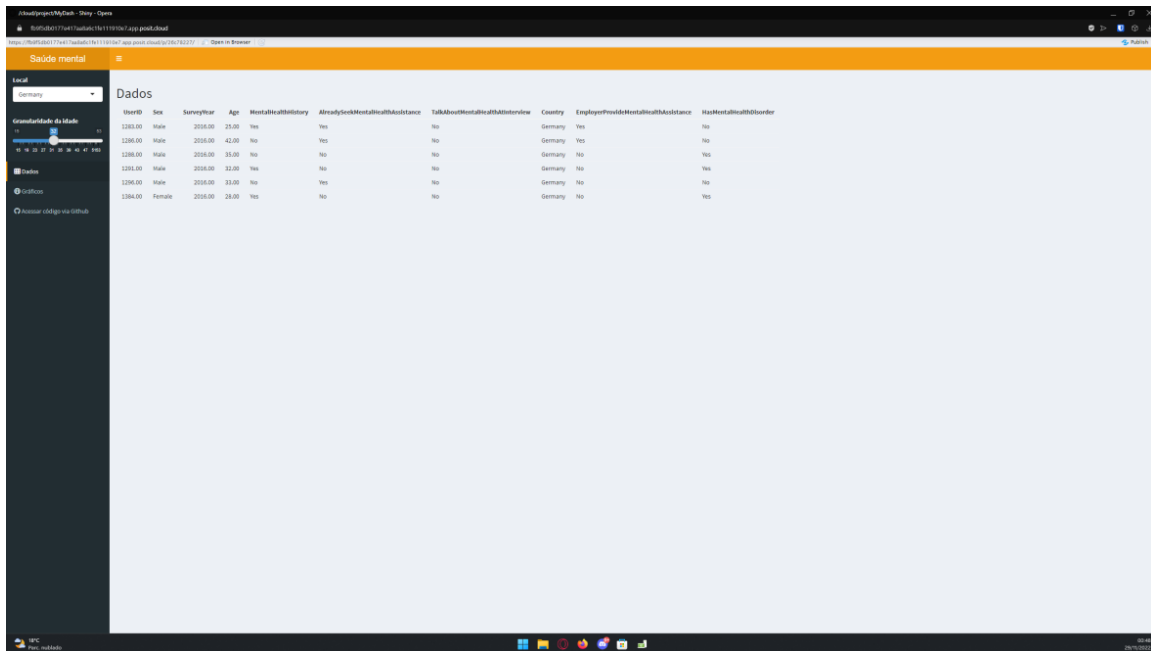
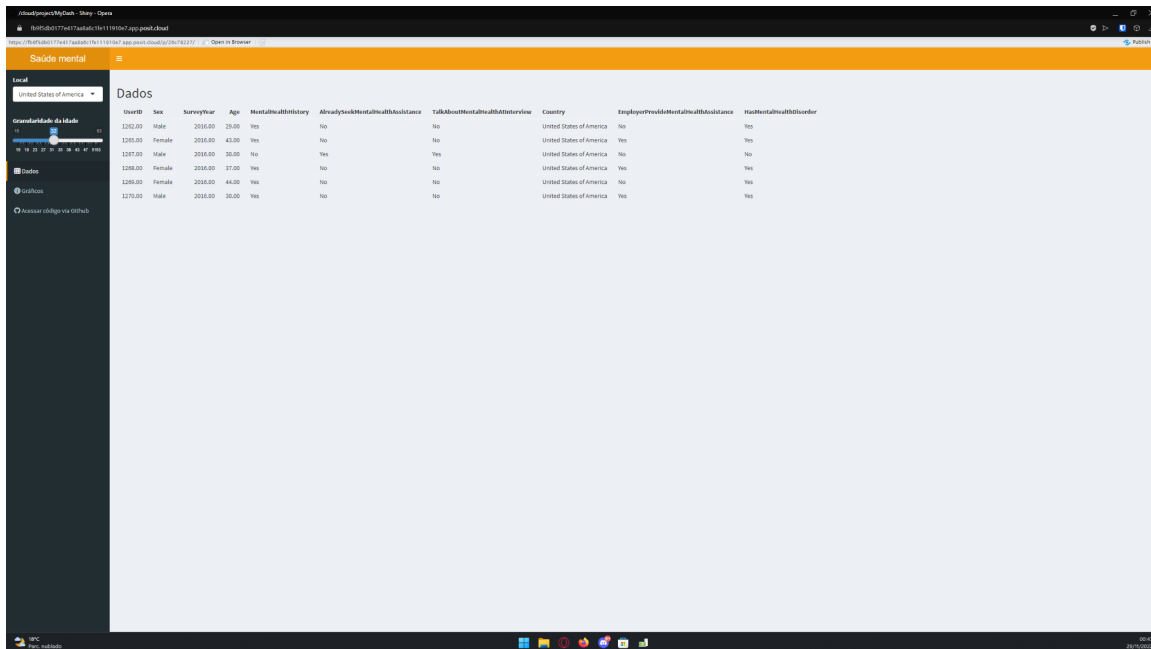
The screenshot shows the Shiny dashboard with the 'Local' dropdown set to 'United Kingdom'. The 'Dados' table displays the following data:

UserID	Sex	SurveyYear	Age	MentalHealthHistory	AlreadySeekMentalHealthAssistance	TalkedtoMentalHealthAdvisor	Country	EmployerProvidesMentalHealthAssistance	HasMentalHealthDisorder
1291.00	Male	2016.00	39.00	No	Yes	No	United Kingdom	No	No
1293.00	Male	2016.00	38.00	No	No	Yes	United Kingdom	No	No
1294.00	Male	2016.00	41.00	No	No	No	United Kingdom	Yes	Yes
1296.00	Male	2016.00	42.00	No	No	No	United Kingdom	No	Yes
1276.00	Male	2016.00	34.00	No	Yes	No	United Kingdom	No	No
1289.00	Female	2016.00	34.00	Yes	No	No	United Kingdom	No	Yes

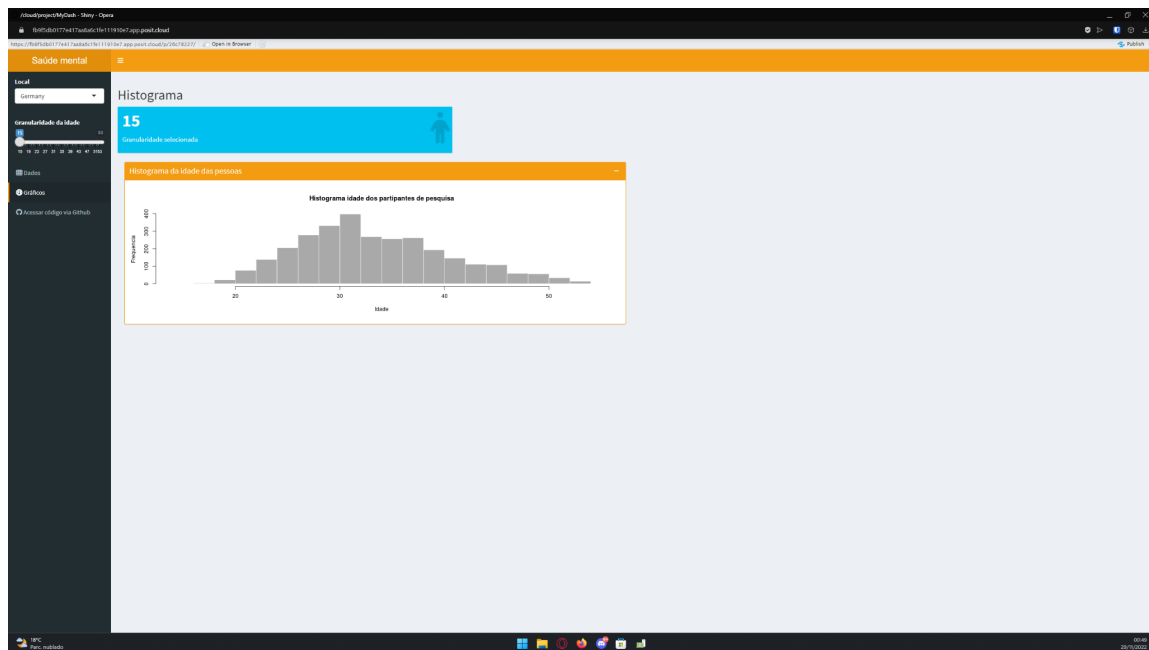
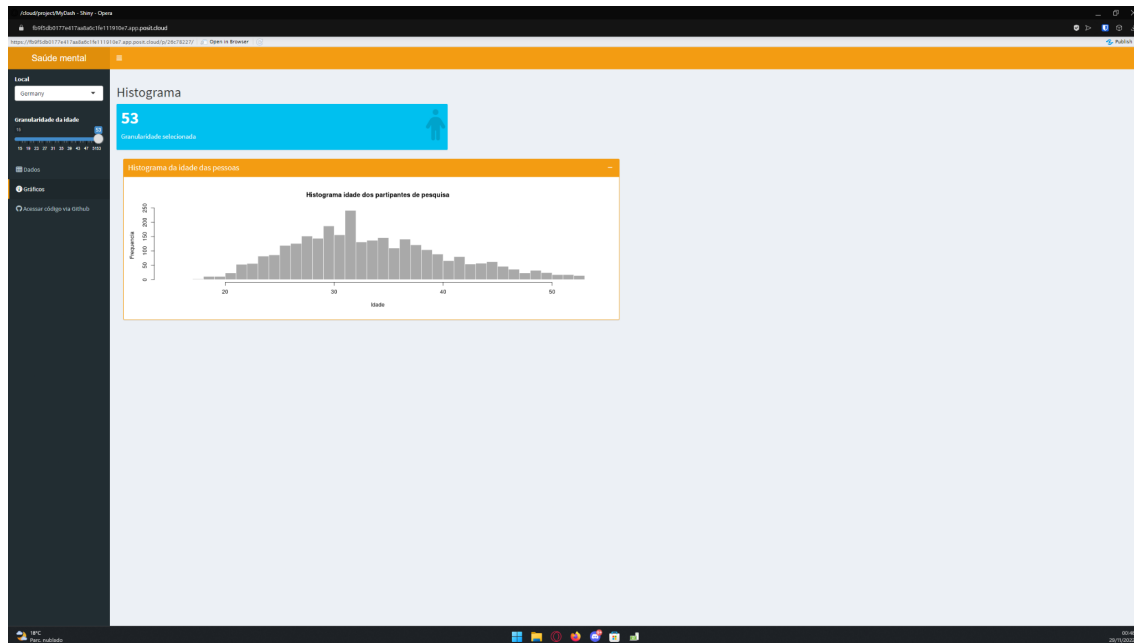


The screenshot shows the Shiny dashboard with the 'Local' dropdown set to 'Canada'. The 'Dados' table displays the following data:

UserID	Sex	SurveyYear	Age	MentalHealthHistory	AlreadySeekMentalHealthAssistance	TalkedtoMentalHealthAdvisor	Country	EmployerProvidesMentalHealthAssistance	HasMentalHealthDisorder
1274.00	Male	2016.00	46.00	No	Yes	No	Canada	Yes	Yes
1277.00	Female	2016.00	32.00	Yes	No	No	Canada	No	Yes
1284.00	Male	2016.00	33.00	Yes	No	No	Canada	Yes	No
1290.00	Male	2016.00	44.00	Yes	Yes	Yes	Canada	Yes	No
1288.00	Male	2016.00	32.00	Yes	No	No	Canada	Yes	Yes
1360.00	Male	2016.00	39.00	Yes	No	No	Canada	Yes	Yes



Histograma com slider que permite o filtro por granularidade das idades presentes na base. Nos screenshots é possível observar os valores máximos e mínimos.





## CONCLUSÃO

Com a motivação inicial de entender melhor se as pessoas na indústria de software têm buscado ajuda relacionada à saúde mental, chegou-se a um modelo preditivo com acurácia de 80% que analisa de forma multifatorial se uma pessoa tem a probabilidade de ter já ter buscado esse tipo de ajuda ou não.

Além do objetivo acadêmico ou comercial que pode ser associado, o trabalho traz à luz uma pauta muitas vezes negligenciada ou até mesmo tratada como tabu por alguns indivíduos. E tal como os dados evidenciaram, esse assunto impacta milhões de pessoas todos os anos, dentre esses, estão profissionais da área de tecnologia.

Com o mercado veloz, tecnologias evoluindo rapidamente e diversas dinâmicas de modelo de trabalho faz sentido se atentar ao aspecto da saúde mental desse grupo e como ele pode ser impactado, bem como quais ações preventivas podem ser tomadas. Dessa forma, o valor de negócio foi alcançado à medida que o modelo está aderente ao seu propósito inicial. Além disso, cada etapa de análise, explícita, exploratória e implícita produziu recursos com o potencial de gerar outras análises e modelos.

## REFERÊNCIAS

OSMI Home :: Open Sourcing Mental Illness - Changing how we talk about mental health in the tech community - Stronger Than Fear. Disponível em: <<https://osmihelp.org/>>.

Types of Data Analysis: A Guide | Built In. Disponível em: <<https://builtin.com/data-science/types-of-data-analysis>>.

WORLD HEALTH ORGANIZATION. Depression. Disponível em: <<https://www.who.int/news-room/fact-sheets/detail/depression>>.

Mental Health in the Tech Industry. Disponível em: <<https://www.kaggle.com/datasets/anth7310/mental-health-in-the-tech-industry>>. Acesso em: 29 nov. 2022.

Health Expenditure and Suicide Rates [2000-2019]. Disponível em: <<https://www.kaggle.com/datasets/fernandoretamales/health-expenditure-and-suicide-rates>>. Acesso em: 29 nov. 2022.

Logistic Regression in R Tutorial. Disponível em:

<<https://www.datacamp.com/tutorial/logistic-regression-R#visualizing-data>>. Acesso em: 29 nov. 2022.

RPubs - Logistic regression using R. Disponível em:

<[https://rpubs.com/juanhklopper/logistic\\_regression](https://rpubs.com/juanhklopper/logistic_regression)>. Acesso em: 29 nov. 2022.

R Documentation and manuals | R Documentation. Disponível em:

<<https://www.rdocumentation.org/>>.

MOURA, K. Ciclo de vida dos dados #2. Disponível em:

<<https://medium.com/@kvmoura/crisp-dm-79580b0d3ac4>>. Acesso em: 29 nov. 2022.

SPERANDEI, S. Understanding logistic regression analysis. *Biochemia Medica*, v. 24, n. 1, p. 12–18, 2014.

Logistic regression. Disponível em:

<[http://www.cookbook-r.com/Statistical\\_analysis/Logistic\\_regression/](http://www.cookbook-r.com/Statistical_analysis/Logistic_regression/)>.

Fitting a logistic regression model. Disponível em:

<[https://cran.r-project.org/web/packages/sjPlot/vignettes/plot\\_model\\_estimates.html](https://cran.r-project.org/web/packages/sjPlot/vignettes/plot_model_estimates.html)>. Acesso em: 29 nov. 2022.

MICHEL. Validação de Modelos em Machine Learning. Disponível em:

<<https://www.3dimensoes.com.br/post/valida%C3%A7%C3%A3o-de-modelos-em-machine-learning>>. Acesso em: 29 nov. 2022.

KULKARNI, A. R Shiny | Build Interactive Models With R Shiny | Modelling with R Shiny. Disponível em:

<<https://www.analyticsvidhya.com/blog/2021/05/build-interactive-models-with-r-shiny/>>. Acesso em: 29 nov. 2022.

SCIENCE, R. STATISTICS FOR P. Analyse Pseudo-R<sup>2</sup>, VIF scores and robust standard errors for generalised linear models in R. Disponível em:

<<https://rforpoliticalscience.com/2021/01/08/add-r2-vif-scores-and-robust-standard-errors-to-generalized-linear-models-in-r/>>. Acesso em: 29 nov. 2022.

KULKARNI, A. R Shiny | Build Interactive Models With R Shiny | Modelling with R Shiny. Disponível em:

<<https://www.analyticsvidhya.com/blog/2021/05/build-interactive-models-with-r-shiny/>>. Acesso em: 29 nov. 2022.