

# DATA SCIENCE & STATISTICAL COMPUTING

Prof. Dr. Carlos André M. Vieira

✉ profcarlos.vieira@fiap.com.br

# QUEM SOU EU?



- Analista de Modelagem Financeira na REAG Investimentos;
- Doutor em Ciências Contábeis (UFPB) na linha de Mercado de Capitais, com ênfase em finanças comportamentais e irracionalidade no mercado financeiro;
- Ex-professor de Cursos de Graduação e Pós-Graduação;
- Pesquisador na área de Contabilidade e Finanças com diversas pesquisas publicadas em eventos e periódicos nacionais;
- Revisor de artigos de revistas científicas na área de Contabilidade e Finanças;
- Currículo completo: <http://lattes.cnpq.br/5575529862603319>

# EMENTA

## 1º Semestre

- Introdução à Ciência de Dados
- Análise Exploratória e Estatística Descritiva
- Distribuições de Probabilidade
- Testes de Hipótese (Paramétricos e Não Paramétricos)
- Correlação e Regressão Linear

## 2º Semestre

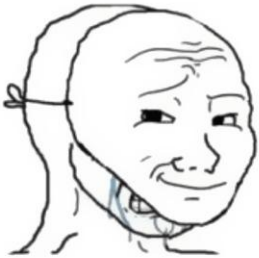
- Limpeza e Manipulação de Dados
  - Séries Temporais
  - Algoritmos de Regressão
  - Algoritmos de Classificação
-

# AVALIAÇÃO

- **Carga Horária:** 80hs
- **Sistema Avaliativo:**
  - ❑ **Checkpoints:** 3x no semestre, formatada pelo professor
  - ❑ **Sprints:** 4 mentorias/ano, 2 entregas/semestre, challenge com empresa parceira
  - ❑ **Global Solution:** Case de avaliação semestral contendo todas as disciplinas
- **Pesos:** 20% Checkpoints, 20% Sprints, 60% Global Solution.
- **Chamada ao início das aulas**

# AVALIAÇÃO

## DESIGNERS



Look, we have similar ideas.



No! You stole my idea.

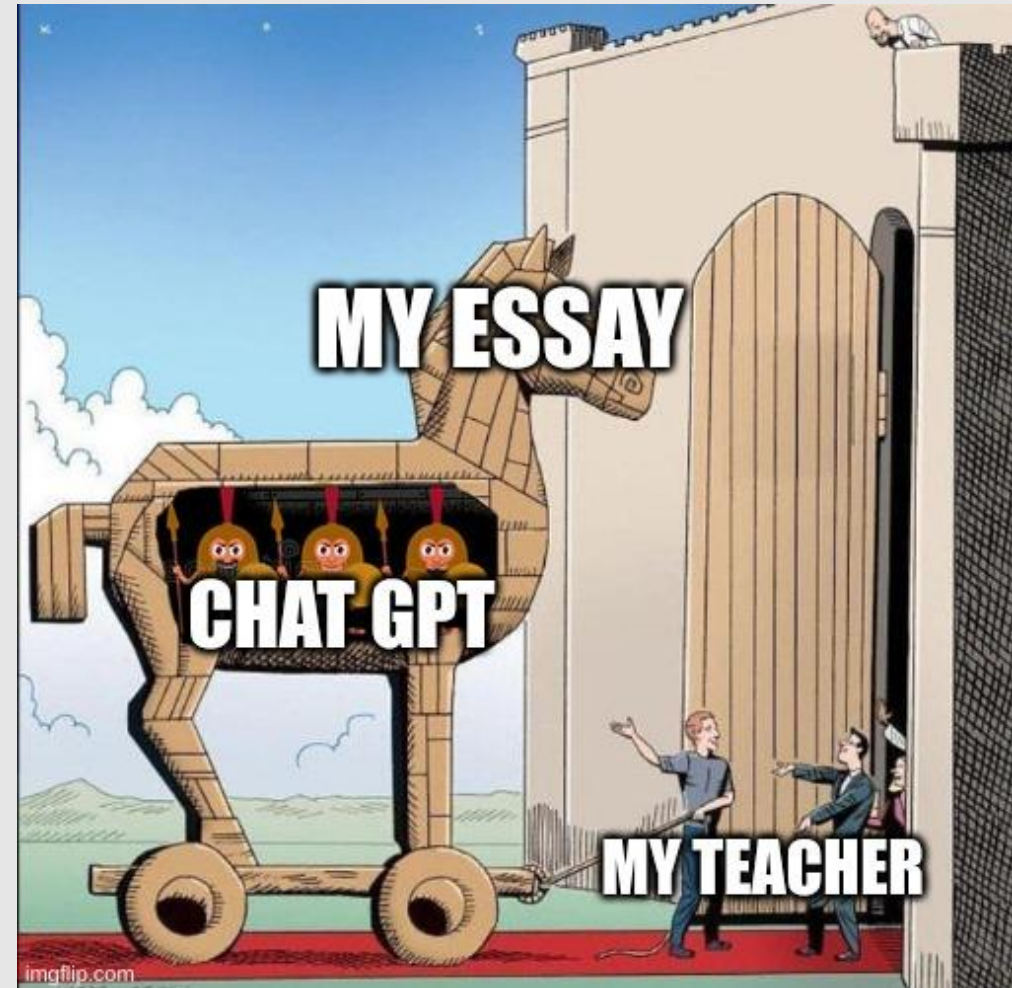
## PROGRAMMERS



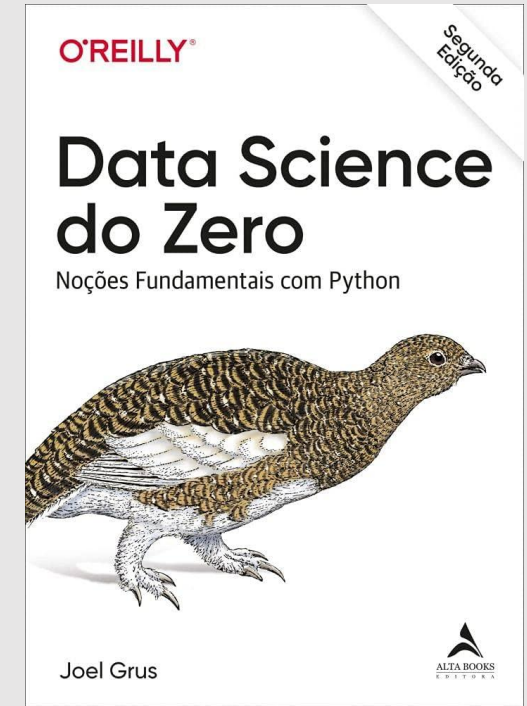
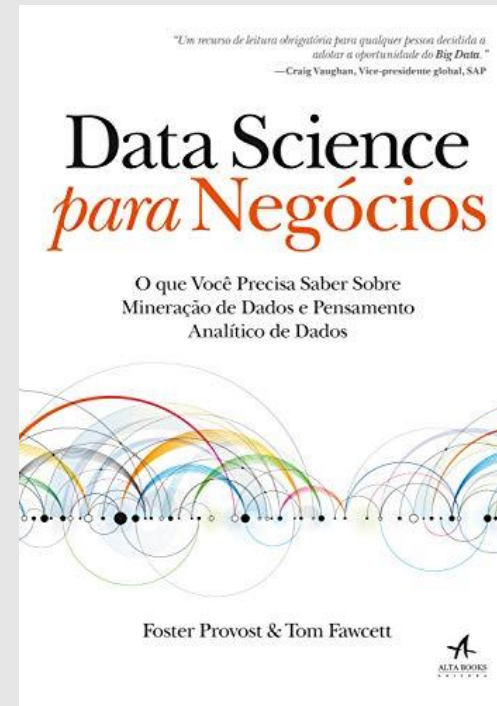
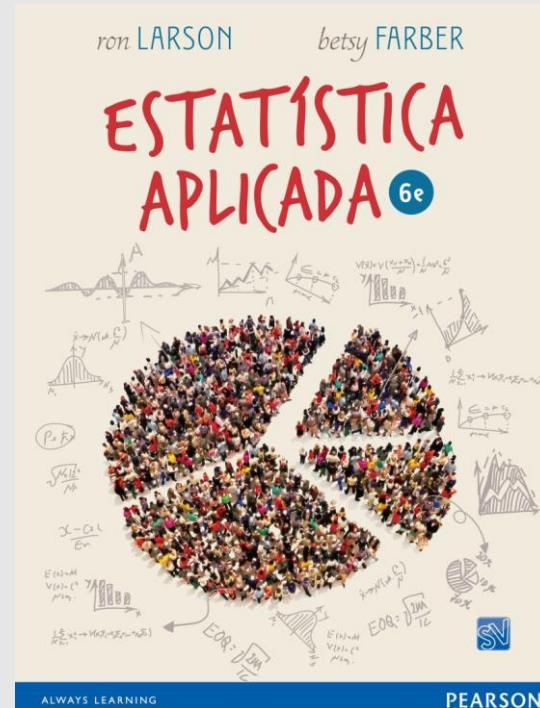
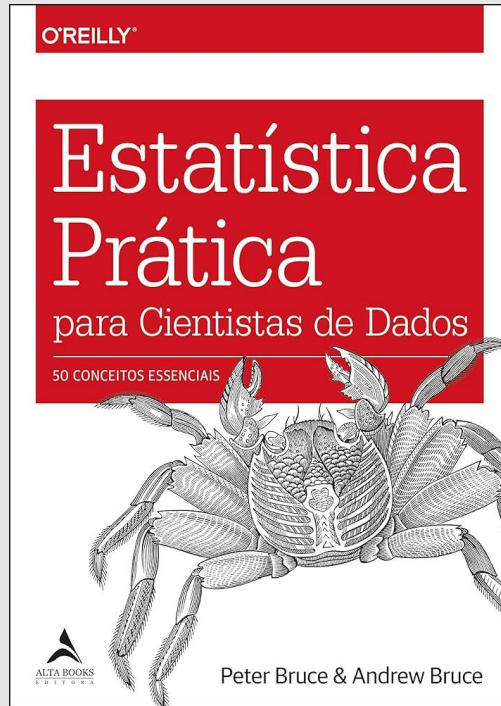
Man, I stole your code.



It's not my code.



# BIBLIOGRAFIA





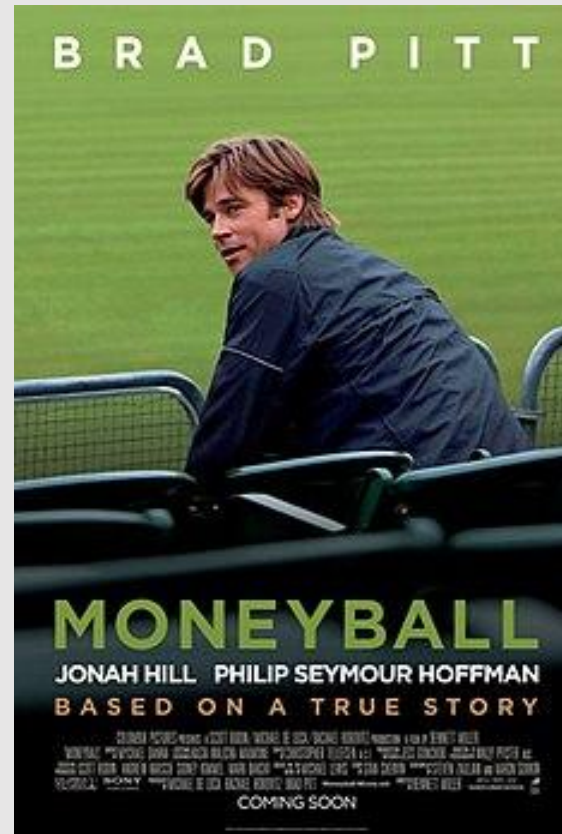
# O QUE É ANÁLISE DE DADOS?

FIAP



# INTRODUÇÃO À ANÁLISE DE DADOS

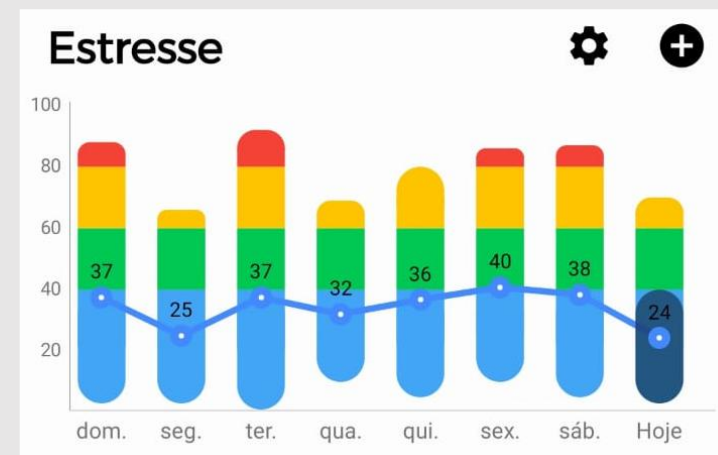
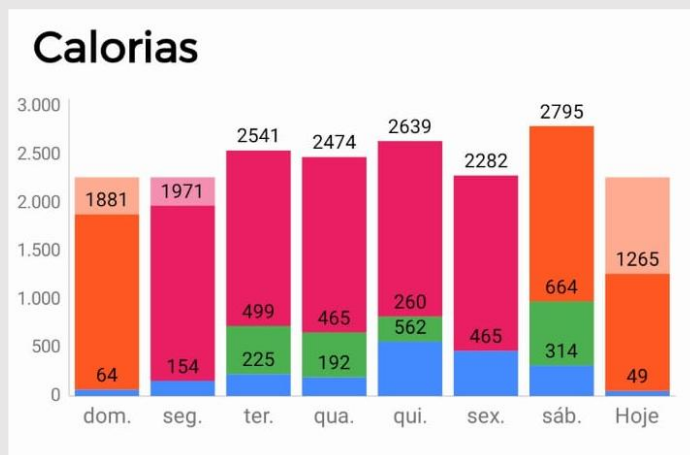
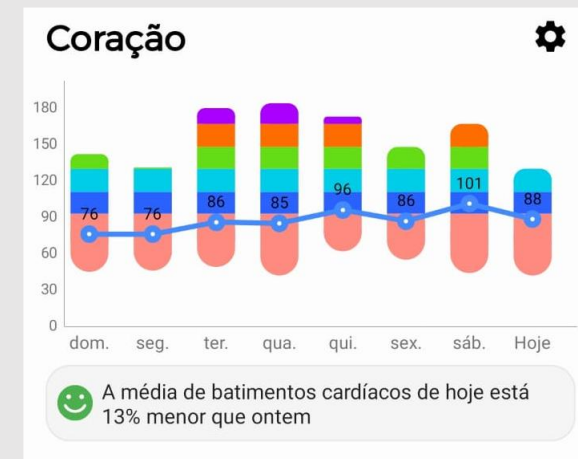
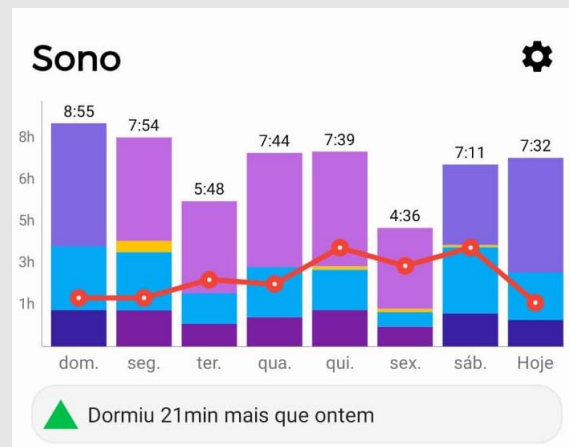
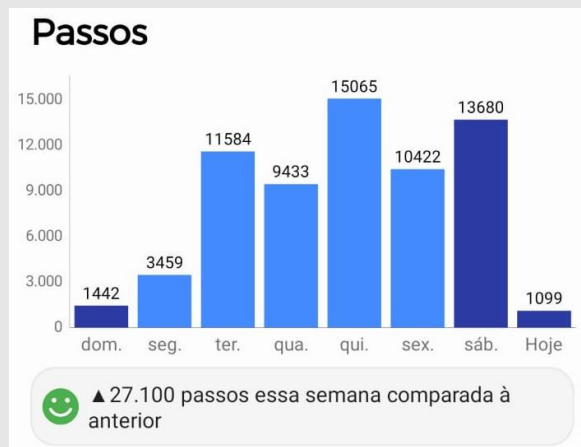
- **Motivação;**
- **Áreas:**
  - ☐ Finanças;
  - ☐ Marketing;
  - ☐ Vendas;
  - ☐ Recursos Humanos;
  - ☐ Esportes.



**“It’s unbelievable how much you don’t know about the game you’ve been playing all your life”**



# ANÁLISE DE DADOS

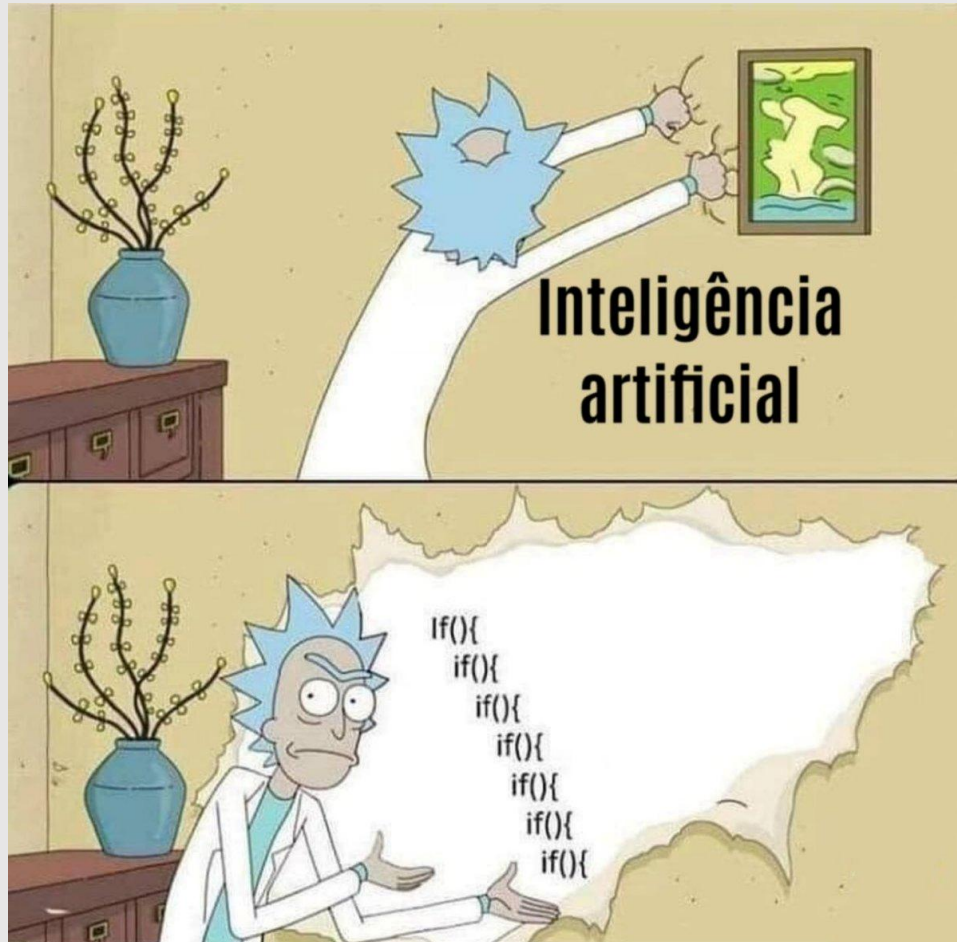


# ANÁLISE DE DADOS

- **Análise/Analytics:** Processo de desenvolvimento de ações ou recomendações para ações baseado nos insights gerados através de dados históricos;
- Combinação do uso da computação, técnicas de gestão e estatística para a resolução de problemas reais.



# ANÁLISE DE DADOS



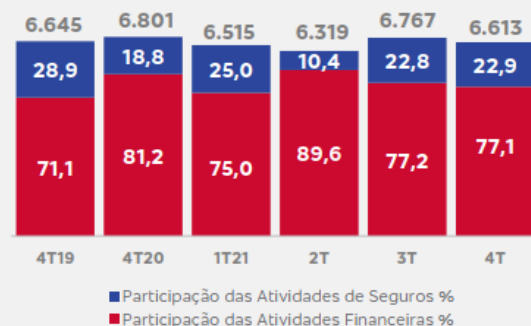
# ANÁLISE DE DADOS

Análise de Dados			
	Descritiva	Preditiva	Prescritiva
Questões	O que aconteceu? O que está acontecendo?	O que vai acontecer? Por que vai acontecer?	O que deveria ser feito? Por que deveria ser feito?
Facilitadores	Relatórios Dashboards Tabelas Data Warehousing	Previsão Data mining Text Mining Web/media mining	Otimização Simulação Modelos de decisão
Resultados	Problemas e oportunidades do negócio	Projeções acuradas de eventos e resultados futuros	As melhores decisões e ações de negócio possíveis

# ANÁLISE DESCRITIVA

Lucro Líquido Recorrente

R\$ milhões



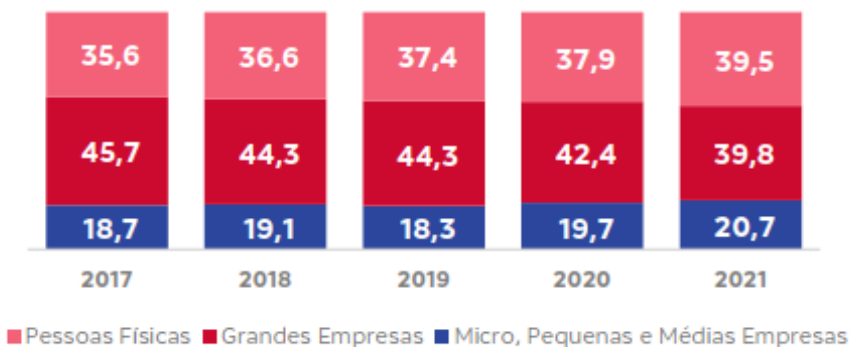
ROAE Acumulado e Trimestral

%



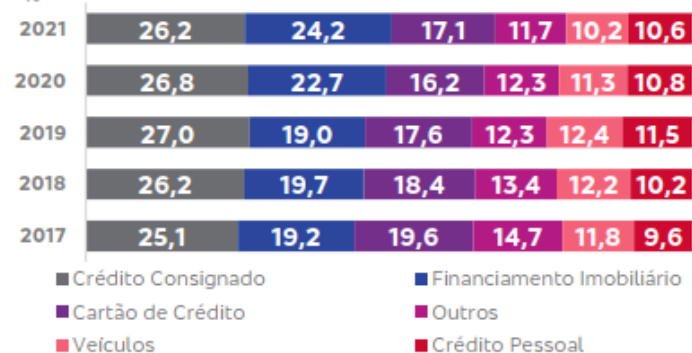
Mix da Carteira Total

%



Mix de Produtos Pessoas Físicas

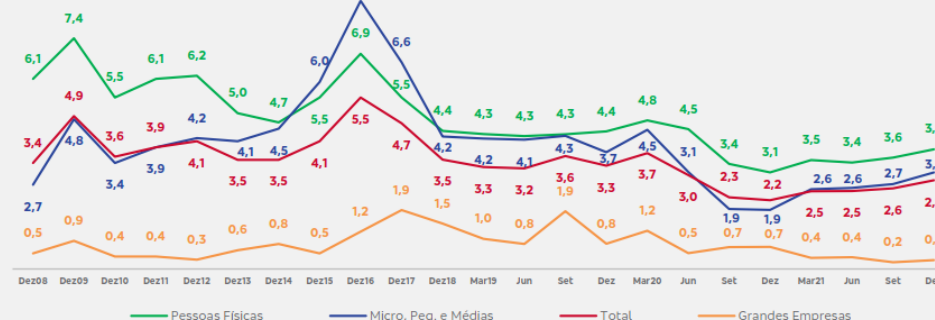
%



Índices de Inadimplência

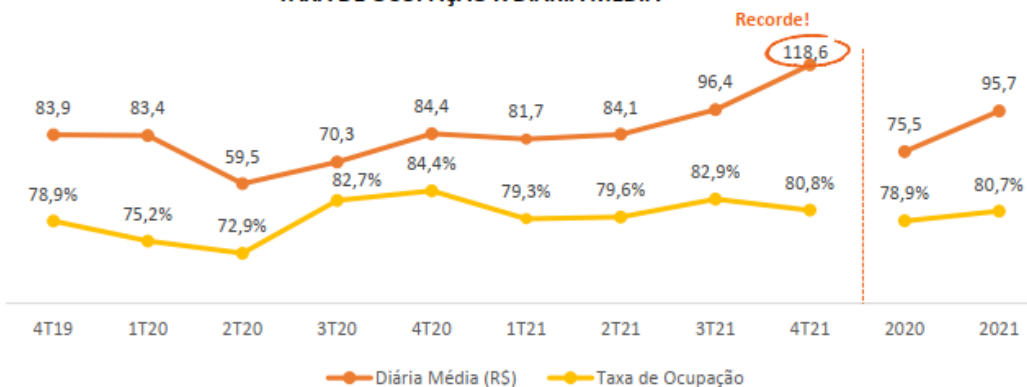
Acima de 90 dias

%

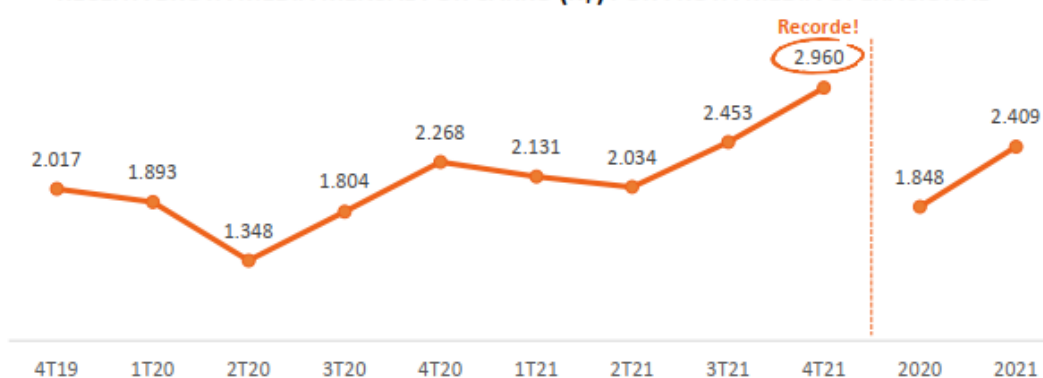


# ANÁLISE DESCRITIVA

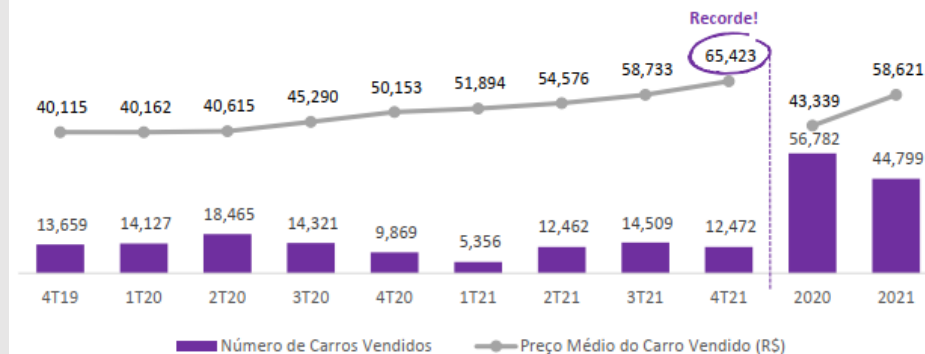
TAXA DE OCUPAÇÃO X DIÁRIA MÉDIA



RECEITA BRUTA MÉDIA MENSAL POR CARRO (R\$) POR FROTA MÉDIA OPERACIONAL



NÚMERO DE CARROS VENDIDOS X PREÇO MÉDIO DO CARRO VENDIDO (R\$)



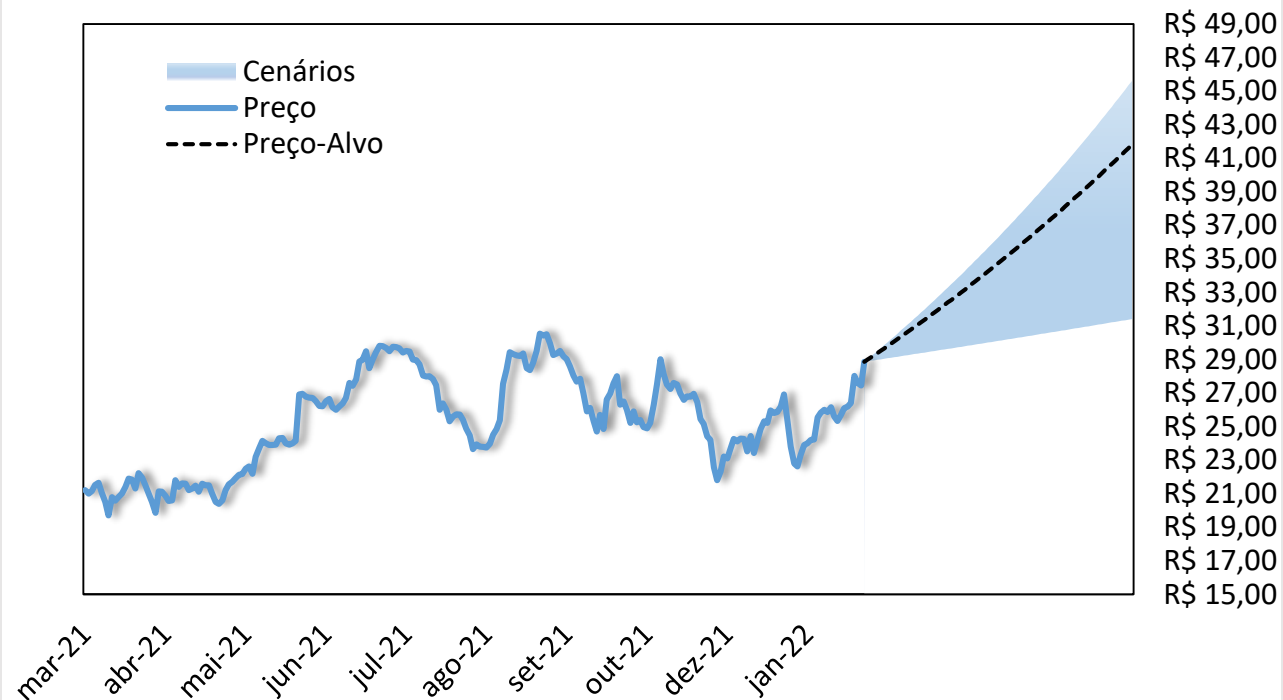


# ANÁLISE PREDITIVA

## Guidance 2022

	Guidance 2022
Crescimento da Carteira de Crédito Expandida <sup>1</sup>	12% a 16%
Crescimento da Carteira de Crédito Expandida do Segmento Middle	40% a 50%
Crescimento de Despesas (Pessoal, Outras Administrativas e PLR)	30% a 40%
Índice de Eficiência (Despesas/Receitas)	36% a 39%

## Expectativas de Cenários



# ANÁLISE PREDITIVA

## Pesquisa FEBRABAN de Economia Bancária e Expectativas - Dezembro de 2021

Legenda: em relação às projeções da pesquisa anterior

revisão positiva

revisão negativa

estabilidade

### Projeções para Saldo de Crédito e Inadimplência (média entre as instituições)

	Efetivos		Pesquisa nov/21		Pesquisa dez/21	
	2019	2020	2021	2022	2021	2022
Carteira Total (var. % nominal, fim de período, total do SFN)	6,5	15,7	12,7	7,3	13,9	6,7
Recursos Direcionados (var. % nominal, fim de período, total do SFN)	-2,4	15,9	8,1	4,1	10,0	4,1
Recursos Livres (var. % nominal, fim de período, total do SFN)	14,0	15,5	14,8	8,0	16,3	8,2
Pessoas Jurídicas - Livre (var. % fim de período)	11,0	21,3	12,7	6,4	13,9	7,4
Pessoas Físicas - Livre (var. % fim de período)	16,5	10,8	16,9	9,6	18,8	9,5
Crédito Pessoal (var. % nominal, inclui consignado)	14,7	13,6	17,6	10,5	18,5	10,2
Aquisição de Veículos (var. % nominal, inclui leasing)	19,6	8,2	11,0	9,2	10,9	9,7
Taxa de Inadimplência - % da Carteira Livre (fim de período, acima de 90 dias)	3,7	2,9	3,4	3,7	3,2	3,8

# ANÁLISE PRESCRITIVA

FIAP

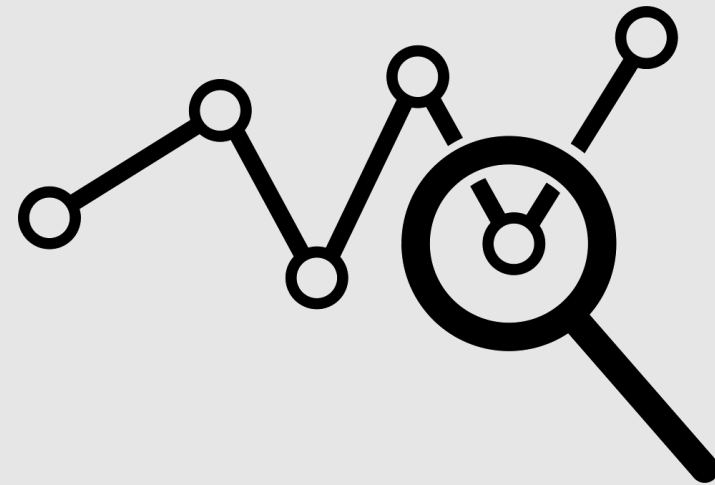


amazon

Cientes que visualizaram este item também visualizaram  
Clientes que compraram este item também compraram  
**Produtos que podem lhe interessar** [Ver mais](#)

# ANÁLISE DE DADOS

- **Análise aplicada a diferentes domínios;**
- **Engenheiro, Analista ou Cientista de Dados?**
- **Ferramentas e medidas para análise de dados:**
  - ☐ Distribuição e Dispersão;
  - ☐ Tendências;
  - ☐ Correlação e Causalidade.



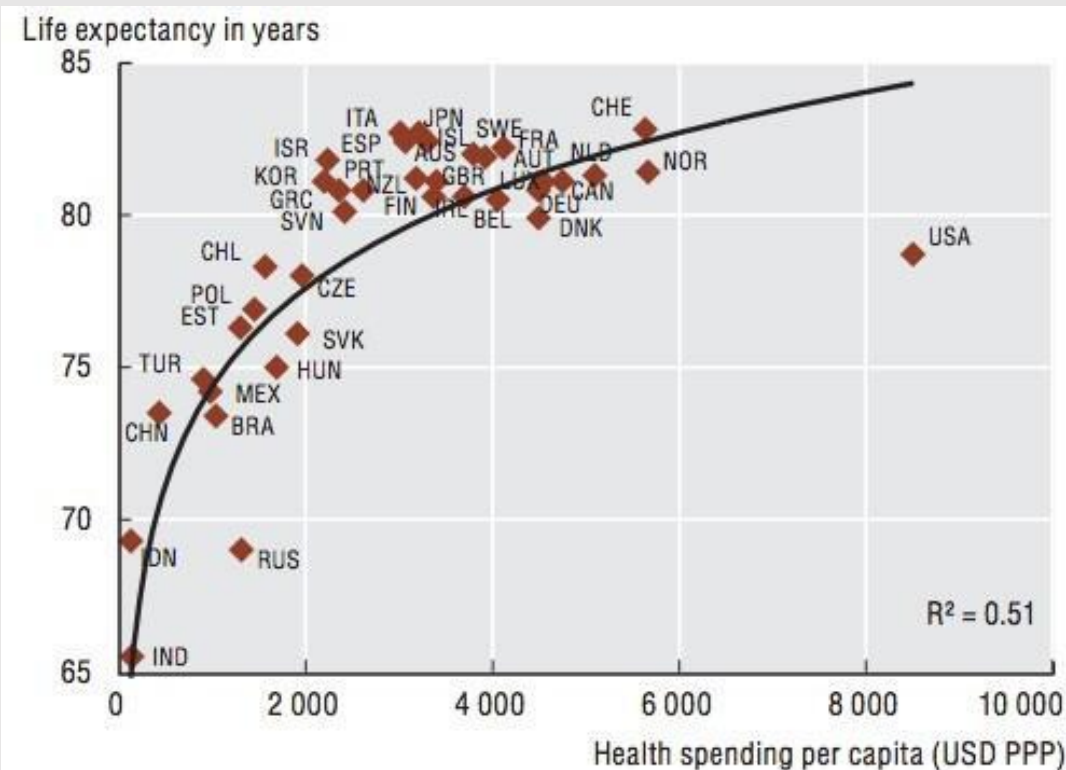
# ANÁLISE DE DADOS

- **Distribuição e Dispersão:**

- ☐ Média
- ☐ Mediana
- ☐ Desvio-Padrão
- ☐ Outliers

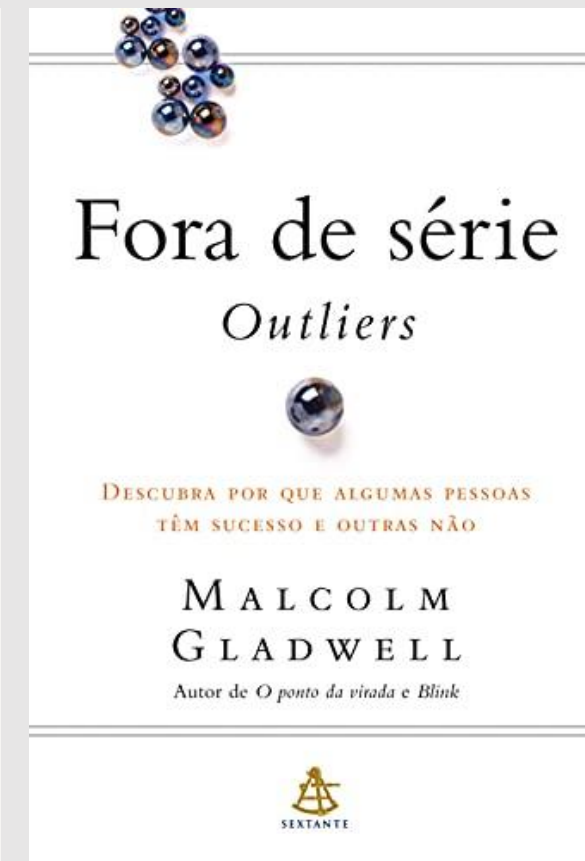
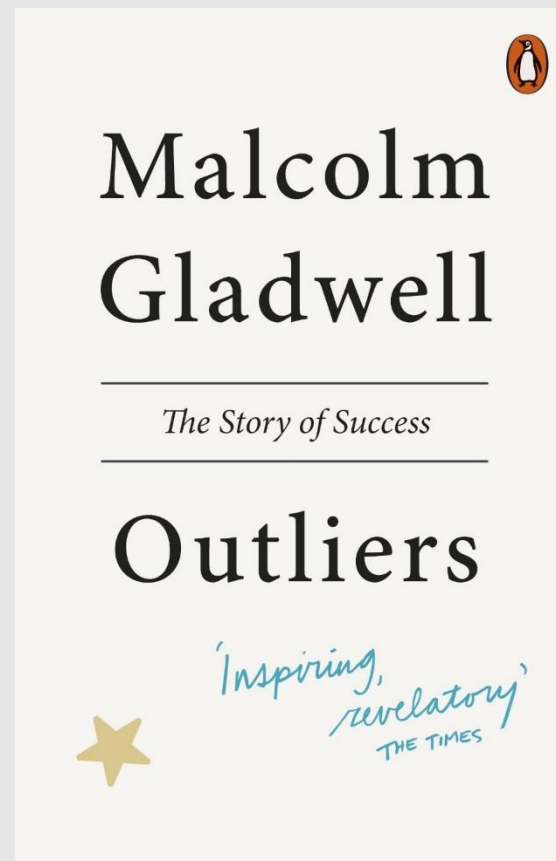
$$\bar{x} = \frac{\sum x_i}{n}$$

$$\sigma = \sqrt{\frac{\sum_{i=1}^n (x_i - \bar{x})^2}{n - 1}}$$



# ANÁLISE DE DADOS

- O que torna algumas pessoas capazes de atingir um sucesso tão extraordinário e peculiar a ponto de serem chamadas de "fora de série"?
- Costumamos acreditar que trajetórias excepcionais, como a dos gênios que revolucionam o mundo dos negócios, das artes, das ciências e dos esportes, devem-se unicamente ao talento. Mas neste livro você verá que o universo das personalidades brilhantes esconde uma lógica muito mais fascinante e complexa do que aparenta.





# ANÁLISE DE DADOS

- **Correlação e Causalidade:**

- ☐ Correlação
- ☐ Causalidade
- ☐ Correlação Espúria

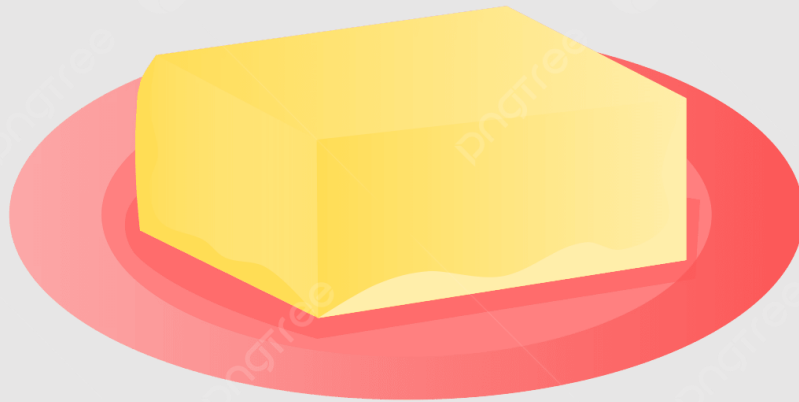
$$COV_{A,B} = \rho_{A,B} \sigma_A \sigma_B$$

$$COV_{A,B} = \frac{\sum (x_A - \bar{x}_A)(x_B - \bar{x}_B)}{n - 1}$$



# CORRELAÇÃO E CAUSALIDADE

- **Correlação Espúria 1:** Há uma correlação 0,99 entre a redução no consumo de margarina por pessoa e a diminuição de divórcios para cada mil pessoas no Estado do Maine (EUA).



# CORRELAÇÃO E CAUSALIDADE

- **Correlação Espúria 2:** Quanto menos se produz e vende mel nos mercados, mais jovens são presos por porte de maconha nos EUA (correlação de -0,93).



# CORRELAÇÃO E CAUSALIDADE

- **Correlação Espúria 3:** Você sabe qual é a correlação entre o número de norte-americanos que se afogam (por ano) ao caírem de barco quando pescavam e a taxa de casamentos em Kentucky (EUA)? É 0,95!



# CORRELAÇÃO E CAUSALIDADE

- **Correlação Espúria 4:** A correlação entre o número de filmes feitos por Nicolas Cage em um ano e a quantidade de gente que morre em acidentes de helicóptero nos EUA foi calculada e indicou uma correlação de -0,82.



# CORRELAÇÃO E CAUSALIDADE

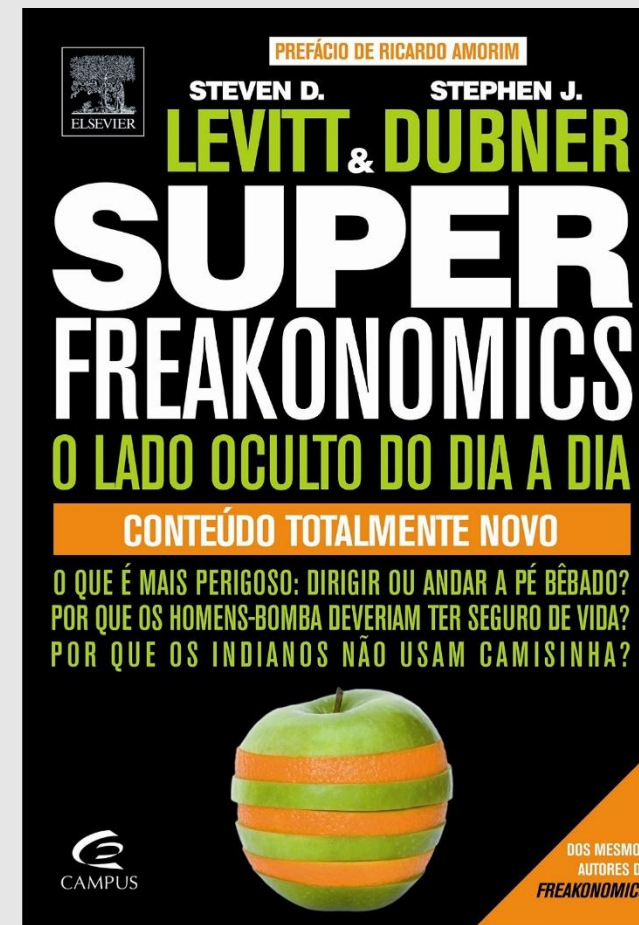
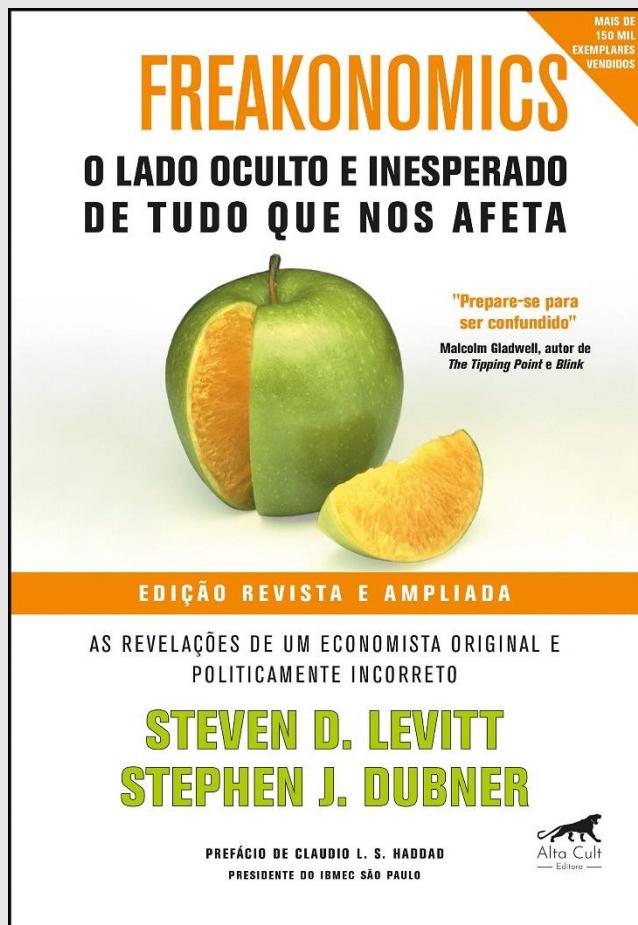
- **Correlação Espúria 5:** Quanto mais velha a Miss América em um determinado ano, mais gente é assassinada nos EUA, pois a correlação entre esses dados é 0,87!





# ANÁLISE DE DADOS

- O que é mais perigoso, uma arma ou uma piscina?
- O que os professores e os lutadores de sumô têm em comum?
- Por que os traficantes de drogas moram com as mães?
- Qual a importância real dos pais?
- Que tipo de impacto teve o caso Roe x Wade, que levou à legalização do aborto nos Estados Unidos, sobre a criminalidade?



# DADOS

- Informações factuais (como medições ou estatísticas) usadas como base para raciocínio, discussão ou cálculo;
- **Formas de organização:**
  - ☐ Tabela de Dados
  - ☐ Conjunto de Dados
  - ☐ Programas/Aplicativos
  - ☐ Identificadores (ID)

# ESTRUTURA DE DADOS

- **Estrutura de Dados:**
  - ❑ Cross-sectional / Cortes transversais
  - ❑ Séries temporais
  - ❑ Painel de dados / Dados longitudinais:
    - ❖ Balanceados
    - ❖ Não Balanceados
    - ❖ Agregação e nível das informações

# QUALIDADE DOS DADOS

- Poucas coisas têm consequências mais devastadoras para a reputação de um analista de dados do que alguém na plateia apontando problemas sérios de medição que o analista não considerou.
- **Validade:** O conteúdo da variável deve ser o mais próximo possível do que ela espera mensurar;
- **Confiabilidade:** A forma de mensuração da variável deve ser estável, levando ao mesmo valor caso seja medida da mesma forma novamente;
- **Comparável:** A variável deve ser medida da mesma maneira para todas as observações.

# COLETA DE DADOS

- Bancos de Dados Abertos;
- Bancos de Dados Fechados;
- Web Scrapping;
- Dados Internos.




# COLETA DE DADOS

- Coletam dados através de entrevistas com pessoas, fazendo perguntas e registrando as suas respostas. Normalmente, as respostas são curtas e facilmente transformadas em variáveis quantitativas ou qualitativas.
- Respondentes;
- Questionário;
- **Tipos:**
  - ☐ Questionários administrados;
  - ☐ Entrevistas.
  - ☐ <https://m.leiaja.com/noticias/2015/01/15/responda-10-perguntas-e-avalie-sua-educacao-financeira/>



# FRED

FIAP


**ECONOMIC RESEARCH**  
FEDERAL RESERVE BANK of ST. LOUIS

Economic Research Resources ▼Switch Products ▼My Account ▼

Release CalendarFRED Tools ▼FRED NewsFRED BlogAbout FRED ▼


Welcome to FRED, your trusted source for economic data since 1991.

# FRED®



Download, graph, and track [816,000 US and international time series](#) from [108 sources](#).

**ECONOMIC DATA | ST. LOUIS FED**

Search FRED data...

Trending Search Terms:

cpi10 yeargdpm2

real gdpinflationyield curve

interest ratelibor10 year treasury

Browse Data By:


[Category](#) [Release](#) [Source](#) [Tag](#)

[Release Calendar](#)

**FRED News**

FRED Adds Overnight AMERIBOR Unsecured Interest Rates

FRED Adds Selected Business and Consumer Loans Data




A picture's worth a thousand words.



Use FRED Graph to **display your**

# FRED

FIAP


 ECONOMIC RESEARCH

Economic Research Resources ▼ Switch Products ▼ My Account ▼

ECONOMIC DATA | ST. LOUIS FED

Your trusted data source  
since 1991.

Search FRED ... 

Release Calendar FRED Tools ▼ FRED News FRED Blog About FRED ▼

## API Keys


+ Request API Key

### About API Keys

The API (Application Programming Interface) is a web service that allows users to write programs and build applications that retrieve economic data from the [FRED®](#), [ALFRED®](#) and [GeoFRED®](#) websites.

- [Terms of use](#)
- API keys are needed to develop programs & applications using FRED data.
- Developers should request a distinct API key for each application they build.
- All users of an application shall use their own API key.
- Data retrieval documentation for [FRED](#) and [GeoFRED](#)

**You do not have any API keys.**

**BANCO CENTRAL  
DO BRASIL**


**SGS - Sistema Gerenciador de Séries Temporais - v2.1**  
Módulo público

Usuário público  
29/03/2022 15:22  
English

Consultar | Minhas listas de séries | Configurações | Ajuda |

Início → Consultar séries → Localizar séries

[SGSFW0101] ?

 **Pesquisa**

Selecione a periodicidade

Todas ▼

Selecione uma opção

Por tema →

Por código →

Por fonte →

Abecip e BCB-Depec ▼

Não há lista(s).  
Para criar clique [aqui](#)

Séries mais pesquisadas →

Séries desativadas →

Pesquisa textual  
(nome da série) →

Pesquisa Avançada →

**Localizar séries - Selecione um dos temas abaixo**

**Atividade econômica**  
Setor real, Mercado de trabalho, Preços

**Economia regional**  
Nível de atividade, mercado de trabalho, preços, setor externo, finanças públicas e crédito por estados e regiões

**Expectativas do mercado**  
Taxa Over-Selic, Taxa de Câmbio, Investimento Estrangeiro Direto, Balança Comercial, Saldo das Transações em Conta Corrente, Preços, Resultados Fiscais, Produção Industrial e PIB

**Inclusão financeira**  
Indicadores de inclusão financeira.

**Indicadores monetários**  
Política monetária, Agregados monetários, Contas analíticas do sistema financeiro

**Mercosul**  
Indicadores de atividade econômica, monetário, fiscal e setor externo dos países do Mercosul

**Setor externo**  
Balanço de pagamentos, Balança comercial, Reservas internacionais, Dívida externa, Taxa de rolagem e Taxas de câmbio

**Tabelas especiais**

**Economia internacional**  
Indicadores da atividade econômica, financeiros e do setor externo de países selecionados

**Estabilidade Financeira**  
Indicadores de estabilidade e solidez do Sistema Financeiro Nacional






**Finanças públicas**  
Dívida líquida e necessidades de financiamento do setor público, Dívida mobiliária, Execução financeira do Tesouro Nacional, Despesa com pessoal da União, Receita dos estados e municípios, etc

**Indicadores de crédito**  
Operações de crédito do sistema financeiro (volume segundo a atividade econômica, segundo a qualidade, segundo origem dos recursos, taxas de juros)

**Mercados financeiros e de capitais**  
Aplicações financeiras, Indicadores do mercado financeiro, Indicadores do mercado de capitais, etc

**Multiplicadores de unificação monetária**  
Conversores de unidade monetária corrente para reais correntes

**Sistema Financeiro Nacional**  
Organização e funcionamento do Sistema Financeiro Nacional



Marcar todas Desmarcar todas Acrescentar séries Consultar séries Incluir em nova lista



# TENDÊNCIAS

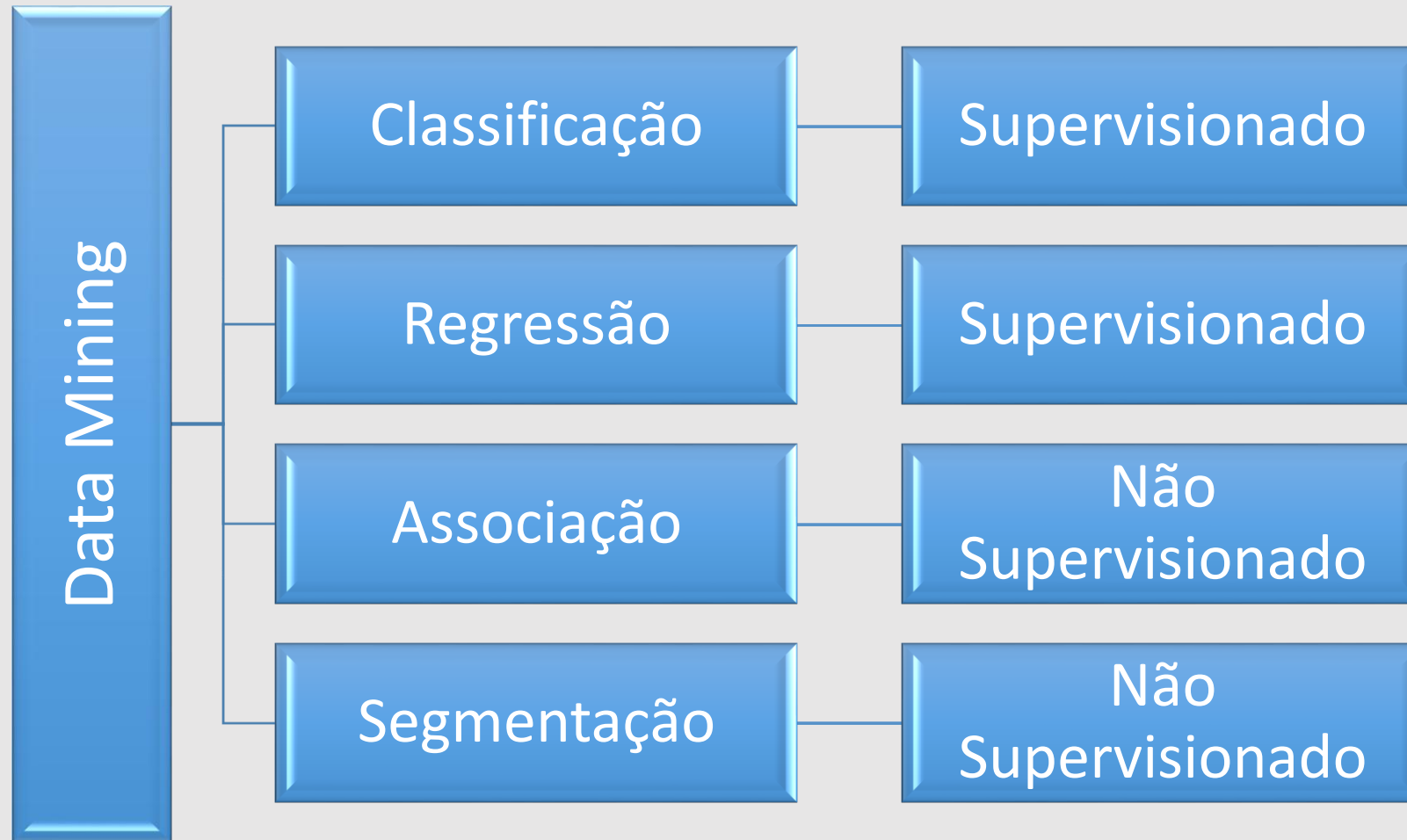
- Data Mining;
- Text Mining / Text Analytics;
- Processamento de Imagem (Image Processing);
- Processamento de Linguagem Natural (Natural Language Processing – NLP);
- Internet of Things (IoT) / Web of Things (WoT);
- Cloud Computing;
- Análise Geoespacial (Location-Based Analytics).

# DATA MINING

- **Data Mining:** Processo a partir do qual são descobertos padrões nos dados anteriormente desconhecidos.



# DATA MINING





# DATA MINING

- **Classificação:** Aprende os padrões após a observação dos dados e busca categorizar novas observações nos seus respectivos grupos;
- **Uso:**
  - ❑ Transações Bancárias Fraudulentas;
  - ❑ Probabilidade de Inadimplência;
  - ❑ Classificação de Enfermidades;
- Dados de Treino e Teste;
- Matriz de Confusão.

		Observado	
		Positivo	Negativo
Previsão	Positivo	Verdadeiro Positivo	Falso Positivo
	Negativo	Falso Negativo	Verdadeiro Negativo

# DATA MINING

- **Regressão:** Modela o comportamento de uma variável (dependente) com base no comportamento de uma ou de um conjunto de outras variáveis (independentes);
- **Teste de hipóteses e Previsão:**
  - Previsão de séries temporais;
  - Análise de políticas;
  - Modelagem de risco.



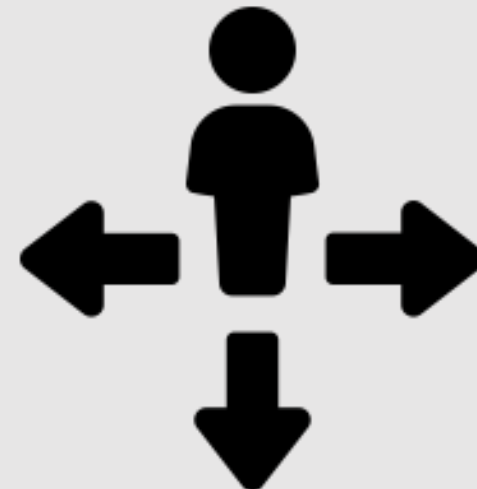
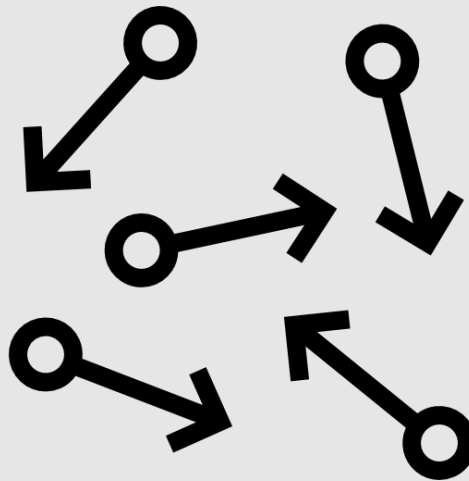
# TEXT MINING/TEXT ANALYTICS

- **Text Analytics:** Inclui a busca de informações, sua coleta e a mineração de dados;
- **Text Mining:** Descoberta de conhecimento dos dados de texto disponíveis;
- Processamento de Linguagem Natural;
- **Aplicações:**
  - ☐ Marketing;
  - ☐ Segurança;
  - ☐ Medicina;
  - ☐ Ciência.



# PROCESSAMENTO DE IMAGEM

- Identificação de Padrões e Análise de Movimentos;
- Construção de Cenários 3D;
- Identificação de Objetos ou Pessoas;
- Detecção ou Ocorrência de Eventos.



# PROCESSAMENTO DE IMAGEM

FIAP

 Buscar

tech**tudo** • ELETRÔNICOS

**GALILEU**

REVISTA DIGITAL NOTÍCIAS MEIO AMBIENTE CULTURA SAÚDE

## Algoritmo de inteligência artificial consegue identificar rostos cobertos

Através de um algoritmo de aprendizagem, a tecnologia reconhece pessoas mesmo com disfarce

Por **Thiago Rocha**, para o TechTudo

14/09/2017 07h01 · Atualizado há 4 anos



TECNOLOGIA

## Reconhecimento facial: prós e contras da tecnologia que veio para ficar

Mesmo com pouco debate e regulamentação, tecnologias de reconhecimento facial já são usadas em massa no mundo inteiro — e se intensificam em meio à pandemia do novo coronavírus

 6 min de leitura

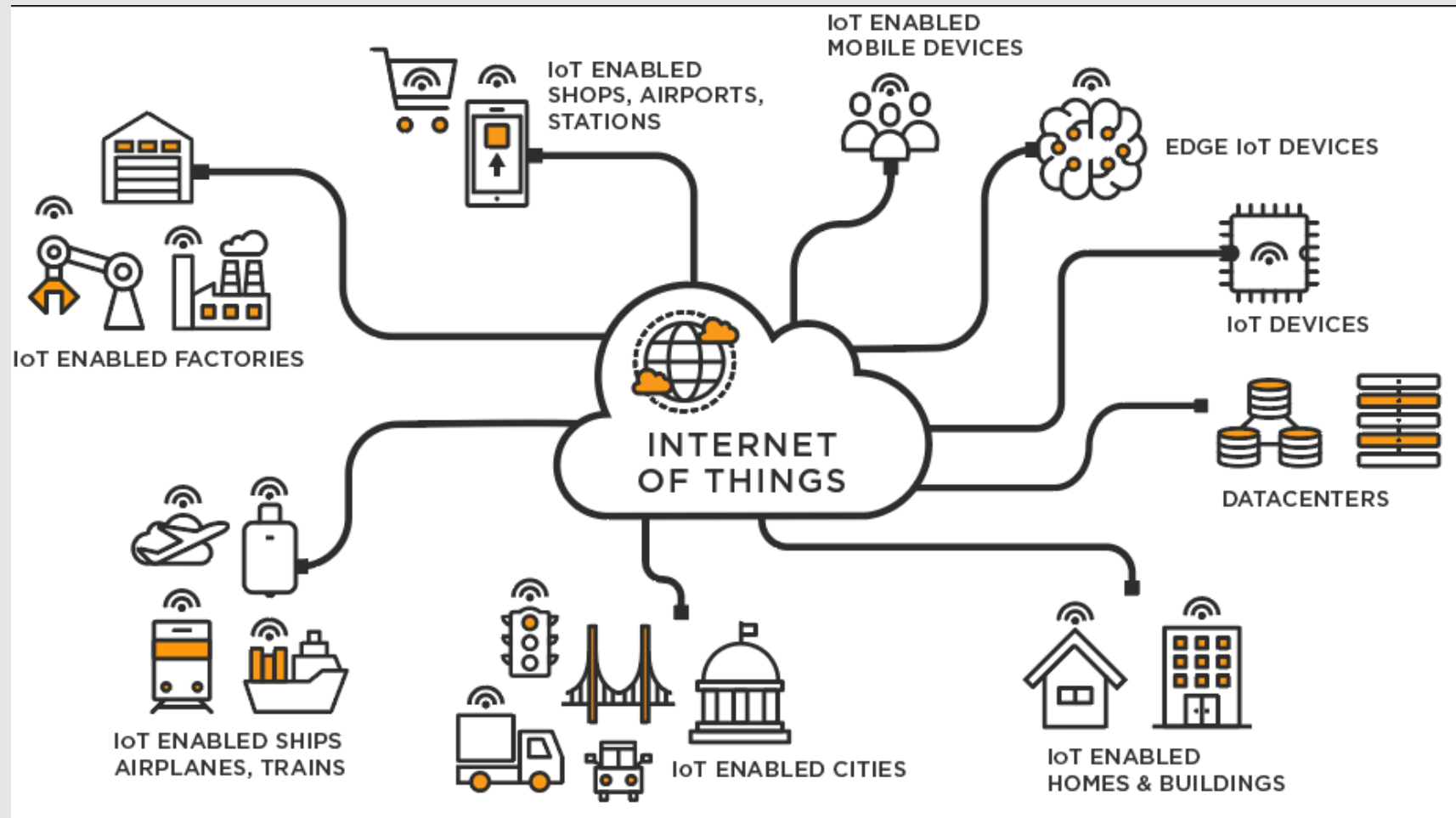
**MARILIA MARASCIULO**  
22 JUN 2020 - 10H05 | ATUALIZADO EM 21 MAR 2022 - 18H16

# PROCESSAMENTO DE LINGUAGEM NATURAL

- **PLN:** Inteligência Artificial aplicada à Linguística Computacional. Busca entender a linguagem natural humana através da conversão de trechos de comunicação em representações mais formais, mais fáceis de serem entendidas e manipuladas por programas de computador;
- **Principais Usos:**
  - ☐ Respostas Automáticas;
  - ☐ Reconhecimento de Discurso;
  - ☐ Tradução;
  - ☐ Análise de Sentimento.



# INTERNET OF THINGS (IoT)





# INTERNET OF THINGS (IoT)

- **IoT:** Conexão do mundo físico com a internet;
- **Exemplos:**
  - ☐ Veículos autônomos;
  - ☐ Casa Inteligente;
  - ☐ Dispositivos fitness;
  - ☐ Aparelhos autônomos;
  - ☐ Realidade Virtual.

## FIAP

FIAP

# CLOUD COMPUTING

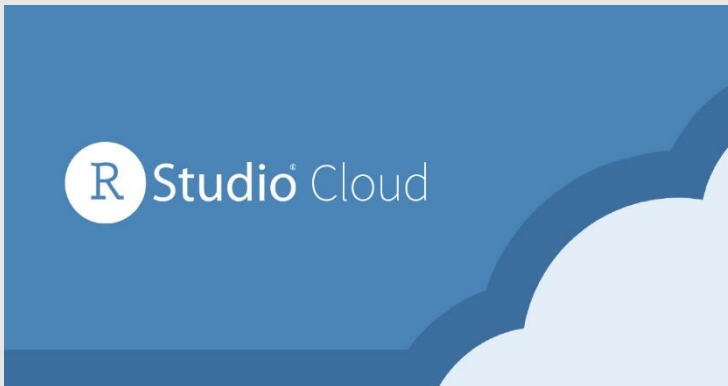
- **Cloud Computing:** Entrega de recursos em TI sob demanda por meio da Internet com definição de preço de pagamento conforme o uso;
- Em vez de comprar, ter e manter datacenters e servidores físicos, você pode acessar serviços de tecnologia, como capacidade computacional, armazenamento e bancos de dados, conforme a necessidade, usando um provedor de nuvem.

# CLOUD COMPUTING

- **Data as a Service (DaaS):** Dados podem ser armazenados em um servidor físico ou em um servidor que está dentro de um ambiente de nuvem;
- **Software as a Service (SaaS):** Consumidores podem utilizar aplicações e softwares que são executados em computadores distantes com arquitetura em nuvem;
- **Platform as a Service (PaaS):** Empresas podem armazenar seus softwares e aplicações na nuvem e permitir o uso pelos seus clientes;
- **Infrastructure as a Service (IaaS):** Recursos de infraestrutura, como redes, armazenamento, servidores e outros recursos computacionais são fornecidos para empresas.

# CLOUD COMPUTING

- **Analytics as a Service (DaaS):** É a combinação de ferramentas de análise de dados e de computação na nuvem. Oferece soluções para organizar, analisar e visualizar dados;



# ANÁLISE GEOESPACIAL

- **Análise Geoespacial:** Adiciona a percepção de tempo e de espaço para os tipos tradicionais de dados e busca construir formas de visualização;
- <http://grindgis.com/blog/gis-applications-uses>
- **Principais Usos:**
  - ☐ Agricultura;
  - ☐ Análises Criminais;
  - ☐ Propagação de Doenças.

