




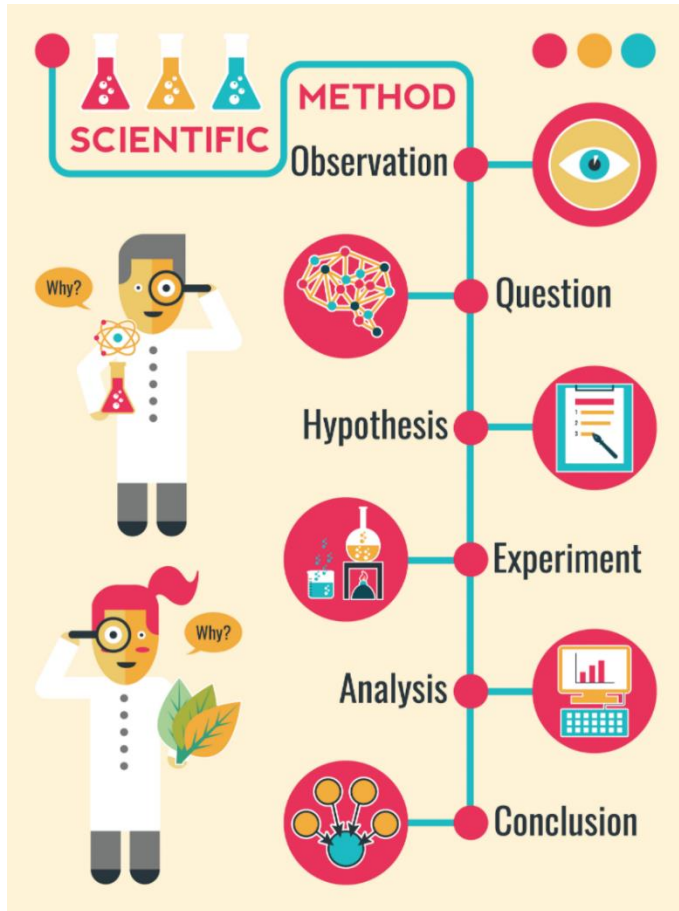
Analytics para área da saúde



O que é Data Science?

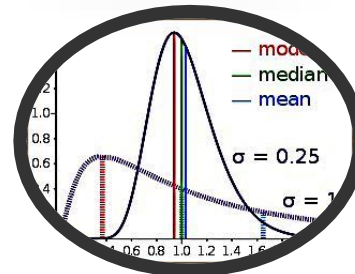
O que é Data Science?

É a utilização de metodologia científica aliada à ferramentas matemáticas, estatísticas e computacionais com o objetivo de solucionar problemas de negócios.



$$\begin{aligned} \frac{\partial \mathcal{L}}{\partial x_1} - \sum_{j=1}^n \frac{\partial}{\partial x_j} \left(\frac{\partial \mathcal{L}}{\partial f_{1,j}} \right) &= 0_1 \\ \frac{\partial \mathcal{L}}{\partial x_2} - \sum_{j=1}^n \frac{\partial}{\partial x_j} \left(\frac{\partial \mathcal{L}}{\partial f_{2,j}} \right) &= 0_2 \\ &\vdots \\ \frac{\partial \mathcal{L}}{\partial x_n} - \sum_{j=1}^n \frac{\partial}{\partial x_j} \left(\frac{\partial \mathcal{L}}{\partial f_{n,j}} \right) &= 0_n \end{aligned}$$

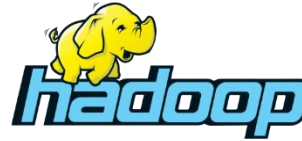
Matemática



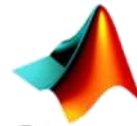
Estatística



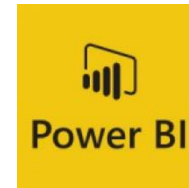
Computação



WolframAlpha



MATLAB



O Primeiro Cientista de Dados da História?

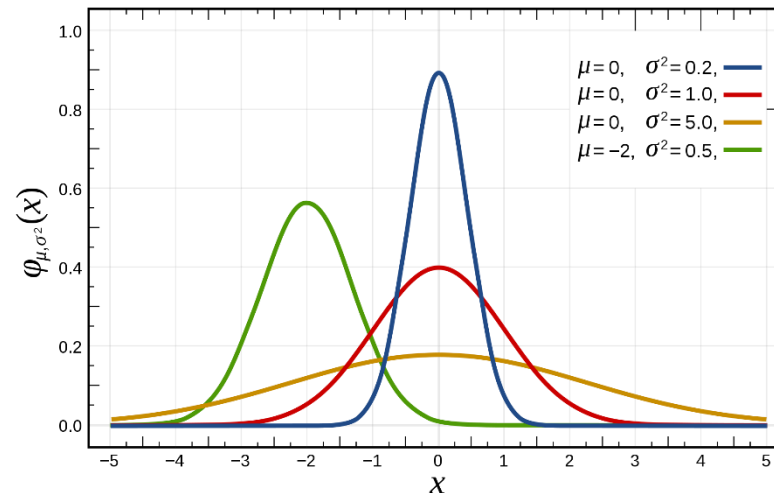
Carl Friedrich Gauss



Nascido na Alemanha em 1777 (faleceu em 1855), conhecido como o príncipe da matemática devido a seu precoce talento, desde os 3 anos quando aprendeu álgebra sozinho

Grandes Contribuições

- * Matemática (Álgebra, Geometria, cálculo,...)
- * Física (Mecânica, Gravitação, Eletromagnetismo,...)
- * Astronomia
- * Estatística



Distribuição Gaussiana
(Estatística)

Análise de Dados

Gauss foi o criador da regressão estatística de dados (método dos mínimos quadrados). Ele utilizou esta ferramenta matemática para criar um modelo de predição da posição de corpos celestes, e auxiliar navegadores em suas viagens, além do estudo das geodésicas e geometria. Mais tarde, este método se mostrou útil no estudo de correlações entre variáveis. Esta invenção de Gauss foi o embrião do que é hoje conhecido por Estatística. Tendo sido um primeiro modelo de otimização e aprendizado, que são hoje tópicos centrais em Ciência de Dados.

Aplicações de Data Science



Retail Analytics

Insights sobre consumidores, logística, otimização de preços, planejamento de estoque, ofertas customizadas, etc.



Risco de Crédito

Score de crédito, ratings, grupos homogêneos, clusterização de perfis de clientes, modelos de collection, behavior, predição de calote, fraude, etc.



Financial Analytics

Modelos preditivos de valores de Ações, opções, índices, estruturados, commodities, fiduciários, colateral, custos, etc.



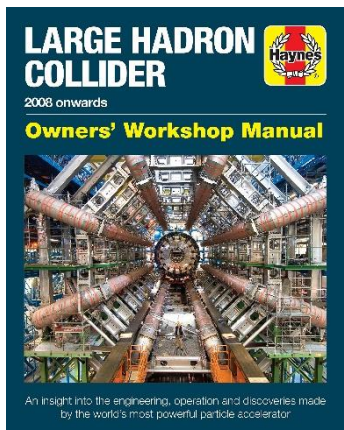
Banking Analytics

CRM, relatórios, P&L, insights e perfis de clientes, planejamento de resultados e metas, plan. numerário, agências, clientes e carteiras, ofertas contextualizadas.



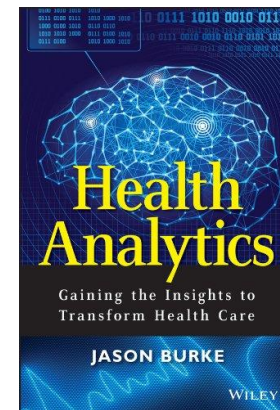
Big Data Analytics

Análise de big data em projetos científicos, tomadas de decisão, descobertas e aprendizado de máquina, validação de resultados, organização, versionamento e debugging.



Health Analytics

Seguradoras, operadoras, saúde pública, diagnósticos, pesquisa acadêmica, melhoria em tratamentos, otimização de rede de atendimento.





Principais Habilidades e alguns papéis em Analytics

Principais Habilidades



Programming
Language

Conhecer a lógica de programação, os algoritmos, sintaxe da linguagem e os paradigmas de computação.
(desenvolvimento, debug, documentação e etc)



Knowledge and
Application

Conhecimento básico de Administração e Economia. Entendimento de processos e metodologias de análise. Raciocínio lógico e analítico e pensamento estratégico.
(Comunicação, Sagacidade, Oportunismo e etc)



Data
Modeling
Engineering

Conhecimento em estruturação de banco de dados. Entendimento de relação entre tabelas.
(Conexão com o banco de dados, Consultas nas tabelas, Otimização, Monitoramento e etc)



Data
Visualization

Prática com software de visualização. Maneiras gráficas de apresentar as informações. Criação em dashboard simples e atrativas. Boas práticas de visualização.
(Storytelling, Apresentação executiva e etc)



Software
Engineering

Conhecimento de aplicação e framework. Conhecimento de metodologias e ferramentas ágeis. Boas práticas de desenvolvimento e processos.
(Concepção, Especificação, Design Pattern, Documentação, Teste, debug e etc)



Data Intuition

Seleção de variáveis relevantes para o negócio, intuição sobre métricas, indicadores, números relevantes, variáveis, etc. Capacidade de validar e estimar números através de raciocínio indutivo.
(Experiência, conhecimento técnico e teórico, criatividade e etc)

Principais Habilidades



Digital Expertise

Entendimento relacionados em ambientes digitais (Mobile ou Web). Conhecimento de ferramentas (Google Ads, Tag Manager, SEO e outros).

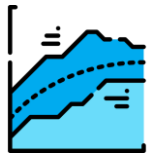
(Tagueamento, Google Analytics e etc)



Data Wrangling

Conhecimento de métodos de transformação e manipulação de dados. Métodos estatísticos de complementação (Missing, outlier e outros).

(Padronização, Normalização dos dados e etc)



Advanced Mathematics and Statistics

Conhecimento de disciplinas avançadas de matemática e estatística.

(Calculo, Algebra, Analise, Algoritmos e etc)



Modeling, Operations and Research

Métodos analíticos de soluções de problemas e tomadas de decisão na gestão empresarial. Utilização de matemática aplicada e computacional e análise estatística.

(Otimização, Econometria, Cadeias de Markov e etc.)



Communication

Comunicação verbal clara e objetiva. Sabe lidar com diferentes perfis.

(soft skills, retórica, concisão, objetividade)



Big Data

Conhecimento de programação distribuída. Conceitos relacionados a Big Data. Conhecimento de ferramentas de desenvolvimento do ecossistema.

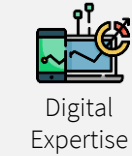
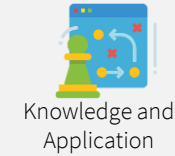
(5V, Spark, Hive, DataFactory e etc)

Alguns papéis em analytics



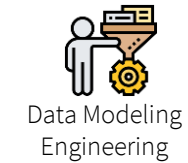
Web Analyst

Trabalha no desenvolvimento, análise, interpretação, mapeamento e implantação de tecnologias digitais como google e adobe analytics.



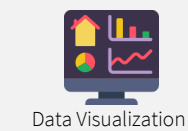
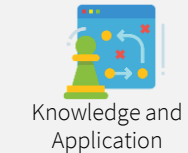
Engenheiro de Dados

Arquitetura e construção do pipeline de dados, ingestão de fontes internas e externas de dados, modelagem dos dados, processos de ETL e automação de processos, integração de serviços cloud com outras tecnologias e API.



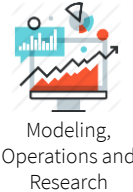
Analista de Negócios

Realiza acompanhamento executivo, conhecimento em análise e visualização de dados, apresentações executivas, alinhamentos, criação de métricas e indicadores de performance.



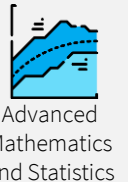
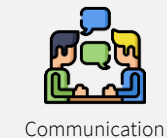
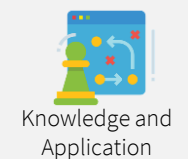
Analista de Dados

Análise de dados mais aprofundada, busca por insights e oportunidades de negócios através dos dados. Possuem aprofundamento técnico e específico moderado. Focado em análises descritivas e prescritivas.



Cientista de Dados

Apoia o engenheiro a modelar os dados, auxilia o analista de negócios na definição de indicadores, e realiza análises avançadas (modelos matemáticos e estatísticos). Auxilia o desenvolvedor na implantar modelos.





Processos em Analytics

Etapas no processo de análises

Valor ao
Negócio

Restrospectiva

Insight

Previsão

Como fazer
acontecer?

O que irá
acontecer?

Por que
aconteceu?

O que
aconteceu?

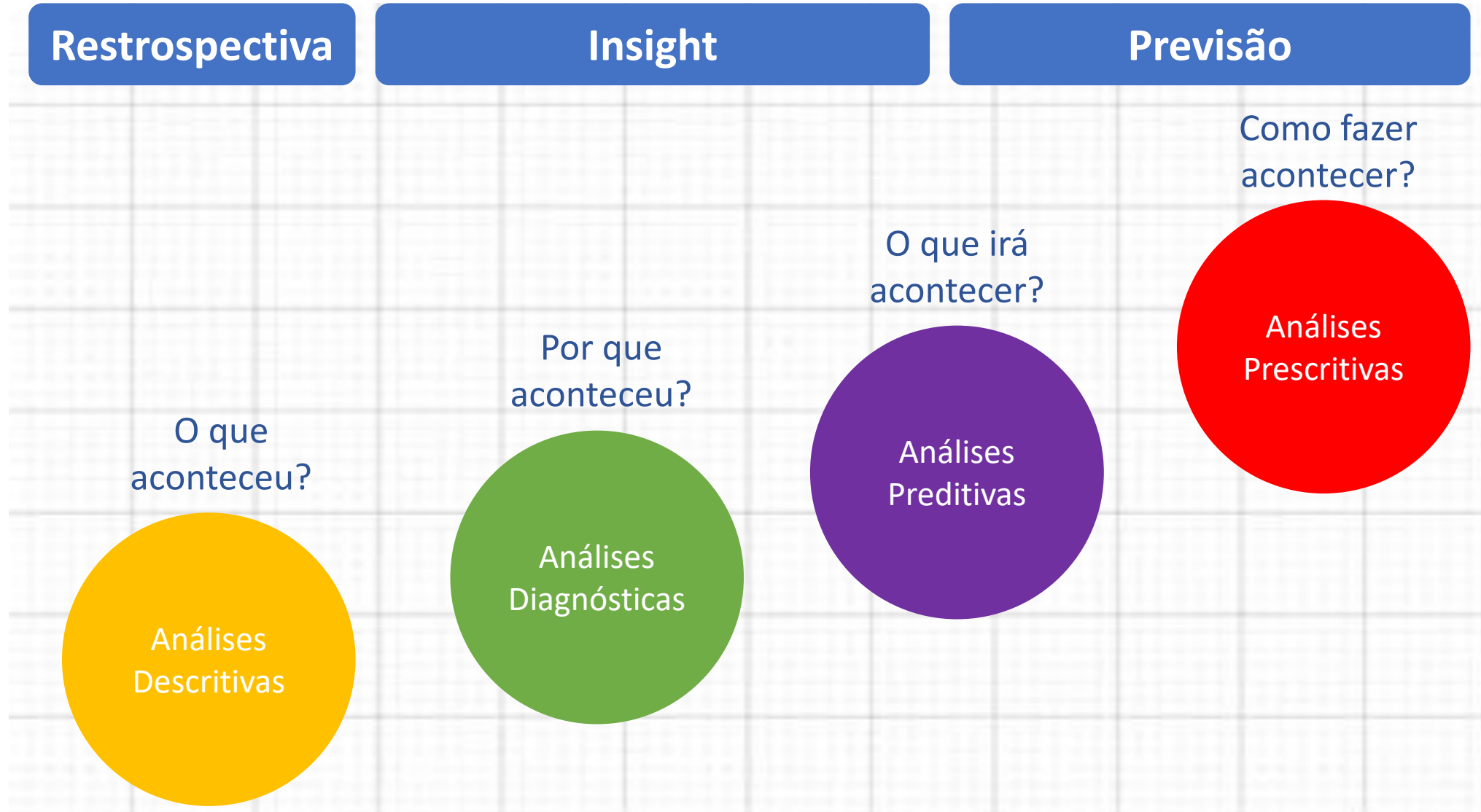
Análises
Prescritivas

Análises
Preditivas

Análises
Diagnósticas

Análises
Descritivas

Complexidade



Etapas no processo de análises

Análises
Descritivas

O que
aconteceu?

Investigação e mapeamento dos dados, sumarização de resultados, criação de painéis e apresentações executivas

Análises
Diagnósticas

Por que
aconteceu?

Investigação mais aprofundada dos dados com objetivo de criar insights e diagnósticos sobre os resultados ocorridos

Análises
Preditivas

O que irá
acontecer?

Análise histórica dos dados para identificação de padrões e criar modelos preditivos sobre resultados futuros e automação de tomadas de decisão

Análises
Prescritivas

Como fazer
acontecer?

Criação de modelos e métodos analíticos avançados para fazer recomendações específicas sobre estratégias de negócio

Etapas no processo de análises

Análises Descritivas

O que
aconteceu?

-
- * Estatística descritiva
 - * Data Mining
 - * Agregação de dados
 - * Cruzamento de bases
 - * Dashboards

Análises Diagnósticas

Por que
aconteceu?

-
- * Estatística
 - * Probabilidade
 - * Séries temporais
 - * Análises de padrões

Análises Preditivas

O que irá
acontecer?

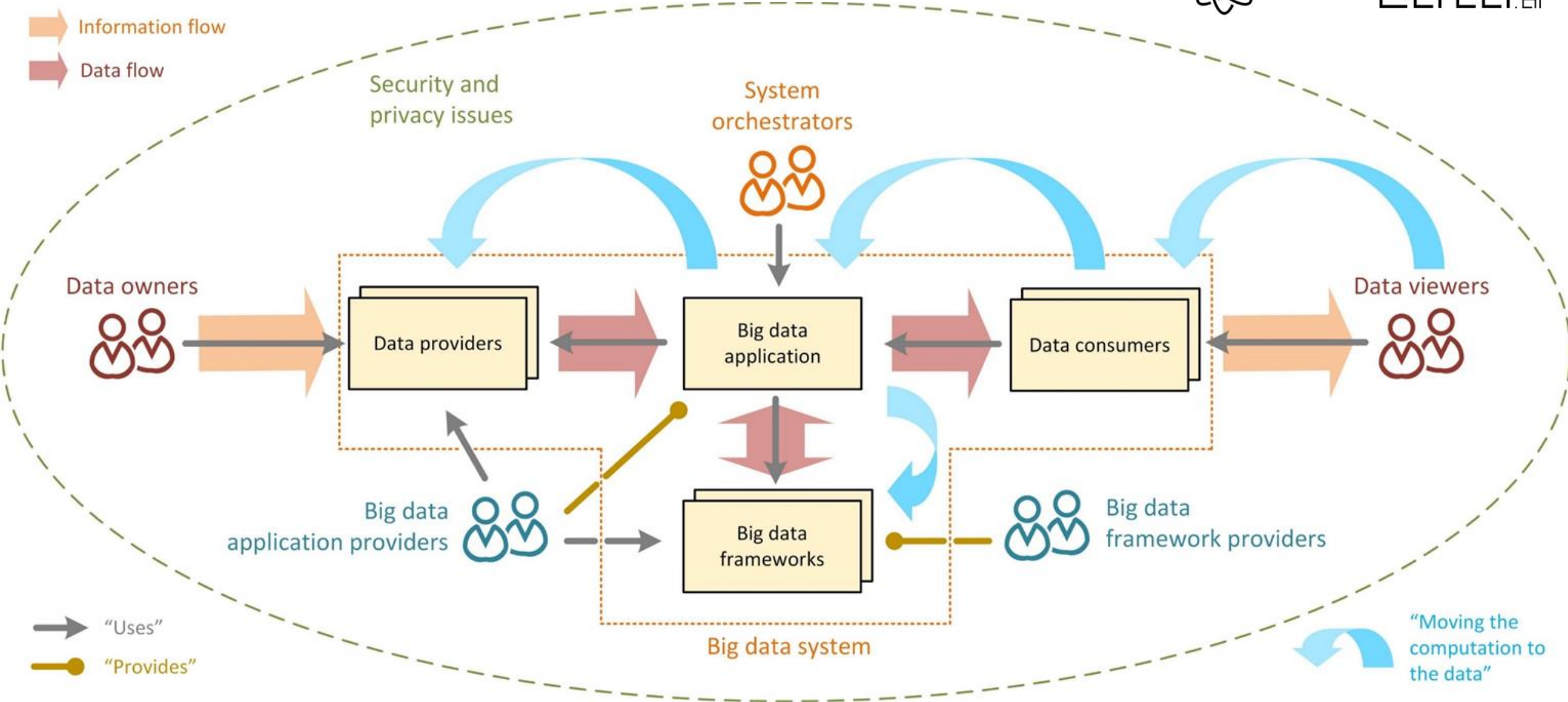
-
- * Modelos estatísticos
 - * Modelos matemáticos
 - * Simulações
 - * Forecasting
 - * Machine Learning

Análises Prescritivas

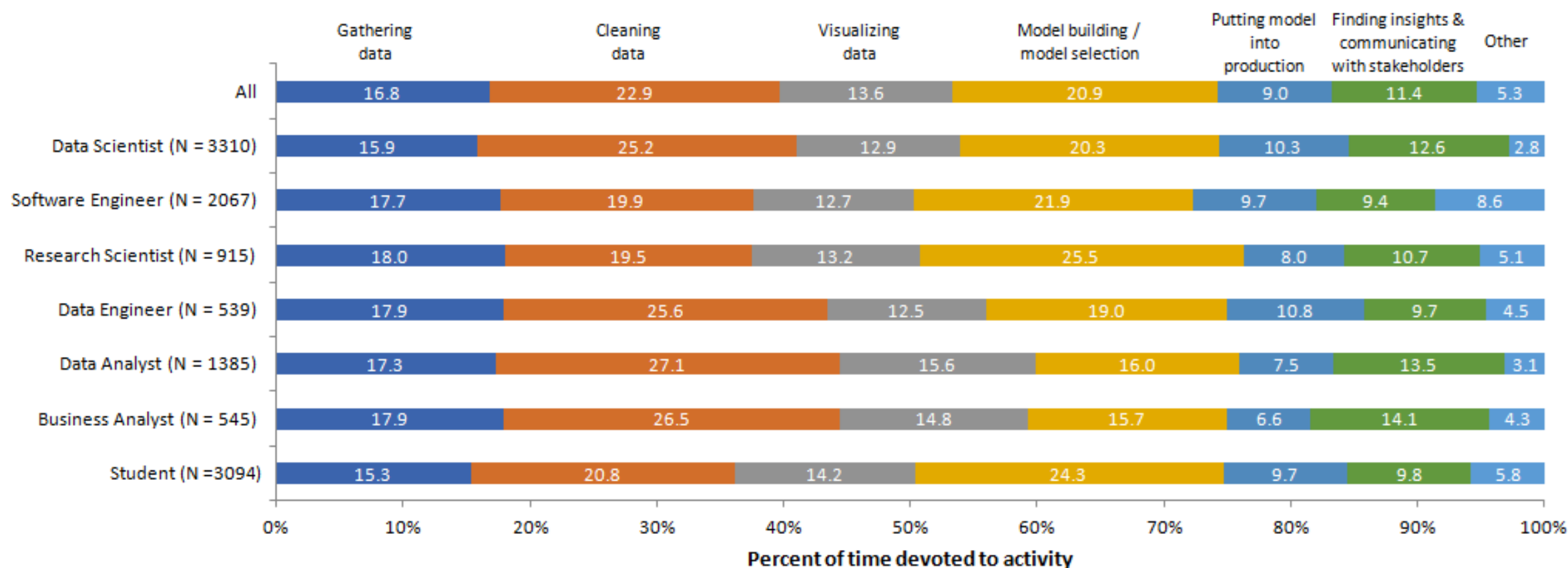
Como fazer
acontecer?

-
- * Otimização
 - * Machine Learning
 - * Deep Learning
 - * Redes Neurais

Data Management



During a typical data science project at work or school, approximately what proportion of your time is devoted to the following?



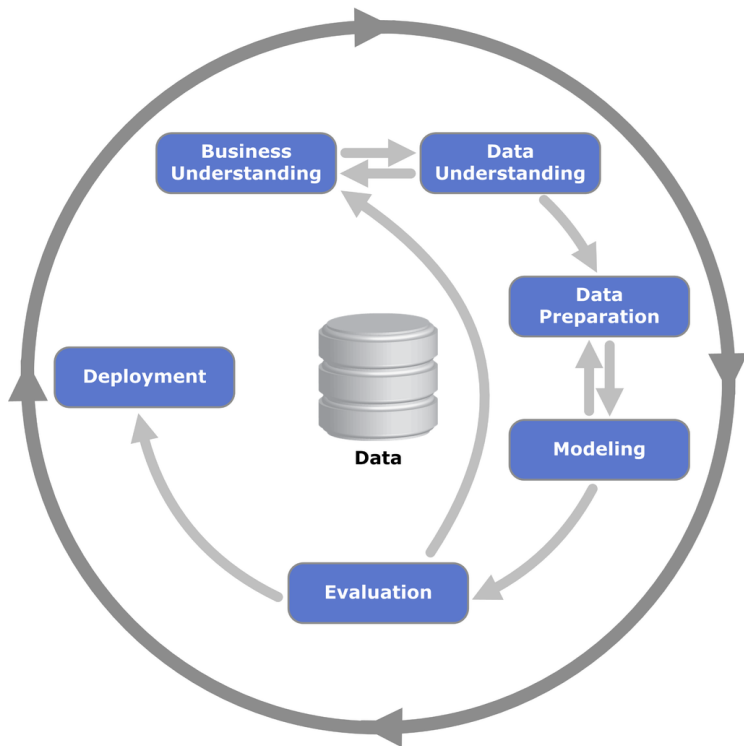
Note: Data are from the 2018 Kaggle ML and Data Science Survey. You can learn more about the study here: <http://www.kaggle.com/kaggle/kaggle-survey-2018>.

A total of 23859 respondents completed the survey; the percentages in the graph are based on a total of 15937 respondents who provided an answer to this question. Only selected job titles are presented.



Metodologias em Analytics

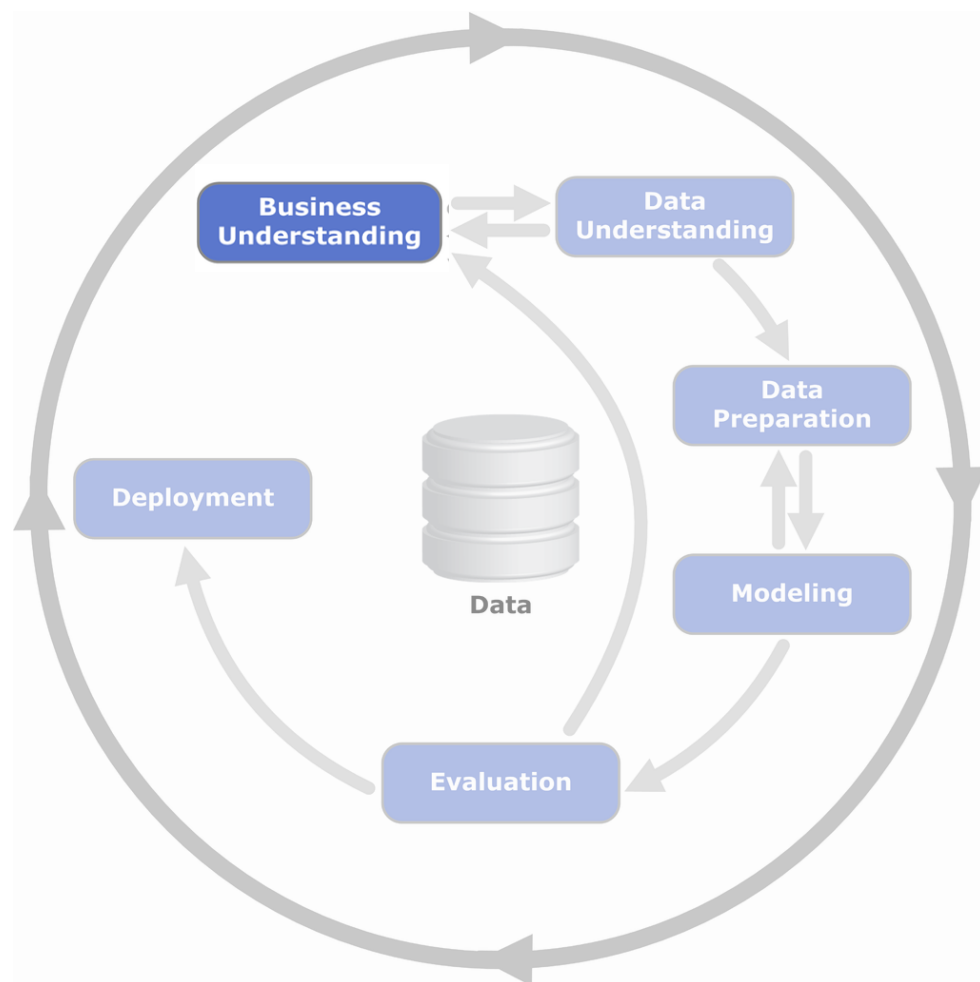
Metodologia: C.R.I.S.P-DM



- A metodologia CRISP-DM* surgiu para apoiar os projetos relacionados com o processamento e a análise de grande volumes de dados.
- A CRISP-DM reúne as melhores práticas para que o processo seja mais produtivo e eficiente.

* *Cross Industry Standard Process for Data Mining*

Business Understanding | Entendendo de Negócios



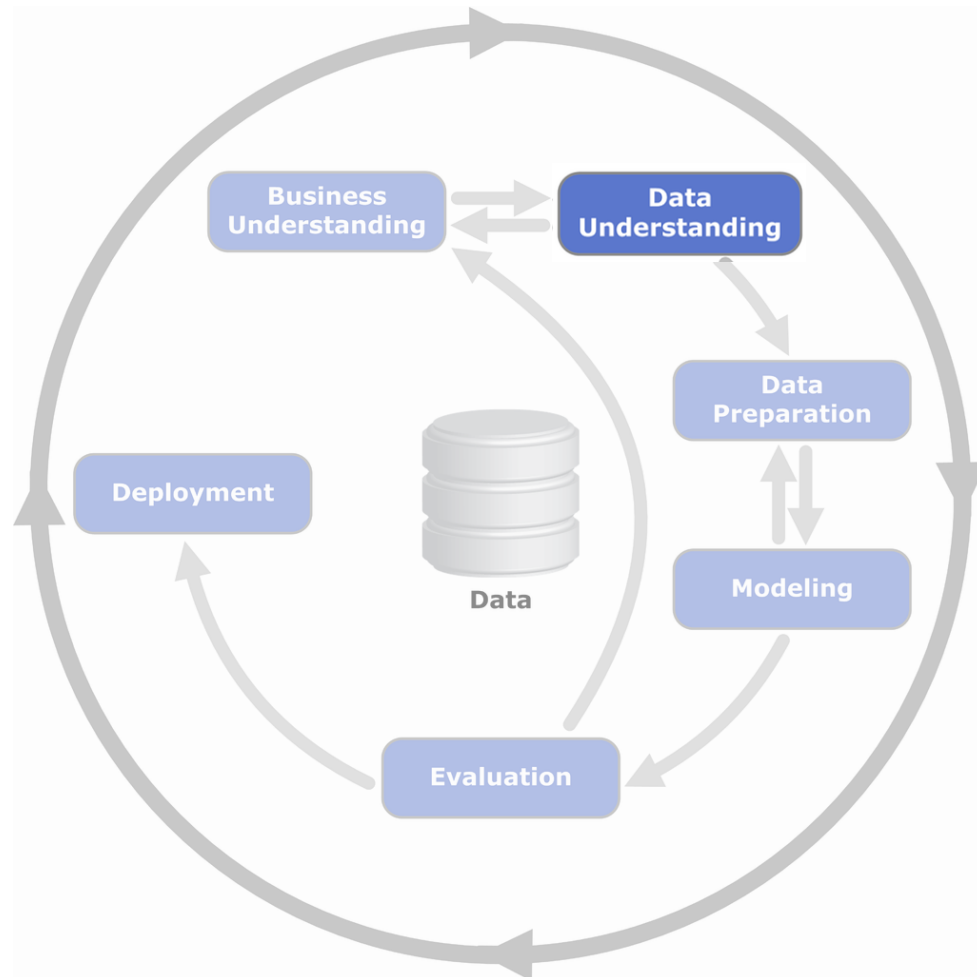
Fase do entendimento das perguntas de negócio/KPIs/ Performance Indicator/Índices de Monitoramento. Alinhamento dos critérios que serão utilizados no levantamento dos dados.

Definição da “espinha dorsal” do projeto, com planos e metas iniciais que darão o direcionamento dos próximos passos.

Mapeamento dos principais resultados já apresentados e projetos implantados pelas áreas de negócio.

Definição dos principais problemas de negócio a serem solucionados. Importante focar nos principais e ter uma lista reduzida com alta prioridade.

Data Understanding | Entendendo os Dados



Mapeamento dos dados já ingeridos, em produção e em “sistemas manuais” (bases acess, planilhas excel, etc.), data lakes, bancos de dados, etc.

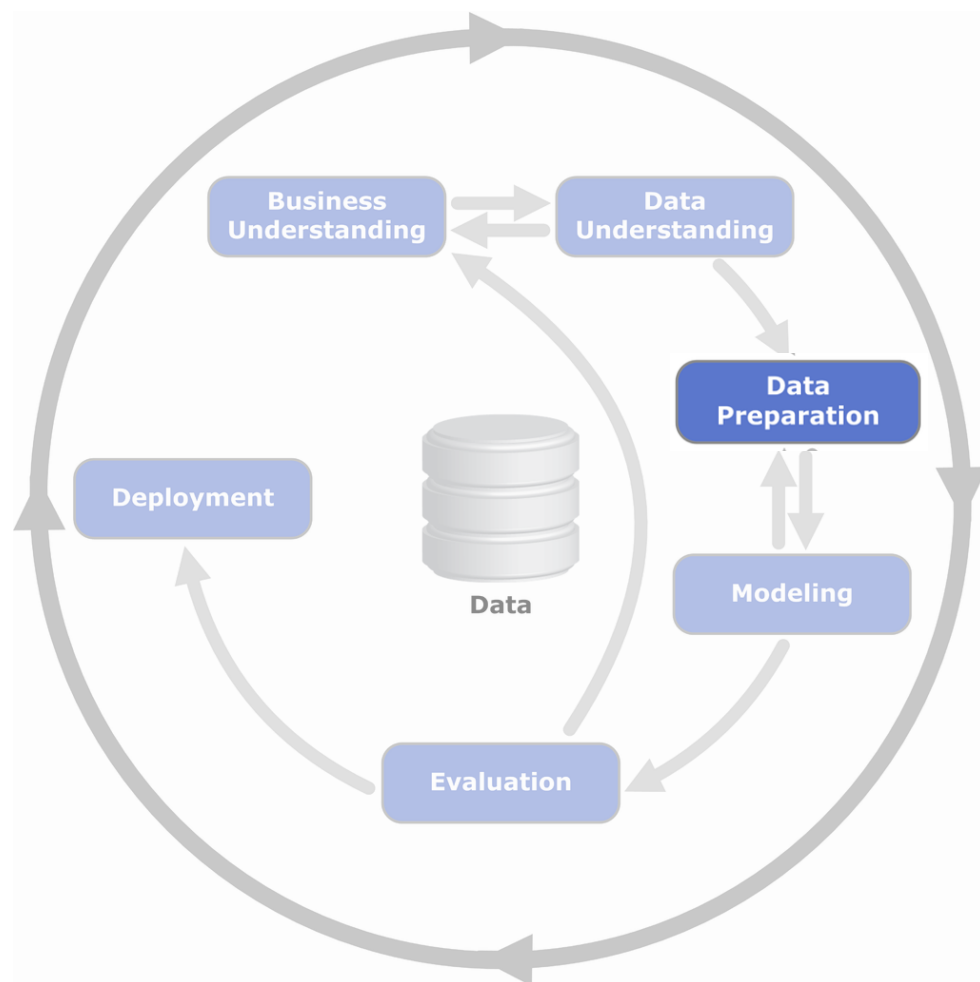
Levantamento dos dados para as perguntas de negócio a serem respondidas. Definição das fontes de dados a serem acessadas.

Avaliação qualitativa dos dados: consistência, integridade, qualidade, problemas estruturais de bases (dados missing, incompletos, nulos, etc.)

Análises primárias (descritiva) e investigações iniciais.

Primeiro checkpoint com PO: apresentação do descritivo e levantamentos iniciais de respostas.

Data Preparation | Preparação dos Dados



Preparação dos procedimentos de bases e seus relacionamentos. Aplicação dos critérios definidos nas etapas anteriores.

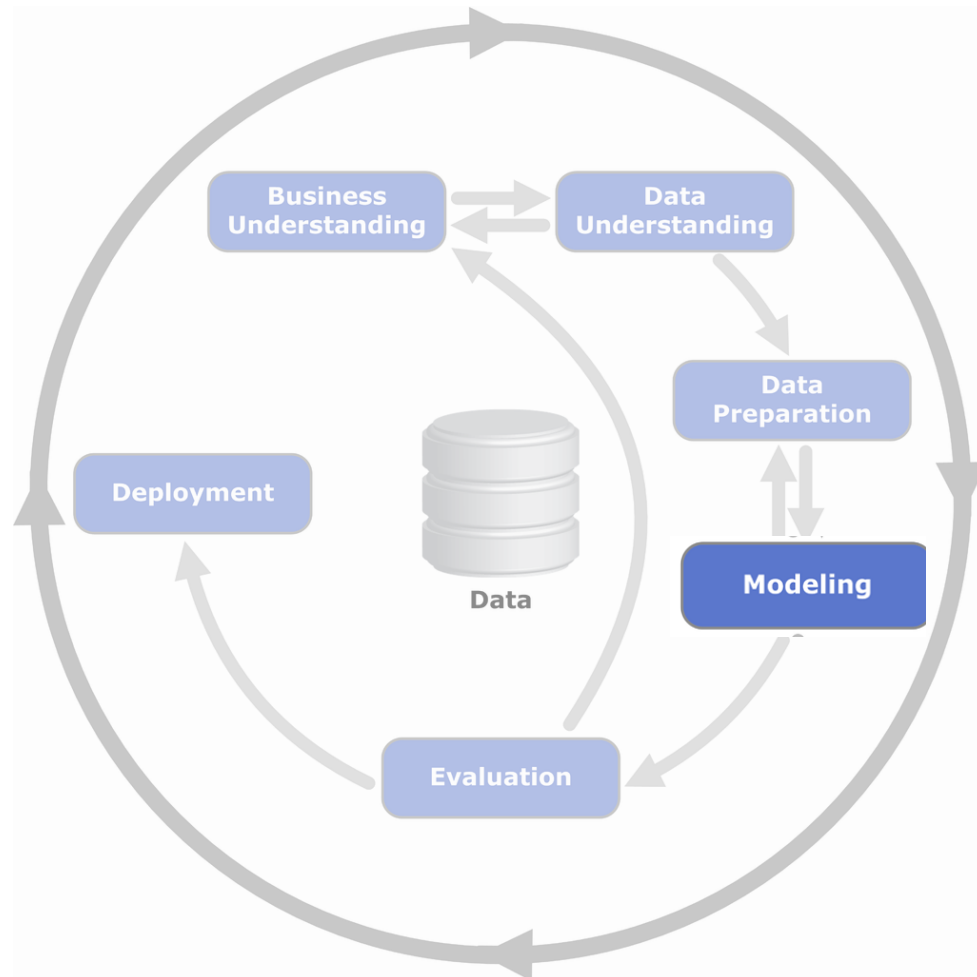
Seleção: escolha das variáveis resposta, variáveis independentes, métricas.

Limpeza: Formatação, padronização e normalização dos dados, tratamentos de dados nulos, faltantes, outliers. Criação de regras de contingenciamento.

Construção: Criação de métricas, indicadores, variáveis auxiliares, flags, etc.

Integração: Criação dos esquemas de relacionamento entre as bases e suas chaves.

Modeling | Modelagem



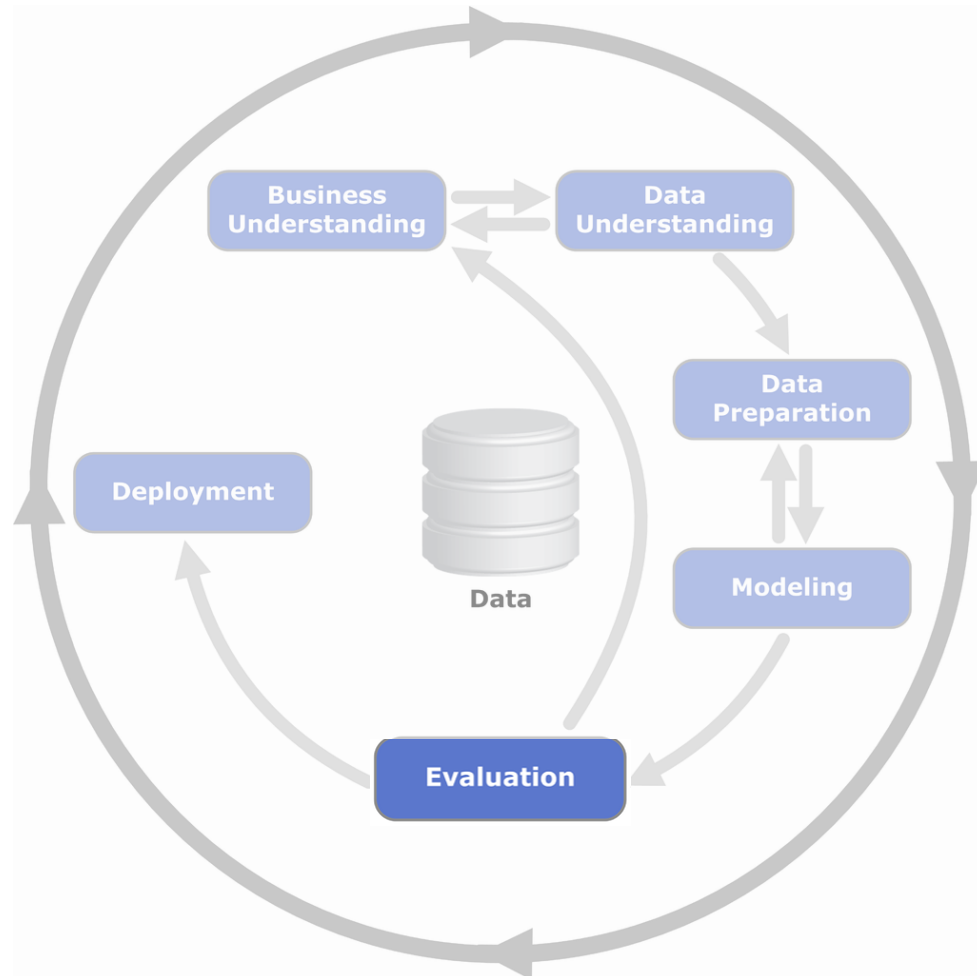
Definição das técnicas em analytics que serão utilizadas na obtenção de respostas e possíveis modelos que possam ser utilizados.

Os modelos podem ser **preditivos** (automação de decisões ou regras de negócios para tomadas de decisão.), ou **prescritivos** nos quais algoritmos de aprendizado de máquina são utilizadas para transformar inputs de negócio em respostas à estas perguntas

Separar bases em desenvolvimento e teste. Possíveis bases amostrais e sintéticas. Desenvolvimento dos modelos selecionados. Testes de robustez e performance dos modelos.

Último checkpoint do ciclo com PO antes da validação

Evaluation | Validação



Validação apartada das metodologias aplicadas e resultados. Testes de robustez, consistência, integridade, coerência e eficiência do modelo.

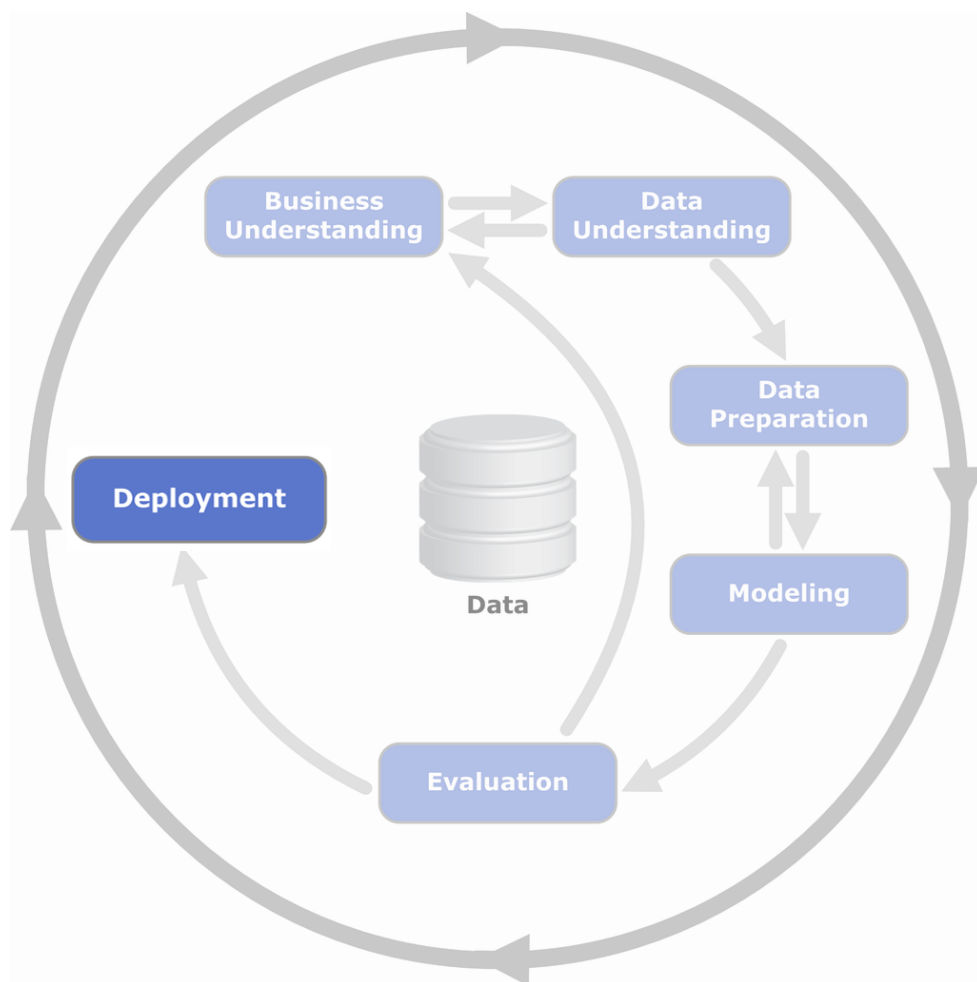
Validação apartada das regras de contingenciamento aplicadas no modelo. Verificação do perfil da base em contingenciamento.

Teste de stress do modelo para performance em cenários extremos e inesperados.

Backtest do modelo (performance fora da amostra e tempo).

Último checkpoint com PO.

Deployment | Implantação



Etapa de desenvolvimento e implantação dos modelos para serem colocados em produção.

Criação de dashboards para acompanhamento dos KPI de negócio e performance do modelo

Implantação dos algoritmos de contingência para tratamento de grupos de risco, outliers e riscos pouco prováveis

Implantação de “circuit breaker” para os casos de risco improváveis e raros (grande impacto de negócio)



Visualização de Dados

Objetivo de Visualização de Dados



Comunicação rápida e eficiente, demonstrando resultados de forma visual e resumida.



Dashboards de acompanhamento de indicadores de performance. Visualização rápida e online de resultados.



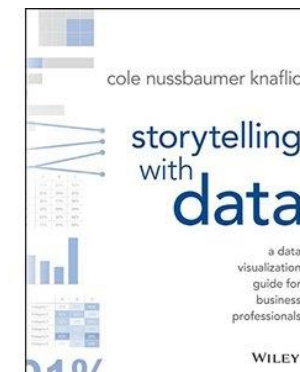
Apresentações executivas com dados e gráficos são maneiras eficientes de formalizar tomadas de decisão e documentar insights.



Gráficos são maneiras eficientes de se analisar dados e encontrar insights e resultados e documentar os processos de descoberta.



Visualização de dados é uma arte. Gráficos bonitos e visualizações agradáveis atraem muito mais a atenção do público.



Gráficos ajudam a contar uma história. A maioria das pessoas são visuais e precisam desse apoio para entender o conteúdo exposto.

Catálogo de Gráficos

<https://datavizcatalogue.com/>



Arc Diagram



Area Graph



Bar Chart



Box & Whisker Plot



Brainstorm



Bubble Chart



Bubble Map



Bullet Graph



Calendar



Candlestick Chart



Chord Diagram



Choropleth Map



Circle Packing



Connection Map



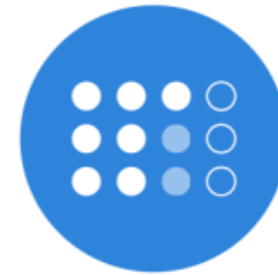
Density Plot



Donut Chart



Dot Map



Dot Matrix Chart

Processo de storytelling a partir de dados

Contexto

Visual

Essencial

Atrair

Design

História



Entenda bem o contexto de seus dados. Quais as perguntas de negócio e hipóteses a serem testadas

Escolha maneiras apropriadas de visualizar seus dados, utilizando displays dinâmicos e atraentes

Elimine das visualizações aquelas informações que não forem essenciais

Seja estratégico, atraia a atenção do seu público para os pontos que você considera como sendo os mais importantes

Pense como um designer e enxergue a experiência do usuário.

Conte uma história interessante

Histogramas

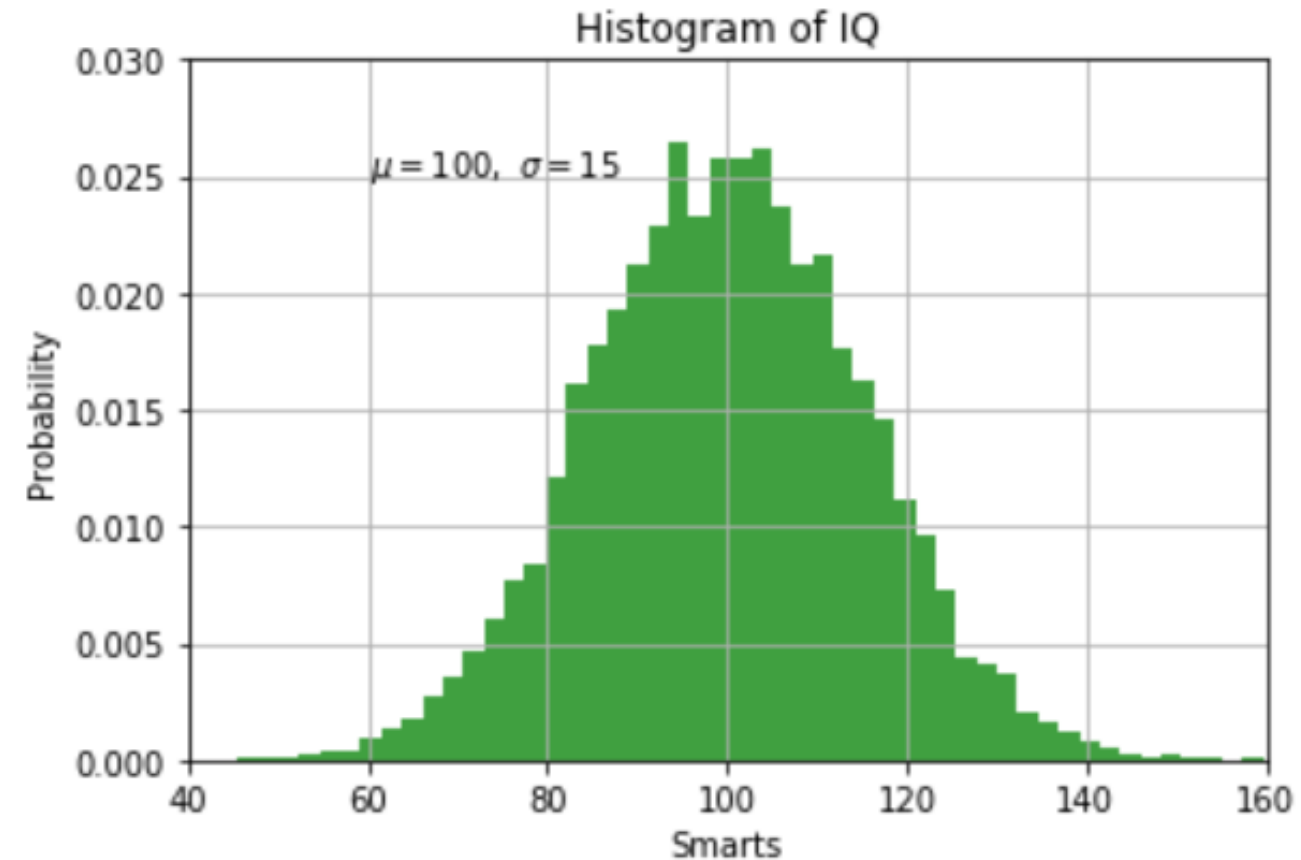


```
import numpy as np
import matplotlib.pyplot as plt
```

```
mu, sigma = 100, 15
x = mu + sigma * np.random.randn(10000)
```

```
# the histogram of the data
n, bins, patches = plt.hist(x, 50, density=1, facecolor='g',
alpha=0.75)
```

```
plt.xlabel('Smarts')
plt.ylabel('Probability')
plt.title('Histogram of IQ')
plt.text(60, .025, r'$\mu=100,\ \sigma=15$')
plt.axis([40, 160, 0, 0.03])
plt.grid(True)
plt.show()
```

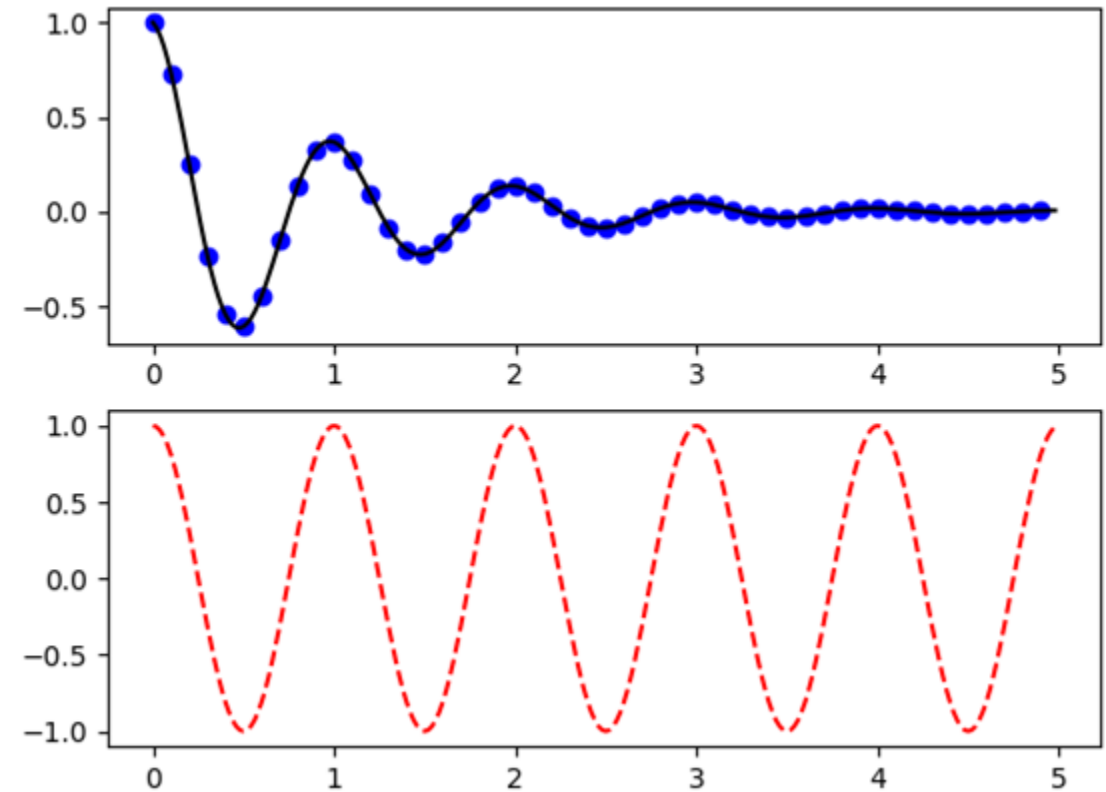


Histogramas são utilizados para visualizar como a frequência de valores de uma variável é distribuída por faixas. Abaixo o gráfico resultante do código ao lado.

Gráficos de Funções Matemáticas

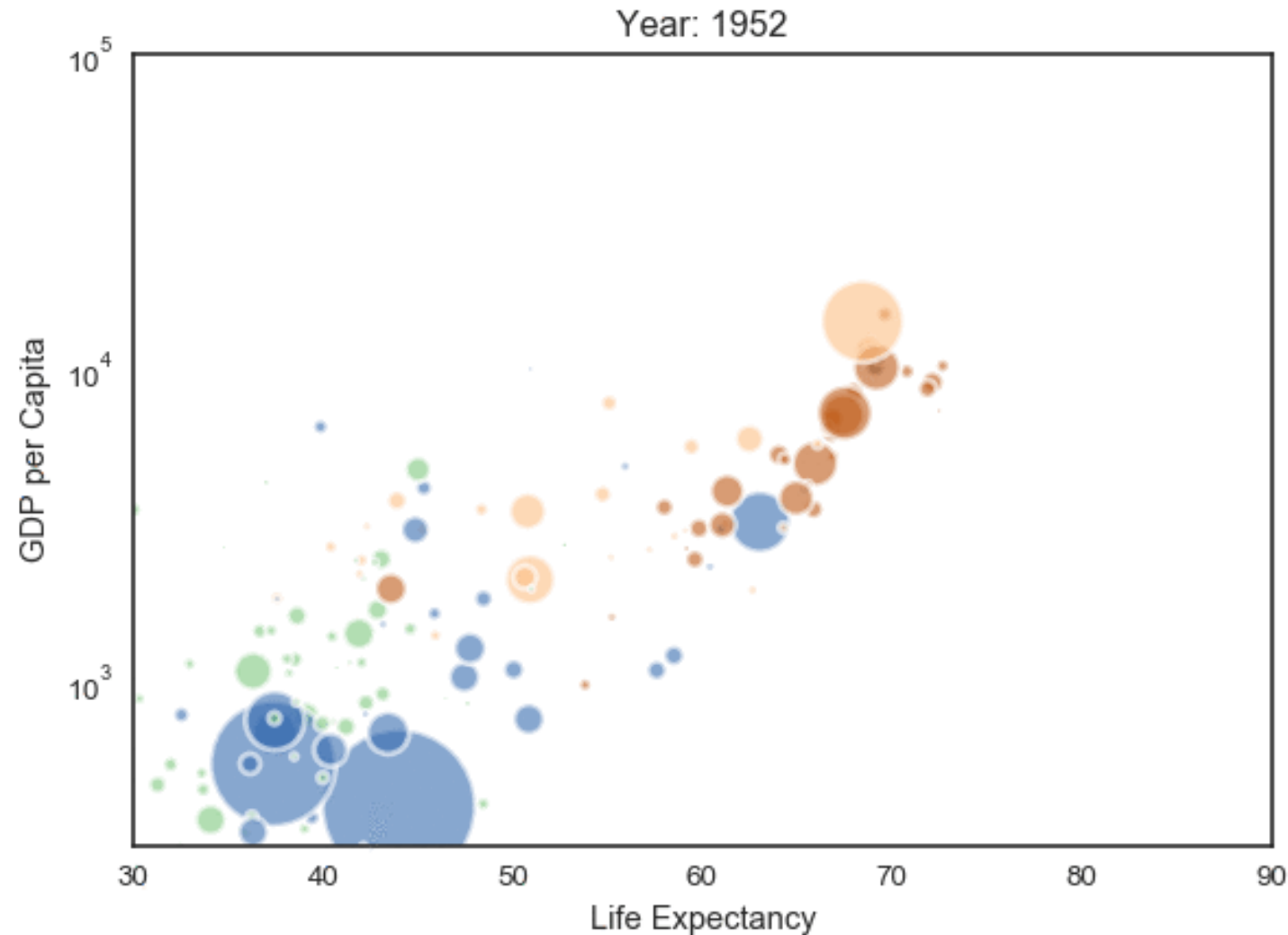


```
def f(t):  
    return np.exp(-t) * np.cos(2*np.pi*t)  
  
t1 = np.arange(0.0, 5.0, 0.1)  
t2 = np.arange(0.0, 5.0, 0.02)  
  
plt.figure()  
plt.subplot(211)  
plt.plot(t1, f(t1), 'bo', t2, f(t2), 'k')  
  
plt.subplot(212)  
plt.plot(t2, np.cos(2*np.pi*t2), 'r--')  
plt.show()
```



Algumas métricas são funções de outras variáveis. Gráficos de funções são importantes na visualização de como essas variáveis se comportam uma em relação a outra.

Gráficos de Bolha (dinâmicos)



Gráficos animados são uma maneira didática e lúdica de demonstrar como variáveis evoluem ao longo do tempo e espaço (análise multivariada).

Além dos eixos ortogonais tradicionais que representam funções entre as variáveis, o tamanho, cores e formatos dos pontos podem representar mais variáveis, e a movimentação destes pontos pode representar como estas variáveis evoluem ao longo de um período de tempo.

Referências



- * <https://python-graph-gallery.com>
- * <https://plot.ly/python/>
- * <https://datavizcatalogue.com/>
- * <https://matplotlib.org/tutorials/introductory/pyplot.html>
- * <https://pandas.pydata.org/pandas-docs/stable/>
- * <https://matplotlib.org/>
- * <https://seaborn.pydata.org/>
- * <https://pbpython.com/visualization-tools-1.html>
- * <https://www.dataquest.io/blog/python-data-visualization-libraries/>
- * <https://towardsdatascience.com/introduction-to-data-visualization-in-python-89a54c97fbed>

