



Curso Aprendizado de Máquina


Aula 1 Introdução e porque este curso é importante

Começando uma carreira em Data Science

- **Ao escolher uma carreira, existem caminhos seguros e arriscados que você pode seguir.** Uma carreira em data science oferece o melhor dos dois mundos. Por um lado, é uma escolha segura – a demanda por talentos na área continua aumentando e não mostra sinais de desaceleração.
- Com análise de dados entre suas habilidades, você será muito **requisitado em praticamente qualquer setor.**
- Por outro lado, é uma área muito desafiadora: estamos produzindo quantidades massivas de dados, e esses dados estão possibilitando coisas incríveis.
- Mas, os métodos e estratégias que precisamos inventar continuamente para aproveitar todos esses dados significam que o futuro está continuamente sendo redefinido em tempo real. Aqueles que estiverem na linha de frente dos dados garantirão a vanguarda do progresso tecnológico.
- A boa notícia é que não importa qual função você desempenhe – a segura, a arriscada ou um meio termo – **existem amplas oportunidades de carreira para todos os interessados em data science.**

Um mundo repleto de dados...

- Toda vez que você envia uma mensagem, escreve um tweet, posta uma foto no Facebook, clica em um link ou compra algo online, **você está gerando dados**.
- E, considerando que existem mais de 3.5 bilhões de usuários de internet no mundo e 2 bilhões de usuários de celular, estamos falando em uma quantidade enorme de dados.



Um mundo repleto de dados... ...esperando para serem analisados!!

- Felizmente, à medida que os dados se multiplicaram, a habilidade de coletar, organizar e analisar também se multiplicou.
- O armazenamento de dados está mais barato, o poder de processamento está mais massivo e as ferramentas estão mais acessíveis do que nunca para extrair os zettabytes de dados disponíveis para inteligência de negócios.
- **Nos últimos anos, a análise de dados fez de tudo, desde prever o preço das ações até evitar incêndios domésticos.**

E quem vai analisá-los ?

- Analisar esses dados requer um exército de mestres dos dados. Traduzindo: **nunca houve uma época melhor para seguir uma carreira em data science.**
- Em 2011, o McKinsey Global Institute previu que, até 2018, os Estados Unidos poderiam enfrentar uma escassez de 1,5 milhão de profissionais que sabem como aproveitar a análise de dados para tomar decisões efetivas – nós já estamos no caminho para exceder esse número.
- Em um relatório recente de ciência de dados da CrowdFlower, um espantoso número de ***“83% dos cientistas de dados pesquisados acham que há uma escassez de cientistas de dados atualmente”***.
- **Entra: você.** O primeiro passo no seu caminho para ser um profissional de data science é conhecer melhor as três principais opções de carreira na área: analista de dados, cientista de dados e engenheiro de dados.

Analista de dados

- Um analista de dados é essencialmente um cientista de dados júnior.
- Esse é o cargo perfeito para começar, se você é novo na carreira de data science e está interessado em saber como ela funciona.
- Os analistas de dados não têm as bases matemáticas ou de pesquisa para inventar novos algoritmos, **mas eles têm uma forte compreensão de como usar as ferramentas existentes para resolver problemas.**





Habilidades e ferramentas dos analistas de dados

- Analistas de dados precisam ter um conhecimento básico de cinco competências: programação, estatística, machine learning, data wrangling e visualização de dados.
- Além das habilidades técnicas, a atenção a detalhes e a habilidade de efetivamente apresentar resultados são igualmente importantes para ser um analista de dados de sucesso.

Q que isso significa ?

- Analistas de dados recebem direcionamento de profissionais com mais experiência nas organizações.
- Baseados no direcionamento, eles obtêm, processam e resumem os dados.
- Analistas de dados são aqueles que garantem a qualidade do data scraping, fazendo consultas em bancos de dados de acordo com a demanda de diferentes áreas da empresa e realizando a triagem nesses dados para chegar a resoluções oportunas.
- Eles também agrupam os dados e os usam para contar histórias ou apresentá-los de maneira visual.
- Ter interesse genuíno por dados e buscar se atualizar sobre as melhores práticas e ferramentas em desenvolvimento servem bem todos os profissionais de dados, não importa o nível de senioridade.

Cientistas de dados

- Algumas empresas consideram os títulos de “cientista de dados” e “analista de dados” como sinônimos. Porém, existem diferenças entre os dois em termos de conjunto de habilidades e experiência.
- Apesar de cientistas de dados e analistas de dados terem a mesma missão em uma organização – coletar informações com base em uma quantidade massiva de dados disponíveis – o trabalho de um cientista de dados requer habilidades mais sofisticadas para trabalhar com um volume maior de dados, garantindo velocidade no processo.
- Propriamente dito, o cientista de dados é alguém que consegue fazer pesquisas não dirigidas e resolver problemas e questões abertas.
- Cientistas de dados normalmente possuem diplomas avançados em uma área quantitativa, como ciência da computação, física, estatística ou matemática aplicada, e possuem conhecimento para inventar novos algoritmos para resolver problemas de dados.
- Cientistas de dados são extremamente valiosos para suas empresas pois seu trabalho pode descobrir novas oportunidades de negócio ou economizar dinheiro da empresa ao identificar padrões escondidos nos dados (por exemplo, destacando comportamentos surpreendentes dos consumidores ou encontrando potenciais falhas nos clusters de armazenamento).

Habilidades e ferramentas

- Enquanto o analista de dados pode considerar os dados de apenas uma única fonte, um cientista de dados explora dados de várias fontes diferentes.
- Cientistas de dados usam ferramentas como o Hadoop (o framework mais utilizado para processamento distribuído de dados), linguagem de programação como Python e R e aplicam procedimentos de matemática e estatística avançadas.



O toolkit do Cientista de Dados

- O exato conjunto de habilidades varia por organização e projeto, mas esse exemplo da Data Science London dá uma noção do quão complexo o conjunto de habilidades do cientista de dados pode ser:



- Java, R, Python (extra: Clojure, Haskell, Scala);
- Hadoop, HDFS, MapReduce (extra: Spark, Storm);
- HBase, Pig, Hive (extra: Shark, Impala, Cascalog);
- ETL, Web scrapers, Flume, Sqoop (extra: Hume);
- SQL, RDBMS, DW, OLAP;
- Knime, Weka, RapidMiner (extra: SciPy, NumPy, Scikit-learn, Pandas);
- D3.js, Gephi, Ggplot2, Tableau, Flare, Shiny;
- SPSS, Matlab, SAS;
- NoSQL, MongoDB, Couchbase, Cassandra;
- E sim: MS-Excel, a ferramenta de data science mais utilizada e subestimada.
- Fonte: Data Science London

O que isso significa ?

- A mais valiosa habilidade não técnica que um cientista de dados traz à mesa é uma intensa curiosidade.
- Cientistas de dados precisam ser motivados a fazer perguntas e buscar soluções e, ao fazê-lo, desenterrar informações que podem transformar o negócio.
- Cientistas de dados, essencialmente, alavancam dados para resolver problemas de negócios. Eles interpretam, extrapolam e prescrevem dados para entregar recomendações práticas.
- Um analista de dados resume o passado; um cientista de dados planeja para o futuro.
- Cientistas de dados podem identificar precisamente:
 - Como otimizar sites para conquistar mais clientes,
 - Como desenvolver produtos de acordo com diferentes comportamentos de consumo ou como ajustar o processo logístico da empresa, priorizando velocidade e reduzindo desperdícios.



Engenheiro de Dados

- Um engenheiro de dados constrói um pipeline de dados robusto e resistente a falhas, limpando, transformando e agregando dados bagunçados e desorganizados em bancos ou fonte de dados.
- Engenheiros de dados são geralmente engenheiros de software por profissão. Em vez de análises, engenheiros de dados são responsáveis por compilar e instalar sistemas de banco de dados, escrever consultas complexas, escalar para múltiplas máquinas e colocar sistemas de recuperação de desastres em funcionamento.
- Engenheiros de dados essencialmente estabelecem as bases para o analista ou o cientista de dados facilmente resgatar os dados necessários para suas avaliações e experimentos.

Habilidades e Ferramentas do Engenheiro de Dados

- Enquanto cientistas de dados extraem valor dos dados, engenheiros de dados são responsáveis por garantir que os dados fluam suavemente de uma fonte para seu destino, para que possam ser processados.
- Sendo assim, engenheiros de dados possuem profundo conhecimento e experiência em:
 - Tecnologias baseadas em Hadoop, como MapReduce, Hive e Pig;
 - Tecnologias baseadas em SQL (Structured Query Language), como PostgreSQL e MySQL;
 - Tecnologias NoSQL como Cassandra e MongoDB;
 - Soluções de armazenamento de dados.

Engenheiro de Machine Learning – Um cargo em ascensão !

- Com o rápido crescimento da Inteligência Artificial, surge uma demanda crescente por Engenheiros de Machine Learning. Software orientado a IA que emprega aprendizado profundo, aprendizado de máquina, reconhecimento de voz, máquinas autônomas e visão computacional são apenas alguns dos fatores.
- Outro fator que impulsiona a ascensão dos Engenheiros de Machine Learning é o déficit de Cientistas de Dados.
- Como resultado, muitas empresas já perceberam que, assim como no desenvolvimento de software, é melhor distribuir o trabalho por várias funções.
- A função de Engenheiro de Machine Learning está entre Engenharia de Software e Ciência de Dados.
- Em equipes maiores, os Engenheiros de Machine Learning liberam os Cientistas de Dados para se concentrarem na modelagem que requer profundo conhecimento científico, como estatísticas ou outras formas de modelagem matemática, deixando o lado da engenharia para os Engenheiros de Machine Learning.

O Que Exatamente é um Engenheiro de Machine Learning?

- Uma rápida pesquisa por “Engenheiro de Machine Learning” mostrará como as habilidades e a experiência são priorizadas em “Qualificações Preferidas”.
- Qualificações como experiência em ciência da computação ou engenharia, habilidades de programação e frameworks de aprendizado de máquina incluídas.
- As habilidades de modelagem matemática, por outro lado, são listadas, mas geralmente não são priorizadas.

Engenheiro de Machine Learning vs Cientista de Dados

- Você encontra descrições das diferenças entre Engenheiro de Machine Learning e Cientistas de Dados que variam de muito boas a completamente erradas, apesar do fato de muitas empresas usarem os termos Engenheiro de Machine Learning e Cientista de Dados de maneira intercambiável. Proponho uma definição um tanto simples de Cientista de Dados:
 - Se você pode programar e criar modelos únicos, utilizáveis e precisos a partir do zero, é um Cientista de Dados.
 - Por outro lado, o que é um Engenheiro de Machine Learning? Isso requer um pouco mais de contexto e entendimento das tendências que estão contribuindo para a Ascensão do Engenheiro de Machine Learning. Acompanhe.

Tendência 1 – Frameworks de Machine Learning e Deep Learning

- Os frameworks de aprendizado de máquina e aprendizado profundo formam grande parte da infraestrutura e fazem a maior parte do trabalho pesado no ecossistema de ciência de dados.
- Nos últimos cinco anos, houve uma série de frameworks lançados.
- Linguagens de programação como Python, R, Julia e até Java têm muitas bibliotecas e pacotes específicos para ML e DL.
- No entanto, é a disponibilidade de código aberto e a facilidade de uso dos frameworks de aprendizado de máquina e aprendizado profundo mais poderosas e ricas em recursos, como TensorFlow, PyTorch, Keras e Spark, que permitem que o papel do Engenheiro de Machine Learning prospere.
- A experiência em pelo menos alguns desses frameworks é um requisito essencial para a função.

Tendência 2 – Modelos Pré-Treinados

- Modelos pré-treinados estão se tornando mais facilmente disponíveis. Esses modelos foram amplamente adotados em redes de Deep Learning, como YOLO e Mask R-CNN, para caixas delimitadoras em detecção de imagens e VGG-Face e FaceNet para reconhecimento facial (todos estudados em nosso curso de Visão Computacional).
- A mesma tendência continua com o Processamento de Linguagem Natural (PLN), o entendimento de linguagem natural (NLU) e a geração de linguagem natural (NLG).
- Modelos pré-treinados estão tornando acessíveis os chatbots inteligentes, o sistema de perguntas e respostas, a tradução de idiomas e muitos outros aplicativos de PLN.
- Alguns dos modelos pré-treinados polivalentes conhecidos incluem BERT, GPT-2, UMLFIT e, especialmente, a biblioteca API Hugging Face Transformers API, que fornece acesso imediato a mais de 32 modelos NLU e NLG pré-treinados.
- Além disso, bibliotecas como spaCy fornecem modelos pré-treinados para fins gerais, capazes de prever entidades nomeadas, tags de parte do discurso e dependências sintáticas.
- Os Engenheiros de Machine Learning aproveitam o fato de que muitos desses modelos podem ser usados imediatamente e são ajustados com relativa facilidade para dados mais específicos e personalizados.

Tendência 3 – Aprendizado de Máquina Automatizado (AutoML)

- Há também uma tendência crescente em direção ao aprendizado de máquina automatizado. De um modo geral, isso inclui seleção automática de recursos, transformação de dados e outras funções especializadas, visando facilitar o trabalho de Cientistas de Dados.
- O AutoML inicialmente era uma ferramenta de produtividade de ciência de dados, ajudando a reduzir o tempo necessário para muitas dessas tarefas.
- Agora, é uma parte essencial do conjunto de habilidades do Engenheiro de Machine Learning, permitindo que eles automatizem a preparação de dados, incluindo seleção de imputação de recursos, e realizando uma melhor pesquisa de modelo com otimização automática de hiperparâmetros.
- As ferramentas AutoML continuarão a ficar mais sofisticadas, permitindo que os Engenheiros de Machine Learning executem mais tarefas que eram da competência dos Cientistas de Dados.

Tendência 4 – MLOps e Engenheiro de Dados

- Longe vai os dias em que os Cientistas de Dados podem criar um modelo localmente e depois implantá-los facilmente na produção.
- Semelhante ao papel que o DevOps e a engenharia de infraestrutura desempenham na engenharia de software, os MLOps e os Engenheiros de Dados estão executando componentes essenciais nos projetos de aprendizado de máquina e aprendizado profundo.
- Os Engenheiros de Machine Learning, por definição, são programadores que possuem as habilidades necessárias para criar os fluxos de trabalho e a infraestrutura de ML necessários para mover projetos do início até a produção.
- Mecanismos de aprendizado de máquinas distribuídos, como o Apache Spark, e plataformas de gerenciamento de fluxo de trabalho, como o Apache Airflow e o Kubeflow, são apenas algumas das ferramentas que os Engenheiros de Machine Learning empregam para criar pipelines de dados.
- Dada a infraestrutura e as ferramentas empregadas, esse tipo de trabalho deve ser realizado na nuvem, não localmente. Assim, o domínio preferido de um Engenheiro de Machine Learning é a nuvem (Cloud Computing).

Tendência 5 – Tendências do Mercado de Trabalho

- A demanda por Cientistas de Dados continua superando a oferta em ordens de magnitude.
- As organizações entendem a necessidade de formar uma equipe em torno de projetos de IA que inclua Cientistas de Dados, Engenheiros de Machine Learning, Engenheiros de Dados, Engenheiros de IA e muito mais.
- Assim, todos os laboratórios de IA, os gigantes da tecnologia Google, Facebook e Uber, as empresas da Fortune 500 como Bloomberg, CitiBank, Biogen, GE e Ford – para não mencionar as startups como Tesla e Airbnb – estão contratando Engenheiros de Machine Learning.
- Com o aumento da demanda, vem o aumento dos salários, o que está atraindo muitos para o campo. E a tendência é de crescimento para os próximos anos.

O que é python

- Python foi criado no final dos anos oitenta(1989) por Guido van Rossum no Centro de Matemática e Tecnologia da Informação (CWI, Centrum Wiskunde e Informatica), na Holanda, como sucessor da linguagem de programação ABC, capaz de lidar com exceções e interagir com o sistema operacional Amoeba.



Curiosidades

- O nome da língua vem do gosto de seu criador pelos humoristas britânicos Monty Python.
- Van Rossum é o principal autor de Python, e seu papel central contínuo na decisão da direção de Python é reconhecido, referindo-se a ele como Ditador de Vida Benevolente (em inglês: Benevolent Dictator for Life, BDFL).
- Python é uma linguagem de programação interpretada cuja filosofia enfatiza uma sintaxe favorecendo um código mais legível, além de ser "free".

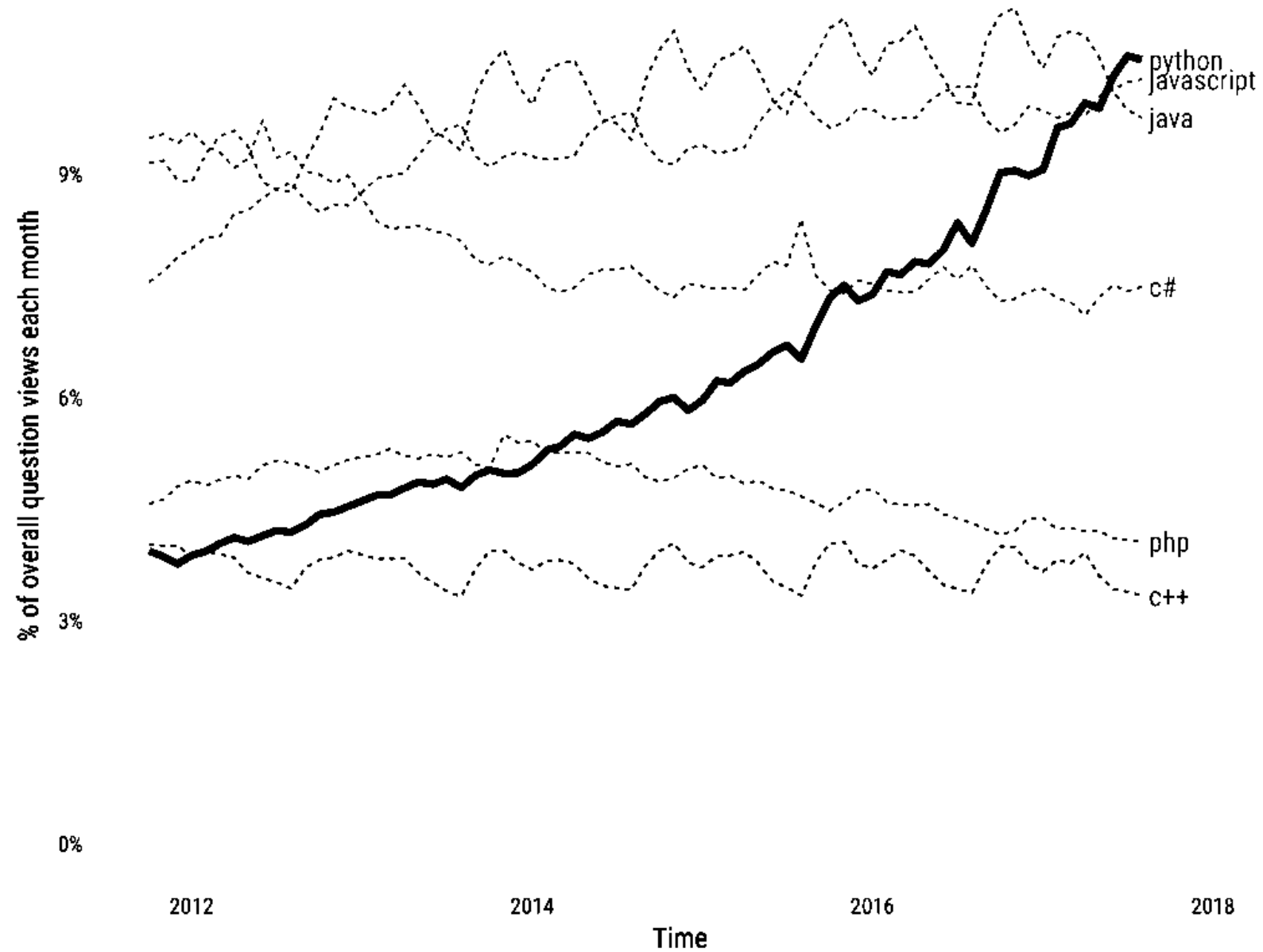


Crescimento da Linguagem Python

- Acompanhando o Stack Overflow Trends fica fácil perceber que a linguagem Python vem crescendo rapidamente nos últimos anos.
- Mas vamos nos concentrar em países de alta renda e considerar as visitas às questões, e não as questões respondidas (isto tende a dar resultados semelhantes, mas tem menos ruído mês a mês, especialmente para tags menores).
- O estudo começa considerando de dados de visualizações a questões sobre Python no final de 2011 e, no período de 2011 a 2017, podemos considerar o crescimento da linguagem Python em relação a outras cinco principais linguagens de programação.
- Observe que esta é, portanto, uma escala de tempo menor do que a ferramenta Stack Overflow Trends, que se remonta a 2008.
- Este gráfico a seguir contém seis das dez tags mais visitadas no Stack Overflow em países de alta renda; as quatro não incluídas são CSS, HTML, Android e JQuery.

Growth of major programming languages

Based on Stack Overflow question views in World Bank high-income countries

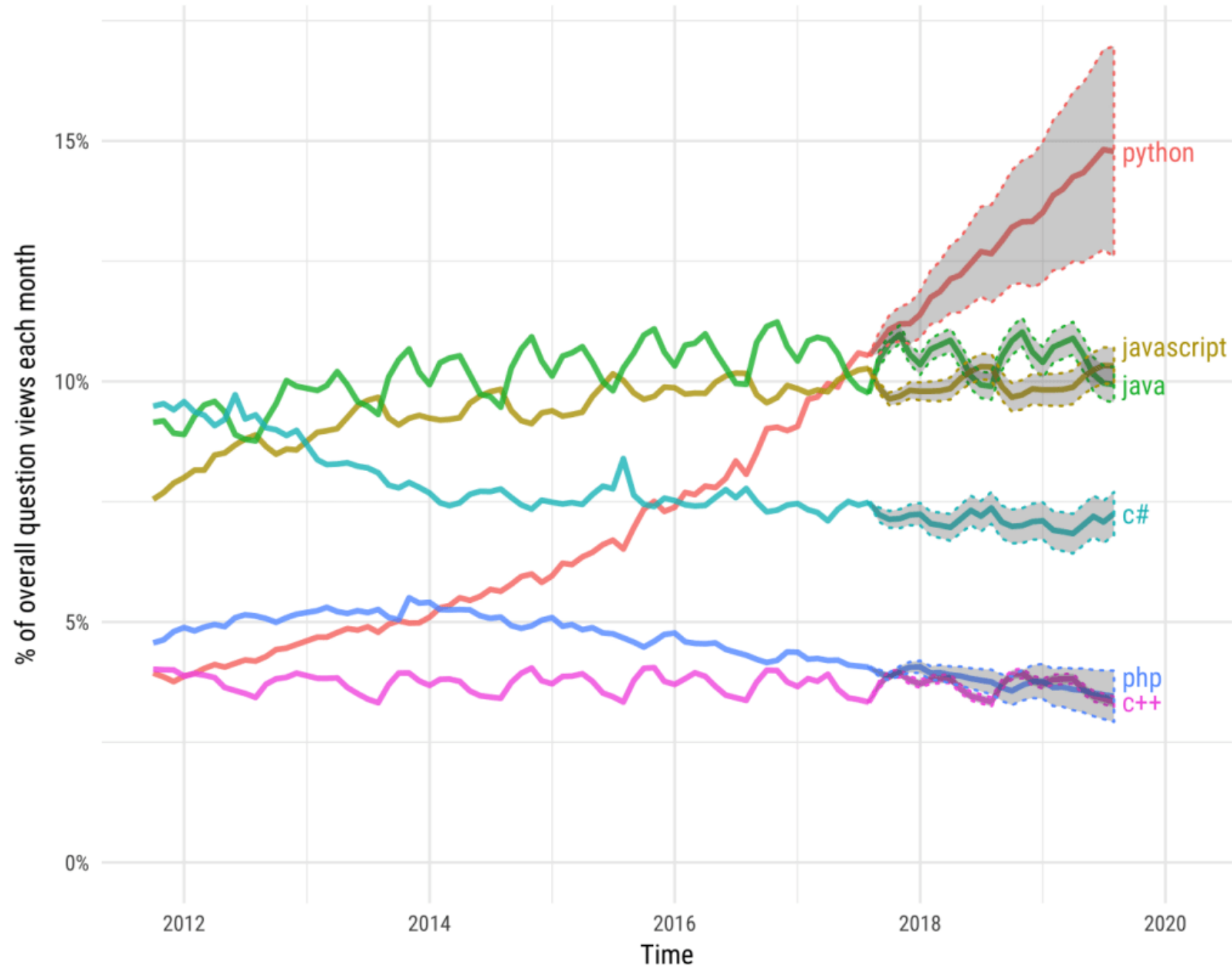


Previsões de crescimento

- Junho de 2017 foi o primeiro mês em que Python foi a tag mais visitada no Stack Overflow em países de alta renda. Isso incluiu ser a tag mais visitada nos EUA e no Reino Unido, e no top 2 em quase todas as outras nações de alta renda (ao lado de Java e JavaScript).
- Isto é especialmente impressionante porque em 2012 a tag Python foi menos visitada do que qualquer uma das outras 5 linguagens, e cresceu 2,5 vezes nesse período.
- Parte disso é devido à natureza sazonal do tráfego para Java. Uma vez que é muito ensinado em cursos de graduação, o tráfego de Java tende a aumentar durante o outono e a primavera e cair durante o verão.
- Será que Java vai se recuperar em relação a linguagem Python novamente até o final do ano?
- Podemos tentar prever os próximos dois anos de crescimento com um modelo chamado "STL", que combina o crescimento com as tendências sazonais para fazer uma previsão sobre os valores futuros.

Projections of future traffic for major programming languages

Future traffic is predicted with an STL model, along with an 80% prediction interval.



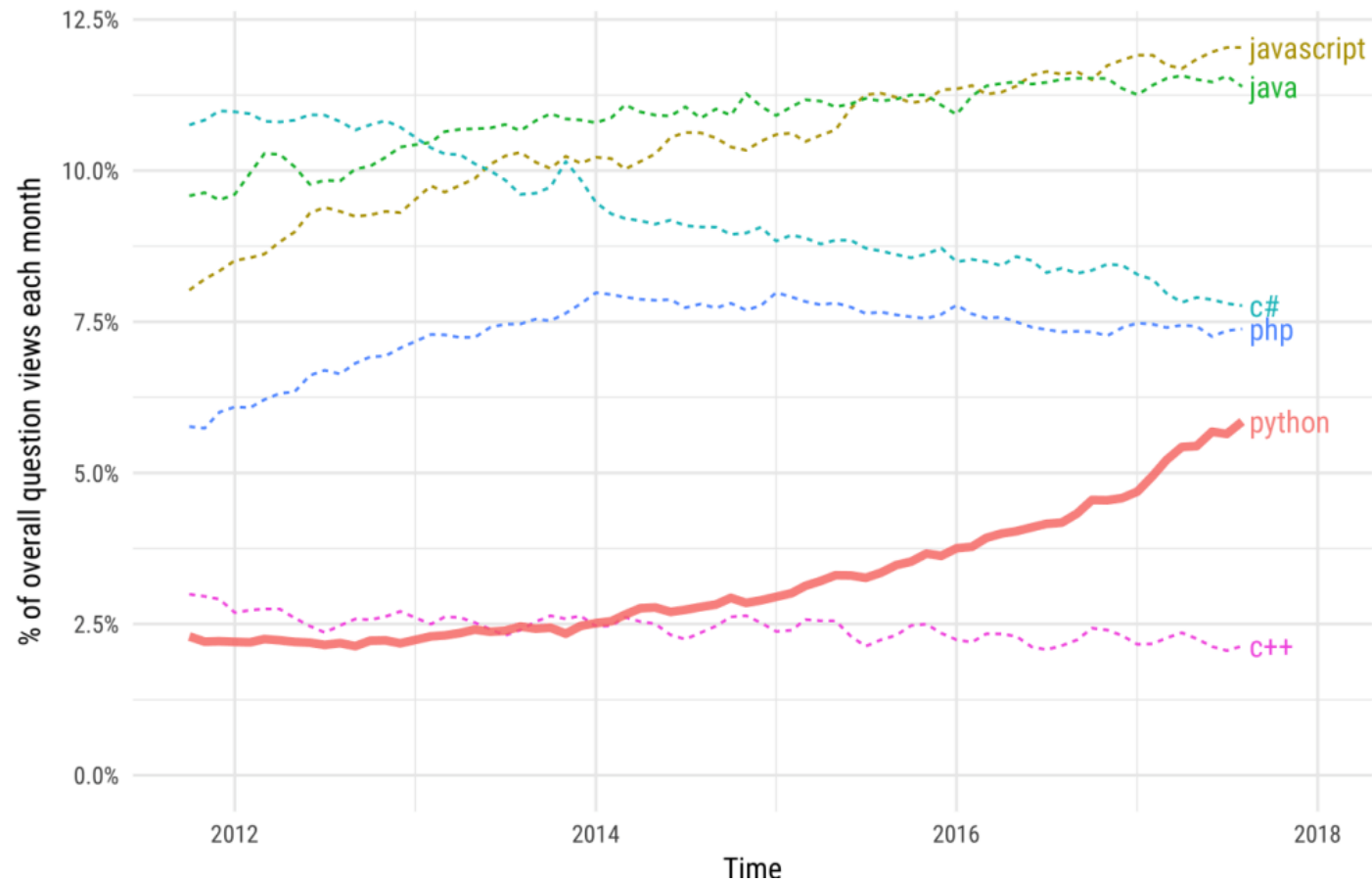
No restante do mundo (incluindo Brasil)

Até aqui, analisamos as tendências nos países de alta renda. Python mostra um crescimento semelhante no resto do mundo, em países como Índia, Brasil, Rússia e China?

Na verdade, sim.

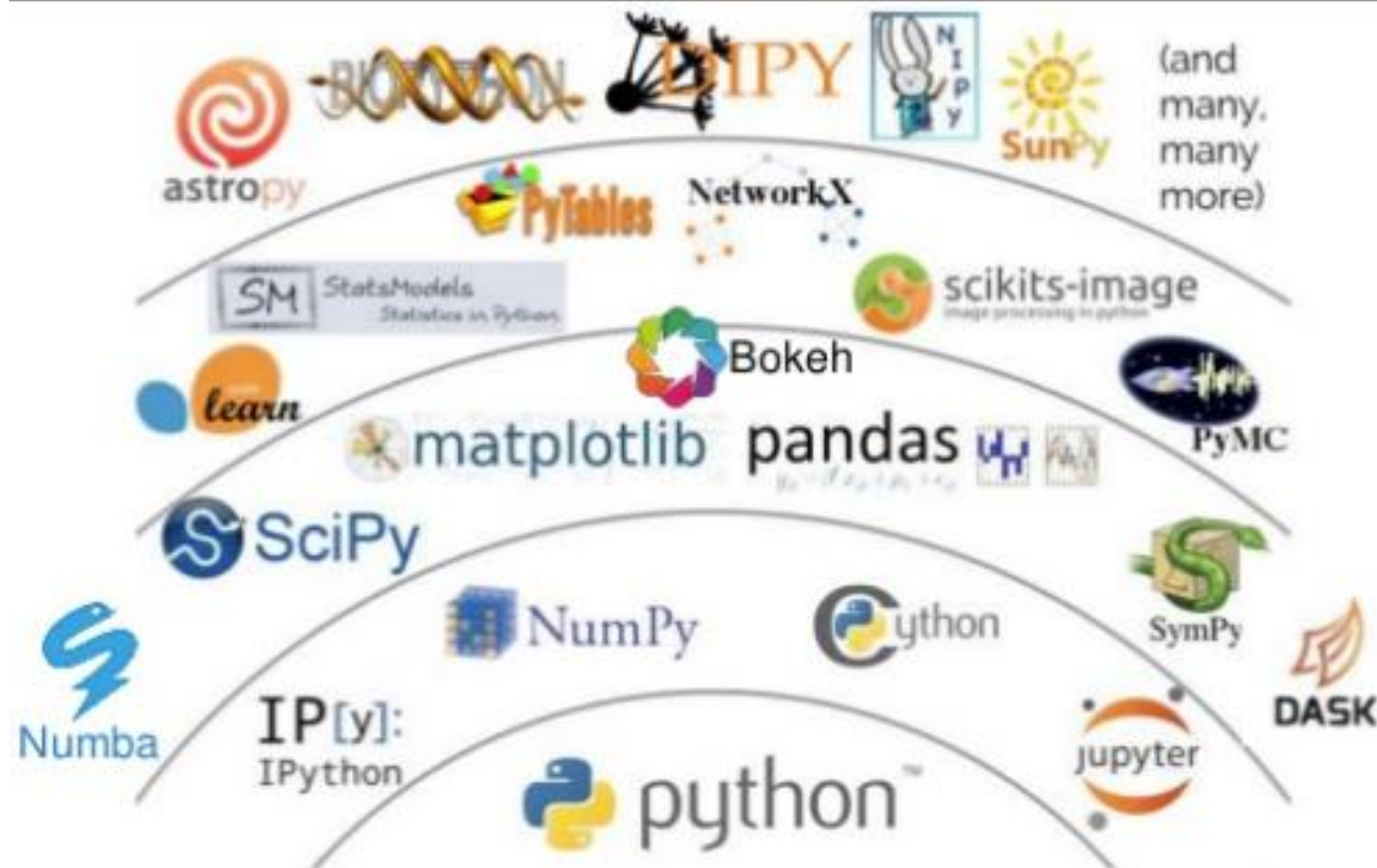
Growth of major programming languages in non-high-income countries

Based on Stack Overflow question views in countries not classified as high-income by the World Bank.



As principais bibliotecas Python

Vamos conhecer as principais bibliotecas que transformam o Python numa poderosa ferramenta de aprendizado de máquina e data science



NUMPY

<http://www.numpy.org/>



NumPy

- Tradicionalmente, começamos nossa lista com as bibliotecas para aplicativos científicos e o NumPy é um dos principais pacotes nessa área.
- Ele é destinado ao processamento de grandes matrizes e matrizes multidimensionais, e uma extensa coleção de funções matemáticas de alto nível e métodos implementados possibilitam a execução de várias operações com esses objetos.
- Durante o último ano, um grande número de melhorias foi feito na biblioteca.
- Além das correções de bugs e problemas de compatibilidade, as mudanças cruciais dizem respeito às possibilidades de estilo, ou seja, o formato de impressão dos objetos NumPy.
- Além disso, algumas funções agora podem manipular arquivos de qualquer codificação disponível no Python.

SCIPY

<https://scipy.org/scipylib/>



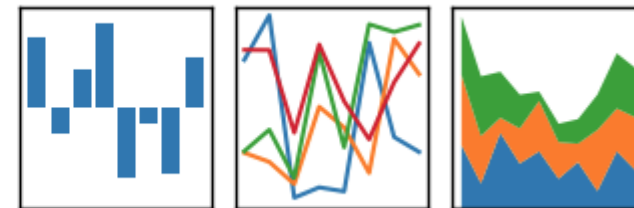
- Outra biblioteca central para computação científica é o SciPy.
- Ele é baseado no NumPy e, portanto, estende seus recursos.
- A principal estrutura de dados SciPy é novamente uma matriz multidimensional, implementada pelo Numpy.
- O pacote contém ferramentas que ajudam a resolver álgebra linear, teoria da probabilidade, cálculo integral e muitas outras tarefas.
- O SciPy recebeu grandes melhorias na construção, na forma de integração contínua em diferentes sistemas operacionais, novas funções e métodos e, o que é especialmente importante - os otimizadores atualizados.
- Além disso, muitas novas funções BLAS e LAPACK foram agrupadas.

PANDAS

<https://pandas.pydata.org/>

pandas

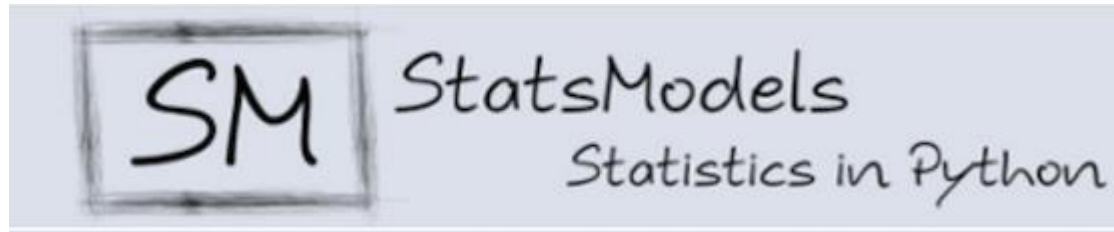
$$y_{it} = \beta' x_{it} + \mu_i + \epsilon_{it}$$



- Pandas é uma biblioteca Python que fornece estruturas de dados de alto nível e uma grande variedade de ferramentas para análise.
- A grande característica deste pacote é a capacidade de traduzir operações bastante complexas com dados em um ou dois comandos.
- O Pandas contém muitos métodos internos para agrupar, filtrar e combinar dados, bem como a funcionalidade de séries temporais.
- Tudo isso é seguido por indicadores de velocidade impressionantes.
- Houve alguns novos lançamentos recentes da biblioteca Pandas, incluindo centenas de novos recursos, aprimoramentos, correções de bugs e alterações de API.
- As melhorias consideram as habilidades do Pandas de agrupar e classificar dados, saída mais adequada para o método apply e o suporte na execução de operações de tipos personalizados.

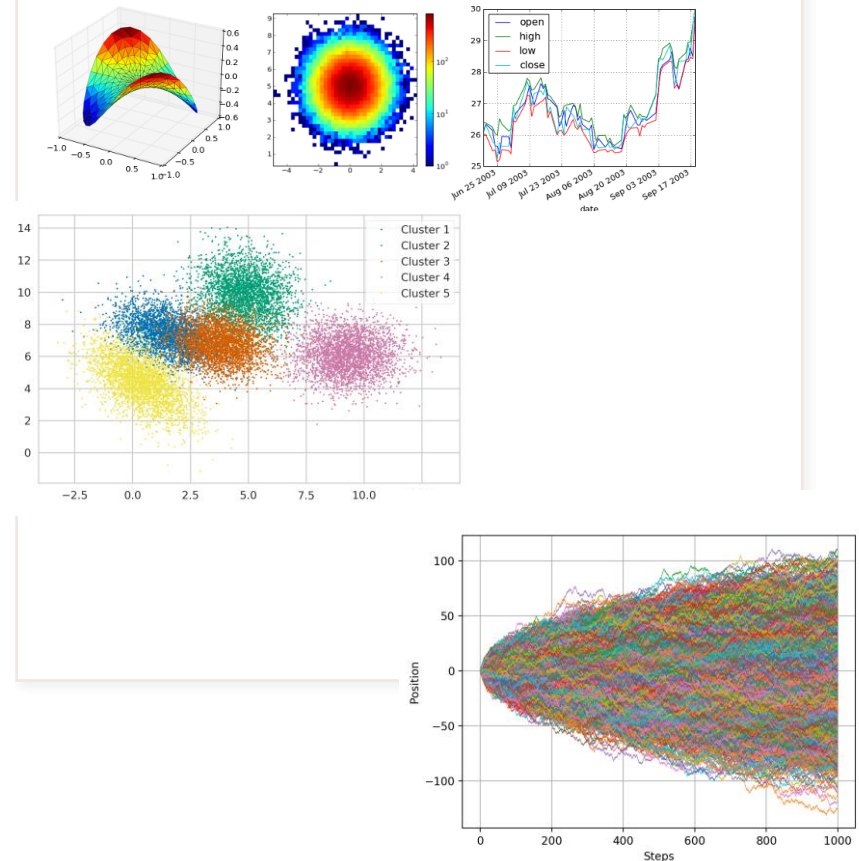
STATSMODELS

<http://www.statsmodels.org/devel/>



- Statsmodels é um módulo Python que oferece muitas oportunidades para análise de dados estatísticos, como a estimação de modelos estatísticos, a realização de testes estatísticos, etc.
- Com este pacote, você pode implementar muitos métodos de aprendizado de máquina e explorar diferentes possibilidades de plotagem.
- A biblioteca está em constante desenvolvimento, enriquecendo e trazendo novas oportunidades.
- Assim, este ano trouxe melhorias de séries temporais e novos modelos de contagem, como o GeneralizedPoisson, modelos NegativeBinomialP, e novos métodos multivariados – análise fatorial, MANOVA e medidas repetidas dentro da ANOVA.

MATPLOTLIB



- O Matplotlib é uma biblioteca de baixo nível para criar diagramas e gráficos bidimensionais.
- Com este pacote, você pode construir gráficos diversos, desde histogramas e gráficos de dispersão a gráficos de coordenadas não cartesianas.
- Além disso, muitas bibliotecas de plotagem populares são projetadas para trabalhar em conjunto com o matplotlib.
- Houve recentemente mudanças de estilo em cores, tamanhos, fontes, legendas, etc.
- Recebeu ainda melhorias no alinhamento automático de legendas nos eixos e, entre melhorias significativas de cores, há um novo ciclo de cores compatível com daltônicos.

<https://matplotlib.org/index.html>



SEABORN

- O Seaborn é essencialmente uma API de alto nível baseada na biblioteca matplotlib.
- Ele contém configurações padrão mais adequadas para o processamento de gráficos.
- Além disso, há uma rica galeria de visualizações, incluindo alguns tipos complexos, como séries temporais, diagramas conjuntos e diagramas de violino.
- As atualizações recentes cobrem principalmente correções de bugs.
- No entanto, houve melhorias na compatibilidade entre FacetGrid ou PairGrid e backends de matplotlib interativos aprimorados, adicionando parâmetros e opções às visualizações.

<https://seaborn.pydata.org/>

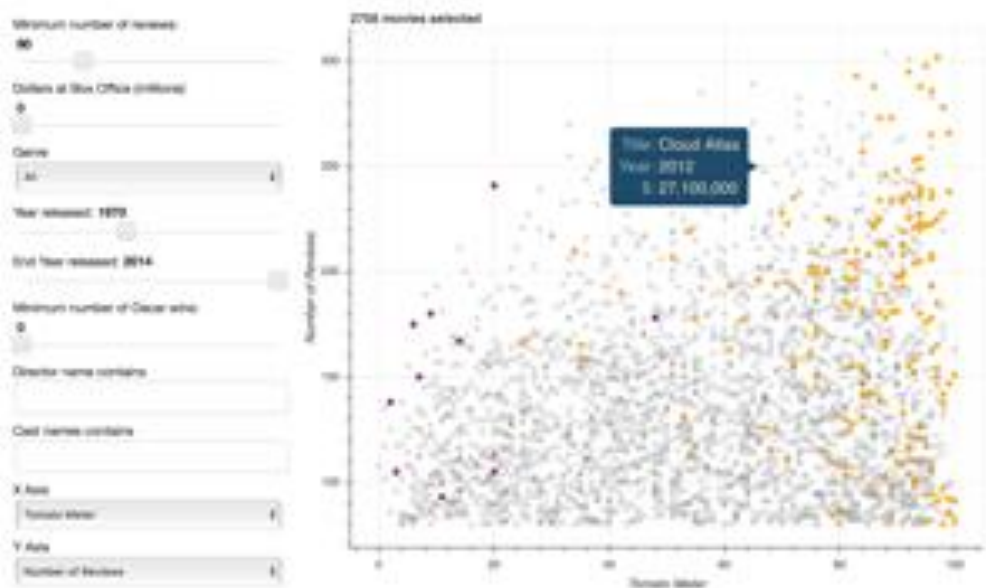
PLOTLY



- Plotly é uma biblioteca popular que permite construir facilmente gráficos sofisticados.
- O pacote é adaptado para trabalhar em aplicativos da web interativos.
- Entre suas visualizações notáveis estão gráficos de contorno, gráficos ternários e gráficos 3D.
- Os aprimoramentos contínuos da biblioteca com novos gráficos e recursos trouxeram o suporte para “várias visualizações vinculadas”, bem como a integração de animação e interferência.

<https://plot.ly/python/>

BOKEH



- A biblioteca Bokeh cria visualizações interativas e escalonáveis em um navegador usando widgets JavaScript.
- A biblioteca oferece uma coleção versátil de gráficos, possibilidades de estilo, habilidades de interação na forma de vincular gráficos, adicionar widgets e definir retornos de chamada, além de muitos outros recursos úteis.
- O Bokeh possui habilidades interativas aprimoradas, como uma rotação de rótulos categóricos, bem como aprimoramentos de pequenos campos de ferramenta de zoom e de dicas de ferramentas personalizadas.

<https://bokeh.pydata.org/en/latest/>

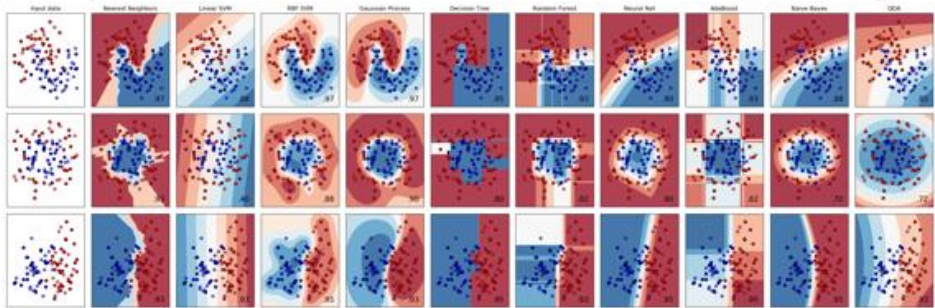


PYDOT

- Pydot é uma biblioteca para gerar grafos complexos orientados e não orientados.
- É uma interface para o Graphviz, escrita em Python puro.
- Com a sua ajuda, é possível mostrar a estrutura dos grafos, que muitas vezes são necessários ao construir algoritmos baseados em redes neurais e árvores de decisão.

SCIKT LEARN

- Este módulo Python baseado no NumPy e SciPy é uma das melhores bibliotecas para trabalhar com dados.
- Ele fornece algoritmos para muitas tarefas padrão de aprendizado de máquina e mineração de dados, como clustering, regressão, classificação, redução de dimensionalidade e seleção de modelo.
- Há vários aprimoramentos feitos na biblioteca. A validação cruzada foi modificada, fornecendo uma capacidade de usar mais de uma métrica.
- Vários métodos de treinamento, como vizinhos mais próximos, e regressões logísticas enfrentaram algumas pequenas melhorias.
- Finalmente, uma das principais atualizações é a realização do Glossário de Termos Comuns e Elementos de API que estão familiarizados com a terminologia e convenções usadas no Scikit-learn.



Google Colab- Guia do Iniciante

- O Google Colab ou “Colaboratório” é um serviço de nuvem gratuito hospedado pelo Google para incentivar a pesquisa de Aprendizado de Máquina e Inteligência Artificial, onde muitas vezes a barreira para o aprendizado e o sucesso é a exigência de um tremendo poder computacional.

Benefícios do Colab















- Além de ser fácil de usar (o que descreverei mais adiante), o Colab é bastante flexível em sua configuração e faz muito do trabalho pesado para você.
- Suporte para Python 2.7 e Python 3.6;
- Aceleração de GPU grátis;
- Bibliotecas pré-instaladas: Todas as principais bibliotecas Python, como o TensorFlow, o Scikit-learn, o Matplotlib, entre muitas outras, estão pré-instaladas e prontas para serem importadas;
- Construído com base no Jupyter Notebook;
- Recurso de colaboração (funciona com uma equipe igual ao Google Docs): o Google Colab permite que os desenvolvedores usem e compartilhem o Jupyter notebook entre si sem precisar baixar, instalar ou executar qualquer coisa que não seja um navegador;
- Suporta comandos bash;
- Os notebooks do Google Colab são armazenados no drive.

Vamos Começar!


Criando um notebook com o Colab



Abra o [Google Colab](https://colab.research.google.com/);

Clique em “novo notebook” e selecione o notebook Python 2 ou o notebook Python 3.

ExamplesRecentGoogle DriveGitHubUpload				
Filter notebooks				
Title	First opened	Last opened		
 Welcome To Colaboratory	Oct 4, 2019	0 minutes ago	 	
 Welcome To Colaboratory	1 hour ago	3 minutes ago		
 Dígitos Manuscritos a mão com Scikit Learn.ipynb	Oct 29, 2019	Dec 15, 2019	 	
 aco.ipynb	Nov 19, 2019	Nov 19, 2019	 	
 overfitting.ipynb	Oct 29, 2019	Oct 29, 2019	 	
				NEW NOTEBOOK CANCEL

<https://colab.research.google.com/notebooks/welcome.ipynb#recent=true>

 **Table of contents** ×

- <> **Introducing Colaboratory**
- Getting Started
-  More Resources
- Machine Learning Examples: Seedbank
-  Section

Welcome to Colaboratory!

Colaboratory is a free Jupyter notebook environment that requires no setup and runs entirely in the cloud.

With Colaboratory you can write and execute code, save and share your analyses, and access powerful computing resources, all for free from your browser.

[] Introducing Colaboratory

This 3-minute video gives an overview of the key features of Colaboratory:



