



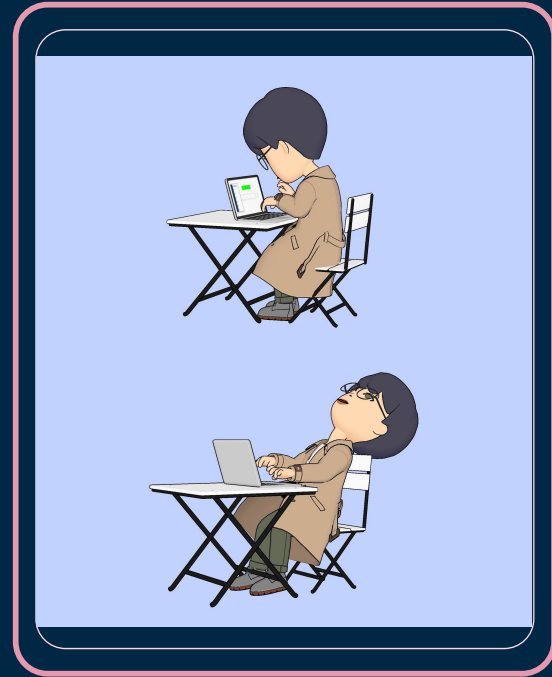
Learn Python from scratch

#MyPathToDataSCIENCE

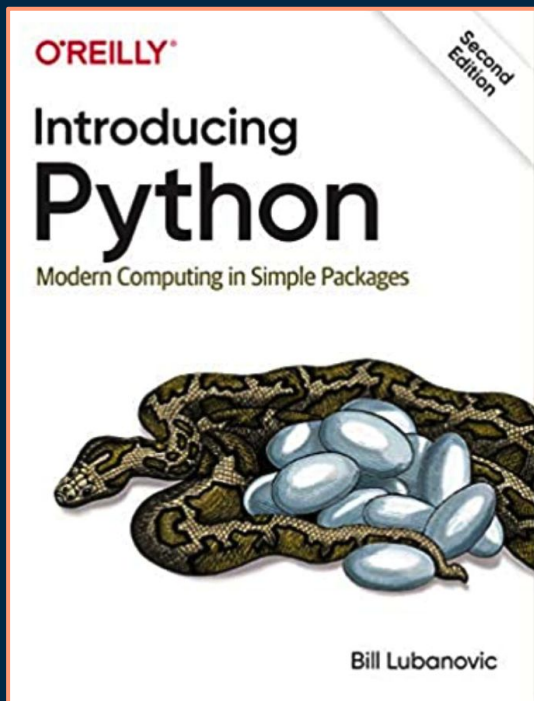
Haowen Jiang, Ph.D.

ABOUT ME

- B.A. in English
- M.A. & Ph.D. in Linguistics
- A muggle in programming until Dec 2019
- An AI engineer since Sep 2020

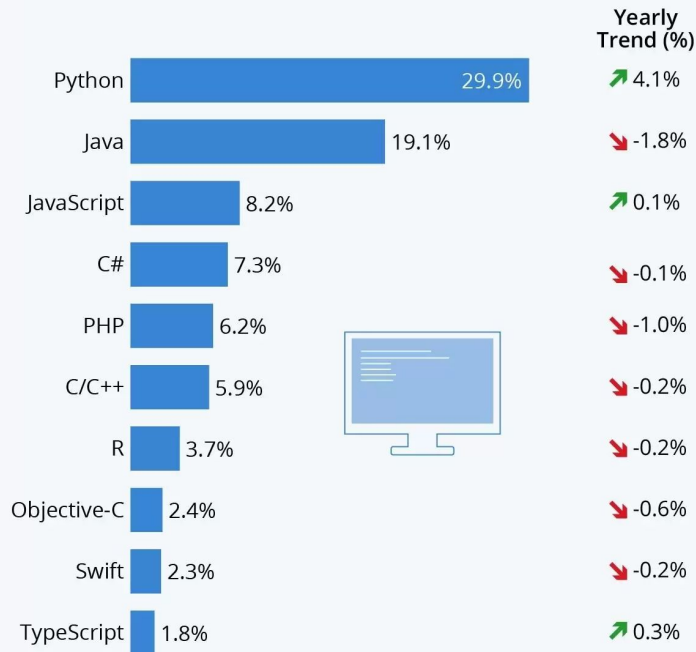


WHAT IS PYTHON



Python Remains Most Popular Programming Language

Popularity of each programming language based on share of tutorial searches in Google



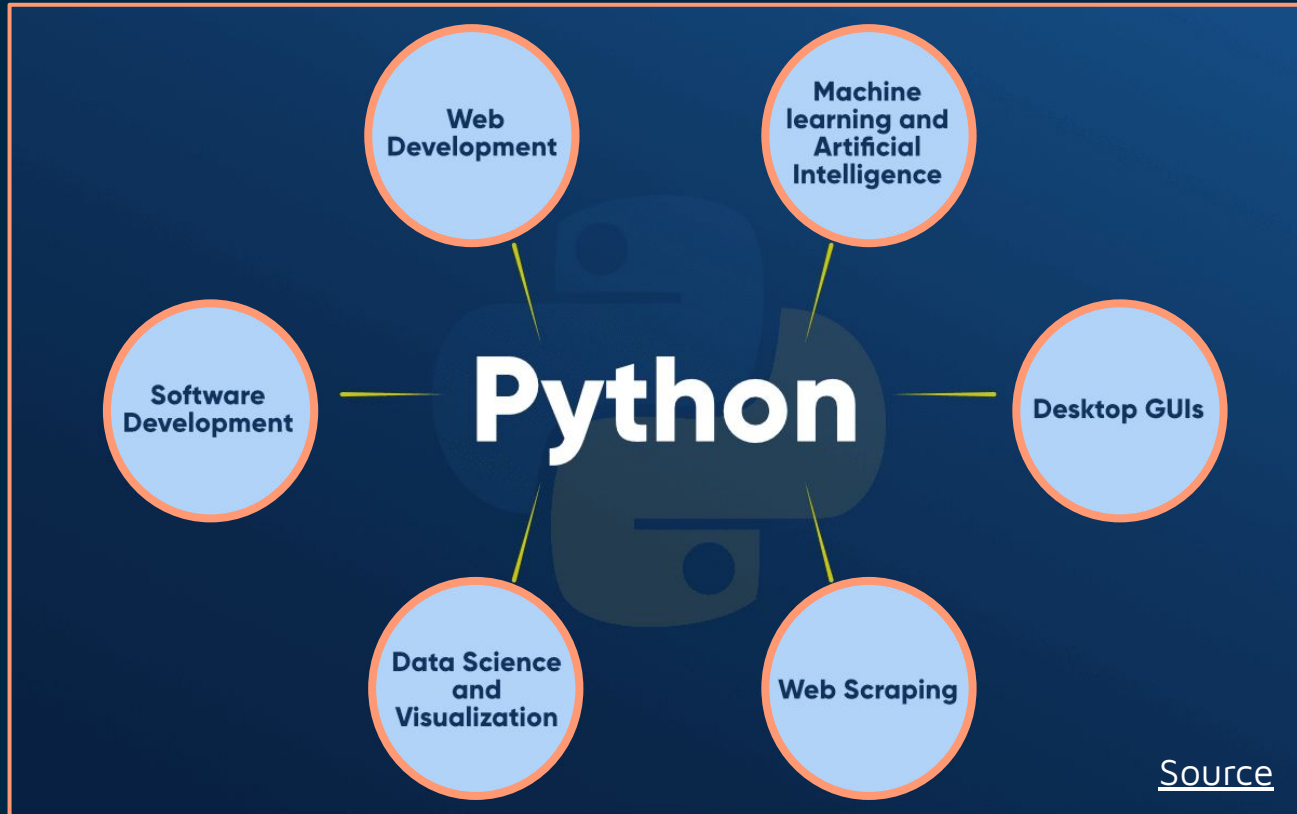
Yearly trend compares percent change from Feb 2019 to Feb 2020
Sources: GitHub, Google Trends



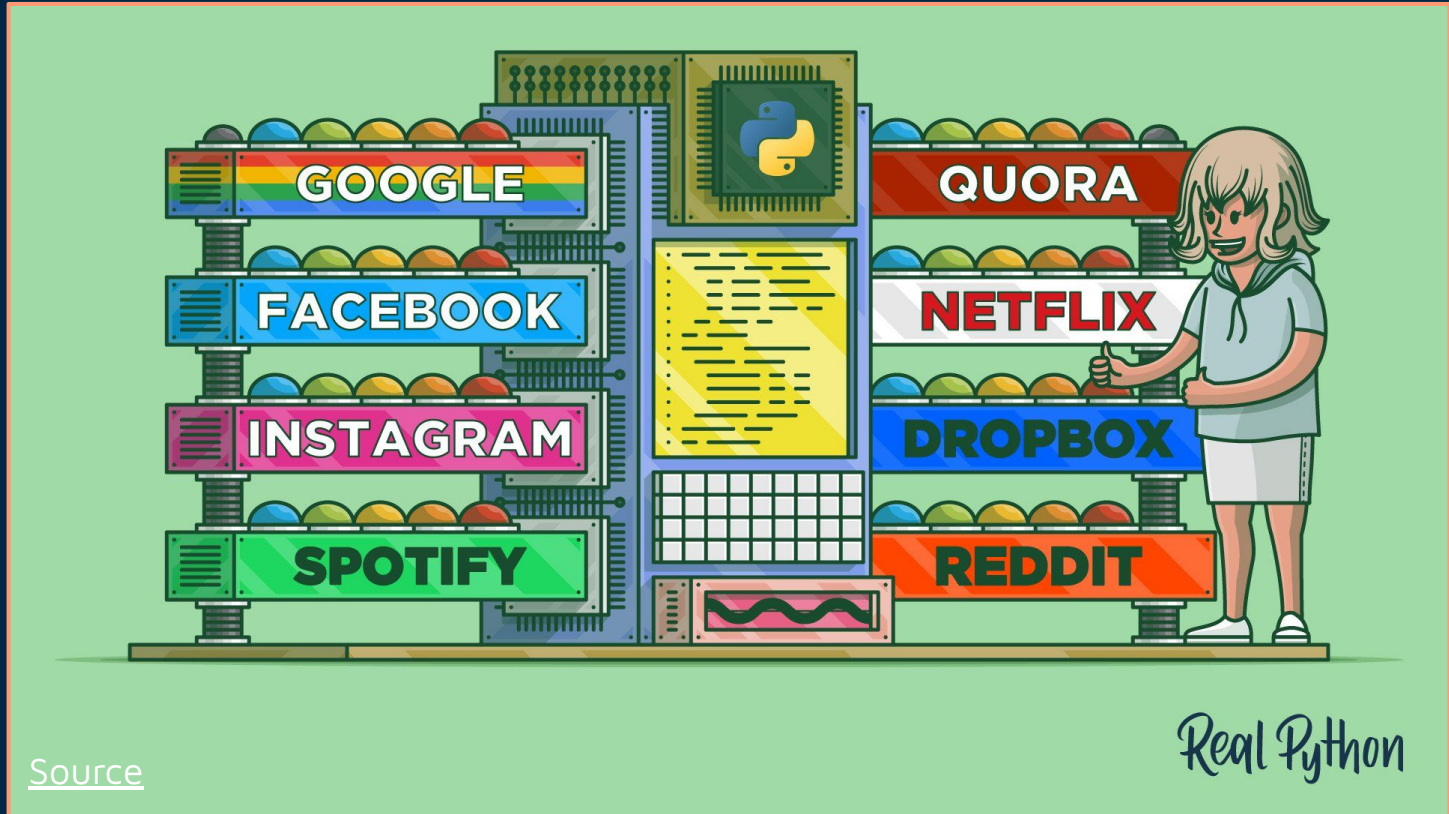
statista

Source

WHY PYTHON



BIG TECHS THAT USE PYTHON



DATA SCIENCE ROLES



Hadoop, NoSQL, Python

Big data,
Mathematics,
Programming

Excel, Tableau, SQL

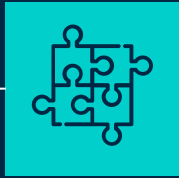
Statistics,
Communication,
Business Knowledge

SQL, Python, R

Communication,
Mathematics,
Programming

Source

ROADMAP



01

MY LEARNING PATH



02

LEARNING TIPS



03

Q&A

My Learning Path

01

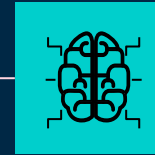
MY LEARNING PATH

01 Web Scraping



02 Text Mining

03 Machine Learning



04 Deep Learning

WEB SCRAPING

ezTravel 易遊網 找住宿 搜機票 瘋台灣 饗玩樂 玩世界 搭遊輪 企業服務 客製旅遊 追蹤 ▾ 查訂單 ▾ 登入

目的地或飯店名稱

高雄

入住日期

2021年05月15日 週六 ▾ 2 晚

退房日期

2021年05月17日 週一 ▾

房間數及人數

1間房, 2位成人, 0位孩童 ▾

搜尋

顯示地圖

共找到 252 家高雄飯店

自訂篩選

熱門篩選

☐ 免費取消

☐ 即時確認

☐ 免費WIFI

☐ 免費早餐

☐ 很棒: 評分4.7以上

星級

☐ 五星/豪華

☐ 四星/高級

☐ 三星/舒適


易遊推薦

價格 ▾

好評優先

飯店星級 ▾

距離遠近



高雄比歐緻居

Brio Hotel

★★★★

三多商圈/五福商圈 距離市中心 2公里

☑ 限時優惠

最低價房型僅剩2間

很棒 4.7


173則評鑑

TWD -3,510

TWD **2,984** 起

平均每晚含稅價

訂購



和逸飯店(高雄中山館)

Hotel Cozzi Zhongshan Kaohsiung

★★★★★ 動人夜景

三多商圈/五福商圈 距離市中心 3公里

年度暢銷飯店

很棒 4.7

2,778則評鑑

TWD **2,874** 起

平均每晚含稅價

訂購

WEB SCRAPING

```
1 max_df.head()
```

	ZH_Name	EN_Name	Hotel_ID	City	Location	Address	Campain	Mask	Hold_Room	Min_Price
0	高雄比歐 緻居	Brio Hotel	22260305	高雄	None	中山一路 14-26號	None	None	2	2584
1	和逸飯店 (高雄中 山館)	Hotel Cozzi Zhongshan Kaohsiung	2985883	高雄	None	中山二路 260號30 樓	None	None	19	2488
2	高雄喜達 絲飯店		2328586	高雄	None	成功一路 286號	None	None	3	3306
3	高雄國賓 大飯店	The Ambassador Hotel Kaohsiung	701744	高雄	None	民生二路 202號	None	None	9999	2097
4	高雄麗尊 酒店	The Lees Hotel	701698	高雄	None	五福一路 105號	None	None	10	2208
5	高雄水京 棧國際酒 店	H2O Hotel	8194467	高雄	None	明華路 366號	None	None	1	4701

TEXT MINING

Manufacturers

- Identify root causes of product issue quicker
- Identify trends in market segments
- Understand competitors products

Government

- Identify fraud
- Understand public sentiments about unmet needs
- Find emerging concerns that can shape policy

Financial Institutions

- Use contact center transcriptions
- Understand customers
- Identify money laundering or other fraudulent situation

Retail

- Identify profitable customers and understand the reasons for their loyalty
- Manage the brand on social media

Legal

- Identify topics and keywords in discovery documents
- Find patterns in defendant's communications

Healthcare

- Find similar patterns in doctor's reports
- Use social media to detect outbreaks earlier
- Identify patterns in patient claims data

Telecommunications

- Prevent customer churn
- Suggest up-sell/cross-sell opportunities by understanding customer comments

Life Sciences

- Identify adverse events in medicines or vaccines
- Recommend appropriate research materials

Insurance

- Identify fraudulent claims
- Track competitive intelligence
- Manage the brand on social media

zencos

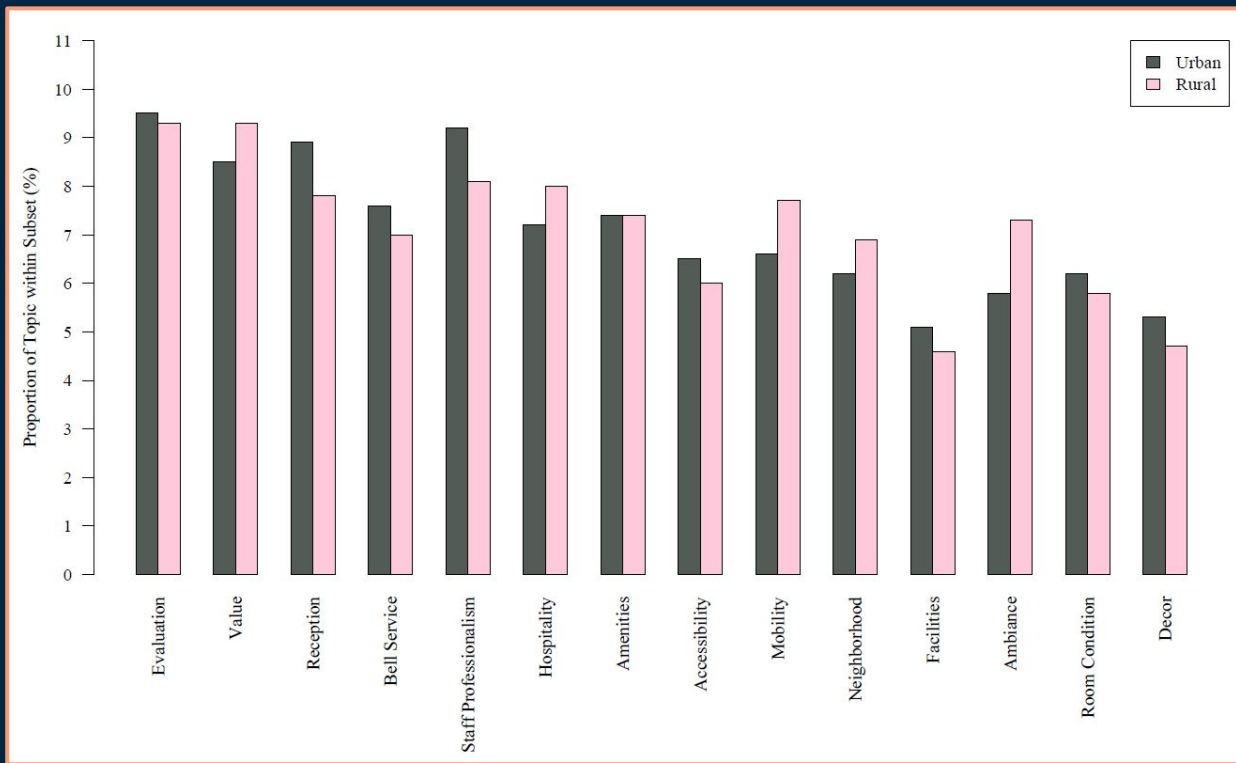
Source

TEXT MINING

Topic	Relevant Keywords
Evaluation	Stay, recommend, back, enjoy, pleasant, experience, happy, satisfy, wonderful, expectation, comfort, love, plan, future, regret, exceed, mention
Value	Location, price, money, facility, cheap, worth, reason, quality, afford, choice, budget, expensive, decent, paid, valuable, competitive, comparison
Reception	Check, book, day, desk, front, late, arrive, pay, quote, reception, morning, call, charge, wait, receptionist, key, card, leave, request, rude, lady, guy
Room Size	Room, small, big, luggage, size, space, floor, large, family, double, lift, person, single, upgrade, move, standard, bigger, carry, stairs, elevator
Staff Professionalism	Staff, friendly, helpful, service, speak, accommodate, kind, friend, polite, communicate, smile, spoke, professional, language, fluent, courteous, language
Hospitality	Hostel, guesthouse, guest, owner, host, kind, feel, house, home, travel, people, atmosphere, experience, love, warm, traditional, meet, beer, care, accommodate
Amenities	Breakfast, free, provide, coffee, water, towel, toiletries, include, available, kitchen, basic, laundry, machine, toast, drink, tea, buffet, amenity, shampoo, fridge, egg
Accessibility	Hotel, airport, bus, stop, convenient, taxi, shuttle, drive, terminal, pick, limousine, road, travel, transfer, international, overnight, express, layover, flight, hour
Mobility	Station, walk, subway, close, convenient, minute, location, distance, store, metro, train, exit, convenience, market, station, nearby, street, road, connect

Topic modeling based on 104,161 online reviews of Korean accommodations

TEXT MINING



Topic distribution by urbanization

ARTIFICIAL INTELLIGENCE

MACHINE LEARNING

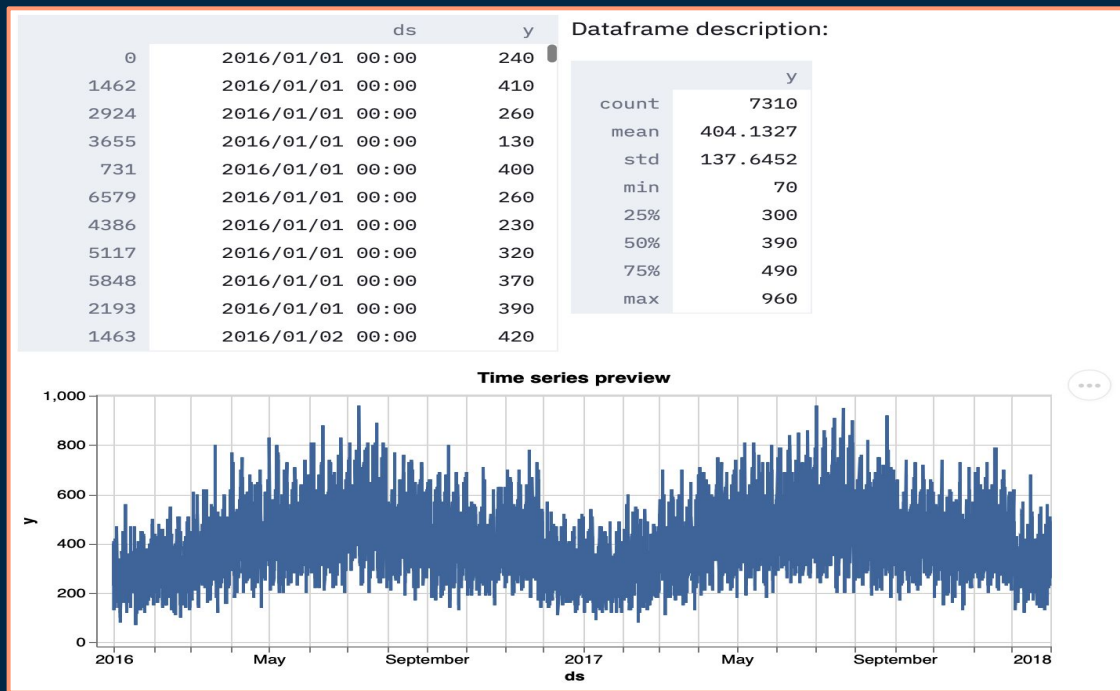
- Probability theory
- Linear Algebra

DEEP LEARNING

- Probability theory
- Linear Algebra
- Calculus

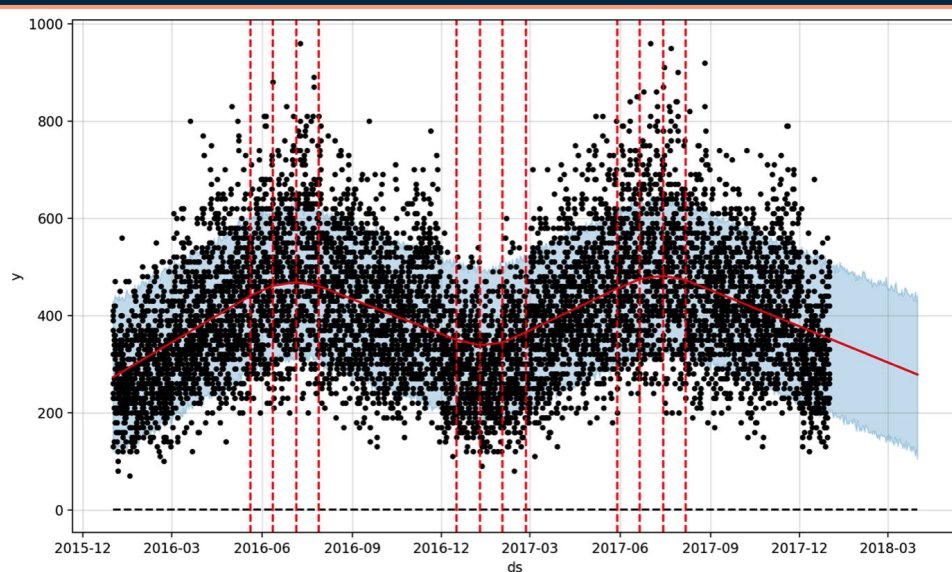


TIME-SERIES DATA

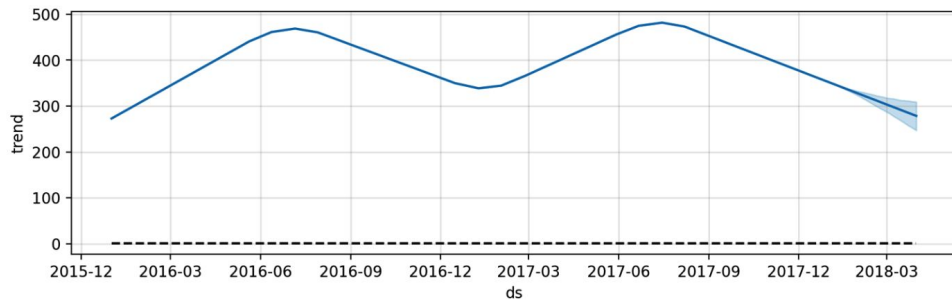


Data source

FORECAST



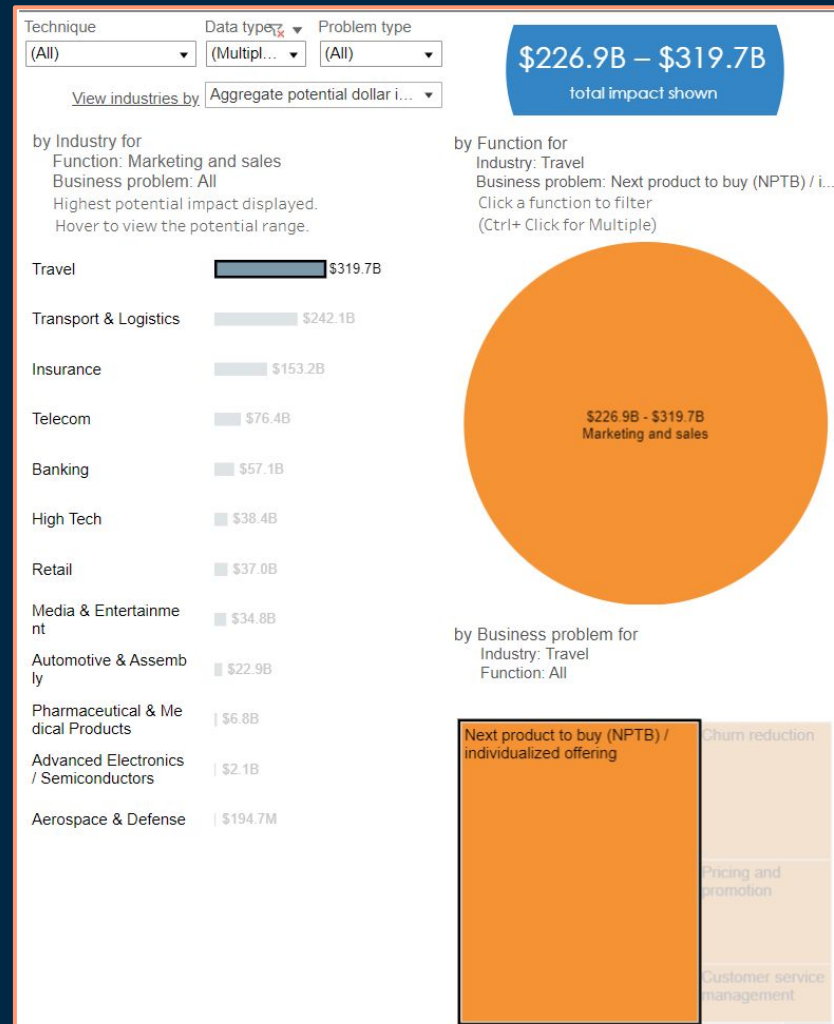
☒ Show components



Try it out [here](#)!

Estimated AI impact by industry (McKinsey)

- Travel
- Text+Time Series

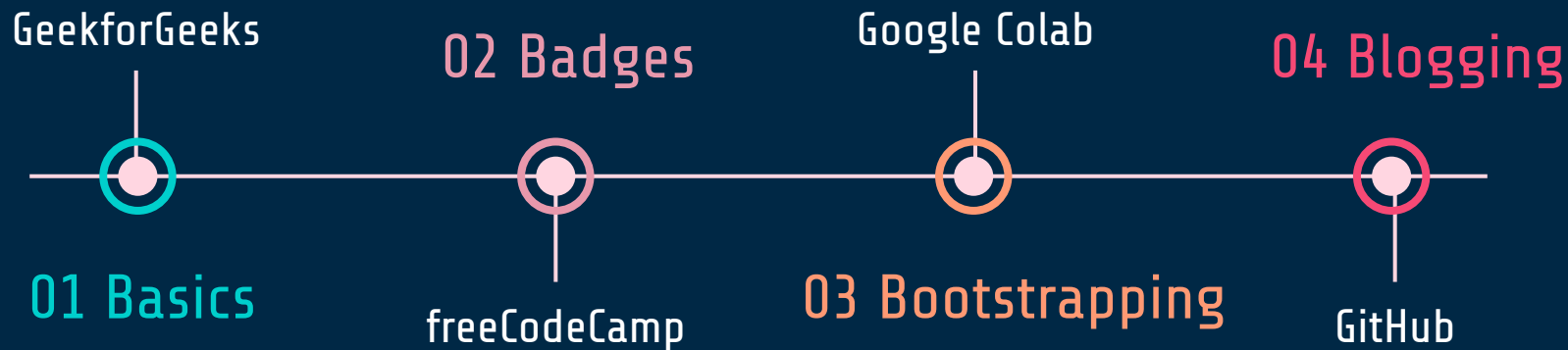


Source

Learning Tips

02

SUGGESTED PATH: 4Bs



Python tutorial on GeekforGeeks

```
# Python program showing
# indentation

site = 'gfg'

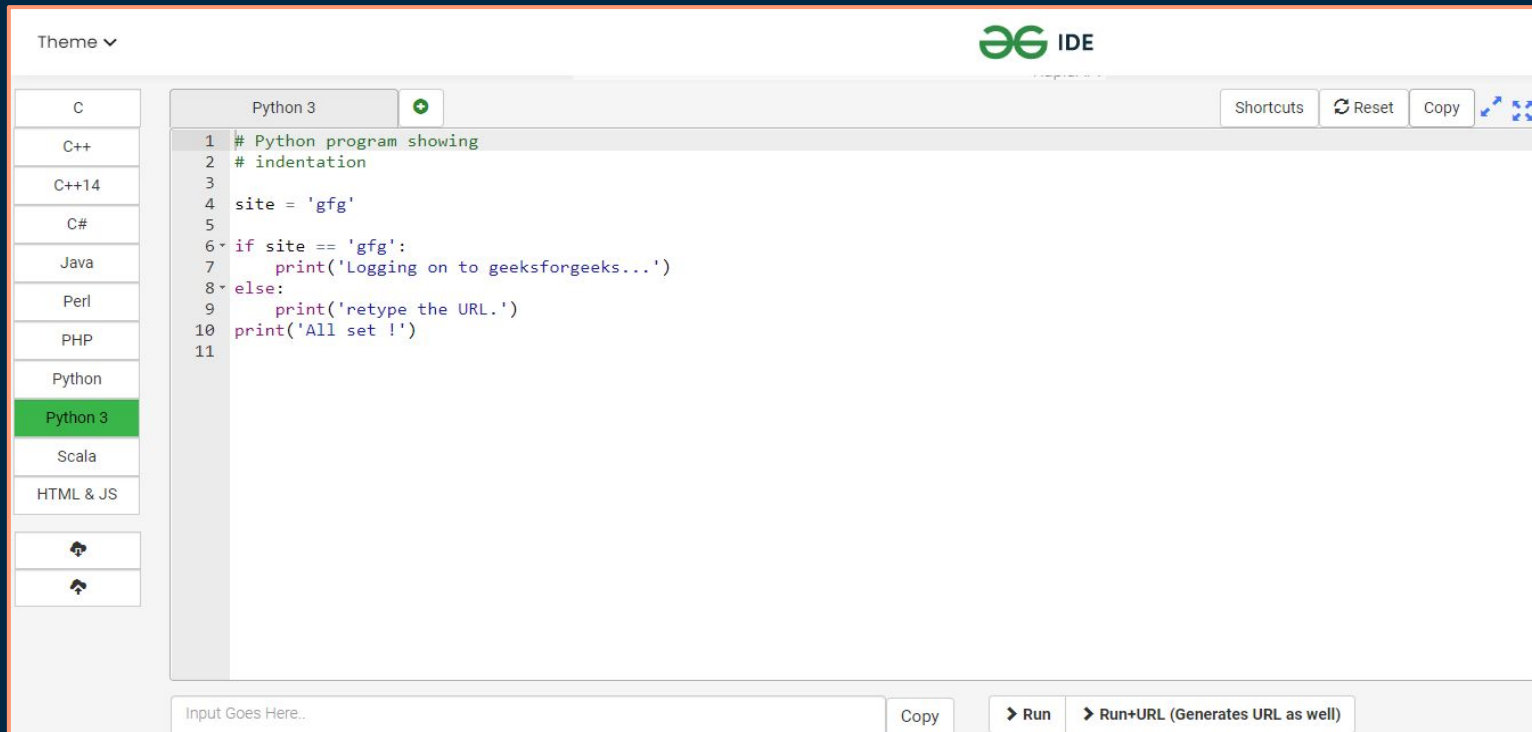
if site == 'gfg':
    print('Logging on to geeksforgeeks...')
else:
    print('retype the URL.')
print('All set !')
```

Run on IDE

Output:

```
Logging on to geeksforgeeks...
All set !
```

IDE on GeekforGeeks



Certification on freeCodeCamp

freeCodeCamp(🔥)	
	Data Visualization Certification (300 hours)
	APIs and Microservices Certification (300 hours)
	Quality Assurance Certification (300 hours)
	Scientific Computing with Python Certification (300 hours)
	Data Analysis with Python Certification (300 hours)
	Information Security Certification (300 hours)
	Machine Learning with Python Certification (300 hours)

YouTube videos from freeCodeCamp

freeCodeCamp (🔥) technical courses

(A) SUPPORT OUR NONPROFIT



freeCodeCamp.org ✓

3.47M subscribers

JOIN

SUBSCRIBE

HOME

VIDEOS

PLAYLISTS

COMMUNITY

CHANNELS

ABOUT



Popular Programming Courses ▶ PLAY ALL



Data Structures Easy to Advanced Course - Full...

freeCodeCamp.org ✓
2.7M views • 1 year ago



Learn Python - Full Course for Beginners [Tutorial]

freeCodeCamp.org ✓
23M views • 2 years ago

CC



Learn JavaScript - Full Course for Beginners

freeCodeCamp.org ✓
5.9M views • 2 years ago

CC



APIs for Beginners - How to use an API (Full Course / ...

freeCodeCamp.org ✓
1.5M views • 1 year ago

CC



Full Ethical Hacking Course - Network Penetration Testin...

freeCodeCamp.org ✓
3.9M views • 1 year ago



Learn Data Science Tutorial - Full Course for Beginners

freeCodeCamp.org ✓
1.2M views • 1 year ago

CC

Projects on freeCodeCamp

Data Analysis with Python Projects

There are many ways to analyze data with Python. By completing these projects, you will demonstrate that you have a good foundational knowledge of data analysis with Python.

Finish them all to claim your Data Analysis with Python certification.

Mean-Variance-Standard Deviation Calculator ☐

Demographic Data Analyzer ☐

Medical Data Visualizer ☐

Page View Time Series Visualizer ☐

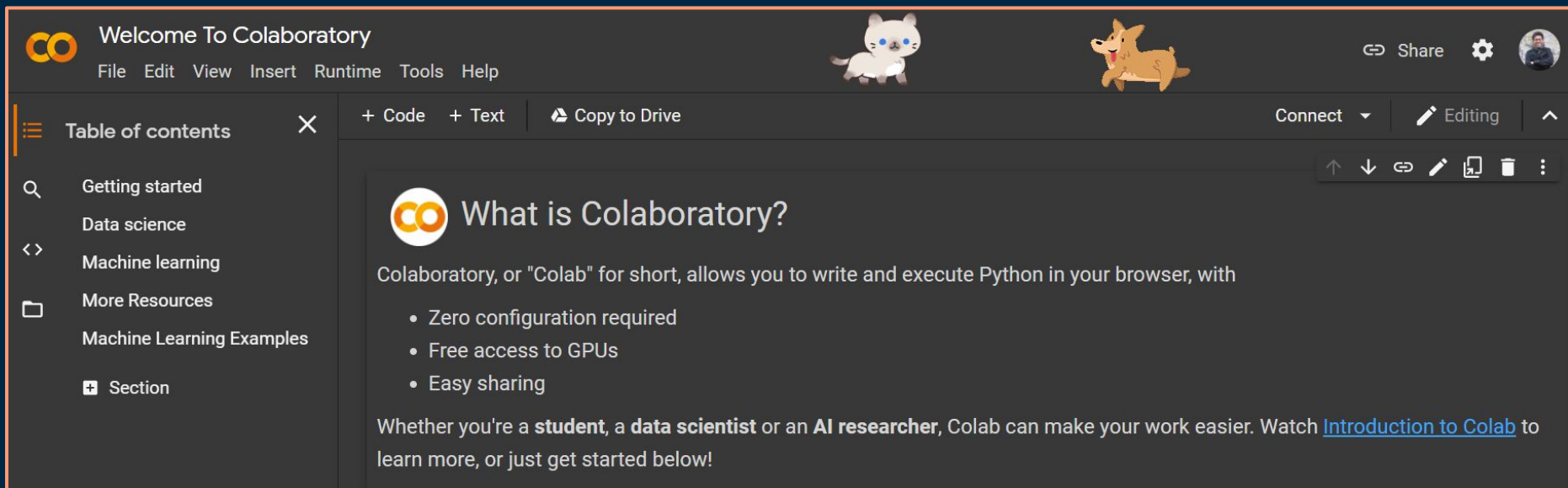
Sea Level Predictor ☐



Data Analysis with Python Certification



Intro to Colab

The image shows the Google Colaboratory web interface. At the top, there's a header with the Colab logo, the text "Welcome To Colaboratory", and a menu with "File", "Edit", "View", "Insert", "Runtime", "Tools", and "Help". On the right of the header, there are icons for a cat and a dog, a "Share" button with a link icon, a settings gear, and a user profile picture. Below the header is a toolbar with buttons for "+ Code", "+ Text", and "Copy to Drive". On the left side, there's a sidebar with a "Table of contents" section, a search icon, and a list of categories: "Getting started", "Data science", "Machine learning", "More Resources", and "Machine Learning Examples". The main content area displays a document titled "What is Colaboratory?". It features the Colab logo, a paragraph explaining that Colab allows writing and executing Python in the browser, and a bulleted list of features: "Zero configuration required", "Free access to GPUs", and "Easy sharing". At the bottom of the main area, there's a paragraph encouraging students, data scientists, and AI researchers to use Colab, with a link to "Introduction to Colab".

co Welcome To Colaboratory

File Edit View Insert Runtime Tools Help

Share ⚙️ 👤

+ Code + Text 📄 Copy to Drive

Connect ▾ Editing ↗

↑ ↓ 🔗 ✎ 📄 🗑️ ⋮

co What is Colaboratory?

Colaboratory, or "Colab" for short, allows you to write and execute Python in your browser, with

- Zero configuration required
- Free access to GPUs
- Easy sharing

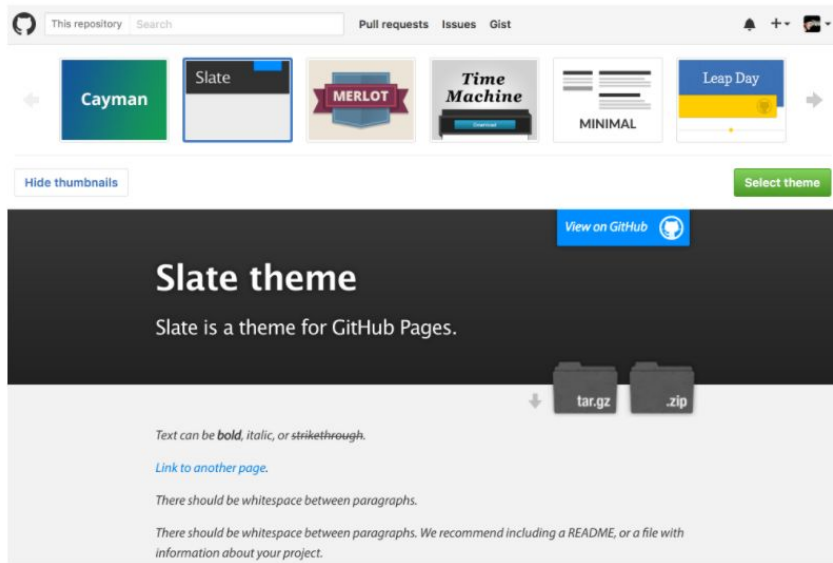
Whether you're a **student**, a **data scientist** or an **AI researcher**, Colab can make your work easier. Watch [Introduction to Colab](#) to learn more, or just get started below!

Blogging with GitHub Pages

3

Pick a theme

Choose one of the themes from the carousel at the top.
When you're done, click **Select theme** on the right.



The screenshot shows the GitHub Pages theme selection interface. At the top, there's a navigation bar with "This repository", a search bar, and links for "Pull requests", "Issues", and "Gist". Below this is a carousel of theme thumbnails: "Cayman", "Slate", "MERLOT", "Time Machine", "MINIMAL", and "Leap Day". A "Hide thumbnails" button is on the left, and a "Select theme" button is on the right. The "Slate" theme is selected and expanded below the carousel. It features a dark header with the text "Slate theme" and "Slate is a theme for GitHub Pages." Below the header, there are download buttons for "tar.gz" and ".zip". The main content area of the theme preview shows examples of text formatting: bold, italic, strikethrough, a link, and paragraphs with varying amounts of whitespace.

This repository Search Pull requests Issues Gist

Cayman Slate MERLOT Time Machine MINIMAL Leap Day

Hide thumbnails Select theme

View on GitHub

Slate theme

Slate is a theme for GitHub Pages.

tar.gz .zip

Text can be **bold**, *italic*, or ~~strikethrough~~.

[Link to another page.](#)

There should be whitespace between paragraphs.

There should be whitespace between paragraphs. We recommend including a README, or a file with information about your project.

My blog: howard-haowen.github.io/blog.ai

Posts



My First Post

For almost everything you'd like to learn, there's an awesome list for that on GitHub. On this blog, I plan to add my personal touch to various tools, documenting not only what I did right to get the ball rolling, but also what I did wrong to save you (or even future me) from abysmal frustration.

Jan 17, 2020



Classifying customer reviews with spaCy v3.0

This post shows how to train models that can automatically identify the sentiment and product category of customer reviews. With the help of the recently released spaCy v3.0, text classification is an absolute breeze.

Mar 12, 2021

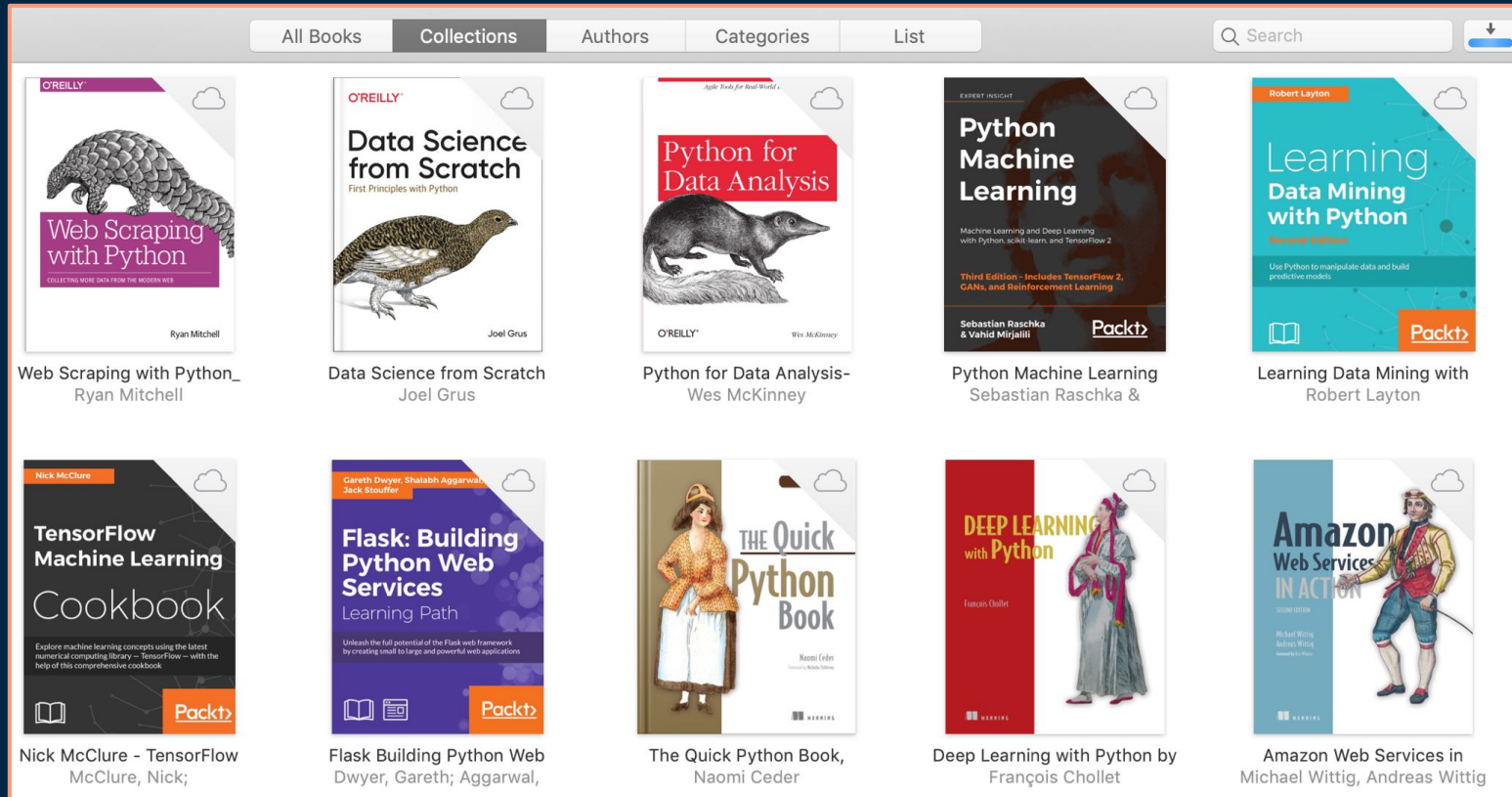


Adding a custom tokenizer to spaCy and extracting keywords

This post shows how to plug in a custom tokenizer to spaCy and gets decent results for the extraction of keywords from texts in traditional Chinese.

Feb 16, 2021

My collection on Apple Books



Q&A

03



```
>>> import this
```

“Now is better than never.
Although never is often
better than right now.”

—Zen of Python

howard.haowen@gmail.com

THANKS



CREDITS: This presentation template was created by [Slidesgo](#),
including icons by [Flaticon](#), and infographics & images by [Freepik](#)
Please keep this slide for attribution