Course Information and Overview

STAT3009 Recommender Systems

by Ben Dai (CUHK-STAT)
on September 4, 2025

» Course Overview and Key Features

- **Contemporary** and **Emerging** Field
  - Recommender systems emerged from industry needs around 2000
  - This course covers cutting-edge methods from **2010 to present**
  - Focus on modern deep learning and machine learning approaches
- **Interdisciplinary** Approach with **Hands-on** Learning
  - **Theory**: Statistics + Machine Learning + Mathematical Optimization
  - **Practice**: Python Programming + Real-world Applications
  - **Delivery**: Interactive lectures + Live Kaggle competitions + Jupyter notebooks

» Course Overview and Key Features

- **Highly Relevant** to Industry Applications
  - **E-commerce**: Amazon, Alibaba product recommendations
  - **Entertainment**: Netflix, Spotify content suggestions
  - **Social Media**: Facebook, LinkedIn connection recommendations
  - **Finance**: Investment portfolio recommendations
- **Distinctive** from Traditional Statistics Courses
  - Emphasis on computational implementation and programming skills
  - Strong mathematical foundation required but practical coding equally important
  - Real-time problem-solving with large-scale datasets

» Course Information and Prerequisites

- **Course Position** in Curriculum
  - STAT3009 is an **elective course** for Statistics majors
  - Also open to students from related quantitative disciplines
  - Bridges theoretical statistics with practical machine learning applications

» Course Information and Prerequisites

- **Course Position** in Curriculum
  - STAT3009 is an **elective course** for Statistics majors
  - Also open to students from related quantitative disciplines
  - Bridges theoretical statistics with practical machine learning applications
- **Prerequisites** and Required Background
  - **Statistical Methods**: Linear regression, ridge regression, basic hypothesis testing
  - **Programming Skills**: Python (NumPy, Pandas), basic scikit-learn usage
  - **Mathematics**: Linear algebra, calculus, probability theory
  - **Note**: Comprehensive tutorials provided for Python libraries

» Class Composition

* **Class Composition** and Diversity
  * **Enrollment**: 49/60 students (as of September 1)
  * **Backgrounds**: STAT, Risk Management, Quantitative Finance, Natural Sciences, CS, Math, ...

Hybrid interdisciplinary + teaching mode

Mix Statistics + Machine Learning + Python + Mathematics

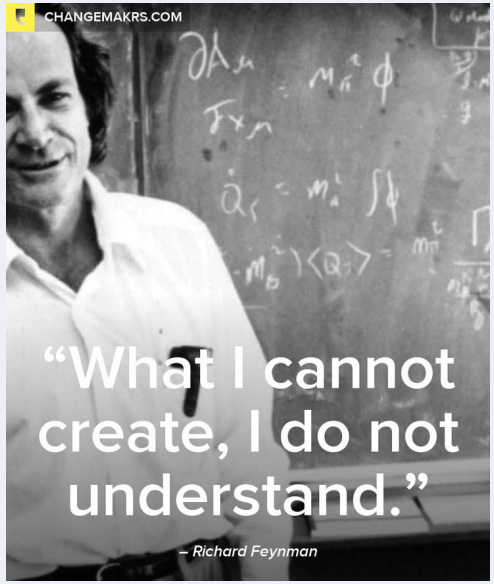
Teaching mode
  * lecture slides/note
  * real time Python programming in colab (Jupyter notebook)
  * InClass practice
  * real time competitions in Kaggle

* **Additional Resources**
  * Course Homepage with lecture materials
  * GitHub Repository with code examples

» Grading policy

- Homework (15%)
- Inclass Kaggle Competition (Open-book InClass Kaggle Competition) (50%)
- Final InClass Quiz (coding and exercise) (35%): Basic Python programming and implementation of recommender systems models (during the final lecture of the semester)

» My suggestion



- practice by yourself
- enjoy debugging: most coding experiences are from that
- independent: 99.9% bugs can be solved by Google
- bad example. Just look at the code, but never practice/implement on your own

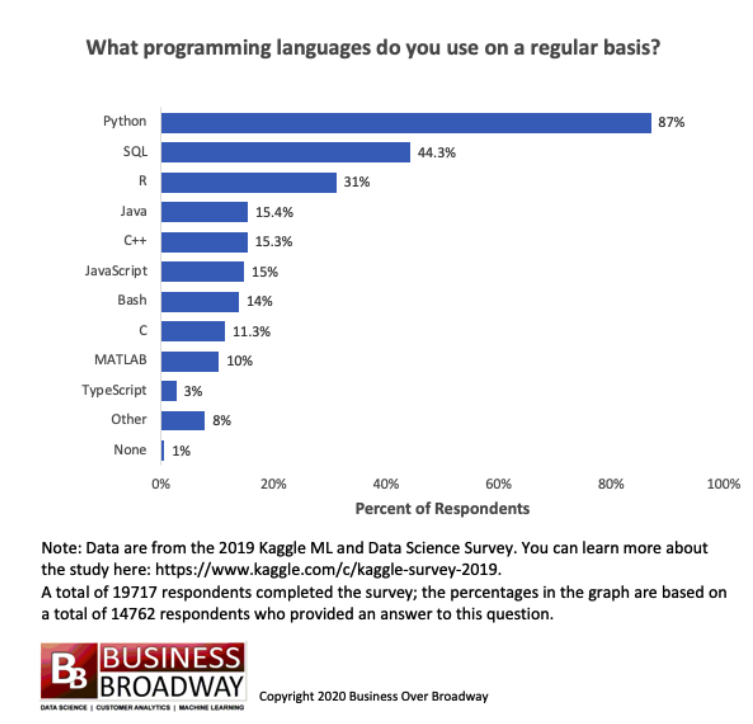
→ You must have a laptop

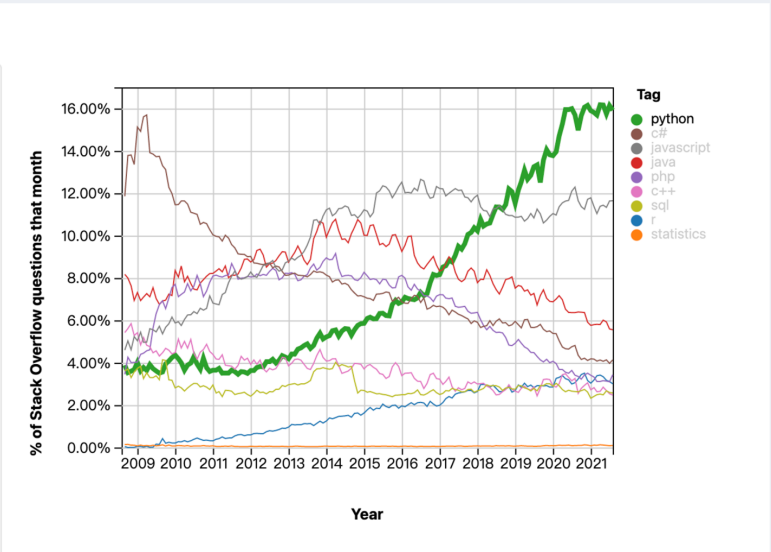
» Preparation

Toolbox we will use:

Python Python is the programming language of choice for data scientists.

* State of Data Science 2021: Popularity of Python



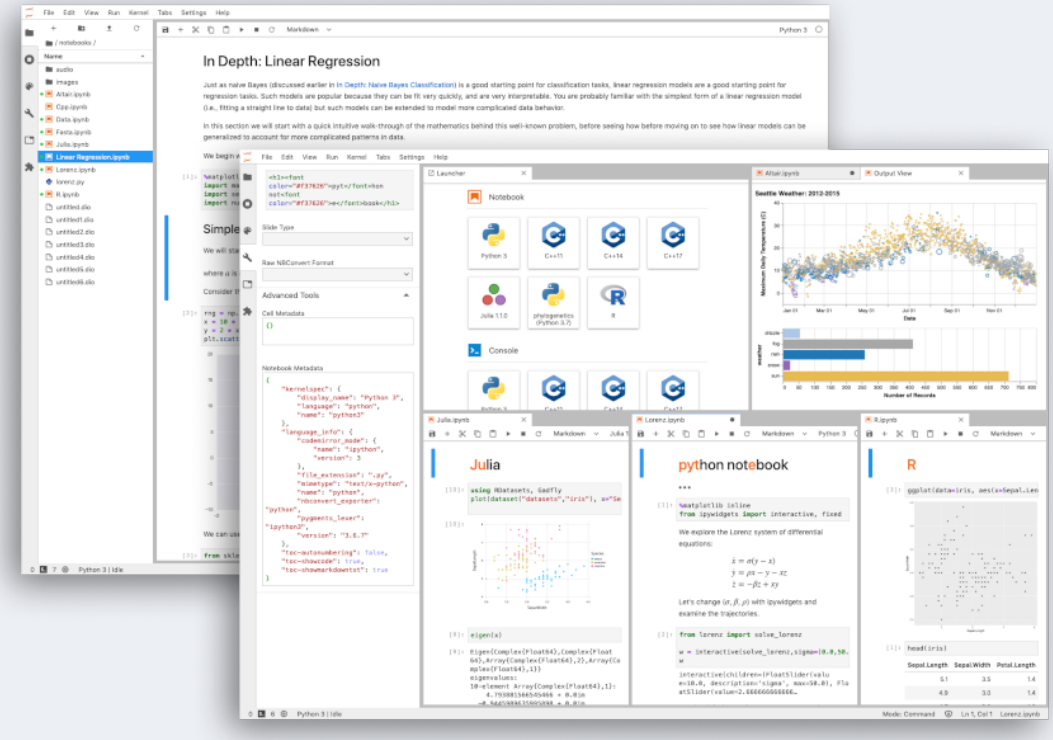


Source: https://businessoverbroadway.com

Source: https://www.alibabacloud.com/blog

Toolbox we will use:

NB Jupyter notebook - create and share documents that contain live code, equations, visualizations and narrative text



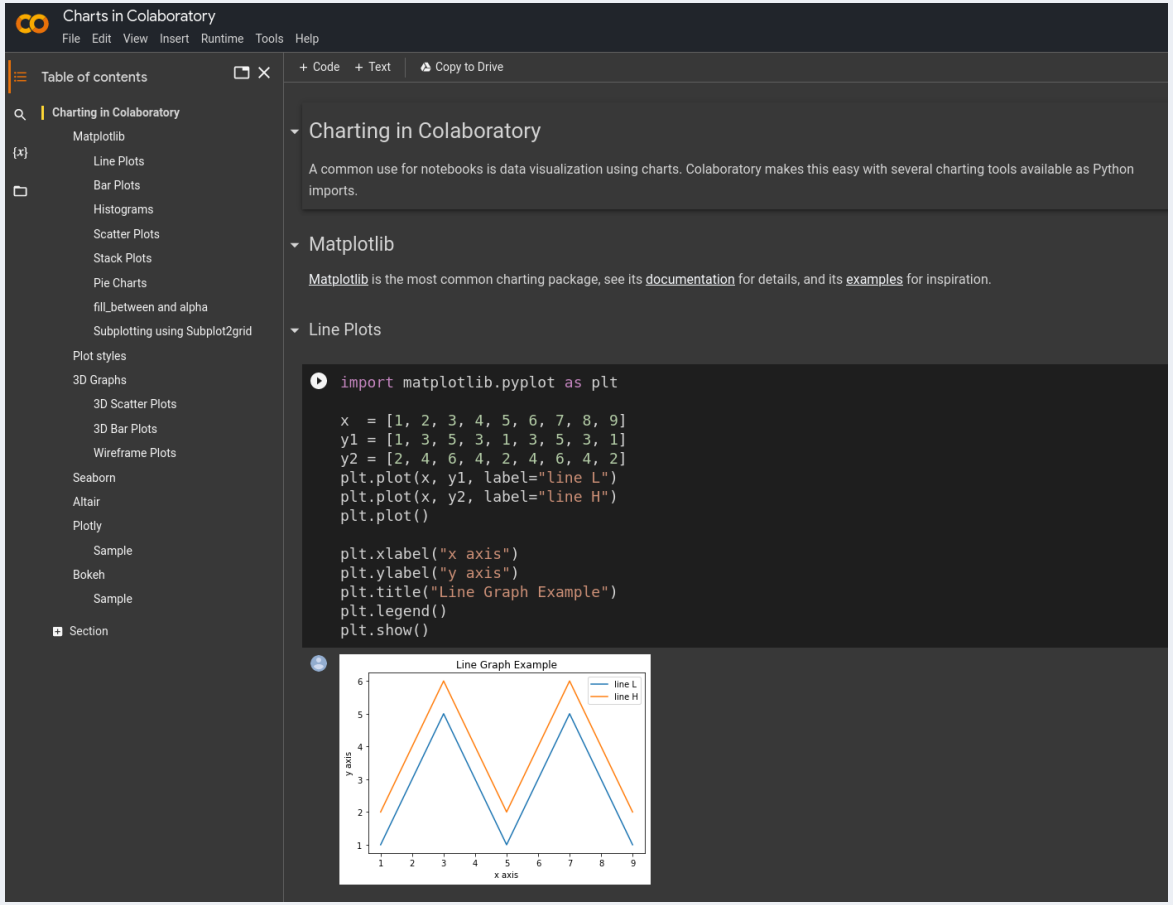
Source: https://jupyter.org/

» Preparation

Toolbox we will use:

Colab Register - Colaboratory (Google account)

* Online Python computing platform
* Python Jupyter notebook



Toolbox we will use:

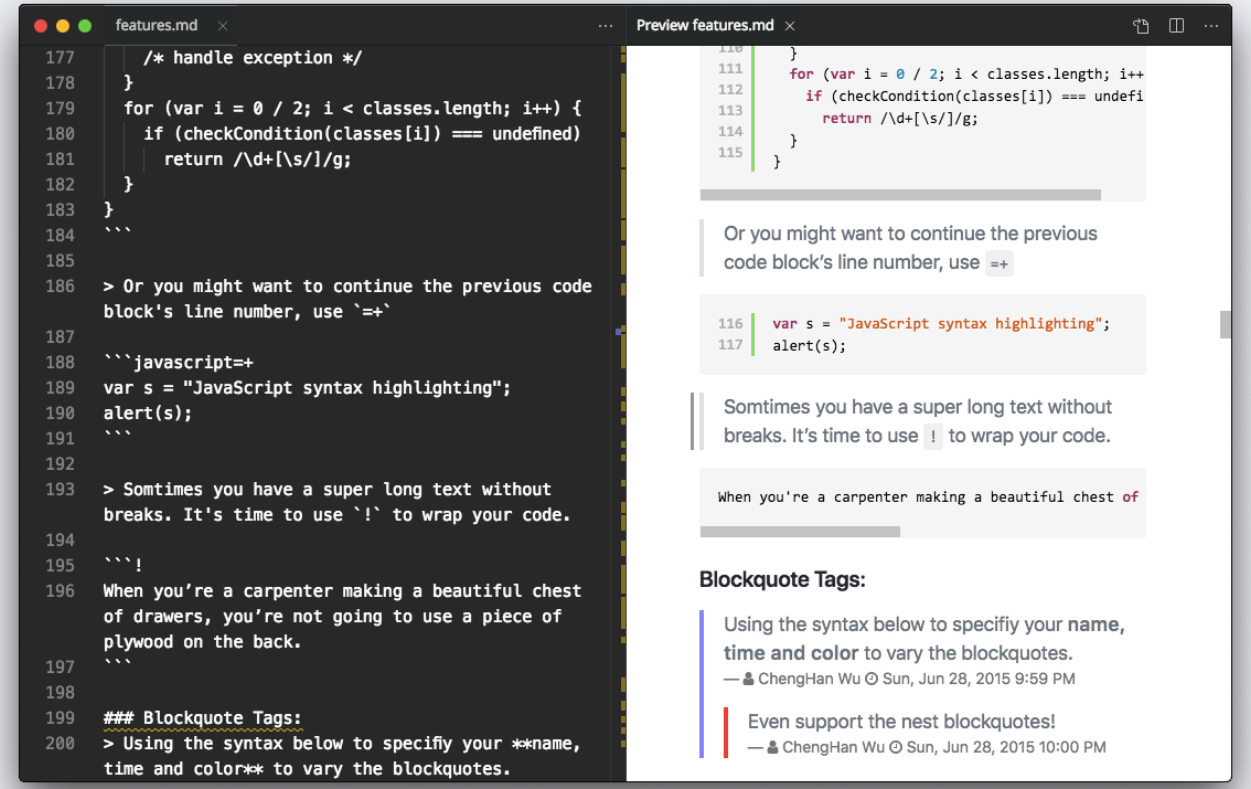
Colab

- Go to `colab.research.google.com` and sign in with your Google account.
- Create a new notebook by clicking on the "New Notebook" button.
- Write Python code in the cells of the notebook, using the same syntax and semantics as you would in a local Python environment.
- Execute the code by clicking on the "Run" button or pressing Shift+Enter.
- See the output of your code in the notebook, including any plots, tables, or text output.

» Preparation

Toolbox we will use:

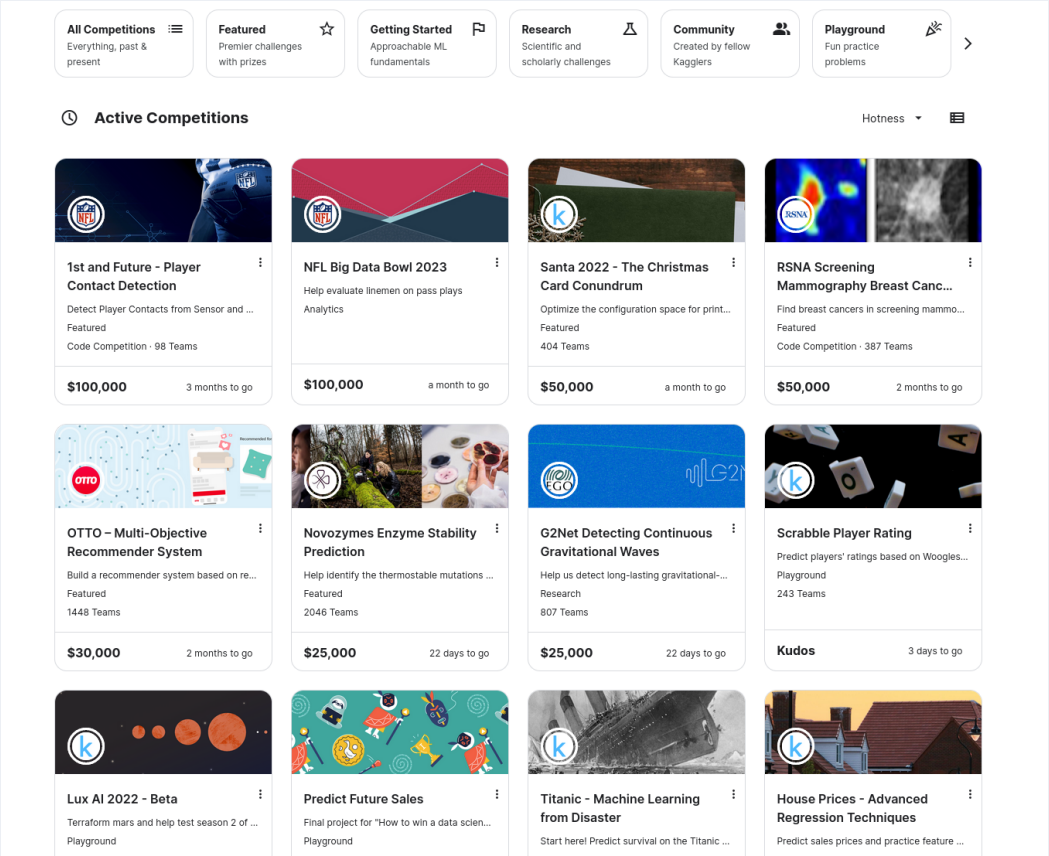
Colab Online Markdown documentation



» Preparation

Toolbox we will use:

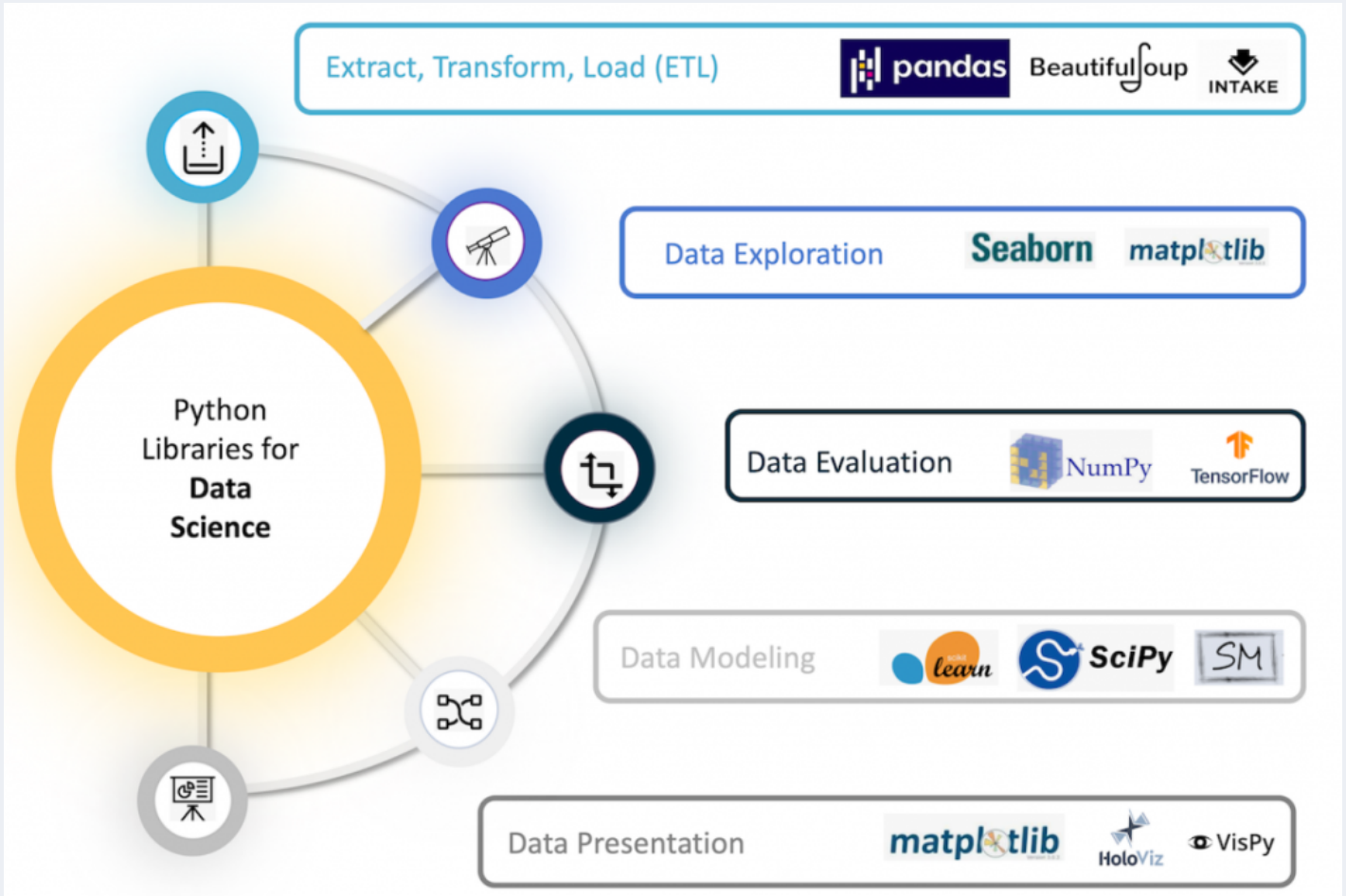
Kaggle - online community of data scientists and machine learning practitioners



Source: https://kaggle.com/

Toolbox we will use:

Libs Python libraries for Data Science



» Preparation

- Python package installation
  - Basic packages: `numpy` + `pandas` + `seaborn` + `scipy` + `scikit-learn`
  - Install packages in Jupyter notebook / colab

install python package

```
!pip install <package name>
```