

Python инструменты

Консоль/REPL

- Стандартный REPL
- bpython
- ptpython

```
>>> print ('Hello world!')
Hello world!

>>> import numpy as np

>>> import pandas as pd

>>> import sklearn
\
>>>
>>> █
```

[F4] Emacs 2898/2898 [F3] History [F6] Paste mode [F2] Menu - CPython 3.4.3

Текстовый редактор/IDE

- Sublime Text/Atom
- PyCharm

Jupyter/IPython notebook

- Редактор кода с автоподсказкой
- Код можно чередовать с текстом в Markdown
- Можно делать отчеты с кодом и сохранять в pdf
- Работает из браузера
- Работает с удаленной машины

PANDAS



Pandas - что за зверь?

Модуль к python, которым при некоторой ловкости можно заменить один Excel.

Pandas - что за зверь?

Модуль к python, которым при некоторой ловкости можно заменить один Excel. И дюжину клерков.

Pandas - что за зверь?

- Dataframe = Лист из книги в Экселе
- Индексы как в базах данных
- Индексированные последовательности
- Временные индексы
- Сводные таблицы
- Запросы с механизмом работы как в SQL
- Простейшие функции статистики
- Более предсказуемое поведение
- Нет графиков
- Для окончательной победы над Excel нужны другие модули

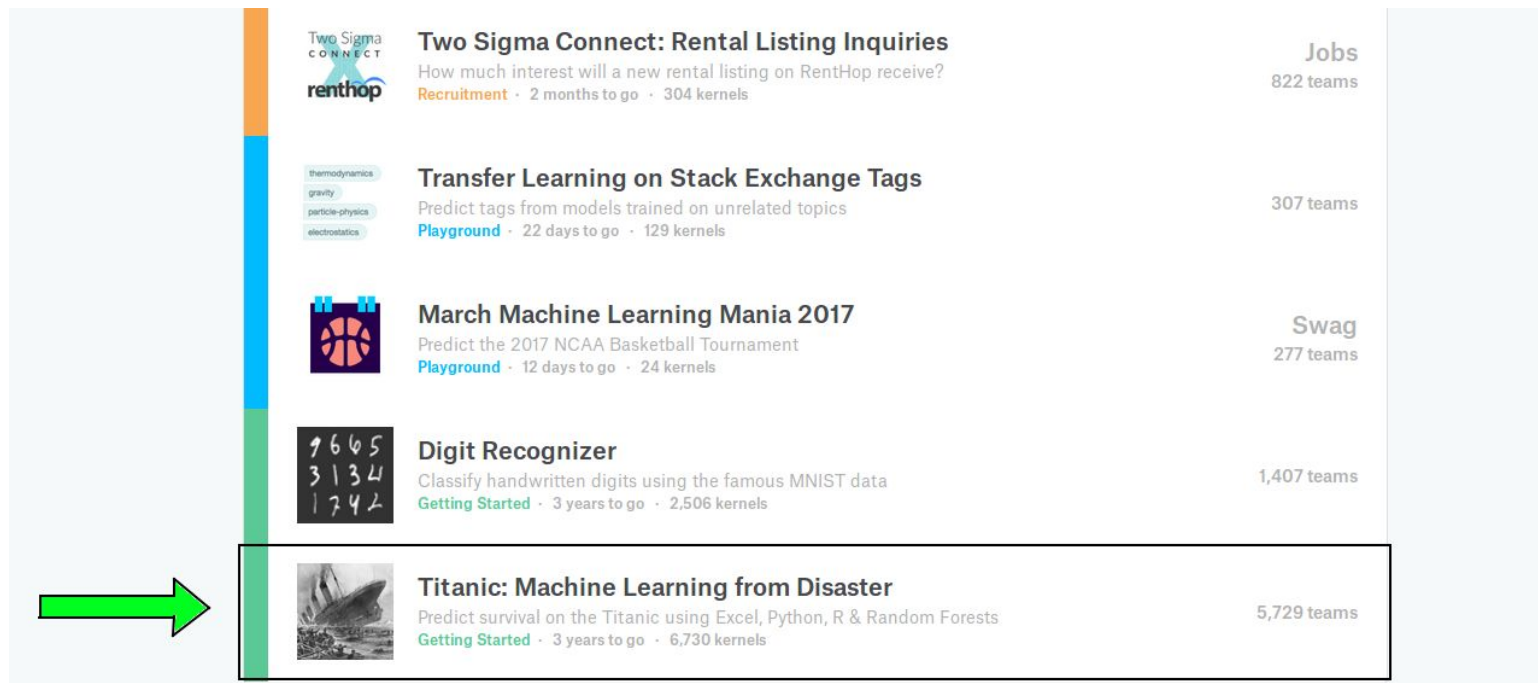
Pandas






Дальнейшее знакомство - в jupyter notebook

Проект!

Титаник

www.kaggle.com/c/titanic



	Two Sigma Connect: Rental Listing Inquiries How much interest will a new rental listing on RentHop receive? Recruitment · 2 months to go · 304 kernels	Jobs 822 teams
	Transfer Learning on Stack Exchange Tags Predict tags from models trained on unrelated topics Playground · 22 days to go · 129 kernels	307 teams
	March Machine Learning Mania 2017 Predict the 2017 NCAA Basketball Tournament Playground · 12 days to go · 24 kernels	Swag 277 teams
	Digit Recognizer Classify handwritten digits using the famous MNIST data Getting Started · 3 years to go · 2,506 kernels	1,407 teams
	Titanic: Machine Learning from Disaster Predict survival on the Titanic using Excel, Python, R & Random Forests Getting Started · 3 years to go · 6,730 kernels	5,729 teams

Титаник

Есть список пассажиров и точно известна их судьба.

Train.csv - часть списка пассажиров, для которых указано, выжили они или нет.

Test.csv - другая часть списка пассажиров - для них такой информации не дано.

Цель - построить модель, которая будет отражать принцип “женщины и дети вперед”