

GeekBrains

# Python для Data Science

Урок 1. Введение в курс.

# Немного о себе



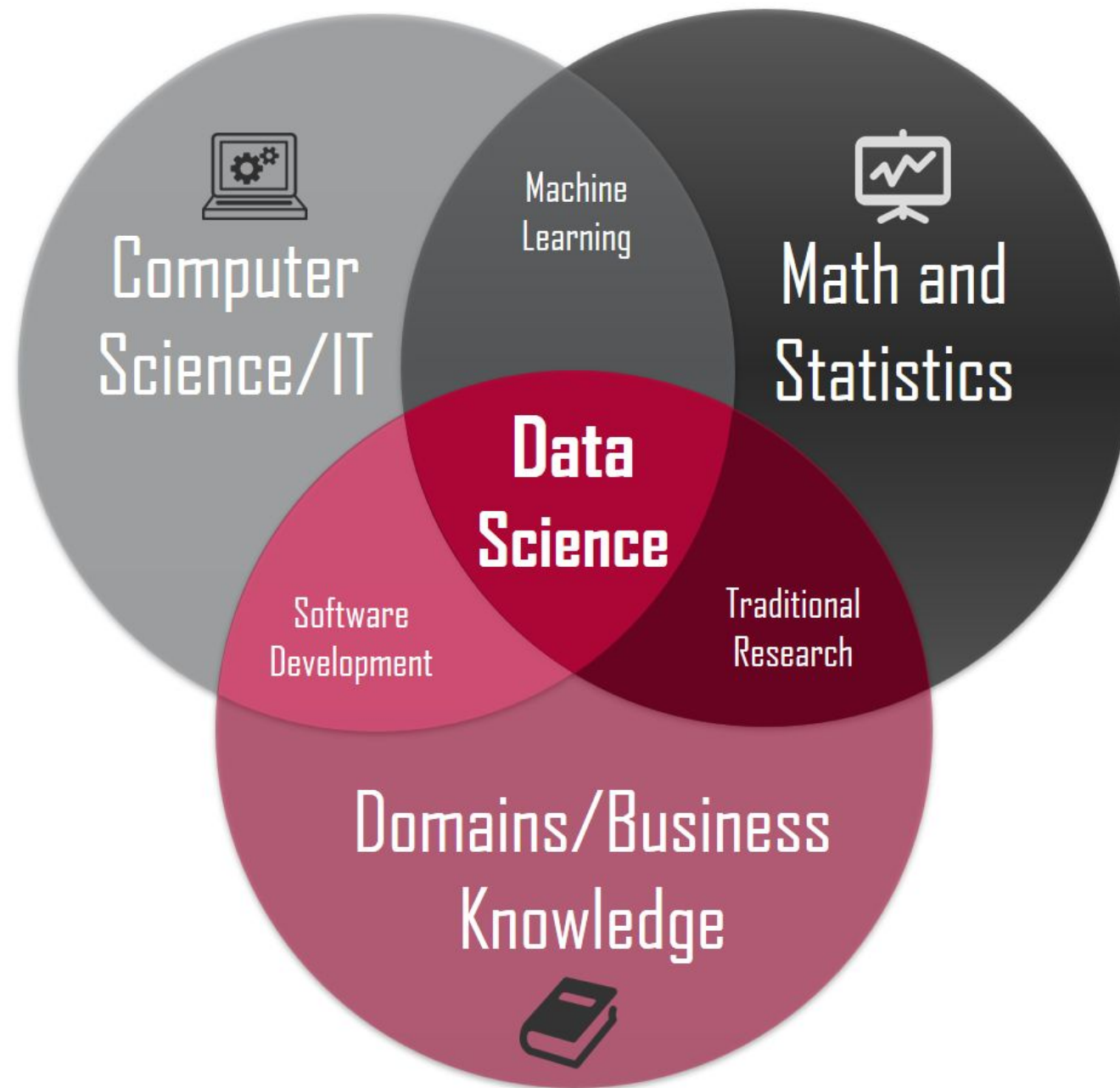
NAPOLEON IT

- ❖ Система распознавания товаров и ценников по фотографии
- ❖ Распознавание людей на фотографии
- ❖ Чат-бот службы поддержки
- ❖ Система расчета рыночной цены объектов недвижимости

# План вебинара

1. Что такое Data Science?
2. Цели данного курса
3. Требования
4. Что мы изучим?
5. Учебный процесс
6. Ответы на вопросы

# Что такое Data Science?



# Профессии и вакансии

- ❖ Data scientist (математик-программист, исследователь данных)
- ❖ Data analyst (аналитик данных)
- ❖ Data engineer (специалист по обработке данных)
- ❖ Machine learning engineer (специалист по машинному обучению)
- ❖ Deep learning engineer NLP (обработка естественного языка)
- ❖ Deep learning engineer CV (компьютерное зрение)

# Senior Data Scientist в Контур.Фокус

з/п не указана

СКБ Контур ✓★

Санкт-Петербург

Откликнуться

Показать контакты



Требуемый опыт работы: 3–6 лет

Полная занятость, гибкий график



## Вы сможете:

- Получать опыт и прокачиваться в команде DS-специалистов. Команда Фокуса тесно взаимодействует с остальными Data Scientist'ами компании. Мы стремимся к воспроизводимости исследований и обмену знаниями.
- Поработать с большими данными, а в нашем случае это более 10 млн организаций, сайтов, новостей СМИ, арбитражных дел и документов из других официальных источников.
- Ездить на IT-конференции, meetup'ы, учебные курсы за счет компании, в том числе спикером.
- Выбрать удобную для работы технику: MacBook, или HP, или Lenovo.

**Технологии DS**, которыми пользуется команда: Python, Pandas, Numpy, Sklearn, Keras, PyTorch.



# Data Scientist в команду IRM General

з/п не указана

Raiffeisenbank ✓

● Технопарк, Москва, проспект Андропова, 18к2



Откликнуться



Требуемый опыт работы: 1–3 года

Полная занятость, полный день

**Эта вакансия для тебя, если ты:**

- Имеешь опыт программирования на Python;
- Знаешь основные распределения из теории вероятностей;
- Понимаешь, что такое A/B-тестирование и зачем его проводить;
- Знаешь базовые библиотеки работы с данными и машинным обучением: `numpy`, `pandas`, `scikit-learn`, `matplotlib`;
- В Python используешь разделение кода на классы и модули, оформляешь код по PEP8, используешь генераторы, коллекции и функциональное программирование;
- Понимаешь, что такое проверка статистических гипотез, критерии и тесты;
- Знаешь про разные модели машинного обучения, чем они отличаются, их плюсы и минусы;
- Знаешь, что такое и зачем нужна кросс-валидация;
- Используешь готовые трансформеры;
- Знаешь, какие метрики для чего используются;
- Можешь подготовить датасет на Spark;
- Используешь git;
- Умеешь самостоятельно работать и презентовать результаты работы;

# Data Scientist (игровое направление)

з/п не указана

Mail.Ru Group, Игры ✓ ☆

● Аэропорт, Москва, Ленинградский проспект, 39с79



Откликнуться



Требуемый опыт работы: 1–3 года

Полная занятость, полный день

## Обязанности:

- разработка и поддержка платформы для автоматической закупки трафика;
- анализ существующих алгоритмов для закупки, поиск точек роста;
- работа с большим объемом неструктурированных данных;
- взаимодействие с разработчиками и дата-инженерами;
- доработка текущих моделей прогноза LTV.

## Требования:

- опыт работы на позиции Data Scientist — от 2 лет;
- знание алгоритмов машинного обучения;
- базовые знания в области теории вероятностей и математической статистики, байесовских методов;
- Python (pandas, numpy, sklearn, scipy, airflow);
- Linux, Bash;
- SQL на уровне написания и оптимизации сложных запросов.



# Цели курса

- ❖ Сформировать начальное представление о сфере Data Science
- ❖ Познакомиться с основными библиотеками для DS и ML
- ❖ Пройтись по ключевым этапам подготовки прогнозных моделей
- ❖ Отточить полученные навыки, в рамках курсового проекта

# Требования

## К навыкам

- ❖ Знание математики на уровне школьной программы
  - Производные
  - Медиана, мода, среднее, квантили
- ❖ Знание основ языка Python

# Требования

## Технические

- ❖ Python 3.6+
- ❖ Anaconda (<https://www.anaconda.com/distribution/>)
- ❖ Либо виртуальное окружение с jupyter notebook (<https://jupyter.org/install>)
- ❖ Либо google colab (<https://colab.research.google.com/>)
- ❖ Профиль на GitHub (<https://github.com/>) или его аналоге
- ❖ Профиль на Kaggle (<https://www.kaggle.com/>)

# Что мы изучим?

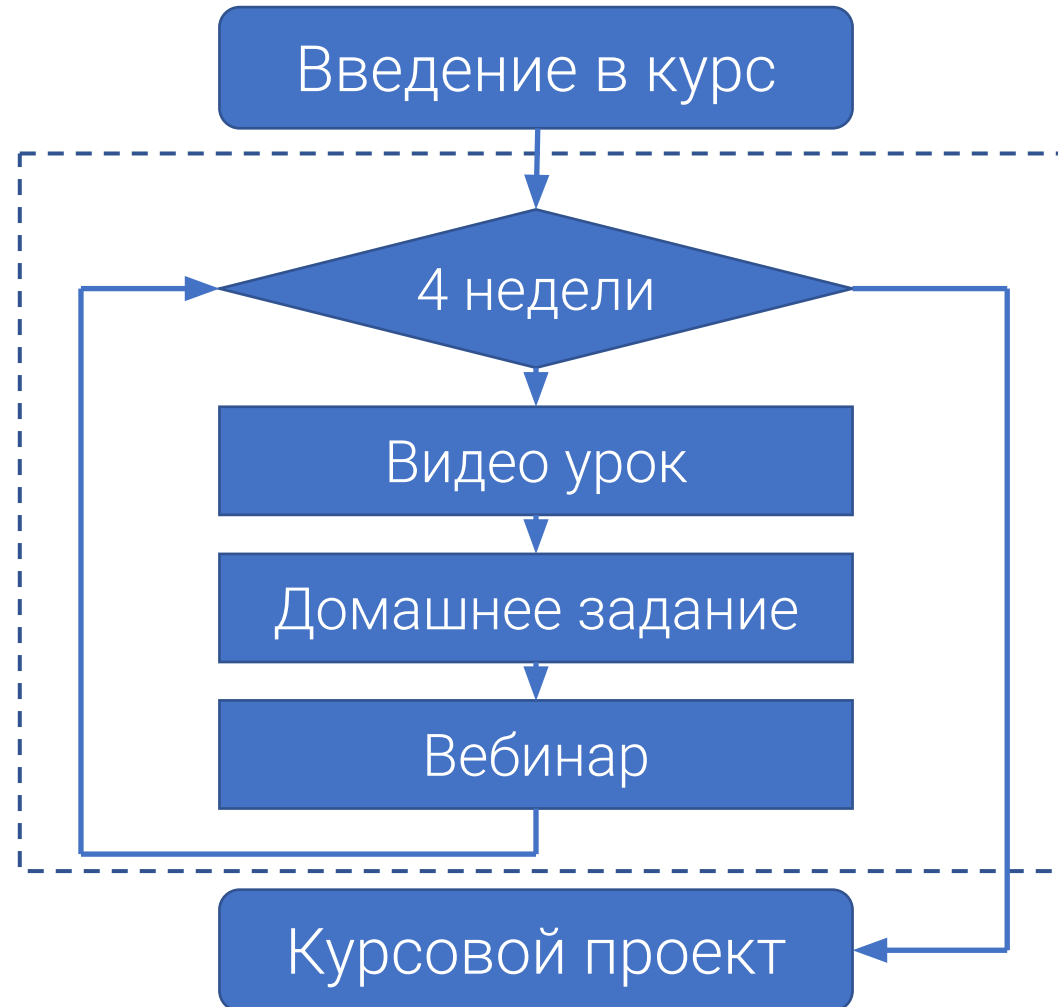
## Библиотеки Python

- ❖ Numpy
- ❖ Pandas
- ❖ Matplotlib / Seaborn
- ❖ SciKit-Learn

## Навыки

- ❖ Обработка данных
- ❖ Визуализация информации
- ❖ Решение задачи классификации
- ❖ Решение задачи регрессии
- ❖ Кластеризация данных
- ❖ Понижение размерности
- ❖ Оценка качества моделей

# Как устроен учебный процесс



# Домашние задания

- ❖ Оформляем решение в тетрадках jupyter notebook
- ❖ Решения прикрепляем к видеоурокам (ссылка на git или файл)
- ❖ Сдавать домашние задания нужно в срок
- ❖ Задания со «звёздочкой», как и задания на повторение материала, сдавать не обязательно, но желательно 😊
- ❖ Если не успели, то можно сдать через службу поддержки ([support@geekbrains.ru](mailto:support@geekbrains.ru))



# Вебинары

- ❖ Обсудим основные этапы построения предсказательных моделей
- ❖ На примере обучающего набора данных отработаем ключевые шаги
- ❖ Подготовимся к сдаче курсового проекта
- ❖ Разберём возникающие вопросы и проблемные моменты
- ❖ После вебинаров домашних заданий сдавать не надо

# Курсовой проект

Тема проекта: «Предсказание цен на жилые дома»

- ❖ Подготовка в течение курса
- ❖ Консультация перед проектом
- ❖ Оценка результатов

# Требования к курсовому проекту

- ❖ Подготовка данных
  - Первичный визуальный анализ
  - Обработка выбросов, пропусков
  - Генерация новых признаков
- ❖ Обучение модели
- ❖ Валидация модели
- ❖ Получить качество работы модели выше заданной планки

# Если возникнут сложности

- ❖ Повторно ознакомиться с материалом (видео уроки, вебинары и др.)
- ❖ Изучить дополнительную информацию в сети Интернет
- ❖ Обратиться с вопросом в чат группы в Telegram
- ❖ Написать в личные сообщения наставнику или преподавателю на портале GeekBrains

# Ответы на вопросы

Спасибо за внимание  
и успехов на курсе!