

1. Знание английского языка
 - a. Чтение литературы
 - b. Разговорный (зависит от компании)
 - c. Написание документации (фриланс, участие в open source, компании)
2. Знание языка
 - a. Тutorials по Python
 - b. Тutorials по R
 - c. Тutorials по Java
 - d. ...
 - e. Организация проекта, принятая в языке
 - f. Техники unit тестирования в языке
 - g. Запуск проекта в виде сервиса
 - h. Раскладка сервиса в боевом окружении
3. В каждом языке - библиотеки для анализа, владение библиотеками.
 - a. Python - numpy
 - b. Python - pandas
 - c. Python - sklearn
 - d. R - ...
 - e. ...
4. Поиск данных для анализа (с бизнес-целью)
 - a. Список источников учебных датасетов
 - b. Список источников готовых датасетов для анализа
 - c. Список ресурсов, где можно собрать неочищенные данные
 - d. Список ресурсов, где есть API
5. Основные математические понятия (применительно к задачам)
 - a. Математика в алгоритмах анализа данных
 - b. Статистика - для оценки результата
6. Улучшение навыков в EDA
 - a. Разбор примеров визуализаций на kaggle
 - b. Чтение статей о восприятии информации и дизайне
 - c. Поиск необычных фактов датасетах, накопление опыта - сбор и исследование данных, оформление в виде законченной статьи