https://dfedorov.spb.ru

ПРОГРАММИСТ - ИНЖЕНЕР	vs.	АНАЛИТИК ДАННЫХ - УЧЕНЫЙ
главное - **алгоритмы** (сложность и скорость работы проверяются на случайных данных) и **языки программирования**, а качество данных - это проблема пользователей		во главе - **данные** (их качество, методы сбора, создание измерительных приборов для сбора данных, проведение экспериментов)
создают собственные виртуальные миры, в которых спокойно живут		пытается понять, как устроен реальный мир, ищет ответы на вопросы, поиск смысла в результатах (полученных значениях)
поиск ошибок типизации, а не в данных		поиск неточностей в данных - это влияет на результат исследования
создание структур и программ		поиск неких вещей
есть истина или ложь		не существует окончательной истины
		- какие данные помогут решить задачу? - какие задачи можно решить с помощью представленных данных?

Аналитик данных — это сотрудник, занимающийся анализом и решением проблем, связанных с данными, типами данных и связями между элементами данных в бизнес-системах или ИТ-системах.
Ниже приведено облако компетенций, которые чаще всего работодатели хотят видеть у специалистов по анализу данных.

Специалисты в своей предметной области: экономисты, юристы, психологи, социологи, культурологи, географы, филологи, лингвисты, биологи, биоинформатики, HR-аналитики, маркетологи и многие другие

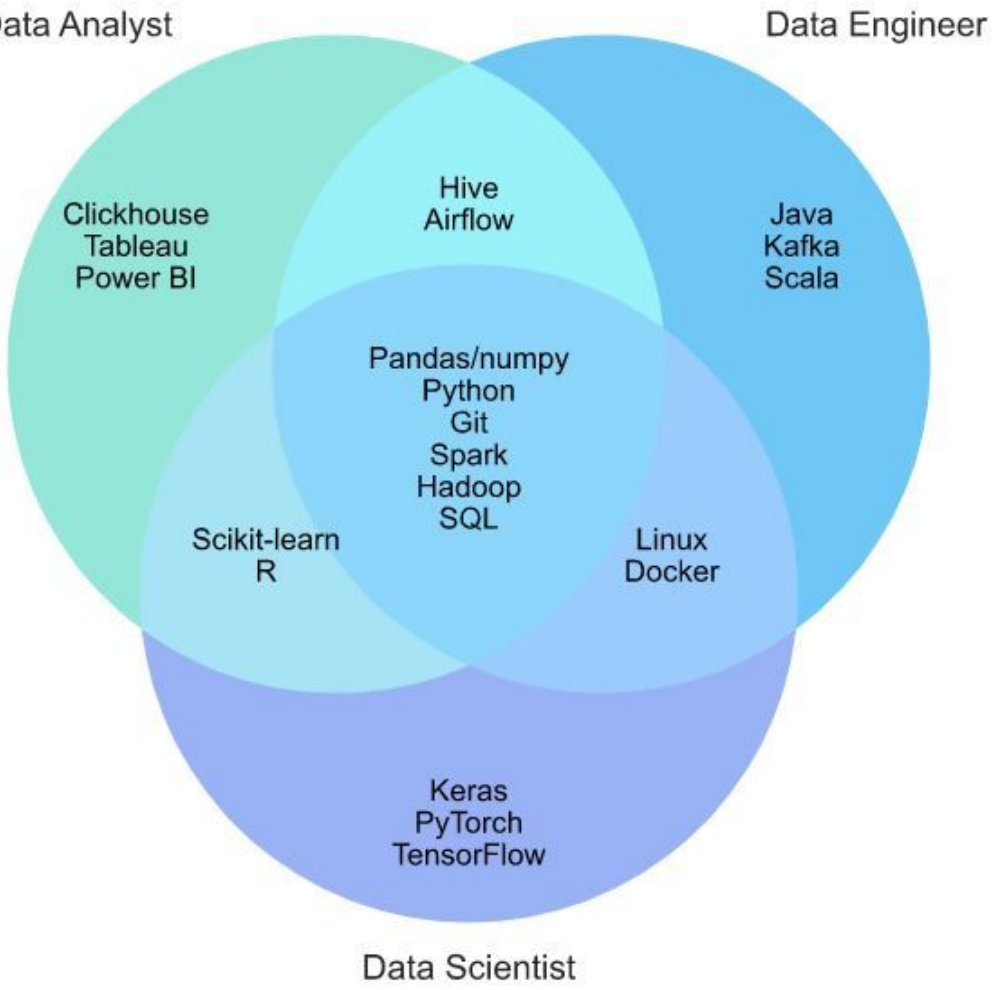
Дата сайентист (ML) — специалист, анализирующий и интерпретирующий большие объемы данных. Это очень широкая специальность, дата сайентистами часто называют и специалистов по ML/DL и просто дата аналитиков.
Ниже приведено облако компетенций, которые чаще всего работодатели хотят видеть у специалистов этого профиля.

Специалисты по математическим моделям, которые умеют их применять для любой предметной области

Аналитик данных

Excel
SQL
Python
R
+
другой софт

Инженер по данным

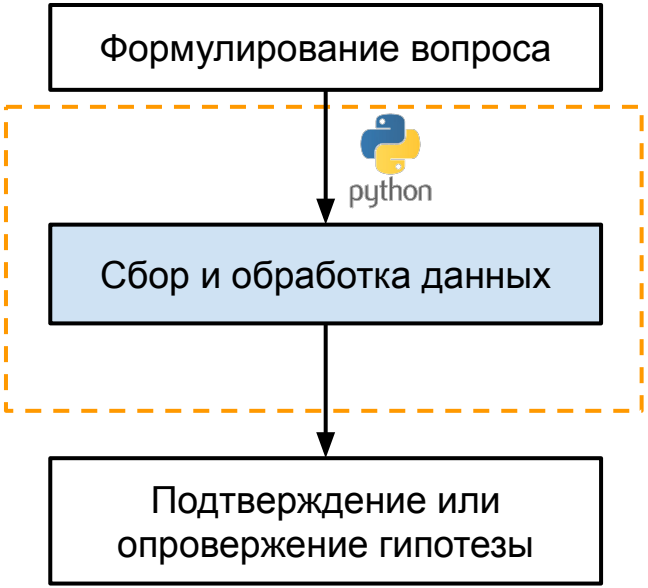


Инженер ML

Источник

Шаги проведения анализа данных

Предсказывать тенденции на основе анализа НН



Excel
СУБД
csv
json

Ключевые навыки в банковской сфере (2018 год)

Тинькофф Банк

Грамотная речь
Пользователь ПК
Телефонные переговоры
Активные продажи
Холодные продажи
B2B Продажи
Работа с юридическими лицами
Консультирование клиентов по телефону
входящие звонки
Грамотность
Прямые продажи
SQL
телефонные продажи
Работа с возражениями
.NET Framework
C#
Взыскание долгов
ASP.NET
ORACLE
AngularJS
JavaScript

Райффайзен БАНК

SQL
Английский язык
XML
Стрессоустойчивость
MS Excel
Linux
DWH
Java
Финансовый анализ
Целеустремленность
Sas
Клиентоориентированность
Коммуникабельность
Ответственность
Android

ВТБ

Поиск и привлечение клиентов
Консультирование клиентов
Активные продажи
Навыки продаж
Проведение презентаций
Грамотная речь
Навыки презентации
Навыки переговоров
Продажа кредитных продуктов
Ведение переговоров
Клиентоориентированность
MS Visio
Анализ трафика
судебная документация
DBlink
View
работа с текущей базой клиентов
SQL
Altova XMLSpy
MQ Explorer
PLSQL Developer
Заключение договоров
Cash Management
Проведение промо акций
Лидогенерация

источник: **pycode.ru** по материалам сайта **hh.ru**

Шаги проведения анализа данных

Оценки студентов

Данные

Выявление проблемных дисциплин в вузе

Формулирование вопроса: что можно узнать?

python

Обработка исходных данных

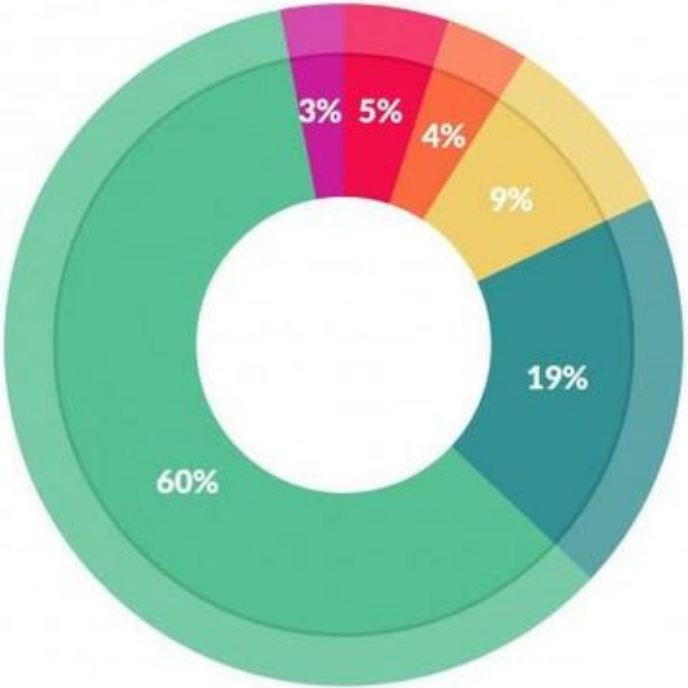
Подтверждение или опровержение гипотезы

Excel
СУБД
csv
json

[12]:
```
courses = pd.DataFrame(df2.groupby('predmet_name')['ball'].mean())
courses = courses.rename(columns={"ball": "mean_grade"})
courses = courses.sort_values('mean_grade', ascending=True)
courses.head()
#Первые 5 дисциплин с наименьшим средним набранным баллом среди студентов
```

[12]:

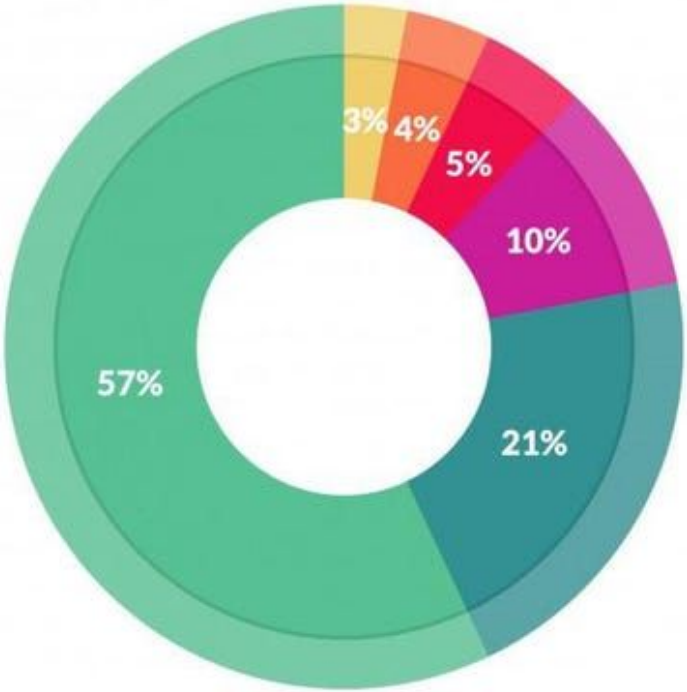
predmet_name	mean_grade
Иностранный язык профессиональный	9.126786
Демография и статистика населения	11.822222
Экономическая теория	12.004232
Линейная алгебра	12.201326
Учетно-аналитическое обеспечение управления организацией	12.740000



What data scientists spend the most time doing

- *Building training sets: 3%*
- *Cleaning and organizing data: 60%*
- *Collecting data sets; 19%*
- *Mining data for patterns: 9%*
- *Refining algorithms: 4%*
- *Other: 5%*

Источник

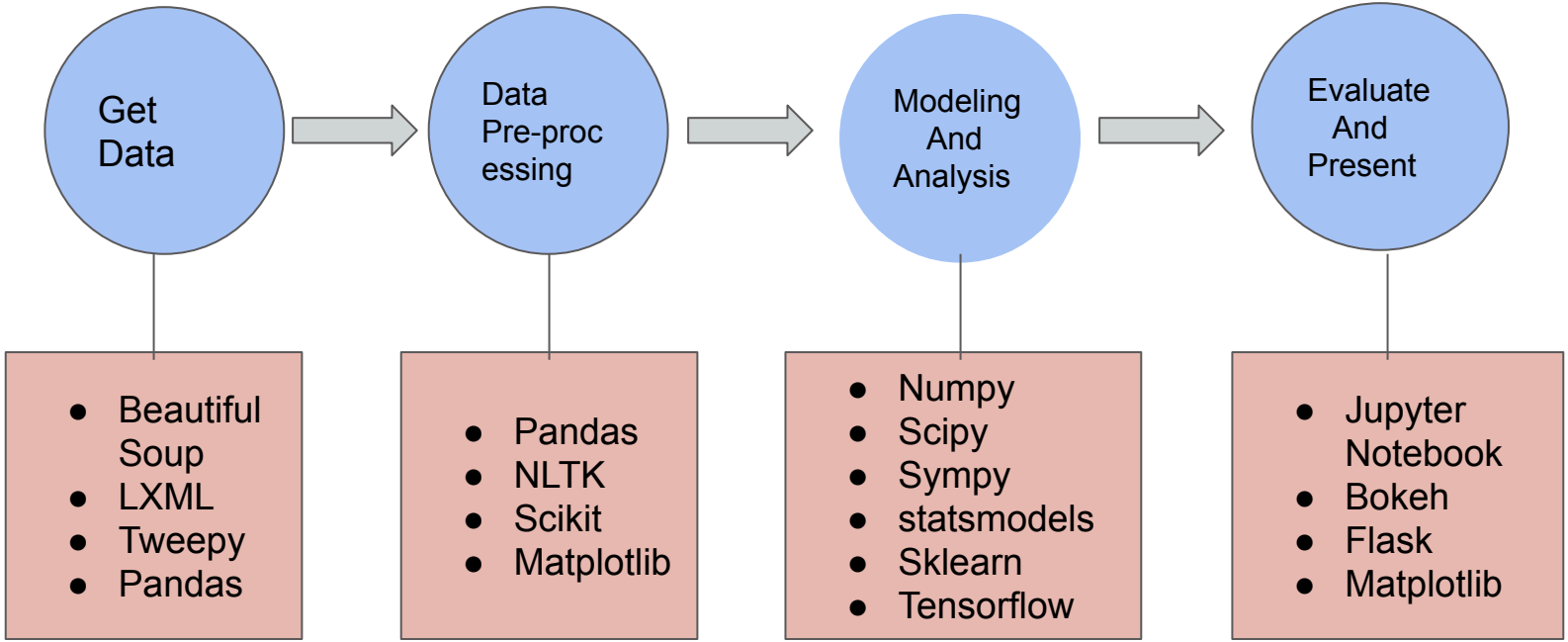


What's the least enjoyable part of data science?

- *Building training sets: 10%*
- *Cleaning and organizing data: 57%*
- *Collecting data sets: 21%*
- *Mining data for patterns: 3%*
- *Refining algorithms: 4%*
- *Other: 5%*

Источник

Data Science with Python



https://dfedorov.spb.ru