



ДОП #1

БЛА-БЛА-БЛА-БЛА-БЛА

ПЛАН ПРЕЗЕНТАЦИИ



Алгоритмы и структуры данных



Excel



R VS Python

ЗАЧЕМ ИЗУЧАЮТ АЛГОРИТМЫ И СТРУКТУРЫ ДАННЫХ?

1. Пройти отбор в FAANG: Facebook, Amazon, Apple, Netflix или Google
2. Найти работу в других компаниях, прежде всего в сфере IT. Пример: Яндекс
3. Знание базовых алгоритмов может помочь лучше подходить к решению задач
4. Это позволяет лучше усвоить язык программирования на задачах, которые ценны сами по себе
5. Общая программистская культура

Список не исчерпывающий.

СТАЖЁР АНАЛИТИК В ЯНДЕКСЕ

Young & Yandex

2 Секция по аналитике с кодом

Это техническое интервью с аналитиком. Оно продлится около часа. Вам предстоит решить одну задачу с написанием кода на Python и 2-3 задачи на любую из следующих тем:

- Теория вероятностей: комбинаторика, дискретная вероятность, условные вероятности, формула полной вероятности, теорема Байеса
- Статистика: точечное и интервальное оценивание, тестирование гипотез, дизайн экспериментов, доверительные интервалы
- Моделирование: различные способы измерения поведения пользователей, основные продуктовые метрики
- Алгоритмы анализа и обработки данных, а также активно применяемые структуры данных (списки, стеки, деки, множества, словари, деревья)
- SQL

Если вы разбираетесь в машинном обучении, вам могут задать несколько вопросов по этой теме.

3 Секция с алгоритмической задачей

Это важный этап отбора: вам предстоит решить одну алгоритмическую задачу за полчаса. Сделать это можно на любом языке программирования, написав код в совместном онлайн-редакторе. Чтобы лучше подготовиться, рекомендуем изучить материалы в нашем [блоге на Habr](#), набить руку на задачах с [Codeforces](#), [LeetCode](#), а также ознакомиться с материалами [тренировок по алгоритмам 1.0](#) и [2.0](#).

3

Секция на алгоритмы и структуры данных

Это один из важнейших этапов отбора: вам предстоит написать код и продемонстрировать знание алгоритмов, решив несколько задач за определённое время. [Здесь](#) можно узнать о секции подробнее. Чтобы получше подготовиться, рекомендуем изучить материалы [тренировок по алгоритмам 1.0 и 2.0](#), а также набить руку на задачах с [LeetCode](#) (уровень medium) и [Codeforces](#) (div2).

ПОЧЕМУ ПОЛЕЗНО ЗНАТЬ EXCEL

«Англоязычный рекрутинговый портал Monster выбрал [топ-7 навыков](#), наиболее востребованных работодателями. Анализ информации — на втором месте. Это умение особенно актуально для тех, кто только начинает карьеру. [По данным портала Burning Glass](#), для 67% стартовых вакансий требуется продвинутое знание Microsoft Excel. «Мы движемся к миру, в котором данные управляют процессами. Но цифры сами по себе бесполезны: нужно правильно их интерпретировать. Поэтому работодатели ищут кандидатов, которые это понимают», — считает Джули Фридман Стил, председатель ассоциации футурологов World Future Society.»

Excel базовый навык, который часто пишут в резюме единственным, когда нечего больше написать 😊

EXCEL

Плюсы Excel:

1. Простота
2. Не нужно писать код
3. Легок в освоении
4. Много где требуется на базовом или «продвинутом» уровне

Минусы Excel:

1. Ограниченность данных (идём за БД в SQL и т.д.)
2. Разные названия формул для разных языков
3. Разные разделители для разных языков
4. Некоторая неудобность привычной навигации (enter смещает вниз и т.д.)
5. Ограниченность применения (нет тонких настроек под конкретную задачу, что дает ЯП)

EXCEL (ЕЩЁ НЕМНОГО ПРО НЕГО)

Частые ошибки в Excel:

1. Забывают ставить \$\$ при копировании формул
2. Пишут = несколько раз в формуле
3. Неаккуратно пользуются Tab и Enter
4. Используют мышку вместо горячих клавиш (Shift, Ctrl, стрелки)

Резюме: Подходит для быстрого решения небольших задач и быстрой предобработки данных.

Для существенных задач часто нужно использовать более серьезные инструменты: SQL, Python, R.

ЯЗЫК ПРОГРАММИРОВАНИЯ R

R — язык математиков (прежде всего статистиков), созданный для математиков. Это одновременно плюс и минус.

«R это среда для статистических расчетов. R задумывался как свободный аналог среды S-Plus, которая, в свою очередь, является коммерческой реализацией языка расчетов S. Язык S довольно старая разработка. Он возник еще в 1976 году в компании Bell Labs и был назван, естественно, по мотивам языка C»

R применяется для

1. Работы с данными: статистика, первичный анализ (графики, таблицы корреляции), продвинутое **математическое** моделирование.
2. Может использоваться вместо программ для математического анализа (MATLAB или Octave)
3. Статистический анализ, эконометрика. От вычисления среднего до вейвлет-преобразований временных рядов.

ЯЗЫК ПРОГРАММИРОВАНИЯ R

R — язык математиков (прежде всего статистиков), созданный для математиков. Это одновременно плюс и минус.

«R это среда для статистических расчетов. R задумывался как свободный аналог среды S-Plus, которая, в свою очередь, является коммерческой реализацией языка расчетов S. Язык S довольно старая разработка. Он возник еще в 1976 году в компании Bell Labs и был назван, естественно, по мотивам языка C»

R применяется для

1. Работы с данными: статистика, первичный анализ (графики, таблицы корреляции), продвинутое **математическое** моделирование.
2. Может использоваться вместо программ для математического анализа (MATLAB или Octave)
3. Статистический анализ, эконометрика. От вычисления среднего до вейвлет-преобразований временных рядов.

ЯЗЫК ПРОГРАММИРОВАНИЯ R

R используется в:

1. Американский и европейских университетах, научно-исследовательских центрах, R&D департаментах крупных корпораций
2. Другие западные компании (Boeing, ANZ, Google, Firefox, BOA, LinkedIn, Flipkart, Amazon и др.)
3. В России R используется в Банке России, научно-исследовательских центрах при Правительстве, Минфине и т.д. для макроэкономических, эконометрических исследований.
4. Используется в России наряду с Python там, где нужен анализ данных без использования продвинутого машинного обучения.

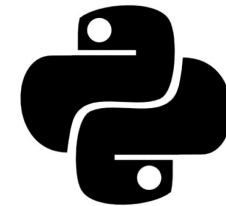
R VS PYTHON

Эконометрика, математическая статистика



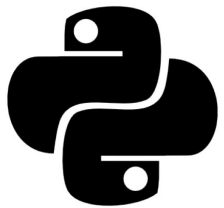
1. Лучше подходит для статистических исследований (проверка гипотез, построение дов. интервалов и т.д.)
2. Из-под капота имеются матрицы, корни, экспоненты, определители, дифференцирование, собственные значения и т.д.
3. Много хороших пакетов для работы с временными рядами и другой эконометрикой. Но нет адекватного МО.

Машинное обучение, нейронные сети



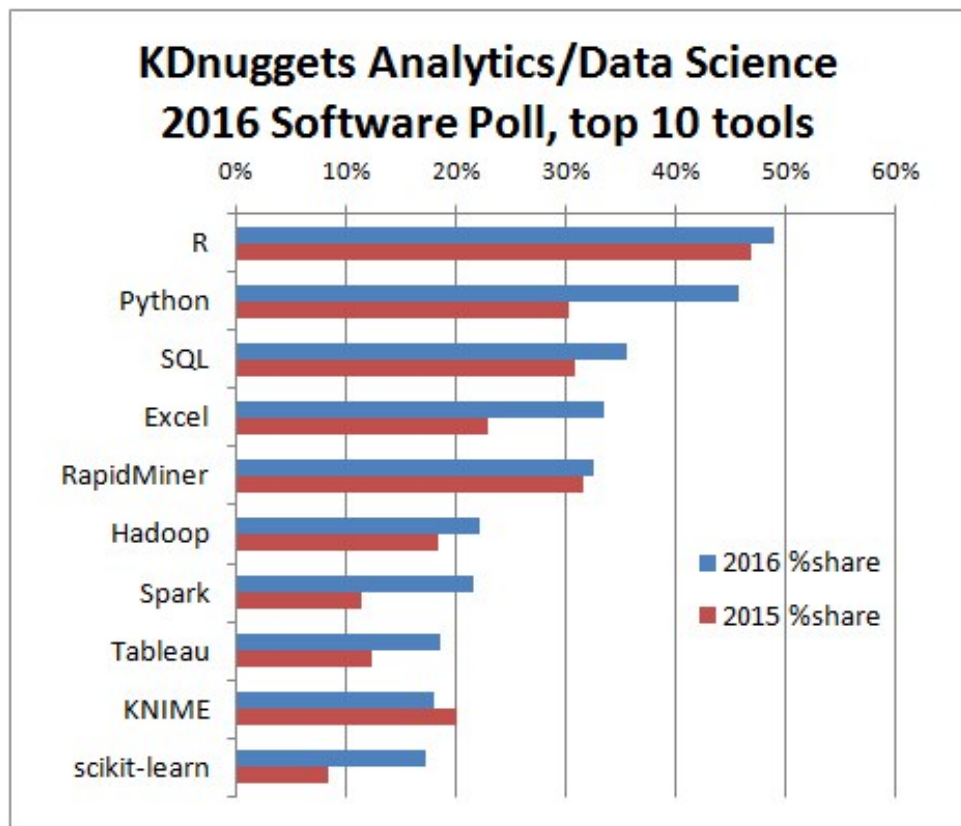
1. Универсальный язык с ООП. Можно писать сайты, разрабатывать приложения и использовать алгоритмы машинного обучения (далее МО)
2. Самый популярный на текущий момент. Реализованы отличные библиотеки по работе с МО, нейросетями.
3. Нет адекватных пакетов для временных рядов, многое математическое нужно подгружать пакетами, которые не всегда хорошие

ОБЩЕЕ У R И PYTHON



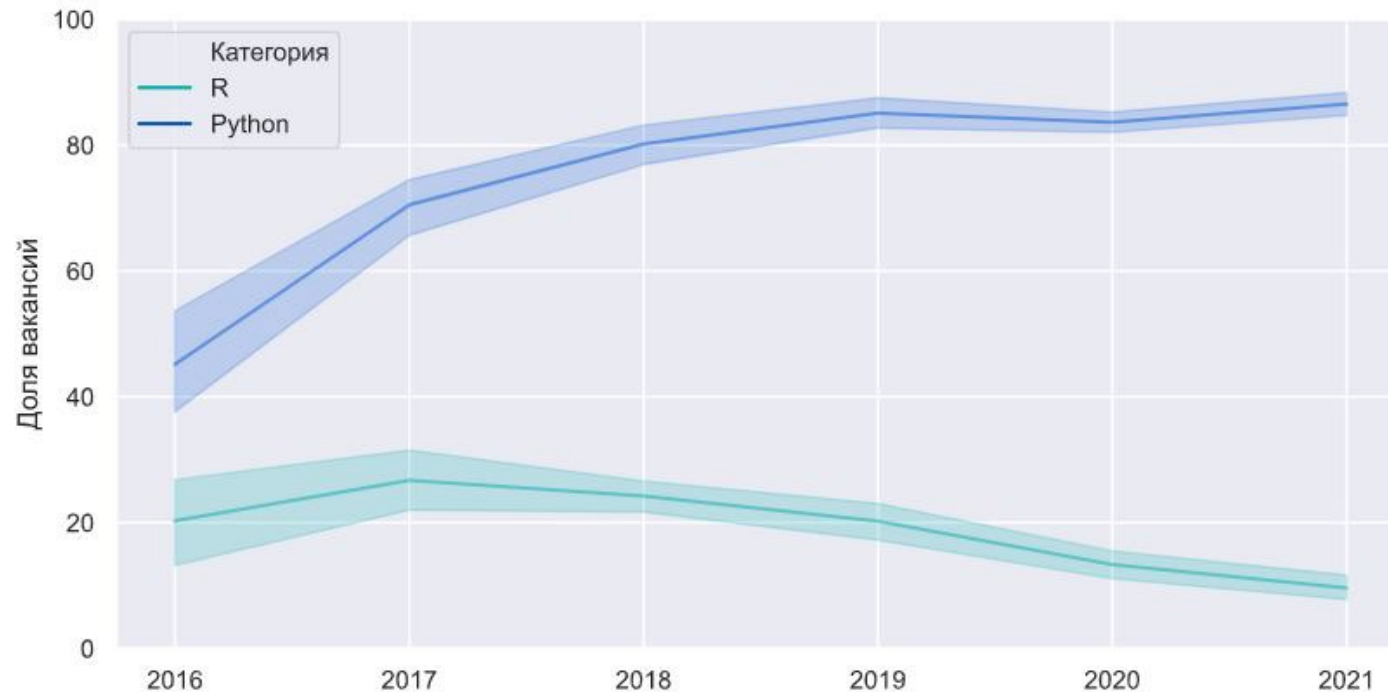
1. Оба языка содержат удобные средства визуализации, построения графиков, диаграмм и т.д.
2. Оба языка являются высокоуровневыми, т.е. код писать достаточно просто и быстро (по сравнению с C, C++ и др.)
3. Похожие структуры данных, типы переменных
4. Оба языка имеют схожие характеристики (типизация, интерпретаторы)
5. Относительно медленные языки
6. Сравнительно просты к освоению
7. Удобные инструменты для работы с табличными данными: вектора, многомерные массивы, дата фреймы

ЯЗЫК ПРОГРАММИРОВАНИЯ R























На основе анализа вакансий в сфере DS в сообществе ODS

Доля вакансий по языкам программирования



Источник: habr.com/ru/company/ods/blog/572264/



	Jan 2022	Jan 2021	Change	Programming Language	Ratings	Change
1	3	3	▲	 Python	13.58%	+1.86%
2	1	1	▼	 C	12.44%	-4.94%
3	2	2	▼	 Java	10.66%	-1.30%
4	4	4		 C++	8.29%	+0.73%
5	5	5		 C#	5.68%	+1.73%
6	6	6		 Visual Basic	4.74%	+0.90%
7	7	7		 JavaScript	2.09%	-0.11%
8	11	11	▲	 Assembly language	1.85%	+0.21%
9	12	12	▲	 SQL	1.80%	+0.19%
10	13	13	▲	 Swift	1.41%	-0.02%
11	8	8	▼	 PHP	1.40%	-0.60%
12	9	9	▼	 R	1.25%	-0.65%
13	14	14	▲	 Go	1.04%	-0.37%
14	19	19	▲▲	 Delphi/Object Pascal	0.99%	+0.20%
15	20	20	▲▲	 Classic Visual Basic	0.98%	+0.19%
16	16	16		 MATLAB	0.96%	-0.19%
17	10	10	▼▼	 Groovy	0.94%	-0.90%
18	15	15	▼	 Ruby	0.88%	-0.43%
19	30	30	▲▲	 Fortran	0.77%	+0.31%
20	17	17	▼	 Perl	0.71%	-0.31%



Rank	Change	Language	Share	Trend
1		Python	28.74 %	-1.8 %
2		Java	18.01 %	+1.2 %
3		JavaScript	9.07 %	+0.6 %
4	▲	C/C++	7.4 %	+1.1 %
5	▼	C#	7.27 %	+0.7 %
6		PHP	6.06 %	+0.0 %
7		R	4.19 %	+0.3 %
8		Objective-C	2.27 %	-1.4 %
9		Swift	1.91 %	-0.2 %
10		TypeScript	1.74 %	-0.0 %
11		Matlab	1.74 %	+0.0 %
12		Kotlin	1.71 %	+0.0 %
13		Go	1.19 %	-0.1 %
14		VBA	1.14 %	-0.0 %
15		Ruby	1.07 %	-0.0 %
16		Rust	0.98 %	-0.0 %
17	▲▲▲▲	Ada	0.75 %	+0.1 %
18	▲▲▲▲	Dart	0.72 %	+0.2 %
19	▲▲▲▲	Abap	0.66 %	+0.2 %
20	▼▼	Visual Basic	0.64 %	-0.0 %
21	▼▼▼▼	Scala	0.48 %	-0.2 %
22	▲▲▲▲	Groovy	0.48 %	+0.1 %
23	▼▼▼▼	Lua	0.42 %	-0.2 %
24	▼	Perl	0.38 %	-0.1 %
25	▲▲	Haskell	0.34 %	+0.1 %
26	▼▼	Julia	0.34 %	-0.0 %



ВАКАНСИИ В ЦБ РФ

Банк России рассматривает кандидатов на текущие вакансии в Департамент Финансовой Стабильности.

Экономист-математик

Основные задачи:

- построение моделей компонент кредитного риска (вероятности дефолта (PD), потерь при дефолте (LGD)) для портфеля ипотечных кредитов (R, SAS, Python);
- разработка предложений по установлению макропруденциальных надбавок к коэффициентам риска по кредитам физических лиц для целей расчета нормативов достаточности капитала;
- подготовка регулярных аналитических отчетов по тенденциям в сегменте банковского кредитования физических лиц (MS Excel, SAS, PowerPoint)
- участие в проведении стресс-тестирования российских банков: обработка больших массивов исходных данных (SQL, SAS EG);
- построение эконометрических моделей (R, SAS, Python).

Требования:

- высшее образование (математика, математические методы в экономике);
- опыт работы приветствуется.

Необходимые теоретические знания:

- знание языков программирования SQL, Python, R;
- знание линейной алгебры, теории вероятностей, математической статистики, основ машинного обучения;
- знания основ банковского дела.

Аналитик финансового рынка

Задачи:

- Участие в разработке системы аналитических показателей;
- Проведение аналитических исследований по отдельным сегментам финансового рынка ;
- Подготовка отраслевых аналитических обзоров по рынку;
- Применение алгоритмов и инструментов для сбора и трансформации данных в удобный для анализа вид;
- Разработка и создание отчетов в среде бизнес аналитики (работа с OLAP-клубами, SQL- запросы);
- Разработка скриптов по выгрузке и аналитике данных.

Требования:

- Высшее образование;
- Умение работать с большими массивами данных;
- Уверенное владение SQL, VBA, Python, R;
- Приветствуется знание математической статистики;
- Желательно наличия аттестата ФСФР, CFA, FRM;
- Наличие опыта в инвестиционных, финансовых или IT компаниях.

ЕЩЁ ВАКАНСИИ

Аналитик (розничные риски) в банке Открытие

Требования:

- Опыт работы в коммерческом банке (предпочтительно в розничных рисках);
- Знание SQL и/или SAS, инструментов предиктивной аналитики;
- Глубокое знание Excel;
- Опыт подготовки презентаций в Power Point.

Будет преимуществом:

- Понимание специфики процесса розничного кредитования;
- Развитые аналитические и коммуникационные навыки;
- Навыки портфельного анализа (vintage analysis; знание основных показателей: LTV, COR, FPD, SPD, TPD, GL, NL, AR, etc);
- Опыт анализа/ валидации/ разработки скоринговых или других эконометрических/ математических/ DS моделей, в т.ч. учебных;
- Знание Python/ R.

Аналитик в Центр макроэкономических исследований

Требования:

- **Высшее образование (экономическое обязательно, математическое является дополнительным плюсом);**
- **Продвинутые знания по макроэкономике и эконометрике** (опыт в области макроэкономического моделирования является дополнительным плюсом);
- Грамотная, структурированная устная и письменная речь;
- Продвинутые знания MS Office;
- **Знание английского языка на свободном уровне обязательно (чтение научных статей и монографий);**
- **Навыки программирования в MATLAB, Dynare, R** (продвинутые знания в программировании являются плюсом);
- Умение работать в команде, инициативность, целеустремленность.

ЕЩЁ ВАКАНСИИ

Требования: Аналитик экономико-математического моделирования

- Высшее/незаконченное высшее образование в области экономической науки (макрэкономика, микроэкономика, эконометрика, экономика отраслевых рынков, экономика государственного сектора), теории вероятностей и математической статистики
- Опыт работы не требуется, вы можете быть выпускником или так же студентом последних курсов ВУЗа с возможностью работать полный рабочий день
- Отличное владение инструментарием эконометрики, включая регрессии и системы уравнений на панельных данных, временных рядах, с ограниченной зависимой переменной
- Уверенное владение не менее чем одним из двух стэков для моделирования: **R** и/или Python (pandas, statsmodels, sklearn)
- Базовые навыки программирования
- Чтение профессиональной литературы на английском языке
- Работа в Excel (функции, солвер)

Рассматриваются в качестве дополнительного преимущества:

- Владение методами «разность-в-разностях», «синтетические контроли»
- Статистические методы проектирования экспериментов
- Опыт работы с пакетами forecast (**R**), pmdarima, prophet/kats (Python)
- Знание SQL и работа с реляционными базами данных
- Знание методов байесовской статистики
- Опыт работы с методами машинного обучения

Сбер (Data Scientist)

Требования

- Высшее образование в области экономики/математики;
- Знание элементов статистики (моменты, распределения и т.п.);
- Базовая культура работы с данными разной природы: структурированные данные, текст и изображения;
- Понимание принципа работы и построения базовых алгоритмов машинного обучения (линейная регрессия, решающее дерево, случайный лес, GBM, MLP, CNN, RNN, Transformer и т.д.);
- Владение одним из следующих языков - Python, **R** (Python предпочтительно);
- Приветствуется опыт участия в соревнованиях по анализу данных и опыт разработки моделей в банковской сфере.

ЛИТЕРАТУРА

- 1. R inferno – очень глубокая книжка с многим тонкостями. Продвинутый уровень
- 2. Введение в R Дружков, Золотых. Маленькое пособие на ~20 страниц. Даёт базовое понимание R
- 3. Наглядная статистика. Используем R! – больше про статистику, но большое приложение по R. Промежуточный уровень



СПАСИБО ЗА
ВНИМАНИЕ!