

Data Science

Утверждаю
Директор школы
Сламов Д.Т.



Курс продолжается в течение 6 месяцев, разделенных на 26 недель.

Курс включает 208 академических часов интенсивного обучения.

Стоимость курса: 500 000 тенге

Этот интенсивный буткэмп по data science сочетает в себе практические и теоретические аспекты обучения. Студенты начнут с основ Python, перейдут к объектно-ориентированному программированию и изучат важные алгоритмы и структуры данных. Основное внимание уделяется использованию библиотек Python, таких как NumPy, Pandas, Matplotlib, Seaborn для анализа и визуализации данных. Курс включает исследовательский анализ данных, подкрепленный математикой, необходимой для машинного обучения, в том числе контролируемым и неконтролируемым обучением. Завершает обучение финальный проект и поддержка в построении карьеры.

Цель курса - обучить студентов применению инструментов и методов data science для извлечения ценной информации из наборов данных.

Задачи курса включают освоение Python, изучение статистических методов, визуализацию данных и создание прогнозных моделей.

Формат обучения: Оффлайн

Формат занятий: Лекции / предзаписанные уроки

Платформы для обучения: Stepik, GitHub, HackerRank, Kaggle, StrataScratch etc.

Периодичность занятий: 4 раза в неделю

.

Python	
Этот модуль буткэмпа по data science охватывает основы Python, включая базовые конструкции, коллекции и управление памятью. Рассматриваются функции, их особенности и принципы объектно-ориентированного программирования. Участники изучат практические навыки программирования, необходимые для эффективной работы с данными и разработки алгоритмов.	4 недели, от 32 часа
Темы - Базовые конструкции Python - Коллекции и работа с памятью - Функции и их особенности в Python - Объектно-ориентированное программирование	Проект Работа над проектом
Дополнительные материалы, гостевая лекция	

Структуры данных и алгоритмы	
Этот модуль буткэмпа по науке о данных посвящен изучению структур данных, включая коллекции, деревья, а также понятия maps и hashing. Участники изучат методы бинарного поиска и различные алгоритмы сортировки.	2 недели, от 16 часов
Темы - Коллекции (списки, массивы, связанные списки, очереди, стек) - Деревья - Maps & Hashing - Бинарный поиск - Алгоритмы сортировки	
Дополнительные материалы, гостевая лекция	

Вероятность и статистика для машинного обучения и науки о данных	
В этом модуле дается подробное введение в фундаментальные концепции вероятности и статистики, которые имеют решающее значение для машинного обучения и науки о данных. Он охватывает целый ряд тем, начиная с основ теории вероятностей и распределений вероятностей и переходя к описанию и анализу распределений с несколькими переменными, выборке, точечному оцениванию, доверительным интервалам и проверке гипотез.	1 неделя, от 8 часов
Темы - Введение в теорию вероятностей и вероятностных распределений - Описание распределений вероятностей и распределений вероятностей с	

несколькими переменными - Выборка и точечная оценка - Доверительные интервалы и проверка гипотез	
Дополнительные материалы	

Линейная алгебра для машинного обучения и науки о данных	
В этом модуле рассматриваются основные понятия линейной алгебры, необходимые для освоения машинного обучения и науки о данных. Он включает в себя понимание и решение систем линейных уравнений, изучение векторов и линейных преобразований, а также понимание детерминантов. Предназначенный для начинающих специалистов по исследованию данных, он обеспечивает прочную математическую основу, облегчающую разработку алгоритмов и интерпретацию многомерных данных.	1 неделя, от 8 часов
Темы - Системы линейных уравнений - Решение систем линейных уравнений - Векторы и линейные преобразования - Детерминанты	
Дополнительные материалы	

Математический анализ для машинного обучения и науки о данных	
Этот модуль предлагает комплексное изучение исчисления, специально разработанное для машинного обучения и приложений науки о данных. Он охватывает основы производных и оптимизации, переходит к градиентам и алгоритму градиентного спуска, а в завершение рассматривает методы оптимизации в нейронных сетях, такие как метод Ньютона. Этот модуль предназначен для тех, кто хочет улучшить свои аналитические навыки. Он предоставляет учащимся математические инструменты, необходимые для оптимизации алгоритмов и моделей машинного обучения.	1 неделя, от 8 часов
Темы - Дериивативы и оптимизация - Градиенты и градиентный спуск - Оптимизация в нейронных сетях и метод Ньютона	
Дополнительные материалы	

Анализ данных

В этом модуле буткэмпа по анализу данных участники изучат основы анализа данных, используя библиотеки NumPy, Pandas, Matplotlib и Seaborn. От подготовки и очистки данных до их анализа и визуализации - модуль научит, как задавать правильные вопросы и принимать обоснованные решения на данных, улучшая обмен информацией.	6 недель, от 48 часов
Темы - Введение в анализ данных - Задавайте вопросы, чтобы принимать решения на основе данных - Подготовка данных к исследованию - Обработка данных от грязных к чистым - Анализируйте данные, чтобы ответить на вопросы - Визуализация данных для облегчения обмена информацией	Проект Работа над проектом
Дополнительные материалы, гостевая лекция	

А/В-тестирование и прогнозирование временных рядов	
В рамках этого модуля буткэмпа по анализу данных участники погрузятся в мир А/В-тестирования, изучив его основы, различные подходы к дизайну тестов, включая случайный дизайн и тесты на соответствие парам. Также будут рассмотрены основы прогнозирования временных рядов с использованием моделей ETS и ARIMA, а также методы анализа и визуализации результатов, обеспечивая полное понимание эффективности решений на основе данных.	2 недели, от 8 часов
Темы - Основы А/В-тестирования - Тесты со случайным дизайном - Тесты на соответствие парам - Основы прогнозирования временных рядов - Модели ETS - Модели Arima - Анализ и визуализация результатов	Проект Работа над проектом
Дополнительные материалы, гостевая лекция	

Машинное обучение	
В этом модуле буткэмпа по анализу данных участники осваивают основы машинного обучения, включая контролируемое обучение с задачами регрессии и классификации, а также продвинутое обучение. Также будут рассмотрены неконтролируемое обучение, системы рекомендаций и обучение с подкреплением, обеспечивая комплексное понимание ключевых концепций и техник машинного обучения.	5 недель, от 40 часов

Темы - Контролируемое машинное обучение: регрессия и классификация - Продвинутые алгоритмы обучения - Неконтролируемое обучение, рекомендатели, обучение с подкреплением	Проект Работа над проектом
Дополнительные материалы, гостевая лекция	

Разработка и реализация Data-проектов	
Переход от теории к практике в этом практическом модуле. Ознакомьтесь с Kaggle, ареной для проведения соревнований по науке о данных. Превратите полученные знания в реальные проекты, демонстрирующие применение в реальном мире. Повысьте свой профессиональный уровень, составив резюме и профиль LinkedIn. В завершение - выход в профессиональную сферу через стажировку, применение своих навыков для решения реальных задач и получение бесценного опыта. Это важный мост от обучения к практическому применению.	4 недели, от 32 часа
Темы - Kaggle - Projects - CV, LinkedIn - Internship	Проект Финальный проект
Дополнительные материалы, гостевая лекция	

По завершению обучения на курсе по data science, студенты могут ожидать приобретения следующих навыков и результатов:

- **Программирование на Python:** Владение синтаксисом Python, написание и отладка кода, работа с библиотеками и модулями.
- **Работа с данными:** Умение обрабатывать, очищать, трансформировать и анализировать данные с использованием Pandas.
- **Визуализация данных:** Создание информативных графиков и диаграмм с использованием библиотек визуализации, таких как Matplotlib и Seaborn.
- **Статистический анализ:** Понимание и применение статистических методов и тестов для анализа и интерпретации данных.
- **Машинное обучение:** Знание основных алгоритмов машинного обучения, умение обучать модели на данных и оценивать их производительность.
- **Математические основы:** Понимание линейной алгебры, теории вероятностей и математического анализа, которые необходимы для глубокого понимания алгоритмов машинного обучения.
- **Решение реальных задач:** Умение применять полученные знания для решения реальных бизнес-задач и анализа данных.
- **Критическое мышление и аналитические навыки:** Способность к критическому анализу данных и выявлению скрытых закономерностей и взаимосвязей.

- **Создание проектов:** Портфолио проектов, демонстрирующих навыки анализа данных, которое можно представить потенциальным работодателям.
- **Профессиональное развитие:** Знание того, как продвигаться в карьере аналитика данных, умение составлять резюме и успешно проходить собеседования.

Приобретение этих навыков делает выпускников курса компетентными в области анализа данных и конкурентоспособными на рынке труда.

По завершению обучения участники курса получают квалификацию *data scientist* с навыками в программировании, статистике и машинном обучении.

Методологическая база курса по *data science* может включать в себя следующие компоненты:

- **Теоретические лекции:** Предоставление теоретических знаний по ключевым концепциям анализа данных, статистики, машинного обучения и программирования.
- **Практические задания:** Решение задач и выполнение проектов, которые помогут студентам применять теоретические знания на практике.
- **Интерактивное программирование:** Использование интерактивных средств, таких как *Jupyter Notebooks*, для написания кода, анализа данных и визуализации результатов.
- **Кейс-стади:** Анализ реальных примеров и сценариев использования анализа данных в различных отраслях и сферах деятельности.
- **Групповые проекты:** Работа в командах над совместными проектами для развития навыков командной работы и решения сложных задач.
- **Обратная связь и рецензирование:** Получение обратной связи от преподавателей и однокурсников, а также возможность рецензировать работы других студентов.
- **Самостоятельное изучение:** Предоставление ресурсов и материалов для самостоятельного изучения, чтобы студенты могли углубить свои знания.
- **Оценка и тестирование:** Проведение тестов и экзаменов для оценки знаний и навыков
- **Карьерное консультирование:** Помощь в составлении резюме, подготовке к собеседованиям и развитии карьеры в области *data science*.

Эта методологическая база создает сбалансированную структуру курса, объединяющую теорию и практику, и способствующую всестороннему развитию навыков в области *data science*.

Соотношение теории к практике: 20/80

Наличие стажировки/практики: TOO Datalysis, TOO FabLab, Full Stack Lab, ALASH Coding, Naiza solutions

Отслеживание прогресса учащихся через регулярные оценки, обратную связь, и аналитику для оптимизации процесса обучения и достижения лучших результатов.

Критерии оценки успеваемости студентов на курсе по *data science* могут включать следующие аспекты:

- **Тесты и экзамены:** Оценка знаний и умений студентов через проведение регулярных тестов и экзаменов, включающих теоретические вопросы и практические задачи.
- **Домашние задания:** Оценка способности студента применять полученные знания на практике через решение задач, анализ данных и написание кода в домашних условиях.
- **Практические проекты:** Оценка навыков анализа данных и программирования на основе выполнения комплексных проектов, возможно, с реальными данными.

- **Участие в дискуссиях:** Активное участие в дискуссиях, семинарах и групповых проектах, что отражает вовлеченность и способность к командной работе.
- **Презентации:** Оценка навыков коммуникации и представления результатов анализа данных через устные презентации и написание отчетов.
- **Самостоятельное обучение:** Оценка инициативы и способности студента к самостоятельному изучению дополнительных материалов и углублению в тему.
- **Соблюдение сроков:** Своевременное выполнение заданий и проектов, что отражает организованность и ответственность студента.
- **Качество кода:** Оценка качества написанного кода в плане чистоты, эффективности и соответствия стандартам программирования.
- **Критическое мышление:** Способность анализировать данные, выявлять скрытые закономерности, критически оценивать результаты и предлагать решения.
- **Обратная связь от преподавателей и однокурсников:** Учет мнения преподавателей и однокурсников по поводу вклада студента в групповые проекты и дискуссии.

Важно, чтобы оценка успеваемости была многоаспектной, чтобы получить полную картину развития навыков студента в течение курса. Также следует предоставлять обратную связь, чтобы студенты могли улучшать свои навыки и корректировать подходы к обучению.

* Сертификат выдается при достижении 75% успеваемости. Для прохождения курса “С отличием”, требуется минимум 90%.

Отбор потенциальных студентов на курс по data science может быть структурирован на нескольких этапах, чтобы гарантировать, что участники обладают необходимыми качествами и мотивацией для успешного обучения.

- **Подача Заявления:** Первый этап, на котором кандидаты подают заявления, включая своё резюме, мотивационное письмо и, возможно, академические результаты.
- **Онлайн Тестирование:** Кандидаты могут быть приглашены пройти онлайн тесты по математике, логике или основам программирования, чтобы оценить их базовые навыки.
- **Собеседование:** На этом этапе проводятся собеседования (очные или онлайн), чтобы оценить мотивацию кандидата, его способность к критическому мышлению и интерес к анализу данных.
- **Задание на Практические Навыки:** Может быть предложено выполнить небольшое практическое задание, связанное с анализом данных, чтобы оценить способности кандидата к решению задач.
- **Финальное Решение:** На основе результатов всех этапов комиссия по приему принимает решение о зачислении кандидатов на курс.

Критерии отбора могут включать:

- **Академическая подготовка:** Наличие предыдущего образования, связанного с математикой, статистикой, информатикой или смежными дисциплинами.
- **Опыт программирования:** Основные знания языка программирования (например, Python) могут быть преимуществом.
- **Мотивация и интерес:** Желание учиться и применять знания в области data science.

- **Критическое мышление и решение проблем:** Способность анализировать информацию и решать задачи.
- **Коммуникативные навыки:** Умение эффективно общаться и работать в команде.

Отборочный процесс должен быть четко структурирован и объективен, чтобы обеспечить справедливый отбор самых подходящих кандидатов для курса.

Критерии отчисления с курса:

- **Низкая посещаемость:** Пропуск более 30% занятий без уважительной причины.
- **Невыполнение домашних заданий:** Систематическое игнорирование заданий или несвоевременная сдача более 50% работ.
- **Низкие академические результаты:** Средний балл по промежуточным и финальным оценкам ниже установленного минимального порога (например, 60%).
- **Отсутствие активности:** Неучастие в командных проектах, обсуждениях и практических заданиях.
- **Нарушение дисциплины:** Грубое поведение, неуважение к преподавателям и другим участникам курса.
- **Мошенничество:** Списывание, плагиат или передача чужих работ под своим именем.



Хайдар Бұлхайыр

Преподаватель
Senior Data Engineer

AI/ML Solution
Architect
AI/ML
Solution Architect
qCloudy

Senior Data Engineer
АО «Bereke Bank»

Главный
DataScientist
Halyk Bank

[Linkedin](#)