

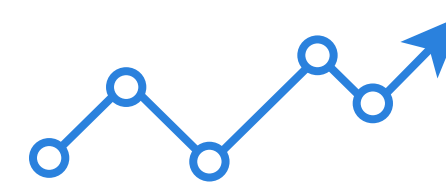
Исследование рынка в Казахстане

# DATA SCIENCE / АНАЛИТИКА

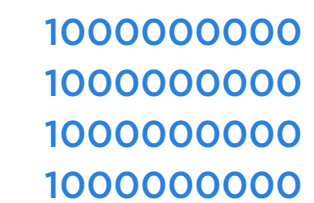
# О чем можно узнать

- Портрет профессии
- Зарплата и мотивация
- Инструменты и задачи

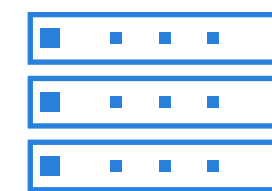
# Респонденты



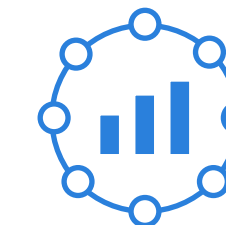
Аналитики



Data Scientists



DWH-инженеры



ML-инженеры

# Цель исследования

## Изучить рынок аналитики и Data Science в Казахстане:

- Как выглядит актуальный портрет казахстанского аналитика
- Сколько зарабатывают аналитики и как оценивают уровень своей зарплаты
- Какова структура аналитических отделов и уровень их развития
- Какие аналитики решают задачи и какие инструменты используют
- Что является мотивацией для представителей профессии
- Каких знаний и навыков им не хватает

Опрос респондентов начался в ноябре 2019 года и продлился два месяца. Для формирования анонимного опросника и конфиденциального сбора данных мы использовали сервис Turiform.

В опроснике принимали участие казахстанские аналитики, Data Science-специалисты, ML- и DWH-инженеры.

# Методика сбора данных

Мы рассылали опросник по внутренней базе аналитиков и дата-сайентистов, таргетировали посты в социальных сетях Kolesa Group на узкий круг интересов, а также распространяли в тематических Telegram-каналах, чатах (DS/ML kz, BI Community) и медиа (BlueScreen). Опросник заполнили 308 респондентов.

Огромное спасибо вам за помощь!

Для получения более репрезентативных данных мы очищали базу от выбросов и пропусков (сильно завышенные или заниженные ответы). После, оперировали частотами ответов, рассчитывали различные метрики центральной тенденции (среднее, медиана), анализировали разброс показателей (стандартное отклонение, коэффициент вариации) и квантили различных уровней.

Затем для более глубокого исследования нюансов профессии и проверки наших выводов и гипотез мы попросили экспертов прокомментировать полученные результаты с точки зрения их опыта.

## Наши эксперты



**Думан Уватаев**

Chief Data Officer  
в Kaspi Bank



**Пётр Царенко**

Chief Data Officer  
в Kolesa Group



**Ануар Аймолдин**

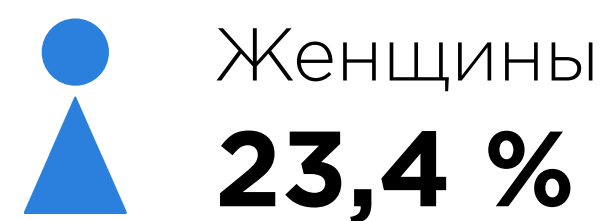
Руководитель отдела ИИ  
в BTS Digital



**Айбек Ергожаев**

Операционный директор  
Frontier KZ

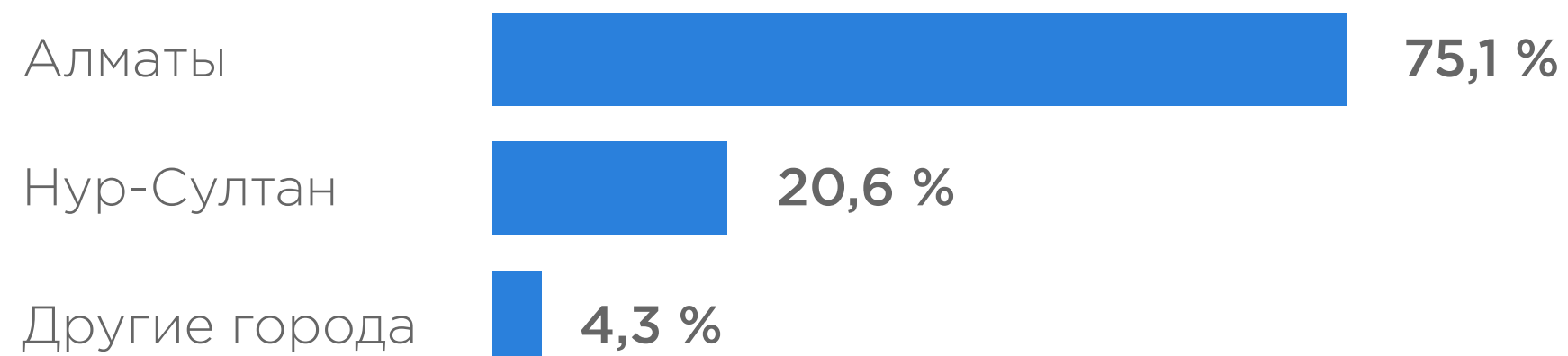
# Портрет респондента



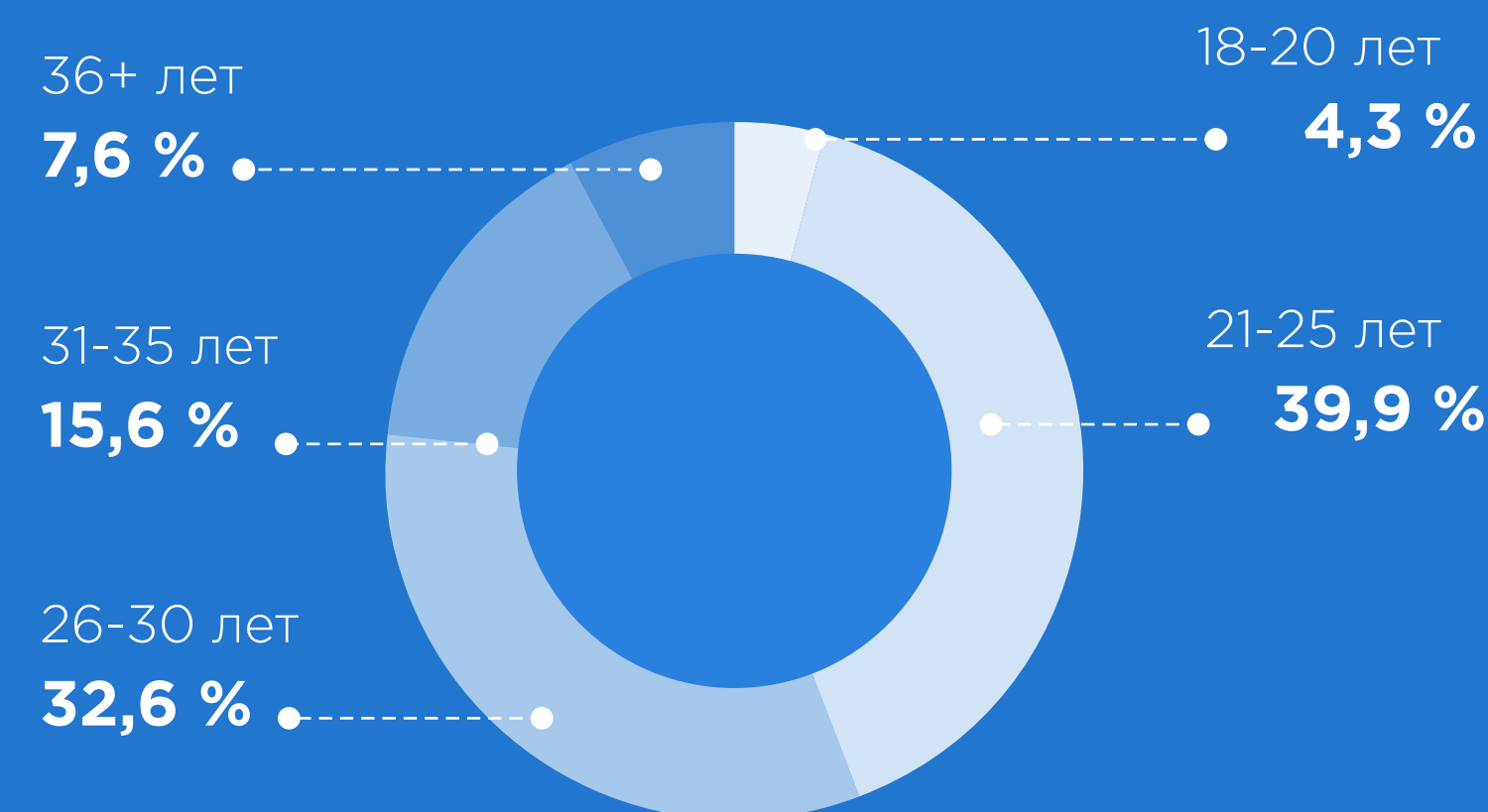
Возраст большинства респондентов

**22-27 лет**

Большинство специалистов, работающих в области аналитики и Data Science, сосредоточены в Алматы

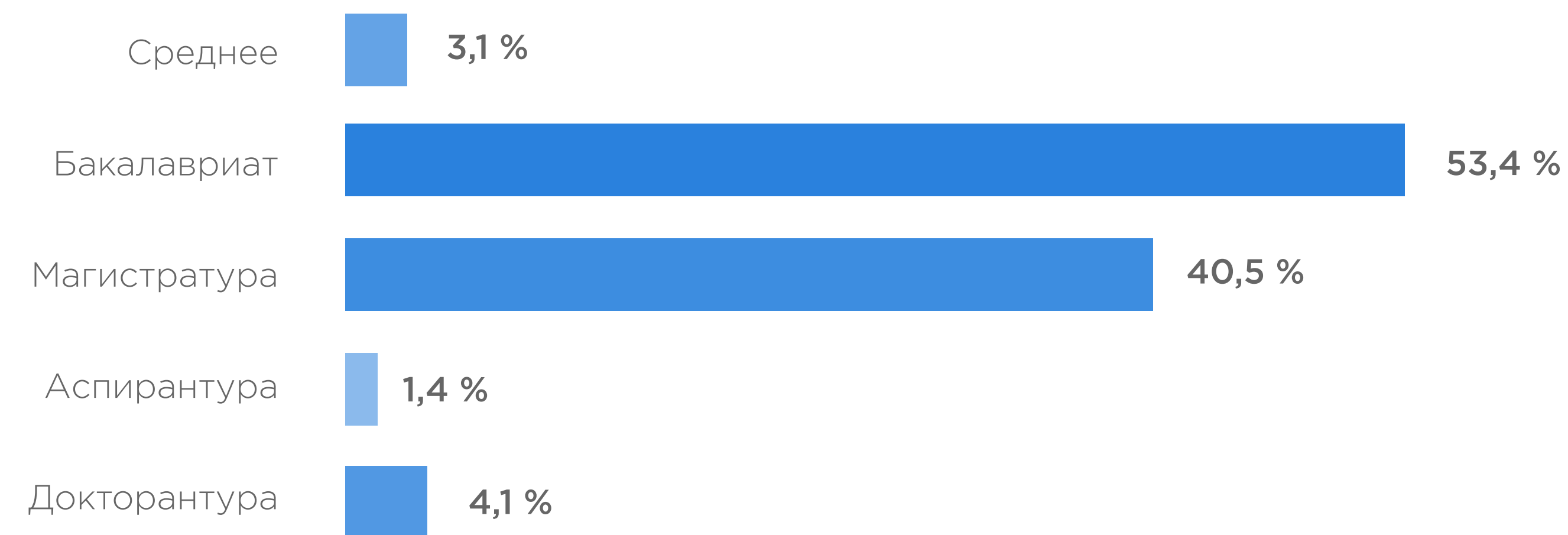


Более 76% респондентов в возрасте до 30 лет (включительно)



# Уровень образования

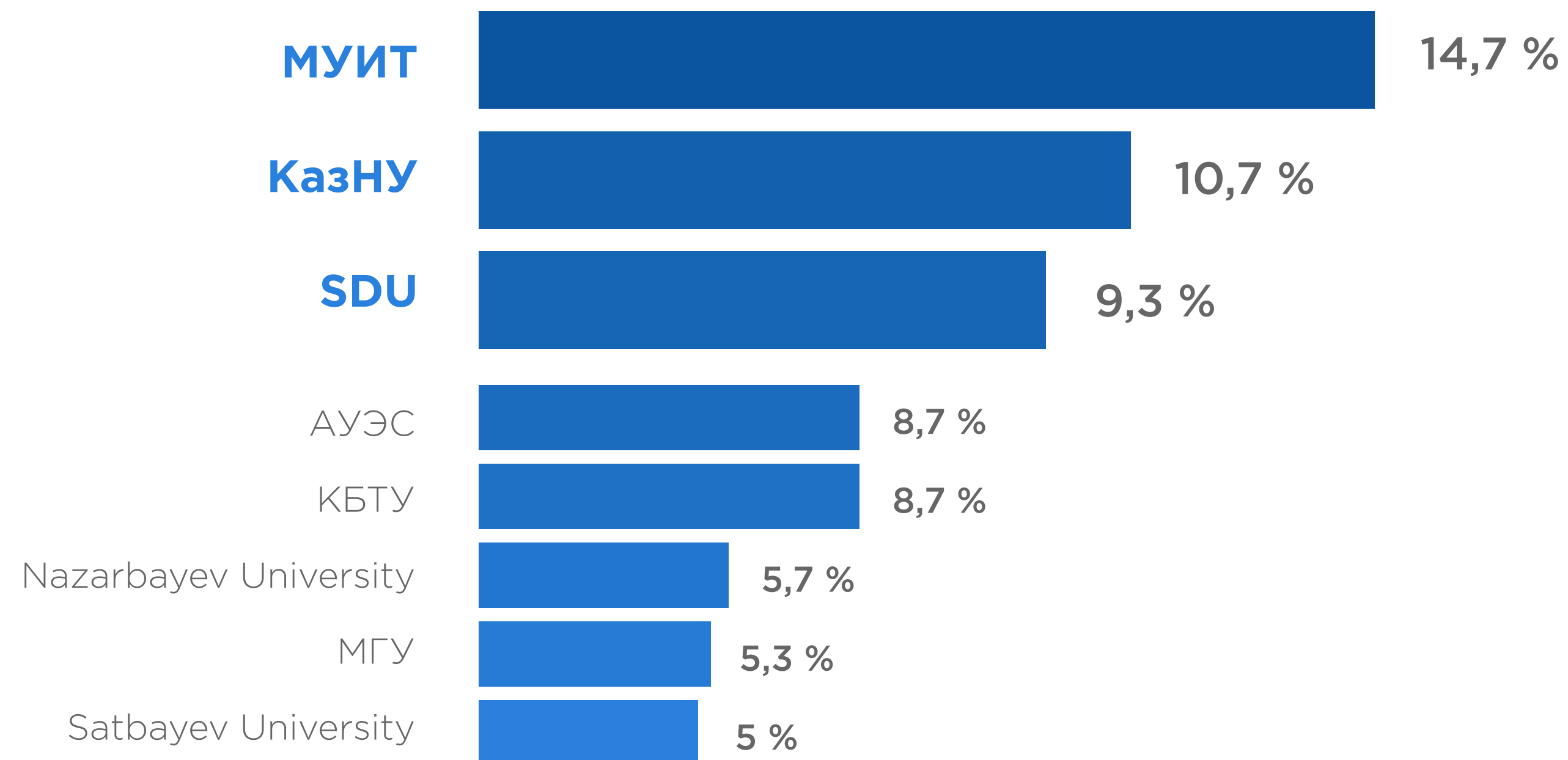
Около половины участников исследования окончили бакалавриат (53 %) ещё около 40 % — магистратуру



**97** %  
имеют высшее образование

# Учебные заведения

Большая часть аналитиков и специалистов по Data Science — выпускники казахстанских вузов. В топ-3 входят МУИТ, КазНУ и SDU



**всего** респондентов получили  
высшее образование  
за пределами СНГ  
**4 %**

## **- Согласно опросу, наибольшее число респондентов выпустились из МУИТ, КазНУ и СДУ. Эти вузы действительно лидеры в подготовке таких специалистов?**

**Пётр Царенко** | Chief Data Officer в Kolesa Group

– Я считаю, что эти вузы дают неплохую базу в области математики и программирования. И если студент дополнительно пройдет узкоспециализированные курсы и получит практический опыт, то сможет стать хорошим специалистом.

Для нашей компании вуз не является определяющим фактором при приеме на работу.

Data Science в Казахстане является растущим направлением, о котором лет пять назад почти никто не знал и не говорил.

Большинство респондентов (около 80 %) работают в анализе данных три года и меньше. Поэтому и программы в вузах по подготовке специалистов в области анализа данных появились совсем недавно, а во многих вузах не появились до сих пор.

Думаю, что эта ситуация скоро изменится. Университеты и бизнес уже начинают работать совместно, чтобы создавать более качественные программы для студентов. А специализация «Анализ данных» в ближайшие пару лет станет одной из самых популярных среди абитуриентов.

## **Как вы оцениваете уровень развития рынка труда в DS/ML/аналитике данных в Казахстане**

**Ануар Аймолдин** | Руководитель отдела ИИ BTS Digital

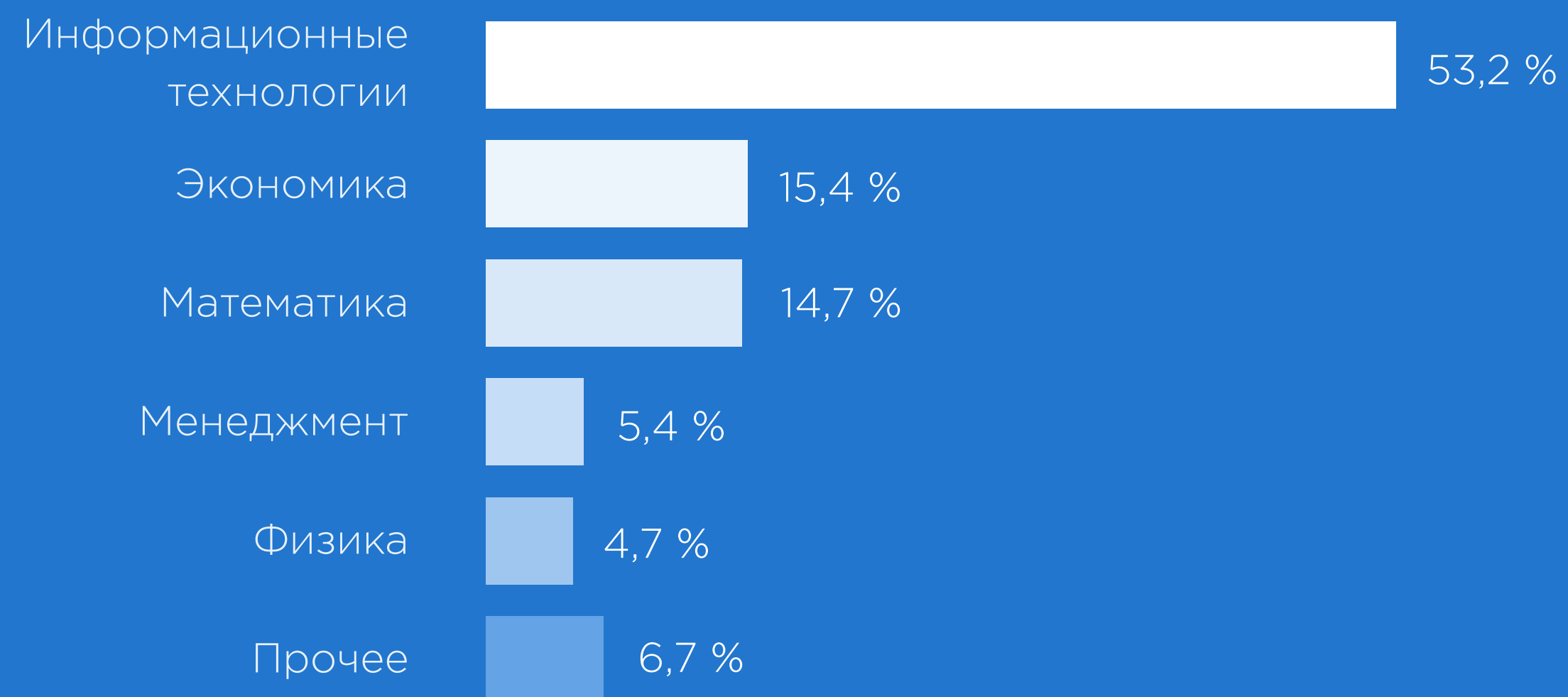
– Рынок начал активно развиваться года три назад. За это время произошел внушительный рост: появляются крупные продуктовые/инфраструктурные/ресерч-команды и AI-driven-стартапы, проводятся многочисленные митапы, а наши ребята успешно выступают на международных конкурсах и пишут интересные исследовательские проекты.

Радует, что в ряде университетов появились специализированные программы по изучению наук о данных (магистерская программа в Назарбаев Университете, КазНУ, КБТУ, Yessenov Data Lab в ALMA University, Astana IT University). При этом по-прежнему ощущается некоторый дефицит middle-, senior-специалистов и людей с опытом руководства DS-компаниями. На ближайшие несколько лет прогнозирую еще более сильную конкуренцию за кадры в области анализа данных, особенно в банковской сфере.



# Специализация

Среди респондентов преобладают выпускники факультетов технических и точных наук, гораздо меньше специалистов с гуманитарным образованием



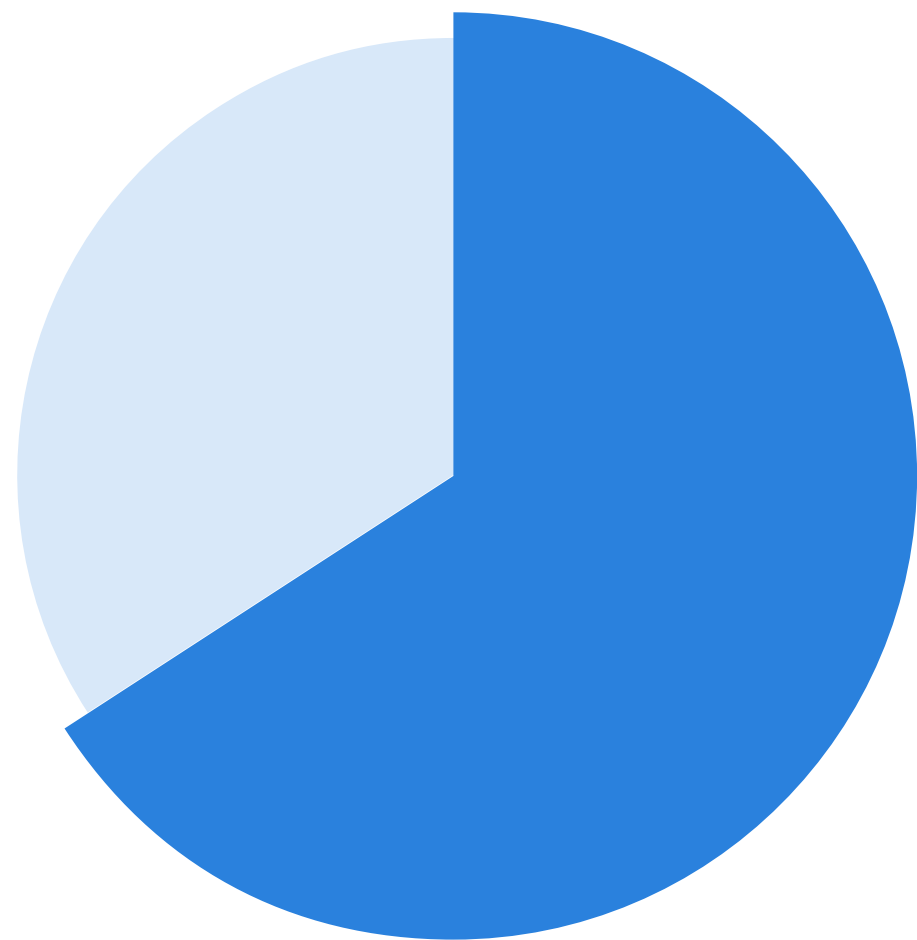
**73** % технических и точных наук

## Непрофильные специализации:

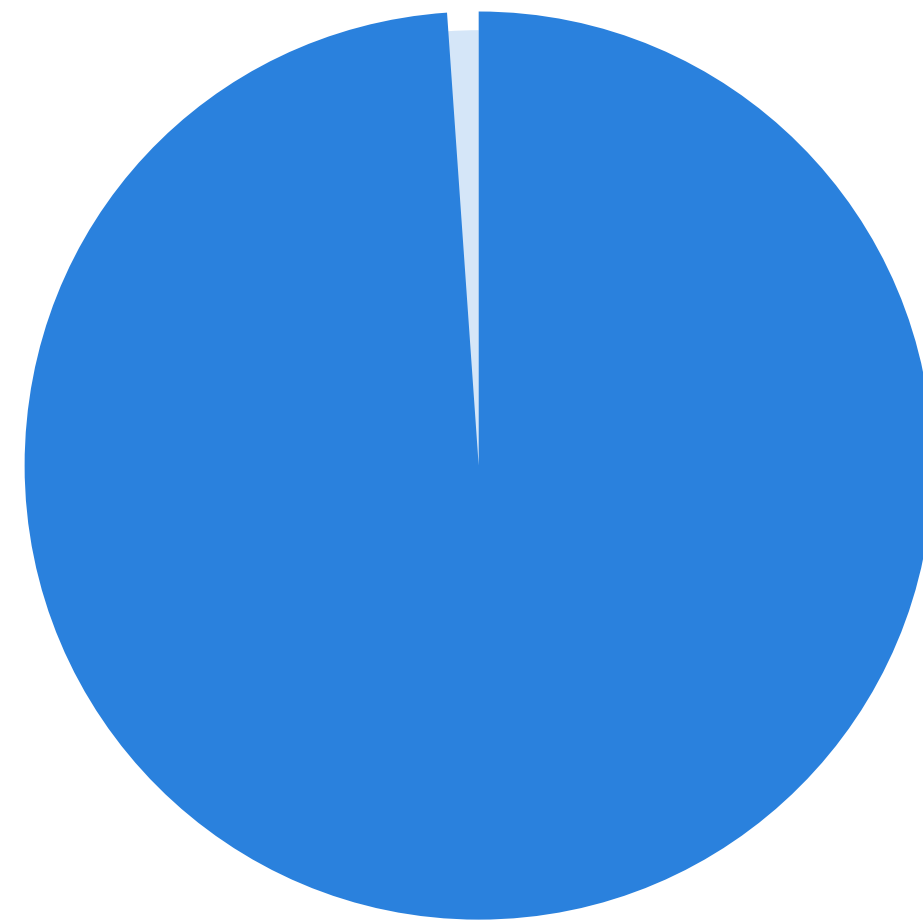
- биология
- экология
- филология
- психология
- лингвистика

# Владение языками

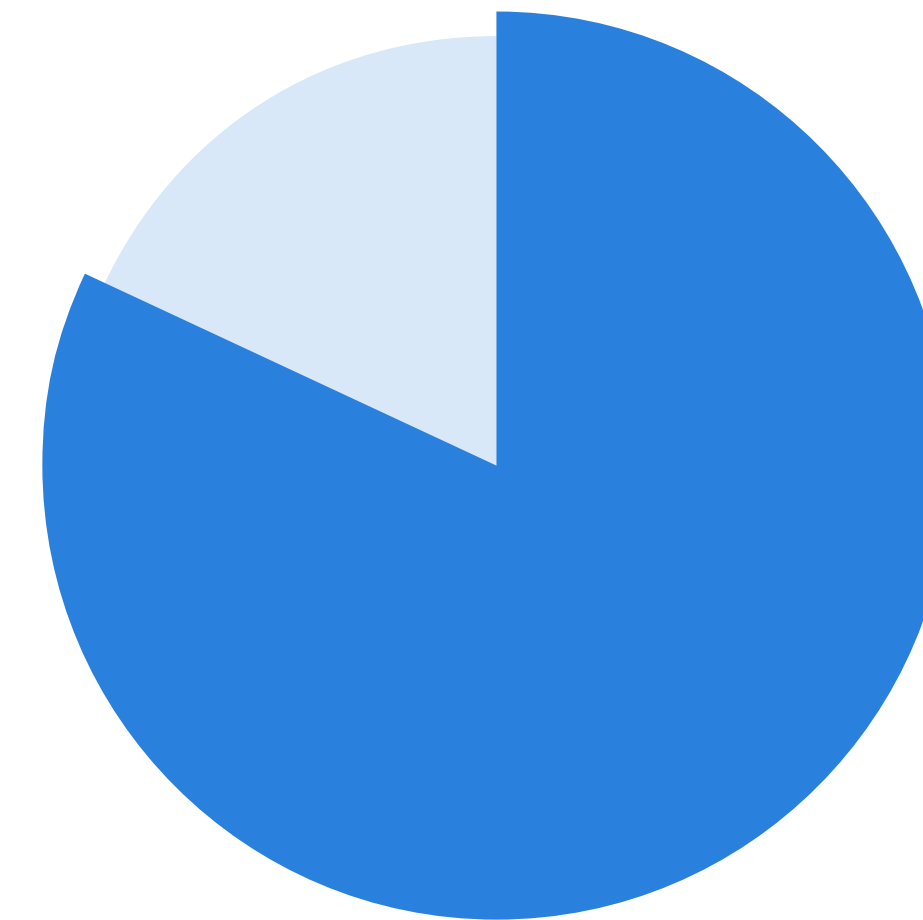
Казахский **66 %**



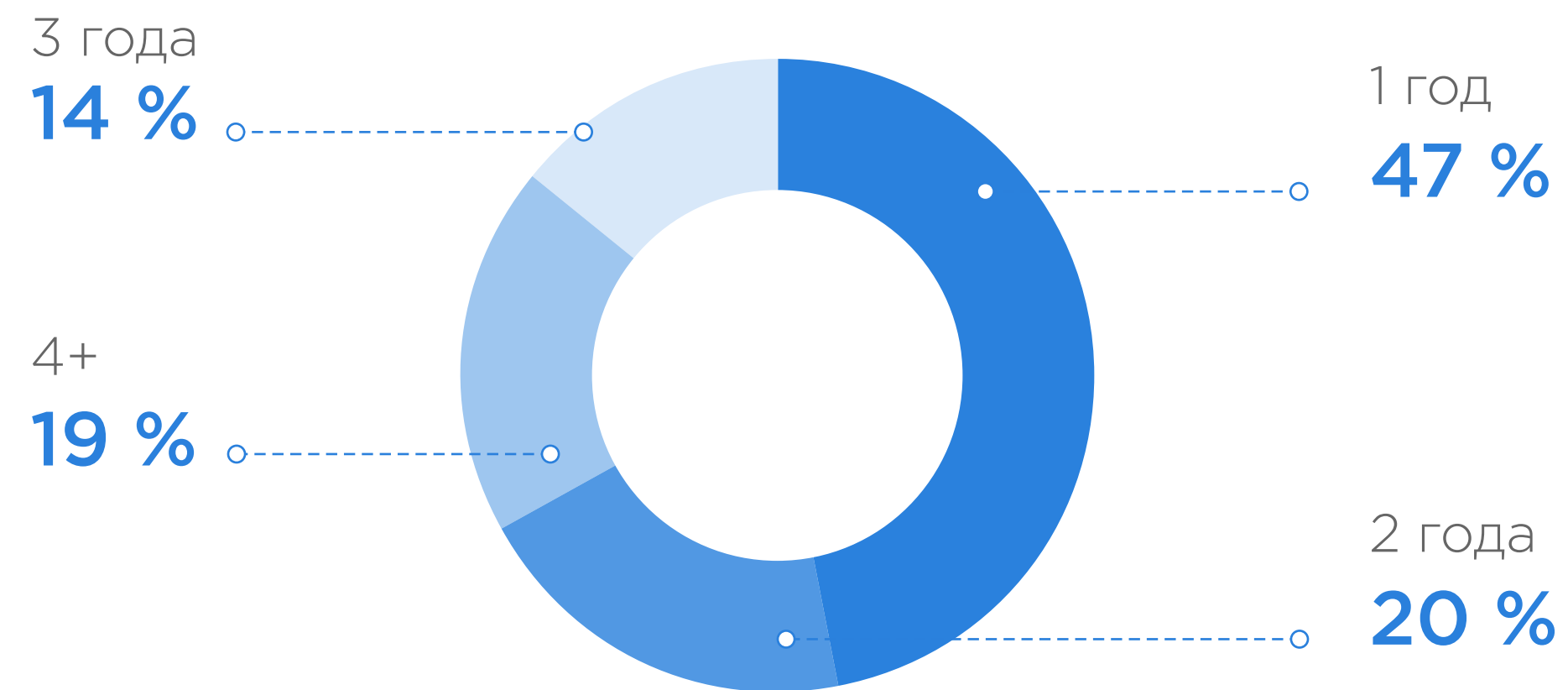
Русский **99 %**



Английский **82 %**



# Стаж в анализе данных

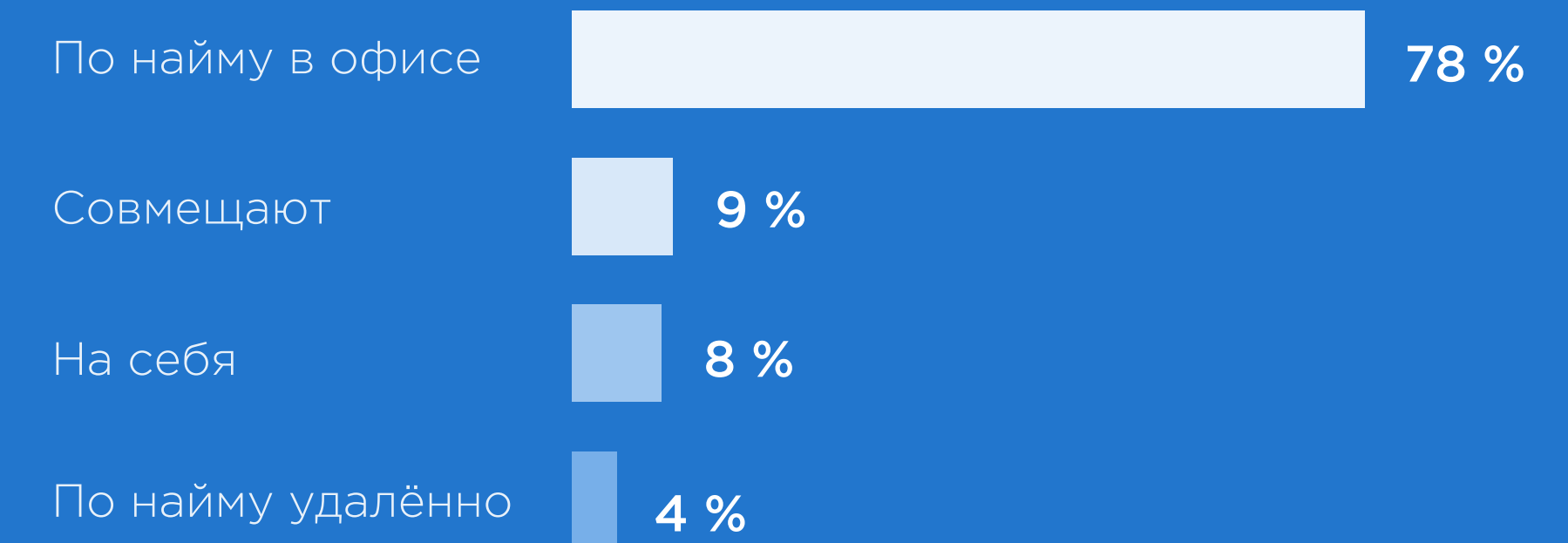


Доля специалистов со стажем менее трёх лет характеризует зрелость рынка Data Science в разных странах.

В Казахстане доля специалистов с опытом работы до трёх лет составляет 80 %, в России – 75 %, в Индии – 63 %, в США – 52 %.

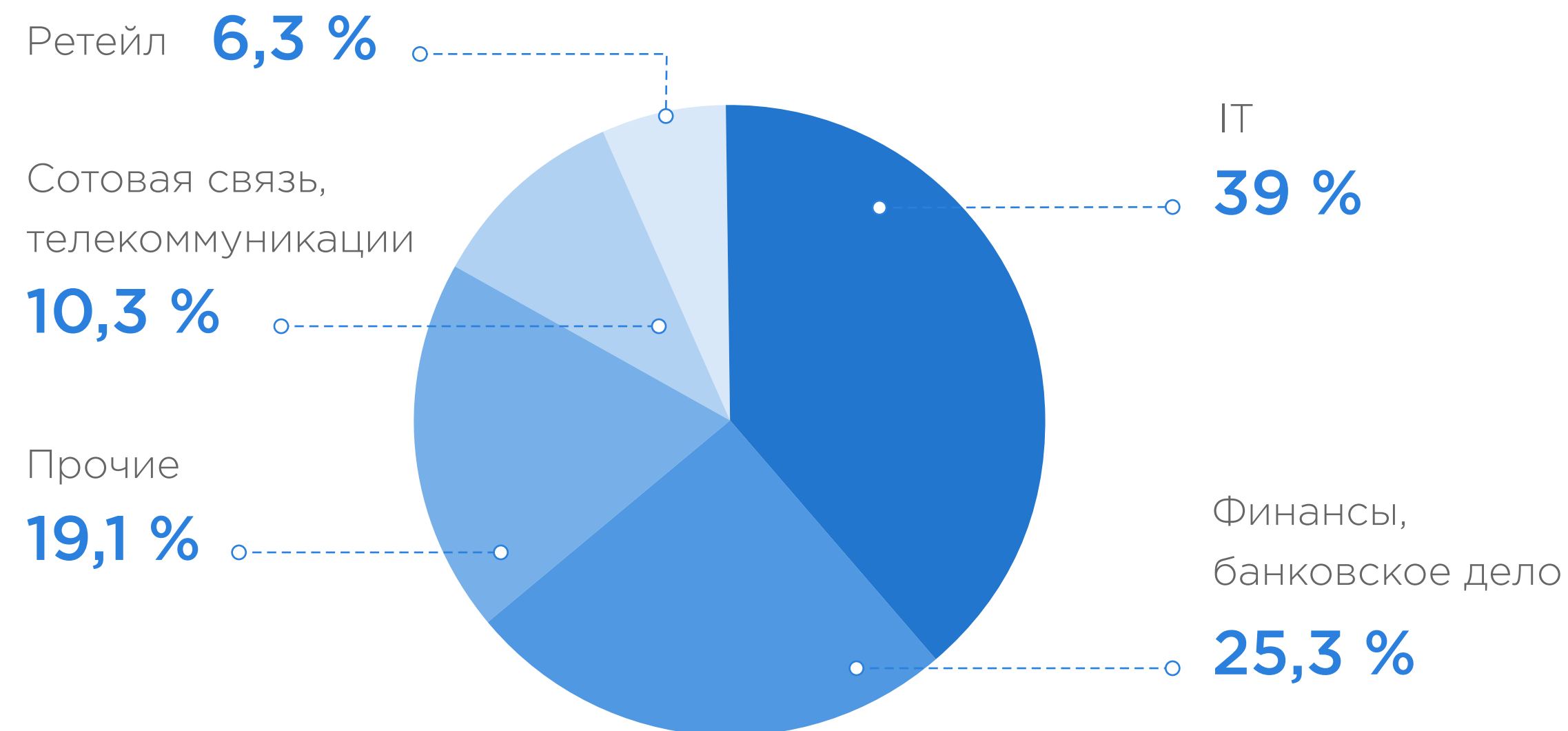
**94,3** % работают в Казахстане

Большинство респондентов работают по найму в офисе



# Индустрии

~ 40 % работают в IT. На втором месте — финансы и банковское дело (около 25 %)



IT, финансы, телеком и госсектор – это основные сферы, в которых в принципе есть данные, а значит, есть и возможность получать от их обработки выгоду для бизнеса в виде сокращения расходов, увеличения доходов или уменьшения рисков.

Могу предположить, что доля специалистов из финансового сектора будет расти активней всего в ближайшие несколько лет.

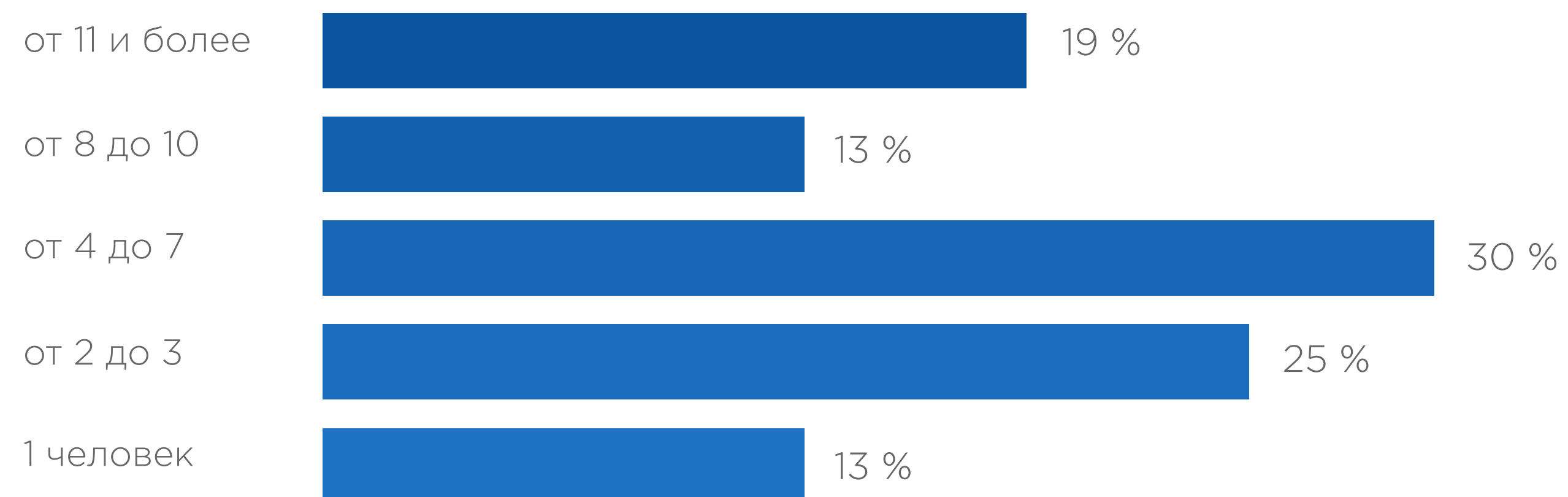
**Пётр Царенко**

Chief Data Officer в Kolesa Group

# Количество человек в отделе

Почти 70 % респондентов работают в небольших командах до 8 человек.

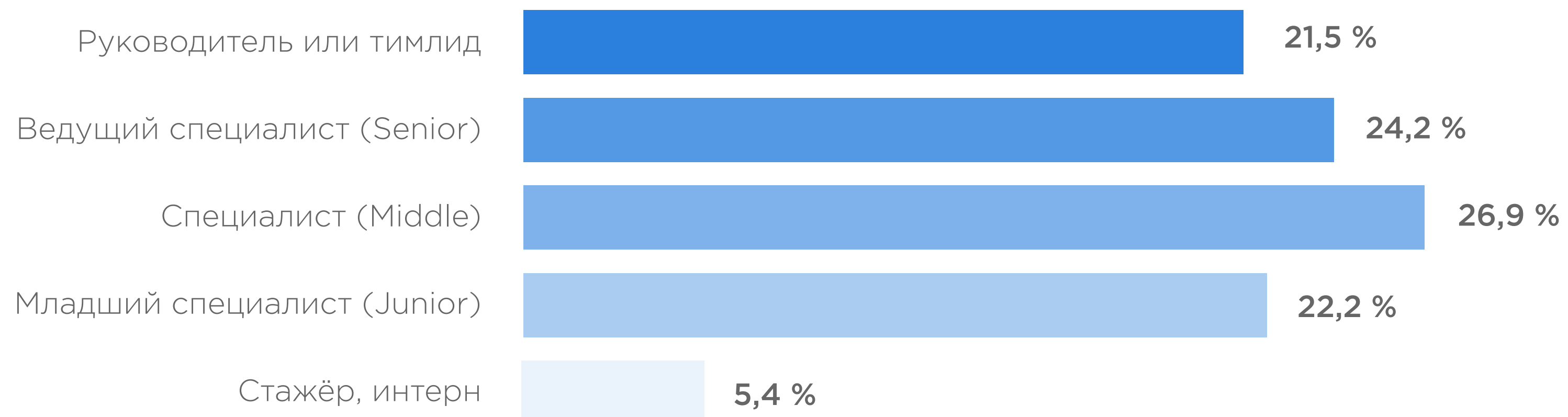
При этом 13 % респондентов указали, что являются единственными сотрудниками по работе с анализом данных в компании



# Уровень должности

Во многих компаниях еще не сформировалась адекватная система оценки: senior или teamlead в одной компании может соответствовать уровню middle в другой.

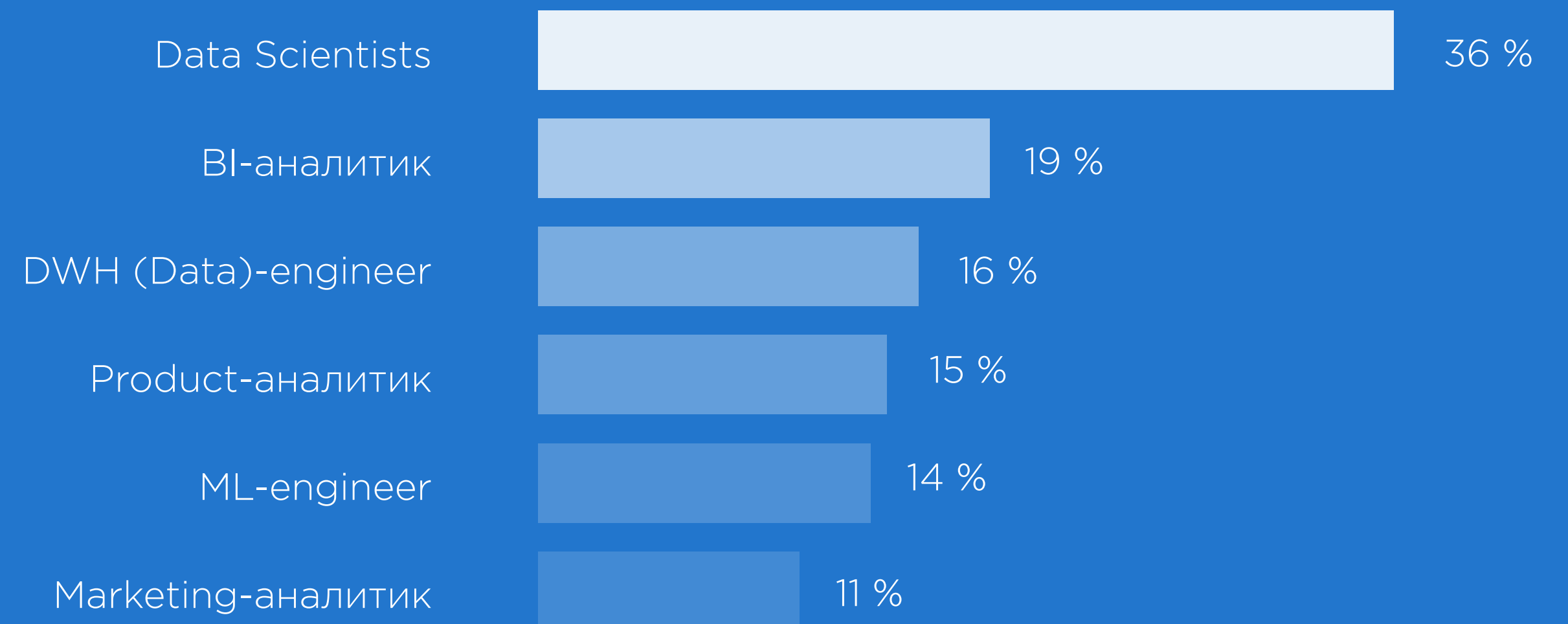
Это подтверждают и ответы респондентов: junior-специалистов почти столько же, сколько руководителей и teamlead-специалистов.



# Специализация

Data Scientist — наиболее распространенная должность среди респондентов (около 36 %). В топ-3 также входят BI-аналитики (19 %) и DWH-инженеры (около 16 %).

**29** %  
совмещают несколько функций



# Откуда приходят в аналитику

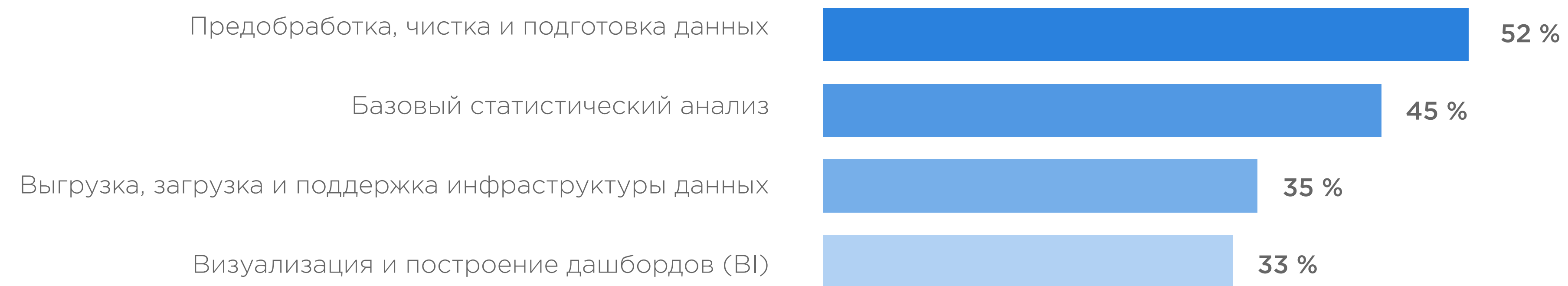
Согласно опросу, более половины респондентов пришли в аналитику и Data Science из менеджмента, экономики, продаж, маркетинга и других нетехнических дисциплин





# Зоны ответственности

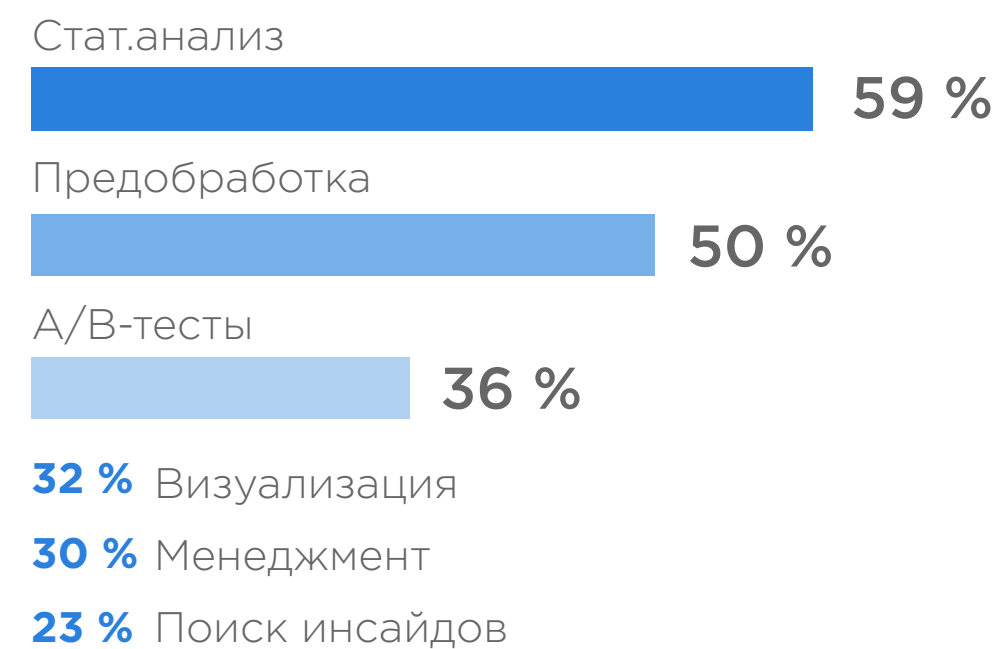
Наиболее популярные рабочие функции среди всех респондентов



# Функции

## В зависимости от специализации

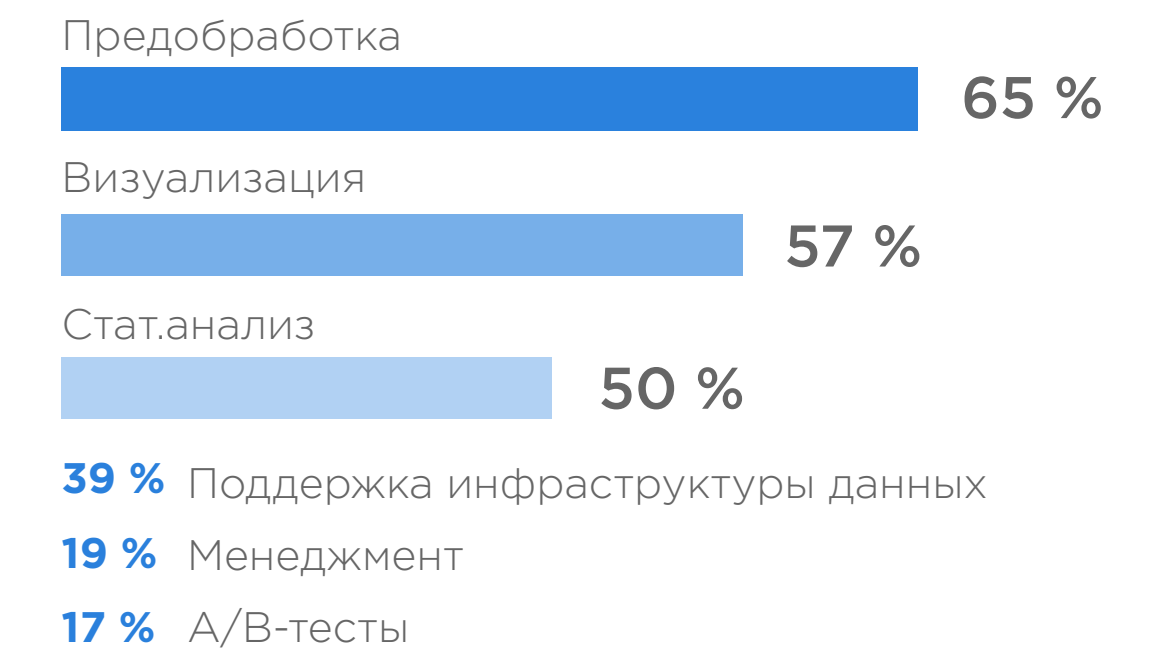
### Product-аналитик



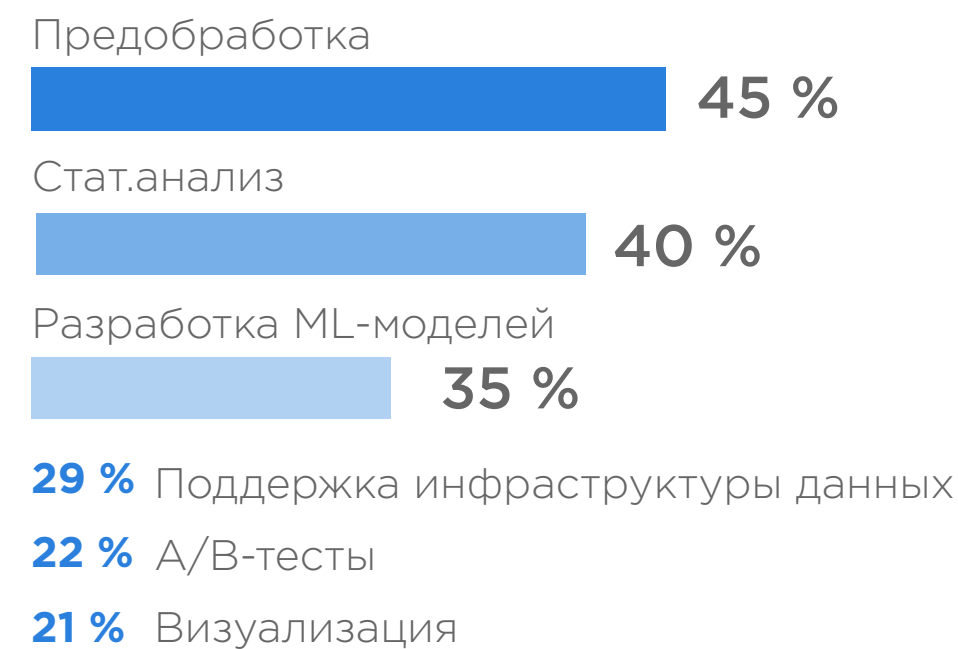
### Marketing-аналитик



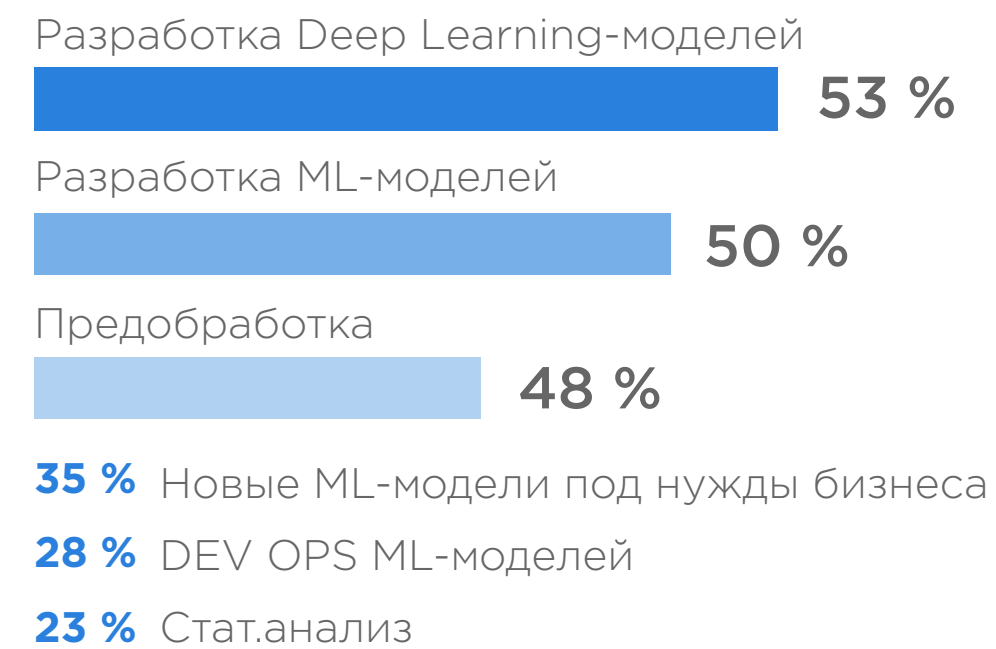
### BI-аналитик



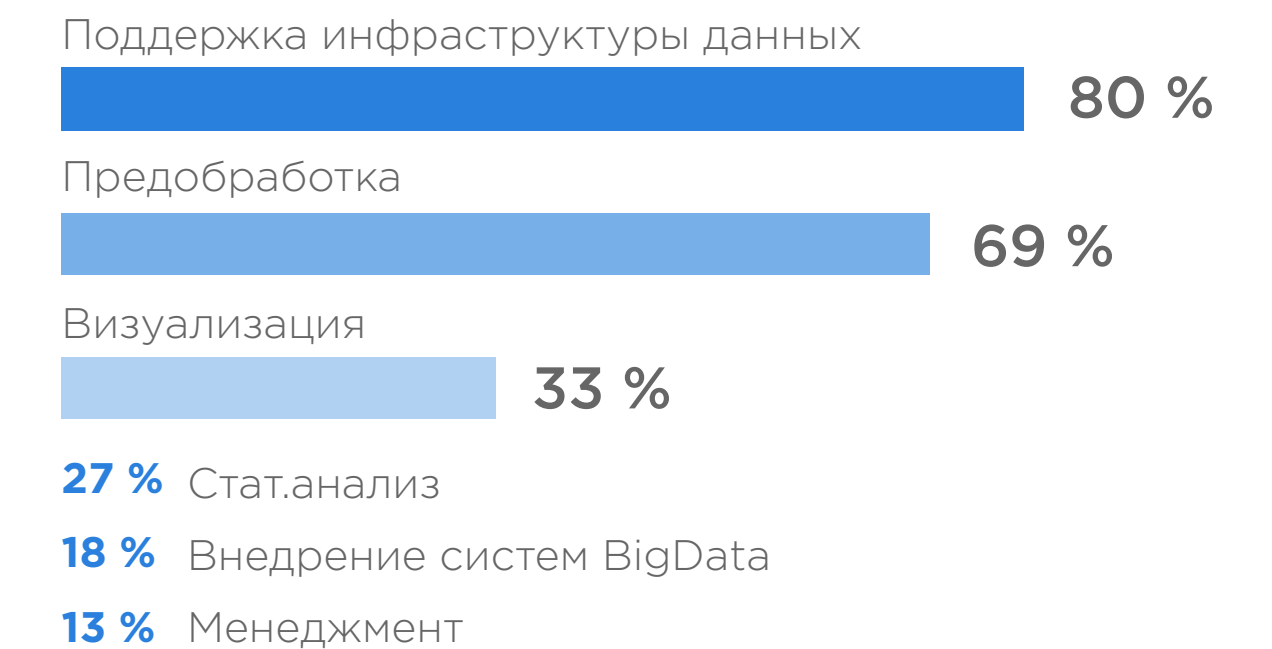
### Data Scientists



### ML-engineer



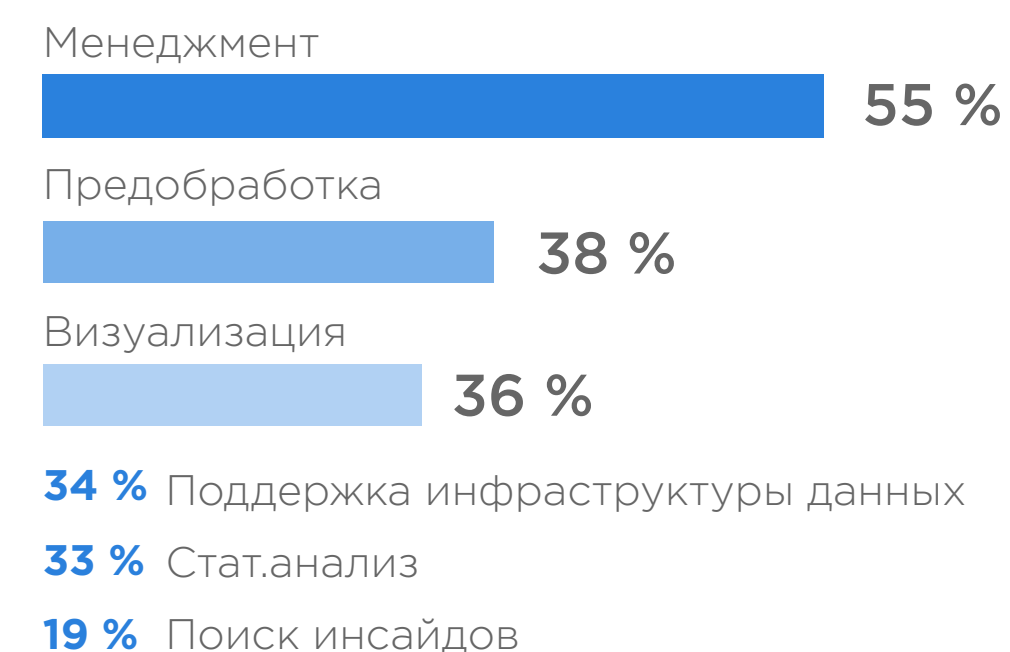
### DWH (Data-engineer)



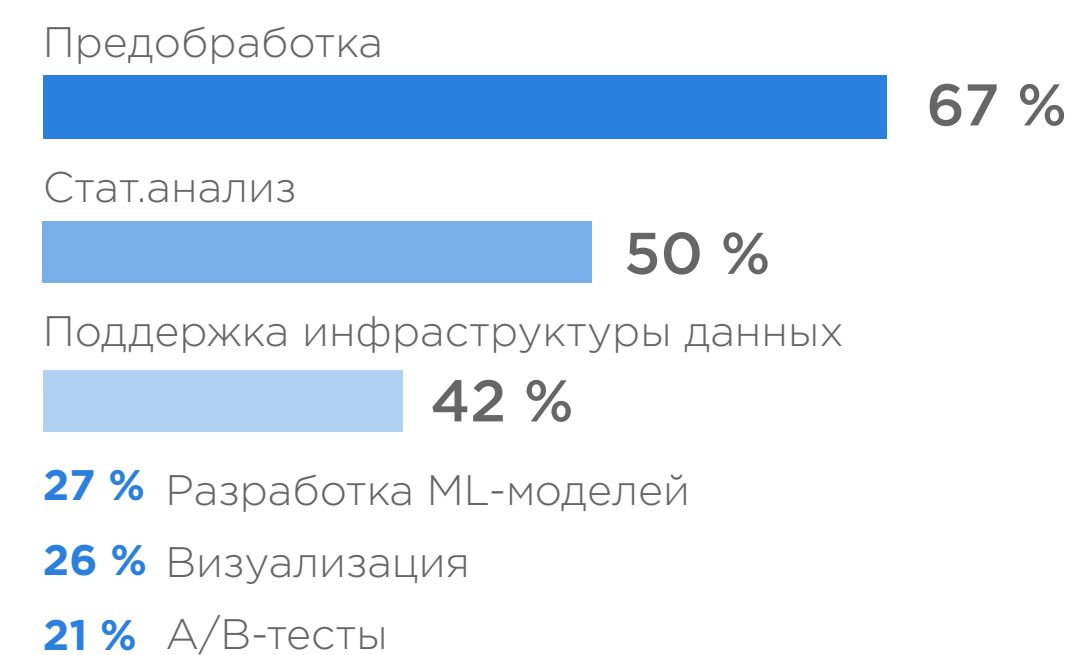
# Функции

## В ЗАВИСИМОСТИ ОТ ДОЛЖНОСТИ

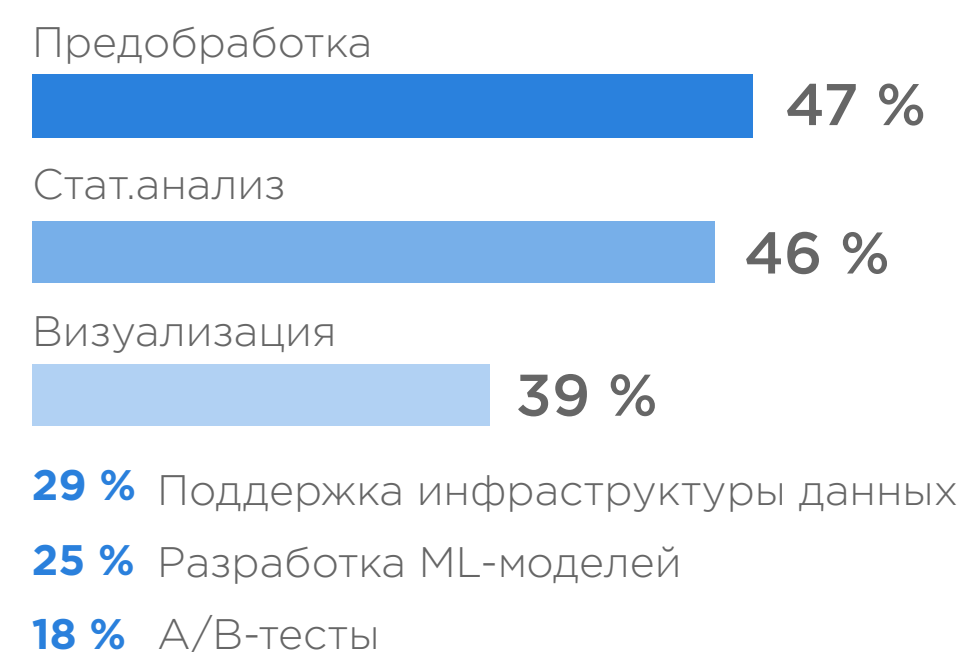
### Руководитель или тимлид



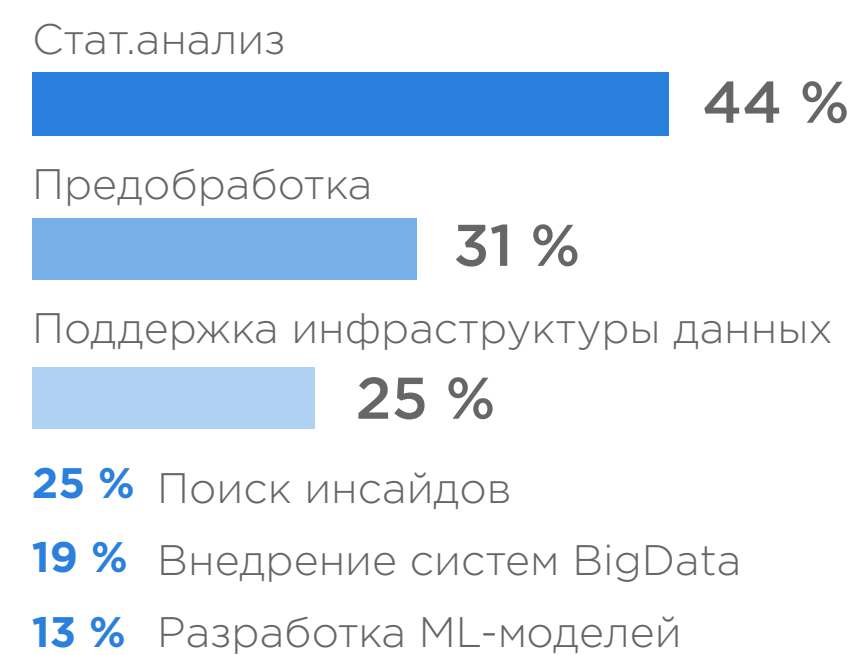
### Junior



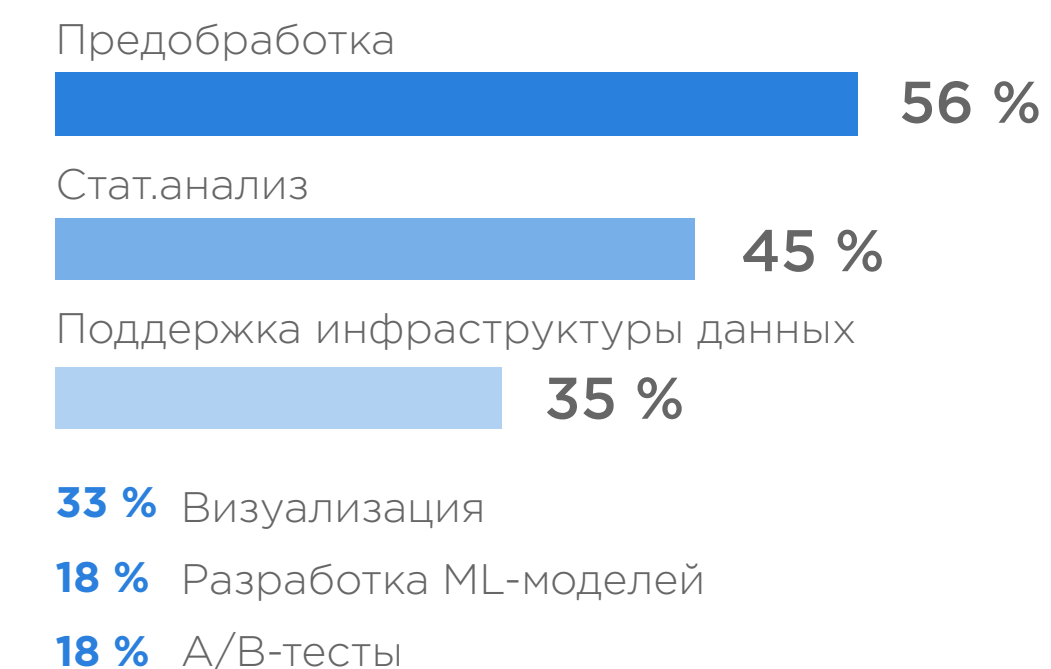
### Senior



### Стажёр, интерн



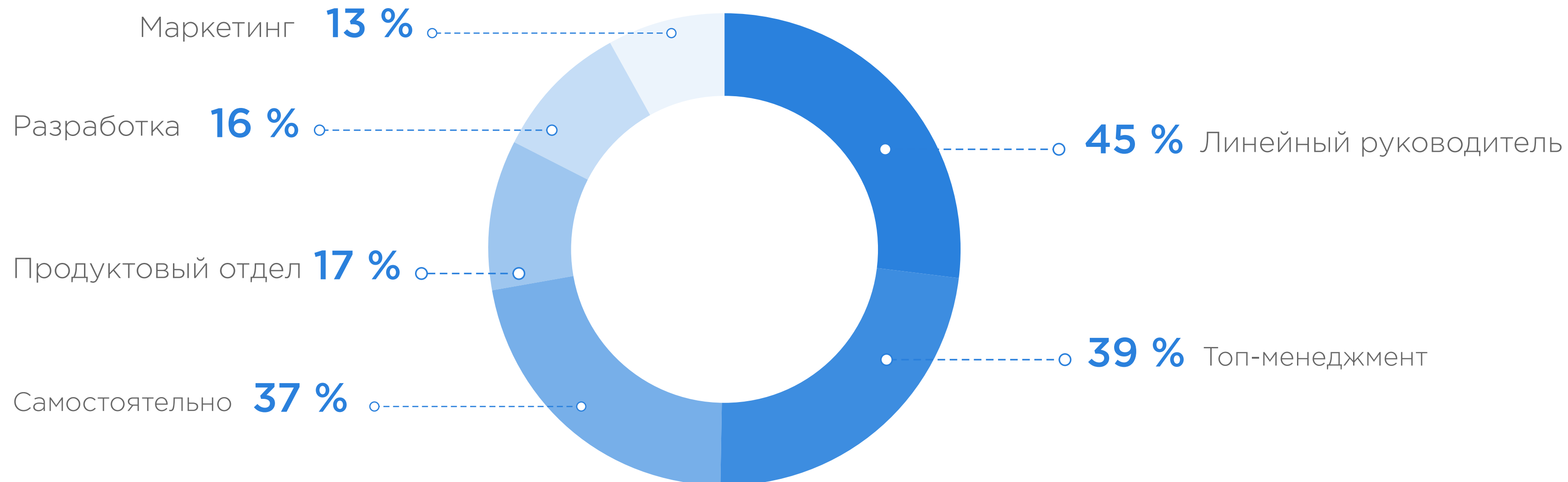
### Middle

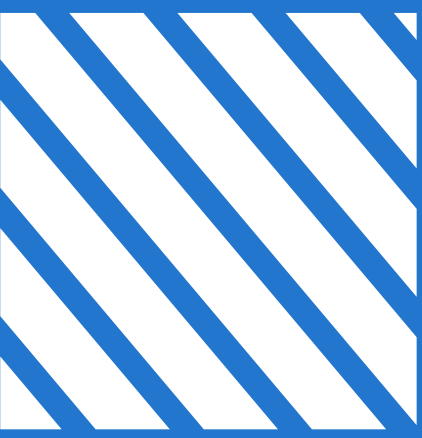




# Источники постановки задач

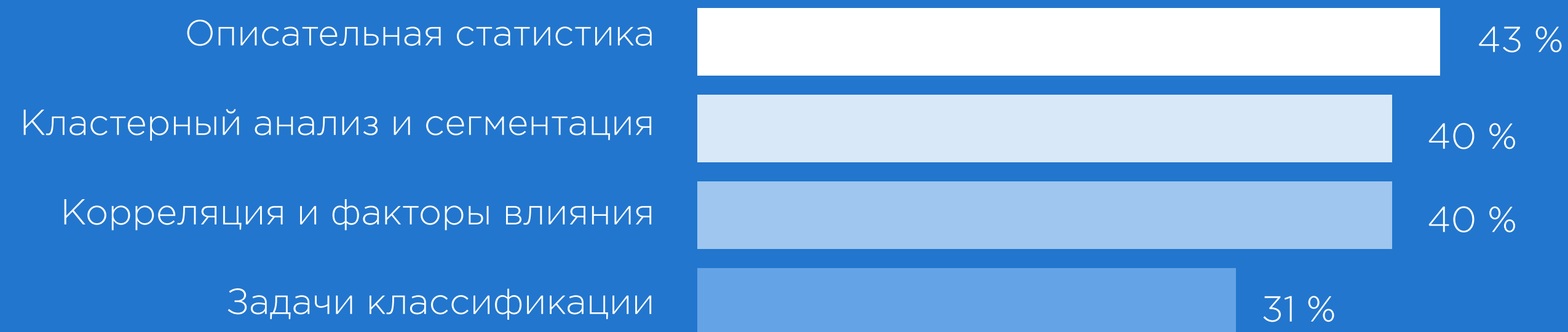
Примечательно, что около трети участников исследования отметили, что ставят себе задачи самостоятельно. Возможно, это связано с тем, что не всегда и не везде бизнес точно знает, чего он ждёт от аналитики и Data Science





# Типы анализа данных

Популярными среди респондентов также оказались A/B-тестирование и дисперсионный анализ. Эти методы всё чаще применяются в казахстанских компаниях для оценки нововведений. Наиболее инновационные глубокие нейросети упомянули 19 % респондентов



- 30 %** проверка гипотез, A/B-тестирование, дисперсионный анализ
- 28 %** анализ временных рядов и прогнозирование на их основе
- 26 %** решение регрессионных задач (классический ML)
- 26 %** поиск аномалий в данных, антифрод
- 19 %** задачи классификации (Deep Learning и нейросети)
- 16 %** EDA
- 11 %** понижение размерностей в данных

**— Согласно опросу, наиболее популярные алгоритмы и методики анализа относятся к классическому ML и анализу данных (описательная статистика — 43 %, кластерный анализ — 40 %, корреляционный анализ и поиск факторов влияния — 40 %). Почему именно эти типы анализа наиболее популярны? Почему Deep Learning менее популярен (всего 19 % респондентов)?**

**Думан Уватаев** | Chief Data Officer в Kaspi Bank

### **1. Уровень зрелости компании**

На этапе внедрения подхода data-driven нет острой необходимости использовать сложные методы ML. Как показывает практика, компании, которые ранее не использовали подобные методы работы с данными, получают впечатляющие результаты при помощи базовых ML-алгоритмов. Многие операционные процессы могут быть оптимизированы при помощи достаточно простых методов интеллектуального анализа данных.

Внедрение и промышленная эксплуатация сложных классов алгоритмов машинного обучения может потребовать много времени, высокого уровня компетенции у задействованных в процессе сотрудников и титанических усилий. Нет смысла строить космолёты, когда можно получить первые результаты по принципу quick win, используя элементарные методы, которые можно внедрить очень быстро.

### **2. Объем анализируемых данных**

Любая компания генерирует данные, но не каждая может похвастаться большими объемами данных, а даже при наличии большого объема данных — готовностью их хранить, анализировать и использовать в работе. Отсутствие необходимой инфраструктуры также может ограничивать возможности аналитиков во внедрении сложных решений.

### **3. Уровень компетенции сотрудников по работе с данными.**

При отсутствии приемлемого уровня компетенций нет возможности перейти к использованию более продвинутых методов. Именно поэтому многие компании поддерживают ранее известные подходы в процессах принятия решений на основании данных.



**Нет смысла строить  
космолёты, когда можно  
получить первые  
результаты по принципу  
quick win, используя  
элементарные методы,  
которые можно внедрить  
очень быстро**

#### **4. Отсутствие экспериментальной базы.**

Чаще всего аналитики заняты текущими задачами и уровень загруженности не позволяет им посвятить время собственному обучению или работе над альтернативными методами решения задач. В результате все придерживаются практически единого стандарта решений задач.

#### **5. Неготовность руководства компании / низкий приоритет / конфликт интересов.**

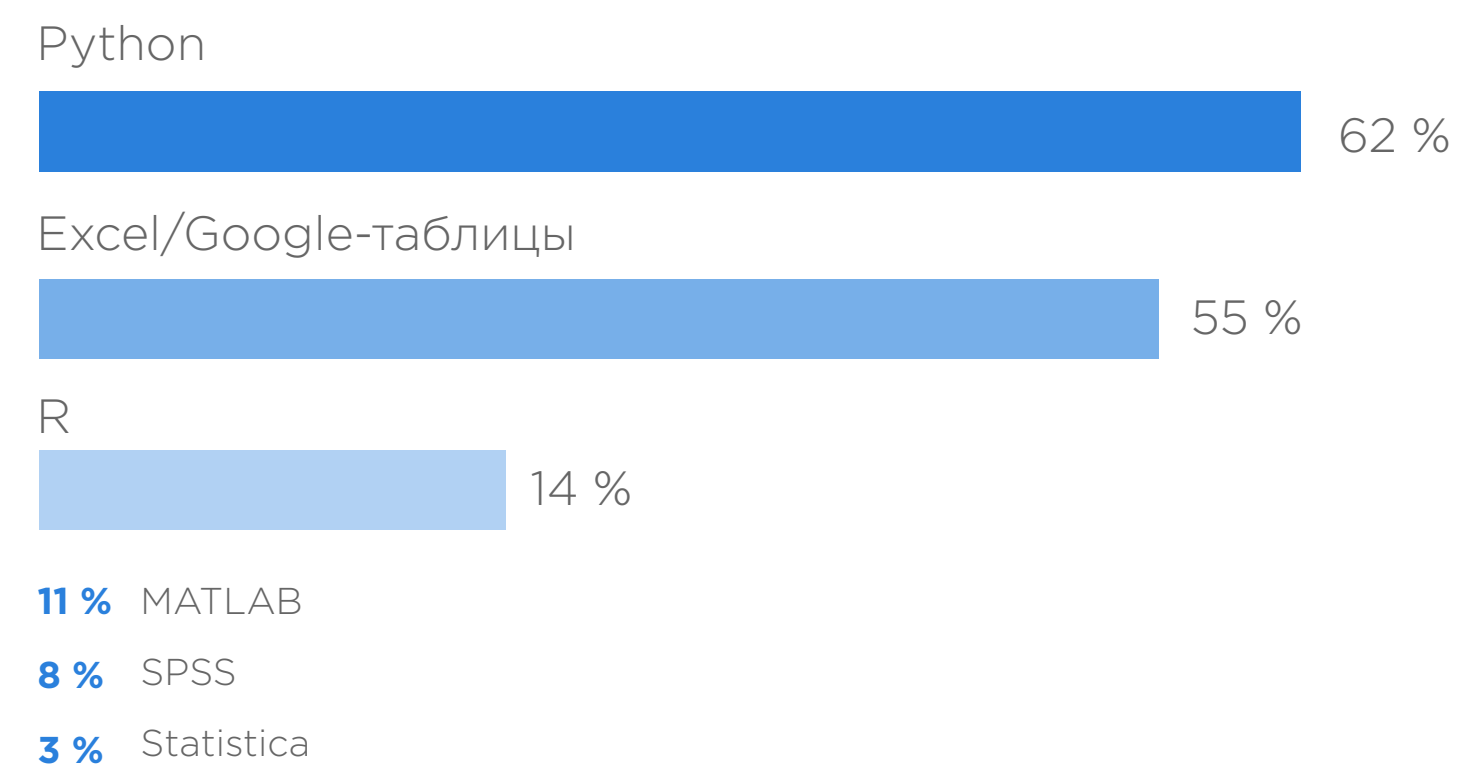
Не каждая компания уделяет должное внимание развитию экспертизы работы с данными, ограничивая роль аналитиков в процессе принятия управленческих решений. Неготовность или нежелание руководства может быть ключевым ограничением, которое не позволяет развить экспертизу и, как следствие, не получить значимых результатов. Следует отметить, что, помимо ограничения со стороны руководства, речь может идти о неготовности владельцев бизнес-процессов, деятельность которых могла бы быть улучшена с применением методов интеллектуального анализа данных. Всесторонняя оценка бизнес-процессов с выделением неэффективных действий/решений может быть весьма болезненной для указанных сотрудников, и именно они могут сопротивляться изменениям.

# Языки программирования и инструменты

## Языки программирования



## Математические и статистические инструменты



Python — общедоступный, абсолютно бесплатный и быстро развивающийся язык программирования. Он обладает четким и последовательным синтаксисом, продуманной модальностью и масштабируемостью, благодаря чему исходный код написанных на Python программ легко читается. Еще одна особенность Python — наличие большого количества готовых библиотек для решения самых разных типов задач. Python популярен и среди гигантов по работе с данными: Spotify, Amazon, Google, Netflix, YouTube и Instagram. Но Python не панацея. Для создания полномасштабных промышленных решений его следует комбинировать с другими предложениями от крупных поставщиков.

**Думан Уватаев**

Chief Data Officer в Kaspı Bank



# Инструменты для визуализации

Специализированная программа Power BI от Microsoft почти в три раза обогнала ближайших конкурентов — Qlik и Tableau

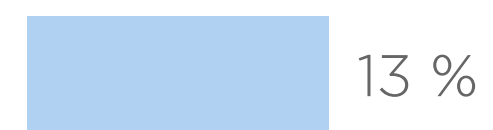
Excel/Google-таблицы



Power BI



Tableau



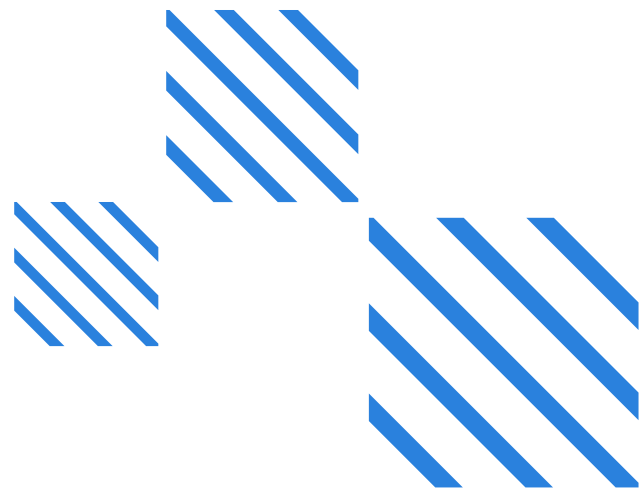
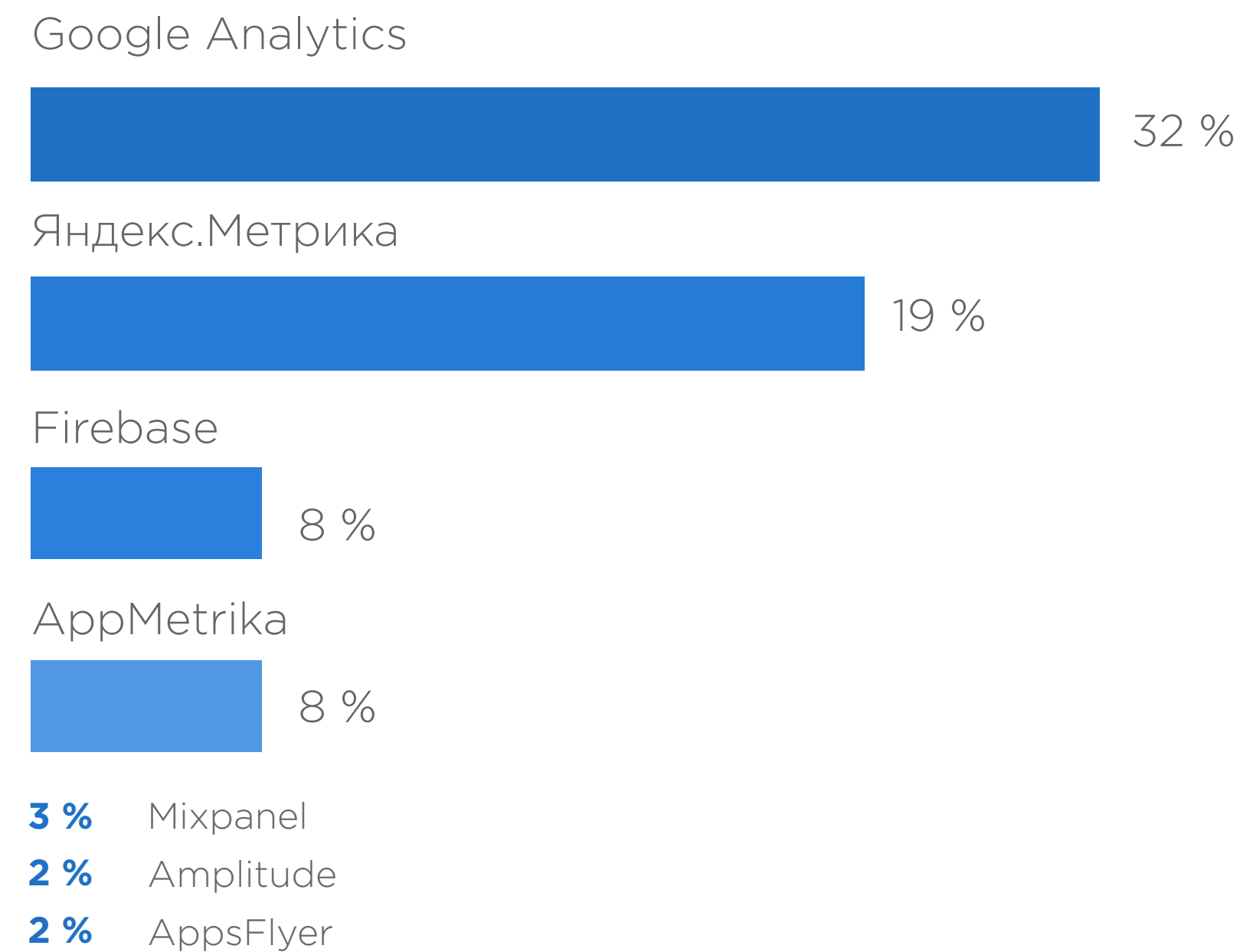
11 % QlikView/Qlik Sense

5 % OLAP

**47** % используют традиционные Excel или Google-таблицы

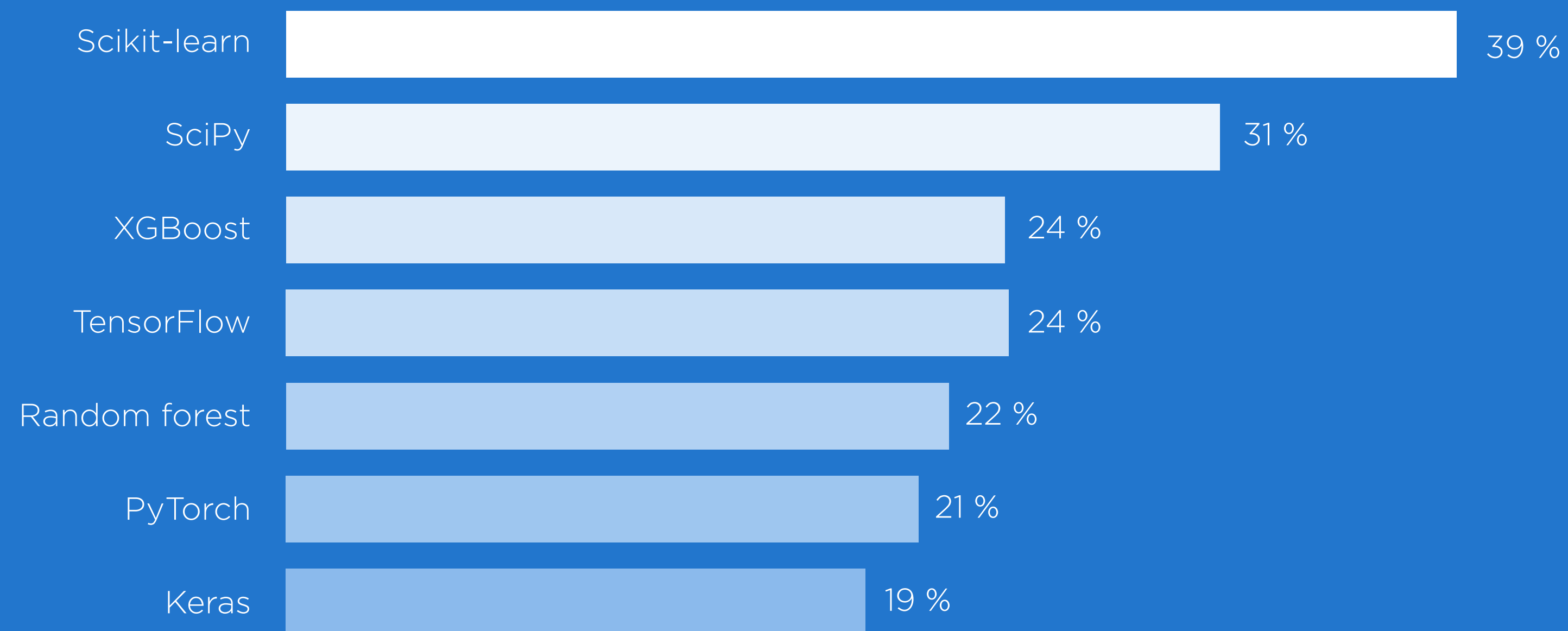
# Системы веб- и мобильной аналитики

Наибольшей популярностью среди респондентов пользуются системы веб-аналитики Google Analytics (32 %) и Яндекс.Метрика (19 %). Системы мобильной аналитики (AppMetrica, AppsFlyer, FireBase) менее популярны, их отметили около 8 % респондентов



# Библиотеки для машинного обучения

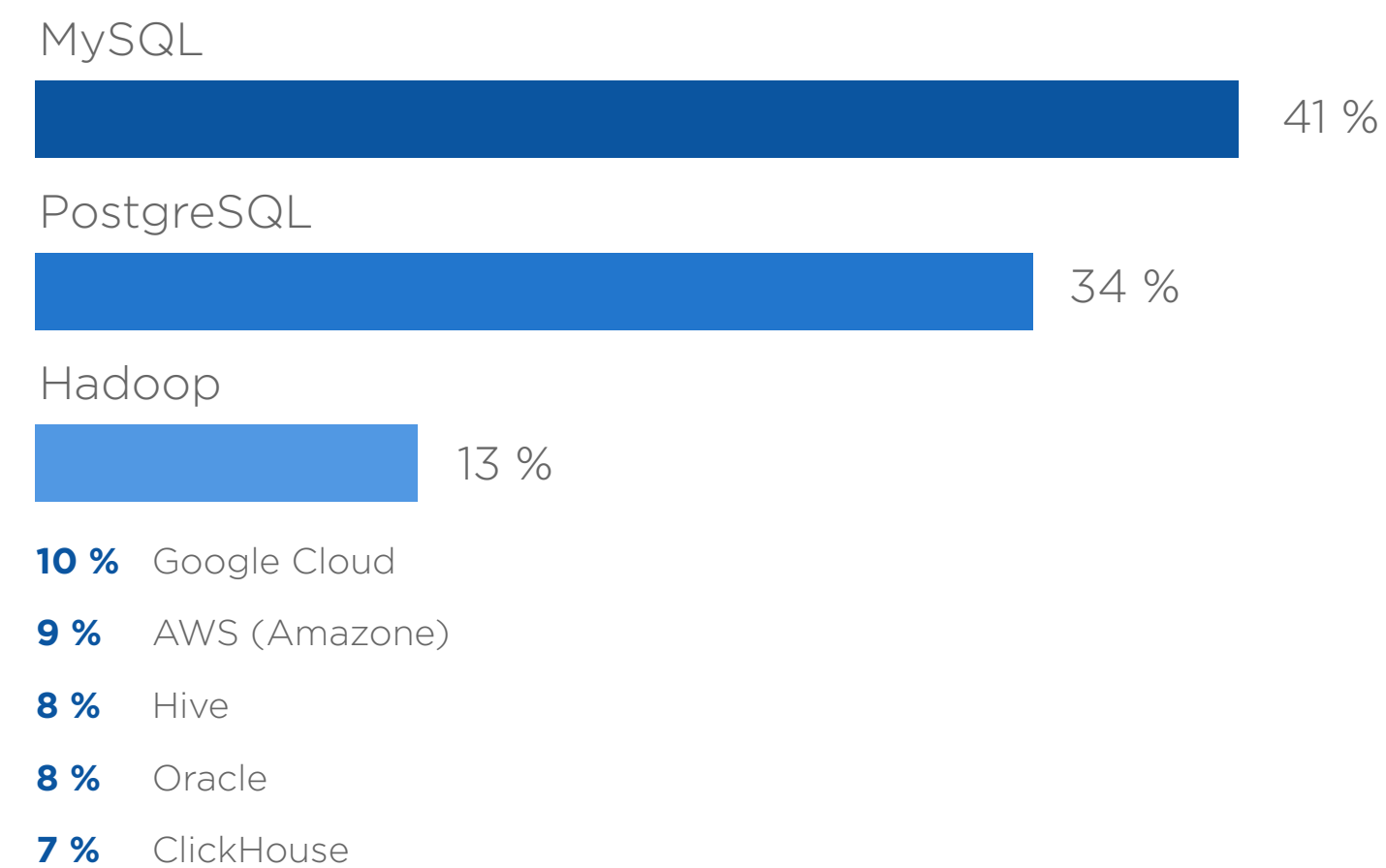
Наиболее популярными библиотеками оказались Scikit-learn и SciPy для Python (39 и 31 % ответов соответственно)



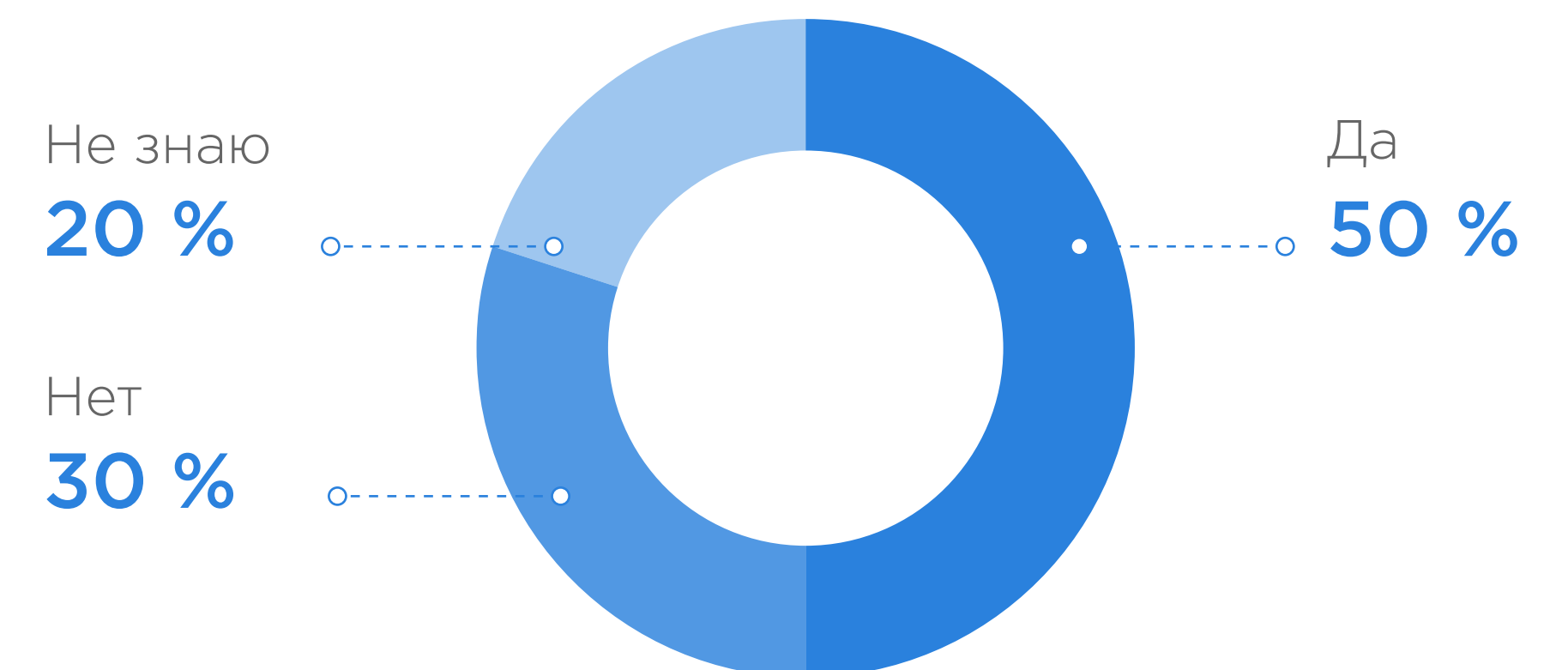
**45** % не используют библиотеки

# Системы управления базами данных (СУБД)

Наиболее популярной СУБД оказалась MySQL (41 %).  
На втором месте — PostgreSQL (34 %). Облачные системы от Google или Amazon Web Services (AWS) менее популярны



## Наличие Data Warehouse (DWH)



# Системы хранения данных

Интересно, что о DWH говорят около половины респондентов, но при этом около 48 % хранят данные в Excel и других spreadsheet — таблицах. Возможно, это связано с тем, что во многих компаниях большая часть данных все-таки представлена в формате small-data.



Среди тех, кто отметил использование облачных сервисов, более популярны решения от Google Cloud

55 %



38 %



19 %

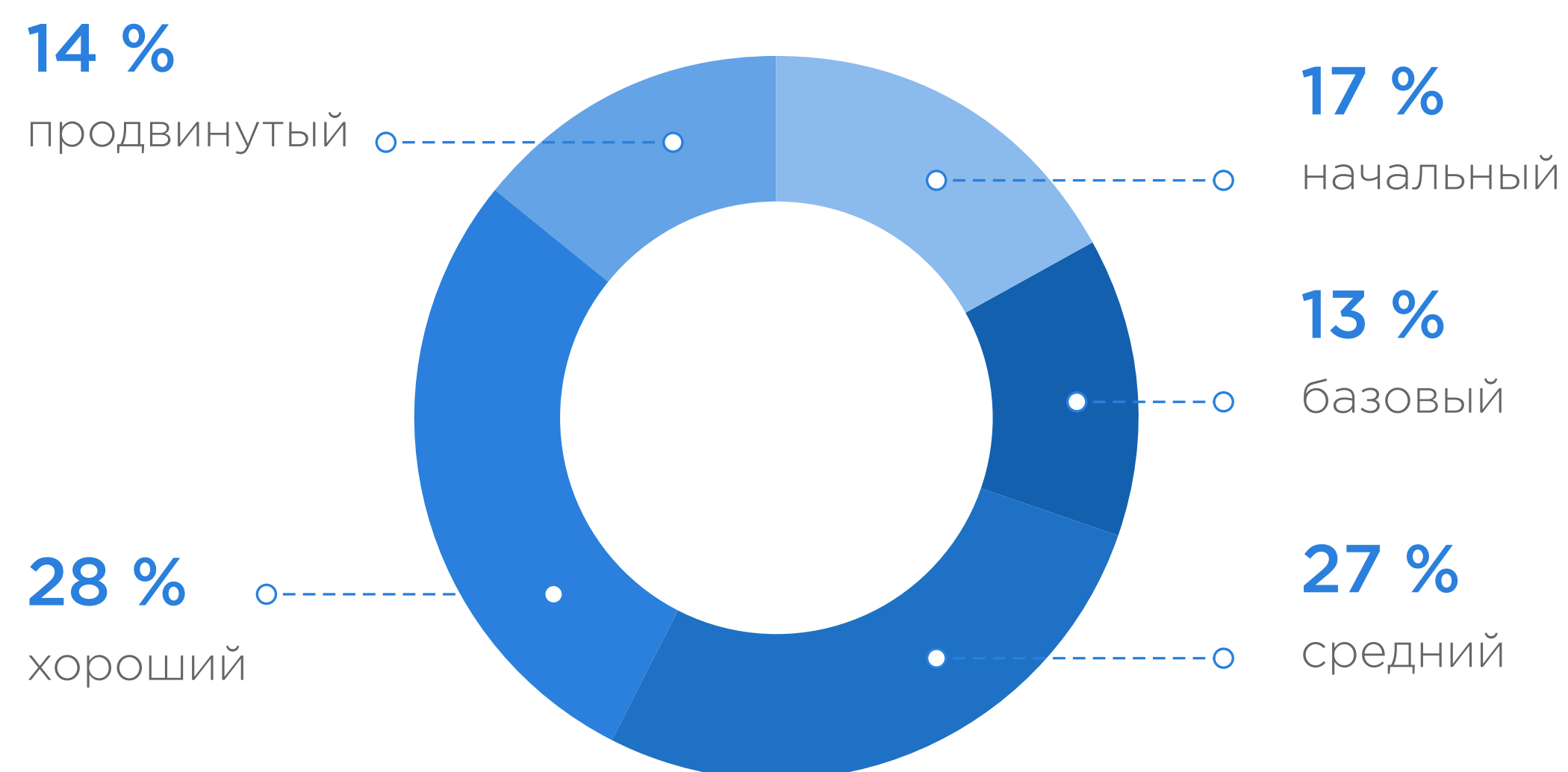


17 %



# Уровень аналитики в вашей компании

Оценки в среднем были выше среди BI-аналитиков, Data Science-специалистов, ML-инженеров и DWH-инженеров. Возможно, команды, в которых есть такие специалисты, работают на более высоком уровне и выполняют более сложные задачи



**Продвинутый уровень.** BI-система работает отлично. Есть DWH и Big Data. Регулярно проводим A/B-тесты. Есть работающие системы ML и DS в production. Решения принимаются только по данным. Отдел работы с данными и Data Science – один из ключевых в компании.

**Хороший уровень.** Периодически проводим сложные Data Science-исследования бизнес-проблем, пробуем внедрять ML-алгоритмы в бизнес. Иногда проводим A/B-тесты. Внедряем предиктивную аналитику. Активно внедряем Data-Driven-подход в управлении.

**Средний уровень.** Простые регулярные отчеты в BI. Простые разовые исследования по требованию. Один или несколько аналитиков.

**Базовый уровень.** Простые отчеты в Excel. Но хочется большего.

**Начальный уровень.** Аналитика в зачаточном состоянии.



## — 55 % респондентов отметили хороший и средний уровень развития аналитики и Data Science в своих компаниях. Вы согласны с такой оценкой?

**Ануар Аймолдин** | Руководитель отдела ИИ в BTS Digital

— Это значит, что больше половины респондентов устраивает уровень развития Data Science и аналитики в их компаниях, что не может не радовать. Отмечу, что приведенная классификация уровней развития аналитики и DS оценивает по большей части инфраструктуру системы аналитики. Например, для AI-driven стартапов это совсем нерелевантно, что также могло слегка пессимизировать результаты опроса.

**Айбек Ергожаев** | Операционный директор Frontier KZ

— Тяжело не согласиться, когда аудитория респондентов, скорее всего, состоит из аналитиков компаний, на практике знакомых с DS. Однако если говорить об уровне развития рынка, то мы наблюдаем уровень продвинутой аналитики только в крупных компаниях с традиционно поставленной экосистемой данных, в частности в банках, телекоме и международных FMCG.



# Уровень аналитики и Data Science в зависимости от величины отдела

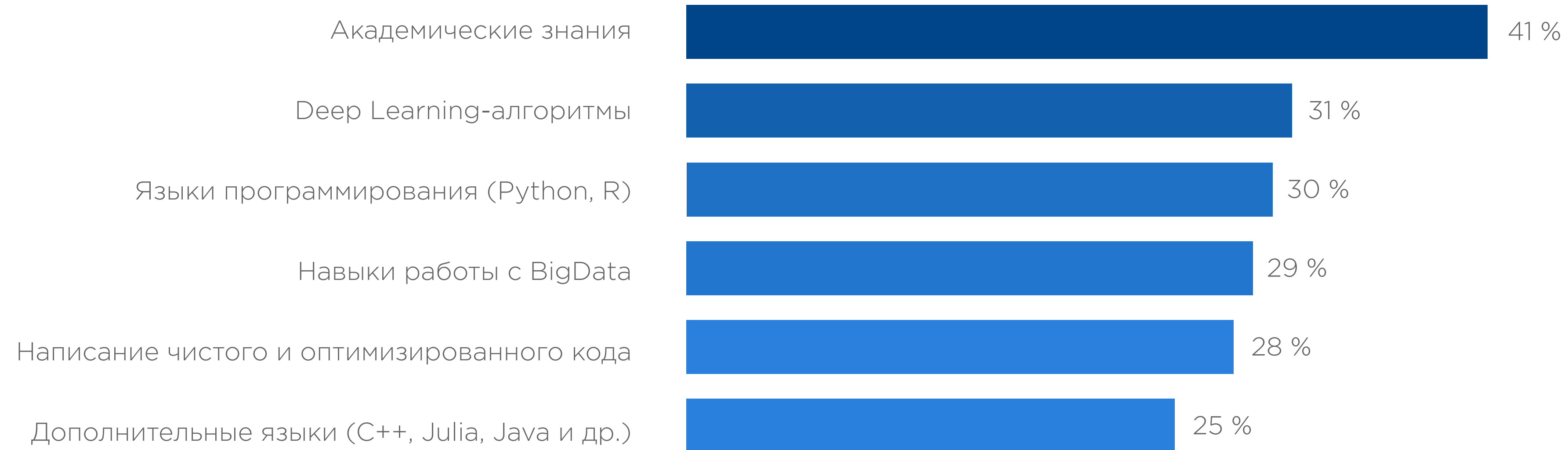
Есть корреляция между размером отдела аналитики/Data Science и оценкой уровня компании в этой области. Так, при отделе свыше трёх человек резко увеличивается доля ответов на хорошем уровне. В отделе из восьми человек и более чаще появляется продвинутый уровень в ответах

	1 человек	от 2 до 3	от 4 до 7	от 8 до 10	от 11 и более
Начальный уровень	<b>44 %</b>	<b>23 %</b>	4 %	<b>13 %</b>	6 %
Базовый уровень	<b>25 %</b>	14 %	<b>13 %</b>	8 %	9 %
Средний уровень	<b>22 %</b>	<b>34 %</b>	<b>38 %</b>	<b>21 %</b>	<b>13 %</b>
Хороший уровень	6 %	17 %	<b>33 %</b>	<b>34 %</b>	<b>45 %</b>
Продвинутый уровень	0 %	10 %	11 %	<b>21 %</b>	<b>25 %</b>



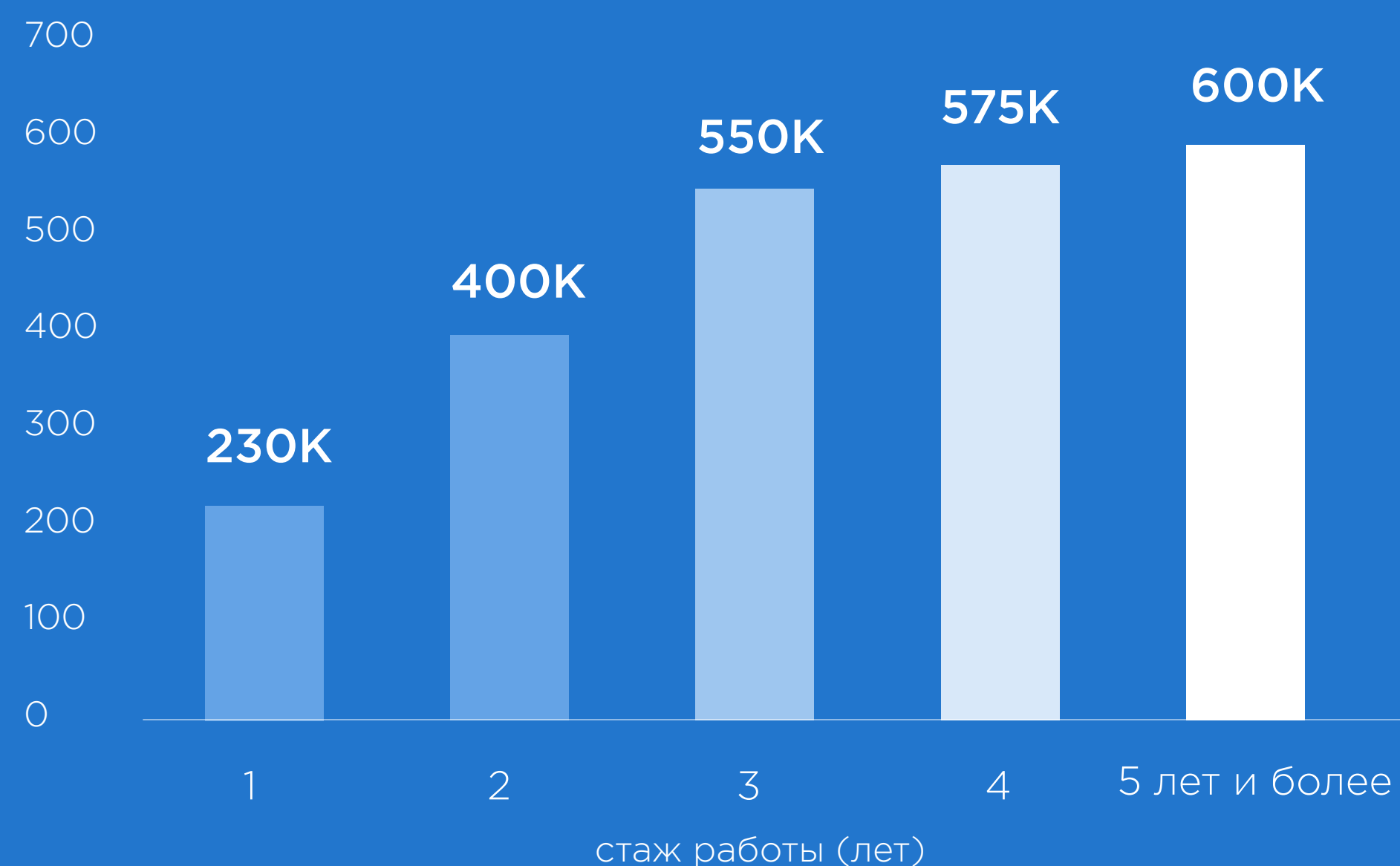


# Недостающие навыки и знания



При ответе на вопрос о недостатках знаний и навыков в основном респонденты говорят о недостатке технических знаний и скиллов. Наибольший недостаток навыков и знаний респонденты ощущают в академических знаниях (41 % респондентов). Около трети респондентов считают, что недостаточно знают алгоритмы Deep Learning. Еще 30 % респондентов отмечают недостаточное знание языков программирования.

# Медианная заработная плата в зависимости от стажа работы



- Для графика мы использовали медианную зарплату, так как она лучше характеризует выборку с большим разбросом и выбросами.
- Специалистов с опытом от 5 лет и более мы объединили в одну группу из-за небольшого количества респондентов

Data Science – молодое развивающееся направление. Поэтому специалисты со стажем от трех лет уже считаются опытными сотрудниками уровня Middle или Senior. Медианная зарплата молодых специалистов без опыта почти на 30 % выше, чем средняя по стране. Работая в сфере анализа данных, за первые три года можно увеличить заработную плату в два с половиной раза

**— Почти 80 % респондентов работают в анализе данных три года или менее. Действительно ли наш рынок такой молодой?**

**Айбек Ергожаев** | Операционный директор Frontier KZ

— Если говорить в целом, о DS, ML и BigData, то казахстанский рынок все еще молод, устойчивый курс на развитие анализа и экосистемы данных среди крупных компаний заметен лишь последние три года, и то в определенных сегментах. Банки активно внедряют продвинутую аналитику, в частности в скоринге и апсейлах, не отстает и телеком, активно применяющий технологии анализа больших данных как для внутренней аналитики, так и для внешней монетизации данных. Следом можно отметить транснациональные FMCG и нефтяные компании. Гос- и квазигоссектор начали активно развиваться последние 1-2 года. Наблюдаем интересные проекты в строительных компаниях. Если говорить о добывающей промышленности, логистике, дистрибуции, страховании и других сферах, то рынок однозначно очень молод.

**— Как Вы считаете, более высокая доля задач, связанных с описательной статистикой, а также с корреляционным и кластерным анализом (43 %, 40 %, 40 % респондентов соответственно), у респондентов соответствует мировым трендам и особенностям? Или это больше особенности казахстанского этапа развития?**

**Ануар Аймолдин** | Руководитель отдела ИИ в BTS Digital

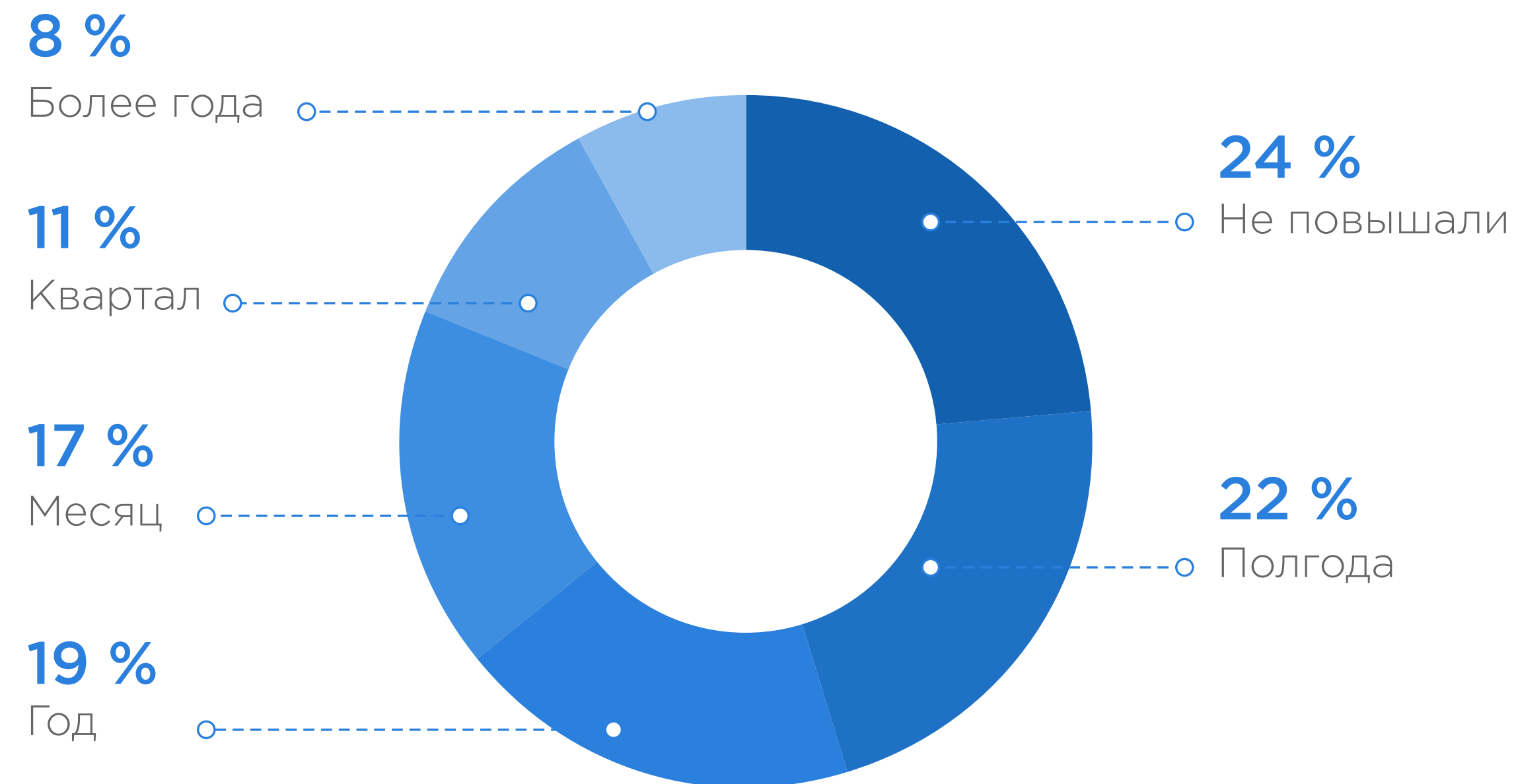
– Я думаю, что это больше история не про уровень развития, а про задачи, которые решают респонденты. Например, вышеприведенными методами могут решаться весьма распространенные у нас задачи клиентской аналитики.



# Последнее повышение зарплаты

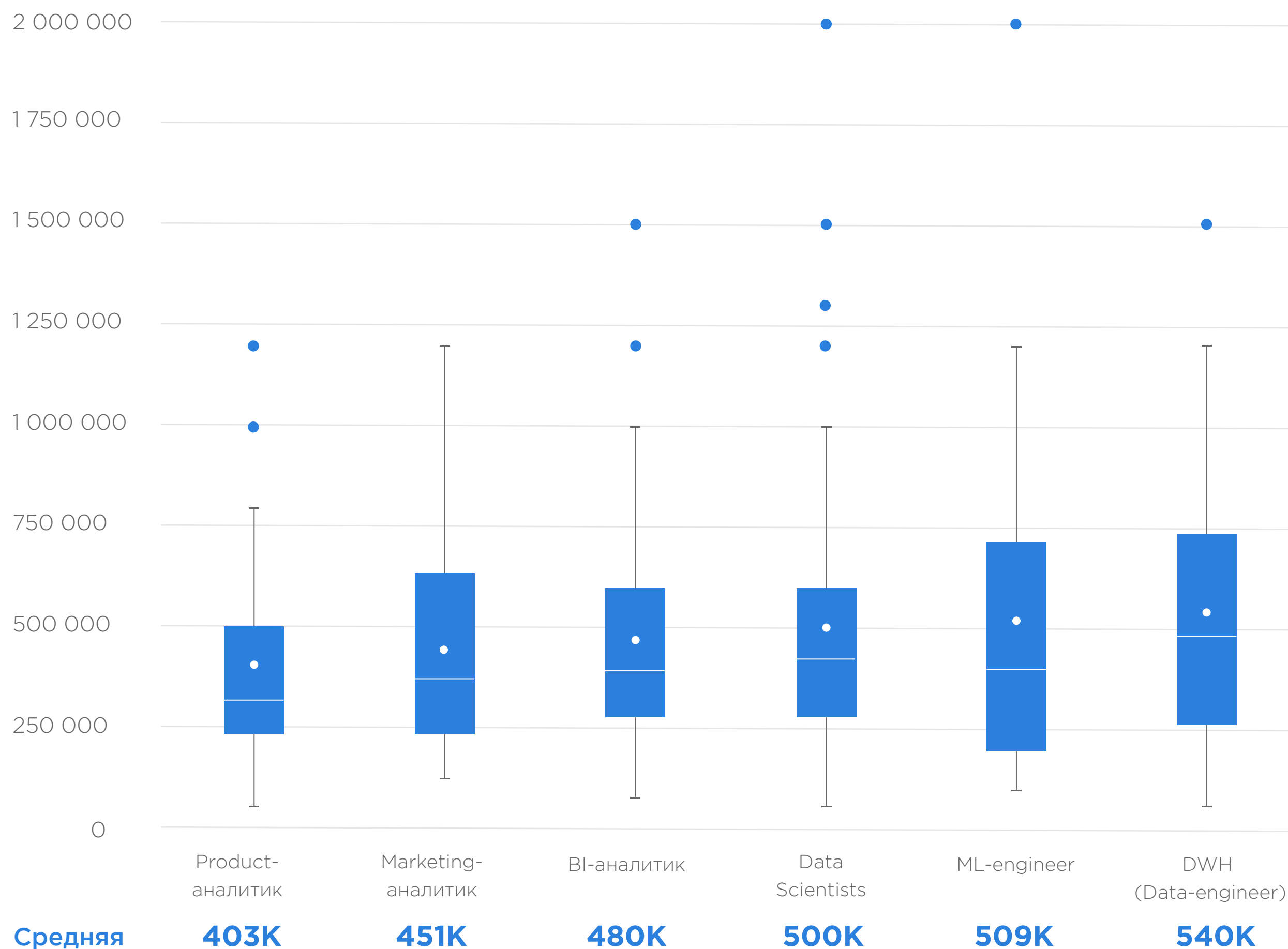
Около 69 % респондентов получали повышение в течение года.

50 % респондентам повышали зарплату в последние полгода.



# Размер заработной платы

## В ЗАВИСИМОСТИ ОТ ПОЗИЦИИ

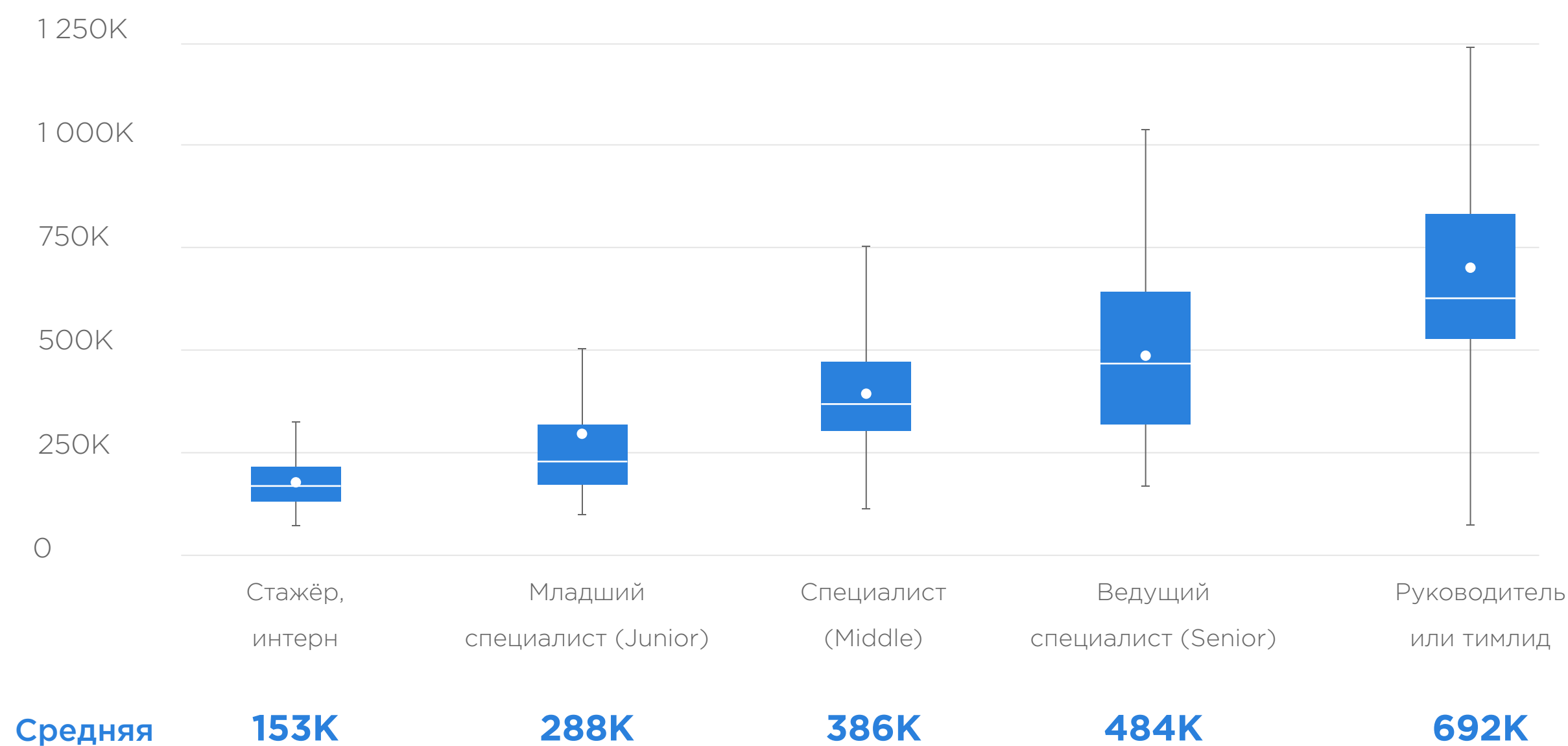


Уровень средней зарплаты (в зависимости от позиции) связан как с требуемыми компетенциями, так и с количеством специалистов на рынке. Компании готовы платить больше редким специалистам, таким как DWH и ML-инженеры.

Рынок труда Data Science в Казахстане находится на ранней стадии развития, поэтому на редкие позиции вроде ML-инженеров и дата-сайентистов можно встретить как очень низкие зарплаты, так и очень высокие.

# Размер заработной платы

В зависимости от уровня должности



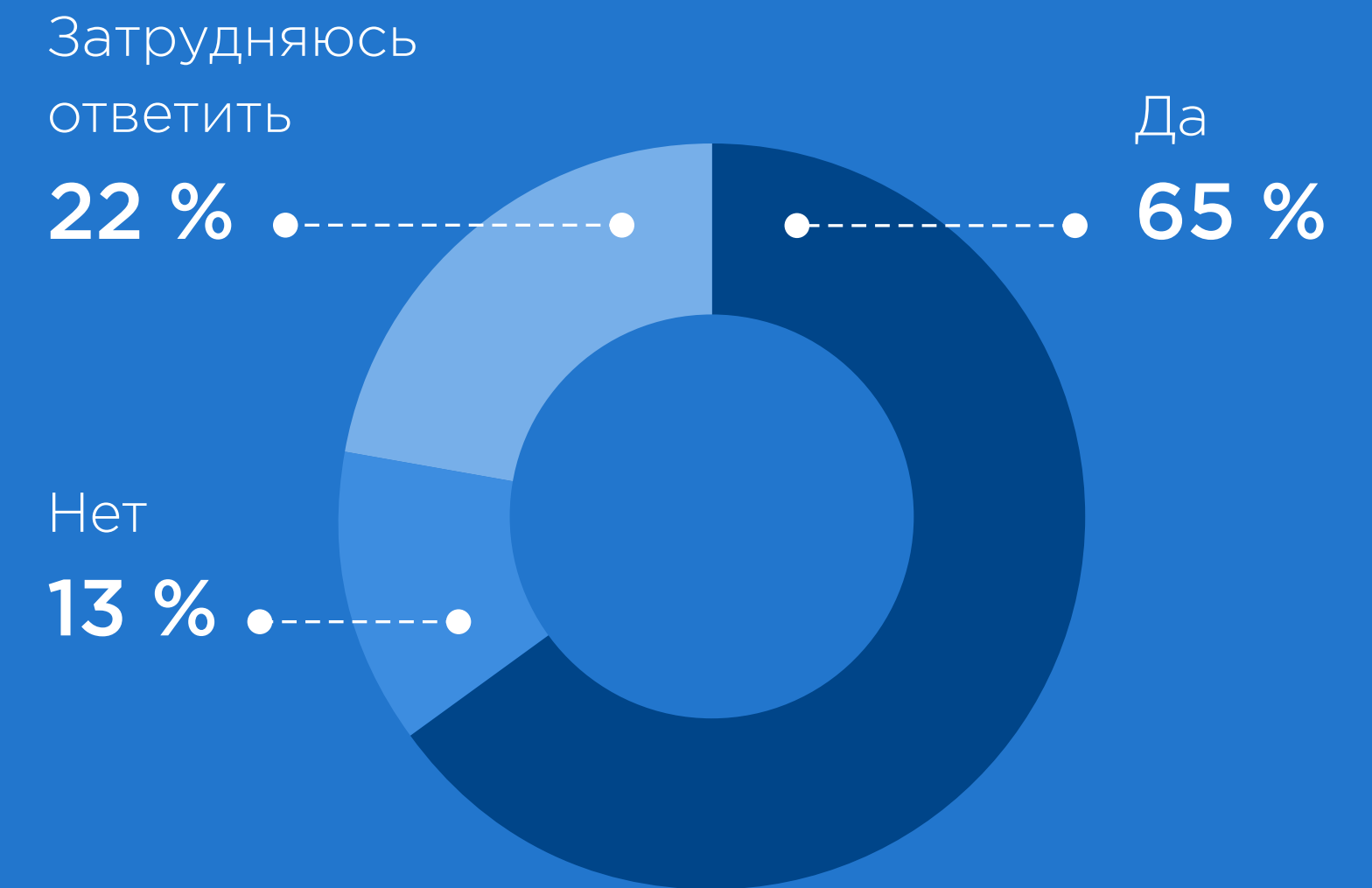
С точки зрения уровня должности самая высокая зарплата у тимлидов и руководителей, в среднем 692 тыс. тенге. Самая низкая — у стажёров, около 150 тыс. тенге.

При этом в каждом из указанных уровней наблюдаются значительные отклонения у единичных специалистов

# Важные пункты при выборе места работы



Готовы ли вы сейчас к смене работы при более выгодном предложении?



Лояльность респондентов невысокая. Почти 65 % готовы к смене работы при более выгодном предложении



# Data Scientist – самая сексуальная профессия XXI века

(c) Harvard Business Review

Данные нашего исследования это подтверждают. Судя по отчету, Data Science – одно из самых растущих и перспективных направлений.

Информация, собранная в отчете, подтверждается нашим опытом: сотней проведенных собеседований, информацией от компаний-партнеров и успехами RnD-отдела Kolesa Group.

Мы приоткрыли занавес и теперь знаем о рынке специалистов данных намного больше. Этими знаниями мы с радостью делимся с вами!