

ДЕПАРТАМЕНТ ОБРАЗОВАНИЯ И НАУКИ ГОРОДА МОСКВЫ

Государственное автономное образовательное учреждение

высшего образования города Москвы

«Московский городской педагогический университет»

(ГАОУ ВО МГПУ)

Институт цифрового образования

Департамент информатики, управления и технологий

Лабораторная работа № 3.1

по дисциплине «Платформы Data Engineering»

Выполнил:

студент группы БД-251м

Направление подготовки/Специальность

38.04.05 - Бизнес-информатика

Быков В.В.

(Ф.И.О.)

Проверил:

доц. к.т.н

(ученая степень, звание)

Босенко Т.М.

(Ф.И.О.)

Москва 2025

## 1. Введение

Цель — пройти полный цикл data-driven исследования: от сбора первичных данных с помощью онлайн-опроса по актуальной теме Data Engineering до создания интерактивного аналитического дашборда для визуализации и интерпретации полученных результатов в Yandex DataLens.

Описание выбранной темы «Облачные решения для работы с данными»:

Облачные решения для работы с данными представляют собой использование удаленных вычислительных ресурсов и сервисов через интернет для хранения, обработки, анализа и управления данными. При создании опроса были включены различные аспекты: выбор облачных платформ компаниями, цели применения облачных решений и др.

## 2. Процесс разработки

### 2.1. Создание опроса

Опрос создан в Google Forms и состоит из 11 вопросов:

- 1 открытый вопрос
- 4 вопроса с одним выбором ответа
- 3 вопроса с использованием шкалы Лайкерта
- 3 вопроса с множественным выбором

### 2.2. Сбор данных

После получения всех ответов от респондентов, происходит выгрузка результатов в формате Excel. До предобработки данных таблица выглядела так:

А	В	С	Д	Е
Отметка времени	Какие функции вы считаете наиболее полезными в облаке	Планируете ли ваша компания расширить использование?	Как часто вы изучаете новые инструменты	Какой тип облачного развертывания вы считаете наиболее подходящим
10.20.2025 21:06:08	Автомасштабирование и управляемые сервисы баз данных	Да	Регулярно	Гибридное облако
10.20.2025 21:07:20	Бессерверные вычисления	Да	Регулярно	Публичное облако
10.20.2025 21:09:30	Инструменты для мониторинга и алертинга	Нет	Редко	Частное облако
10.20.2025 21:10:56	Высокая доступность и отказоустойчивость	Да	Регулярно	Гибридное облако
10.20.2025 21:12:05	Широкий выбор готовых ML-моделей и сервисов AI	Да	Регулярно	Мультиоблако
10.20.2025 21:33:22	Функции автоматического резервного копирования	Затрудняюсь ответить	Редко	Затрудняюсь ответить
10.20.2025 21:37:55	Контейнеризация и оркестрация	Да	Регулярно	Публичное облако
10.20.2025 21:39:20	Гибридные возможности для подключения к локальной инфраструктуре	Да	Редко	Гибридное облако
10.20.2025 21:40:57	Интегрированные среды разработки	Нет	Никогда	Частное облако
10.20.2025 21:42:07	Инструменты для Data Lake и Big Data	Да	Регулярно	Публичное облако
10.21.2025 8:43:14	Платформы как услуга (PaaS), чтобы не управлять инфраструктурой	Да	Регулярно	Мультиоблако
10.21.2025 8:44:19	Сервисы управления идентификацией и доступом	Затрудняюсь ответить	Редко	Гибридное облако
10.21.2025 8:45:51	Возможность быстрого прототипирования и развертывания	Да	Регулярно	Публичное облако
10.21.2025 8:48:02	Глобальная сеть доставки контента	Нет	Редко	Частное облако

После обработки данных таблица выглядит следующим образом:

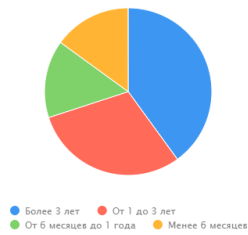
А	В	С	Д	Е
Номер респондента	Какие функции вы считаете наиболее полезными в облаке	Планируете ли расширить использование?	Как часто изучаете новые инструменты	Наиболее подходящий тип облака
1	Автомасштабирование и управляемые сервисы баз данных	Да	Регулярно	Гибридное облако
2	Бессерверные вычисления	Да	Регулярно	Публичное облако
3	Инструменты для мониторинга и алертинга	Нет	Редко	Частное облако
4	Высокая доступность и отказоустойчивость	Да	Регулярно	Гибридное облако
5	Широкий выбор готовых ML-моделей и сервисов AI	Да	Регулярно	Мультиоблако
6	Функции автоматического резервного копирования	Затрудняюсь ответить	Редко	Затрудняюсь ответить
7	Контейнеризация и оркестрация	Да	Регулярно	Публичное облако
8	Гибридные возможности для подключения к локальной инфраструктуре	Да	Редко	Гибридное облако
9	Интегрированные среды разработки	Нет	Никогда	Частное облако
10	Инструменты для Data Lake и Big Data	Да	Регулярно	Публичное облако
11	Платформы как услуга (PaaS), чтобы не управлять инфраструктурой	Да	Регулярно	Мультиоблако
12	Сервисы управления идентификацией и доступом	Затрудняюсь ответить	Редко	Гибридное облако
13	Возможность быстрого прототипирования и развертывания	Да	Регулярно	Публичное облако
14	Глобальная сеть доставки контента	Нет	Редко	Частное облако
15	Встроенные инструменты безопасности и комплаенса	Да	Регулярно	Гибридное облако
16	Сервисы потоковой обработки данных	Да	Регулярно	Мультиоблако
17	Удобный биллинг и детализация расходов	Да	Редко	Публичное облако
18	Managed Kubernetes Service	Затрудняюсь ответить	Редко	Затрудняюсь ответить
19	Serverless базы данных	Да	Регулярно	Публичное облако
20	Выделенные инструменты для миграции данных	Да	Регулярно	Гибридное облако

Были изменены некоторые названия столбцов и добавлен столбец «Номер респондента» для лучшей визуализации результатов.

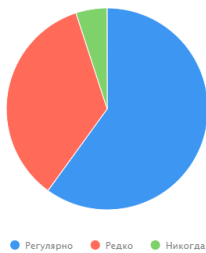
## 2.3. Разработка дашборда

## Облачные решения для работы с данными

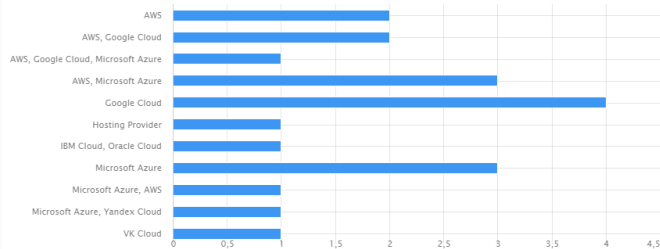
Опыт использования облачных решений



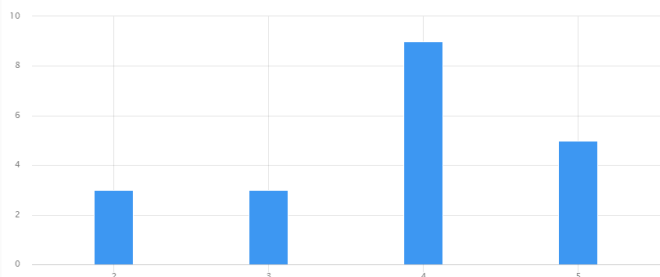
Изучение новых инструментов



Используемые платформы



Удобство интеграции облачных сервисов (оценка)



Таблица

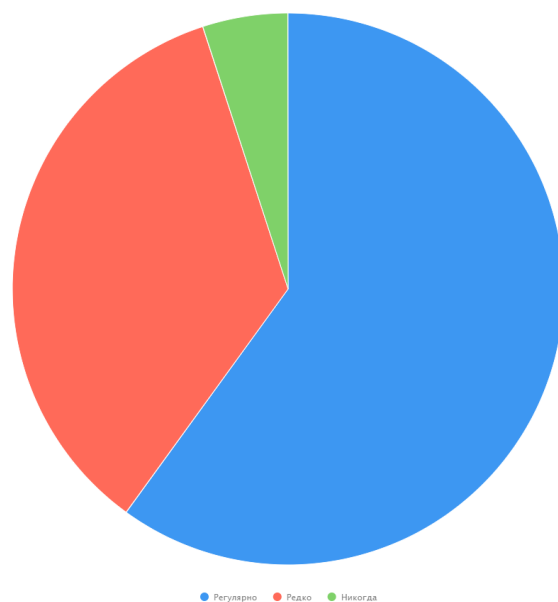
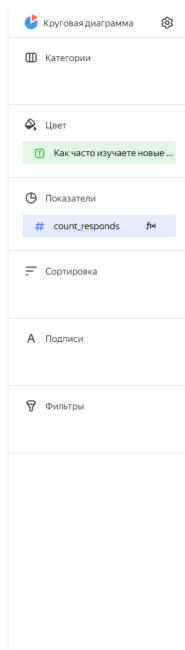
Какие функции вы считаете наиболее полезными в облачных платформах?
Managed Kubernetes Service
Serverless базы данных
Автомасштабирование и управляемые сервисы баз данных
Бессерверные вычисления
Возможность быстрого прототипирования и развертывания
Встроенные инструменты безопасности и комплаенс
Выделенные инструменты для миграции данных
Высокая доступность и отказоустойчивость
Гибридные возможности для подключения к локальной инфраструктуре
Глобальная сеть доставки контента
Инструменты для Data Lake и Big Data
Инструменты для мониторинга и алертинга
Интегрированные среды разработки
Контейнеризация и оркестрация
Платформы как услуга (PaaS), чтобы не управлять инфраструктурой
Сервисы потоковой обработки данных
Сервисы управления идентификацией и доступом
Удобный биллинг и детализация расходов
Функции автоматического резервного копирования
Широкий выбор готовых ML-моделей и сервисов AI

В левой части дашборда располагаются круговые диаграммы для того, чтобы выделить ключевую информацию. Посередине расположены столбчатая и линейная диаграммы, справа — таблица с полезными функциями облачных платформ по мнению респондентов.

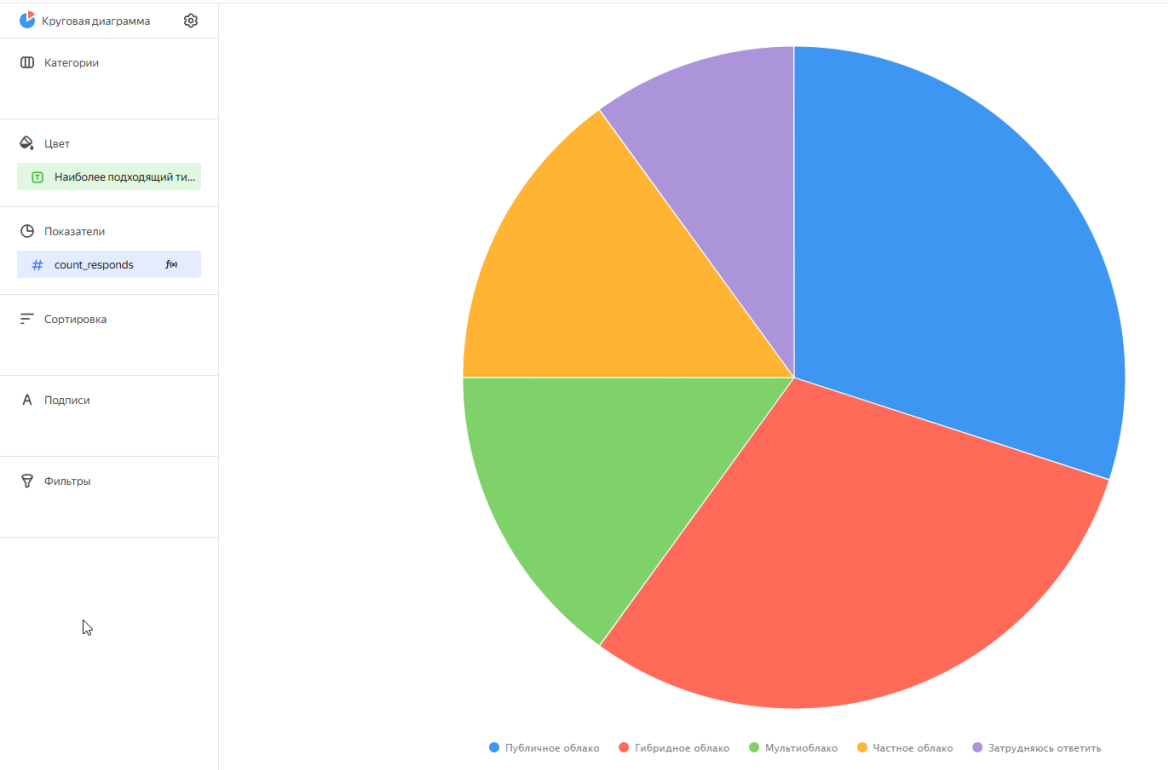
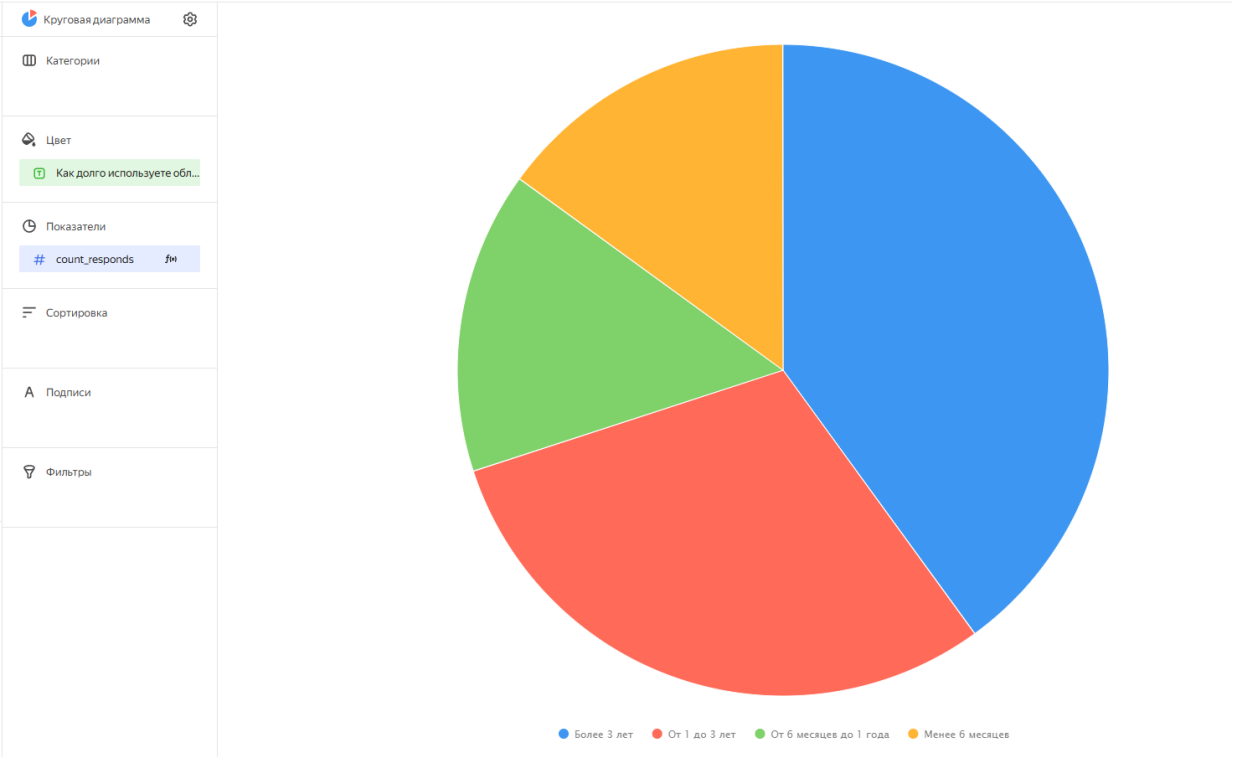
## 3. Анализ результатов

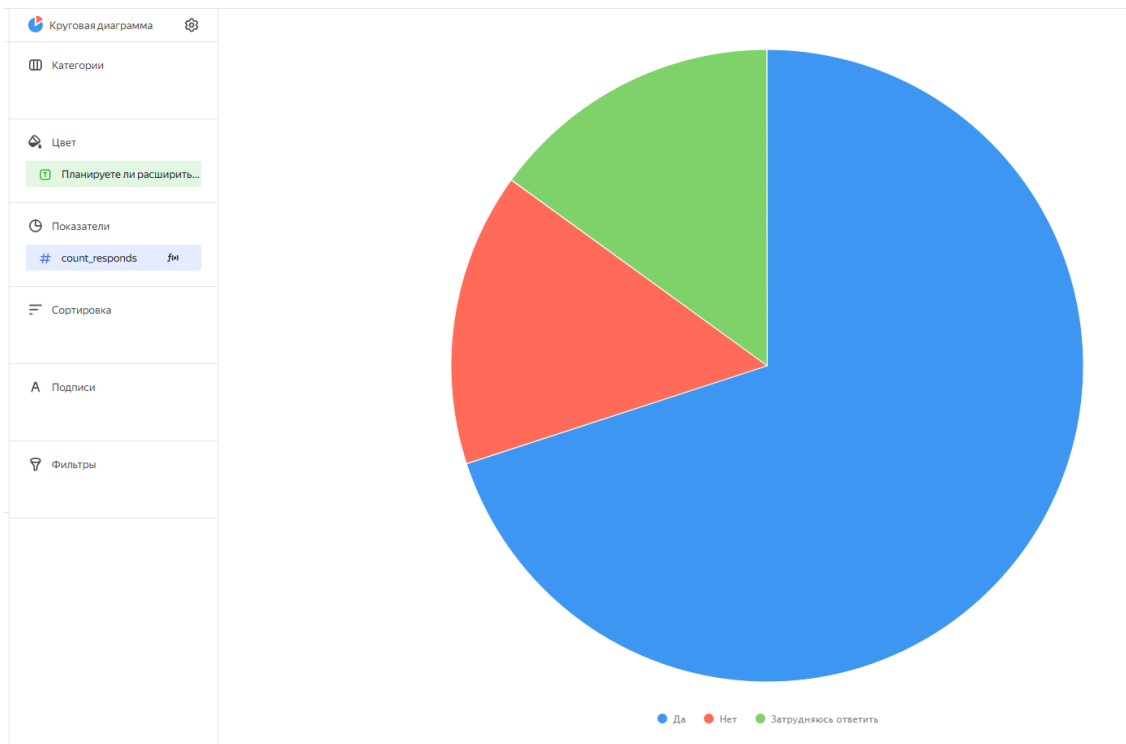
### 3.1. Круговые диаграммы

Круговые диаграммы были разработаны для вопросов с одним вариантом ответа.



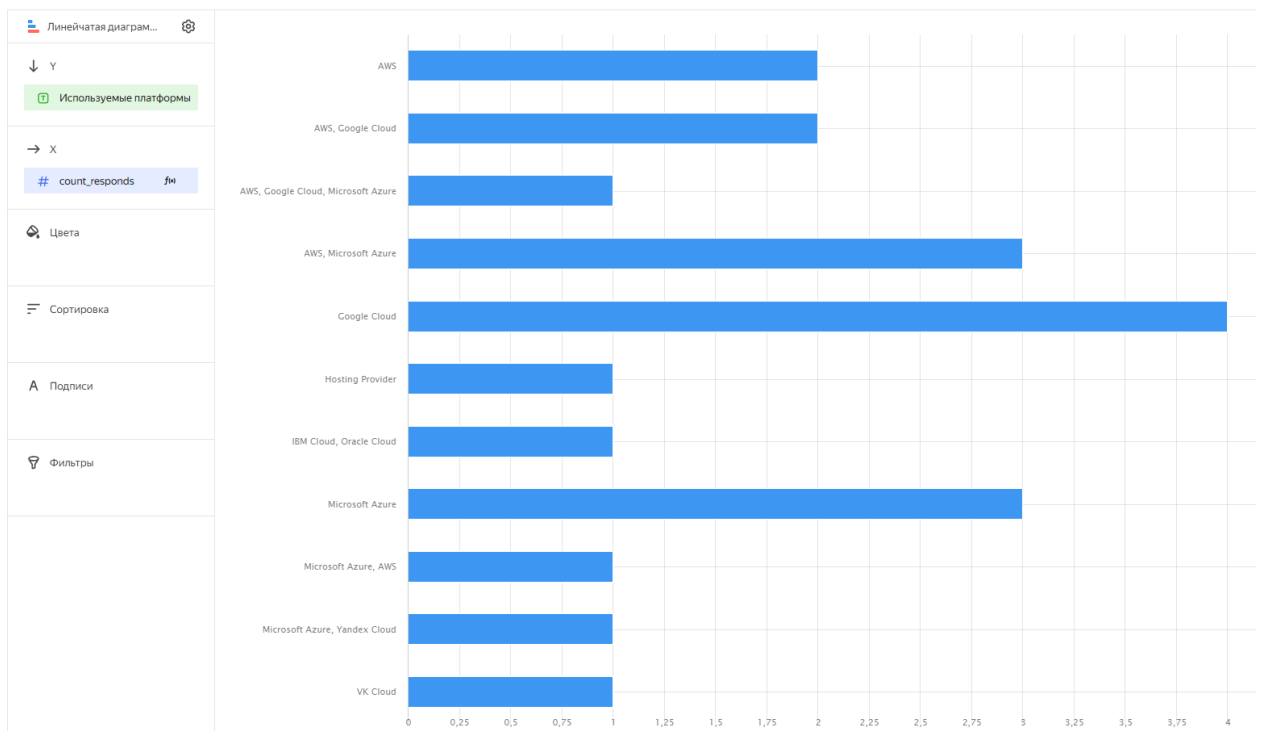
На картинке видно, что больше половины опрошенных изучают новые инструменты регулярно.

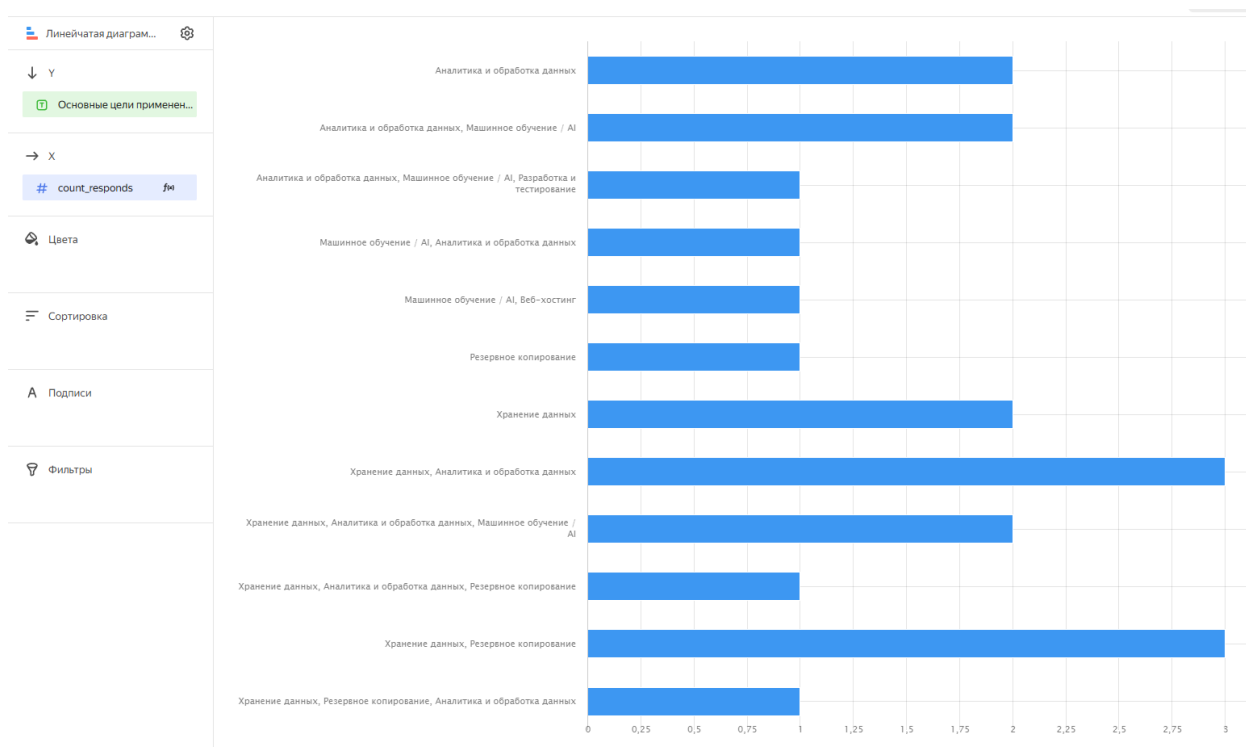
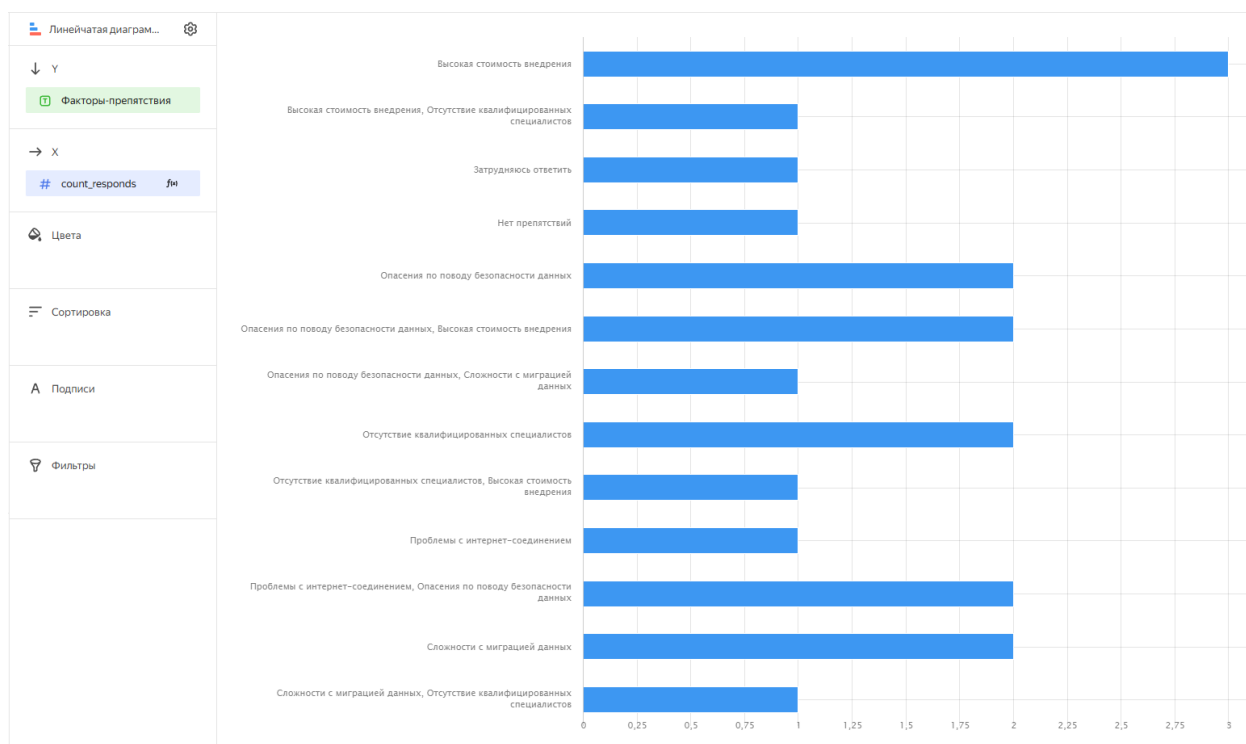




### 3.2. Линейчатые диаграммы

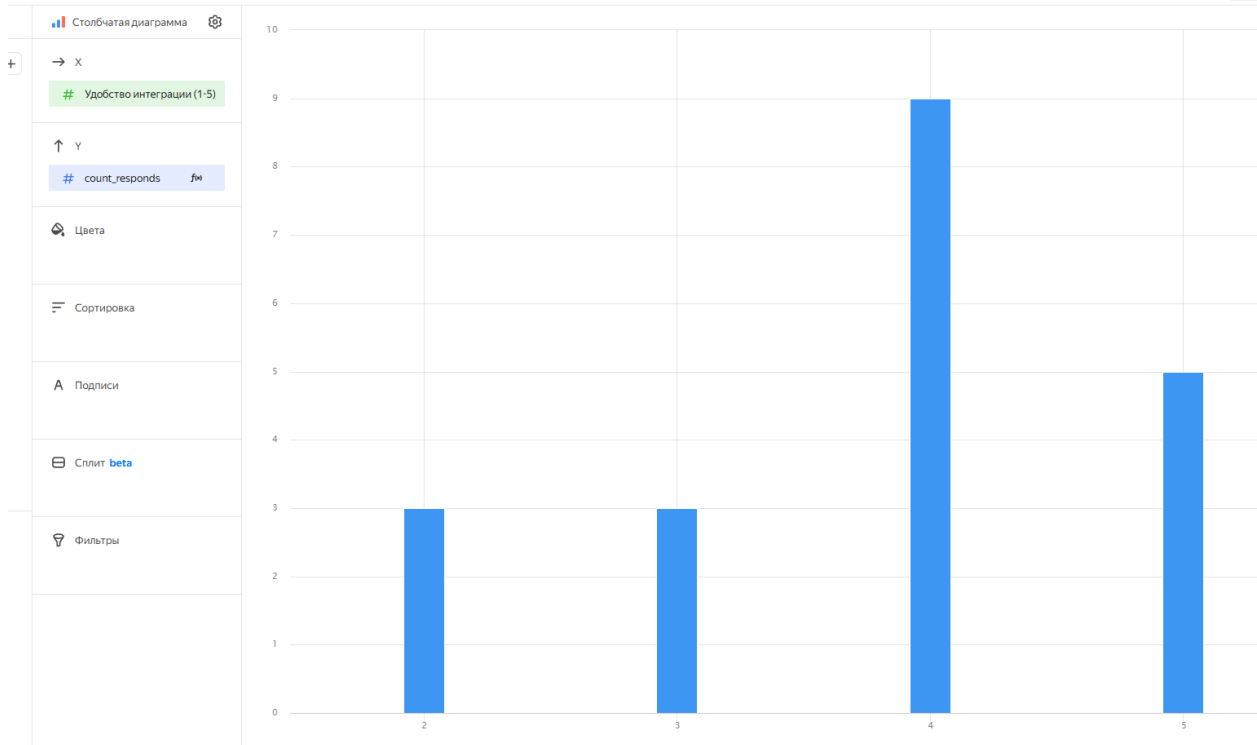
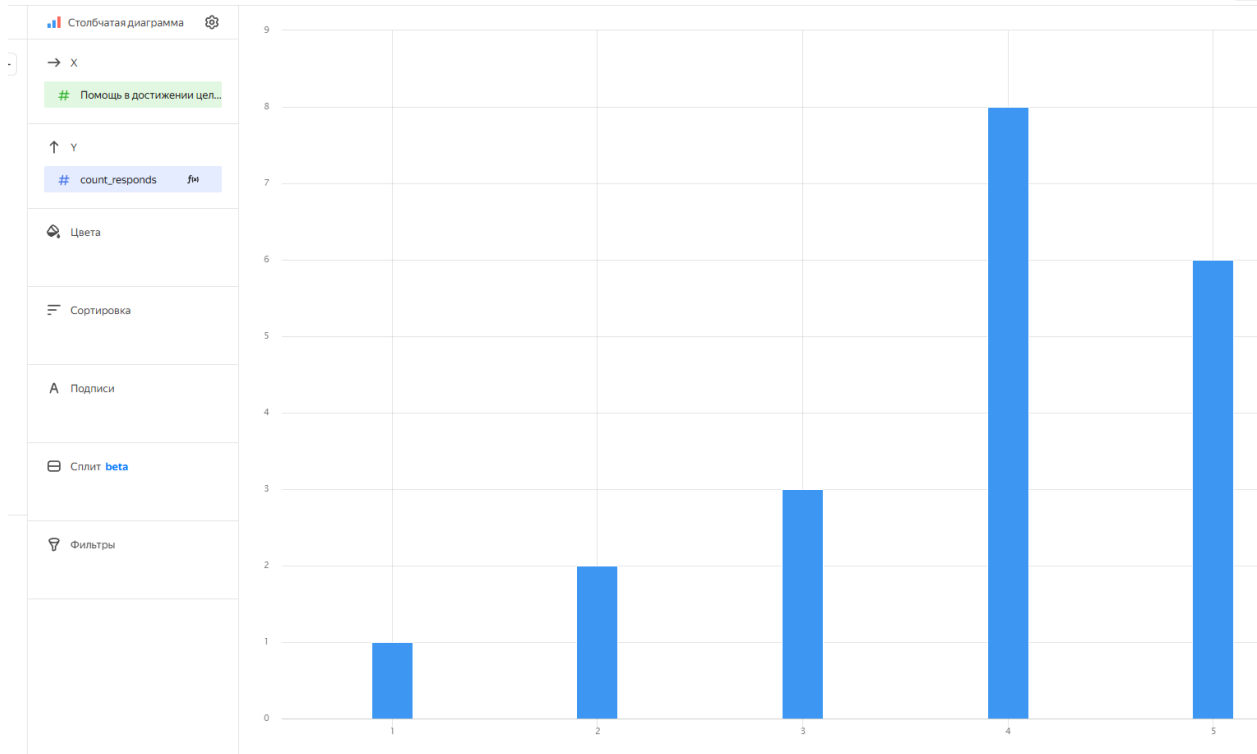
Линейчатые диаграммы были разработаны для вопросов с множественным выбором ответа.



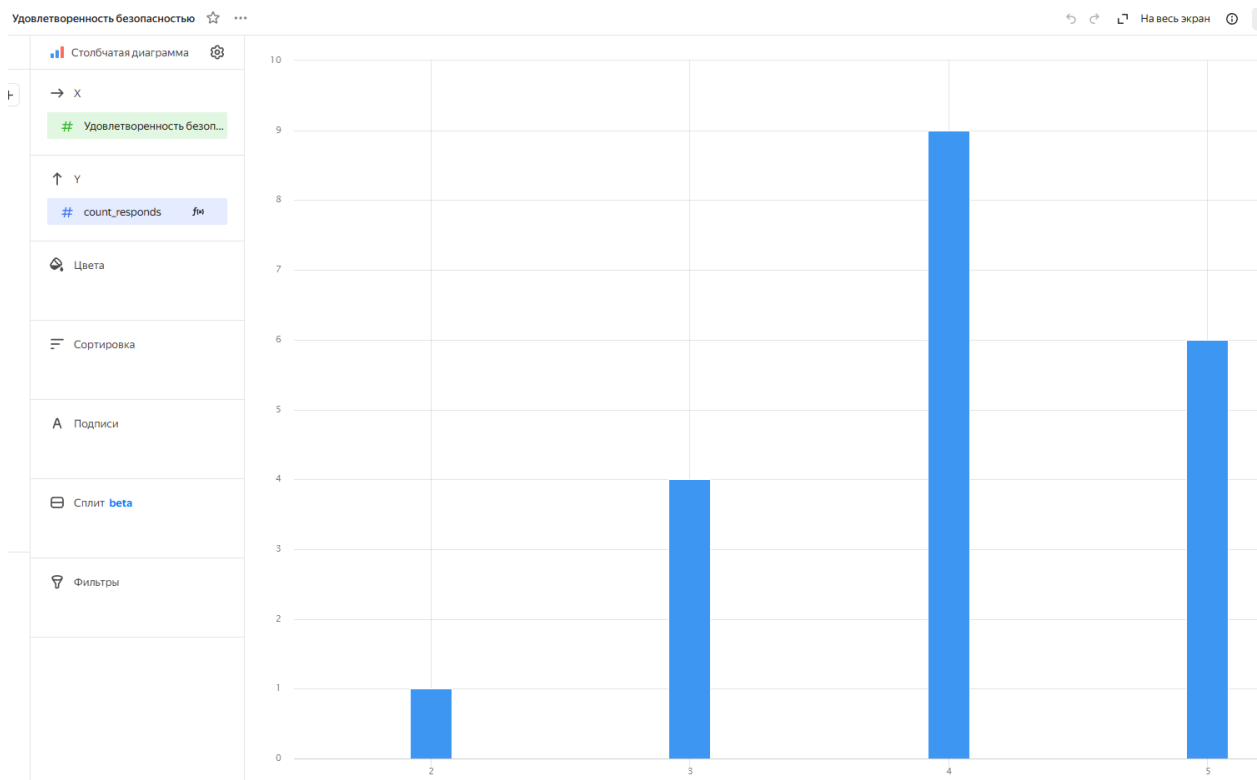


### 3.3. Столбчатые диаграммы

Столбчатые диаграммы были разработаны для вопросов с использованием шкалы Лайкерта.







#### 4. Выводы

- 1) По результатам опроса было выявлено, что 70% респондентов более 1 года используют облачные решения для работы с данными
- 2) Топ 3 самых популярных облачных платформ для работы с данными, которые используют компании оказались: AWS, Microsoft Azure и Google Cloud
- 3) Больше половины респондентов считают (70%), что облачные решения помогли их компаниям в достижении поставленных целей
- 4) 60% респондентов регулярно изучают новые инструменты для работы с облачными данными

#### 5. Заключение

Был разработан дашборд в Yandex Datalens на основе созданного опроса для респондентов. Дашборд включает различные типы визуализаций, позволяющие наглядно оценить распределение ответов и выявить зависимости. Получен опыт в разработке аналитического дашборда на основе опроса и опыт в формулировке ключевых выводов.

## **6. Ссылки**

<https://github.com/Scortlin/DEP-MGPU/tree/main/Module6>