

Анализ автоматизации процессов работы с данными

Цель: Оценить уровень зрелости автоматизации в Data Engineering
Выборка: 50 специалистов по работе с данными
Методология: Online-опрос среди Data Engineers, Data Scientists, Analytics Engineers и ML Engineers

- Ключевые метрики:
- Уровень автоматизации процессов (1-5)
 - Эффективность внедрения (1-5)
 - Сэкономленное время (часов/неделю)
 - Распространенность инструментов

Часов сэкономлено в неделю

AVG(CASE [time_saved] WHEN 'More than 20 hours/week' THEN 20 WHEN '10-20 hours/week' THEN 10 WHEN '5-10 hours/week' THEN 5 WHEN '1-5 hours/week' THEN 1 WHEN 'Less than 1 hour/week' THEN 0)

3,83

Средний уровень автоматизации

AVG (FLOAT([etl_auto] + [transform_auto] + [orchestration_auto] + [quality_auto] + [testing_auto] + [deployment_auto]) / 6)

3,17

Измеряют ROI

is_roi_measured

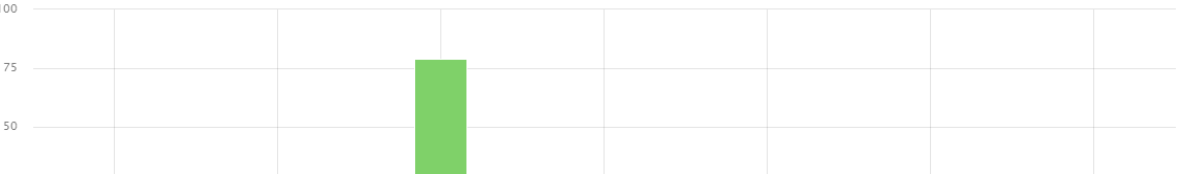
1

Использует Airflow

uses_airflow

2

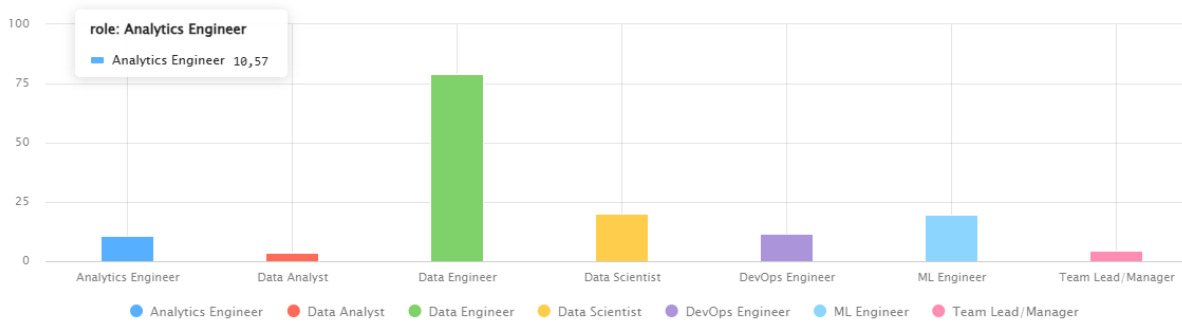
Уровень атоматизации по ролям



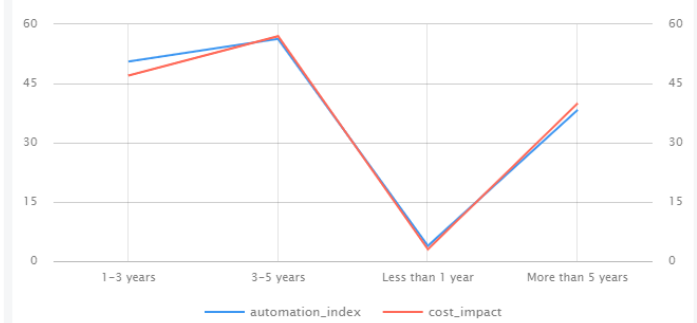
Сравнение уровня автоматизации и ее влияния на рабочие процессы



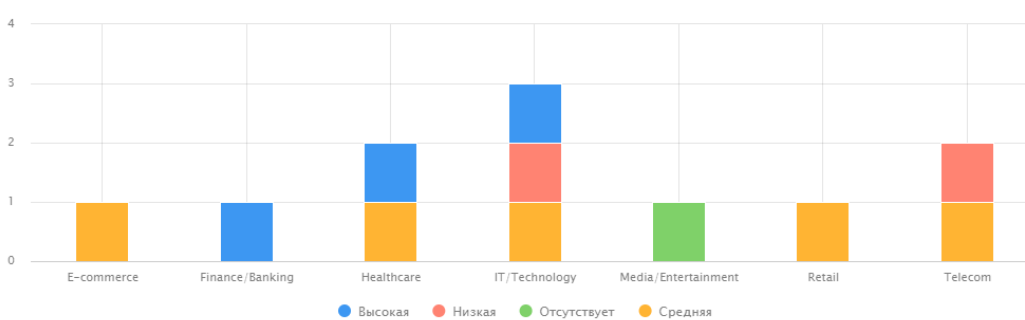
Уровень автоматизации по ролям



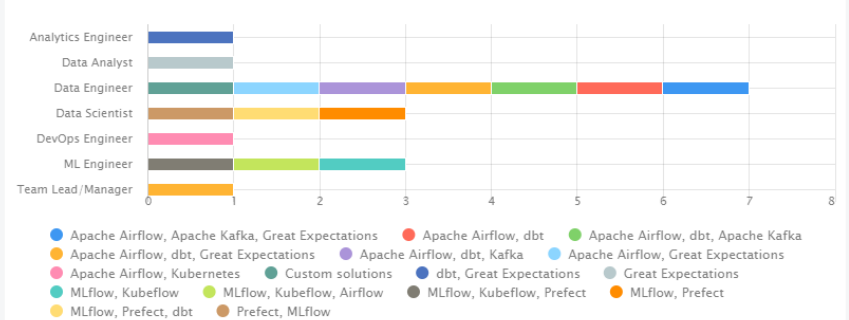
Сравнение уровня автоматизации и ее влияния на рабочие процессы



Анализ используемого стека технологий и зрелости DataOps практик



Популярность инструментов



Детали по респондентам

...

Measure Names	1-3 years					3-5 years					Less than 1 year	More than 5 years	
	Analytics Engineer	Data Analyst	Data Engineer	Data Scientist	ML Engineer	Analytics Engineer	Data Engineer	Data Scientist	DevOps Engineer	ML Engineer	Data Engineer	Data Engineer	Team Lead/Manager
automation_index	2,43	3,71	27,43	9,29	7,71	8,14	13,86	10,71	11,43	12,14	3,86	33,71	4,57
dataops_maturity	1	1	1	1	1	1	2	1	1	2	1	1	1
tools_used	1	1	2	1	1	1	2	2	1	2	1	4	1

Подробные данные по респондентам для глубокого анализа

role

Нет выбранных значений

experience

Нет выбранных значений

industry

Нет выбранных значений

Ключевые выводы

- 1. Data Engineers демонстрируют самый высокий уровень автоматизации
- 2. Компании с внедренными DataOps практиками экономят 15+ часов в неделю
- 3. Apache Airflow - стандарт де-факто для оркестрации пайплайнов
- 4. Высокий уровень автоматизации напрямую влияет на надежность процессов
- 5. Основные барьеры: сложность инструментов и нехватка экспертизы