



ДОКУМЕНТАЦИЯ

**«Анализ сотрудников компании и
факторов увольнений (поиск
инсайтов, составление рекомендаций
стейкхолдерам)»**

Аналитик данных
Лаврушева Елена Викторовна

г. Москва, 2025

ОГЛАВЛЕНИЕ

Введение	3
Блок 1 Описание исходного датасета и типов данных	5
Блок 2 Подготовка и очистка данных	7
2.1. Проверка датасета	7
2.2. Преобразования	7
2.3. Результат итогового обработанного датасета	8
Блок 3 Анализ данных и рекомендации для стейкхолдеров по результатам	11
3.1. Анализ уволенных сотрудников	11
3.1.1. Сегментированный анализ уволенных сотрудников по числовым параметрам	12
3.1.2. Сегментированный анализ уволенных сотрудников по категориальным параметрам	15
3.2. Сравнительный анализ групп уволенных и работающих сотрудников относительно различных параметров	17
3.2.1. Анализ с помощью KDE (Kernel Density Estimate)	18
3.2.2. Анализ сотрудников в процентном и долевым отношениях	28
3.2.3. Анализ с помощью корреляционной матрицы	30
Блок 4 Построение модели	32
4.1. Модель машинного обучения. Логистическая регрессия	32
4.2. Топ-5 наиболее важных признаков для увольнений	34
4.3. Анализ распределения работающих по группам риска	36
Итоги проекта и заключение	38

ВВЕДЕНИЕ

Актуальность: анализ сотрудников компании и факторов увольнений является критически важным для успешного функционирования и развития организации.

Он поможет компании понять причины увольнений и разработать стратегии для улучшения условий работы и удержания сотрудников. Позволит компании не только снизить финансовые потери, но и повысить производительность, удержать таланты и улучшить корпоративную культуру.

Это ключевой инструмент для создания устойчивой и конкурентоспособной организации.

Цель проекта: проведение анализа данных сотрудников с целью оптимизации управления персоналом и бизнес-процессами компании.

Бизнес-задачи:

1. Проведение анализа данных сотрудников и определение ключевых факторов, влияющих на увольнения.
2. Разработка рекомендаций по улучшению кадровой политики компании с учетом выявленных рисков.
3. Построение модели, предсказывающей, уволится ли сотрудник в ближайшее время.

Стейкхолдерами результатов анализа могут являться:

1. Руководство компании (Топ-менеджмент):
2. Генеральный директор, CFO, COO – используют данные для стратегических решений по управлению персоналом и снижению текучести.
3. Директор по персоналу (HRD) – отвечает за кадровую политику и внедрение изменений на основе анализа.
4. Менеджеры среднего звена:
5. Руководители подразделений – заинтересованы в понимании причин увольнений в своих отделах и улучшении микроклимата.
6. HR-бизнес-партнеры – помогают адаптировать HR-стратегию под нужды конкретных команд.
7. Служба по управлению персоналом (HR-отдел):
8. Специалисты по подбору (рекрутеры) – корректируют стратегии найма, если увольнения связаны с ошибками при отборе.
9. Специалисты по обучению и развитию (L&D) – выявляют потребности в обучении, если уход сотрудников связан с недостатком роста.
10. HR-аналитики – непосредственно проводят анализ и интерпретируют данные.
11. Сотрудники компании:

12. Рабочие коллективы – косвенно заинтересованы, так как результаты могут повлиять на условия труда, мотивацию и корпоративную культуру.
13. Профсоюзы / Совет сотрудников – если есть, могут использовать данные для переговоров с руководством.
14. Внешние стейкхолдеры:
15. Акционеры / инвесторы – высокие показатели текучести могут сигнализировать о проблемах в компании, влияющих на её устойчивость.
16. HR-консалтинговые компании – если анализ проводится с привлечением внешних экспертов.
17. Государственные органы (в некоторых случаях) – например, если анализ связан с проверками условий труда.

БЛОК 1 ОПИСАНИЕ ИСХОДНОГО ДАТАСЕТА И ТИПОВ ДАННЫХ

Для исследования был взят датасет «[IBM HR Analytics Employee Attrition & Performance](#)». Он содержит информацию о сотрудниках компании, связанную с текучестью кадров (attrition). Он включает различные характеристики сотрудников, такие как возраст, образование, уровень дохода, должность, удовлетворенность работой, стаж, семейное положение и другие факторы. Основная цель датасета — анализ и прогнозирование увольнений сотрудников на основе предоставленных данных.

Датасет содержит 35 столбцов (из них 26 имеют целочисленный тип данных, 9 – тип данных текстовый или смешанный числовой и нечисловой), 1470 строк, 51450 значений, использует 402.1+ КВ памяти.

Описание содержания признаков:

№	Имя столбца	Описание	Тип данных
1.	Age	Возраст сотрудников (в годах)	int64
2.	Attrition	Уволен	object
3.	BusinessTravel	Командировки	object
4.	DailyRate	Дневная ставка оплаты	int64
5.	Department	Отдел	object
6.	DistanceFromHome	Расстояние от дома до работы (в км/милях)	int64
7.	Education	Образование	int64
8.	EducationField	Область образования	object
9.	EmployeeCount	Количество сотрудников	int64
10.	EmployeeNumber	Номер сотрудника	int64
11.	EnvironmentSatisfaction	Удовлетворенность условиями работы	int64
12.	Gender	Пол	object
13.	HourlyRate	Почасовая ставка оплаты	int64
14.	JobInvolvement	Вовлеченность в работу	int64
15.	JobLevel	Уровень работы	int64
16.	JobRole	Должность	object
17.	JobSatisfaction	Удовлетворенность работой	int64
18.	MaritalStatus	Семейное положение	object
19.	MonthlyIncome	Ежемесячный доход	int64
20.	MonthlyRate	Ежемесячная ставка оплаты труда	int64
21.	NumCompaniesWorked	Количество лет, отработанных в компании	int64
22.	Over18	Старше 18 лет	object
23.	OverTime	Переработка	object

24.	PercentSalaryHike	Процент повышения зарплаты	int64
25.	« PerformanceRating »	« Рейтинг производительности »	int64
26.	« RelationshipSatisfaction »	« Удовлетворенность отношениями »	int64
27.	StandardHours	Стандартное количество рабочих часов в неделю, установленное компанией для сотрудников	int64
28.	« StockOptionLevel »	« Бонусы »	int64
29.	TotalWorkingYears	Общий стаж работы	int64
30.	TrainingTimesLastYear	Количество часов обучения в прошлом году	int64
31.	« WorkLifeBalance »	« Баланс между работой и личной жизнью »	int64
32.	YearsAtCompany	Количество лет в компании	int64
33.	YearsInCurrentRole	Количество лет в текущей должности	int64
34.	YearsSinceLastPromotion	Количество лет с последнего повышения	int64
35.	YearsWithCurrManager	Количество лет с текущим менеджером	int64

Возраст работников: средний – составляет 37 лет, минимум – 18 и максимум – 60.

Расстояние от дома до работы в среднем 9,19 миль, при этом некоторые сотрудники живут всего в 1 миле, а другие в 29 милях от своего дома.

Средняя дневная ставка – 802,49 с широким диапазоном от 102,00 до 1499,00.

Средний ежемесячный доход составляет примерно 6502,93, минимально – 1009,00, максимально – 19999,00.

В среднем сотрудники имеют около 11,28 лет общего рабочего стажа, некоторые только начали, а другие имеют до 40 лет опыта.

Сроки работы в компании: в среднем 7 лет (от 3-х до 9-и).

БЛОК 2 ПОДГОТОВКА И ОЧИСТКА ДАННЫХ

Приведение данных в структурированный, непротиворечивый и пригодный для анализа вид

2.1. Проверка датасета:

Проверка на	Результат
соответствие типов признаков их смысловому содержанию	соответствуют, однако целесообразно некоторым категориальным параметрам изменить тип данных с целью улучшения совместимости
наличие пропущенных значений	пропуски значений отсутствуют
значение со знаком «-»	отрицательные значения отсутствуют
уникальность строк и наличие дубликатов	все строки уникальны, дубли отсутствуют
названий столбцов	некорректные названия не выявлены
выбросы в параметрах	наличие выбросов обнаружено в: <ul style="list-style-type: none">- MonthlyIncome- NumCompaniesWorked- PerformanceRating- StockOptionLevel- TotalWorkingYears- TrainingTimesLastYear- YearsAtCompany- YearsInCurrentRole- YearsSinceLastPromotion- YearsWithCurrManager
неинформативные признаки	более 95% строк признака содержат одно и то же значение: <ul style="list-style-type: none">- EmployeeCount- Over18- StandardHours
нерелевантные признаки	- EmployeeNumber
возможные опечатки символов в признаках имеющих тип данных текстовый или смешанный числовой и нечисловой	не обнаружены

2.2. Преобразования:

1. Столбцам Attrition, OverTime изменён тип данных;
2. Оставлены только параметры, имеющие значение для анализа;
3. Столбцы расставлены в удобном для анализа порядке;
4. Выбросы в параметрах оставлены без изменений, т.к. не оказывают большого влияния. Однако данные приняты во внимание;
5. Тип данных в остальных категориальных столбцах остаются неизменны.



2.3. Результат итогового обработанного датасета:

Датасет содержит 31 столбец (из них 26 имеют целочисленный тип данных, 5 – тип данных текстовый или смешанный числовой и нечисловой, относящихся к номинальным), 1470 строк, 51450 значений, использует 356.1+ КВ памяти.

№	Имя столбца	Описание	Тип данных
1.	Attrition: 1 0	Уволен: Да – 1 Нет – 0	int64
2.	Department: Sales Research & Developmen Human Resource	Отдел: Продажи Исследования и разработки Кадровые ресурсы	object
3.	JobRole: Sales Executive Research Scientist Laboratory Technician Manufacturing Director Healthcare Representative Manager Sales Representative Research Director Human Resources	Должность: Директор по продажам Научный сотрудник Лаборант-технолог Директор по производству Представитель здравоохранения Менеджер Торговый представитель Директор по исследованиям Рекрутер	object
4.	JobLevel: 1 2 3 4 5	Уровень работы: 1 2 3 4 5	int64
5.	BusinessTravel: 1 2 0	Командировки: Путешествую редко - 1 Путешествую часто - 2 Не путешествую - 0	int64
6.	Age	Возраст сотрудников (в годах)	int64
7.	Gender: Female Male	Пол: Женский Мужской	object
8.	Education: 1 2 3 4 5	Образование: Ниже колледжа – 1 Колледж – 2 Бакалавр – 3 Магистр – 4 Доктор – 5	int64
9.	EducationField: Life Sciencesv Medical Marketing Technical Degree Human Resources Other	Область образования: Науки о жизни Медицина Маркетинг Техническая степень Кадровые ресурсы Другое	object

10.	TrainingTimesLastYear	Количество часов обучения в прошлом году	int64
11.	MaritalStatus: Single Married Divorced	Семейное положение: Холост Женат Разведен	object
12.	DistanceFromHome	Расстояние от дома до работы (в км/милях)	int64
13.	HourlyRate	Почасовая ставка	int64
14.	DailyRate	Дневная ставка	int64
15.	MonthlyRate	Ежемесячная ставка оплаты труда	int64
16.	MonthlyIncome	Ежемесячный доход	int64
17.	OverTime: 1 0	Переработка: Да – 1 Нет - 0	int64
18.	PercentSalaryHike	Процент повышения зарплаты	int64
19.	StockOptionLevel: 0 1 2 3	Бонусы: 0 1 2 3	int64
20.	NumCompaniesWorked	Количество предыдущих мест работы	int64
21.	PerformanceRating: 1 2 3 4	Рейтинг производительности: Низкий – 1 Хороший – 2 Отличный – 3 Превосходный – 4	int64
22.	JobInvolvement: 1 2 3 4	Вовлеченность в работу: Низкая – 1 Средняя – 2 Высокая – 3 Очень высокая – 4	int64
23.	EnvironmentSatisfaction: 1 2 3 4	Удовлетворенность условиями работы: Низкая – 1 Средняя – 2 Высокая – 3 Очень высокая – 4	int64
24.	JobSatisfaction: 1 2 3 4	Удовлетворенность работой: Низкая – 1 Средняя – 2 Высокая – 3 Очень высокая – 4	int64
25.	RelationshipSatisfaction: 1 2 3 4	Удовлетворенность отношениями: Низкая – 1 Средняя – 2 Высокая – 3 Очень высокая – 4	int64
26.	WorkLifeBalance: 1 2 3	Баланс между работой и личной жизнью: Плохо – 1 Хорошо – 2 Лучше – 3	int64

	4	Лучший – 4	
27.	TotalWorkingYears	Общий стаж работы	int64
28.	YearsAtCompany	Количество лет в компании	int64
29.	YearsInCurrentRole	Количество лет в текущей должности	int64
30.	YearsSinceLastPromotion	Количество лет с последнего повышения	int64
31.	YearsWithCurrManager	Количество лет с текущим менеджером	int64

БЛОК 3 АНАЛИЗ ДАННЫХ И РЕКОМЕНДАЦИИ ДЛЯ СТЕЙКХОЛДЕРОВ ПО РЕЗУЛЬТАТАМ

1.1. Анализ уволенных сотрудников

Всего = 1470

Количество уволенных сотрудников = 237

Процент уволенных сотрудников = 16.122448979591837 %

Количество работающих сотрудников = 1233

Процент работающих сотрудников = 83.87755102040816 %

Сравнение статистических показателей уволенных и работающих сотрудников:

Демографические и профессиональные характеристики:

- Возраст (Age) – уволенные в среднем на 4 года моложе;
- Уровень должности (JobLevel) – уволенные чаще занимают более низкие должности;
- Образование (Education) – почти одинаково.

Удовлетворенность и вовлеченность:

- Удовлетворенность средой (EnvironmentSatisfaction) – уволенные менее удовлетворены условиями работы;
- Удовлетворенность работой (JobSatisfaction) – уволенные меньше довольны работой;
- Баланс между работой и личной (WorkLifeBalance) – разница не значительна, но уволенные чаще испытывали дисбаланс.

Опыт и карьера:

- Общий стаж (TotalWorkingYears) – уволенные менее опытны на 3.62 года;
- Стаж в компании (YearsAtCompany) – уволенные работали меньше на 2.24 года;
- Количество лет с последнего повышения (YearsSinceLastPromotion) – разница не значительная, однако уволенные реже получали повышение.

Условия работы:

- Переработки (OverTime) – уволенные в 2,3 раза чаще перерабатывали;
- Дистанция до работы (DistanceFromHome) – уволенные жили дальше на 1.71 (км/миль).

Финансовые показатели:

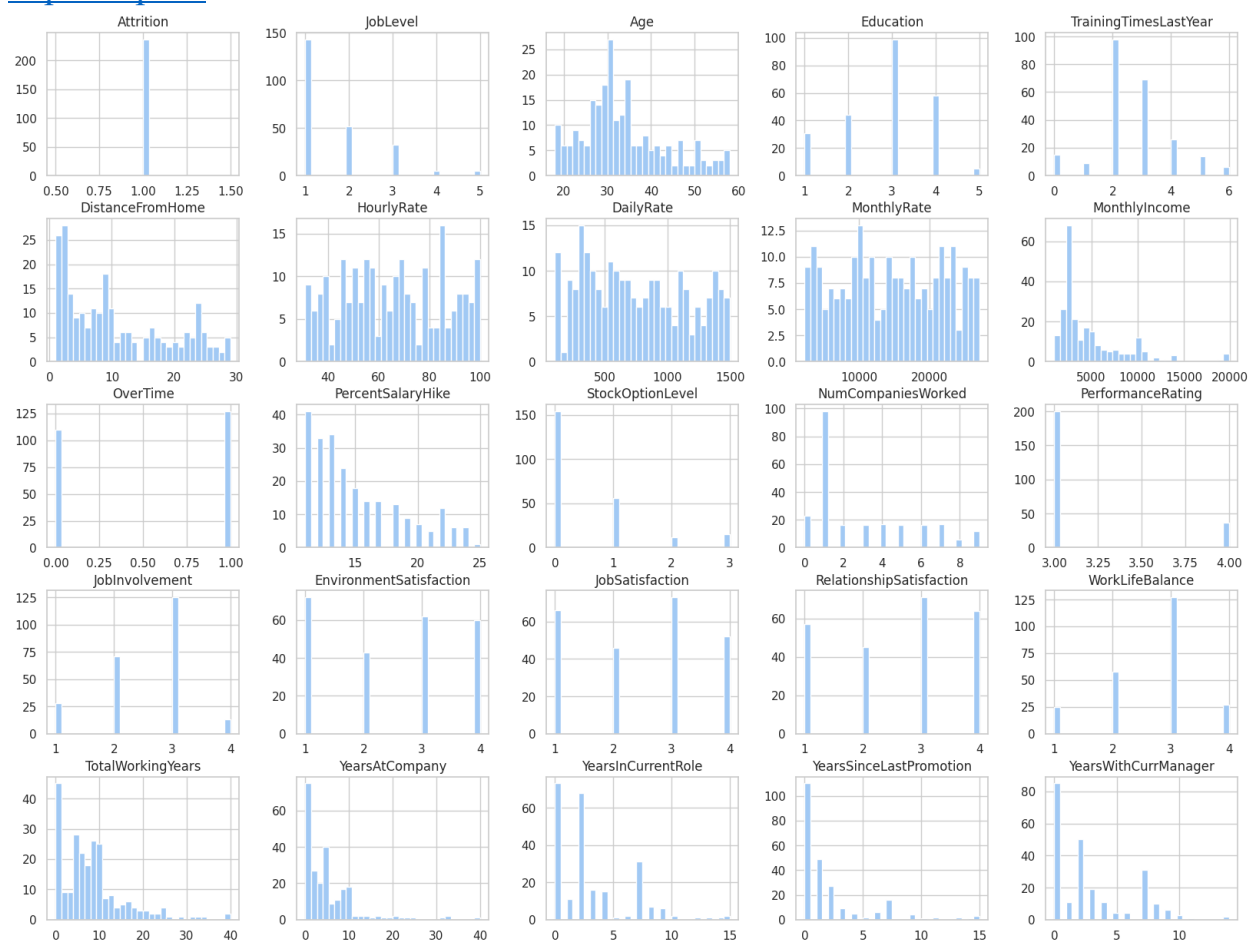
- Зарплата (MonthlyIncome) – уволенные получали примерно на 30% меньше;
- Процент повышения зарплаты (PercentSalaryHike) – почти одинаково;
- Бонусы (StockOptionLevel) – уволенные реже получали бонусы.

Обучение и развитие:

- Количество часов обучения в прошлом году (TrainingTimesLastYear) – разница незначительная, однако уволенные реже проходили обучение.

Уволенные сотрудники	Работающие сотрудники
Более молодые, занимали более низкие должности	Более опытные, занимают выше должности
Менее удовлетворены работой и условиями	Больше удовлетворены работой
Чаще работали сверхурочно, но получали меньше	Реже перерабатывают, но имеют более высокий доход
Жили дальше от работы	Чаще получают бонусы и обучение
Имели меньший стаж и реже получали повышения	

3.1.1. Сегментированный анализ уволенных сотрудников по числовым параметрам:



Интерпретация полученных результатов:

1) По уровню работы (JobLevel) – чаще всего уходят люди первого уровня. Это может быть связано с высокой текучестью кадров среди младших сотрудников, которые либо не удовлетворены условиями труда, либо используют эти позиции как временный этап в карьере. Чем выше уровень, тем меньше увольнений, что объясняется редкой сменой позиций среди топ-специалистов или руководителей.

Рекомендации: основная зона риска — младшие должности с низкими ставками. Улучшение условий для этих сотрудников может значительно снизить количество увольнений. Для снижения текучести кадров на уровнях

1–2 стоит пересмотреть систему оплаты, добавить бонусы или возможности карьерного роста. На уровнях 3–4 важно поддерживать баланс между зарплатой и нематериальной мотивацией (например, гибкий график, обучение). Уровень 5 остаётся стабильным, но требует регулярного мониторинга, чтобы избежать потери ключевых специалистов.

2) По возрасту (Age) – сотрудники в возрастном диапазоне от 25 до 35 лет более склонны к смене работы. Возможно, они сталкиваются с карьерными изменениями или не удовлетворены условиями труда. После 40 лет количество увольнений постепенно снижается, достигая минимума к 60 годам. Это может быть связано с тем, что старшие сотрудники более стабильны в своих позициях или реже ищут новые возможности.

Рекомендации: Компании стоит обратить внимание на причины увольнений среди сотрудников 30–40 лет, так как это наиболее активная и опытная часть коллектива. Возможно, потребуется пересмотреть условия труда, карьерные возможности или систему мотивации для этой возрастной группы.

3) По уровню образования (Education) – чаще остальных увольняются работники с образованием бакалавриат, чуть меньше – магистратура. Люди с докторской степенью почти не увольняются. Возможно, причины связаны с тем, что сотрудники с низким уровнем образования чаще увольняются из-за неудовлетворённости зарплатой, условиями труда или отсутствием карьерного роста. Высокий уровень образования, как правило связан с более высокой должностью, оплатой и стабильностью, что снижает вероятность увольнений.

Рекомендации: инвестиции в их развитие и улучшение условий работы могут значительно снизить количество увольнений: разработка программ мотивации; обеспечение прозрачности карьерных перспектив; мониторинг потребностей, для сохранения ключевых специалистов.

4) По количеству часов обучения в прошлом году (повышению квалификации) (TrainingTimesLastYear) больше всего уволенных прошли среднее количество часов обучения. Возможными причинами увольнения при таком парадоксе могут быть: почувствовали себя "достаточно квалифицированными" для поиска работы в другом месте, испытали разочарование, если обучение не привело к карьерному росту, получили предложения от других компаний благодаря новым сертификатам.

Рекомендации: пересмотреть содержание обучения для группы 2-3: добавить программу закрепления знаний; связать обучение с конкретными карьерными шагами; улучшить систему пост-тренинговой поддержки. Для группы 6+: продолжать текущую стратегию, так как она работает. Для группы 0: ввести минимальный обязательный план обучения.

5) По расстоянию от дома до работы (DistanceFromHome). Большинство уволенных проживают в непосредственной близости от места работы. По данному параметру требуется дополнительное исследование причин.

6) Относительно почасовой, дневной и ежемесячной ставки (HourlyRate, DailyRate, MonthlyRate) наблюдается нелинейная зависимость. В данном случае требуется более детальное исследование параметров.

7) По ежемесячному доходу (MonthlyIncome) – чем выше доход сотрудника, тем реже происходят увольнения. Существующая система мотивации для топ-персонала эффективна. Требуется пересмотр политики компенсаций для младшего персонала.

Рекомендации: оптимизировать базовые оклады на основании анализа рыночных зарплат, рассмотреть возможность введения индивидуальных пакетов.

8) По переработкам (OverTime) – небольшая разница между теми, у кого есть сверхурочные и у кого их нет. В то же время, те у кого есть переработка увольняются чаще.

9) По проценту повышения заработных плат (PercentSalaryHike) и по бонусам (StockOptionLevel) – похожая ситуация. Большинство уволенных сотрудников – это люди, не получавшие процентов и бонусов. С повышением отмечается снижение увольнений.

Рекомендации: обеспечение прозрачности, понятности и достижимости системы KPI для всех уровней сотрудников.

10) По количеству предыдущих мест работы (NumCompaniesWorked) – большинство увольнений отмечается среди сотрудников, для которых данная компания является вторым местом работы. Причины такого распределения стоит искать во взаимосвязях с другими параметрами. Данный факт явно относится к группе риска.

11) По рейтингу производительности (PerformanceRating) – отмечается тот факт, что среди уволенных сотрудников люди с уровнем «отличный» и «превосходный». Что свидетельствует о потере работоспособных кадров.

Рекомендации: разработать и внедрить специальные программы удержания высокоэффективных специалистов; регулярно мониторить удовлетворенности (раз в квартал) и баланса work-life.

12) По вовлеченности в работу (JobInvolvement) – чем выше вовлеченность сотрудников в работу, тем больше отмечается увольнений. Вероятно, данный факт связан с выгоранием. Однако люди с очень высокой вовлеченностью показывают наименьшее количество увольнений. Скорее всего таких людей в организации очень мало, поэтому и показатель среди уволенных низкий.

Рекомендации: проанализировать адекватность критериев достижения 3 и 4 уровня вовлеченности.

13) По удовлетворенности условиями работы (EnvironmentSatisfaction) – люди с низкой удовлетворенностью увольняются немного чаще остальных групп. Однако, разница незначительна.

14) Относительно удовлетворенности работой (JobSatisfaction), отношениями (RelationshipSatisfaction) и балансом между работой и личной жизнью (WorkLifeBalance) наблюдается пик увольнений на «высоком» уровне. Причинами такого распределения на графике могут быть некорректно собранные данные либо лукавство людей при мониторинге.

15) Чем меньше общий стаж работы (TotalWorkingYears), тем больше уволенных. То же касается количества лет в текущей компании (YearsAtCompany), количества лет в текущей должности (YearsInCurrentRole), количества лет с последнего повышения (YearsSinceLastPromotion) и количества лет с текущим менеджером (YearsWithCurrManager).

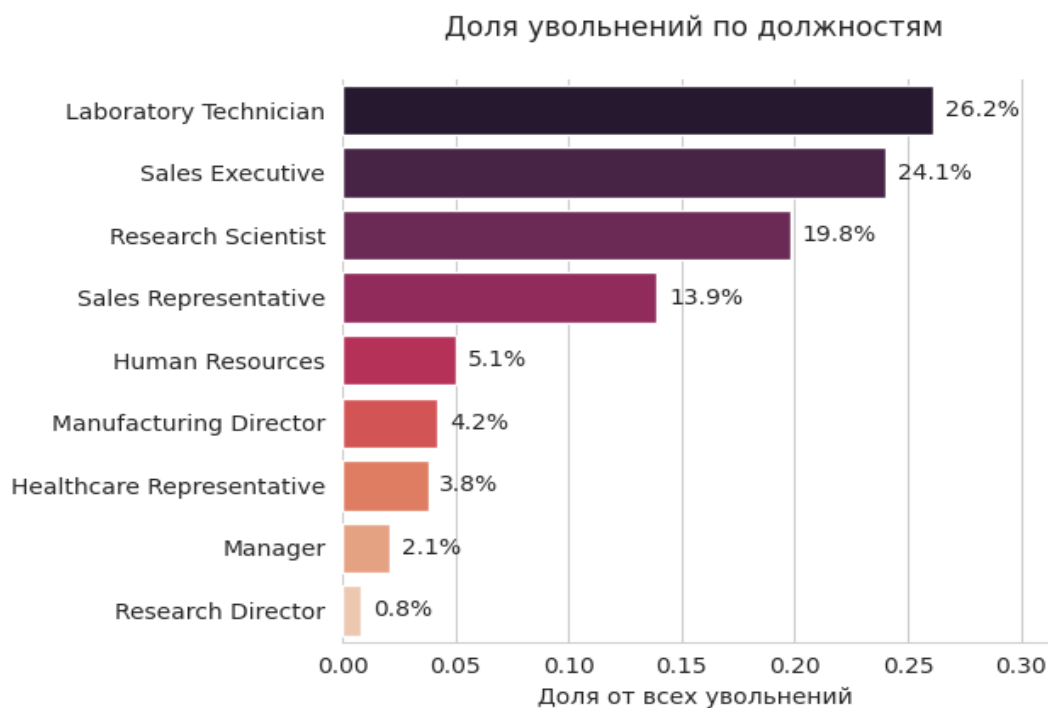
3.1.2. Сегментированный анализ уволенных сотрудников по категориальным параметрам:

1) По отделам: лидером увольнений является отдел исследований и разработок. Однако, стоит отметить, что и штатных единиц в данном отделе больше остальных.



Рекомендации: стабильные отделы стоит изучать на предмет лучших практик для распространения в проблемные зоны. Рекомендуется провести глубинные интервью с уволившимися сотрудниками для точной диагностики причин.

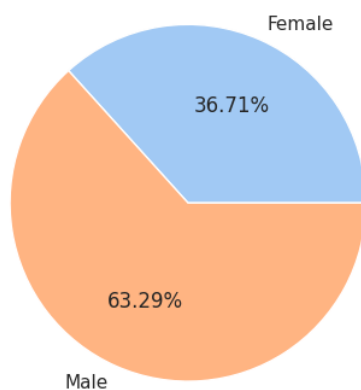
2) По должностям: наибольшая текучесть кадров наблюдается в отделах Laboratory Technician, Sales Executive, Research Scientist, Sales Representative. Возможные причины для техников – рутинный характер работы, низкая зарплата, ограниченные возможности карьерного роста; для директоров по продажам – высокий стресс из-за плановых показателей, конкурентная среда; для научных сотрудников – это может быть более интересные проекты в других компаниях; для торговых представителей скорее всего выгорание из-за высокой нагрузки.



Рекомендации: разработка и введение программ ротации, дополнительное обучение для карьерного развития, рассмотреть возможность оптимизации системы KPI или мотивации за её пределами, провести оценку оборудования на соответствие необходимым требованиям и произвести необходимую модернизацию.

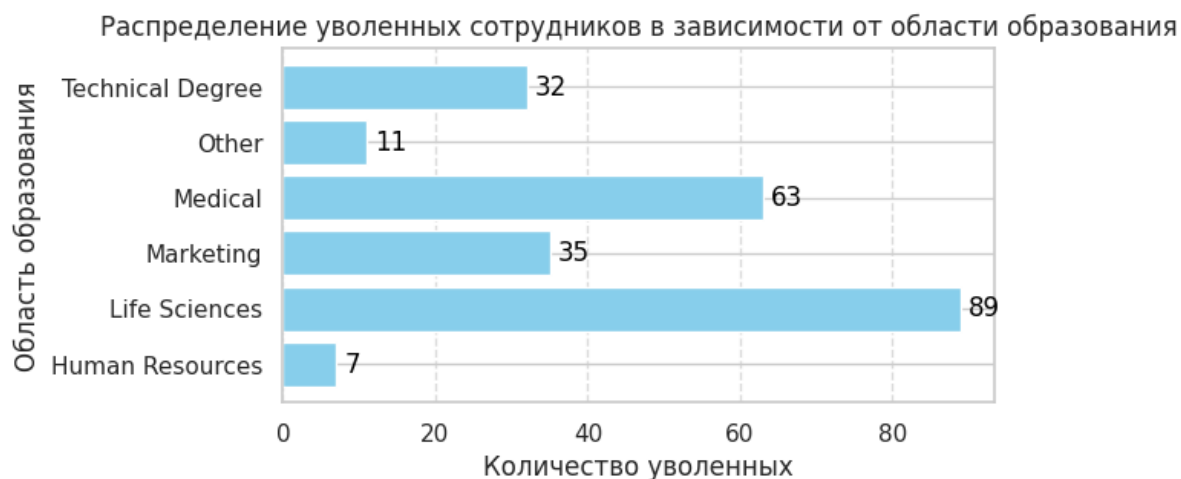
3) По гендеру: женщины увольняются примерно в два раза реже мужчин. Возможно, это связано с гендерным перекосом в штатном составе (больше мужчин).

Распределение уволенных сотрудников в зависимости от пола



Рекомендации: рассмотреть возможность разработки и внедрения Diversity & Inclusion политики. Обучить руководителей управлению разнородными командами.

4) По областям образования: наблюдается резкая дифференциация текучести по образовательным направлениям.



Явные лидеры по текучести:

Life Sciences (89 увольнений) – абсолютный максимум (в 2.8 раза выше среднего показателя);

Medical (63) – второй результат, требует особого внимания;

Marketing (35) и Technical Degree (32) – умеренные, но значимые показатели.

Стабильные группы:

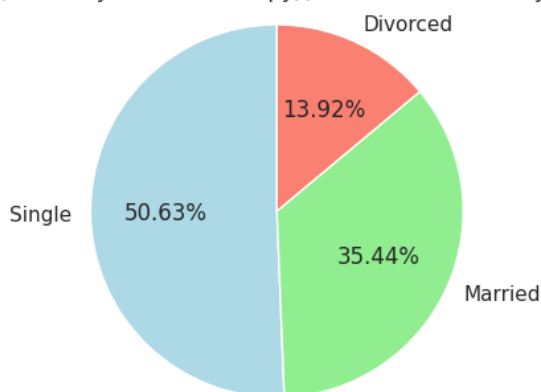
Human Resources (7) – минимальная текучесть (в 12.7 раз ниже, чем у лидера);

Other (11) – низкий показатель, но требует уточнения категорий.

Рекомендации: провести аудит рабочих мест, пересмотреть нагрузки. Рассмотреть возможность проведения творческих челленджей, хакатонов и т.п.

5) По семейному статусу: примерно половина уволившихся сотрудников холосты. Они более мобильны и готовы к изменениям, у них меньше обязательств, что облегчает решение об уходе.

Распределение уволенных сотрудников по семейному статусу



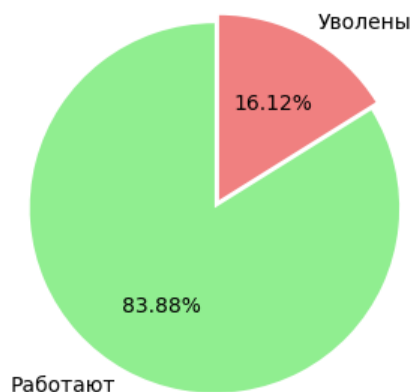
Рекомендации: разработать и внедрить карьерное планирование; обеспечить финансовое консультирование и психологическую поддержку; организовать социальные активности в коллективе.

3.2. Сравнительный анализ групп уволенных и работающих сотрудников относительно различных параметров:

Всего = 1470

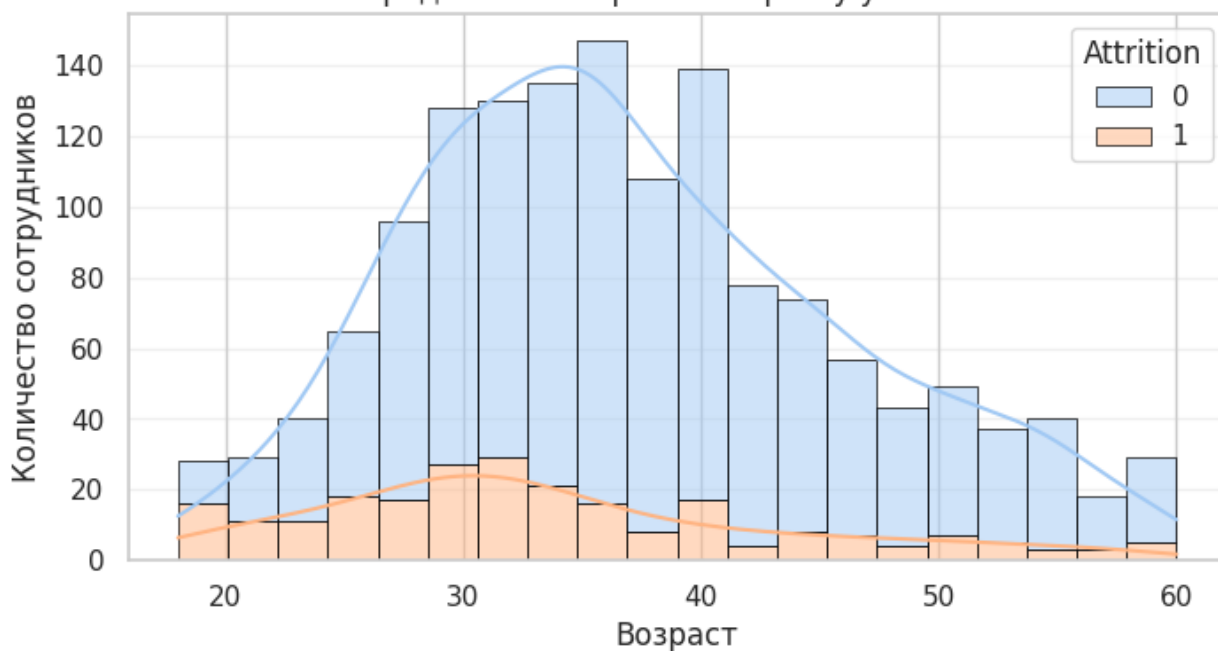
Количество уволенных сотрудников = 237
Количество работающих сотрудников = 1233

Соотношение количества работающих и уволенных сотрудников



Коэффициент текучести кадров = 16,12. В сфере торгово-промышленного производства коэффициент нормы текучести кадров в пределах нормы. Что является показателем стабильного функционирования организации. Для оценки использованы материалы статьи - [ссылка](#).

Распределение возраста по факту увольнения



На графике видно, что для обеих групп (уволившихся и оставшихся) наибольшая концентрация сотрудников наблюдается в возрасте около 30-40 лет. Это типично для большинства компаний.

3.2.1. Анализ с помощью KDE (Kernel Density Estimate)

Kernel Density Estimation (KDE) — это мощный статистический инструмент, который позволяет гибко и наглядно оценить функцию плотности вероятности. Применяя KDE к данным по уволенным и работающим сотрудникам, проанализируем распределение ключевых показателей и

определим, какие факторы оказывают наиболее значимое влияние на увольнение. Проведем проверку гипотез. Это даст четкое понимание взаимосвязей и поможет сделать выводы на основе визуализированных результатов.

Построение гипотез и интерпретация строится по данному шаблону:

Нулевая гипотеза (H_0): между показателями уволенных и работающих нет значимых различий.

Альтернативная гипотеза (H_1): различия между уволенными и работающими имеют разные распределения.

Критерий принятия решения:

Если $p\text{-value} < 0,05 \rightarrow$ отвергаем H_0 .

\Rightarrow Распределения различаются статистически значимо.

Если $p\text{-value} \geq 0.05 \rightarrow$ нет оснований отвергать H_0 .

\Rightarrow Распределения не различаются значимо.

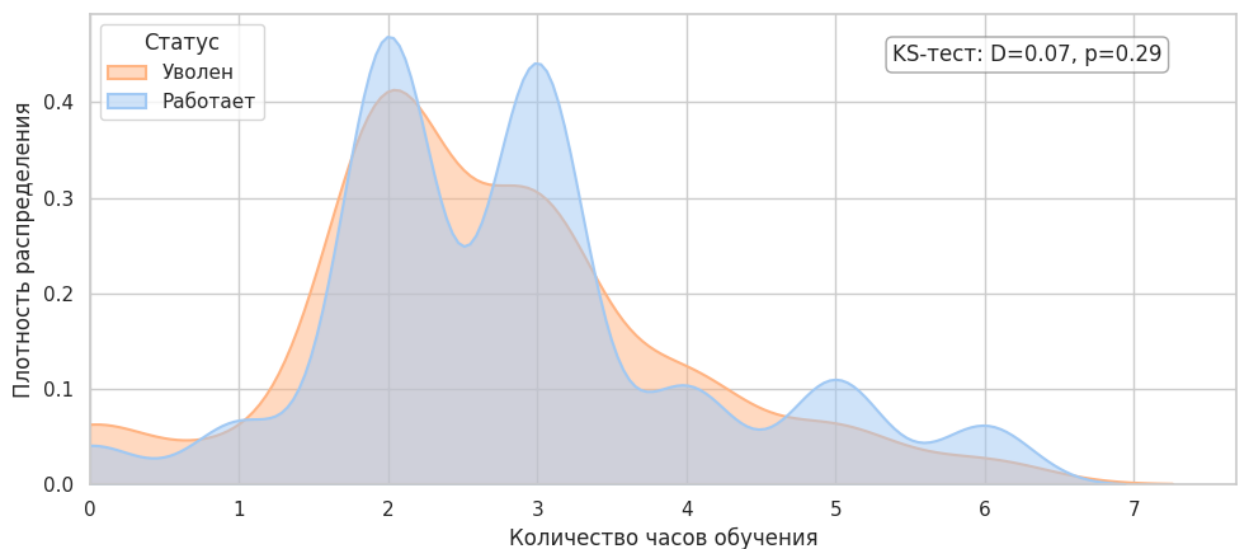
1) Влияние часов обучения:

Выдвигаем гипотезы:

H_0 – между количеством часов обучения уволенных и работающих нет значимых различий.

H_1 – различия количества часов обучения между уволенными и работающими имеют разные распределения.

Распределение количества часов обучения для работающих и уволенных сотрудников



$p\text{-value} = 0,29$ (это $\geq 0,05$) нет оснований отвергать $H_0 \Rightarrow$ распределения не различаются значимо.

Количество часов обучения между уволенными и работающими **значимо не различаются**.

Однако, стоит отметить, что на графике количество часов обучения для работающих и уволенных сотрудников имеются **различия в распределении**:

у уволенных сотрудников меньшее количество часов обучения по сравнению с работающими.

плотность распределения: пик плотности для уволенных сотрудников находится в области меньшего количества часов, в то время как у работающих сотрудников распределение смещено в сторону большего количества часов. С одной стороны это подчеркивает важность обучения для удержания сотрудников, с другой стоит помнить о том, что сотрудники увольнялись в течение года и соответственно проходили меньше обучения.

рекомендации: оптимизировать планирование обучения сотрудников. Необходимо рассмотреть возможность внедрения персонифицированного подхода к данному вопросу, таким образом, чтобы направлять на обучение людей в соответствии с проблематикой достижения поставленных целей. Это позволит сократить затраты на обучение и повысить эффективность труда. А также, будет являться рычагом к удержанию сотрудников через мотивацию к горизонтальному и вертикальному карьерному росту.

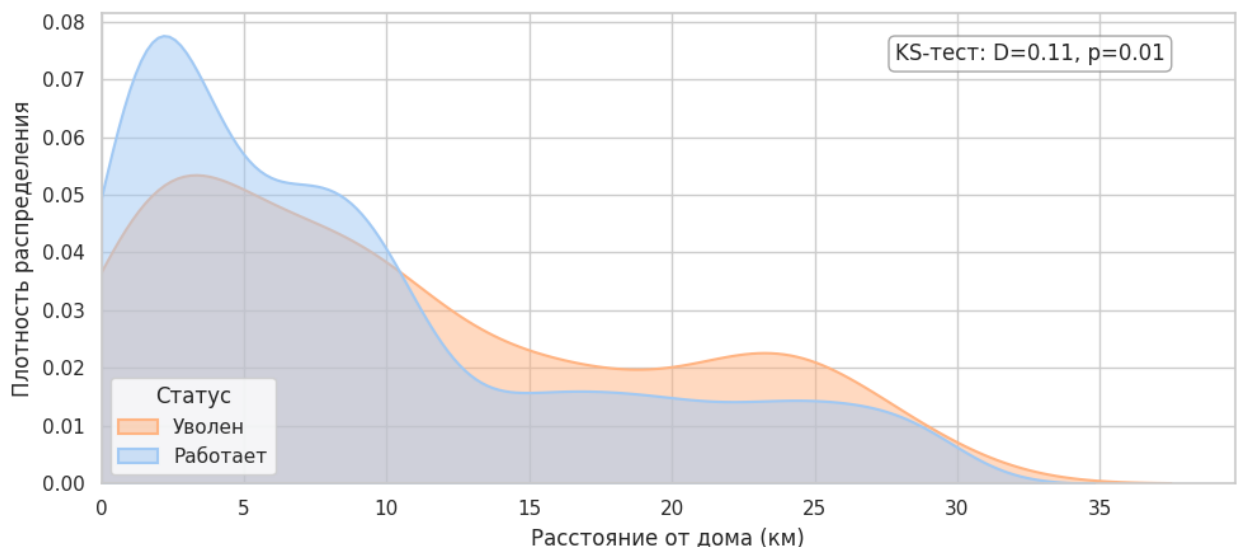
2) Влияние удаленности места работы от дома:

Выдвигаем гипотезы:

H_0 – между количеством км/миль от дома до работы уволенных и работающих нет значимых различий.

H_1 – различия количества км/миль от дома до работы между уволенными и работающими имеют разные распределения.

Распределение расстояния от дома до работы для работающих и уволенных сотрудников



$p\text{-value} = 0,01$ (это $< 0,05$) \Rightarrow Распределения различаются статистически значимо.

Расстояние от дома до работы статистически **значимо** связано с увольнением. Стоит рассмотреть введение мер поддержки.

На основе графика распределения расстояния от дома до работы для работающих и уволенных сотрудников можно сделать следующие выводы:

различия в распределении:

- уволенные сотрудники чаще проживают на большем расстоянии от работы (пик плотности смещен вправо по сравнению с работающими).
- работающие сотрудники в среднем живут ближе к месту работы (основная масса данных сосредоточена в диапазоне 0–13 км).

влияние расстояния на увольнения:

- большое расстояние от дома до работы может быть одним из факторов, способствующих увольнению, так как длительные поездки утомляют и снижают удовлетворенность работой.
- возможно, сотрудники, живущие дальше, чаще сталкиваются с логистическими трудностями (транспорт, опоздания), что повышает вероятность их ухода.

рекомендации:

- для снижения текучести кадров можно рассмотреть гибкий график работы, удаленный формат или компенсацию транспортных расходов для сотрудников, проживающих далеко.
- при найме новых сотрудников стоит учитывать их локацию, чтобы минимизировать риски, связанные с удаленностью.

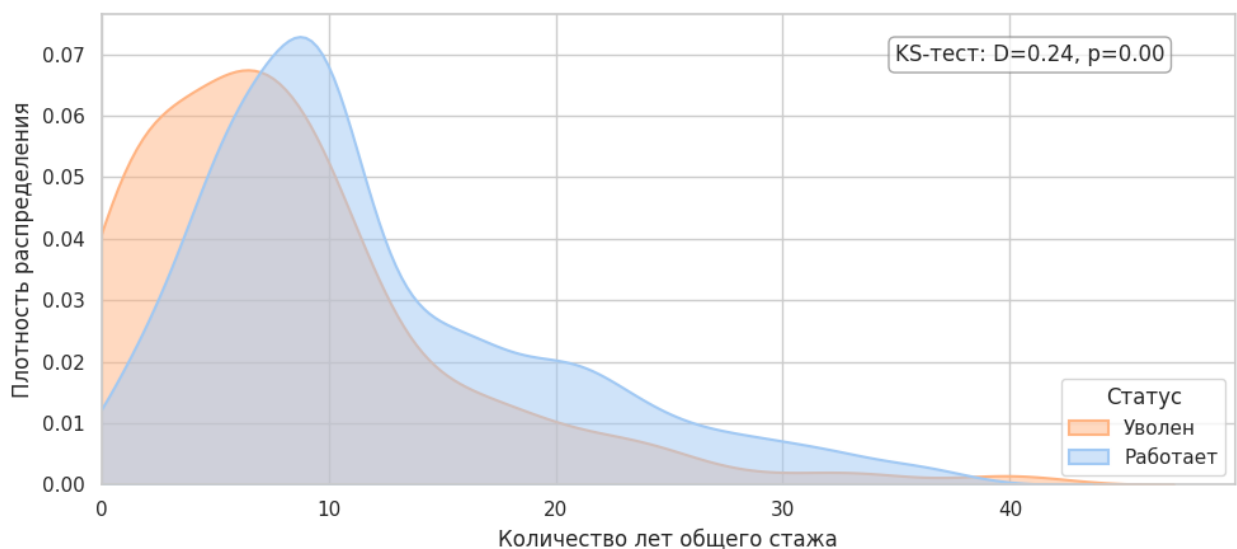
3) Влияние общего стажа:

Выдвигаем гипотезы:

H_0 – между количеством лет общего стажа уволенных и работающих нет значимых различий.

H_1 – различия количества лет общего стажа между уволенными и работающими имеют разные распределения.

Распределение общего стажа для работающих и уволенных сотрудников



$p\text{-value} = 0,00$ (это $< 0,05$) \Rightarrow Распределения различаются статистически значимо.

Общий стаж является **значимым** фактором текучести кадров.

На основе графика распределения общего стажа для работающих и уволенных сотрудников можно сделать следующие выводы:

различия в распределении стажа:

- работающие сотрудники демонстрируют более равномерное распределение с пиками на определенных значениях стажа, что может указывать на группы с высокой лояльностью (например, сотрудники с 5–10 годами стажа).
- уволенные сотрудники чаще имеют либо очень малый стаж (0–3 года), либо, наоборот, значительный (15+ лет). Это может говорить о двух ключевых группах риска:
 - новички — возможно, недостаточная адаптация или несоответствие ожиданиям.
 - опытные сотрудники — вероятно, связан с выгоранием, отсутствием карьерного роста или изменением условий работы.

возможные причины увольнений:

- малый стаж (0–3 года):
 - неудачный найм или слабая onboarding-программа.
 - несоответствие корпоративной культуре или требованиям роли.
- большой стаж (15+ лет):
 - отсутствие мотивации или развития.
 - конфликты с руководством/коллегами.
 - естественная ротация (пенсия, смена сферы деятельности).

рекомендации: компании стоит фокусироваться на двух крайних группах — сотрудниках с малым и большим стажем, чтобы снизить риски их увольнения.

- для новичков:
 - усилить программу адаптации.
 - регулярно собирать обратную связь в первые месяцы работы.
- для ветеранов:
 - предлагать новые challenges (горизонтальные перемещения, менторство).
 - пересматривать систему мотивации (бонусы, гибкий график).
- для всех:
 - анализировать причины увольнений в каждой группе стажа (exit-интервью).

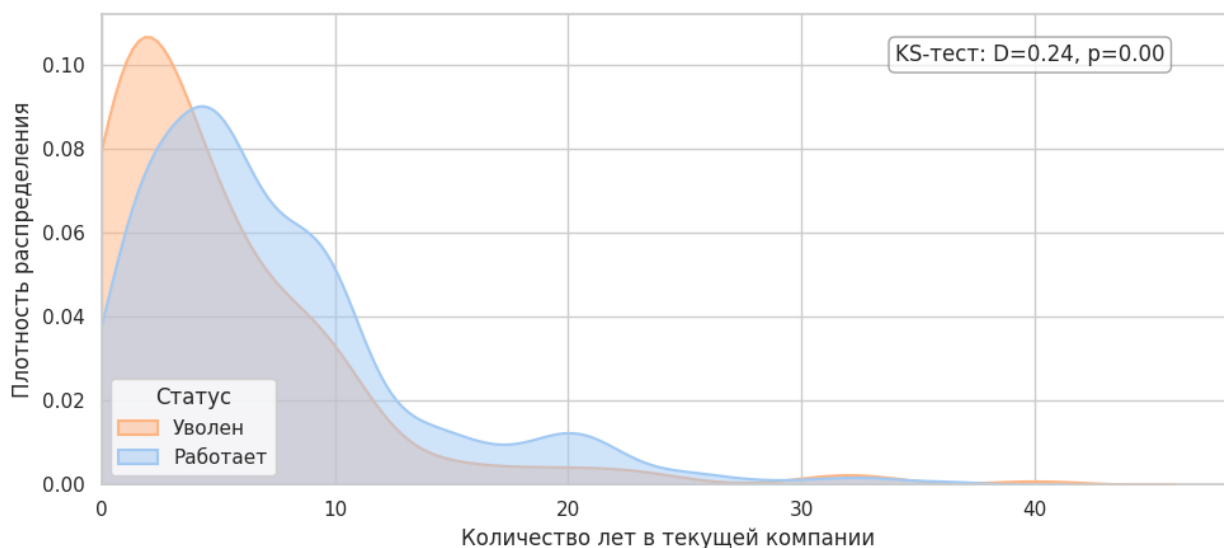
4) Влияния стажа в текущей компании:

Выдвигаем гипотезы:

H_0 — между количеством лет стажа в текущей компании уволенных и работающих нет значимых различий.

H_1 — различия количества лет в текущей компании между уволенными и работающими имеют разные распределения.

Распределение количества лет в текущей компании для работающих и уволенных сотрудников



$p\text{-value} = 0,00$ (это $< 0,05$) \Rightarrow Распределения различаются статистически значимо.

Количество лет в текущей компании является **значимым** фактором текучести кадров.

На основании графика можно сделать следующие выводы:

ключевые закономерности распределения:

- для уволенных сотрудников наблюдается явный пик в первые 1-3 года работы в компании. Это указывает на наиболее высокий риск увольнения в начальный период.
- для работающих сотрудников распределение более равномерное с небольшими пиками на определенных сроках работы (возможно, 5-7 лет), что может свидетельствовать о периодах повышенной лояльности.

интерпретация результатов:

- период наибольшего риска (0-3 года):
 - проблемы адаптации и вхождения в должность
 - несоответствие ожиданий и реальности
 - ошибки в подборе персонала
- после 3 лет работы:
 - риск увольнения существенно снижается
 - сотрудники, перешагнувшие этот рубеж, демонстрируют большую приверженность компании.

рекомендации по управлению персоналом: первые 1-3 года работы в компании критически важный период с наибольшим риском увольнения. Компании следует сосредоточить усилия на улучшении процессов адаптации и поддержки новых сотрудников в этот период, что позволит существенно снизить текучесть кадров. Для сотрудников, проработавших более 3 лет, важно поддерживать мотивацию и предоставлять возможности развития:

- усилить программу адаптации:
 - ввести систему наставничества;
 - улучшить onboarding-процессы;

- регулярно собирать обратную связь в первые месяцы;
- разработать меры удержания для критического периода:
 - четкое планирование карьеры после испытательного срока;
 - система мотивации для молодых специалистов;
 - программы развития и обучения;
- анализ причин увольнения:
 - проводить детальные exit-интервью;
 - выявлять системные проблемы в подборе и адаптации.

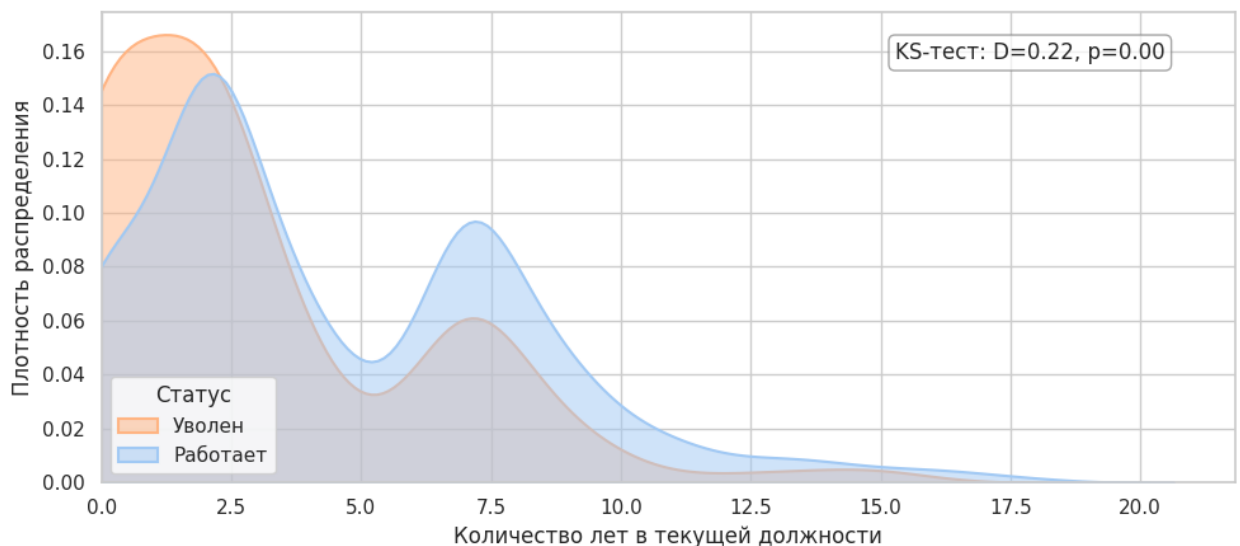
5) Влияние стажа в текущей должности:

Выдвигаем гипотезы:

H_0 – между количеством лет стажа в текущей должности уволенных и работающих нет значимых различий.

H_1 – различия количества лет в текущей должности между уволенными и работающими имеют разные распределения.

Распределение количества лет в текущей должности для работающих и уволенных сотрудников



$p\text{-value} = 0,00$ (это $< 0,05$) \Rightarrow Распределения различаются статистически значимо.

Количество лет в текущей должности является **значимым** фактором текучести кадров.

На основе графика можно выделить следующие закономерности и сделать выводы:

основные тенденции распределения:

- пик увольнений приходится на первые 1-3 года работы в должности, с резким спадом после этого периода.
- работающие сотрудники демонстрируют более равномерное распределение с небольшими пиками на 5-7 и 10-12 лет, что может указывать на периоды стабильности.
- после 7 лет в одной должности количество увольнений существенно снижается, но небольшой рост наблюдается на 15+ годах.

критические периоды и их интерпретация:

высокий риск увольнения (0-3 года): период адаптации:

- несоответствие ожиданий и реальности должности;
- сложности в освоении новых обязанностей;
- недостаточная поддержка в начале работы;

средний риск (3-7 лет) период профессиональной зрелости:

- возможное выгорание;
- отсутствие карьерных перспектив;
- потребность в новых вызовах;

низкий риск (7-15 лет) период стабильности:

- высокая лояльность компании;
- удовлетворенность положением;
- хорошая адаптация к корпоративной культуре;

повторный рост риска (15+ лет) поздний карьерный этап:

- возможное профессиональное выгорание;
- изменение личных приоритетов;
- естественная текучесть перед пенсией;

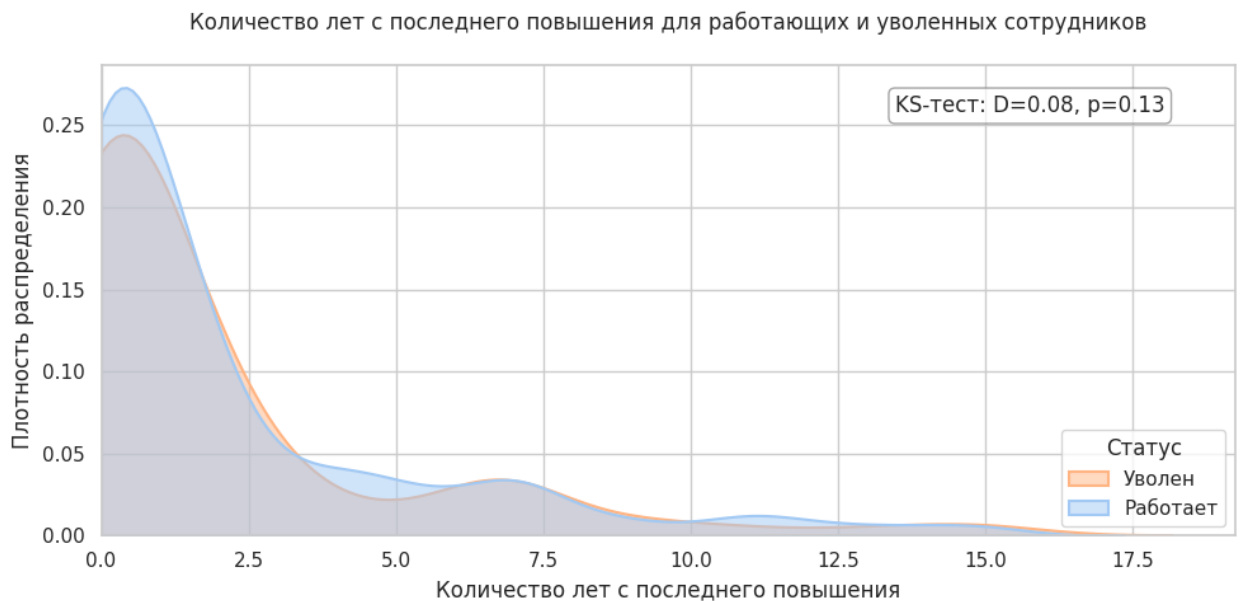
рекомендации: компании следует уделять особое внимание процессам адаптации и поддержки новых сотрудников, а также разрабатывать программы развития для тех, кто находится в должности 3-7 лет. Для долгосрочных сотрудников важны программы признания их заслуг и возможности передачи опыта. Такой дифференцированный подход позволит значительно снизить текучесть кадров на всех этапах работы в должности.

- для новых сотрудников в должности (0-3 года):
 - разработать программу адаптации;
 - ввести систему регулярной обратной связи;
 - обеспечить четкое понимание карьерных перспектив;
- для сотрудников 3-7 лет:
 - предлагать ротацию или горизонтальные перемещения;
 - ввести программы профессионального развития;
 - обеспечить участие в интересных проектах;
- для долгосрочных сотрудников (7+ лет):
 - разработать программы наставничества;
 - предоставлять возможность обучать других;
 - рассмотреть возможность вертикального роста;
- общие меры:
 - Внедрить систему мониторинга удовлетворенности;
 - Проводить регулярные карьерные собеседования;
 - Анализировать причины увольнений по стажу.

6) Влияние периода с последнего повышения:

H_0 – между количеством лет с последнего повышения уволенных и работающих нет значимых различий.

H_1 – различия количества лет с последнего повышения между уволенными и работающими имеют разные распределения.



$p\text{-value} = 0,29$ (это $\geq 0,05$) нет оснований отвергать H_0 . \Rightarrow Распределения не различаются значимо.

Количество лет с последнего повышения различается статистически **не значимо**.

Рекомендации: компании все же стоит выстроить систему, где сотрудники будут получать регулярные карьерные изменения (не обязательно вертикальные) с интервалом не более 3-4 лет (оптимально 2-3 года):

- внедрить систему регулярного карьерного планирования (не реже 1 раза в год);
- установить прозрачные критерии повышений;
- для "застойных" сотрудников (5+ лет без повышения):
 - проводить специальные карьерные собеседования;
 - предлагать альтернативные формы профессионального роста;
 - рассматривать возможность ротации.

Рассмотреть возможность внедрения альтернативных форм признания (для случаев, когда повышение невозможно):

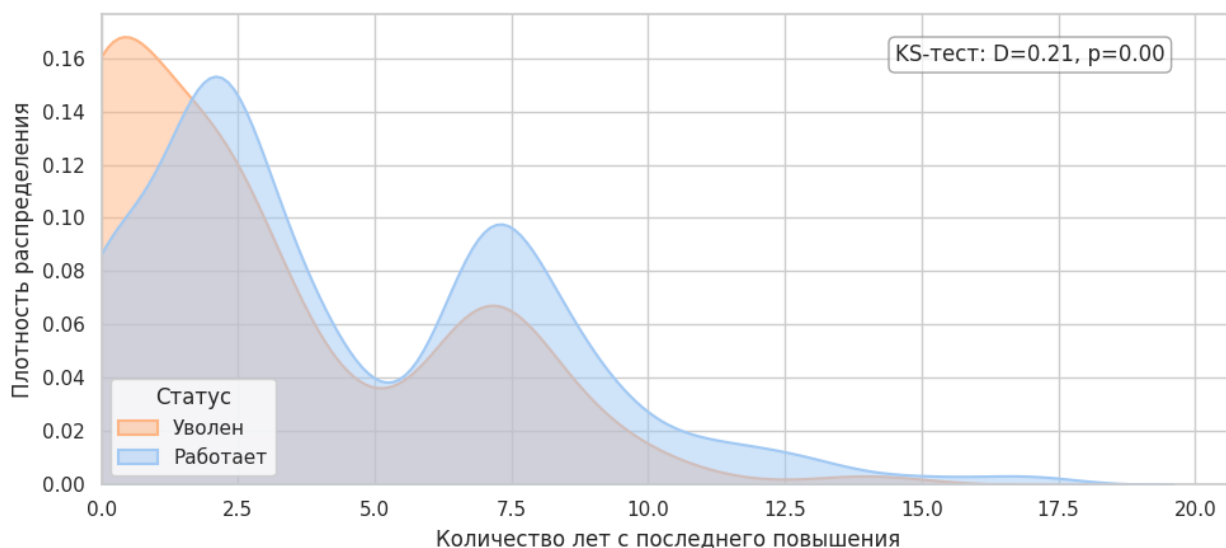
- горизонтальные перемещения;
- участие в специальных проектах;
- обучение и развитие новых навыков;
- нематериальная мотивация.

7) Влияние работы с одним менеджером:

H_0 – между количеством лет работы с одним менеджером уволенных и работающих нет значимых различий.

H_1 – различия количества лет работы с одним менеджером между уволенными и работающими имеют разные распределения.

Распределение лет работы с одним менеджером для работающих и уволенных сотрудников



$p\text{-value} = 0,00$ (это $< 0,05$) \Rightarrow Распределения различаются статистически значимо.

Количество лет с одним менеджером является **значимым** фактором текучести кадров.

На основе графика можно выделить следующие закономерности и сделать выводы:

основные закономерности:

- пик увольнений наблюдается в первые 1-3 года работы с одним руководителем
- работающие сотрудники демонстрируют более равномерное распределение с двумя выраженными пиками:
 - первый пик: 0,5-2 года (период адаптации к менеджеру);
 - второй пик: 5-7 лет (стабильные рабочие отношения);
 - после 7 лет работы с одним менеджером количество увольнений существенно снижается.

практические выводы:

адаптационный период (0-6 месяцев):

- внедрить систему взаимной адаптации сотрудника и менеджера;
- проводить регулярные встречи по обратной связи;
- четко формулировать ожидания с обеих сторон

период 1-3 года:

- развивать программы менторства и коучинга;
- обеспечивать карьерное планирование;
- мониторить удовлетворенность сотрудников стилем руководства;

долгосрочные отношения (5+ лет):

- разработать программы ротации менеджеров;
- предотвращать профессиональное выгорание;
- создавать кросс-функциональные проекты для разнообразия

рекомендации: первые 3 года работы с одним менеджером являются наиболее критическими с точки зрения увольнений. Компании следует уделять особое

внимание процессу адаптации сотрудников к новым руководителям и активно работать с парами "сотрудник-менеджер" в течение первых лет совместной работы. Долгосрочные рабочие отношения (5+ лет) показывают высокую стабильность, однако требуют специальных мер по поддержанию мотивации и предотвращению профессионального застоя:

- внедрить систему оценки совместимости сотрудника и менеджера;
- разработать программы развития управленческих навыков;
- обеспечить возможность смены руководителя при необходимости;
- регулярно измерять индекс удовлетворенности сотрудников своим руководством.

3.2.2. Анализ сотрудников в процентном и долевым отношениях

Осуществляется с помощью формулы `calculate_attrition_stats`, которая предоставляет возможность анализировать сотрудников в сравнении (уволенных и работающих) по интересующим параметрам в процентах и долях. Дополнительно в функцию заложена возможность применения различных метрик и фильтрации. Такая вариативность позволяет ответить практически на любой вопрос заказчиков и/или стейкхолдеров.

Проанализируем процент увольнений в разрезе отделов и гендера:

	Department	Gender	Attrition	count	count_share	count_share_pct
0	Human Resources	Female	0	14	0.70	70.00
1	Human Resources	Female	1	6	0.30	30.00
2	Human Resources	Male	0	37	0.86	86.05
3	Human Resources	Male	1	6	0.14	13.95
4	Research & Development	Female	0	336	0.89	88.65
5	Research & Development	Female	1	43	0.11	11.35
6	Research & Development	Male	0	492	0.85	84.54
7	Research & Development	Male	1	90	0.15	15.46
8	Sales	Female	0	151	0.80	79.89
9	Sales	Female	1	38	0.20	20.11
10	Sales	Male	0	203	0.79	78.99
11	Sales	Male	1	54	0.21	21.01

Общая картина:

- самый высокий процент увольнений наблюдается в отделе продаж (Sales):
 - 20,11% среди женщин (Female)
 - 21,01% среди мужчин (Male)

- самый низкий процент увольнений в отделе исследований и разработке (R&D):
 - 11,35% среди женщин
 - 15,46% среди мужчин
- отдел кадровых ресурсов (HR) занимает промежуточное положение:
 - 30% увольнений среди женщин (высокий показатель!)
 - 13,95% среди мужчин

Вывод: в продажах текучесть кадров максимальна, а в R&D сотрудники остаются чаще. HR-женщины увольняются в 2 раза чаще, чем мужчины в том же департаменте.

Гендерные различия в увольнениях:

- наибольший гендерный разрыв в HR (разница ~16%) – возможно, проблемы с условиями работы, карьерными возможностями или балансом работы и личной жизни.
- в Продажах гендерный разрыв минимален (~1%), что говорит о схожих условиях для мужчин и женщин.

Вывод:

- HR-отделу стоит обратить внимание на женский коллектив – почему женщины уходят в 2 раза чаще?
- В R&D и Продажах гендерный дисбаланс в увольнениях незначителен.

Анализ по отделам:

Отдел	Женщины (Attrition)	Мужчины (Attrition)	Общий вывод
HR	30% (высокий)	13,95% (средний)	Женщины уходят в 2 раза чаще
R&D	11,35% (низкий)	15,46% (средний)	Самый стабильный отдел, мужчины уходят чуть чаще
Sales	20,11% (высокий)	21,01% (высокий)	Максимальная текучесть, гендерный паритет

Вывод:

- Sales – зона риска по текучести кадров (вероятно, высокий стресс, KPI, конкуренция).
- HR-женщины – особая группа риска (нужен дополнительный анализ причин).
- R&D – самый стабильный департамент, но мужчины уходят немного чаще женщин.

Рекомендации:

Отдел продаж:

- внедрить программы удержания (бонусы, менторство, снижение стресса);
- провести exit-интервью, чтобы понять причины увольнений

Отдел кадровых ресурсов (особенно женская часть):

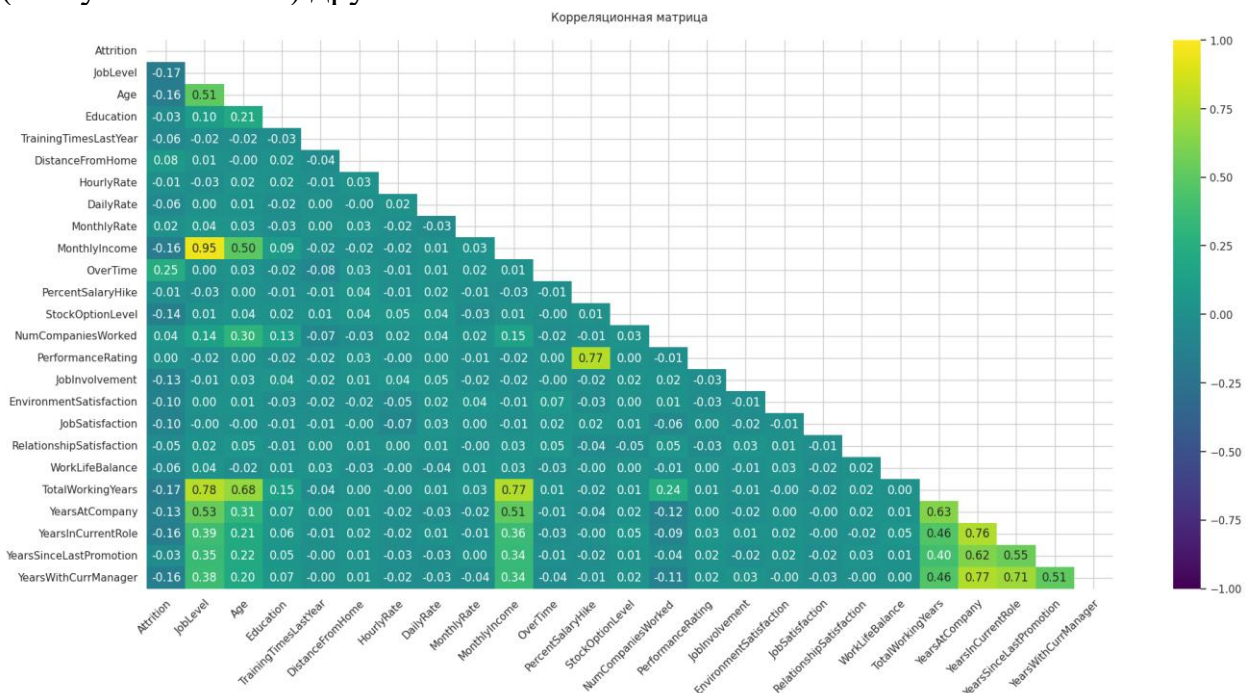
- провести опрос среди женщин HR: что не устраивает?
- рассмотреть гибкий график, карьерные возможности, оценку нагрузки

Отдел исследования и разработки:

- несмотря на низкую текучесть, мужчины уходят чаще – изучить причины;
- возможно, влияет зарплата или отсутствие карьерного роста.

3.2.3. Анализ с помощью корреляционной матрицы

Анализ корреляционной матрицы помогает понять, как различные переменные в наборе данных связаны друг с другом. Увеличение (или уменьшение) значений одного показателя ведет к закономерному увеличению (или уменьшению) другого показателя.



В данном датасете установлен ряд зависимостей (положительная корреляция) между:

«Уровнем работы» и «Ежемесячным доходом» - 0,95 – *очень высокая корреляция*;

высокая корреляция:

«Общим стажем работы» и «Уровнем работы» - 0,78;

«Общим стажем работы» и «Ежемесячным доходом» - 0,77;

«Рейтингом производительности» и «Процентом повышения зарплаты» - 0,77;

средняя корреляция:

«Количеством лет в компании» и «Уровнем работы» - 0,53;

«Возрастом» и «Уровнем работы» - 0,51;

«Ежемесячным доходом» и «Возрастом» - 0,50;

«Количеством лет в компании» и «Ежемесячным доходом» - 0,51

Для оценки использованы данные [статьи](#).

ЗНАЧЕНИЕ (по модулю)	ИНТЕРПРЕТАЦИЯ
до 0,2	очень слабая корреляция
до 0,5	слабая корреляция
до 0,7	средняя корреляция
до 0,9	высокая корреляция
свыше 0,9	очень высокая корреляция

БЛОК 4 ПОСТРОЕНИЕ МОДЕЛИ

4.1. Модель машинного обучения. Логистическая регрессия

Цель построения данной модели — прогнозирование вероятности увольнения сотрудников (Attrition) на основе их характеристик и факторов рабочей среды. Прогнозирование риска увольнения.

1) Датафрейм содержит следующие ключевые признаки:

Категориальные: Department, JobRole, BusinessTravel, Gender, EducationField, MaritalStatus, OverTime.

Числовые: Age, DistanceFromHome, MonthlyIncome, YearsAtCompany, YearsSinceLastPromotion, JobLevel, Education, TrainingTimesLastYear, HourlyRate, DailyRate, MonthlyRate, PercentSalaryHike, StockOptionLevel, NumCompaniesWorked, PerformanceRating, JobInvolvement, EnvironmentSatisfaction, JobSatisfaction, RelationshipSatisfaction, WorkLifeBalance, TotalWorkingYears, YearsInCurrentRole, YearsWithCurrManager

Целевая переменная: Attrition (бинарная: Yes/No)

Предобработка включает в себя:

- кодирование категориальных признаков: применён One-Hot Encoding с drop_first=True для устранения мультиколлинеарности;
- масштабирование числовых признаков: стандартизация через StandardScaler.

2) Построение модели:

Разделяем данные на: 80% — тренировочная выборка, 20% — тестовая (random_state=42). Стратификация по целевой переменной для сохранения баланса.

Обучение модели логистической регрессии с max_iter=1000 для гарантии сходимости.

3) Оценка качества прогнозной модели логистической регрессии:

Confusion Matrix:

```
[[244  11]
 [ 22  17]]
```

Classification Report:

	precision	recall	f1-score	support
False	0.92	0.96	0.94	255
True	0.61	0.44	0.51	39
accuracy			0.89	294
macro avg	0.76	0.70	0.72	294
weighted avg	0.88	0.89	0.88	294

	Metric	Value
0	Train Accuracy	0.89
1	Test Accuracy	0.89
2	Precision (class 1)	0.61
3	Recall (class 1)	0.44
4	F1 (class 1)	0.51
5	ROC-AUC Score	0.80

Accuracy (Точность) – модель корректно предсказывает 89% случаев. Однако, при дисбалансе классов (например, мало увольнений) ассигасу может вводить в заблуждение.

Классы распределены неравномерно:

- False (класс 0) — 255 примеров,
- True (класс 1) — 39 примеров.

Это влияет на метрики для класса True, так как модель склонна лучше предсказывать

доминирующий класс.

Precision (Точность предсказаний) – из всех предсказанных увольнений 61% действительно были увольнениями. Это означает, что модель часто ошибается, предсказывая True (ложные срабатывания).

Recall (Полнота) – модель обнаруживает только 44% реальных увольнений, что указывает на высокую долю пропущенных случаев (FN).

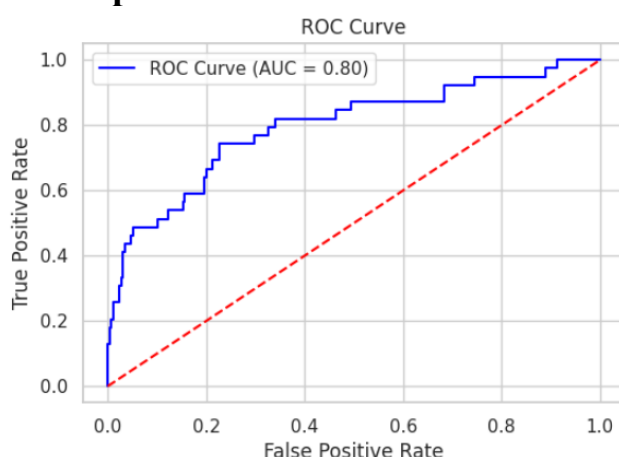
F1 Score – гармоническое среднее precision и recall. Низкое значение F1 указывает на дисбаланс между точностью и полнотой.

ROC-AUC Score - AUC показывает хорошую способность модели разделять классы, т.е. модель имеет хорошую предсказательную способность, но не идеальную. Это согласуется с низким recall для True.

Train и **Test** Ассигасу совпадают (89%), что указывает на отсутствие переобучения. Однако это может быть связано с доминированием класса False.

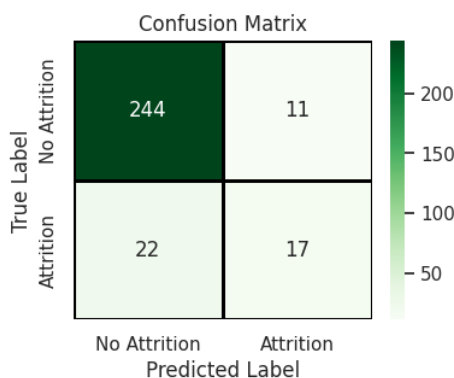
Итог: Модель хорошо предсказывает класс False, но плохо справляется с классом True.

ROC-кривая AUC:



Результат ROC-кривой AUC (Area Under the Curve) равное 0,80 указывает на то, что модель имеет высокую предсказательную силу, так как значение близко к 1,0 (идеальный классификатор).

Матрица неточностей:



True Negative (TN): 244 (корректные предсказания отсутствия увольнения)

False Positive (FP): 11 (ошибочные предсказания увольнения)

False Negative (FN): 22 (ошибочные предсказания отсутствия увольнения)

True Positive (TP): 17 (корректные предсказания увольнения)

Сильные стороны модели:

Высокий ROC-AUC (0,80) указывает на хорошую способность модели различать классы.

Умеренная точность (precision = 0,61) означает, что, когда модель предсказывает увольнение, она часто права.

Слабые стороны:

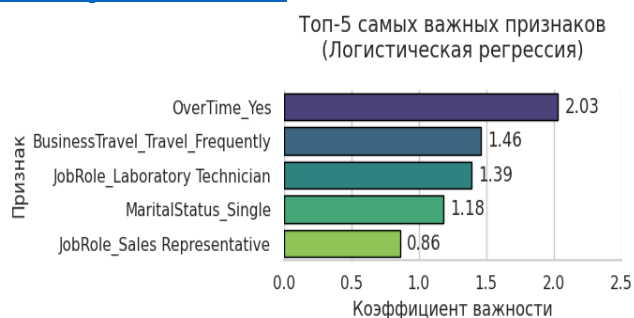
Низкий recall (0,44) — модель пропускает много реальных увольнений (FN = 22).

Низкий F1 Score (0,51) указывает на несбалансированность precision и recall.

Итог: Модель имеет приемлемое качество, но требует доработки для снижения количества пропущенных увольнений (улучшение recall).

4.2. Топ-5 наиболее важных признаков для увольнений:

	Feature	Coefficient
43	OverTime_Yes	1.79
33	BusinessTravel_Travel_Frequently	1.55
26	JobRole_Laboratory Technician	1.31
32	JobRole_Sales Representative	0.96
42	MaritalStatus_Single	0.68



Сверхурочная работа (OverTime_Yes) — самый значимый фактор. Сотрудники, работающие сверхурочно чаще увольняются.

Рекомендация: пересмотреть нагрузку, ввести гибкий график или компенсации за переработки.

Частые командировки (BusinessTravel_Travel_Frequently) — второй по важности фактор. Регулярные разъезды повышают риск ухода, вероятно, из-за усталости или дисбаланса работы и личной жизни.

Рекомендация: оптимизировать график командировок, добавить дополнительные дни отдыха или бонусы за мобильность.

Должность "Лаборант" (JobRole_Laboratory Technician) – третий ключевой фактор. Специалисты в лабораториях уходят чаще других. Возможные причины: низкая зарплата, монотонность работы, недостаток карьерного роста.

Рекомендация: проанализировать условия труда, зарплаты и возможности развития для этой роли.

Должность "Торговый представитель" (JobRole_Sales Representative). У сотрудников на этой позиции риск увольнения почти вдвое выше (по сравнению с низкорисковыми группами). Возможные причины: высокий стресс, нестабильный доход, давление KPI.

Рекомендация: улучшить систему мотивации, ввести поддержку менторов, снизить turnover за счет адаптации новичков.

Холостые сотрудники (MaritalStatus_Single). Неженатые/незамужние сотрудники увольняются чаще. Возможные причины: меньшая привязанность к месту (легче сменить работу), отсутствие финансовых обязательств (ипотека, семья).

Рекомендация: для удержания таких сотрудников важны нематериальные факторы – корпоративная культура, развитие, комфортная атмосфера.

Общий вывод:

Основные драйверы увольнений связаны с перегрузками (сверхурочные, командировки) и специфическими должностями (лаборанты, продавцы). Также важна личная ситуация сотрудников (семейное положение).

Рекомендации:

- 1) Снижать нагрузку на сотрудников в зоне риска.
- 2) Пересмотреть политику командировок и переработок.
- 3) Улучшать условия для "проблемных" должностей (лаборантов, продавцов).
- 4) Укреплять вовлеченность холостых сотрудников через развитие и корпоративные ценности.
- 5) Если компания сфокусируется на этих факторах, можно значительно снизить текучесть кадров.

4.3. Анализ распределения работающих по группам риска:

Для всего набора данных результат таков:

Risk_Group	
Low_Risk	85.9%
Medium Risk	9.9%
High Risk	4.1%

Однако, нас больше интересуют риски среди работающих:

Risk_Group	
Low_Risk	94.0%
Medium Risk	5.8%
High Risk	0.2%

Выводы:

Подавляющее большинство сотрудников находятся в зоне низкого риска увольнения. Это говорит о стабильности кадрового состава.

Средний и высокий риск охватывают 6% работающих сотрудников. Это доля, требующая внимания:

High Risk (0,2%) – критическая группа, где увольнение вероятно в ближайшее время.

Количество сотрудников в зоне высокого риска увольнения: 3

Средняя вероятность увольнения: 76.32%

Средний уровень удовлетворенности работой: 3.0

Отделы в группе высокого риска:

Research & Development	2
Sales	1

Должности в группе высокого риска:

Research Director	1
Sales Representative	1
Research Scientist	1

Рекомендации: необходим анализ причин (конфликты, перегрузки, низкая зарплата) и индивидуальные меры удержания (например, пересмотр обязанностей, бонусы).

Medium Risk (5,8%) – группа, где возможны признаки неудовлетворенности.

Количество сотрудников в зоне среднего риска увольнения: 71

Средняя вероятность увольнения: 50.60%

Средний уровень удовлетворенности работой: 2.380281690140845

Отделы в группе среднего риска:

Research & Development	40
Sales	28
Human Resources	3

Должности в группе среднего риска:

Sales Executive	19
Research Scientist	17
Laboratory Technician	16
Sales Representative	8
Manufacturing Director	5
Human Resources	3
Manager	2
Healthcare Representative	1

Рекомендации: провести анкетирование или интервью, чтобы выявить и устранить/минимизировать причины влияющие на решение увольнения.

Профилактика важнее «тушения пожаров». Хотя доля High Risk мала, их потеря может быть критичной для ключевых ролей. Для Medium Risk полезны программы развития и вовлеченности, чтобы предотвратить переход в High Risk.

Данная модель логистической регрессии может быть эффективно использована **для единичного нового наблюдения**.

При указании значений параметров модель предсказывает будет ли увольнение или нет и рассчитывает процент риска.

ИТОГИ ПРОЕКТА И ЗАКЛЮЧЕНИЕ

В ходе проведенного анализа данных сотрудников компании были выявлены ключевые закономерности, причины текучести кадров и потенциальные направления для улучшения ситуации.

Резюмируя полученные данные, можно составить портрет наиболее явного кандидата на увольнение: молодой человек (до 40 лет), занимающий невысокую должность, выполняющий однообразную работу, требующую большого расхода эмоционально-физических ресурсов. У данного индивида явно не высокая заработная плата и имеются сложности в личной жизни.

Остальные параметры имеющиеся в датафрейме безусловно оказывают влияние на принадлежность сотрудника к организации и каждый из них заслуживает внимания. Рекомендации стейкхолдерам прописаны в документации по каждому критерию предметно. Обобщая их, можно предложить ряд стратегических мер позволяющих компании достичь желаемых результатов, за счет снижения текучести кадров и перенаправить коллектив в русло развития и достижений:

1. Разработать программу поддержки для удаленных работников;
2. Запустить программу интеграции и раскрытия потенциала молодых сотрудников через систему тимбилдинга, mentorship, карьерных челленджей и культурных кодов компании, формируя долгосрочную лояльность;
3. Разработать персонализированную систему обучения для карьерного роста с опциями корпоративного/частного финансирования, обязательной отработкой и сохранением права на результаты работы за организацией;
4. Внедрить систему мотивационных командировок с элементами геймификации, гибкими условиями и возможностью профессионального нетворкинга и бонусами за достижения;
5. Внедрить KPI по принципу «30 секунд»: если сотрудник не может объяснить свои показатели за полминуты – система подлежит пересмотру. Этапы: аудит → совместная разработка с отделами → пилот в одном департаменте + фидбек → корректировка и масштабирование на всю компанию;
6. Разработать и внедрить программу экспертности, включающую в себя систему поддержки и мотивацию сотрудников со стажем: кадровый резерв на стратегические роли (советники, аудиторы); персональные бонусы за передачу экспертизы (например, за подготовку методичек); гибкий график или удалёнка по запросу; возможность выбрать себе командировку-мечту; ежегодный «День эксперта» - когда стажисты выбирают тему и проводят мастер-класс для всей компании (с трансляцией и памятным сертификатами) и т.п.

Для улучшения качества данных датафрейма необходимо доработать отчет и дополнить параметрами со следующей информацией:

- **дата принятия на работу и дата увольнения** – даст возможность отследить наиболее пиковые моменты, вероятность сезонности. И на основе этого подготовить необходимые меры.
- **наличие иждивенцев** – данный показатель может влиять на работоспособность сотрудника, это позволит оптимизировать баланс работы и личной жизни и улучшить адаптацию.
- **даты повышения квалификации (обучения)** – позволит сбалансировать план обучения без ущерба для компании.
- **причина увольнения** в двух проекциях: со слов работника, по мнению организации – позволит выявить причины увольнения с субъективной точки зрения сотрудников, а также можно будет установить корреляционные связи с иными параметрами датасета. Анализируя мнение организации возможно будут выявлены новые метрики, влияющие на удержание работников в компании.