

HR-аналитика: увольнение сотрудников

Данный блокнот создан, чтобы выяснить причины увольнения сотрудников, основываясь на 34 показателях, таких как: возраст, удовлетворение от работы, уровень зарплаты, баланс работы и личной жизни и т.д.

(Это вымышленный набор данных, созданный специалистами по данным IBM. Данные открытые. Ссылка на датасет: <https://www.kaggle.com/pavansubhasht/ibm-hr-analytics-attrition-dataset>)

Посмотрим все данные, с которыми работаем:

a

SELECT
*
FROM
ibm_hr.hr_red

Untitled

Age	Attrition	BusinessTravel	DailyRate	Department	DistanceFromHome
48	Work	Travel_Rarely	530	Sales	29
45	Work	Travel_Frequently	364	Research & Development	25
41	Work	Travel_Rarely	896	Sales	6
57	Work	Travel_Rarely	210	Sales	29
42	Work	Travel_Rarely	265	Sales	5

35 columns, 1 470 rows

Некоторые категориальные данные в таблице заданы числами и расшифровываются так:

Образование

- 1 «Ниже колледжа»
- 2 «Колледж»
- 3 «Бакалавр»
- 4 «Магистр»
- 5 «Доктор»

Удовлетворение от условий работы

- 1 «Низкая»
- 2 «Средняя»
- 3 «Высокая»
- 4 «Очень высокая»

Вовлеченность в работу

- 1 «Низкая»
- 2 «Средняя»
- 3 «Высокая»
- 4 «Очень высокая»

Удовлетворение от работы

- 1 «Низкая»
- 2 «Средняя»
- 3 «Высокая»
- 4 «Очень высокая»

Оценка работы

- 1 «Низкая»
- 2 «Хорошо»
- 3 «Отлично»
- 4 «Великолепно»

Удовлетворение от отношений

- 1 «Низкая»
- 2 «Средняя»
- 3 «Высокая»
- 4 «Очень высокая»

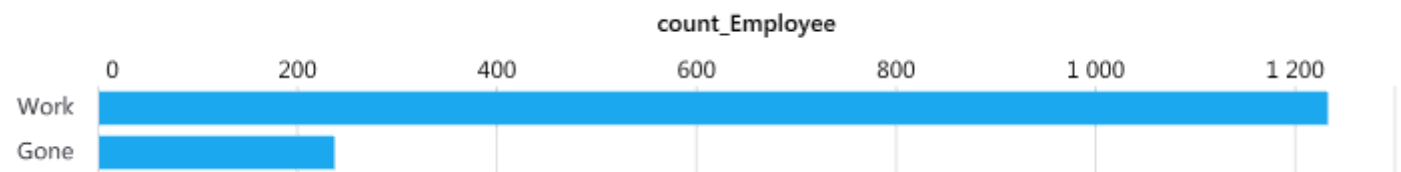
WorkLifeBalance

- 1 «Плохо»
- 2 «Хорошо»
- 3 «Чуть лучше»
- 4 «Лучше всех»

m

```
select Attrition, count(*) as count_Employee
from ibm_hr.hr_red
group by Attrition
```

Количество уволившихся и работающих



В базе данных представлены данные о 237 уволившихся и 1233 еще работающих сотрудниках.

Посмотрим средние значения у уволившихся сотрудников:

b

```
SELECT
  avg(Age) AS avg_Age, avg(MonthlyIncome) as avg_MonthlyIncome, avg(DistanceFromHome) as
avg_DistanceFromHome, avg(YearsAtCompany) as avg_YearsAtCompany, avg(YearsInCurrentRole) as
avg_YearsInCurrentRole, avg(YearsWithCurrManager) as avg_YearsWithCurrManager
FROM
  ibm_hr.hr_red
WHERE
  (Attrition = 'Gone')
```

Средние значения уволившихся сотрудников:

avg_Age	avg_MonthlyIncome	avg_DistanceFromHome	avg_YearsAtCompany	avg_YearsInCurrentRole
33,608	4 787,093	10,633	5,131	2,903
6 columns, 1 row				

И средние значения еще работающих сотрудников:

r

```
SELECT
  avg(Age) AS avg_Age, avg(MonthlyIncome) as avg_MonthlyIncome, avg(DistanceFromHome) as
avg_DistanceFromHome, avg(YearsAtCompany) as avg_YearsAtCompany, avg(YearsInCurrentRole) as
avg_YearsInCurrentRole, avg(YearsWithCurrManager) as avg_YearsWithCurrManager
FROM
  ibm_hr.hr_red
WHERE
  (Attrition = 'Work')
```

Untitled

avg_Age	avg_MonthlyIncome	avg_DistanceFromHome	avg_YearsAtCompany	avg_YearsInCurrentRole
37,561	6 832,74	8,916	7,369	4,484
6 columns, 1 row				

Судя по средним значениям, можно увидеть различие между уволившимися сотрудниками и работающими:

- увольняются более молодые люди,
- уволившиеся сотрудники получают в среднем меньше,
- уволившиеся сотрудники в среднем живут дальше от работы,
- уволившиеся сотрудники в среднем проработали меньше в компании.

Посмотрим, сотрудники какого возраста чаще увольняются:

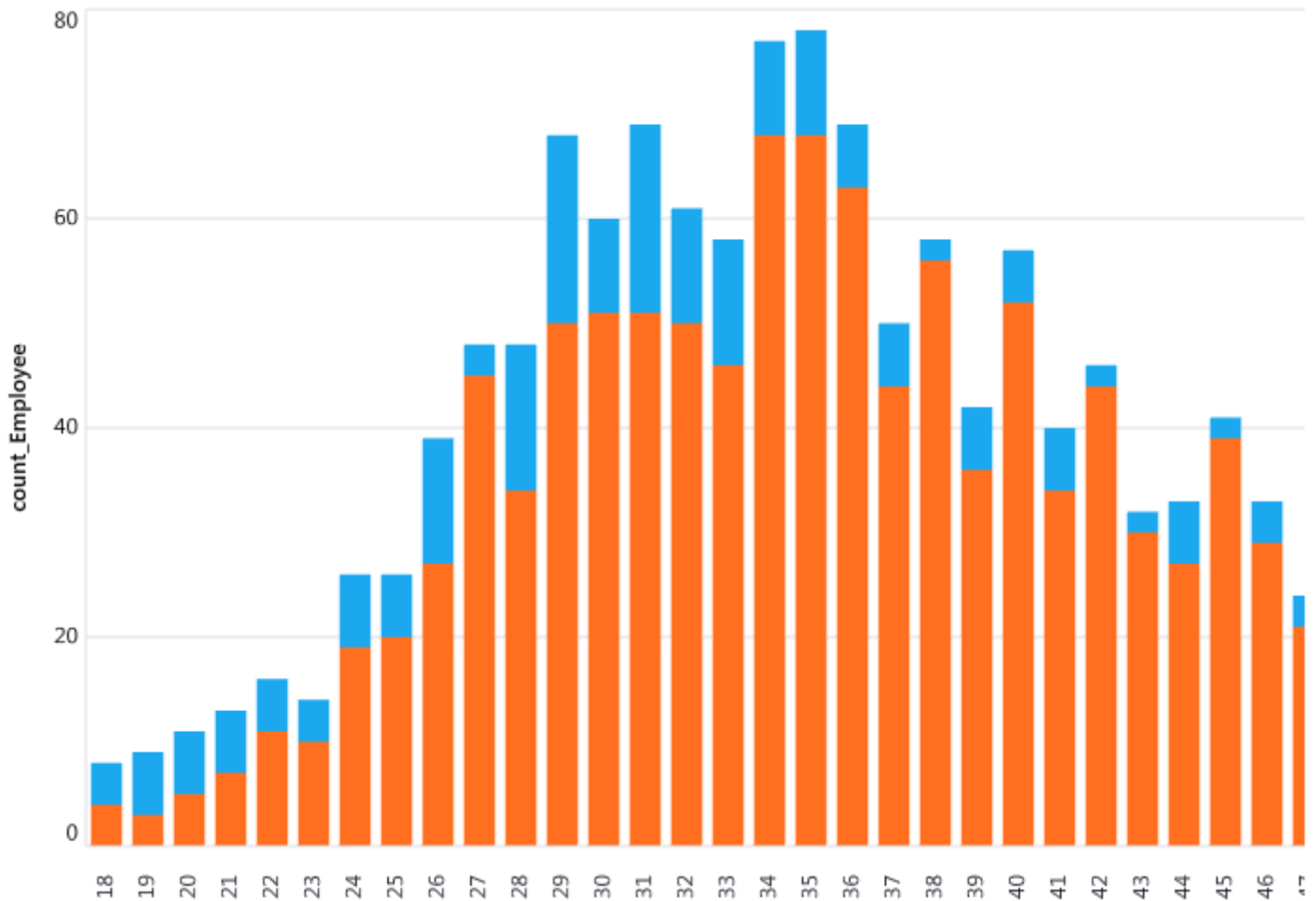
n

```

SELECT
  hr_red.Attrition AS Attrition,
  hr_red.Age AS Age,
  count(*) as count_Employee
FROM
  ibm_hr.hr_red
GROUP BY
  Attrition,
  Age

```

Уволившиеся и работающие по возрасту



В соотношении уволившихся к работающим больше всего уволившихся сотрудников в возрасте 18-21 (примерно 50%). Далее процент уволившихся уменьшается до 30% и после 33 лет резко падает.

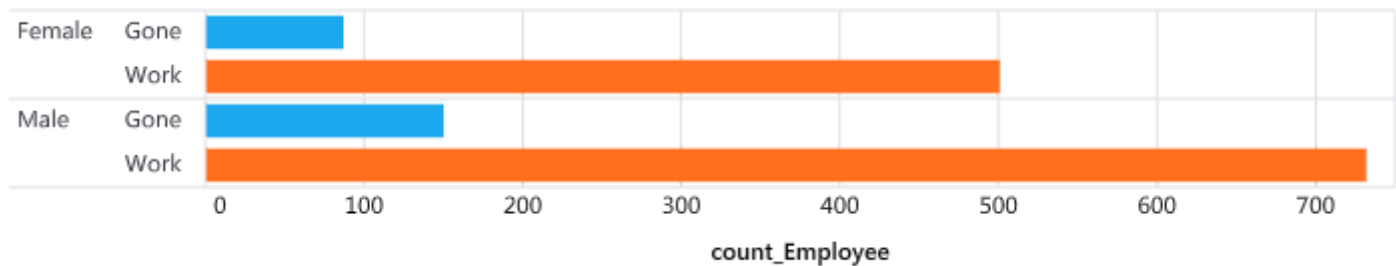
с

```

SELECT
  hr_red.Attrition AS Attrition,
  hr_red.Gender AS Gender,
  count(*) as count_Employee
FROM
  ibm_hr.hr_red
GROUP BY
  Attrition,
  Gender

```

Уволившиеся и работающие в разрезе по полу



В компании преобладают сотрудники мужского пола и среди уволившихся также преобладают мужчины.

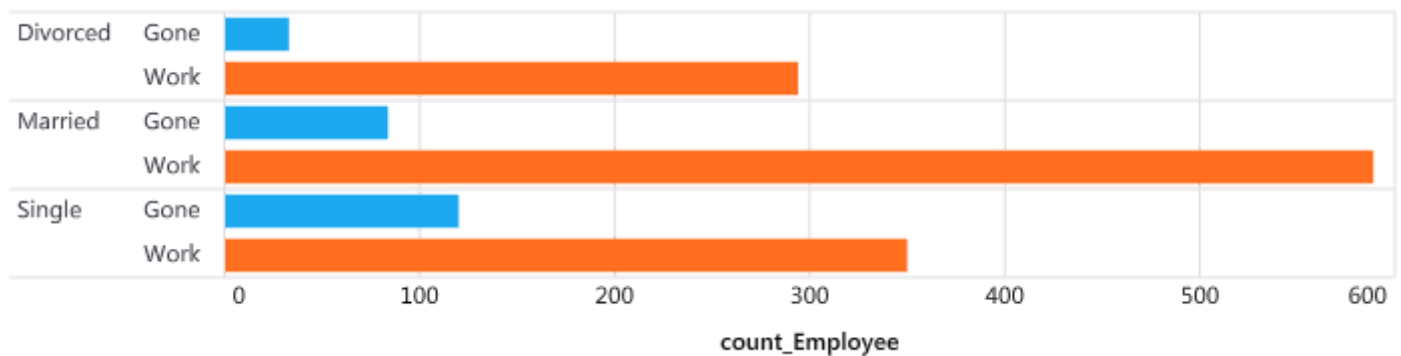
р

```

SELECT
  hr_red.Attrition AS Attrition,
  hr_red.MaritalStatus AS MaritalStatus,
  count(*) as count_Employee
FROM
  ibm_hr.hr_red
GROUP BY
  Attrition,
  MaritalStatus

```

Уволившиеся и работающие в разрезе по семейному положению

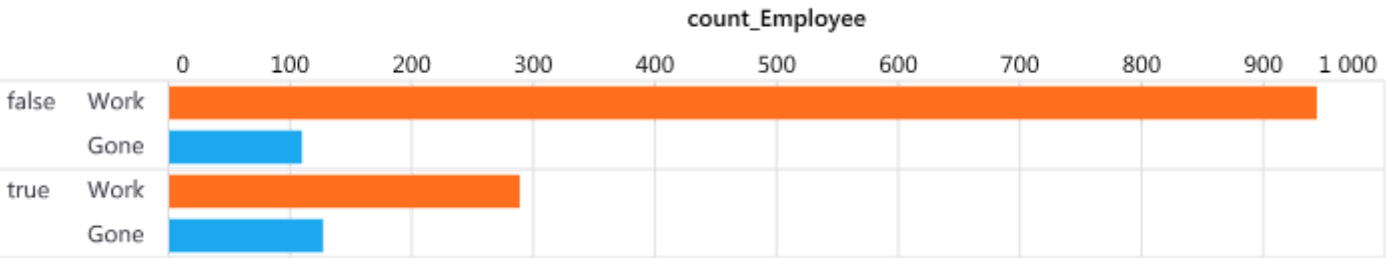


В компании больше всего работает женатых/замужних сотрудников, а среди уволившихся больше всего одиноких.

q

```
SELECT
  hr_red.Attrition AS Attrition,
  hr_red.OverTime AS OverTime,
  count(*) as count_Employee
FROM
  ibm_hr.hr_red
GROUP BY
  Attrition,
  OverTime
```

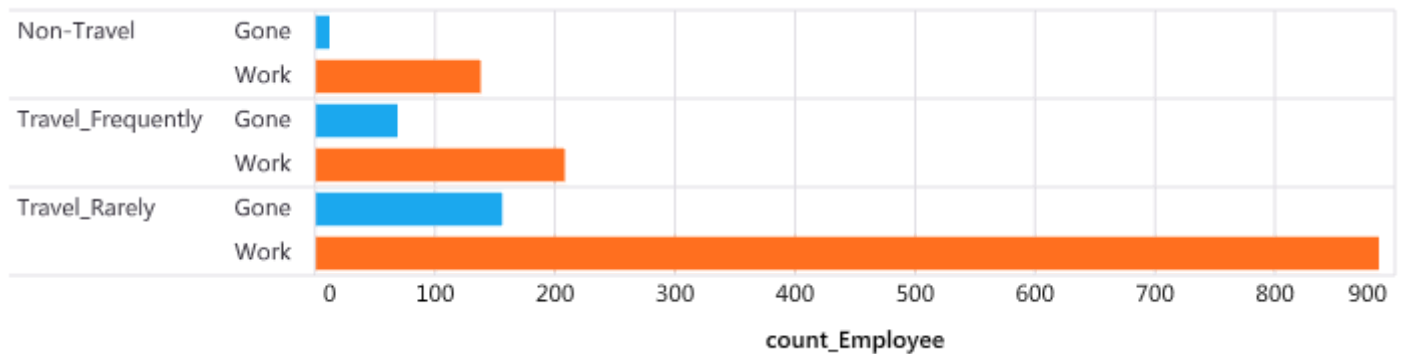
Уволившиеся и работающие в разрезе по переработкам



Если рассматривать данные по переработкам, то видно, что большинство работающих сотрудников не перерабатывает. А среди уволившихся наоборот больше тех, кто перерабатывал.

g

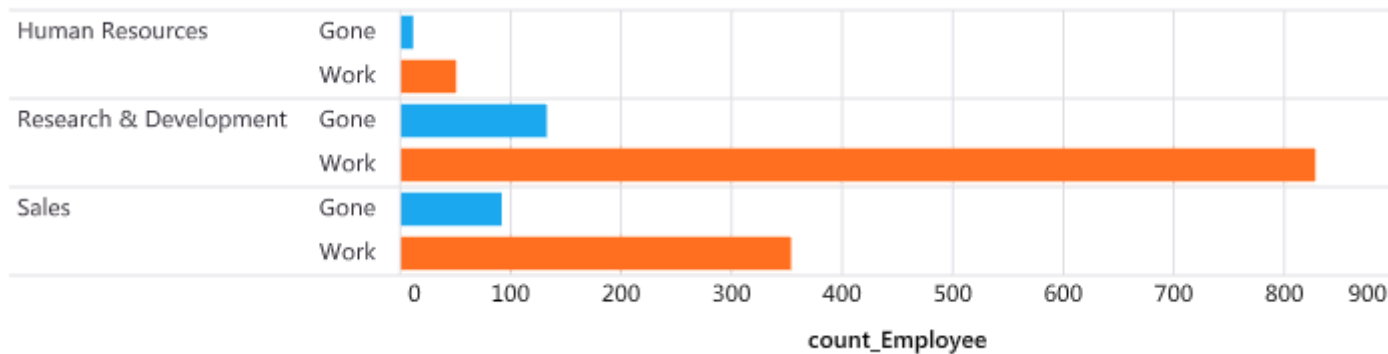
```
SELECT
  hr_red.Attrition AS Attrition,
  hr_red.BusinessTravel AS BusinessTravel,
  count(*) as count_Employee
FROM
  ibm_hr.hr_red
GROUP BY
  Attrition,
  BusinessTravel
```

Уволившиеся и работающие по частоте командировок

В разрезе командировок дополнительную закономерность не видно: больше всего уволившихся так же как и большинство работающих редко ездили в командировки.

d

```
SELECT
  hr_red.Attrition AS Attrition,
  hr_red.Department AS Department,
  count(*) as count_Employee
FROM
  ibm_hr.hr_red
GROUP BY
  Attrition,
  Department
```

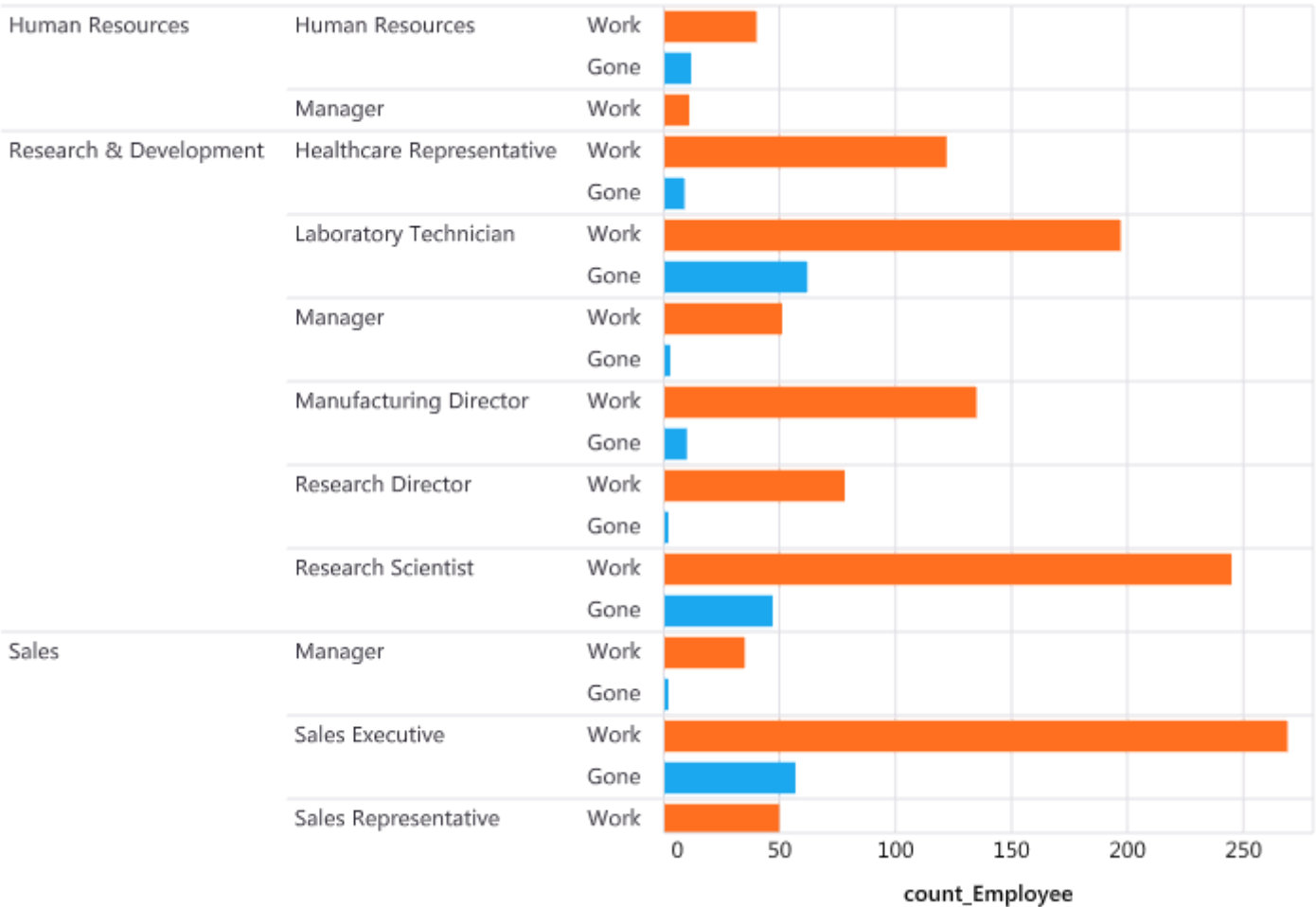
Уволившиеся и работающие по отделам

В разрезе по отделам больше всего уволившихся так же как и большинство работающих были в отделе Research&Development. Следующий по количеству работающих и уволившихся отдел Sales, на последнем месте отдел Human Resources.

t

```
SELECT
  hr_red.Attrition AS Attrition,
  hr_red.JobRole as JobRole,
  hr_red.Department AS Department,
  count(*) as count_Employee
FROM
  ibm_hr.hr_red
GROUP BY
  Attrition,
  JobRole,
  Department
```

Уволившиеся и работающие по должностям

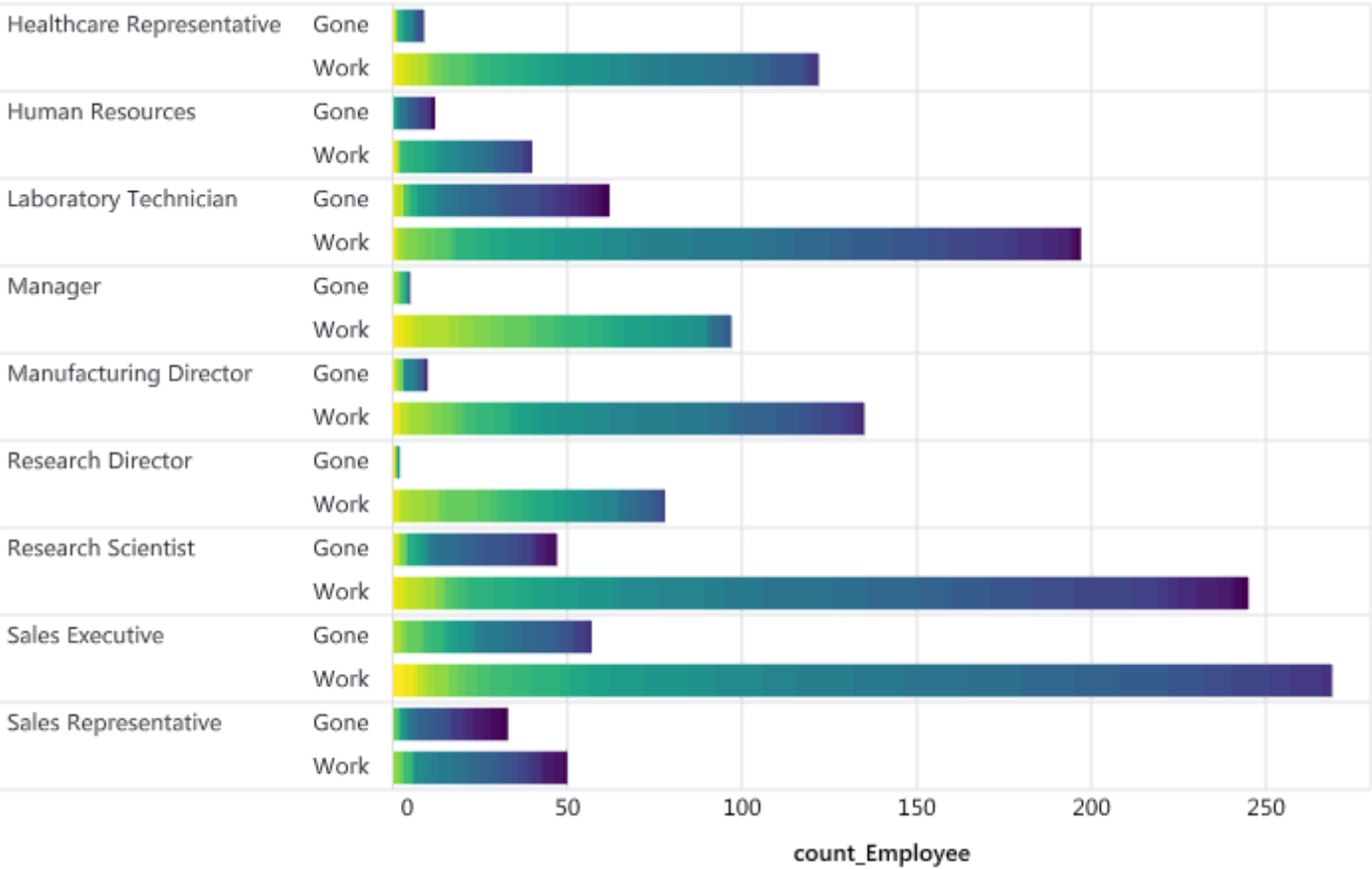


В разрезе по должностям можно увидеть, что большинство уволившихся были на должности Laboratory Technician. А большинство еще работающих сотрудников работает на должности Sales Executive.

u

```
SELECT
  hr_red.Attrition AS Attrition,
  hr_red.JobRole AS JobRole,
  hr_red.Age as Age,
  count(*) as count_Employee
FROM
  ibm_hr.hr_red
GROUP BY
  Attrition,
  JobRole, age
order by count_Employee desc
```

Уволившиеся и работающие по должностям и возрасту



Судя по визуализации данных на всех должностях увольняются более молодые сотрудники, а более возрастные продолжают работать.

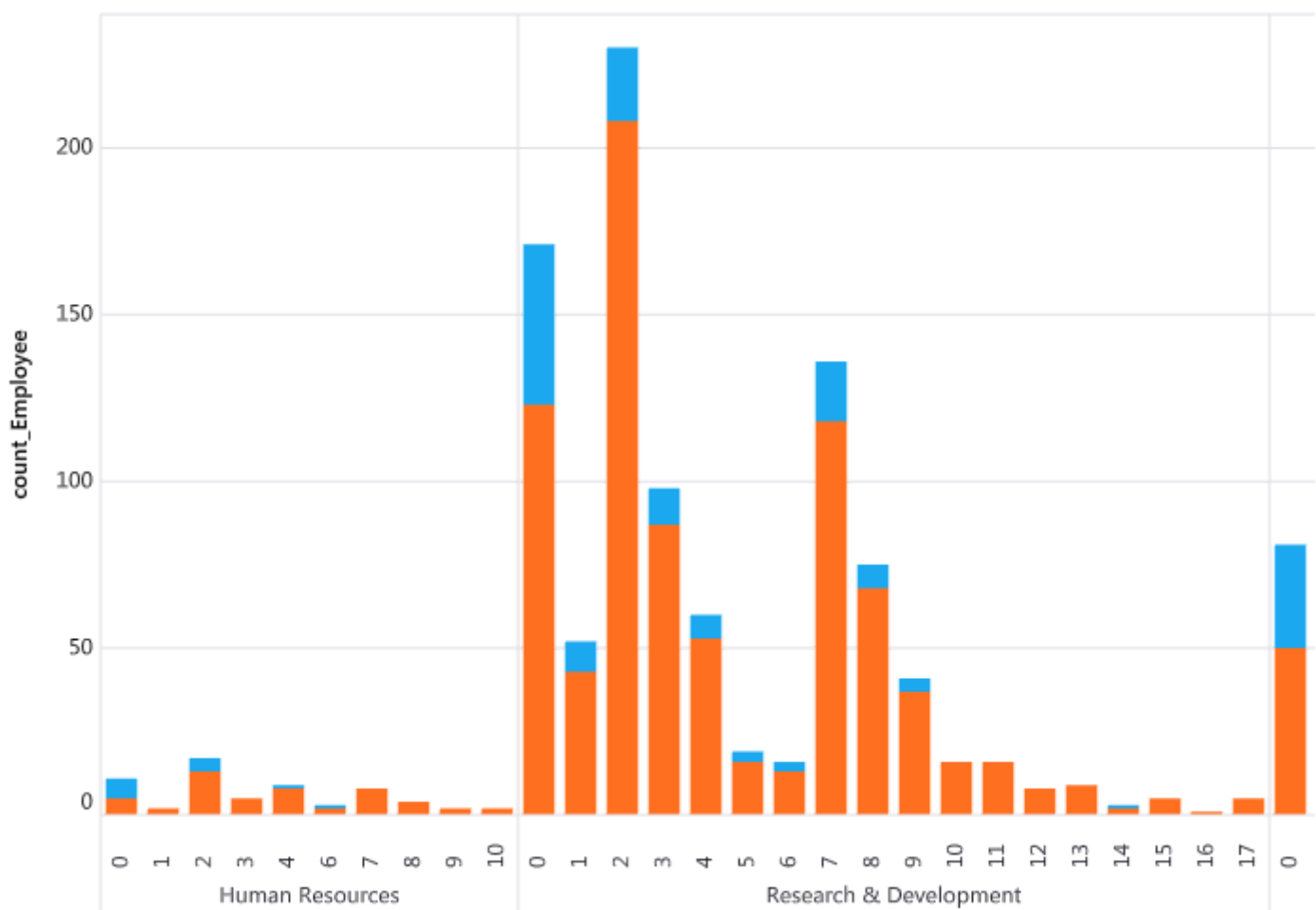
k

```

SELECT
  hr_red.Attrition AS Attrition,
  hr_red.YearsWithCurrManager AS YearsWithCurrManager,
  hr_red.Department as Department,
  count(*) as count_Employee
FROM
  ibm_hr.hr_red
GROUP BY
  Attrition,
  YearsWithCurrManager,
  Department

```

Уволившиеся и работающие по отделам и количеству лет работы с одним руководителем

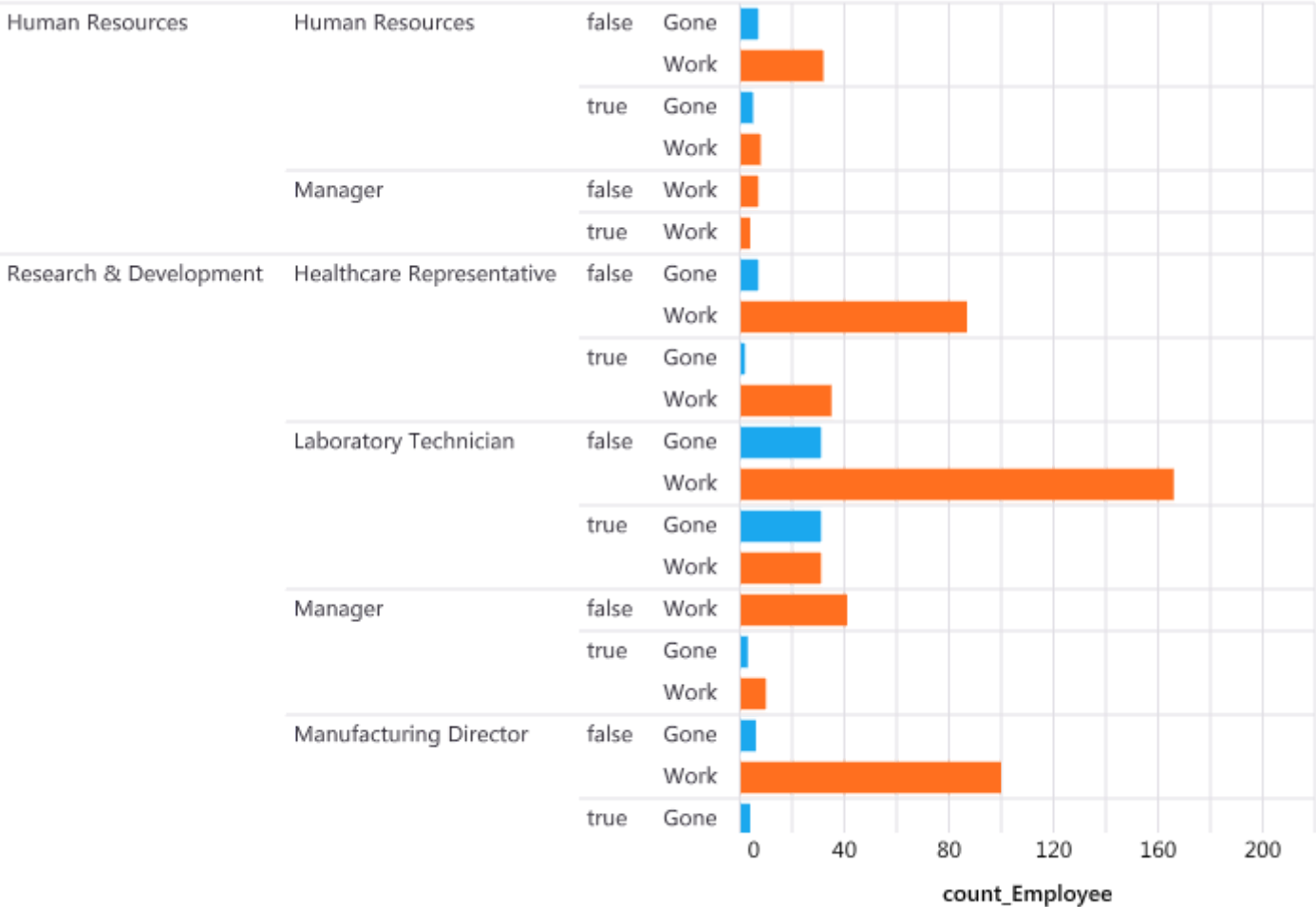


Судя по визуализации в первый год работы под началом руководителей из компании увольняются 50, 48 и 38% сотрудников в отделах Human Resources, Sales и Research&Development.

ab

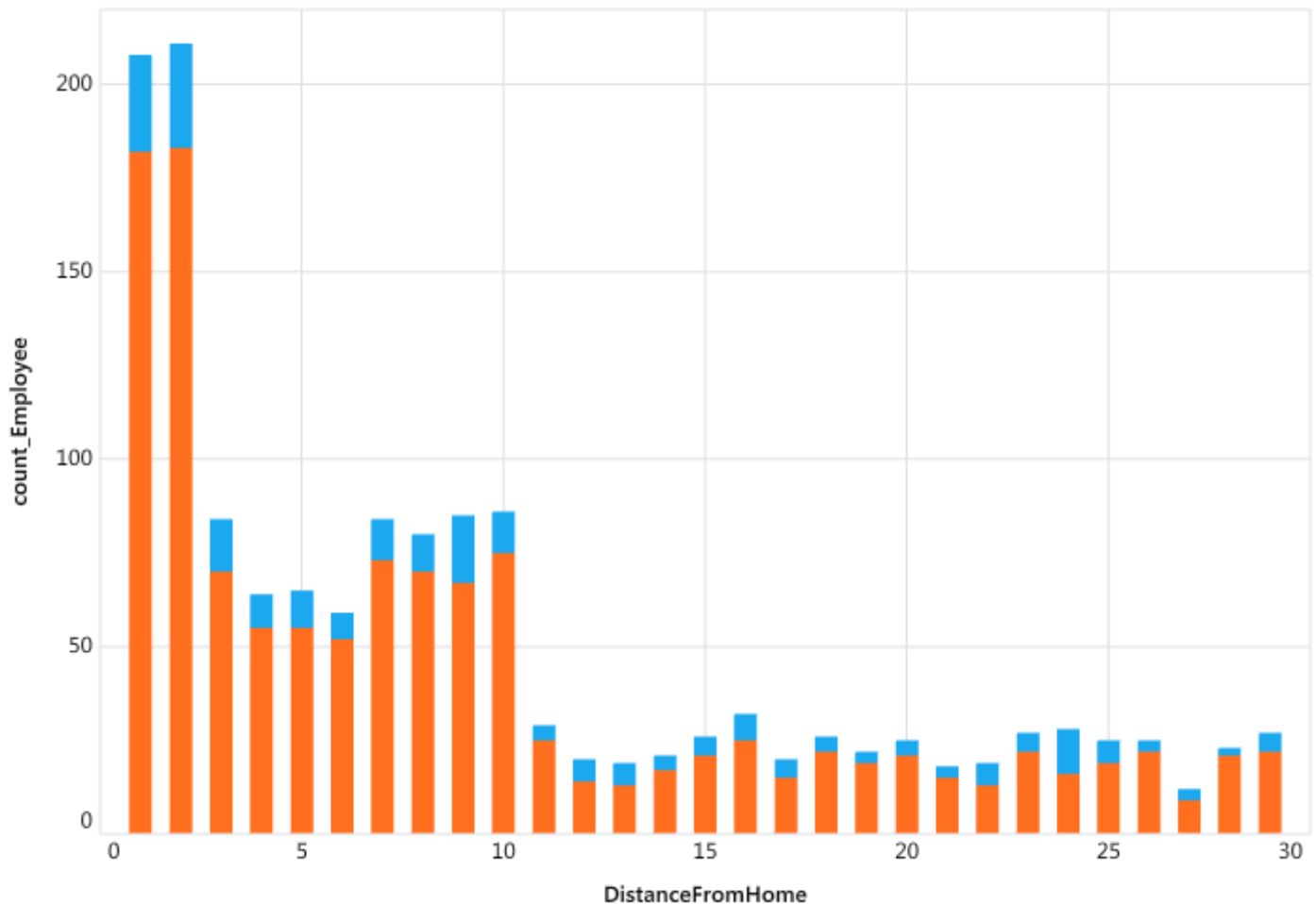
```
SELECT
  hr_red.Attrition AS Attrition,
  hr_red.JobRole as JobRole,
  hr_red.Department AS Department,
  hr_red.OverTime as OverTime,
  count(*) as count_Employee
FROM
  ibm_hr.hr_red
GROUP BY
  Attrition,
  JobRole,
  OverTime,
  Department
```

Уволившиеся и работающие по отделам, должности и переработкам



e

```
SELECT
  hr_red.Attrition AS Attrition,
  hr_red.DistanceFromHome AS DistanceFromHome,
  count(*) as count_Employee
FROM
  ibm_hr.hr_red
GROUP BY
  Attrition,
  DistanceFromHome
```

Уволившиеся и работающие по расстоянию от работы до дома

По параметру "расстояние от дома до работы" уволившиеся сотрудники распределены почти равномерно, но из-за того, что большинство работающих сотрудников живут меньше чем в 10км от работы, получается, что доля уволившихся среди тех, кто живет далеко от работы выше, чем среди тех, кто живет рядом с работой.

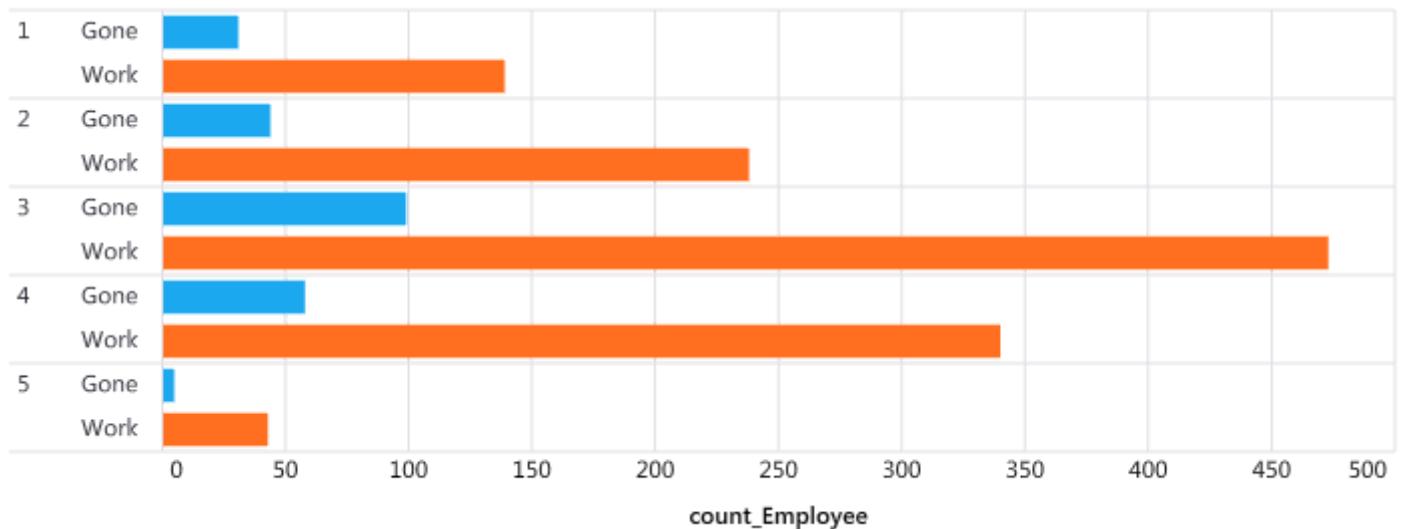
f

```

SELECT
  hr_red.Attrition AS Attrition,
  hr_red.Education AS Education,
  count(*) as count_Employee
FROM
  ibm_hr.hr_red
GROUP BY
  Attrition,
  Education

```

Уволившиеся и работающие в разрезе по образованию



Образование

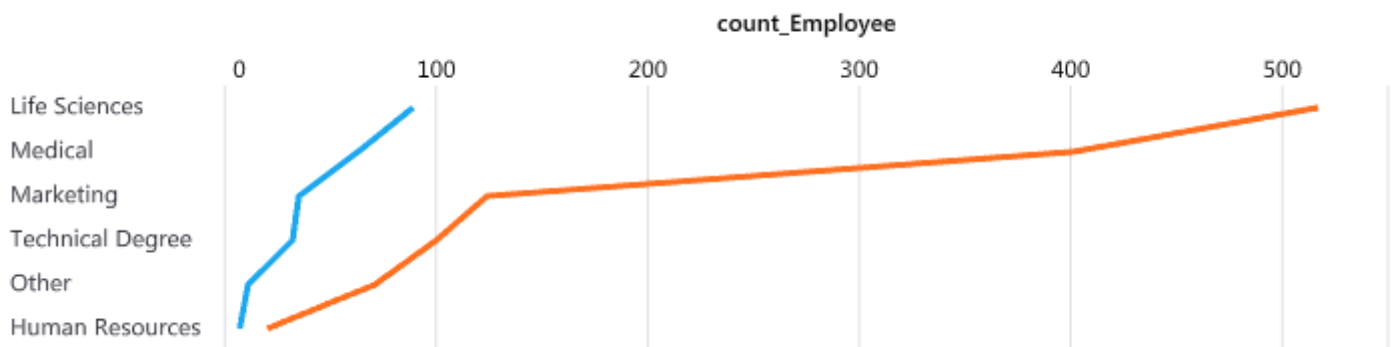
- 1 «Ниже колледжа»
- 2 «Колледж»
- 3 «Бакалавр»
- 4 «Магистр»
- 5 «Доктор»

В разрезе по образованию больше всего уволившихся так же как и большинство работающих имеют степень Бакалавра. Следующий по количеству работающих и уволившихся степень Магистра. В целом по уровню образования количество уволившихся пропорционально количеству работающих сотрудников.

h

```
SELECT
  hr_red.Attrition AS Attrition,
  hr_red.EducationField AS EducationField,
  count(*) as count_Employee
FROM
  ibm_hr.hr_red
GROUP BY
  Attrition,
  EducationField
```

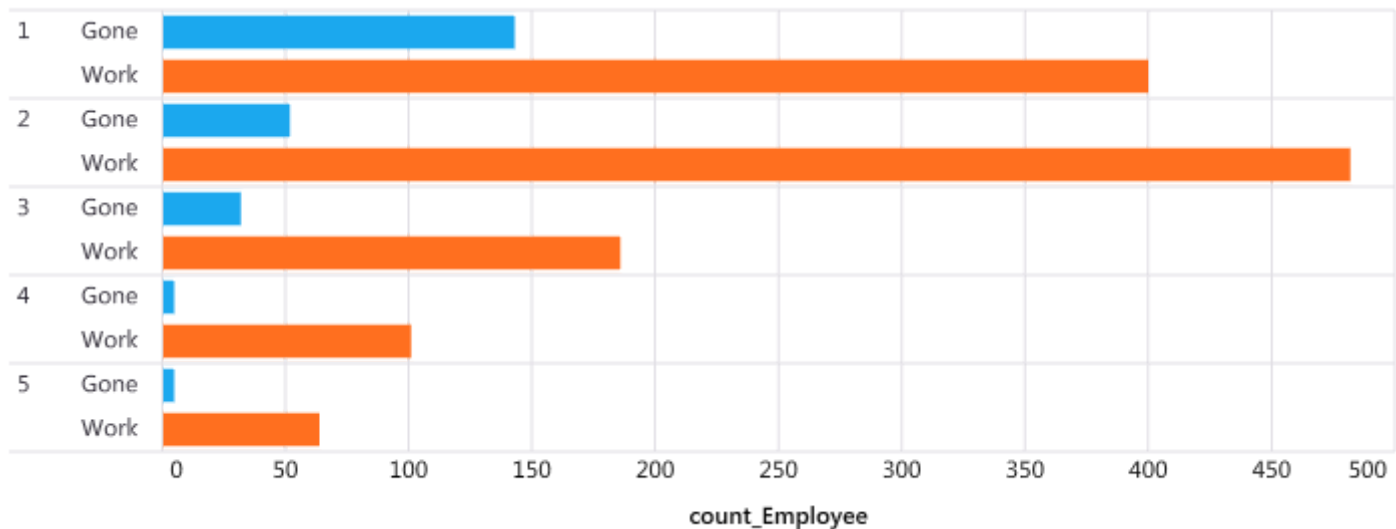
Уволившиеся и работающие в разрезе по сфере обучения



Распределение уволившихся и работающих по сферам обучения прямо пропорционально: чем больше сотрудников относится к каждой сфере, тем больше уволившихся сотрудников из данной сферы.

0

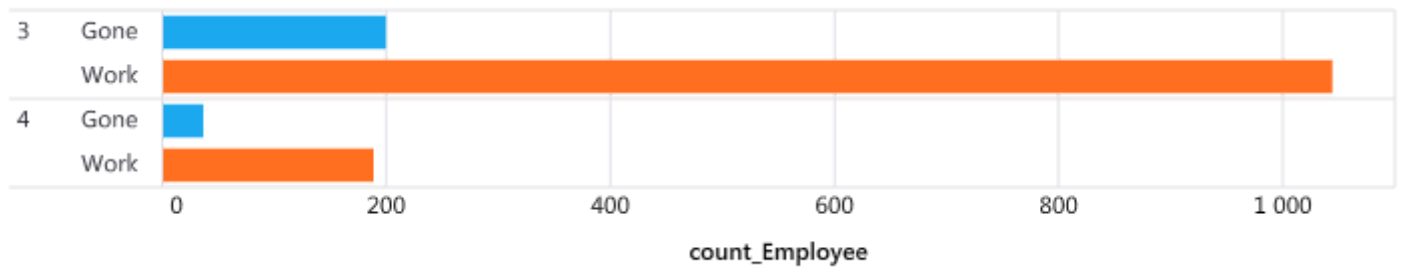
```
SELECT
  hr_red.Attrition AS Attrition,
  hr_red.JobLevel AS JobLevel,
  count(*) as count_Employee
FROM
  ibm_hr.hr_red
GROUP BY
  Attrition,
  JobLevel
```

Уволившиеся и работающие в разрезе по уровню работы

Среди уволившихся сотрудников больше всего тех, кто относится к 1-ому уровню работы. Поскольку нет расшифровки что бы это могло значить, можно предположить, что сотрудники плохо адаптируются и им не удается продвигаться дальше по карьерной лестнице. На втором уровне работы работает большинство сотрудников.

j

```
SELECT
  hr_red.Attrition AS Attrition,
  hr_red.PerformanceRating AS PerformanceRating,
  count(*) as count_Employee
FROM
  ibm_hr.hr_red
GROUP BY
  Attrition,
  PerformanceRating
```

Уволившиеся и работающие в разрезе по оценке**Оценка**

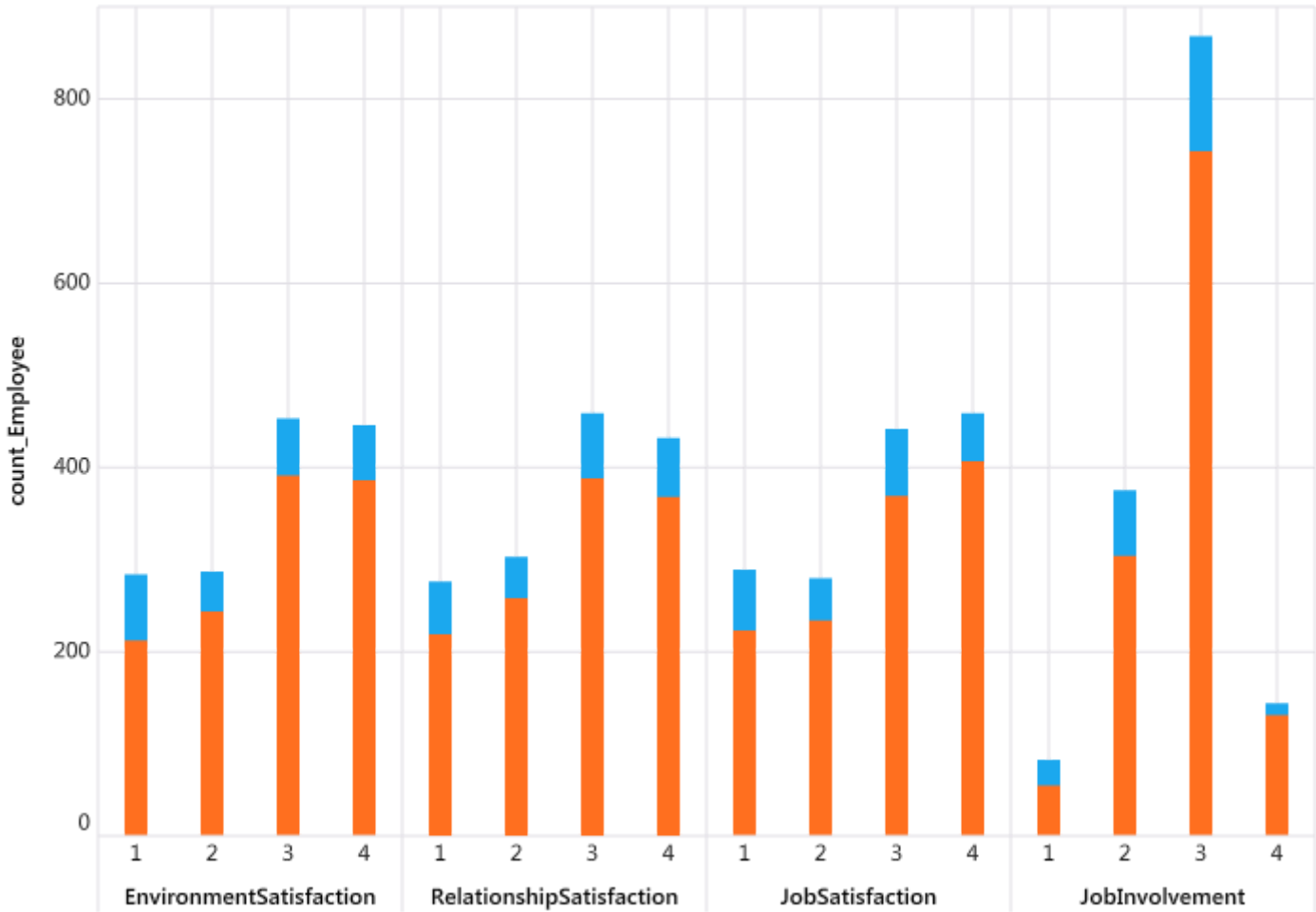
- 1 «Низкая»
- 2 «Хорошо»
- 3 «Отлично»
- 4 «Великолепно»

По этим данным видно, что все сотрудники справляются с работой на уровне "Отлично" и "Великолепно". При этом тех, кто работает отлично и тех, кто уволился с оценкой за работу "Отлично" - большинство.

i

```
SELECT
  hr_red.Attrition AS Attrition,
  hr_red.EnvironmentSatisfaction AS EnvironmentSatisfaction,
  hr_red.JobInvolvement as JobInvolvement,
  hr_red.JobSatisfaction as JobSatisfaction,
  hr_red.RelationshipSatisfaction as RelationshipSatisfaction,
  count(EnvironmentSatisfaction) as count_EnvironmentSatisfaction,
  count(*) as count_Employee
FROM
  ibm_hr.hr_red
GROUP BY
  Attrition,
  EnvironmentSatisfaction,
  JobInvolvement,
  JobSatisfaction,
  RelationshipSatisfaction
```

Уволившиеся и работающие в разрезах по уровню удовлетворения условиями работы, отношениями, работой и уровню вовлеченности в работу



Удовлетворение от условий работы

- 1 «Низкая»
- 2 «Средняя»
- 3 «Высокая»
- 4 «Очень высокая»

Вовлеченность в работу

- 1 «Низкая»
- 2 «Средняя»
- 3 «Высокая»
- 4 «Очень высокая»

Удовлетворение от работы

- 1 «Низкая»
- 2 «Средняя»
- 3 «Высокая»
- 4 «Очень высокая»

Удовлетворение от отношений

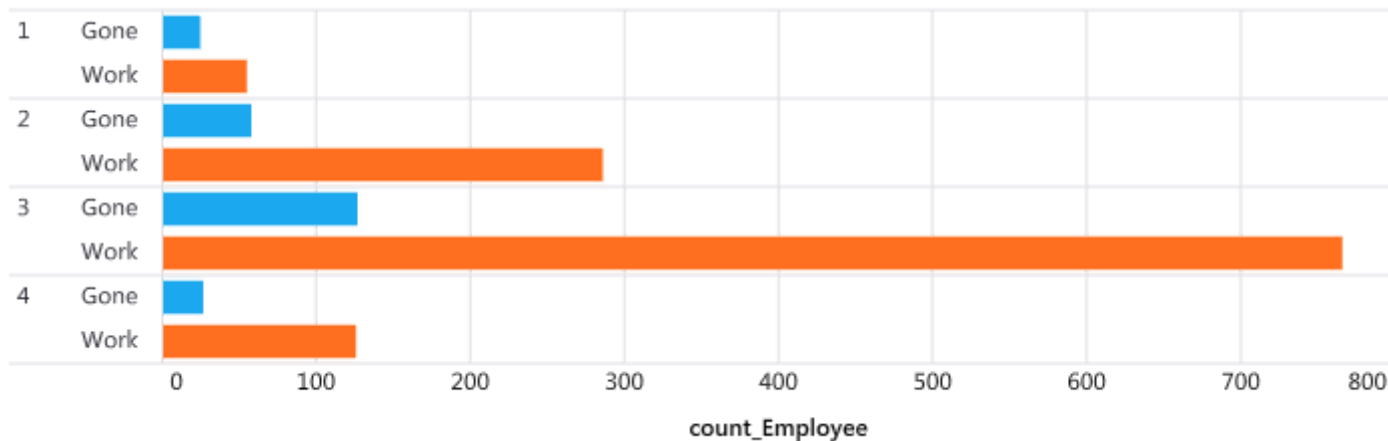
- 1 «Низкая»
- 2 «Средняя»
- 3 «Высокая»
- 4 «Очень высокая»

Судя по данным удовлетворение от условий работы, отношениями на работе, самой работой и вовлеченность в работу у уволившихся и работающих сотрудников одинаковые.

S

```
SELECT
  hr_red.Attrition AS Attrition,
  hr_red.WorkLifeBalance AS WorkLifeBalance,
  count(*) as count_Employee
FROM
  ibm_hr.hr_red
GROUP BY
  Attrition,
  WorkLifeBalance
```

Уволившиеся и работающие по уровню баланса жизни и работы



WorkLifeBalance

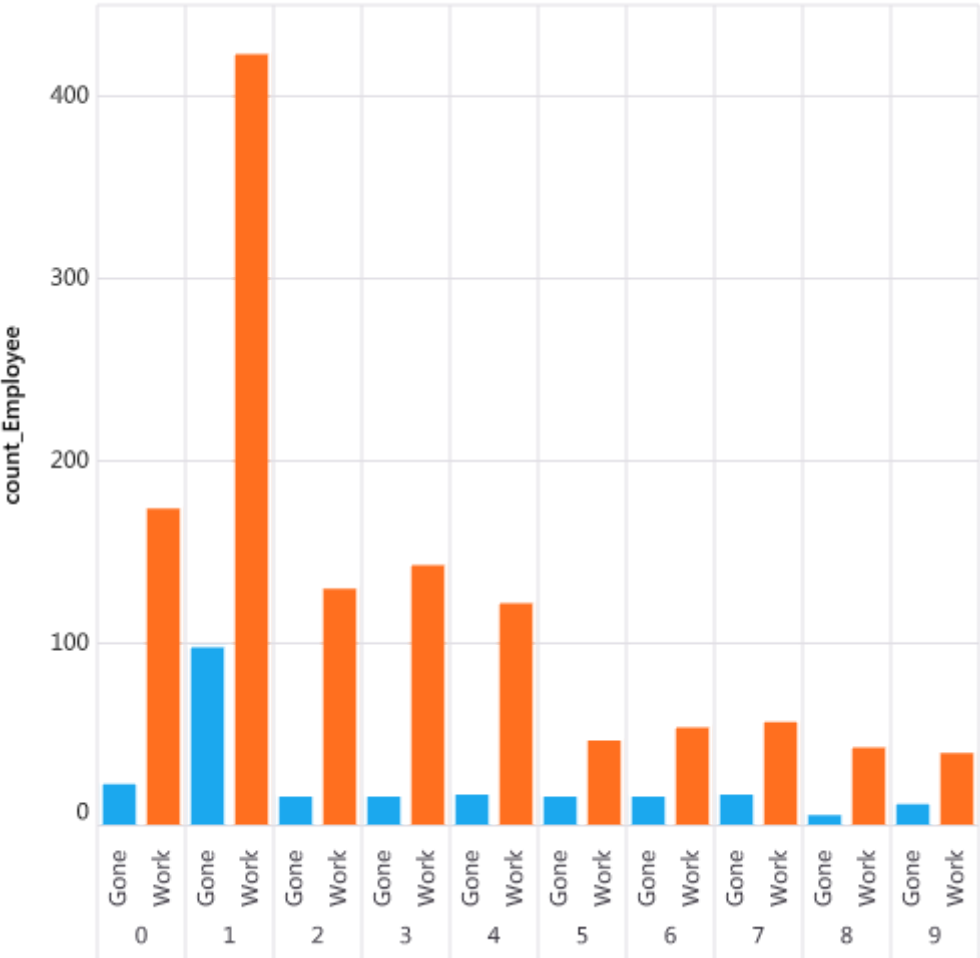
- 1 «Плохо»
- 2 «Хорошо»
- 3 «Отлично»
- 4 «Лучше всех»

По уровням баланса жизни и работы соотношение уволившихся и работающих прямо пропорционально. Большинство работающих и уволившихся сотрудников оценивают свой баланс как "Отлично".

v

```
SELECT
  hr_red.Attrition AS Attrition,
  hr_red.NumCompaniesWorked AS NumCompaniesWorked,
  count(*) as count_Employee
FROM
  ibm_hr.hr_red
GROUP BY
  Attrition,
  NumCompaniesWorked
```

Уволившиеся и работающие по количеству компаний, в которых они работали



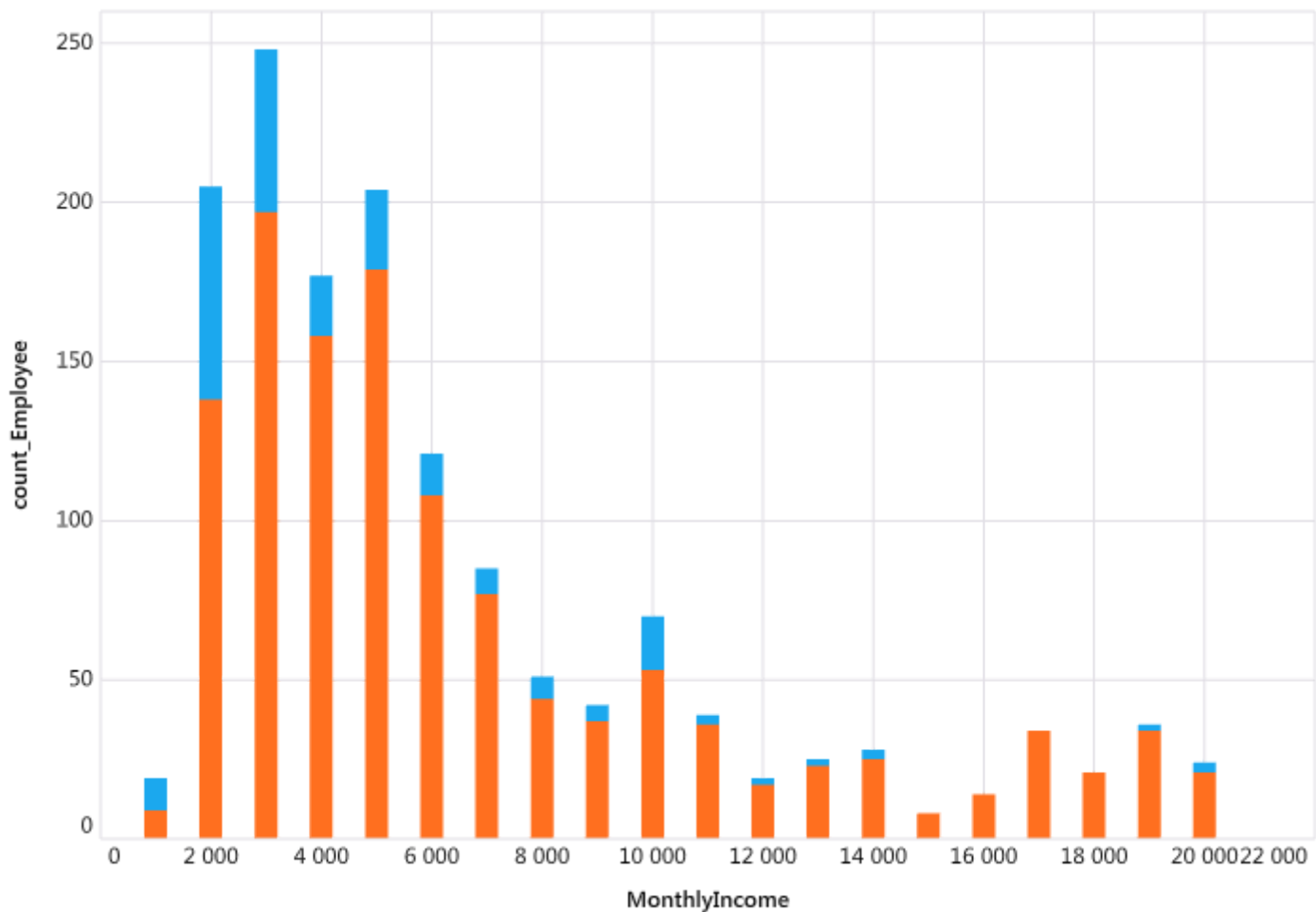
Видно, что для большинства сотрудников (как уволившихся, так и продолжающих работать), данная компания вторая (до этого момента они работали в одной компании).

y

```
SELECT
  hr_red.Attrition as Attrition,
  round (hr_red.MonthlyIncome,-3) AS MonthlyIncome,
  count(*) as count_Employee
```

```
FROM
  ibm_hr.hr_red
group by MonthlyIncome, Attrition
```

Уволившиеся и работающие по уровню зарплаты



Доля уволившихся сотрудников по отношению к работающим с зарплатой в 1000 и 2000 составляет 53% и 33% соответственно, далее доля уволившихся сотрудников уменьшается, а доля работающих растет.

aa

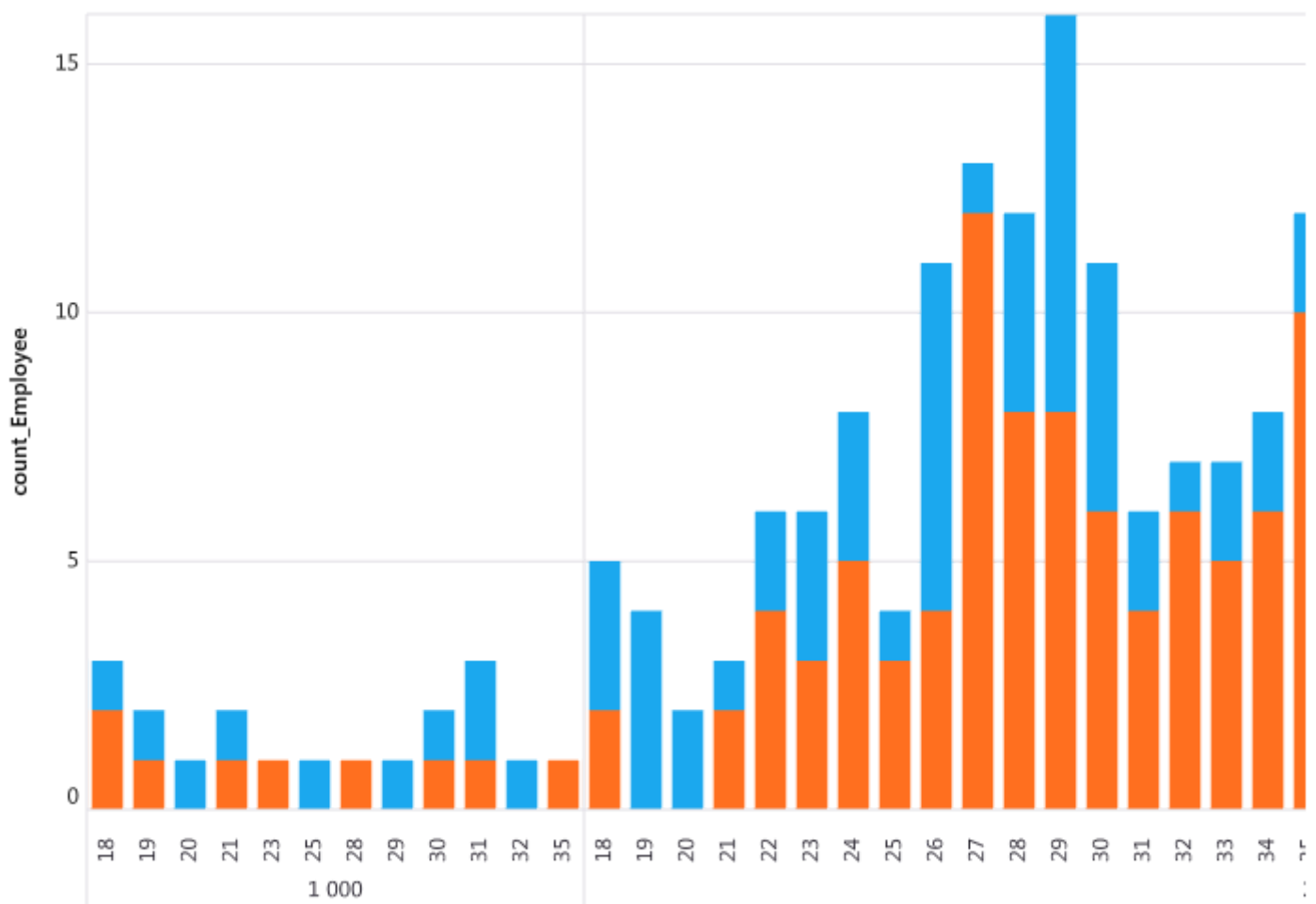
```

SELECT
  hr_red.Attrition AS Attrition,
  hr_red.Age AS Age,
  round(hr_red.MonthlyIncome,-3) as MonthlyIncome,
  count(*) as count_Employee
FROM
  ibm_hr.hr_red

GROUP BY
  Attrition,
  Age, MonthlyIncome

```

Уволившиеся и работающие по уровню зарплаты и возрасту



Данная визуализация подтверждает закономерности, выявленные выше: чаще всего увольняются молодые сотрудники (до 33 лет) и сотрудники с невысокой зарплатой (до 3000)

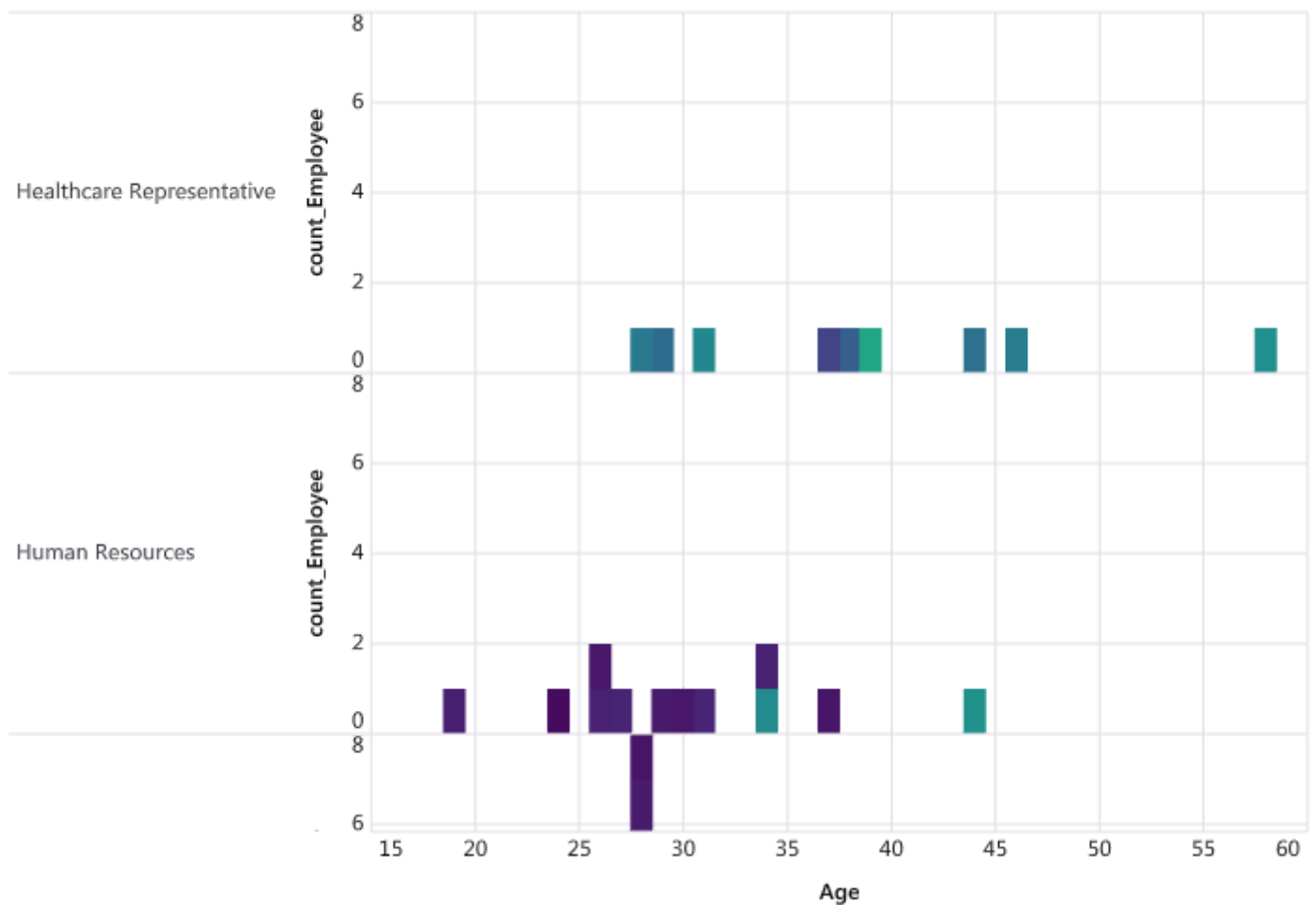
x

```

SELECT
  hr_red.Attrition AS Attrition,
  hr_red.Age AS Age,
  hr_red.JobRole as JobRole,
  hr_red.MonthlyIncome as MonthlyIncome,
  count(*) as count_Employee
FROM
  ibm_hr.hr_red
where
  (Attrition = 'Gone')
GROUP BY
  Attrition,
  Age, JobRole, MonthlyIncome

```

Уволившиеся по возрасту, уровню зарплаты и должности



Данная визуализация подтверждает закономерности, выявленные выше: чаще увольняются сотрудники с должностью Laboratory Technician и Research Scientist и сотрудники с зарплатой до 3000.

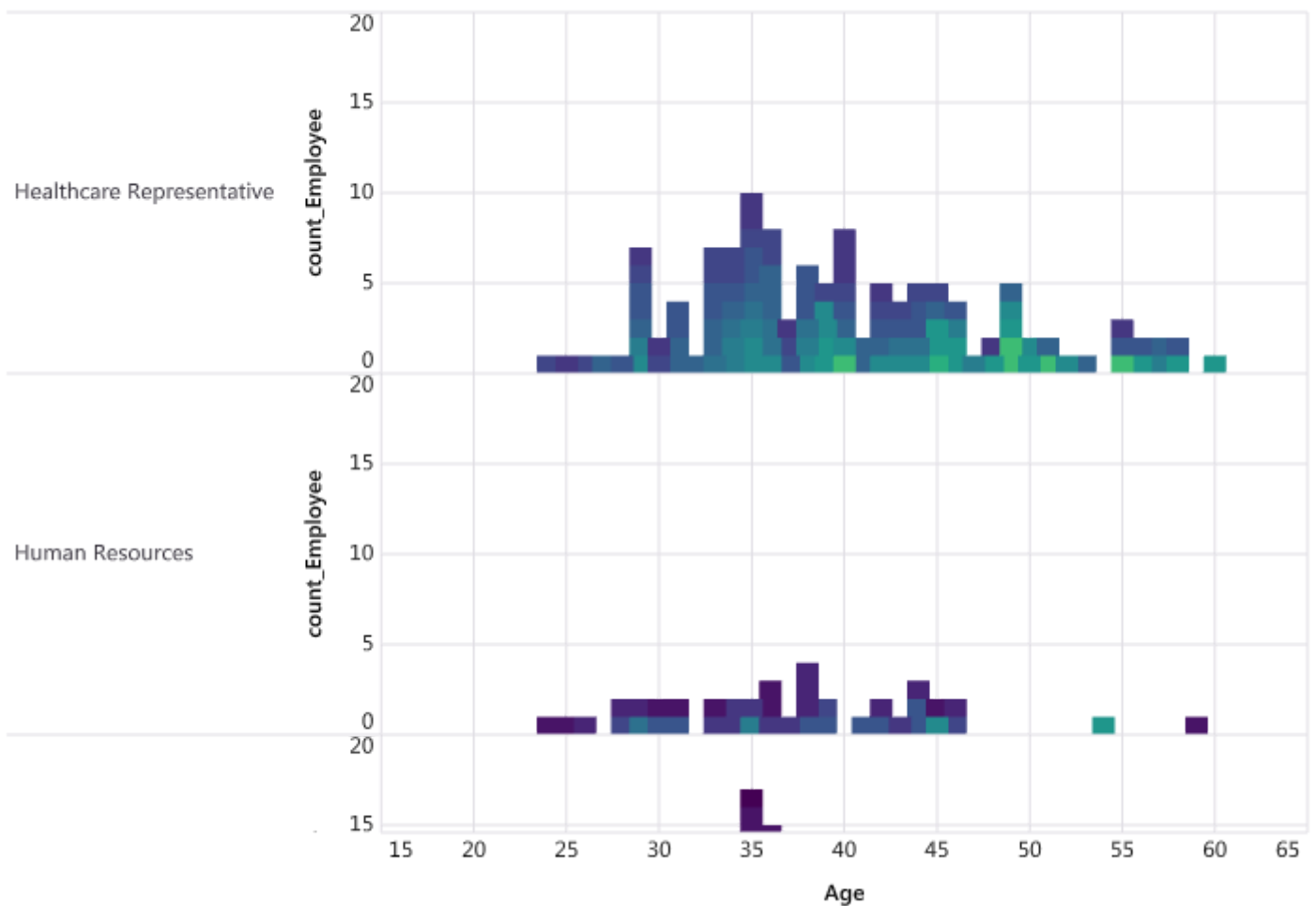
z

```

SELECT
  hr_red.Attrition AS Attrition,
  hr_red.Age AS Age,
  hr_red.JobRole as JobRole,
  round(hr_red.MonthlyIncome,-3) as MonthlyIncome,
  count(*) as count_Employee
FROM
  ibm_hr.hr_red
where
  (Attrition = 'Work')
GROUP BY
  Attrition,
  Age, JobRole, MonthlyIncome

```

Работающие по возрасту, уровню зарплаты и должности



ac

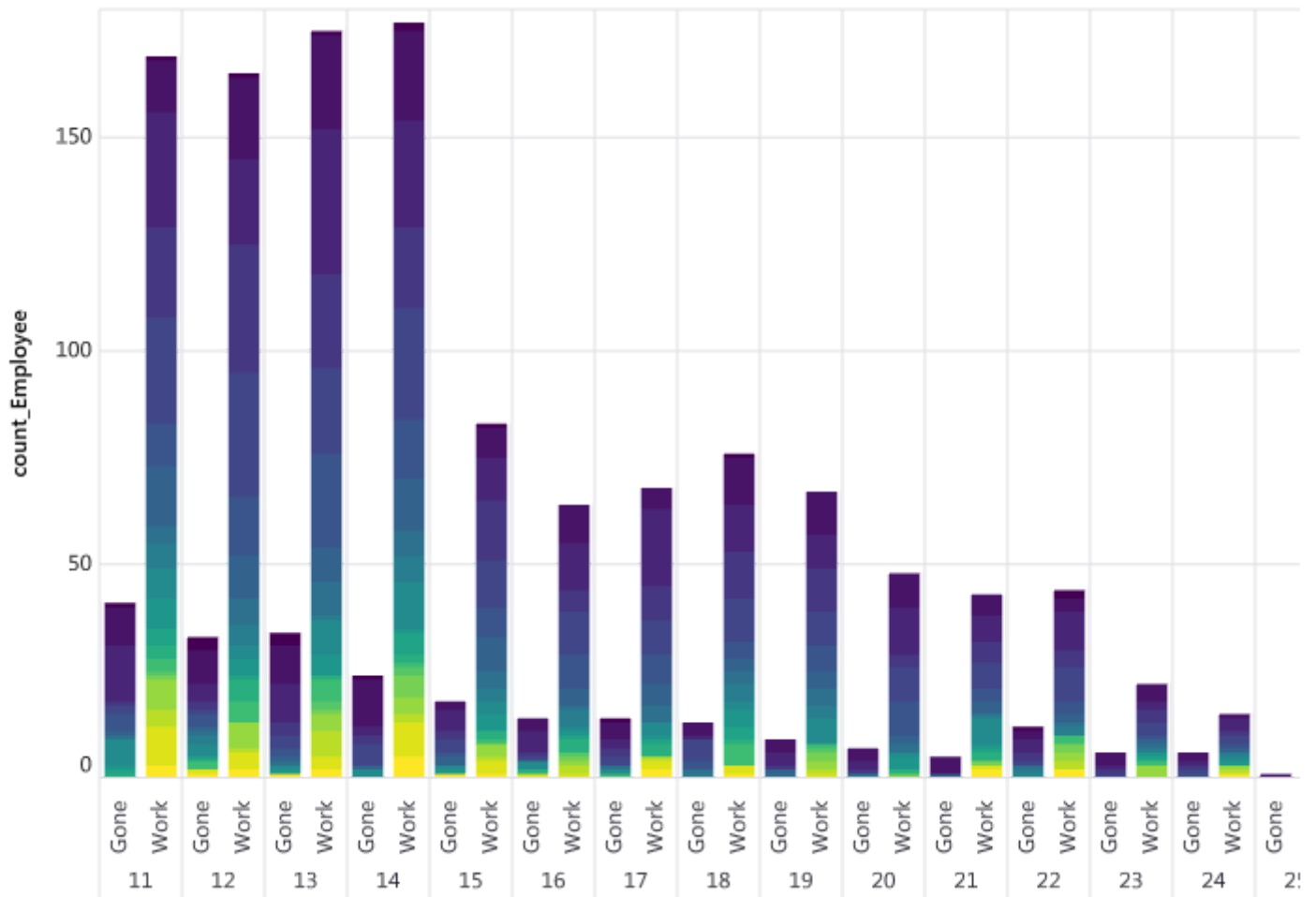
SELECT

```
hr_red.Attrition as Attrition,
hr_red.PercentSalaryHike AS PercentSalaryHike,
round (hr_red.MonthlyIncome,-3) AS MonthlyIncome,
count(*) as count_Employee
```

FROM

```
ibm_hr.hr_red
group by PercentSalaryHike, MonthlyIncome, Attrition
```

Уволившиеся и работающие по уровню зарплаты и проценту повышения зарплаты



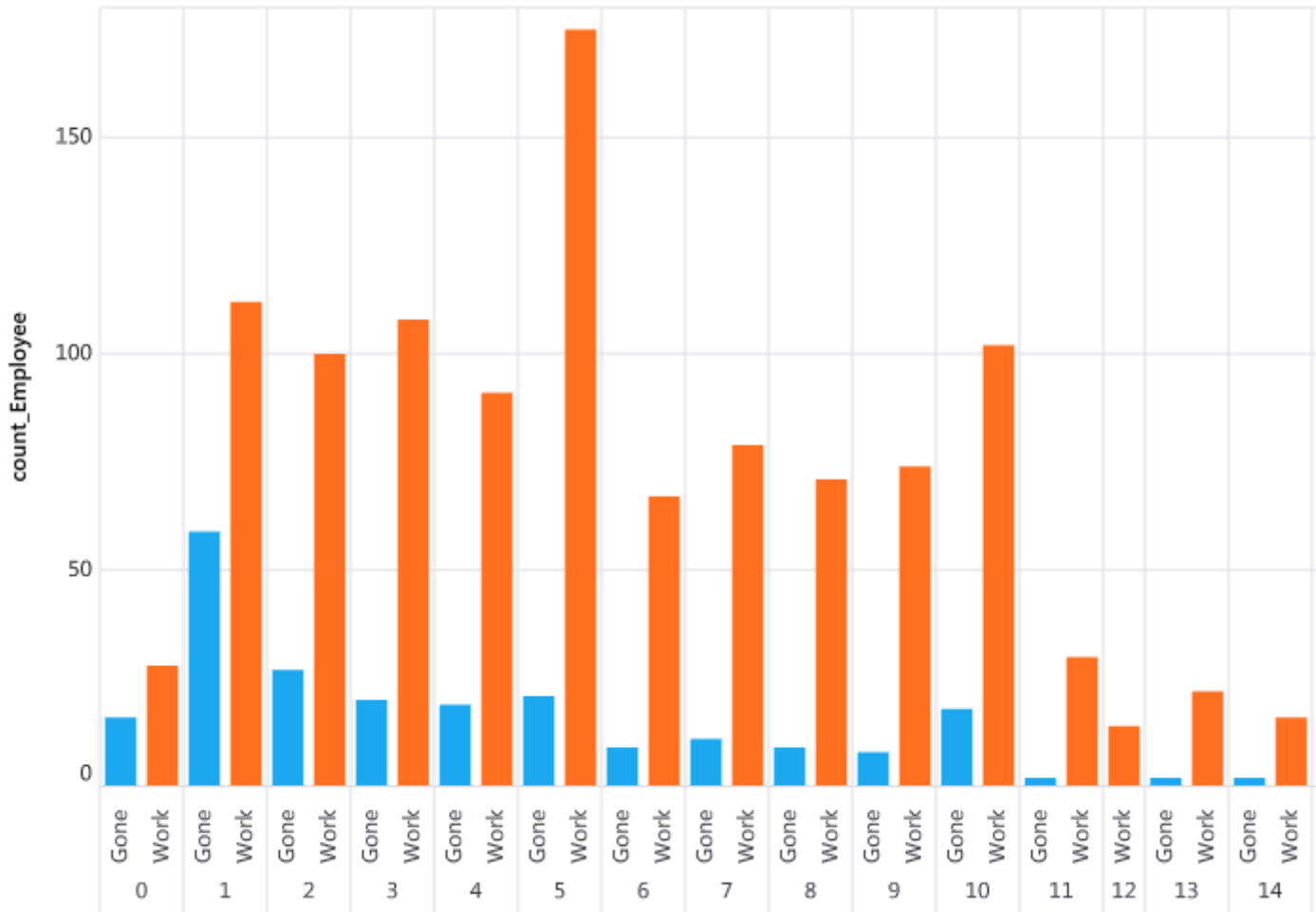
Здесь видно, что есть закономерность между зарплатой и тем процентом, на который повышают зарплату: чем выше зарплата, тем меньше процент повышения зарплаты. Уволившимся сотрудникам повышали зарплаты аналогично тем сотрудникам, которые остались работать.

W

```
SELECT
  hr_red.Attrition as Attrition,
  hr_red.YearsAtCompany as YearsAtCompany,
  count(*) as count_Employee

FROM
  ibm_hr.hr_red
group by YearsAtCompany, Attrition
```

Уволившиеся и работающие в разрезе по годам работы в компании



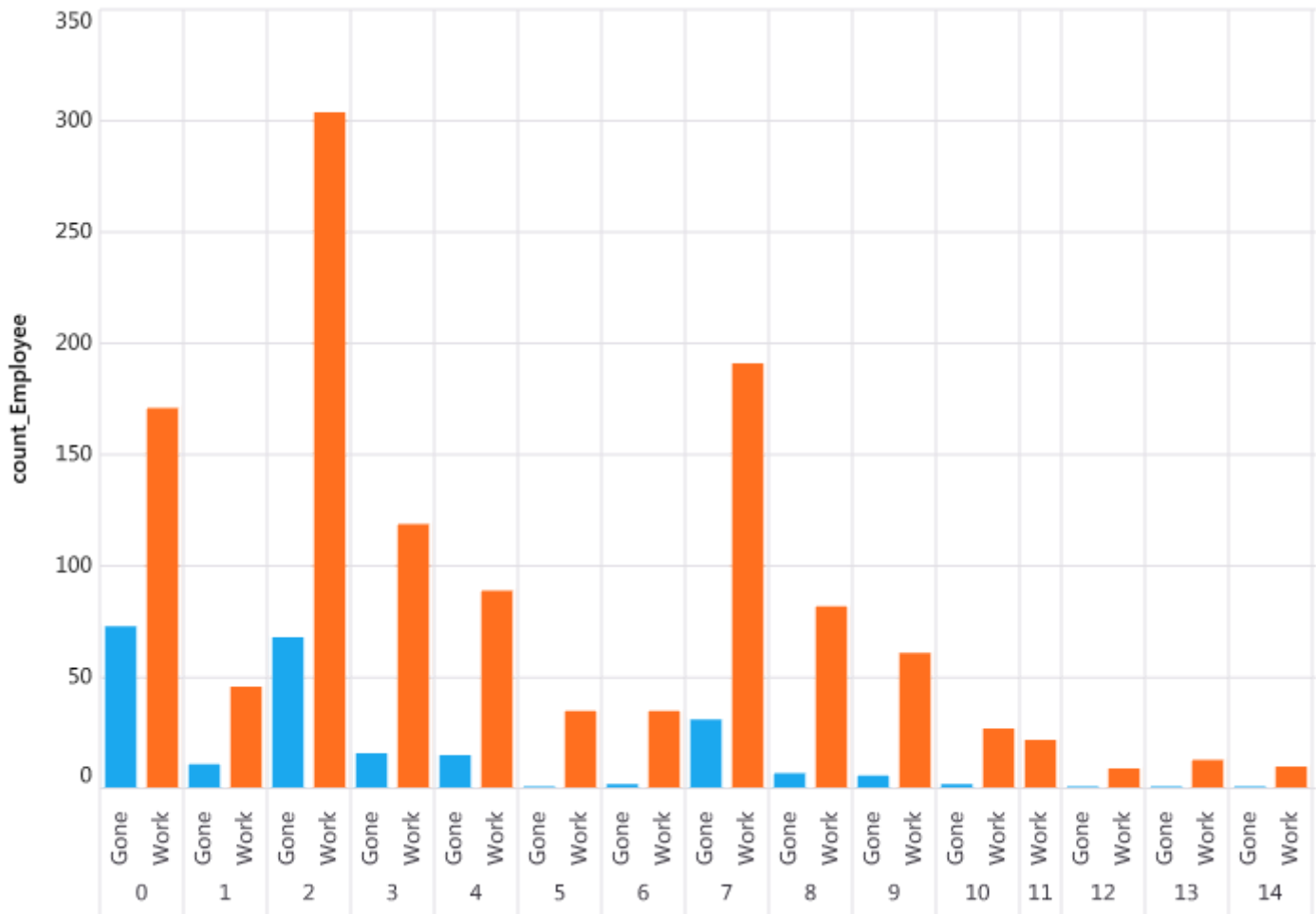
Видно, что большая доля сотрудников (по отношению к работающим) увольняется в первый и второй год работы в компании (36% и 35%).

1

```
SELECT
  hr_red.Attrition as Attrition,
  hr_red.YearsInCurrentRole AS YearsInCurrentRole,
  count(*) as count_Employee

FROM
  ibm_hr.hr_red
group by  YearsInCurrentRole, Attrition
```

Уволившиеся и работающие в разрезе по годам работы в данной должности



На данной визуализации видно, что чаще всего сотрудники увольняются в первый год работы в должности и в третий.

ad

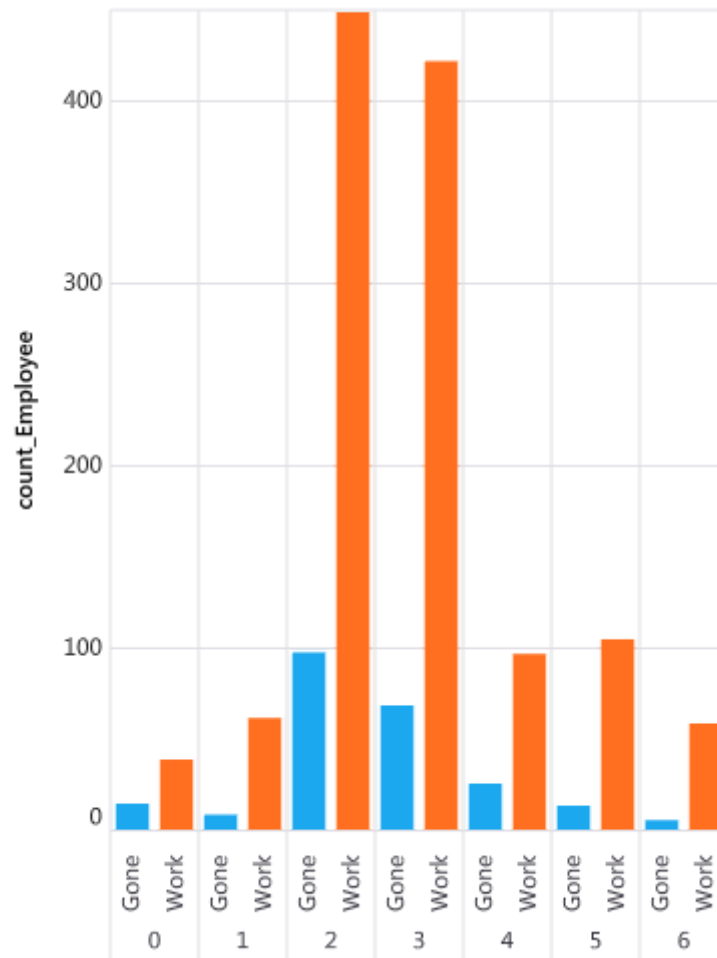
```

SELECT
count(*) as count_Employee,
  hr_red.TrainingTimesLastYear AS TrainingTimesLastYear,
hr_red.Attrition as Attrition

FROM
  ibm_hr.hr_red
group by TrainingTimesLastYear, Attrition

```

Уволившиеся и работающие по количеству обучения в прошлом году



Больше всего уволившихся среди тех, кто проходил 2 и 3 обучения в прошлом году. Это можно объяснить тем, что в компании больше всего людей прошли 2 и 3 обучения в прошлом году.

Выводы:

В данных видно взаимосвязь увольнения сотрудников со следующими факторами: возраст, зарплата, должность, переработки, уровень работы, сколько лет проработал в компании.

Гипотезы:

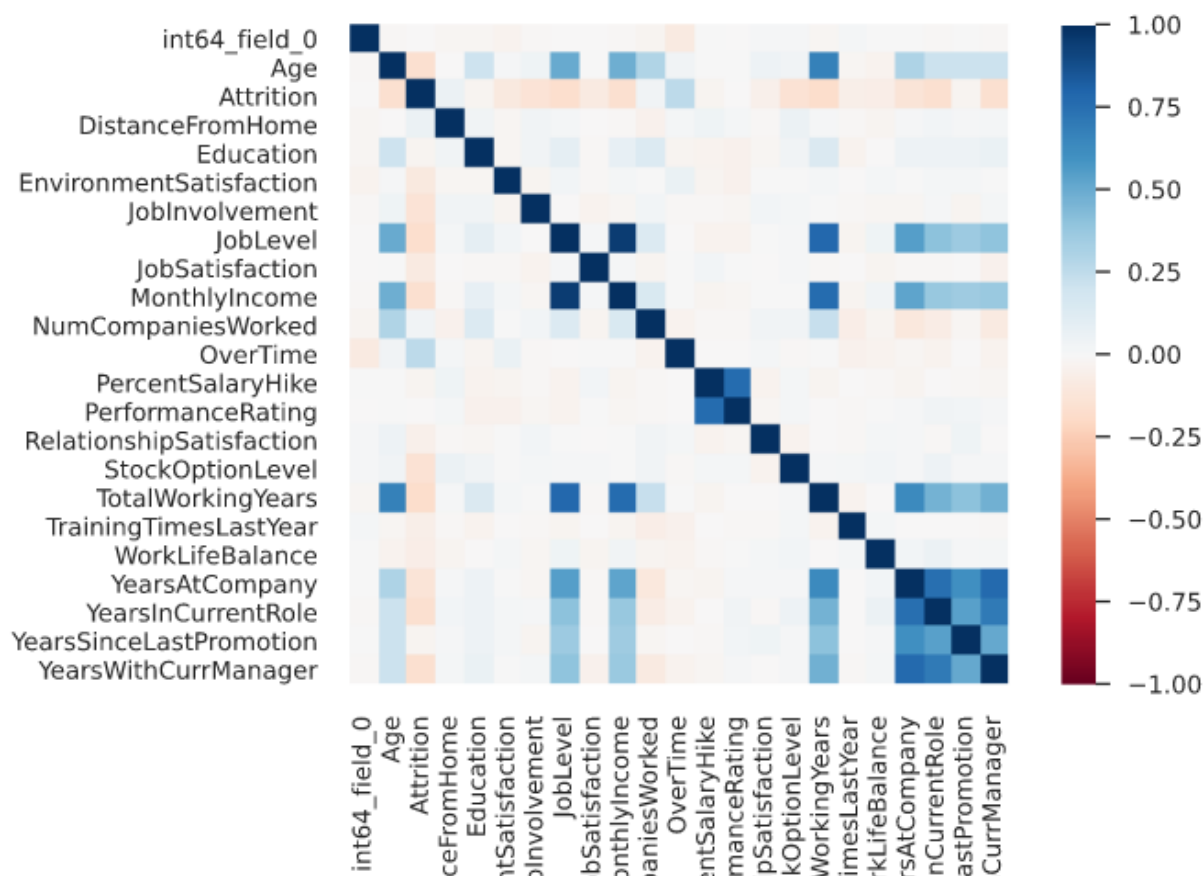
- Адаптационных мероприятий при устройстве на работу в компанию и при переходе на новую должность недостаточно для удержания специалистов, возможно их нужно доработать.
- Сотрудникам не хватает возможностей для карьерного развития внутри компании. Необходимо разработать карьерную карту.
- Возможно руководителям не хватает лидерских компетенций. Стоит пристальнее присмотреться к их стилю руководства и возможно провести дополнительное обучение
- Имеет место выгорание сотрудников. Необходимо понять почему сотрудникам приходится перерабатывать и разработать мероприятия по решению проблем с выгоранием

Гипотезы основаны на том, что:

- большая доля сотрудников (по отношению к работающим) увольняется в первый и второй год работы в компании (36% и 35%),
- чаще всего сотрудники увольняются в первый год работы в должности и в третий,
- в первый год работы под началом руководителей из компании увольняются 38-50% сотрудников (по соотношению с теми, кто остался работать) ,
- чаще всего увольняются молодые сотрудники с невысокой зарплатой,
- сотрудники, которые перерабатывают чаще увольняются

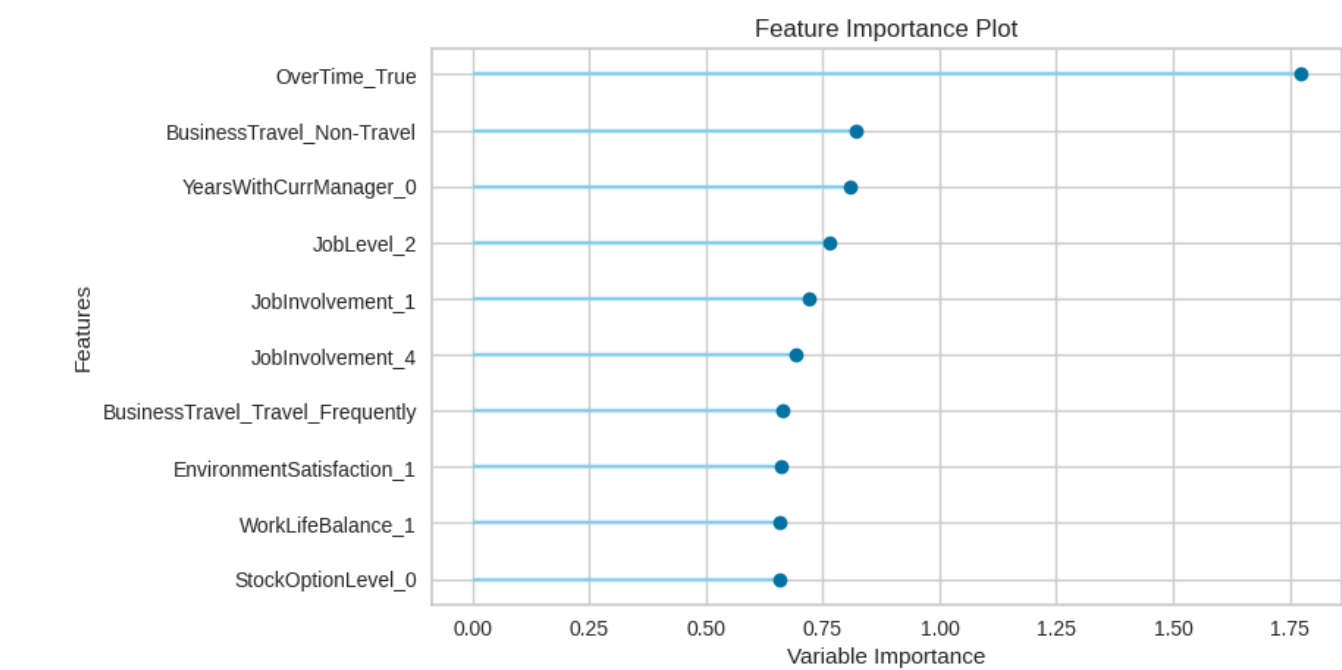
Дополнение

Исследование данных с помощью Python показало, какие данные с какими коррелируют:

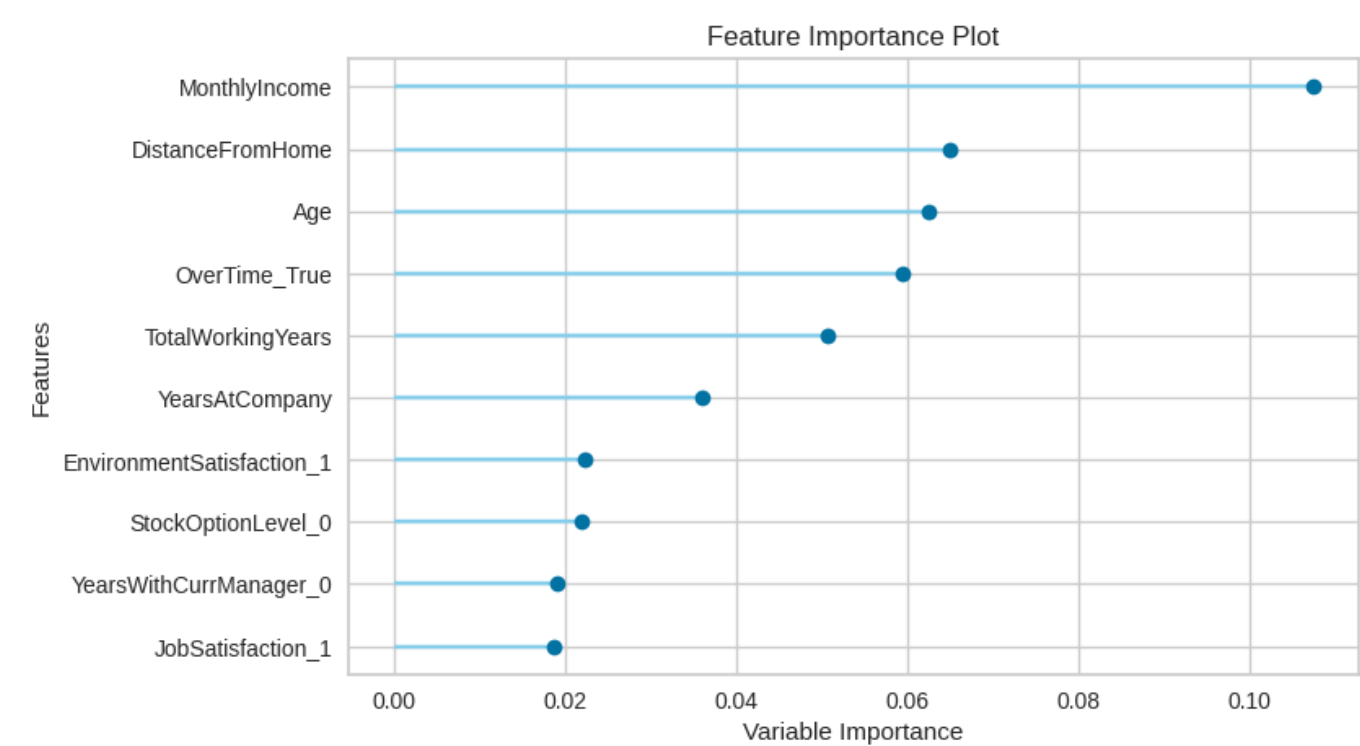


А машинное обучение с помощью Python показало важные параметры с точки зрения моделей:

```
[22] plot_model(final_lr_pycaret, plot='feature')
```



```
plot_model(final_gbc_pycaret, plot='feature')
```



Эти данные в целом подтверждают открытые закономерности в данных, что увольнение сотрудника зависит от:

- Переработок
- Зарботной платы
- Возраст
- Сколько лет проработал в компании/сколько лет вообще работает
-