• Ухудшается ли здоровье работника? Происходит выгорание? • Или он предпочитает больше времени проводить семьей и с близкими людьми? • Возможно причина в удаленности работы от места проживания, и время пребывания дома сильно сокращается? • Есть другие причины? 1. Во второй группе основными характеристиками оттока являются высокая оценка труда и низкая удволетворенность, но для выявления причин ухода, хорошо бы исследовать вопросы: • Имеет ли значение возраст? • Поможет ли увеличение часов работы по желанию работника, снизить количество увольнений? • Какова официальная причина ухода сотрудника? С учетом имеющихся данных, можно сказать, что для более глубокого исследования оттока не хватает сведений о работниках. Таких как: • должность, • возраст, • состояние здоровья, • семейное положение, • место проживания или время поездки от дома до работы и обратно, • официальная причина ухода. Возможные рекомендации На основе же имеющихся данных отделу кадров возможно рекомендовать следующее: 1. Полностью пересмотреть подход к распределению рабочих часов в компании. Сейчас нагрузка работников увеличивается пропорционально стажу. Но дело в том, что физиология и потенциал у всех людей разные и не увеличиваются с опытом работы по той же линейно-пропорциональной зависимости. Следовательно, какие-то работники способны выдерживать большие нагрузки, какие-то нет. Именно поэтому сотрудники с меньшим стажем должны иметь возможность выбора большего количества рабочих часов, а более опытные - меньшего. 2. В компании очень мало новичков. За 4 года былл принято на работу лишь 578 человек (6% от состава). В связи с этим, можно пожелать руководству сделать компанию более открытой для новых трудовых ресурсов, что тоже поспособствуют балансированию количества рабочих часов и возможно снимет перегрузку с более опытных сотрудников. Отчет об исследовании оттока работников компании import plotly plotly.offline.init notebook mode(connected=True) import plotly.express as px import plotly.graph objs as go from plotly.subplots import make subplots import pandas as pd df = pd.read csv('employee churn data.csv') В каком департаменте наблюдается наибольший отток работников? В каком - наименьший? bar = pd.crosstab(df.department, df.left, normalize='index').sort values('yes') pie = pd.crosstab(df.department, df.left) figure = make subplots(rows=1, cols=2, column widths=[0.6, 0.4], specs=[[{"type": "bar"}, {"type": "pie"}]], subplot titles=("","2784 former employees")) colors = px.colors.sequential.Reds[2:8] figure.add traces( [go.Bar(name='No', x=bar.no\*100, y=bar.index, orientation='h', marker color='rgb(129, 180, 179)', opacity=0.5), go.Bar(name='Yes', x=bar.yes\*100, y=bar.index, orientation='h', marker\_color='tomato', opacity=0.8), go.Pie(name='Left', labels=pie.index, values=pie.yes, marker\_colors=colors\*2, opacity=0.8, textinfo='label+percent', textposition='inside', hole=.2, showlegend=False)], rows=[1, 1, 1], cols=[1, 1, 2]) figure.update layout(title text='[A] Employee turnover by the departments', template='plotly white', barmode='stack', legend=dict(orientation="h", yanchor="top", y=1.1, xanchor="left", x=0.25, title\_text='Left, %:')) plotly.offline.iplot(figure) [A] Employee turnover by the departments Left, %: Yes No 2784 former employees ΙT logistics retail sales 16.9% retail 19.3% marketing financ 3.88% support engineering ΙT 15.7% 3.95% engineering logistics operations sales operations support 15.7% 7.61% admin marketing 8.73% finance 20 60 80 0 40 100 Данные представляют сведения о 9540 работниках, включая уволившихся сотрудников 2784 (29%). В компании 10 подразделений и на первый взгляд можно заметить, что наибольшее количество ушедших сотрудников в департаментах: sales, retail, engineering, operations, поскольку они являются наиболее крупными в компании. Но если обратить внимание на пропорцоинальное соотношение в каждом департаменте, то становится очевидным, что процент оттока примерно одинаковый. И специфика работы отдела не играет большой роли в текучести кадров. Имеет ли значение стаж работника? tenure\_ = pd.crosstab(df.tenure, df.left) max hours = df.groupby('tenure').avg hrs month.max().round(2).values fig = go.Figure() fig.add\_traces( [go.Bar(name='Remaining employees', y=tenure\_.no, x=tenure\_.index, marker color='rgb(129, 180, 179)', opacity=0.5), go.Bar(name='Left employees', y=tenure\_.yes, x=tenure\_.index, marker\_color='tomato', opacity=0.8), go.Scatter(x=tenure\_.index+0.45, y=[2100]\*11, name='Max monthly hours', mode="markers+text", marker=dict(color='goldenrod', symbol='triangle-right', size=20), text=list(max hours.astype('object')), textposition="top center")]) fig.update\_layout( title\_text='[B] Employee tenure and distribution of the avg. working hours monthly', template='ggplot2', #barmode='stack', xaxis title = 'Years', xaxis\_nticks = 12) plotly.offline.iplot(fig) [B] Employee tenure and distribution of the avg. working hours monthly 172.3 178.5 175.47 181.5 184.5 187.5 190.49 193.46 196.41 199.38 200.86 Remaining employees Left employees 2000 -Max monthly hours 1500 -1000 -500-0 -8 10 11 12 2 13 Years Исследуя стаж работников, определенно выявляется, что: • В компании заметно малое количество сотрудников 578 (**6% от всего состава компании**) с низким стажем (**до 5 лет**). **За последний год не было принято ни одного!** • С увеличением стажа возрастает количество рабочих часов в месяц на 3 часа. • Наибольший отток происходит на 7-8 годах стажа. Попробуем выяснить причины ухода, изучая удовлетворенность сотрудников: fig = px.scatter(df, x='avg\_hrs\_month', y='satisfaction', color= 'left', opacity=0.2, color discrete sequence=['rgb(129, 180, 179)', 'tomato'], marginal x = 'histogram') fig.update\_layout(title = '[C] Employee satisfaction vs. monthly work hours', template='plotly white', width = 900, height = 650, yaxis\_title = 'Satisfaction', yaxis nticks = 20, xaxis\_title = 'The average hours the employee worked in a month', xaxis nticks = 10, legend=dict(orientation="h", yanchor="top", y=0.99, xanchor="right", x=0.9, title\_text='Left:')) x0 = df.query('tenure == 7.0').avg\_hrs\_month.min() x1 = df.query('tenure == 8.0').avg hrs month.max() fig.add\_shape(fillcolor='goldenrod', opacity=0.2, x0=x0, x1=x1, y0=0, y1=1, xref="x", yref="paper", layer="below", line width=1) fig.add\_annotation(x=190, y=0.85, text="tenure: 7-8 years", showarrow=True, font=dict(family="Courier New, monospace", size=18), xanchor="left", arrowhead=2, arrowsize=2, ax=20, ay=-20,) plotly.offline.iplot(fig) [C] Employee satisfaction vs. monthly work hours Left: tenure: 7-8 years 0.9 0.8 0.7 Satisfaction 0.6 0.5 0.4 0.3 0.2 0.1 170 175 180 185 190 195 200 The average hours the employee worked in a month На визуализации мы можем увидеть две группы: • Первая группа представляет сотрудников со стажем более 7 лет. С повышением количества рабочих часов от 186 до 190 наблюдается внушительное количество увольняющихся со снижением степени удволетворенности работой. • Вторая группа - сотрудники со стажем до 7 лет, в основном с низкой степенью удовлетворенности. И если на данном этапе исследования, мы можем предположить, что причиной текучки в первой группе является увеличение количества рабочих часов, то про вторую пока сложно сказать что-либо определенное. Возможно, дело в зарплате? Рассматривая возможную зависимость оттока от стажа, в первую очередь можем заметить подтверждение нашего предположения о том, что увольняющихся работников более всего на 7 (37%) и 8 (43.6%) годах стажа. # Dividing data into the two groups first = df[df.avg\_hrs\_month > 186] first.left.value\_counts() Out[88]: no Name: left, dtype: int64 second = df[df.avg\_hrs\_month < 186]</pre> second.left.value counts() 4344 Out[58]: no 1321 yes Name: left, dtype: int64 In [5]: # Dividing data into the two groups first = df[df.avg hrs month > 186] second = df[df.avg\_hrs\_month < 186]</pre> figure2 = make subplots(rows=1, cols=2, subplot\_titles=("First group", "Second group")) figure2.add traces( [go.Violin(x=first['salary'][first.left == 'no'], y=first['satisfaction'][first.left == 'no'], name='Remaining', side='negative', line\_color='rgb(129, 180, 179)'), go.Violin(x=first['salary'][first.left == 'yes'], y=first['satisfaction'][first.left == 'yes'], name='Left', side='positive', line\_color='tomato'), go.Violin(x=second['salary'][second.left == 'no'], y=second['satisfaction'][second.left == 'no'], showlegend=False, side='negative', line color='rgb(129, 180, 179)'), go.Violin(x=second['salary'][second.left == 'yes'], y=second['satisfaction'][second.left == 'yes'], showlegend=False, side='positive', line\_color='tomato')], rows=[1, 1, 1, 1], cols=[1, 1, 2, 2]) figure2.update\_traces(meanline\_visible=True, scalemode='count') figure2.update\_xaxes(categoryarray=['low', 'medium', 'high'], title = 'Salary') figure2.update\_layout(title\_text='[D] Employee salary and satisfaction', width = 950, height = 500, violingap=0, violinmode='overlay', template='ggplot2') plotly.offline.iplot(figure2) [D] Employee salary and satisfaction Second group First group Remaining 1 -Left 1 -0.8-0.8-0.6-0.6-0.4-0.4-0.2-0.2-0 -0 – medium low high low medium high Salary Salary Что ж, определенно можем сказать, что зарплата большинства работников относится к уровню медиум и в целом уровень оплаты труда не характеризует удовлетворенность или утечку кадров. Но данное отображение показывает вновь различие между степенью удволетвореннсти ушедших и оставшихся в обеих группах: • в первой группе уходят работники со средней степенью удовлетворенности 0.58, остаются - с 0.45; • во второй группе уходят сотрудники с 0.41, остаются - с 0.54. Могут ли помочь данные об оценке труда понять, что вынуждает покидать компанию работников второй группы? fig = px.scatter(df, x='avg\_hrs\_month', y='review', color= 'left', opacity=0.1, color\_discrete\_sequence=['rgb(129, 180, 179)', 'darkorange'], marginal\_x = 'histogram', marginal\_y = 'histogram') fig.update\_layout(title = '[E] Employee review score', template='plotly\_white', width = 950, height = 650, yaxis\_title = 'Review score', yaxis\_nticks = 10, xaxis\_title = 'The average hours the employee worked in a month', xaxis\_nticks = 10, legend=dict(orientation="h", yanchor="top", y=0.9, xanchor="right", x=0.99, title\_text='Left:')) fig.add\_shape(type="line", line\_color='tomato', line\_width=3, opacity=1, line\_dash="dot", x0=186, x1=186, y0=0.3, y1=1) fig.add\_annotation(x=186, y=0.9, text="186 hours", showarrow=True, font=dict(family="Courier New, monospace", size=18), xanchor="left", arrowhead=2, arrowsize=2, ax=20, ay=-20,) plotly.offline.iplot(fig) [E] Employee review score Left: 186 hours 0.9 0.8 Review score 0.7 0.6 0.5 0.4 0.3 175 180 185 190 170 195 200 The average hours the employee worked in a month Итак, первое, что бросается в глаза - тенденция, обратная удовлетворенности. То есть, складывается впечатление обратной зависимости между удовлетворенностью работника и его профессиональной оценкой. Второе - большая часть увольняющихся имеют высокую оценку, и особенно это проявлено во второй группе. Возможно ушедшие не получают достаточного количества проектов для реализации своих способностей? def plot\_violin(data, data\_name): In [7]: fig = px.violin(data, x='projects', y='review', color\_discrete\_sequence=['rgb(129, 180, 179)', 'darkorange'], box=True) fig.update\_layout(title = '[F] ' + str(data\_name) + ' employees workload', template='ggplot2', #'plotly\_white', width = 900, height = 500, yaxis\_title = 'Review score', xaxis\_title = 'Number of projects the employee is involved in', legend=dict(title\_text='Left:')) fig.update\_traces(scalemode='count') plotly.offline.iplot(fig) plot\_violin(first, 'First group') plot\_violin(second, 'Second group') [F] First group employees workload Left: no 1 yes 0.9-0.8-Review score 0.7-0.6-0.5-0.4-0.3-3 2 5 Number of projects the employee is involved in [F] Second group employees workload Left: 1 -0.9-0.8-Review score 0.7-0.6-0.5-0.4-0.3-5 Number of projects the employee is involved in В компании в обоих группах большая часть работников (примерно 93%) задействованы в 3-4 проектах одновременно. Поэтому предположение о влиянии количества проектов на текучку, скорее всего не актуально. Но при этом, мы видим еще одно подтверждение тому, что ушедшие по сравнению с оставшимися во второй группе определенно имеют в среднем более высокую оценку по последней аттестации. Что удалось выяснить на основе имеющихся данных? tab = go.Figure(data=[go.Table( In [9]: header=dict( values=['<b>Employees</b>','<b>First group</b>', '<b>Second group</b>', '<b>Visualisation</b>'], line color='white', fill\_color= 'tan', #rgb(129, 180, 179)', 'cadetblue', align=['left','center'], font=dict(color='white', size=12) ), cells=dict( values=[ ['<b>All: 9540</b>', 'Tenure', 'Avg.work hours monthly', 'Promoted', 'Bonus', '<b>Left: 2784</b>', 'Satisfaction (avg.)', 'Review score (avg.)', '<b>Prominent</b>'], ['<b>3875</b>', '>= 7 years', '> 186 hours', '2.7 %', '21 %', '<b>1463 (37.75%)</b>', 0.58, 0.64, 'tenure: 7-8 years, <br><b>186-190 work hours</b>, <br>decreasing satisfaction'], ['<b>5665</b>', '<= 7 years', '< 186 hours', '3.3 %', '21 %', '<b>1321 (23.3%)</b>', 0.41, 0.76, 'perceptibly low number with <br/>tenure: < 5 years, \ <br><b>high review score</b>,<br>low satisfaction'], ['A, C, E', 'B', 'C', '', '', 'A, C, E', 'C, D', 'E, F', ''] line\_color='white', fill\_color = [(['lightblue'] + ['azure']\*4 + ['lightsalmon'] + ['linen']\*3)\*3], #navajowhite align = ['left', 'center'], font = dict(color = 'darkslategray', size = 12), height=30 )) ]) tab.update\_layout(title = 'Exploration results:') plotly.offline.iplot(tab) Exploration results: **Second group** Visualisation **Employees** First group A, C, E All: 9540 3875 5665 В Tenure <= 7 years>= 7 years C Avg.work hours monthly > 186 hours < 186 hours 2.7 % Promoted 3.3 % Bonus 21 % 21 % A, C, E Left: 2784 1463 (37.75%) 1321 (23.3%)

0.58

0.64

**186-190** work hours,

decreasing satisfaction

tenure: 7-8 years,

0.41

0.76

perceptibly low number with

tenure: < 5 years,

low satisfaction

high review score,

Satisfaction (avg.)

Review score (avg.)

**Prominent** 

C, D

E, F

Инсайты

принято ни одного человека!

часов становятся критичными?

• Наибольший отток происходит на 7-8 годах стажа.

• Данные представляют сведения о 9540 работниках, включая уволившихся - 2784 (29%). В компании 10 подразделений, и все же процентное соотношение оттока по департаментам

• Первая группа, со стажем выше 7 лет, теряет наибольшее количество сотрудников (37.75% группы) с повышением рабочих часов. Учитывая, что работники с высокой степенью

• Также, не влияет на отток количество актуальных проектов, в которых участвует сотрудник. Большая часть работников (примерно 93%) задействованы в 3-4 проектах одновременно.

• Необходимо сказать, что в данном отчете выявлены лишь проблемные участки, но для более полного понимания причин оттока, нужны дополнительные данные и исследования.

1. Прежде чем выбрать предиктором оттока количество рабочих часов в первой группе, желательно рассмотреть дополнительные данные, чтобы ответить на вопрос: почему 186-190

• В целом, стаж сотрудников компании составляет от 2 до 12 лет включительно. Заметно малое количество работников (6% штата) с низким стажем (до 5 лет). За последний год не было

примерно одинаковое. То есть специфика работы отдела не играет большой роли в текучести кадров.

• При этом повышение за последние 2 года получили около 3% от всего состава компании.

Данные, необходимые для дальнейших исследований

• Среднее количество рабочих часов в месяц повышается на 3 часа пропорционально возрастанию стажа на 1 год.

• Для лучшего понимания, состав компании желательно рассматривать как две группы с разными причинами ухода сотрудников.

удовлетворенности начинают покидать компанию, можно предположить, что 186-190 часов становятся для большинства критичным.

• Вторую группу, со стажем менее 7 лет, оставляют работники (23.3% группы) с высокой оценкой труда 0.76 и низкой степенью удволетворенности.

• Зарплата большинства работников (69%) относится к медиум и в целом уровень оплаты труда не характеризует удовлетворенность или утечку кадров.