

Perpetuum Mobile

Фокин Вячеслав

Суворов Дмитрий

Дудникова Надежда

Структура данных и очистка

- Дана информация о сотрудниках call-центра и их работе в разные дни
- Данные представлены подневно для каждого сотрудника
- Данные за 5 месяцев

- Удалили бесполезные строки, где:
 - `all_system_fact == 0`
 - `work_activity_plan == 0`
- Итоговое количество строк:
170774 -> 141603

Исследовательский вопрос и гипотеза



Как выглядит
портрет
эффективного
сотрудника?



Люди с высоким
стажем, и не
отвлекающиеся на
рабочем месте,
более эффективны

skill_group_exp

1.00 0.18 -0.04 -0.07

efficiency

0.18 1.00 -0.11 -0.35

break_fact

-0.04 -0.11 1.00 -0.04

additional_fact

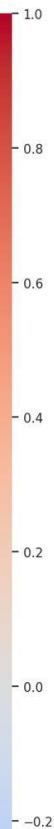
-0.07 -0.35 -0.04 1.00

skill_group_exp

efficiency

break_fact

additional_fact



Что такое эффективность?

- Определили переменную эффективности как сумму полезного времени: всё avail_time и useful_time, делённую на всё время, проведённое на работе: all_system_fact

всё useful time

всё avail time

$$\frac{(\text{useful_calls_fact} + \text{useful_chats_fact}) + (\text{avail_chat_call_fact} + \text{avail_chats_fact} + \text{avail_calls_fact})}{\text{all_system_fact}}$$

Очистка данных 2.0

- Обработали выбросы по эффективности:
 - Удалили строки со значениями эффективности > 95 перцентиля и строки со значениями эффективности < 5 перцентиля
- Итоговая очистка данных:
 - $17074 \Rightarrow 141603 \Rightarrow ???$

Математическая модель:

$$\text{Эффективность} = \beta_0 - (\beta_1 \times \text{skill_group_exp}) + (\beta_2 \times \text{education_fact}) + (\beta_3 \times \text{additional_fact}) - (\beta_4 \times \text{break_fact})$$

Результат регрессии:

OLS Regression Results

```
=====
Dep. Variable:          efficiency    R-squared:                0.585
Model:                  OLS          Adj. R-squared:           0.585
Method:                 Least Squares  F-statistic:             2.608e+04
Date:                  Sun, 16 Mar 2025  Prob (F-statistic):       0.00
Time:                  12:53:43       Log-Likelihood:          53910.
No. Observations:      92367         AIC:                    -1.078e+05
Df Residuals:          92361         BIC:                    -1.078e+05
Df Model:               5
Covariance Type:       nonrobust
=====
```

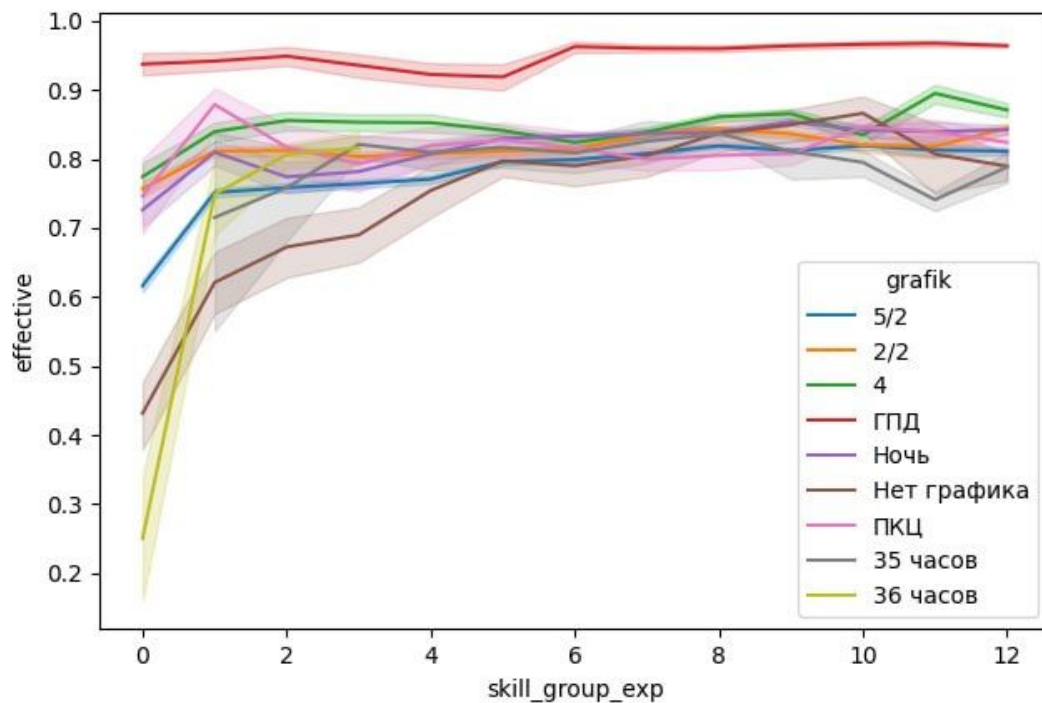
```
=====
               coef    std err          t      P>|t|      [0.025      0.975]
-----
const           0.7527      0.001    773.909      0.000      0.751      0.755
education_fact  -0.6387      0.011   -59.571      0.000     -0.660     -0.618
dlg_time_chat   0.5121      0.002   207.769      0.000      0.507      0.517
education_plan  -0.1679      0.007   -25.472      0.000     -0.181     -0.155
avail_calls_fact -0.2460      0.006   -39.770      0.000     -0.258     -0.234
all_smena_plan  -0.4709      0.002  -221.582      0.000     -0.475     -0.467
=====
```

```
=====
Omnibus:            18009.634    Durbin-Watson:           1.994
Prob(Omnibus):      0.000      Jarque-Bera (JB):        43909.962
Skew:               -1.089      Prob(JB):                0.00
Kurtosis:           5.581      Cond. No.                30.3
=====
```

Механизм



Устойчивость матмодели



Интерпретация и ограничения

Интерпретация:

-

Ограничения:

- Возможно, неподтверждаемость гипотезы для ГПД связана с зарплатой, процентом людей на удалёнке и тд, это непроверяемое с данным датасетом
- Более точное определение эффективности должно отталкиваться от отзывов клиентов

Применение

Одним из важных критериев приёма на работу - опыт работника в данной скилл-группе