

Анализ базы данных проекта «Московское долголетие»

5 команда “Нижегородские аналитики”

Особенности данных/зависимости и шаблоны

Особенности данных:

- **Выбросы:** участники моложе 44 (молодые люди), участники старше 90 лет (долгожители), мужчины (количество мужчин в датасете в разы меньше количества женщин)
- **Пропуски:** отсутствуют данные по месяцу март; имеются граждане, которые лишь зарегистрировались, но не ходили впоследствии
- **Неравномерность по подгруппам:** количество мужчин в разы меньше (в 9 раз), чем женщин

Зависимости/Паттерны:

- **Гендерные различия в направлениях:**
 - Женщины в основном предпочитают спортивные и творческие направления.
 - Мужчины чаще занимаются спортом.

Введение

Гипотеза

- Люди, которые выбирают одно направление, чаще уходят из программы, чем те, кто выбирают больше одного направления

Механизм гипотезы

- Наша гипотеза может выполняться потому что люди, идущие на одну программу, не знают о других, более интересных направлениях и, как следствие, отключаются

Переменные

- Возраст, кол-во направлений,
- Основные направления у мужчин и женщин,
- Количество отключившихся от программы

Методы исследования

- Как показывает анализ датасета, наибольшее кол-во участников выбирает более одного направления, что стимулирует их посещать занятия чаще, чем те, кто выбирает только одно занятие.
- Это связано с тем, что есть популярные курсы с наименьшим количеством отключений и наименее популярные курсы
- Мы проанализировали и нашли самые популярные направления у мужчин и женщин.



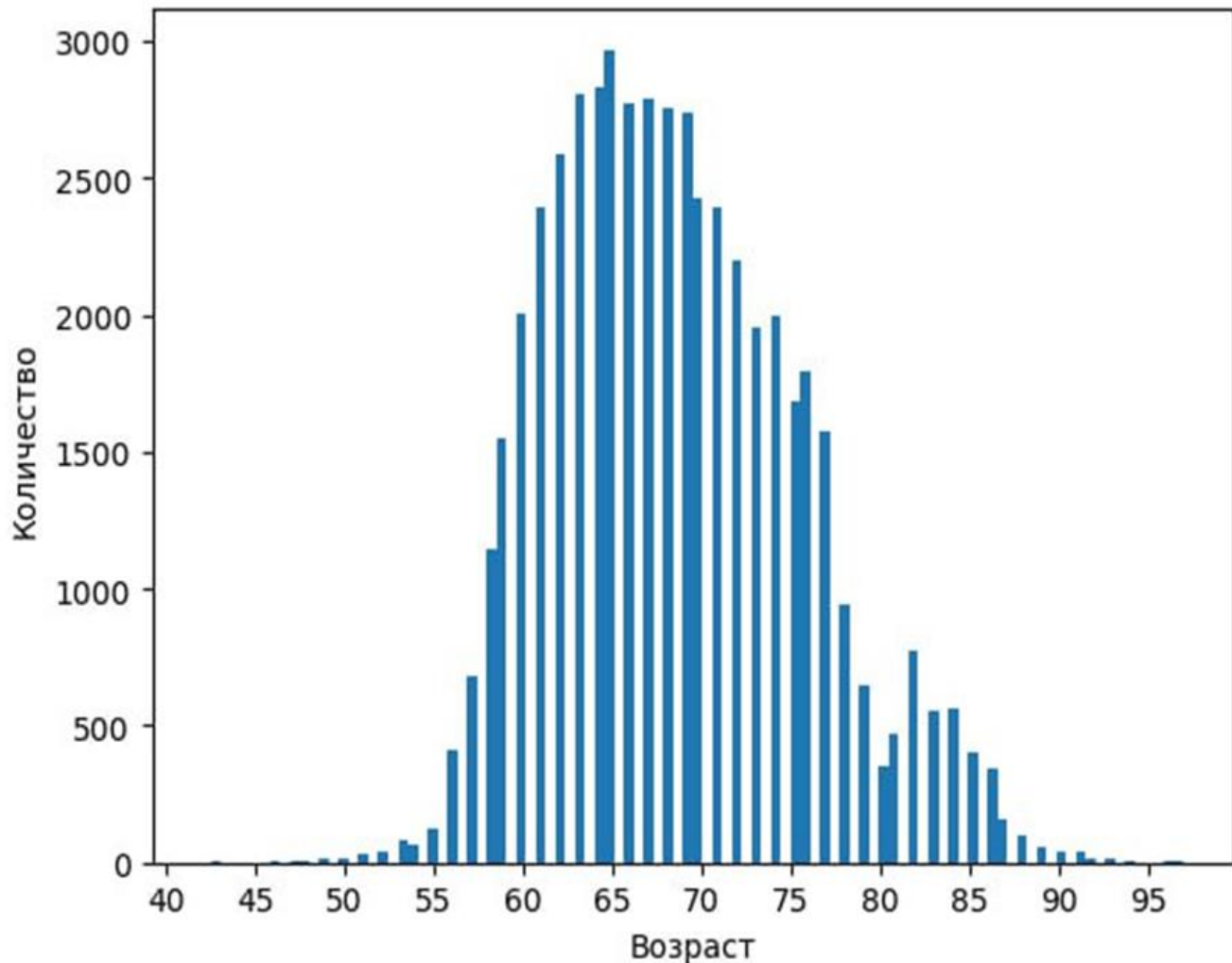
Женщина



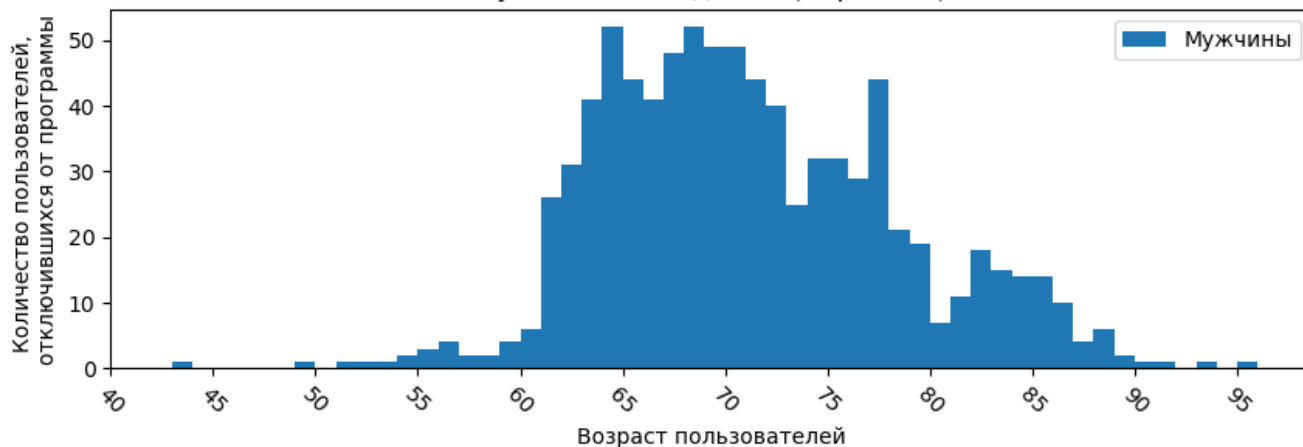
Мужчина



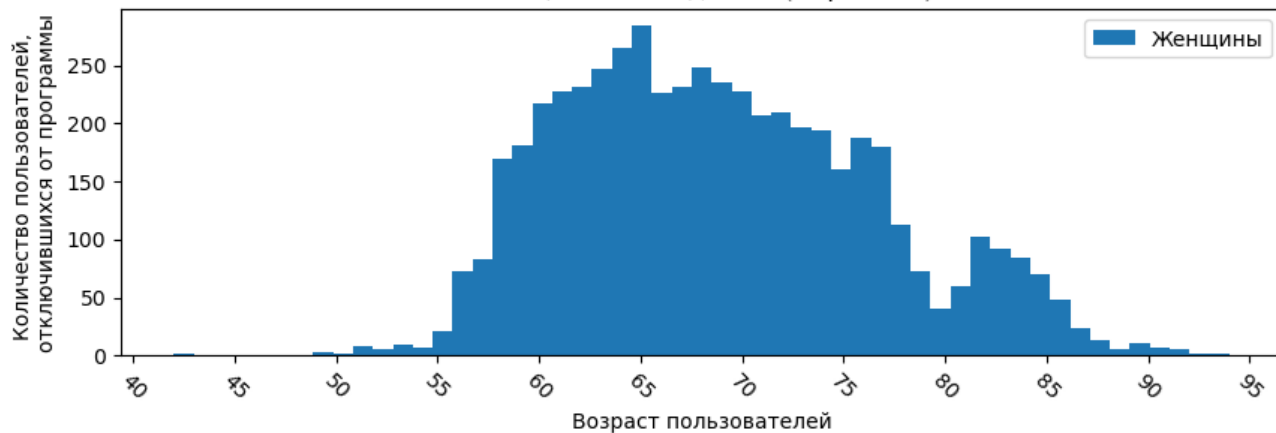
- Согласно Социальному фонду России, в Москве проживают 3.2 млн пенсионеров. Среди них в программе участвуют 380 тысяч человек, что примерно 12% от всех пожилых людей.
- К тому же, наибольшая доля участников приходится на возраст от 60 до 70 лет (см. график):



Мужчины от 40 до 100 (Порог = 1)

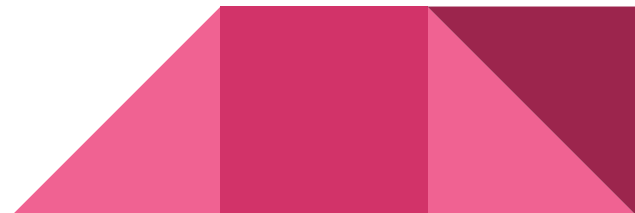


Женщины от 40 до 100 (Порог = 1)



- Более того, данные возрастные группы имеют самый высокий показатель отключения от программы (см. графики)

- По такой причине формируется две круговые диаграммы самых непопулярных направлений среди мужчин и женщин (см. диаграммы):



Женщина

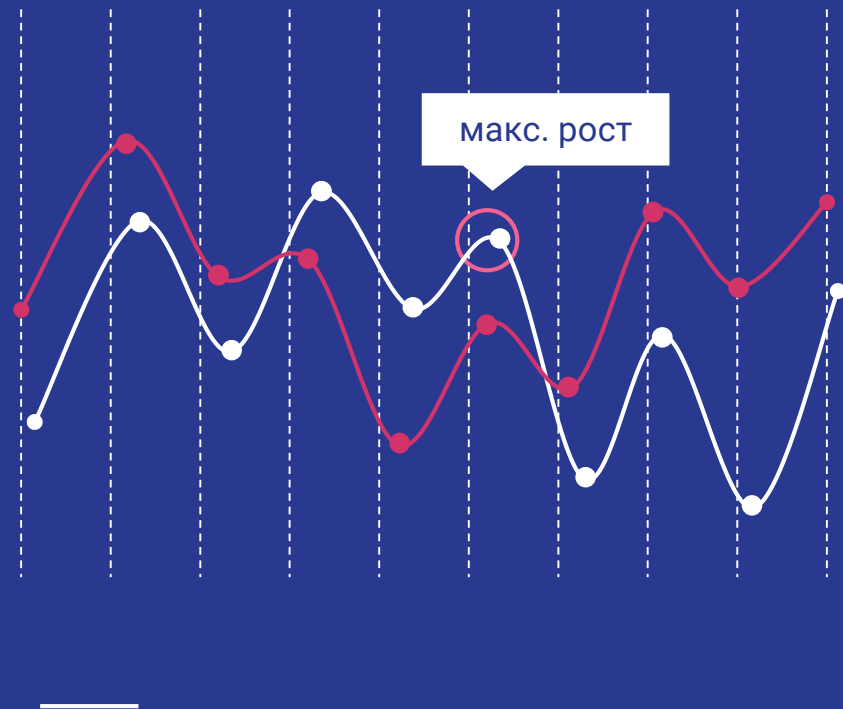


Мужчина



Выводы

- В данном исследовании датасета мы затронули актуальность программ для жителей Москвы преклонного возраста.
- В свете данной информации требуется дать определенные рекомендации для привлечения и удержания большего количества участников среди людей пенсионного возраста города Москвы.



Рекомендации

Разнообразие программ:

- Рекламируйте разнообразие программ в газетах и на национальном телевидении, подчеркивая спорт, интеллектуальные и творческие направления.

Попробуйте разное:

- Советуйте новым участникам попробовать занятия из различных сфер, чтобы они могли открыть для себя разные интересы и таланты.

Акцент на здоровье:

- Подчеркивайте важность интеллектуальных и творческих направлений для тех, кто из-за здоровья не может участвовать в спорте регулярно. Это поможет им оставаться активными и вовлеченными в программу.

Долгосрочное участие:

- Советуйте участникам, находящимся в программе уже длительное время, обращать внимание на интеллектуальные и творческие направления, чтобы разнообразить свой опыт и продлить интерес к участию.

Новое поколение:

- Обращайтесь к новому поколению и поколению заботящихся о здоровье людей, предлагая им не только физические, но и умственные и культурные вызовы в рамках программы "Московское долголетие".

```

In [125]: %%time
import numpy as np
import matplotlib.pyplot as plt
from scipy.stats import gaussian_kde

fig = plt.figure()
ax = fig.add_subplot()
sex_test = "Женщина"

plt.title(f"Женщины и Мужчины от {low_age} до {high_age} (порог = {porog})")
plt.ylabel("Отношение пользователей, \отключившихся от программы")
plt.xlabel("Возраст пользователей")

g_women = []
g_men = []

for i in res_u:
    if i[0] == "Женщина":
        g_women.append(i[1])
    elif i[0] == "Мужчина":
        g_men.append(i[1])

# Оценка плотности распределения для женщин
kde_women = gaussian_kde(g_women)
x_women = np.linspace(min(g_women), max(g_women), 100)
y_women = kde_women(x_women)

# Оценка плотности распределения для мужчин
kde_men = gaussian_kde(g_men)
x_men = np.linspace(min(g_men), max(g_men), 100)
y_men = kde_men(x_men)

# Отображение гистограмм и огибающих
plt.hist(g_women, label="Женщины", bins=high_age-low_age-7, density=True, alpha=0.7)
plt.hist(g_men, label="Мужчины", bins=high_age-low_age-7, density=True, alpha=0.7)
plt.plot(x_women, y_women, 'b-', linewidth=2)
plt.plot(x_men, y_men, 'r-', linewidth=2)

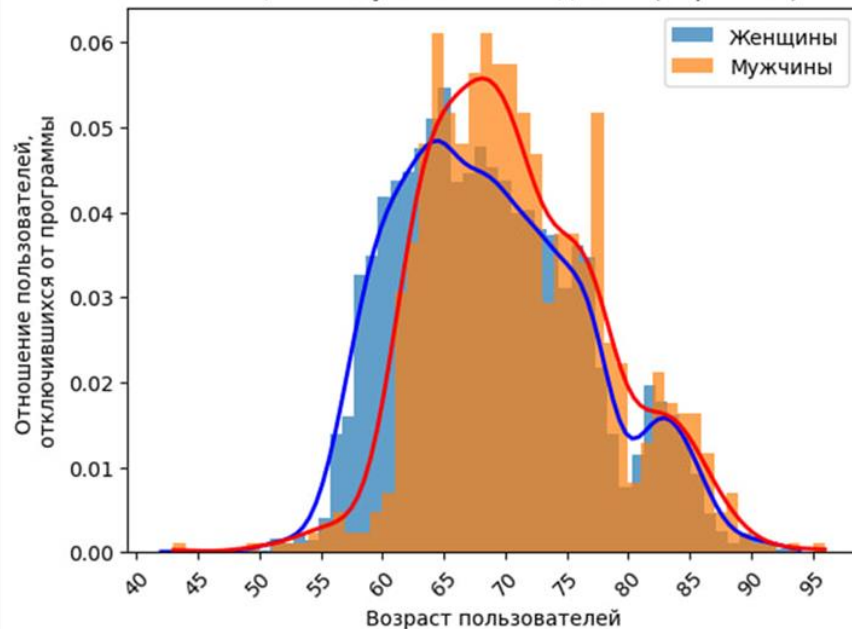
plt.xticks(list(range(low_age, high_age, 5)))
for label in plt.gca().xaxis.get_ticklabels():
    label.set_rotation(45)

plt.legend()
plt.show()

```

Q&A

Женщины и Мужчины от 40 до 100 (Порог = 1)



```
In [34]: %%time
import numpy as np
import matplotlib.pyplot as plt
```

```
fig = plt.figure()
ax = fig.add_subplot()

print(count_day, age)
plt.ylabel("Количество")
plt.xlabel("Возраст")

plt.hist(age, bins=100)
plt.xticks(list(range(40, 100, 5)))

plt.show()
```

```
[124 72 24 ... 28 27 90] [64 65 67 ... 73 76 69]
```

Q&A

