

Исследование потенциала
госпитального рынка для
продвижения госпитального
портфеля компании

Дмитриева Н.Ю.

Июль 2021



Предпосылки проекта

- Компания, имеющая свой штат продвижения, открывает линию госпитального продвижения
- Для успешного выхода на рынок необходимо выделить лечебно-профилактические учреждения (ЛПУ), которые обеспечивают 70 - 80% тендерного госпитального бизнеса на территориях ответственности каждого из сотрудников
- Анализ данных на основе тендерных закупок невозможен из-за специфичности препарата
- Для сегментации ЛПУ выбран объем одного из видов анестезий, которая проводится с применением препарата компании
- Так как в распоряжении компании есть информация об объемах анестезий не во всех ЛПУ, но в полном объеме собраны данные по коечному фонду, было предложено спрогнозировать недостающие данные по анестезиям на основе регрессионной модели

Цель и задачи проекта

Цель проекта

- Исследование потенциала госпитального рынка Палексии и гинекологического портфеля компании с сегментацией таргетных ЛПУ (ТА)

Задачи проекта

- Провести сегментацию таргетных аудиторий госпитального рынка Палексии и гинекологического портфеля компании;
- Разработать план распределения Регионов ответственности в команде госпитального продвижения с их распределением по ЛПУ и при возможности формирования индивидуальных баз клиентов.

Исходные данные

Источник информации:

Ежегодная отчетность бюджетных ЛПУ, передаваемая в региональные организации медицинской статистики

Собираемая отчетность:

Форма № 14 - «Сведения о деятельности стационара»

- таблица 2000 – состав пациентов
- таблица 4000 – хирургическая работа стационара
- таблица 4100 – виды анестезий
- таблица 4201 – трансплантология

Форма № 30 - «Сведения о медицинской организации»

- таблица 1001 – отделения, кабинеты
- таблица 1100 – медицинский персонал
- таблица 3100 – коечный фонд и его использование
- таблица 3200 – трансфузионная помощь

Детализация форм отчетности стационаров

Состав пациентов (Ф.№14 т.2000)

- Все классы болезней
- Классификация до 5-го уровня

Показатели:

- Поступило, выписано, умерло пациентов (взр., дет.)
- Выписано: из них доставленных по экстренным показаниям (взр., дет.)
- Выписано: из них доставленных СМП (взр., дет.)
- Проведено пациентами койко-дней (взр., дет.)

Хирургическая работа стационара (Ф.№14 т.4000, т.4100, т.4201)

- Все виды операций (вкл. трансплантации и все виды анестезий)
- Классификация до 5-го уровня

Показатели:

- Число операций (взр., дет.)
- С применением ВМТ
- С осложнениями
- Число операций по поводу злокачественных образований
- **Число анестезий**

Коечный фонд (Ф.№30 т.3100)

- Все профили коек
- Классификация до 4-го уровня

Показатели:

- Кол-во среднегодовых коек
- Поступило пациентов (взр., дет.)
- Выписано, умерло
- Проведено пациентами койко-дней

Трансфузионная помощь (Ф.№30 т.3200)

Все виды трансфузионных средств

Показатели:

- Число пациентов
- Число переливаний
- Объем ТФС (л)
- Число посттрансфузионных осложнений

Дополнительная информация (Ф.№30 т.1001, т.1100)

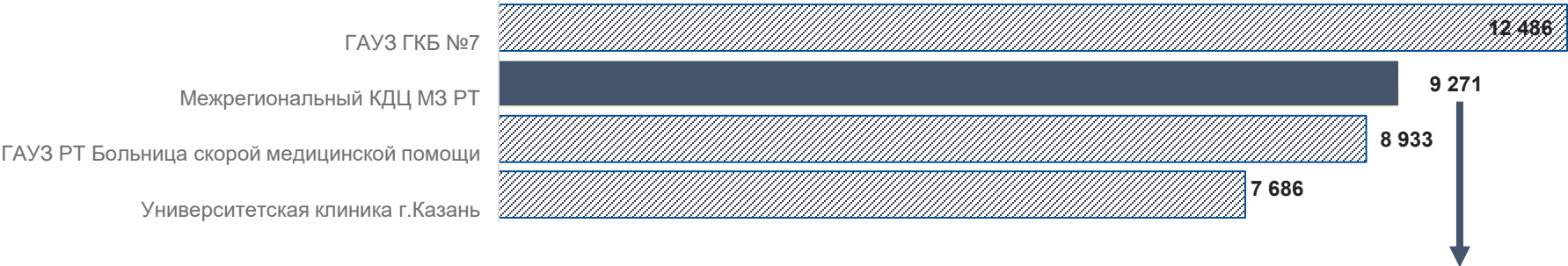
- Все наименования отделений (кабинетов)
- Все наименования должностей врачей

Показатели:

- Число подразделений/отделений, кабинетов
- Число кабинетов
- Число врачей в поликлиниках, в стационарах

Структура данных на примере Республики Татарстан

Крупнейшие ЛПУ по количеству пациентов с болезнями системы кровообращения



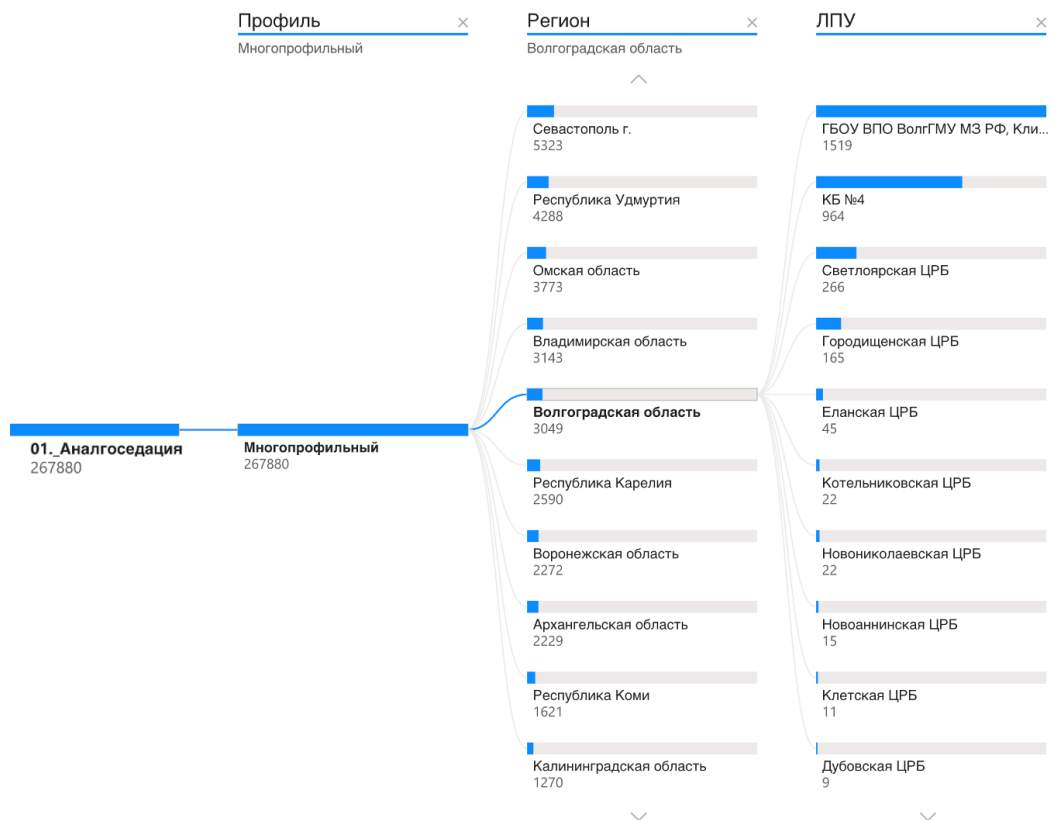
Название болезни	Кол-во пациентов за год
Болезни системы кровообращения	9 271
хронические ревматические болезни сердца	21
болезни, характеризующиеся повышенным кровяным давлением	1 047
эссенциальная гипертензия	225
гипертензивная болезнь сердца (гипертоническая болезнь с преимущественным поражением сердца)	813
гипертензивная болезнь почек (гипертоническая болезнь с преимущественным поражением почек)	5
гипертензивная болезнь сердца и почек (гипертоническая болезнь с преимущественным поражением сердца и почек)	4
ишемические болезни сердца	4 007
из них: стенокардия	2 629
из нее: нестабильная стенокардия	2 064

Этапы сегментации ЛПУ

Данные дают возможность спрогнозировать объем анестезий и провести объективную сегментацию ЛПУ при отсутствии данных продаж в них

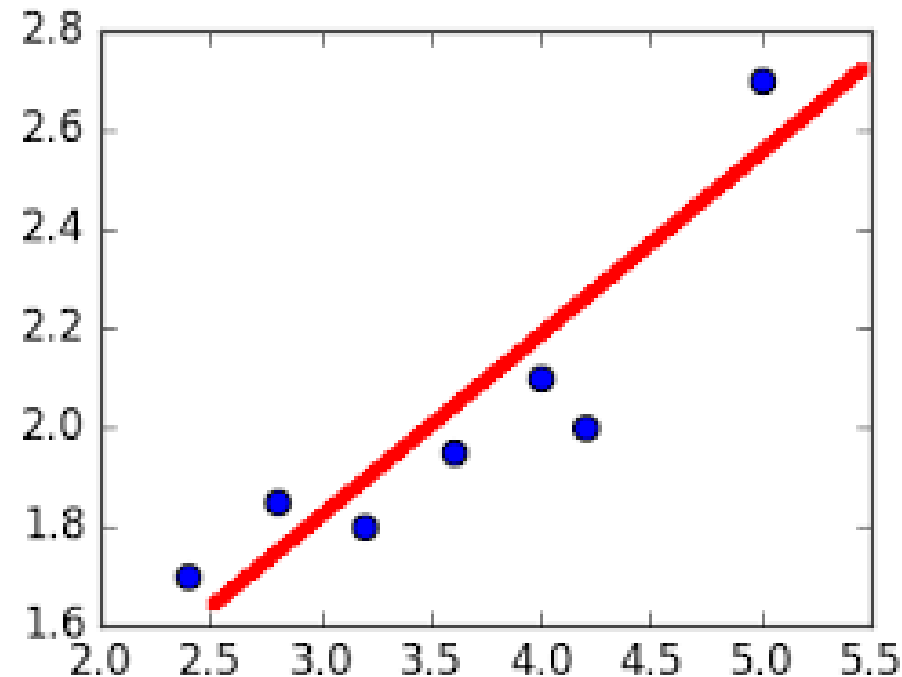
Шаг 1

Сенсус / Данные выгрузки аналгоседаций в ЛПУ



Шаг 2

Формируем БД аналгоседаций на основании фактических данных, и данных полученных на основе регрессионной модели, построенной на основе объемов анестезий и данных о количестве коек определенного профиля



Моделирование



Коечный фонд

К исходным данным были добавлены группирующие профили, независящие от возрастной категории

Коечный фонд (исходные данные)		
01_ Неуточненные данные	27_ онкологические для детей	53_ травматологические для взрослых
02_ аллергологические для взрослых	28_ оториноларингологические для взрослых	54_ травматологические для детей
03_ аллергологические для детей	29_ оториноларингологические для детей	55_ ортопедические для взрослых
04_ для беременных и рожениц (кроме патологии беременности)	30_ офтальмологические для взрослых	56_ ортопедические для детей
05_ для патологии беременности	31_ офтальмологические для детей	57_ туберкулезные для взрослых
06_ гинекологические для взрослых	32_ ожоговые	58_ туберкулезные для детей
07_ гинекологические для детей	33_ паллиативные для взрослых	59_ урологические для взрослых
08_ гастроэнтерологические для взрослых	34_ паллиативные для детей	60_ урологические для детей
09_ гастроэнтерологические для детей	35_ педиатрические соматические	61_ хирургические для взрослых
10_ гематологические для взрослых	36_ проктологические	62_ хирургии абдоминальной
11_ гематологические для детей	37_ психиатрические для взрослых	63_ хирургические для детей
12_ геронтологические	38_ психиатрические для детей	64_ нейрохирургические для взрослых
13_ дерматологические для взрослых	39_ профпатологические	65_ нейрохирургические для детей
14_ дерматологические для детей	40_ пульмонологические для взрослых	66_ хирургии торакальной для взрослых
15_ венерологические для взрослых	41_ пульмонологические для детей	67_ хирургии торакальной для детей
16_ венерологические для детей	42_ радиологические	68_ кардиохирургические
17_ инфекционные для взрослых	43_ реабилитационные соматические для взрослых	69_ хирургии сосудистой
18_ инфекционные для детей	44_ реабилитационные соматические для детей	70_ хирургические гнойные для взрослых
19_ кардиологические для взрослых	45_ реанимационные	71_ хирургические гнойные для детей
20_ кардиологические для детей	46_ ревматологические для взрослых	72_ хирургии челюстно-лицевой
21_ наркологические	47_ ревматологические для детей	73_ стоматологические для детей
22_ неврологические для взрослых	48_ сестринского ухода	74_ эндокринологические для взрослых
23_ неврологические для детей	49_ скорой медицинской помощи краткосрочного пребывания	75_ эндокринологические для детей
24_ нефрологические для взрослых	50_ скорой медицинской помощи суточного пребывания	76_ прочие койки для взрослых
25_ нефрологические для детей	51_ терапевтические	77_ прочие койки для детей
26_ онкологические для взрослых	52_ токсикологические	

Дополнительные переменные

Категориальные:

- Федеральные округа (Москва и все ФО)
 - Добавлена переменная Москва (Да, Нет)
 - Добавлена переменная с частотой встречаемости каждого ФО в выборке
- Подчинение ЛПУ (муниципальное, региональное, субрегиональное, федеральное)
 - Были выделены переменные, отвечающие за каждый тип
- Поселение (городское, сельское)
 - Были выделены переменные, отвечающие за каждый тип

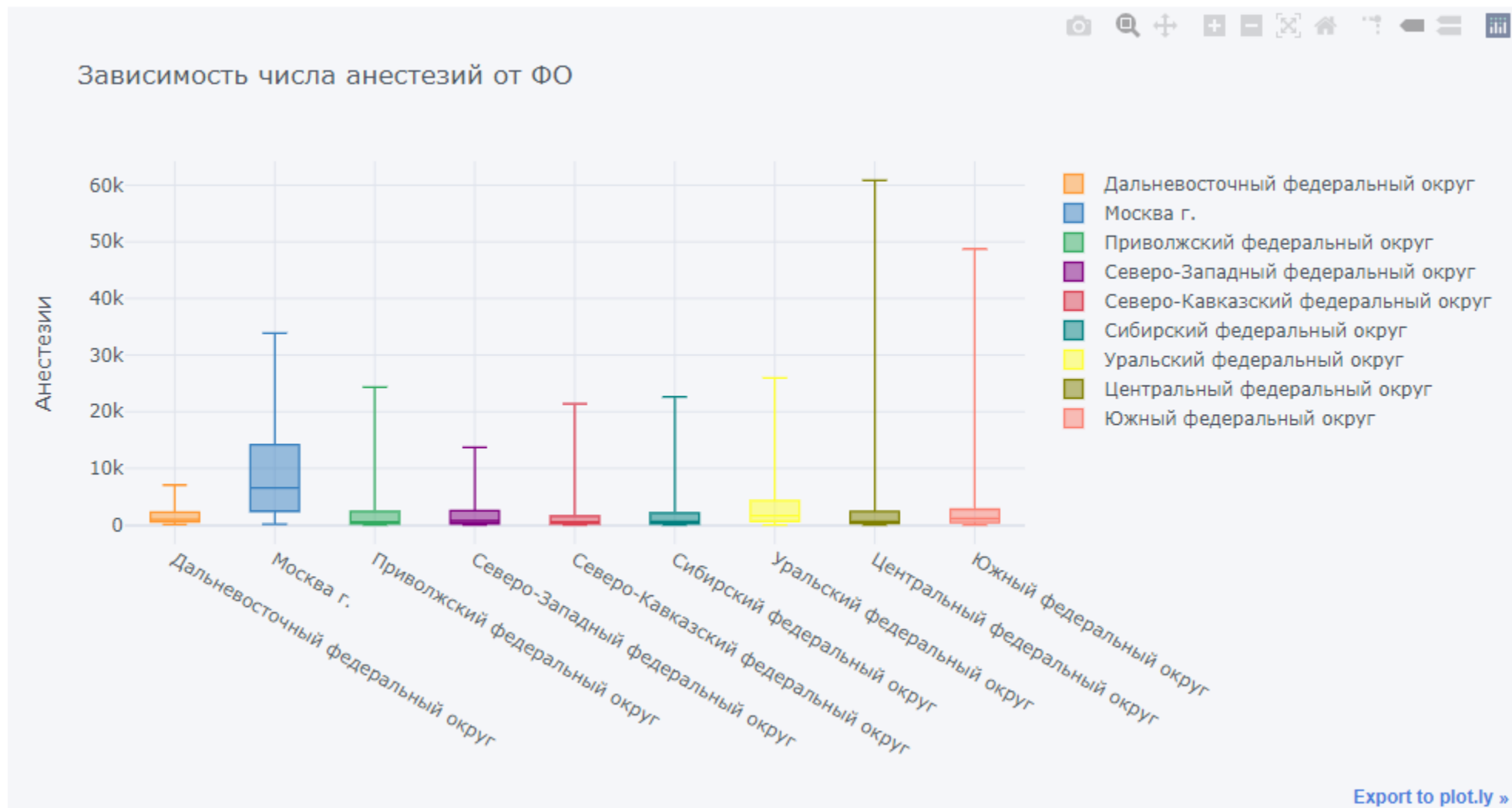
Внутри каждой переменной выявлена хотя бы одна группа значимо отличающиеся от других

Количественные

- Объем территориальных программ государственных гарантий (ТПГГ), отражающий финансирование медицинской помощи в регионе

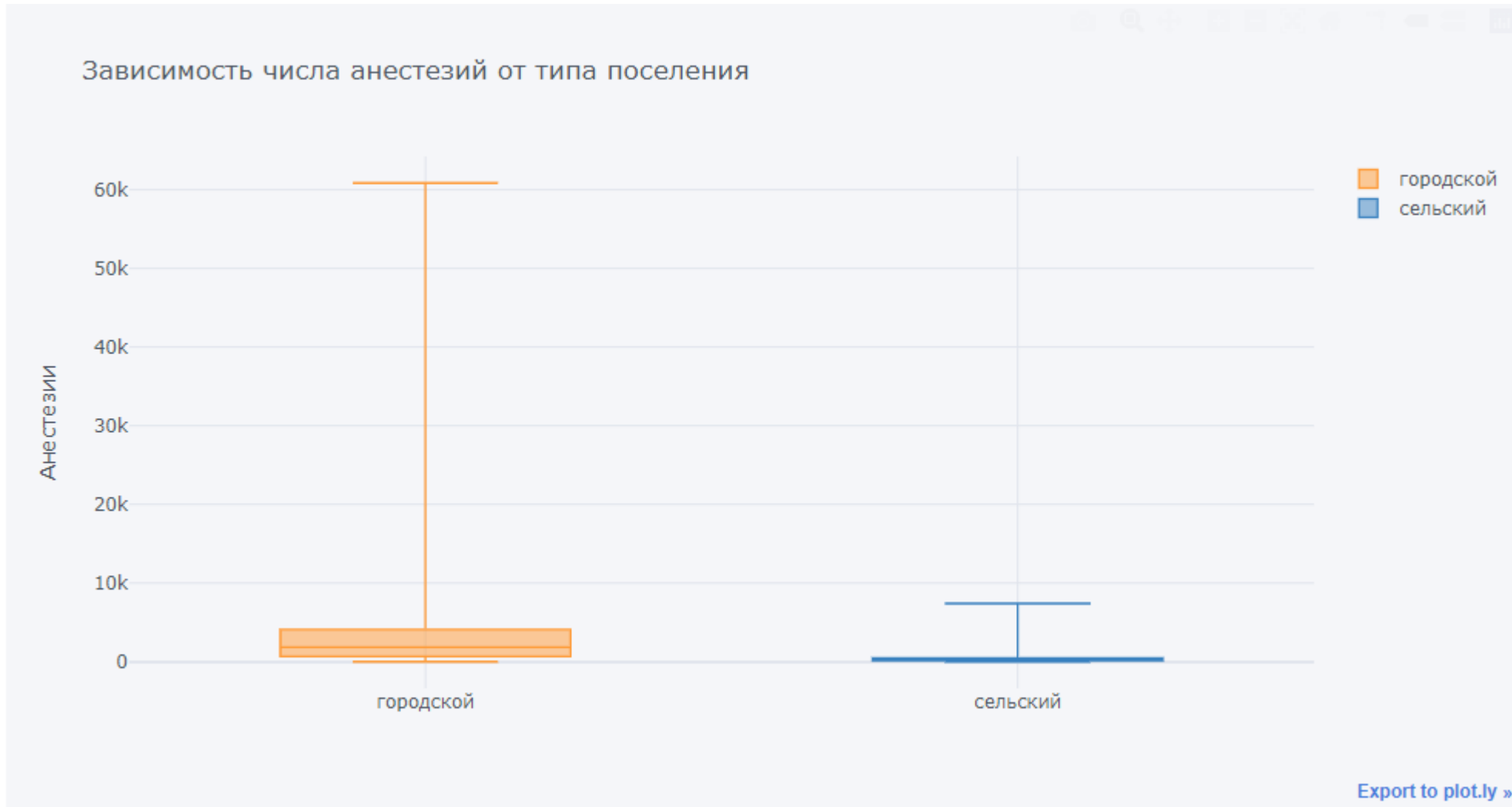
Федеральные округа

Москва значимо отличается от других регионов



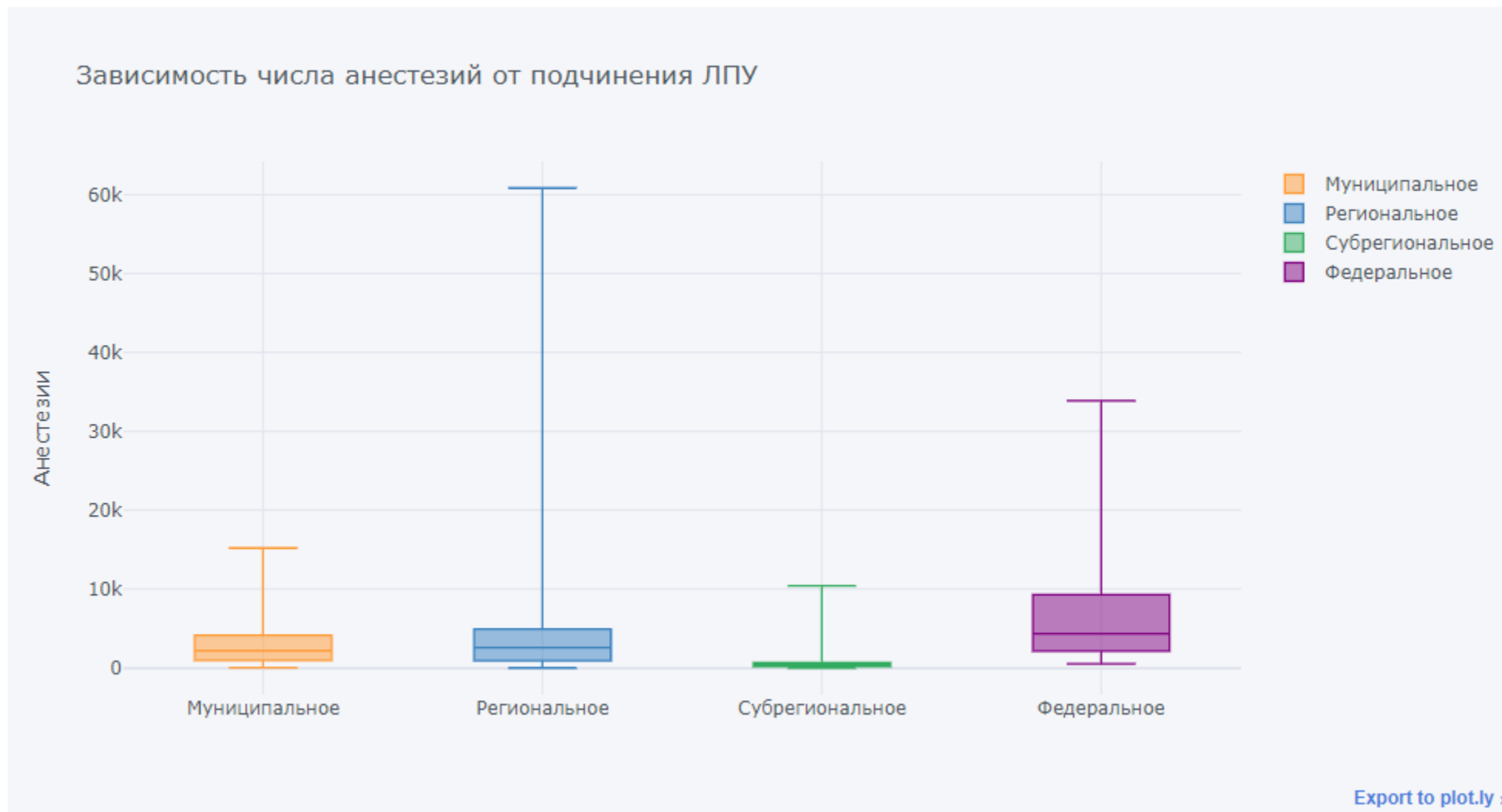
Тип поселения

Объемы анестезий в городских ЛПУ значительно выше, чем в сельских

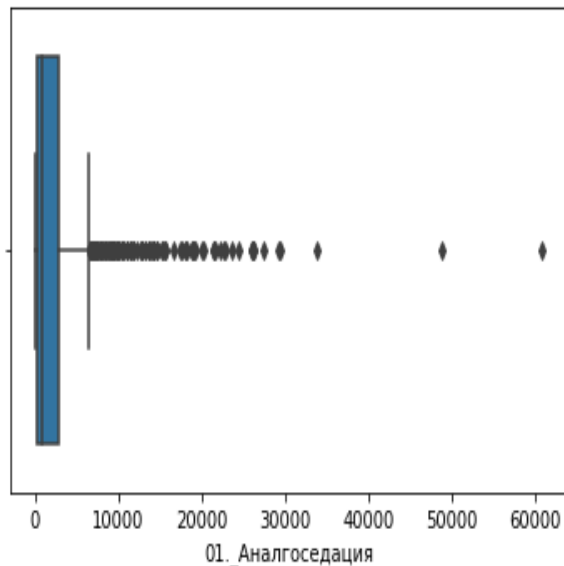
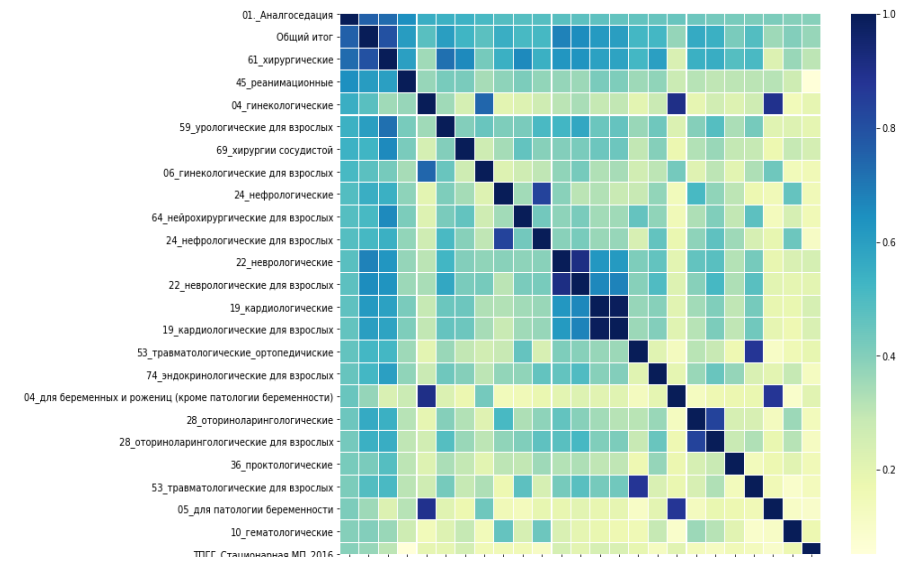
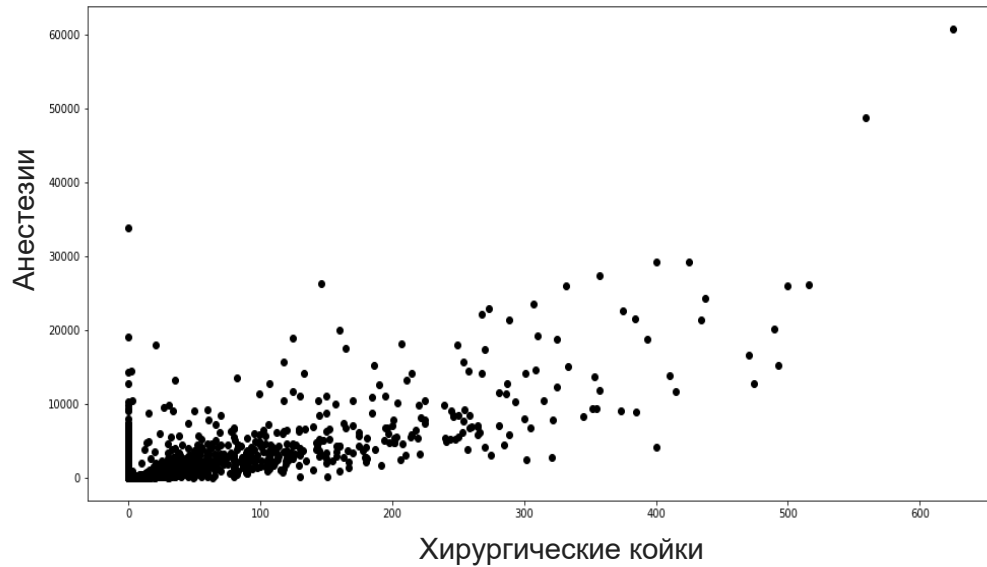


Тип подчинения

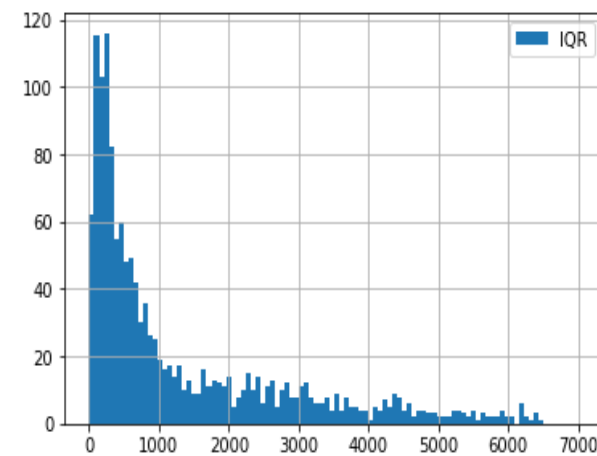
Финансирование федеральных ЛПУ значимо выше



Предварительная обработка данных



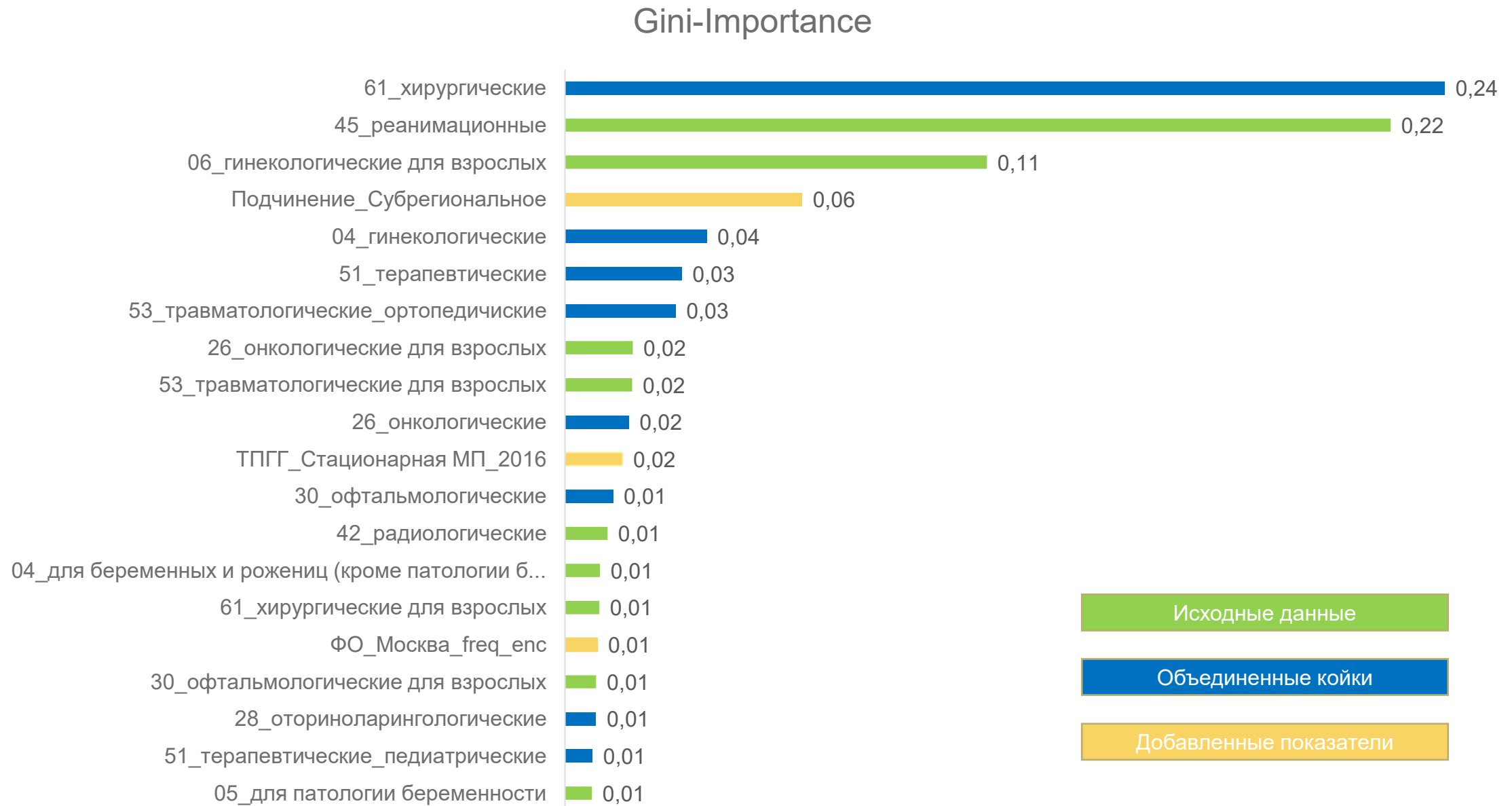
25-й перцентиль: 270.0, 75-й перцентиль: 2761.0, IQR: 2491.0, Границы выбросов: [-3466.5, 6497.5].



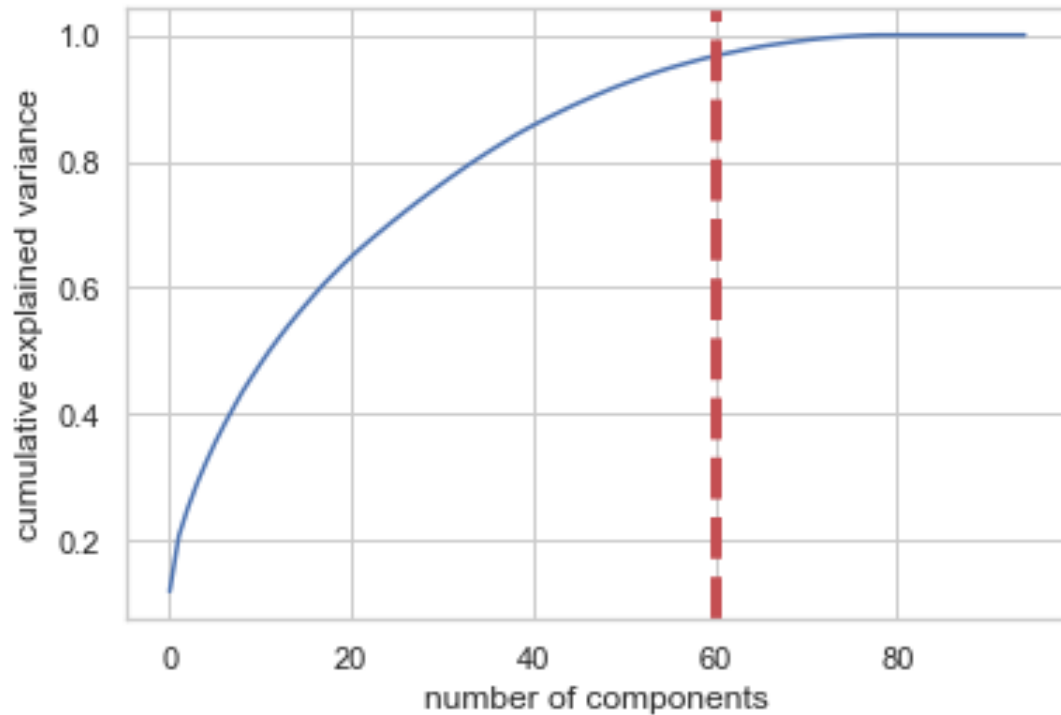
Модели

Модель	MAE_Train	MSE_Train	RMSE_Train	R2_Train	MAE_Test	MSE_Test	RMSE_Test	R2_Test
RandomForestRegressor	165,43	75 866,77	275,44	0,966	449,08	481 291,60	693,75	0,797
XGBRegressor	49,91	6 847,07	82,75	0,997	470,66	515 013,70	717,64	0,783
LassoCV	446,41	481 996,51	694,26	0,784	472,72	523 836,91	723,77	0,779
ElasticNetCV	477,90	520 281,46	721,31	0,767	508,98	591 471,07	769,07	0,750
RidgeCV	409,96	412 552,12	642,30	0,815	502,10	636 123,95	797,57	0,732
LinearRegression	409,90	412 394,24	642,18	0,815	504,12	644 793,73	802,99	0,728
KNeighborsRegressor	465,08	610 905,97	781,60	0,726	565,11	845 826,47	919,69	0,643
DecisionTreeRegressor	1,37	755,74	27,49	1,000	584,99	898 984,36	948,15	0,621

Random Forest (feature importances – Top20)



Метод главных компонент



'Baseline Random Forest r2 score'

0.797

'Baseline Random Forest With PCA r2 score'

0.712

'Hyperparameter Tuned Random Forest With PCA Reduced Dimensionality r2 score'

0.729

Optuna (RandomForestRegressor vs XGBRegressor)

- Подбор параметров с помощью Optuna позволил улучшить качество модели.
- Для модели градиентного бустинга улучшение получилось более значимое

Model	Best_params	R2 test
RandomForestRegressor	<code>{'max_features': 'auto', 'bootstrap': True, 'max_depth': 15, 'n_estimators': 100, 'min_samples_leaf': 1, 'min_samples_split': 2}</code>	0,811
xgb.XGBRegressor	<code>{'max_depth': 6, 'subsample': 0.38626904699216386, 'n_estimators': 1082, 'lambda': 0.001290443878149748, 'learning_rate': 0.013652804188290686, 'min_child_weight': 1, 'colsample_bytree': 0.3526960129865251, 'colsample_bynode': 0.8050966469634867, 'colsample_bylevel': 0.6449096845125648}</code>	0,836

Выводы

1. Добавленные дополнительные признаки, объемы ТПГГ, и разделение на ФО и Москву оказались значимыми
2. Лучший результат получился на основе модели градиентного бустинга, параметры подбирались с помощью Optuna
3. Полученная модель показала, что использование данных коечного фонда для оценки объема анестезий дает хорошую прогностическую точность $R^2=0,84$, и таким образом позволяет выделить ЛПУ, необходимые для посещения медицинскими представителями.
4. Данное решение показывает, что в случае отсутствия данных по тендерным закупкам ЛПУ, показатели статистической отчетности о деятельности медицинских учреждений можно использовать для сегментирования ЛПУ для посещения медицинскими представителями той или иной компании

СПАСИБО ЗА ВНИМАНИЕ!