

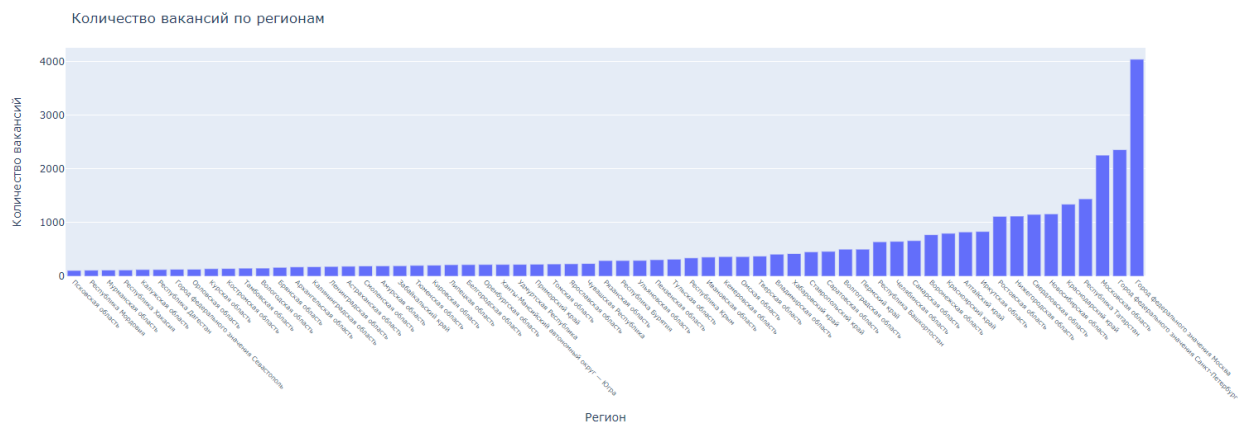


Влияние региона на зп

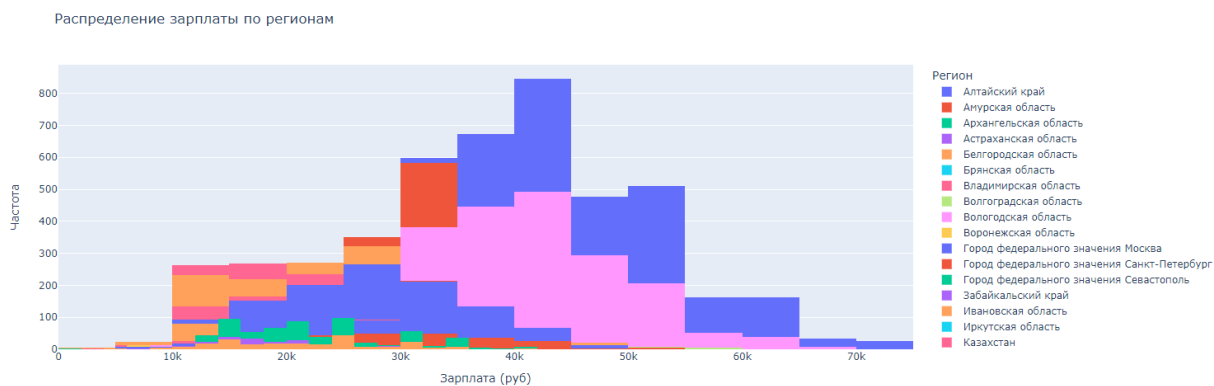
<https://colab.research.google.com/drive/1MkGHk-yn6LVfKbFMSdTBgO77PJRjIMQQ?usp=sharing> (вкладка # ВЛИЯНИЕ РЕГИОНА НА ДЕМО)

public.v3_vacancies_all_type_id_116_part_1_20230621 (оператор) смотрела вот этот файл, пока что без навыков, Антон сделает - добавлю, и как-то учту.

оставляю только те регионы, где кол-во вакансий хотя бы 100



распределение зп по регионам (после удаления выбросов)



Потом пробовала рассчитывать зп для всех регионов на основе базовых. Для получения коэффициентов обучила CatBoostRegressor, получилось пока что вот так:

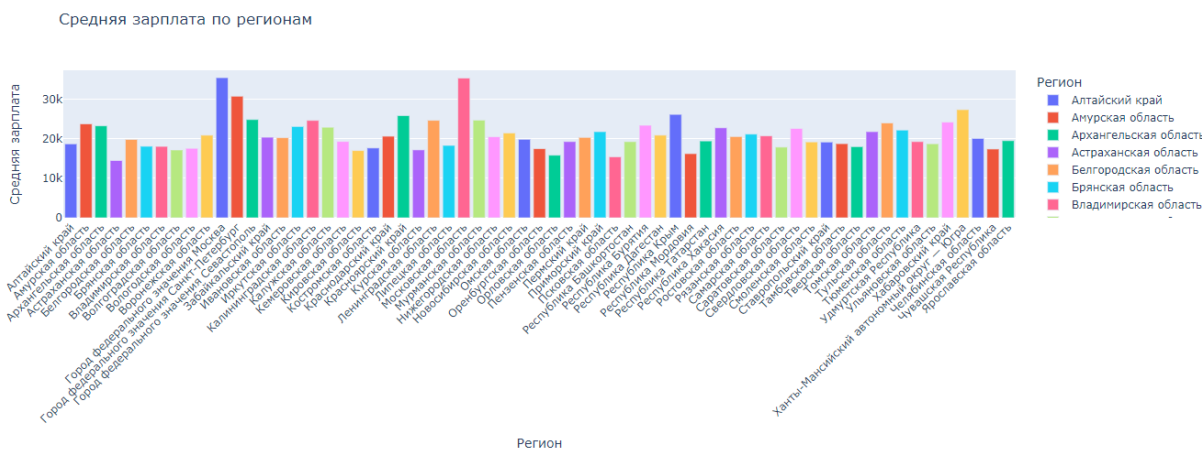
```
MAPE (train): 0.4044906595829629
MAPE (test): 0.7905586954385849
```

т.к. нам для каждой профессии надо делать свою модель, то надо будет подбирать параметры (мб потом это можно как-то перебором это сделать)

в итоге получили коэффициент значимости каждого региона (когда коэффициент в пределах от 0 до 1, то это норм, а когда вроде Москвы коэф равен 56, нужно либо другой подход выбирать, либо как-то пересчитывать).

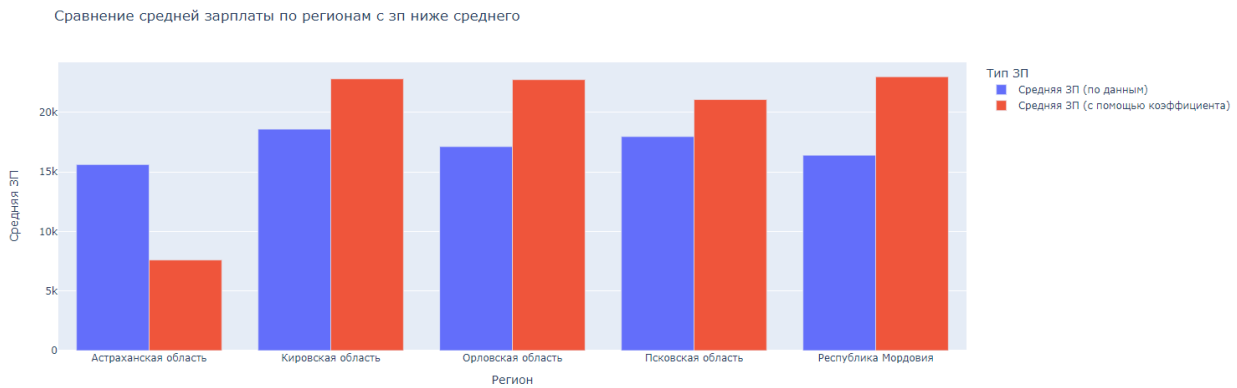
	Feature	Importance
10	encoder__region_name_Город федерального значен...	56.164990
28	encoder__region_name_Московская область	18.188335
11	encoder__region_name_Город федерального значен...	15.492070
70	remainder__experience_id	1.225207
29	encoder__region_name_Мурманская область	1.098452
16	encoder__region_name_Казахстан	0.887809
0	encoder__region_name_Алтайский край	0.779266
7	encoder__region_name_Волгоградская область	0.716686
3	encoder__region_name_Астраханская область	0.566189
66	remainder__is_shift	0.410934
63	encoder__region_name_Чувашская Республика	0.376013
12	encoder__region_name_Город федерального значен...	0.304426
50	encoder__region_name_Свердловская область	0.291776
30	encoder__region_name_Нижегородская область	0.281634
68	remainder__is_parttime	0.274292
35	encoder__region_name_Пензенская область	0.250583
36	encoder__region_name_Пермский край	0.245871
67	remainder__is_distance	0.221693
27	encoder__region_name_Минск	0.195732
56	encoder__region_name_Тульская область	0.194935
13	encoder__region_name_Забайкальский край	0.186321
23	encoder__region_name_Красноярский край	0.185751
39	encoder__region_name_Республика Башкортостан	0.167184
37	encoder__region_name_Приморский край	0.134187
6	encoder__region_name_Владимирская область	0.105737
64	encoder__region_name_Ярославская область	0.091354
52	encoder__region_name_Ставропольский край	0.064788
4	encoder__region_name_Белгородская область	0.063275

Т.к. у нас достаточно много регионов, а потом будет ещё больше, я решила выделить 3 группы: средние, ниже средней зп, выше средней зп (типо +- std от средней зп по всем регионам, вроде логично). Для подсчёта брала только вакансии без всего (типо опыт - 0, вахта - 0 и пр.). Когда Антон добавит признаки, наверное, имеет смысл брать без всех навыков или же вычитать их). В целом на графике видно, что среднее значение зп где-то 18-19 к.



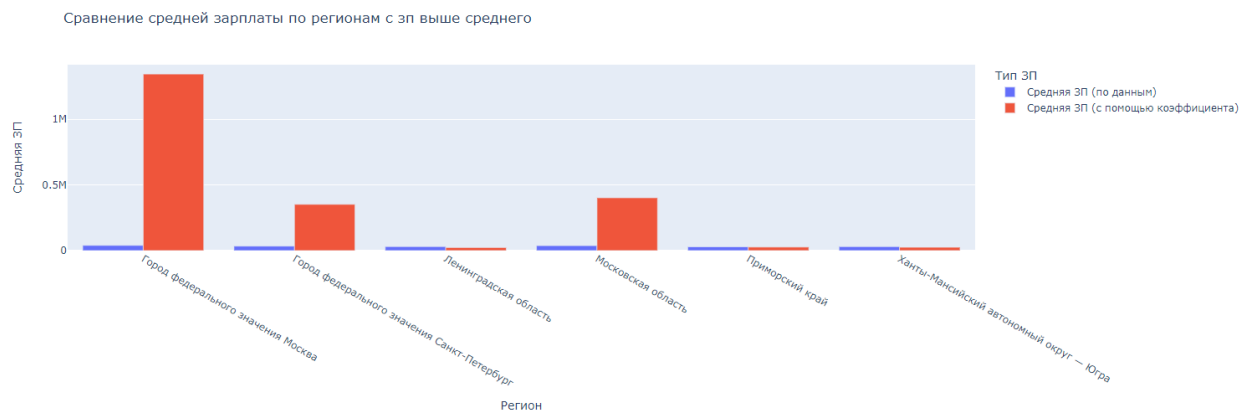
Для регионов ниже среднего

	Регион	Средняя ЗП (по данным)	Коэффициент	Средняя ЗП (с помощью коэффициента)
0	Астраханская область	15616.502762	0.671213	7601.983630
1	Кировская область	18586.278049	0.013460	22810.128358
2	Орловская область	17122.830769	0.016448	22741.038575
3	Псковская область	17964.672897	0.088710	21070.240294
4	Республика Мордовия	16398.477876	0.005834	22986.435276



Регионы выше среднего (там где коэф > 1 получился кринж, остальные вроде норм, см в файле)

	Регион	Средняя ЗП (по данным)	Коэффициент	Средняя ЗП (с помощью коэффициента)
0	Город федерального значения Москва	39417.598254	57.102490	1343407
1	Город федерального значения Санкт-Петербург	33701.664959	14.216246	351819
2	Ленинградская область	30353.043956	0.005414	23246
3	Московская область	37666.471218	16.378624	401816
4	Приморский край	28681.408889	0.155622	26719
5	Ханты-Мансийский автономный округ — Югра	30182.493151	0.096155	25344



Средние регионы

	Регион	Средняя ЗП (по данным)	Коэффициент	Средняя ЗП (с помощью коэффициента)
0	Алтайский край	19855.018248	0.999135	46222
1	Амурская область	25552.860825	0.042828	24111
2	Архангельская область	23586.321839	0.010113	23355
3	Белгородская область	21157.660465	0.171719	27091
4	Брянская область	19528.804878	0.097689	25380
5	Владимирская область	21390.114914	0.243623	28754
6	Волгоградская область	19290.280000	0.933733	44710
7	Вологодская область	21630.302632	0.023780	23671
8	Воронежская область	23211.391361	0.010372	23361
9	Город федерального значения Севастополь	27074.023256	0.314247	30387
10	Забайкальский край	21825.664948	0.150422	26599
11	Ивановская область	22417.273239	0.058190	24466
12	Иркутская область	24229.176116	0.022959	23652
13	Калининградская область	25963.301136	0.006942	23281
14	Калужская область	23554.824000	0.008735	23323
15	Кемеровская область	21494.799451	0.021070	23608
16	Костромская область	19747.604167	0.375166	31795
17	Краснодарский край	24354.422704	0.022047	23631
18	Красноярский край	26911.092994	0.154623	26696
19	Курская область	20348.525180	0.006938	23281

Сравнение средней зарплаты по регионам с средней зп

