



НАЦИОНАЛЬНЫЙ ИССЛЕДОВАТЕЛЬСКИЙ  
УНИВЕРСИТЕТ

Факультет компьютерных наук

# Исследование факторов, влияющих на долю МСП в регионах России

Попов Н.С., Фролов К.В., Ганзюк Н.А.

Москва, 2023



- Таргет - доля МСП в ВРП регионов
- Макроэкономические факторы: уровень безработицы, численность трудоспособного населения, отраслевая структура, объем субсидий, поручительств и тд
- Собрали из открытых источников: Росстат, КонсультантПлюс

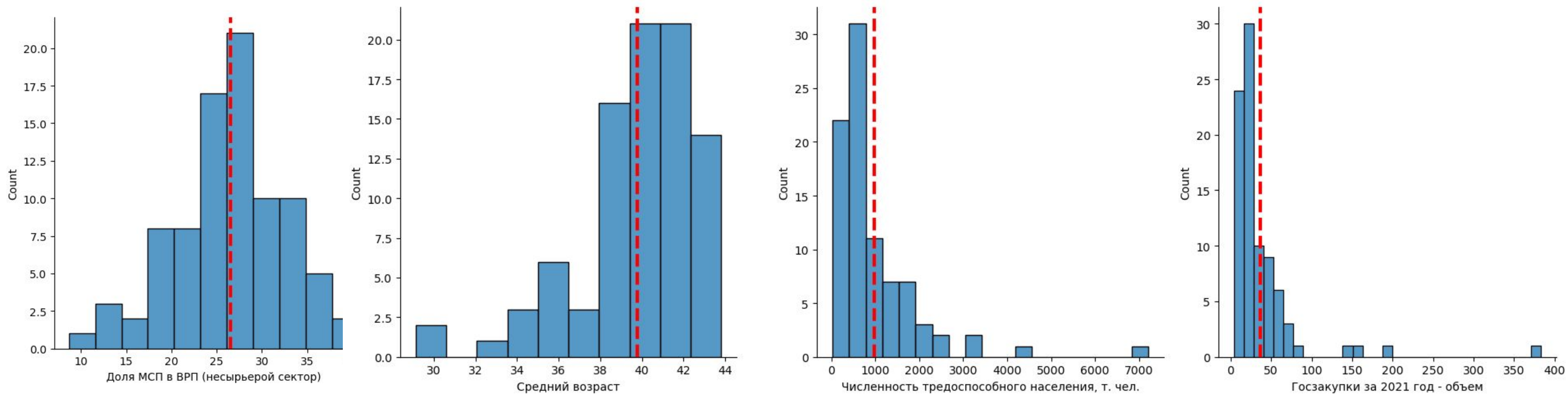


1. Доля МСП зависит от отраслевой структуры региона. Например, объем добычи полезных ископаемых, кол-во обрабатывающих производств отрицательно влияет на долю МСП в регионе
2. Средний возраст в регионе негативно влияет на долю МСП в регионе



# Статистические характеристики: гистограммы

Факультет компьютерных наук





# Статистические характеристики: корреляция таргета с каждым признаком

Факультет компьютерных наук







# Регрессия с помощью МНК

Факультет компьютерных наук

	coef	std err	t	P> t	[0.025	0.975]
const	21.2438	17.337	1.225	0.225	-13.448	55.936
Среднемесячная номинальная начисленная заработная плата работников по полному кругу организаций в целом по экономике	-6.695e-05	0.000	-0.213	0.832	-0.001	0.001
Объем субсидий_МЭР, выделенный из федерального бюджета бюджетам субъектов РФ, т.р.	5.009e-06	5.95e-06	0.842	0.403	-6.89e-06	1.69e-05
Уровень безработицы	-0.1655	0.220	-0.753	0.454	-0.605	0.274
Объем поддержки, оказанной Сбером - кол-во ИП/ООО	0.0029	0.005	0.629	0.532	-0.006	0.012
Место в рейтинге СЭП, (РИА рейтинг)	-0.0170	0.047	-0.363	0.718	-0.111	0.077
Госзакупки за 2021 год - кол-во	1.264e-05	7.57e-05	0.167	0.868	-0.000	0.000
Кол-во МСП	-3.891e-05	3.76e-05	-1.036	0.304	-0.000	3.62e-05
Введено в действие общей площади жилых домов на 1000 человек населения	-0.0008	0.002	-0.377	0.708	-0.005	0.004
Объем поддержки, оказанной Сбером - млн	0.0002	0.000	1.181	0.242	-0.000	0.001
Численность трудоспособного населения, т. чел.	0.0013	0.003	0.391	0.697	-0.005	0.008
Количество моногородов	-0.0809	0.142	-0.569	0.571	-0.365	0.203
Отраслевая структура: Обрабатывающие производства	-0.3274	0.092	-3.548	0.001	-0.512	-0.143
Кол-во центров Мой бизнес	-0.0258	0.075	-0.342	0.734	-0.177	0.125
Средний возраст	0.3341	0.382	0.874	0.386	-0.431	1.099
Среднемесячная начисленная заработная плата наёмных работников в организациях, у индивидуальных предпринимателей и физических лиц	-1.424e-05	0.000	-0.039	0.969	-0.001	0.001
Размер выданных поручительств и (или) независимых гарантий РГО (млн. руб)	-0.0004	0.002	-0.236	0.814	-0.004	0.003
Госзакупки за 2021 год - объем	-0.0378	0.045	-0.846	0.401	-0.127	0.052
Отраслевая структура: Добыча полезных ископаемых	-0.3094	0.076	-4.079	0.000	-0.461	-0.158
Отраслевая структура: Сельское хозяйство	-0.1769	0.127	-1.394	0.169	-0.431	0.077
Количество выданных поручительств и (или) независимых гарантий РГО (единиц)	0.0126	0.010	1.202	0.234	-0.008	0.034
Площадь объектов недвижимости, находящихся на территории индустриальных (промышленных) парков, кв. м	3.609e-07	1.44e-06	0.251	0.803	-2.52e-06	3.24e-06
наличие ОЗЗ	1.2077	1.051	1.149	0.255	-0.896	3.311
Главная отрасль: _Деятельность по операциям с недвижимым имуществом	6.5229	2.862	2.279	0.026	0.795	12.251
Главная отрасль: _Добыча полезных ископаемых	3.4121	3.352	1.018	0.313	-3.296	10.120
Главная отрасль: _Обрабатывающие производства	5.3597	2.810	1.907	0.061	-0.263	10.982
Главная отрасль: _Сельское, лесное хозяйство, охота, рыболовство и рыбоводство	6.3695	2.887	2.206	0.031	0.592	12.147
Главная отрасль: _Торговля оптовая и розничная; ремонт автотранспортных средств и мотоциклов	5.6467	3.510	1.609	0.113	-1.377	12.670



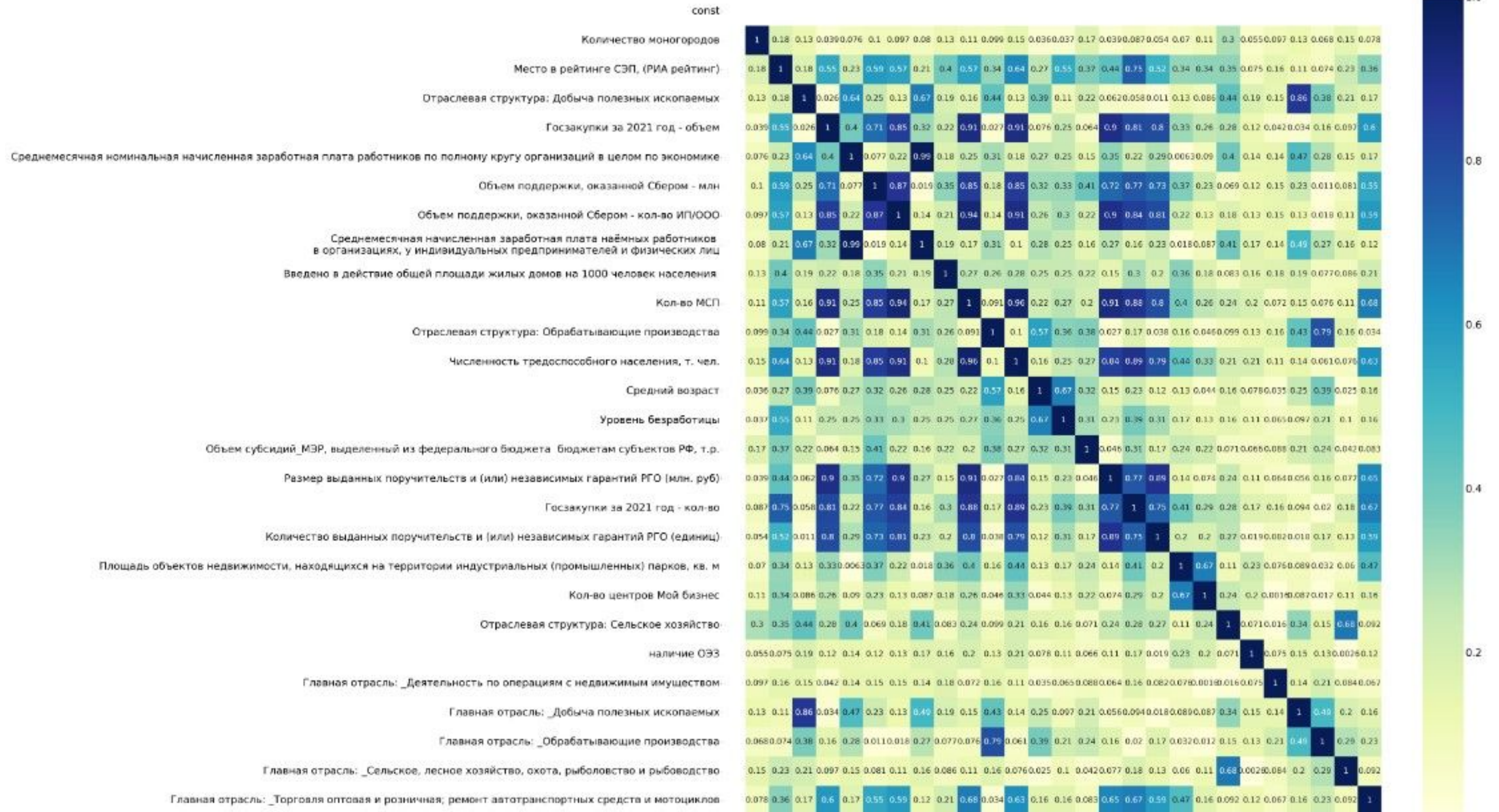


<b>Dep. Variable:</b>	Доля МСП в ВРП	<b>R-squared:</b>	0.835
<b>Model:</b>	OLS	<b>Adj. R-squared:</b>	0.759
<b>Method:</b>	Least Squares	<b>F-statistic:</b>	11.04
<b>Date:</b>	Mon, 04 Dec 2023	<b>Prob (F-statistic):</b>	1.68e-14
<b>Time:</b>	04:46:56	<b>Log-Likelihood:</b>	-227.17
<b>No. Observations:</b>	87	<b>AIC:</b>	510.3
<b>Df Residuals:</b>	59	<b>BIC:</b>	579.4
<b>Df Model:</b>	27		
<b>Covariance Type:</b>	nonrobust		

$$RMSE = 6.35065$$

$$MAE = 5.06101$$









## Признаки с высокими значениями корреляции ( $> 0.8$ )

- 'Госзакупки за 2021 год - кол-во'
- 'Кол-во МСП'
- 'Объем поддержки, оказанной Сбером - млн'
- 'Численность трудоспособного населения, т. чел.'
- 'Среднемесячная начисленная заработная плата наёмных работников \пв организациях, у индивидуальных предпринимателей и физических лиц'
- 'Размер выданных поручительств и (или) независимых гарантий РГО (млн. руб)'
- 'Госзакупки за 2021 год - объем'
- 'Количество выданных поручительств и (или) независимых гарантий РГО (единиц)'
- 'Главная отрасль: \_Добыча полезных ископаемых'

13 признаков с VIF>10

## Оценка VIF

Госзакупки за 2021 год - кол-во	Кол-во МСП	Объем поддержки, оказанной Сбером - млн	Численность трудоспособного населения, т. чел.	Среднемесячная начисленная заработная плата наёмных работников \пв организациях, у индивидуальных предпринимателей и физических лиц	Размер выданных поручительств и (или) независимых гарантий РГО (млн. руб)	Госзакупки за 2021 год - объем	Количество выданных поручительств и (или) независимых гарантий РГО (единиц)	Главная отрасль: _Добыча полезных ископаемых
11.015847	77.230505	10.447829	61.550028	238.331994	54.432764	25.458881	12.660222	11.541751



## Метрики качества в модели скоррелированных признаков

Dep. Variable:	Доля МСП в ВРП	R-squared:	0.795
Model:	OLS	Adj. R-squared:	0.741
Method:	Least Squares	F-statistic:	14.69
Date:	Mon, 04 Dec 2023	Prob (F-statistic):	7.46e-17
Time:	06:55:22	Log-Likelihood:	-236.45
No. Observations:	87	AIC:	510.9
Df Residuals:	68	BIC:	557.8
Df Model:	18		
Covariance Type:	nonrobust		





# Мультиколлинеарность

Факультет компьютерных наук

	coef	std err	t	P> t	[0.025	0.975]
const	26.9711	14.133	1.908	0.061	-1.231	55.173
Среднемесячная номинальная начисленная заработная плата работников по полному кругу организаций в целом по экономике	-0.0001	3.5e-05	-3.836	0.000	-0.000	-6.44e-05
Объем субсидий_МЭР, выделенный из федерального бюджета бюджетам субъектов РФ, т.р.	1.191e-05	5.21e-06	2.283	0.026	1.5e-06	2.23e-05
Уровень безработицы	-0.3334	0.213	-1.566	0.122	-0.758	0.092
Объем поддержки, оказанной Сбером - кол-во ИП/ООО	-0.0001	0.002	-0.076	0.940	-0.003	0.003
Место в рейтинге СЭП, (РИА рейтинг)	-0.0085	0.036	-0.238	0.813	-0.080	0.063
Введено в действие общей площади жилых домов на 1000 человек населения	-0.0006	0.002	-0.291	0.772	-0.005	0.004
Количество моногородов	-0.0548	0.138	-0.398	0.692	-0.329	0.220
Отраслевая структура: Обрабатывающие производства	-0.2970	0.082	-3.627	0.001	-0.460	-0.134
Кол-во центров Мой бизнес	0.0036	0.075	0.049	0.961	-0.145	0.153
Средний возраст	0.2798	0.323	0.867	0.389	-0.364	0.924
Отраслевая структура: Добыча полезных ископаемых	-0.2384	0.046	-5.146	0.000	-0.331	-0.146
Отраслевая структура: Сельское хозяйство	-0.0692	0.114	-0.606	0.547	-0.297	0.159
Площадь объектов недвижимости, находящихся на территории индустриальных (промышленных) парков, кв. м	-5.05e-07	1.15e-06	-0.439	0.662	-2.8e-06	1.79e-06
наличие ОЭЗ	1.1394	1.015	1.123	0.266	-0.886	3.165
Главная отрасль: _Деятельность по операциям с недвижимым имуществом	3.0381	2.442	1.244	0.218	-1.836	7.912
Главная отрасль: _Обрабатывающие производства	2.9525	1.981	1.490	0.141	-1.001	6.906
Главная отрасль: _Сельское, лесное хозяйство, охота, рыболовство и рыбоводство	3.1508	2.273	1.386	0.170	-1.385	7.687
Главная отрасль: _Торговля оптовая и розничная; ремонт автотранспортных средств и мотоциклов	4.1853	2.932	1.427	0.158	-1.666	10.036





	lm	lm_pvalue	fvalue	f_pvalue	feature
0	0.387966	0.533371	0.380745	0.538854	Размер выданных поручительств и (или) независи...
0	3.252938	0.071296	3.301605	0.072737	Объем поддержки, оказанной Сбером - млн
0	0.583203	0.445059	0.573642	0.450910	Площадь объектов недвижимости, находящихся на ...
0	1.802773	0.179378	1.798599	0.183455	Численность трудоспособного населения, т. чел.
0	1.073031	0.300262	1.061455	0.305806	Количество выданных поручительств и (или) неза...
0	0.961303	0.326859	0.949698	0.332561	Объем субсидий_МЭР, выделенный из федерального...
0	1.385278	0.239204	1.375332	0.244174	Госзакупки за 2021 год - объем
0	1.392743	0.237942	1.382863	0.242897	Госзакупки за 2021 год - кол-во
0	0.988422	0.320128	0.976797	0.325796	Отраслевая структура: Сельское хозяйство
0	0.363560	0.546536	0.356693	0.551936	Кол-во центров Мой бизнес
0	1.922348	0.165598	1.920594	0.169416	Уровень безработицы
0	0.303426	0.581743	0.297488	0.586890	Среднемесячная номинальная начисленная заработ...
0	0.132535	0.715818	0.129686	0.719651	Отраслевая структура: Обрабатывающие производства
0	0.334633	0.562944	0.328203	0.568232	Среднемесячная начисленная заработная плата на...

Нулевая гипотеза о **гомоскедастичности** отвергается

На **автокорреляцию** проверять наши данные не имеет смысла, т.к. в них отсутствует временная структура





'Количество выданных поручительств и (или) независимых гарантий РГО (единиц)',

'Отраслевая структура: Сельское хозяйство',

'Объем поддержки, оказанной Сбером - кол-во ИП/ООО',

'Отраслевая структура: Обрабатывающие производства',

'Госзакупки за 2021 год - объем',

'Количество моногородов'



## Логарифмируем

<b>Dep. Variable:</b>	Доля МСП в ВРП	<b>R-squared:</b>	0.803
<b>Model:</b>	OLS	<b>Adj. R-squared:</b>	0.713
<b>Method:</b>	Least Squares	<b>F-statistic:</b>	8.926
<b>Date:</b>	Mon, 04 Dec 2023	<b>Prob (F-statistic):</b>	1.79e-12
<b>Time:</b>	13:59:23	<b>Log-Likelihood:</b>	-234.73
<b>No. Observations:</b>	87	<b>AIC:</b>	525.5
<b>Df Residuals:</b>	59	<b>BIC:</b>	594.5
<b>Df Model:</b>	27		
<b>Covariance Type:</b>	nonrobust		





# Тестовые прогнозы

```
X_train, X_test, y_train, y_test = train_test_split(to_log, Y, test_size=0.3, random_state=42)
```

```
pred = ksSplitted_res.predict(X_test)
mean_squared_error(y_test, pred, squared=False)
```

6.350655804049782

```
mean_absolute_error(y_test, pred)
```

5.061018160047688

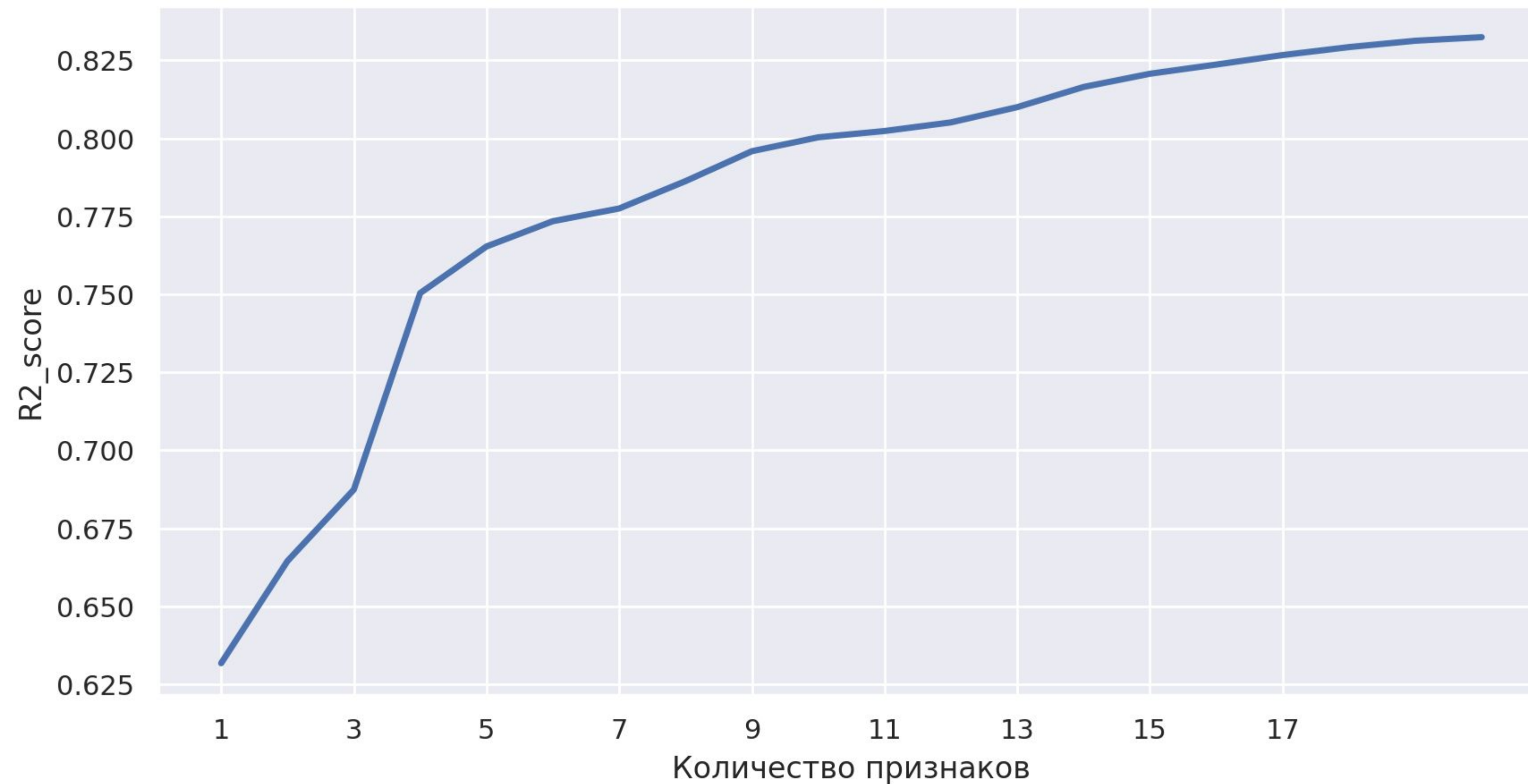
```
pred = ksSplitted_res.get_prediction(sm.add_constant(X_test)).summary_frame()
pred['target'] = y_test
pred
```

	mean	mean_se	mean_ci_lower	mean_ci_upper	obs_ci_lower	obs_ci_upper	target
78	19.667702	5.228295	9.018014	30.317390	6.472180	32.863224	12.487504
2	23.521170	2.577051	18.271890	28.770451	14.126444	32.915897	40.168474
28	21.289722	3.810211	13.528577	29.050867	10.292390	32.287055	28.075798
24	24.661065	2.151032	20.279556	29.042574	15.722185	33.599944	28.710987
14	25.024047	5.532951	13.753795	36.294299	11.322787	38.725307	32.093048
69	26.741034	2.191149	22.277810	31.204258	17.761818	35.720250	17.846206
12	25.771478	3.621306	18.395120	33.147836	15.042237	36.500719	32.596827
20	30.235685	4.516047	21.036799	39.434571	18.180585	42.290786	29.747166
6	27.045730	3.316927	20.289370	33.802089	16.732905	37.358554	33.963277
70	19.521867	2.244586	14.949795	24.093939	10.488053	28.555680	17.492357
86	4.388260	4.450896	-4.677918	13.454438	-7.565883	16.342403	5.442218
67	27.959795	2.550532	22.764531	33.155058	18.595143	37.324447	18.144793
55	23.343450	4.910660	13.340764	33.346137	10.664346	36.022555	21.808965
82	8.568885	3.719618	0.992272	16.145499	-2.299004	19.436775	10.193762
85	13.393884	8.033299	-2.969411	29.757179	-4.729675	31.517442	6.373876
66	16.002256	4.214998	7.416587	24.587925	4.408307	27.596206	18.290082
35	27.968842	2.363449	23.154654	32.783029	18.810111	37.127572	26.863559
81	21.507467	3.418204	14.544813	28.470122	11.058328	31.956606	10.478529
57	23.518152	1.851103	19.747580	27.288725	14.862335	32.173969	21.114727
75	11.965832	3.240769	5.364602	18.567062	1.753967	22.177696	14.591632
13	25.749775	2.453655	20.751843	30.747707	16.493142	35.006409	32.469115
32	17.708522	3.307185	10.972006	24.445038	7.408687	28.008357	27.136746
42	27.322363	3.657851	19.871563	34.773163	16.541808	38.102918	25.004963
30	26.382472	4.477515	17.262073	35.502871	14.387156	38.377789	27.196735
11	28.293496	1.847502	24.530258	32.056734	19.640872	36.946120	32.803372
68	21.793558	2.665684	16.363737	27.223379	12.296774	31.290342	18.046877





R2-score при разном количестве лучших признаков



## ТОП 5:

- 'Отраслевая структура: Добыча полезных ископаемых',
- 'Среднемесячная начисленная заработная плата наёмных работников в организациях, у индивидуальных предпринимателей и физических лиц',
- 'Отраслевая структура: Обрабатывающие производства',
- 'Уровень безработицы',
- 'Объем субсидий\_МЭР, выделенный из федерального бюджета бюджетам субъектов РФ, т.р.'



Модель неплохо объясняет долю МСП в ВПР. На таргет главным образом влияет отраслевая структура региона. Для специализирующихся на обрабатывающей промышленности/добыче ископаемых доля МСП значительно меньше. Средний возраст в регионе оказался незначимой переменной.





НАЦИОНАЛЬНЫЙ ИССЛЕДОВАТЕЛЬСКИЙ  
УНИВЕРСИТЕТ