

Федеральное государственное автономное образовательное учреждение высшего образования
**“НАЦИОНАЛЬНЫЙ ИССЛЕДОВАТЕЛЬСКИЙ УНИВЕРСИТЕТ
“ВЫСШАЯ ШКОЛА ЭКОНОМИКИ”**
Факультет экономических наук
Образовательная программа “Экономика”

Проект по Эконометрике-2

Исследование зависимости заработной платы
в России в 2022 году от различных признаков

Выполнили студенты группы БЭК2110:
Некрасова Мария
Серочкин Егор

Преподаватель:
Станкевич Иван Павлович

Москва 2024г.

1. Введение

В данном проекте мы будем исследовать зависимость заработной платы в России от различных признаков.

Актуальность исследования заключается в том, что российский рынок труда в последнее время отличается нестабильностью с точки зрения распределения зарплат и низким уровнем безработицы. Более того, имеют место множество трендов со стороны работников, которые сильно влияют на показатели рынка труда. Например, сейчас в России явно дает о себе знать смещение интереса в сторону сферы ИТ и инфобизнеса. На ситуации шока Российская экономика обычно реагирует колебанием цен, в том числе и зарплат, нежели изменением уровня безработицы. Именно поэтому мы считаем, что необходимо подробно исследовать вопрос моделирования заработной платы, так как это внесет ясность о том, какие факторы в наибольшей степени влияют на ее колебания.

Российский рынок труда характеризуется нестандартным поведением во времена экономических шоков: при снижении деловой активности работники соглашаются сохранять свои рабочие места за значительное снижение заработной платы или за счет уменьшения рабочих часов. Беря во внимание все вышеперечисленное, становится очевидно, что тематика российских заработных плат является одной из наиболее интересных и не описываемых стандартными экономическими теориями.

2. Обзор литературы

Исследование, описанное в статье [1] показывает, что заработная плата в России имеет нестандартную траекторию динамики роста и падения на протяжении трудового опыта работников. Например, своего пика она достигает на ранних этапах развития карьеры и затем стремительно падает. При стаже работы 35-39 лет доходы падают даже ниже начального уровня. Также исследование показывает, что сейчас наибольшую зарплату получают люди в возрасте 34-38 лет. Однако нельзя утверждать, что данная работа в точности характеризует реальную ситуацию на рынке труда, так как для анализа использовались данные 2000-2019 года. Мы считаем, что с тех пор ситуация на рынке труда значительно изменилась из-за многих факторов: изменился спектр действующих компаний в России, сместился фокус на ИТ-сферу как у работодателей, так и у работников, вырос приток трудовых мигрантов и другое. Более того, выборка включает в себя только мужчин 20-60 лет, имеющих стаж работы до 40 лет, причем исключая самозанятых. В своем исследовании мы планируем рассматривать данные по более широкой выборке респондентов.

Несомненно, еще одним важным аспектом на рынке труда является уровень образования работников. В нашем исследовании этот показатель разбит на множество когорт:

- 0 классов школы
- 5 классов школы
- 7 классов школы
- 8 классов школы
- 9 классов школы
- 7-9 классов школы (незак. средн) и менее 2 лет в техникуме
- 7-9 классов школы (незак. средн) + ПТУ без диплома
- 7-9 классов школы (незак. средн) + ПТУ с дипломом
- 10+ классов школы без аттестата
- есть аттестат о среднем образовании (11 классов)
- 10+ классов школы и техникум без диплома
- 10+ классов школы и какое-либо професс. обр. без диплома
- 10+ классов школы и какое-либо професс. обр. с дипломом

- техникум с дипломом
- 1-2 года в ВУЗе
- 3+ лет в ВУЗе
- есть диплом о высшем образовании
- аспирантура и т.п. с дипломом
- аспирантура и т.п. без диплома.

В статье [2] обсуждается влияние грамотности на трудоустроенность и заработок. Вероятнее всего, эту самую грамотность можно оценить через уровень образования работника. Основным результатом данной статьи является вывод о том, что навыки эффективной обработки информации имеют огромное влияние на трудоустройство.

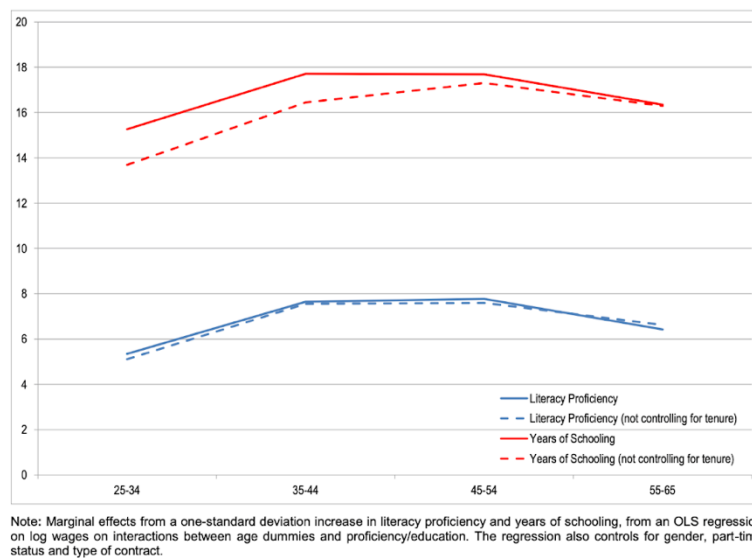


Рисунок 1. Распределение грамотности и лет образования по возрастам.

Источник: Survey of Adult Skills (PIAAC) (2012),
www.oecd.org/site/piaac/publicdataandanalysis.htm

Кроме того, в исследовании показано, что более высокий уровень грамотности способствует более позднему выходу на пенсию. Таким образом, у работников появляется шанс на более долгосрочном горизонте получать опыт работы и, вероятно, требовать более высокую зарплату.

Еще одним интересующим нас феноменом стала взаимосвязь курения и размера заработной платы. По этому поводу в статье [3] авторы, исследующие отличия в образовании, занятости и зарплате между курильщиками и некурящими в Китае, пришли к выводу о том, что курение напрямую не значимо влияет на зарплату. Однако исследование показало, что в сельской среде социально-экономический статус ниже, чем в городах, и тенденция к курению там (в поселках) выше. Также была обнаружена взаимосвязь между курением и занимаемой должностью работника. В статье говорится о том, что высокий заработок и важная позиция в компании имеют положительную взаимосвязь с курением среди мужчин. Кроме того, у людей без образования, но умеющих читать и писать, вероятность наличия зависимости от курения намного выше. У тех же, кто имел образование выше школьного, тяга к курению значительно уменьшалась вне зависимости от пола или места проживания.

Также мы хотим проверить влияние количества детей у индивидов на их заработную плату. В статье [4], где автор также исследует данные РМЭЗ 18-22 волн, показано, что наличие ребенка до 3-х лет у работающих женщин уменьшает их зарплату на 11,8% на 10% уровне значимости. При этом наличие ребенка от 3 до 7 лет никак не влияет на зарплату. “Для модели

со структурным уравнением самоотбора наличие одного ребенка значимо отрицательно влияет на заработки женщин (на 10%-ном уровне значимости для уравнения без учета самоотбора и на 5%-ном для уравнения с его учетом),” – говорится в статье. Под самоотбором автор подразумевает сознательное решение женщины выходить на рынок труда, что приводит к эндогенности, так как только для них мы можем наблюдать заработную плату.

В целом посмотреть, какую форму имеют эконометрические модели для предсказания зарплаты, нам позволила статья [5]. Несмотря на то, что описанная там модель включает не так много переменных, как ожидается в нашей, результаты все же будут полезны для дальнейшей работы. Автор использует такие переменные как:

- продолжительность обучения (лет)
- продолжительность рабочего стажа (лет)
- пол
- этничность (1 - белый, 0 - иначе)
- полный график (1 - если да, 0 - иначе)
- замужество (1 - в браке, 0 - иначе)
- локация (1 - Лондон, 0 - иначе)
- вера (1 - христианство, 0 - иначе)
- национальность (1 - британец, 0 - иначе)
- финансы (1 - работает в сфере бизнес/финансах/computer science/R&D, 0 - иначе)

Несомненно, многие формулировки переменных вызывают вопросы: например, локация и национальность (брались данные по индивидам в Великобритании за 2007 год).

3. Данные

Наш проект основывается на данных РМЭЗ НИУ ВШЭ. Мы будем использовать полную выборку данных по индивидам за 2022 год (волна 31). Собрав необходимые переменные и очистив их от пропусков, мы получили около 4500 наблюдений.

В качестве переменных мы выбрали следующие данные:

- кол-во жителей в населенном пункте;
- возраст;
- количество рабочих часов в день;
- стаж работы;
- курение (1 - курит, 0 - иначе)
- гос. компания или частная
- тип трудоустройства;
- вредность производства;
- профессия;
- сфера деятельности;
- пол;
- образование;
- квалификация;
- семейное положение

Среди респондентов оказалось 2435 женщин и 2078 мужчин в возрасте от 16 до 83 лет. В приложении мы представили [графики](#) по нашим данным и [расшифровку переменных](#).

В нашей модели задействованы следующие эндогенные переменные:

- квалификация (высшая, средняя, низшая)
- образование
- семейное положение
- курение

Причины эндогенности образования в моделировании зарплаты очевидны: мы упускаем переменную, описывающую талант индивида. В качестве инструмента для образования мы выбрали переменную типа школы, где учился.

Квалификацию мы рассматриваем в трех категориях: высшая, средняя и низкая (неквалифицированный специалист). Инструментом для высшей квалификации является пол индивида, для средней и низшей — ответ на вопрос: Вы испытываете удовлетворенность жизнью?: «и да, и нет» и «нет» соответственно. Мы выбрали такой инструмент, так как более низкая квалификация работника говорит о том, что он скорее всего имеет более низкую оплату и более плохие условия труда, что в свою очередь взаимосвязано с его жизнеощущением.

Семейное положение является эндогенной переменной, так как люди в браке зарабатывают больше, чтобы обеспечивать семью. При этом индивиды, которые получают большую зарплату и не состоят в браке, более склонны к тому, чтобы завести семью, чем те, кто зарабатывает мало. Инструментом для семейного положения мы хотели бы выбрать статус отношений родителей, однако за неимением этих данных мы выбрали наличие детей: у кого есть дети, тот и в браке (примерно).

Курение также является эндогенной переменной по той причине, что курят в основном те, кто либо зарабатывает мало, либо люди на высоких должностях, которые много нервничают. Инструментом для курения мы выбрали вредность производства, на котором работает индивид. По нашей логике, люди склонны к курению в более вредных для здоровья сферах труда.

4. Гипотезы

В ходе изучения данных, различных объясняющих, а также множества эконометрических статей мы решили проверить несколько скучных и интересных гипотез для модели, предсказывающей зарплату, а именно:

- В больших городах (в городах с большей численностью) больше возможностей для самореализации человеческого потенциала, поэтому коэффициент при переменной численности населения должен быть положительным.
- Более квалифицированные специалисты обладают большими знаниями и навыками, поэтому получают зарплату больше (коэффициент при образовании положительный).
- Согласно найденным статьям, сам факт курения положительно влияет на доход индивида, поэтому коэффициент при курении должен быть положительным.

5. Спецификация

Для построения моделей далее будет использоваться язык программирования Python и статистические библиотеки statsmodels и linearmodels.

Для начала нам необходимо оценить регрессию с исходными переменными, допуская отсутствие эндогенности во всех выбранных признаках. МНК-модель показала следующие результаты (Таб. 1).

Model:	OLS	Adj. R-squared:	0.469
Dependent Variable:	sal	AIC:	7058.2728
Date:	2024-05-11 15:42	BIC:	7110.8532
No. Observations:	316	Log-Likelihood:	-3515.1
Df Model:	13	F-statistic:	22.42
Df Residuals:	302	Prob (F-statistic):	3.87e-37
R-squared:	0.491	Scale:	2.8136e+08

	Coef.	Std.Err.	t	P> t	[0.025	0.975]
const	33328.6226	7931.3346	4.2021	0.0000	17720.9439	48936.3012
popul	0.0187	0.0227	0.8253	0.4098	-0.0259	0.0633
popul_ln	-333.3458	598.7702	-0.5567	0.5781	-1511.6358	844.9443
popul_sq	-0.0000	0.0000	-0.4534	0.6506	-0.0000	0.0000
w_years	1374.3306	357.3951	3.8454	0.0001	671.0307	2077.6306
w_years_sq	-17.7358	6.0813	-2.9165	0.0038	-29.7028	-5.7688
smoke	6093.3545	2667.2291	2.2845	0.0230	844.6471	11342.0620
occup_ВЫСШАЯ_КВАЛИФИКАЦИЯ	14351.0936	3836.4943	3.7407	0.0002	6801.4475	21900.7397
occup_КВАЛ_РАБОЧИЕ	21245.8373	3173.2999	6.6952	0.0000	15001.2583	27490.4163
occup_НЕКВАЛ_РАБОЧИЕ	5419.1982	5366.7066	1.0098	0.3134	-5141.6765	15980.0729
educ_ВЫСШЕЕ	5402.7542	2737.5161	1.9736	0.0493	15.7326	10789.7757
marriage_cocmoum	-107.0565	1987.6616	-0.0539	0.9571	-4018.4768	3804.3638
age	-798.6536	120.0291	-6.6538	0.0000	-1034.8529	-562.4543
ill	50.3393	128.9734	0.3903	0.6966	-203.4611	304.1397

Omnibus:	171.892	Durbin-Watson:	1.738
Prob(Omnibus):	0.000	Jarque-Bera (JB):	1232.508
Skew:	2.168	Prob(JB):	0.000
Kurtosis:	11.649	Condition No.:	5846562826171

Notes: [1] Standard Errors assume that the covariance matrix of the errors is correctly specified. [2] The condition number is large, 5.39e+12. This might indicate that there are strong multicollinearity or other numerical problems.

Таб. 1

Обратим внимание на второй пункт заметки, который говорит о большом количестве числе условий (Cond. No.), что может свидетельствовать о строгой мультиколлинеарности или других числовых проблемах. Очевидно, что строгую мультиколлинеарность можно исключить в силу получения результатов модели (при ее наличии модель не оценилась бы). Для получения данных о нестрогой мультиколлинеарности мы прибегли к использованию VIF-коэффициента. Результаты получились грустными (Таб. 2)

Очевидно, что в данных наблюдается коллинеарность, так как мы можем наблюдать $VIF > 10$ опыта работы (w_years) и его квадрата.

Примечательно, что при использовании не абсолютного значения непрерывных переменных, а их натуральных логарифмов заметка о числовых проблемах исчезает, оставляя свидетельство о мультиколлинеарности ($VIF > 10$) без изменений. В силу этого, было решено продолжить использование абсолютных значений для простоты интерпретации, а для решения проблемы мультиколлинеарности не принимать никаких дальнейших действий.

Variable	VIF
const	59,452859
popul_ln	2,1269734
popul_sq	2,2021473
w_years	22,815419
w_years_sq	19,326748
smoke	1,2586627
occup_ВЫСШАЯ_КВАЛИФИКАЦИЯ	1,4810066
occup_КВАЛ_РАБОЧИЕ	1,3003416
occup_НЕКВАЛ_РАБОЧИЕ	1,0860908
educ_ВЫСШЕЕ	1,6062926
marriage_состоит	1,0936071
age	3,8335922
ill	1,0574196

Таб. 2

Обратимся к регрессорам. Во-первых, в дальнейшем будет использоваться уровень значимости $\alpha = 0.05$. Во-вторых, предполагая отсутствие эндогенности, проанализируем результат МНК-модели.

Коэффициент при численности населения не является статистически значимым ($p\text{-value} > \alpha$), поэтому гипотеза №1 о положительном влиянии численности населения на заработную плату отвергается. Коэффициент при неквалифицированных сотрудниках не является статистически значимым, однако наличие квалификации является статистически значимым, при этом просто квалификация делает больший вклад в заработную плату, чем высшая квалификация на 6894,7437 рублей. Поэтому гипотеза №2 о росте заработной платы при повышении квалификации отвергается.

Коэффициент при факте курения является статистически значимым ($p\text{-value} < \alpha$) и положительным, что подтверждает гипотезу 3, то есть курящие люди получают большую заработную плату.

Как было сказано выше, для решения проблемы эндогенности для признаков факта курения, квалификации, получения высшего образования и факта замужества будут использованы инструменты: удовлетворенность жизнью, пол, факт обучения в школе с углубленным изучением какого-либо предмета, количество детей и вредность производства. В такой постановке задачи число инструментов оказывается больше, чем число эндогенных признаков, однако стоит взять во внимание, что квалификация является бинарным признаком, который распадется в 3 признака (высшая квалификация, квалифицированный работник и без квалификации), что приводит к тому, что число инструментов оказывается равным числу эндогенных признаков, что нарушает условие для проверки инструментов на валидность (усл.: число инструментов $>$ число эндогенных признаков) и лишает возможности проверки инструментов на это свойство. Однако мы можем проверить инструменты на релевантность, модель должна иметь F-статистику > 10 .

В качестве модели для борьбы с эндогенностью был выбран двухшаговый МНК (2SLS) (Таб. 3)

Сразу обратим внимание на F-статистику, которая получилась больше 15, что говорит о том, что инструменты релевантны. Для 2МНК вышло меньшее количество условий,

однако следует также взять во внимание довольно малый R^2 , что говорит о том, что модель описывает данные чуть лучше, чем средняя от таргета.

Model:	IV2SLS	R-squared:	0.049
Dependent Variable:	sal	Adj. R-squared:	0.008
Date:	2024-05-11 15:42	F-statistic:	13.92
No. Observations:	316	Prob (F-statistic):	2.93e-24
Df Model:	13	Scale:	5.2606e+08
Df Residuals:	302		

	Coef.	Std.Err.	t	P> t	[0.025	0.975]
const	-1340.4239	26992.3525	-0.0497	0.9604	-54457.3309	51776.4831
popul	0.0162	0.0919	0.1764	0.8601	-0.1646	0.1971
popul_ln	-1124.9146	1800.9766	-0.6246	0.5327	-4668.9668	2419.1376
popul_sq	-0.0000	0.0000	-0.1392	0.8894	-0.0000	0.0000
w_years	574.6506	1010.2583	0.5688	0.5699	-1413.3864	2562.6876
w_years_sq	-10.4041	19.2520	-0.5404	0.5893	-48.2891	27.4809
smoke	10400.9679	29336.7343	0.3545	0.7232	-47329.3313	68131.2672
occup_ВЫСШАЯ_КВАЛИФИКАЦИЯ	35801.2873	66746.1534	0.5364	0.5921	-95545.1460	167147.7205
occup_КВАЛ_РАБОЧИЕ	44307.0970	37004.9087	1.1973	0.2321	-28513.0209	117127.2149
occup_НЕКВАЛ_РАБОЧИЕ	62458.4524	58510.2061	1.0675	0.2866	-52680.8700	177597.7749
educ_ВЫСШЕЕ	26102.8119	51077.0470	0.5110	0.6097	-74409.1667	126614.7906
marriage_состоит	8666.6286	14507.7255	0.5974	0.5507	-19882.4023	37215.6595
age	-80.7196	301.0309	-0.2681	0.7888	-673.1034	511.6641
ill	-90.1316	288.7235	-0.3122	0.7551	-658.2962	478.0330

Omnibus:	32.598	Durbin-Watson:	1.761
Prob(Omnibus):	0.000	Jarque-Bera (JB):	73.878
Skew:	0.518	Prob(JB):	0.000
Kurtosis:	5.131	Condition No.:	5844912789236

Таб. 3

Для удобства сравнения коэффициентов создадим общую таблицу (Таб. 4).

Стоит также проверить необходимости борьбы с эндогенностью, для этого проведем тест Хаусмана-Ву. Основная гипотеза H_0 : в данных отсутствует эндогенность. С помощью библиотеки `linearmodels` рассчитаем статистику:

Wu-Hausman test of exogeneity
 H_0 : All endogenous variables are exogenous
Statistic: 11.8058
P-value: 0.0000
Distributed: F(6,296)

Получили $p\text{-value} = 0$, что означает отвержение гипотезы H_0 на любом разумном уровне значимости, следовательно, в данных присутствует эндогенность, а значит, что сделанные выводы на основании МНК-оценок не являются состоятельными.

Исходя из модели 2SLS все коэффициенты являются статистически незначимыми, поэтому все обозначенными нами гипотезы отвергаются.

Вероятнее всего, такие результаты получились в связи с малым количеством наблюдений, а также недостаточность содержательных признаков, которые смогли бы лучше описать данные по заработной плате.

	OLS	2SLS
R-squared	0.4900	0.0455
	OLS	2SLS
R-squared	0.4912	0.0486
R-squared Adj.	0.4693	0.0077
age	-798.6536 (120.0291)	-80.7196 (301.0309)
const	33328.6226 (7931.3346)	-1340.4239 (26992.3525)
educ_ВЫСШЕЕ	5402.7542 (2737.5161)	26102.8119 (51077.0470)
ill	50.3393 (128.9734)	-90.1316 (288.7235)
marriage_состояние	-107.0565 (1987.6616)	8666.6286 (14507.7255)
occup_ВЫСШАЯ_КВАЛИФИКАЦИЯ	14351.0936 (3836.4943)	35801.2873 (66746.1534)
occup_КВАЛ_РАБОЧИЕ	21245.8373 (3173.2999)	44307.0970 (37004.9087)
occup_НЕКВАЛ_РАБОЧИЕ	5419.1982 (5366.7066)	62458.4524 (58510.2061)
popul	0.0187 (0.0227)	0.0162 (0.0919)
popul_ln	-333.3458 (598.7702)	-1124.9146 (1800.9766)
popul_sq	-0.0000 (0.0000)	-0.0000 (0.0000)
smoke	6093.3545 (2667.2291)	10400.9679 (29336.7343)

Таб. 4

6. Расшифровка переменных

sal - зарплата (таргет)
popul - численность населения
w_years - опыт работы
smoke - курение (1 - да, 0 - нет)
occup - квал
educ - образование
marriage - семейное положение
satisf - удовлетворенность жизнью
sex - пол (1 - мужчина, 0 - женщина)
age - возраст
school - вид школы
kids - количество детей

harm - вредность производства (1 - вредно, 0 - нет)
ill - количество дней, пропущенных по болезни

7. Список литературы

[1] Eugenia Chernina, Vladimir Gimpelson “Do wages grow with experience? Deciphering the Russian puzzle” // Journal of Comparative Economics. - 2023. - Volume 51, Issue 2.

[2] Marco Paccagnella “Age, Ageing and Skills: Results from the Survey of Adult Skills” // OECD Education Working Papers. - 2017. - №132.

[3] Wang Q, Shen JJ, Sotero M, Li CA, Hou Z (2018) Income, occupation and education: Are they related to smoking behaviours in China? PLoS ONE 13(2): e0192571. <https://doi.org/10.1371/journal.Pone.0192571>

[4] Кадрева О.Н. “Влияние количества и возраста детей на заработки работающих женщин” // Прикладная эконометрика, 2016, т. 41, с. 62–77.

[5] Shari J. Eli “What Determines Our Wage: The Econometric Analysis of Male-Female Wage Gap” // Economics 196, Special Topics in Economic Research. - 2011. https://eml.berkeley.edu/~webfac/eichengreen/e196_fall/econ196_paper4.pdf

8. Первоначальный Project-Proposal

1. Введение

В данном проекте мы будем исследовать зависимость заработной платы в России от различных признаков.

Актуальность исследования заключается в том, что российский рынок труда в последнее время отличается нестабильностью с точки зрения распределения зарплат и низким уровнем безработицы. Более того, имеют место множество трендов со стороны работников, которые сильно влияют на показатели рынка труда. Например, сейчас в России явно дает о себе знать смещение интереса в сторону сферы IT и инфобизнеса. На ситуации шока Российская экономика обычно реагирует колебанием цен, в том числе и зарплат, нежели изменением уровня безработицы. Именно поэтому мы считаем, что необходимо подробно исследовать вопрос моделирования заработной платы, так как это внесет ясность о том, какие факторы в наибольшей степени влияют на ее колебания.

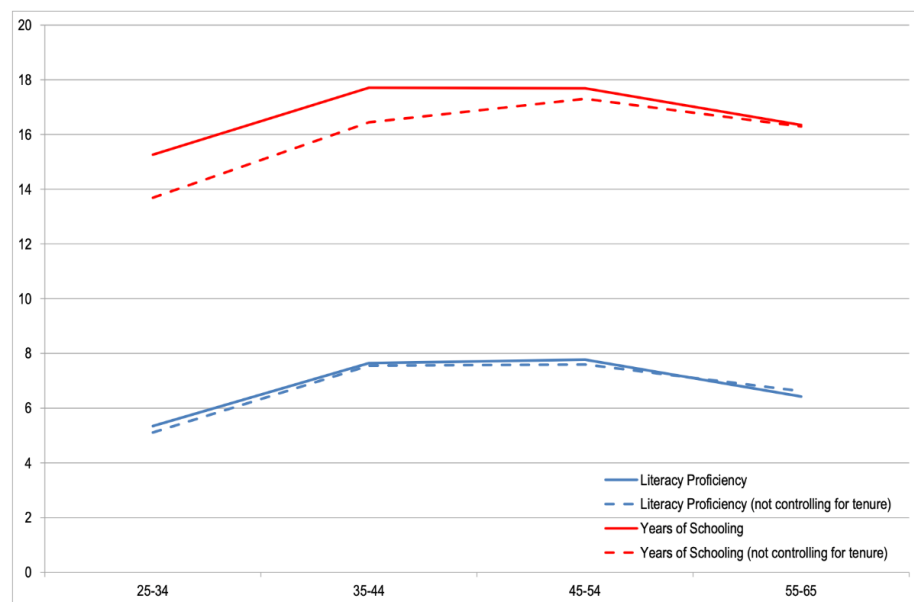
2. Обзор литературы

Исследование, описанное в статье [1] показывает, что заработная плата в России имеет нестандартную траекторию динамики роста и падения на протяжении трудового опыта работников. Например, своего пика она достигает на ранних этапах развития карьеры и затем стремительно падает. При стаже работы 35-39 лет доходы падают даже ниже начального уровня. Также исследование показывает, что сейчас наибольшую зарплату получают люди в возрасте 34-38 лет. Однако нельзя утверждать, что данная работа в точности характеризует реальную ситуацию на рынке труда, так как для анализа использовались данные 2000-2019 года. Мы считаем, что с тех пор ситуация на рынке труда значительно изменилась из-за многих факторов: изменился спектр действующих компаний в России, сместился фокус на IT-сферу как у работодателей, так и у работников, вырос приток трудовых мигрантов и другое. Более того, выборка включает в себя только мужчин 20-60 лет, имеющих стаж работы до 40 лет, причем исключая самозанятых. В своем исследовании мы планируем рассматривать данные по более широкой выборке респондентов.

Несомненно, еще одним важным аспектом на рынке труда является уровень образования работников. В нашем исследовании этот показатель разбит на множество когорт:

- 0 классов школы
- 5 классов школы
- 7 классов школы
- 8 классов школы
- 9 классов школы
- 7-9 классов школы (незак. средн) и менее 2 лет в техникуме
- 7-9 классов школы (незак. средн) + ПТУ без диплома
- 7-9 классов школы (незак. средн) + ПТУ с дипломом
- 10+ классов школы без аттестата
- есть аттестат о среднем образовании (11 классов)
- 10+ классов школы и техникум без диплома
- 10+ классов школы и какое-либо професси. обр. без диплома
- 10+ классов школы и какое-либо професси. обр. с дипломом
- техникум с дипломом
- 1-2 года в ВУЗе
- 3+ лет в ВУЗе
- есть диплом о высшем образовании
- аспирантура и т.п. с дипломом
- аспирантура и т.п. без диплома.

В статье [2] обсуждается влияние грамотности на трудоустроенность и заработок. Вероятнее всего, эту самую грамотность можно оценить через уровень образования работника. Основным результатом данной статьи является вывод о том, что навыки эффективной обработки информации имеют огромное влияние на трудоустройство.



Note: Marginal effects from a one-standard deviation increase in literacy proficiency and years of schooling, from an OLS regression on log wages on interactions between age dummies and proficiency/education. The regression also controls for gender, part-time status and type of contract.

Рисунок 1. Распределение грамотности и лет образования по возрастам.

Источник: Survey of Adult Skills (PIAAC) (2012),

www.oecd.org/site/piaac/publicdataandanalysis.htm

Кроме того, в исследовании показано, что более высокий уровень грамотности способствует более позднему выходу на пенсию. Таким образом, у работников появляется шанс на более долгосрочном горизонте получать опыт работы и, вероятно, требовать более высокую зарплату.

Еще одним интересующим нас феноменом стала взаимосвязь курения и размера заработной платы. По этому поводу в статье [3] авторы, исследующие отличия в образовании, занятости и зарплате между курильщиками и некурящими в Китае, пришли к выводу о том, что курение напрямую не значимо влияет на зарплату. Однако исследование показало, что в сельской среде социально-экономический статус ниже, чем в городах, и тенденция к курению там (в поселках) выше. Также была обнаружена взаимосвязь между курением и занимаемой должностью работника. В статье говорится о том, что высокий заработок и важная позиция в компании имеют положительную взаимосвязь с курением среди мужчин. Кроме того, у людей без образования, но умеющих читать и писать, вероятность наличия зависимости от курения намного выше. У тех же, кто имел образование выше школьного, тяга к курению значительно уменьшалась вне зависимости от пола или места проживания.

Также мы хотим проверить влияние количества детей у индивидов на их заработную плату. В статье [4], где автор также исследует данные РМЭЗ 18-22 волн, показано, что наличие ребенка до 3-х лет у работающих женщин уменьшает их зарплату на 11,8% на 10% уровне значимости. При этом наличие ребенка от 3 до 7 лет никак не влияет на зарплату. “Для модели со структурным уравнением самоотбора наличие одного ребенка значимо отрицательно влияет на заработки женщин (на 10%-ном уровне значимости для уравнения без учета самоотбора и на 5%-ном для уравнения с его учетом),” – говорится в статье. Под самоотбором автор подразумевает сознательное решение женщины выходить на рынок труда, что приводит к эндогенности, так как только для них мы можем наблюдать заработную плату.

В целом посмотреть, какую форму имеют эконометрические модели для предсказания зарплаты, нам позволила статья [5]. Несмотря на то, что описанная там модель включает не так

много переменных, как ожидается в нашей, результаты все же будут полезны для дальнейшей работы. Автор использует такие переменные как:

- продолжительность обучения (лет)
- продолжительность рабочего стажа (лет)
- пол
- этничность (1 - белый, 0 - иначе)
- полный график (1 - если да, 0 - иначе)
- замужество (1 - в браке, 0 - иначе)
- локация (1 - Лондон, 0 - иначе)
- вера (1 - христианство, 0 - иначе)
- национальность (1 - британец, 0 - иначе)
- финансы (1 - работает в сфере бизнес/финансах/computer science/R&D, 0 - иначе)

Несомненно, многие формулировки переменных вызывают вопросы: например, локация и национальность (брались данные по индивидам в Великобритании за 2007 год).

3. Данные

Наш проект основывается на данных РМЭЗ НИУ ВШЭ¹. Мы будем использовать полную выборку данных по индивидам за 2022 год (волна 31). Собрав необходимые переменные и очистив их от пропусков, мы получили около 4500 наблюдений.

В качестве переменных мы выбрали следующие данные:

- тип населенного пункта (где человек родился)
- кол-во жителей в населенном пункте
- проживает ли на данный момент там, где родился
- семейное положение
- квалификация (специалисты высшего уровня квалификации, служащие офисные и по обслуживанию клиентов, работники сферы торговли и услуг и другие)
- образование
- возраст
- пол
- национальность
- занятость (работает/в декрете/в отпуске и тд)
- удовлетворенность жизнью (скорее удовлетворены/полностью удовлетворены и т.д.)
- профессия
- сфера деятельности
- наличие подчиненных
- продолжительность рабочего дня
- тип трудоустройства
- задолженность со стороны работодателя по з/п
- вредность работы
- гос. компания или частная
- владеет ли человек компанией, в которой работает
- ожидания по уровню жизни в течение 12 месяцев

¹ <https://www.hse.ru/rlms/>

- оценка собственной степени богатства (1-9)
- курение (1 - курит, 0 - иначе)

Среди респондентов оказалось 2435 женщин и 2078 мужчин в возрасте от 16 до 83 лет. В приложении мы представили [графики](#) по нашим данным и [расшифровку переменных](#).

4. Гипотезы

В ходе изучения данных, различных объясняющих, а также множества эконометрических статей мы решили проверить несколько скучных и интересных гипотез для модели, предсказывающей зарплату,, а именно:

- В больших городах (в городах с большей численностью) больше возможностей для самореализации человеческого потенциала, поэтому коэффициент при переменной численности населения должен быть положительным.
- Более квалифицированные специалисты обладают большими знаниями и навыками, поэтому получают зарплату больше (коэффициент при образовании положительный).
- Согласно найденным статьям, сам факт курения положительно влияет на доход индивида, поэтому коэффициент при курении должен быть положительным.

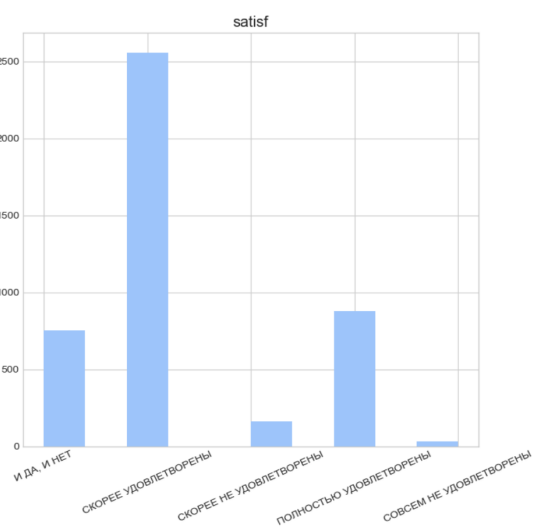
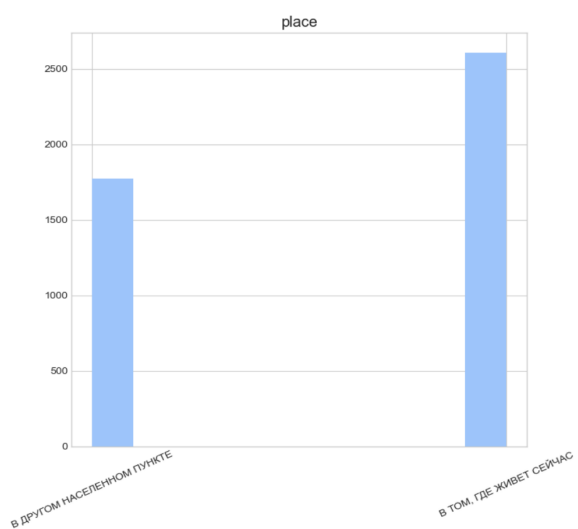
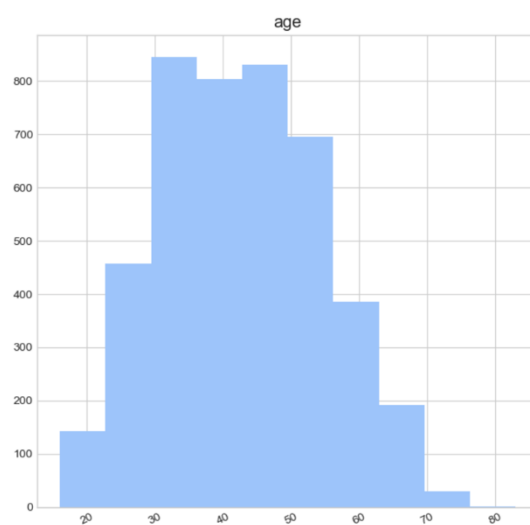
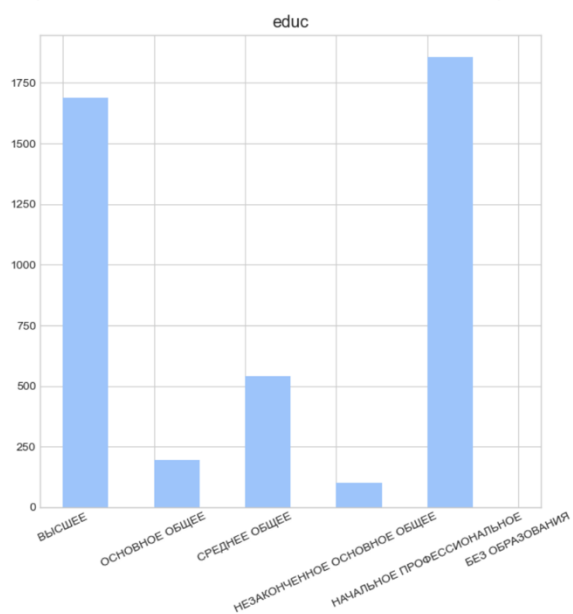
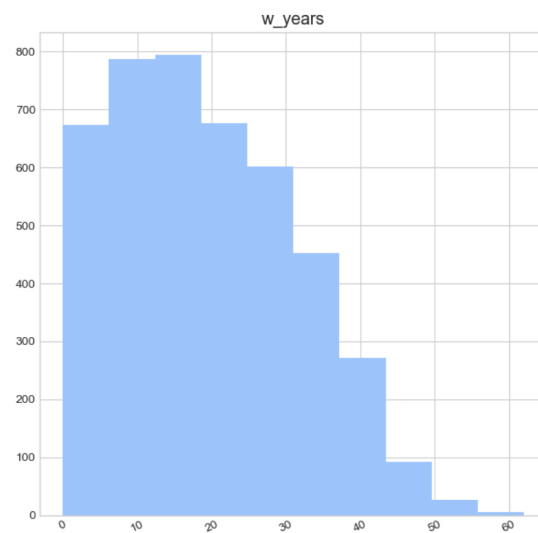
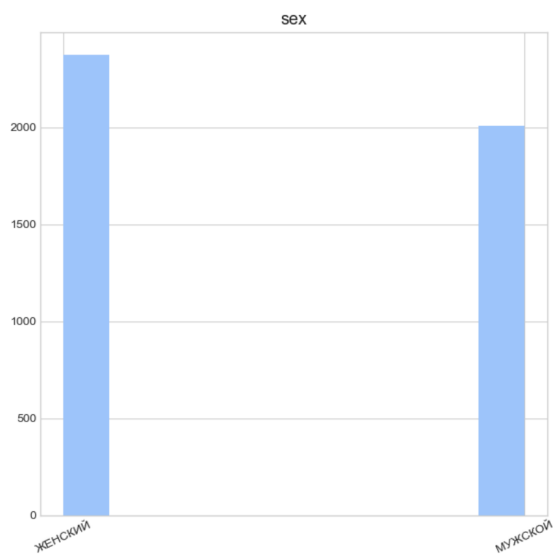
5. Спецификация модели (пробная)

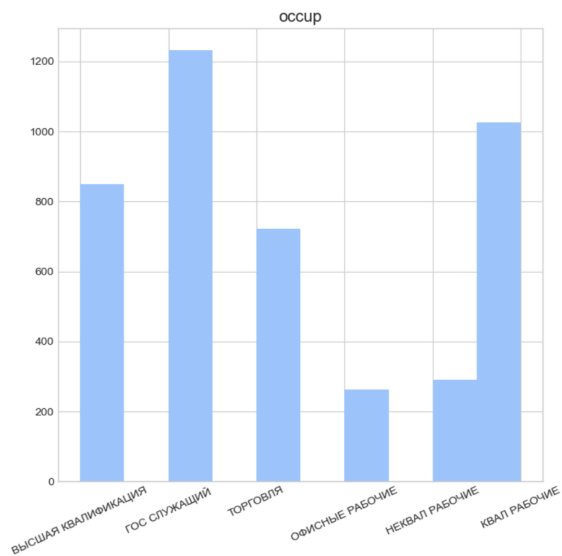
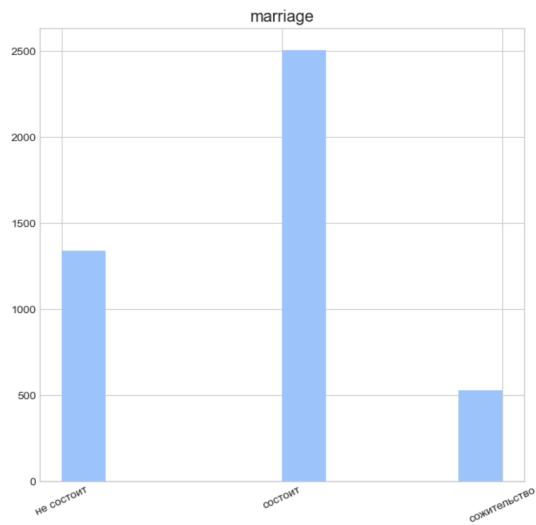
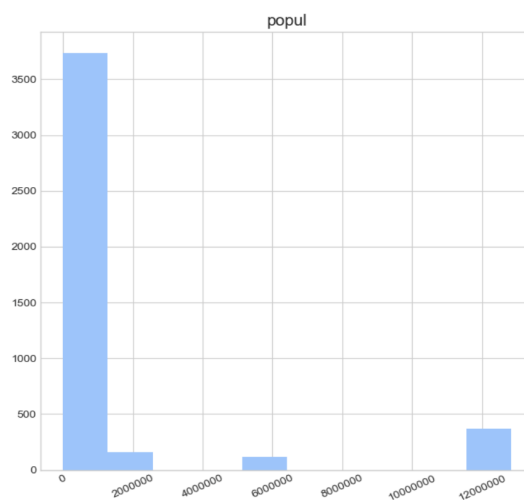
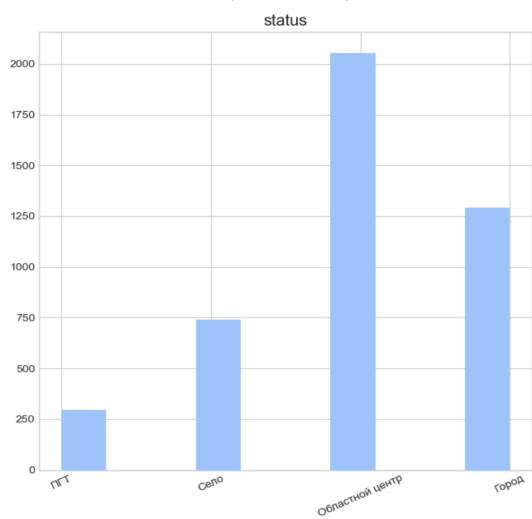
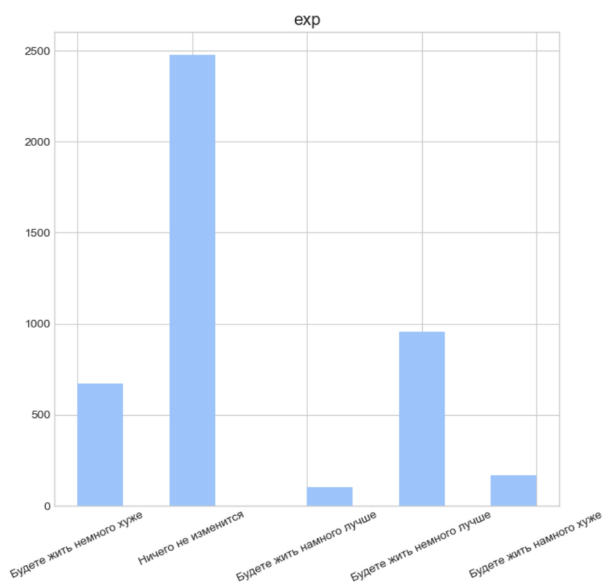
Проанализировав корреляционные матрицы и воспользовавшись здравым смыслом, мы сделали несколько выводов относительно взаимосвязи зарплаты и объясняющих переменных в нашей будущей модели. Выводы представлены в [Таблице 1](#).

Эндогенность наверняка присутствует в образовании, занятости (из-за самоотбора), регионе (двусторонняя связь).

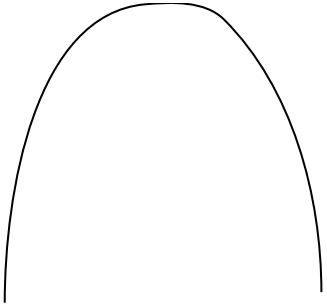
6. Приложения

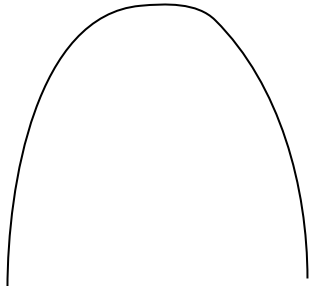
1. Графики

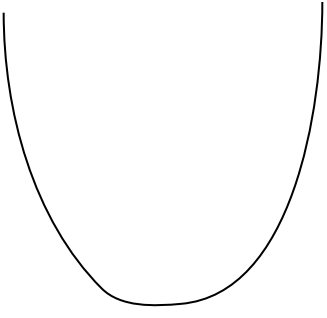




2. Таблица 1

Переменная	Предполагаемая связь	Тип взаимосвязи
Возраст	<p>Ни для кого не секрет, что пика з/п достигает в среднем возрасте, поэтому связь нелинейная и что-то типа параболы ветвями вниз</p> <p>Кстати, корреляция: - 0.146230</p>	
Кол-во рабочих часов в день	<p>По логике, чем больше человек работает, тем больше у него должна быть з/п. Но к нашему удивлению, корреляция между з/п и whours оказалась отрицательной (хоть и маленькой по модулю: - 0.014554)</p>	-
Стаж работы	<p>Тут тоже вроде неожиданно, а вроде и логично: корреляция -0.074957)</p>	-
Оценка богатства	<p>Очевидно, кто считает себя богатым, тот и есть богатый скорее всего. Корреляция это отчасти подтверждает: 0.200319</p>	+
Тип населенного пункта	<p>Скорее всего, в городе/областном центре з/п будет больше, чем в ПГТ/селе.</p>	<p>+</p> <p>(если идти от меньшего к большему)</p>
Семейное положение	<p>Вероятно, у женатиков з/п должна быть больше для содержания семьи</p>	<p>+</p> <p>(у тех, кто в браке)</p>
Квалификация	<p>Тут скорее всего надо как-то проранжировать квалификации, тогда, возможно, мы увидим, кто зарабатывает больше: военные / чиновники / работники торговли и услуг и др.</p>	?

Образование	Тут, очевидно, будет положительное влияние на зарплату, однако нелинейное. Иван Павлович говорил, что у кандидатов и докторов редко зарплаты бывают крутые, так что ожидаем что-то типа параболы ветвями вниз?	<p>?</p> 
Пол	Ожидаем статистически незначимое различие между мужчинами и женщинами	Нет значимого влияния
Национальность	Т.к. у нас в выборке практически 100% считают себя русскими, сложно сейчас сказать, как другая национальность влияет на з/п. Если ввести дамми (1 - русский, 0 - нет), то пожалуй, у коэффициент будет положительным (?)	+?
Занятость	Очевидно, мамочки в декрете / люди в отпуске получают меньше тех, кто ходит на работу	<p>+</p> <p>(для тех, кто занят)</p>
Сфера деятельности	Смеем предположить, что больше всех будут зарабатывать айтишники, нефтегаз, финансы и еще некоторые	Пока не знаем, как отранжировать значения :(
Наличие подчиненных	Очевидно, у кого есть подчиненные, занимает более высокую должность=выше з/п	+
Официальность трудоустройства	Вот тут тяжело предположить, но если аппроксимировать неофициально трудящихся к мигрантам, которые в основном водятся на стройках и рынках, то пожалуй, официально трудоустроенные работники	<p>+</p> <p>(для официально трудящихся)</p>

	получают больше	
Вредность производства	По идее, за вредность дают надбавку к зарплате	+
Количество детей	В статье [4] рассказывали, что кол-во не влияет, а вот наличие - да	\pm
Курение	С одной стороны, больше курят в селах, где меньше з/п. С другой стороны, крутые дяди любят сигары. Наверное, снова парабола, но ветвями вверх	

7. Расшифровка переменных

Целевая:

sal заработная плата в рублях

Категориальные:

status тип населенного пункта (где человек родился):

Город; Областной центр; ПГТ; Село

marriage семейное положение:

- Вдовец (вдова);
- Живете вместе, но не зарегистрированы;
- Никогда в браке не состояли;
- ОФИЦИАЛЬНО ЗАРЕГИСТРИРОВАНЫ, НО ВМЕСТЕ НЕ ПРОЖИВАЮТ;
- Разведены и в браке не состоите;
- Состоите в зарегистрированном браке

occup квалификация:

- военнослужащие;
- законодатели, крупные чиновники, руководители высш. и сред. звена;
- квалифицированные работники сельского, лесного хоз-ва и рыбоводства;
- квалифицированные рабочие, занятые ручным трудом;
- квалифицированные рабочие, использующие машины и механизмы;
- неквалифицированные рабочие всех отраслей;
- работники сферы торговли и услуг;
- служащие офисные и по обслуживанию клиентов;
- специалисты высшего уровня квалификации;
- специалисты среднего уровня квалификации;
- чиновники

educ образование:

- 0 классов школы
- 5 классов школы
- 7 классов школы
- 8 классов школы
- 9 классов школы
- 7-9 классов школы (незак. средн) и менее 2 лет в техникуме
- 7-9 классов школы (незак. средн) + ПТУ без диплома
- 7-9 классов школы (незак. средн) + ПТУ с дипломом
- 10+ классов школы без аттестата
- есть аттестат о среднем образовании (11 классов)
- 10+ классов школы и техникум без диплома
- 10+ классов школы и какое-либо профес. обр. без диплома
- 10+ классов школы и какое-либо профес. обр. с дипломом
- техникум с дипломом
- 1-2 года в ВУЗе

- 3+ лет в ВУЗе
- есть диплом о высшем образовании
- аспирантура и т.п. с дипломом
- аспирантура и т.п. без диплома.

sex пол (мужской, женский)

place проживает там, где родился или в другом месте

nation национальность

employ занятость:

Вы находитесь в любом другом оплачиваемом отпуске;

Вы находитесь в отпуске - декретном или по уходу за ребенком до 3-х лет;

Вы сейчас работаете

satisf степень удовлетворенности

И ДА, И НЕТ;

ПОЛНОСТЬЮ УДОВЛЕТВОРЕННЫ;

СКОРЕЕ НЕ УДОВЛЕТВОРЕННЫ;

СКОРЕЕ УДОВЛЕТВОРЕННЫ;

СОВСЕМ НЕ УДОВЛЕТВОРЕННЫ

job профессия (список включает 315 профессий)

field сфера деятельности:

ИТ, ИНФОРМАЦИОННЫЕ ТЕХНОЛОГИИ;

АРМИЯ, МВД, ОРГАНЫ БЕЗОПАСНОСТИ;

ВОЕННО-ПРОМЫШЛЕННЫЙ КОМПЛЕКС;

ГРАЖДАНСКОЕ МАШИНОСТРОЕНИЕ;

ДЕРЕВООБРАБАТЫВАЮЩАЯ ПРОМЫШЛЕННОСТЬ, ЛЕСНОЕ ХОЗЯЙСТВО;

ДРУГАЯ ОТРАСЛЬ ТЯЖЕЛОЙ ПРОМЫШЛЕННОСТИ;

ЖИЛИЩНО-КОММУНАЛЬНОЕ ХОЗЯЙСТВО;

ЗДРАВООХРАНЕНИЕ;

ЛЕГКАЯ, ПИЩЕВАЯ ПРОМЫШЛЕННОСТЬ;

НАУКА, КУЛЬТУРА;

НЕФТЕГАЗОВАЯ ПРОМЫШЛЕННОСТЬ;

ОБРАЗОВАНИЕ;

ОБЩЕСТВЕННЫЕ ОРГАНИЗАЦИИ, СОВЕТ ВЕТЕРАНОВ И ПР.;

ОПЕРАЦИИ С НЕДВИЖИМОСТЬЮ;

ОРГАНИЗАЦИЯ ОБЩЕСТВЕННОГО ПИТАНИЯ;

ОРГАНЫ УПРАВЛЕНИЯ;

РЕКЛАМА, МАРКЕТИНГ;

СЕЛЬСКОЕ ХОЗЯЙСТВО;

СМИ, ИЗДАТЕЛЬСТВО, ПЕЧАТЬ, ТЕЛЕКОММУНИКАЦИИ;

СОЦИАЛЬНОЕ ОБСЛУЖИВАНИЕ;

СПОРТ, ТУРИЗМ, РАЗВЛЕЧЕНИЯ;

СТРОИТЕЛЬСТВО;

ТОРГОВЛЯ, БЫТОВОЕ ОБСЛУЖИВАНИЕ;

ТРАНСПОРТ, СВЯЗЬ;

УСЛУГИ НАСЕЛЕНИЮ;

ФИНАНСЫ;

ХИМИЧЕСКАЯ ПРОМЫШЛЕННОСТЬ;

ЭКОЛОГИЯ, ЗАЩИТА ОКРУЖАЮЩЕЙ СРЕДЫ;

ЭНЕРГЕТИЧЕСКАЯ ПРОМЫШЛЕННОСТЬ;

ЮРИСПРУДЕНЦИЯ

staff наличие подчиненных (есть; нет)

official тип трудоустройства (официально; неофициально)

debt наличие задолженности по зарплате со стороны работодателя (есть; нет)

harm вредность производства (вредно; нет)

gov является ли место работы гос.компанией (да; нет)

own является ли индивид владельцем компании, где трудоустроен (да; нет)

exp ожидания относительно уровня жизни через 12 месяцев:
 Будете жить намного лучше;
 Будете жить намного хуже;
 Будете жить немного лучше;
 Будете жить немного хуже;
 Ничего не изменится

wealth оценка собственного богатства (1; 2; 3 ... 9)

region регион проживания

kids количество детей

smoke сколько сигарет курит в день

vape сколько вейпов курит в день

Непрерывные:

popul кол-во жителей в населенном пункте (непрерывная)

age возраст

whour кол-во рабочих часов в день

wyears стаж работы (кол-во лет)

sal - зарплата (таргет)

popul - численность населения

w_years - опыт работы

smoke - курение (1 - да, 0 - нет)

occup - квал

educ - образование

marriage - семейное положение

satisf - удовлетворенность жизнью

sex - пол (1 - мужчина, 0 - женщина)

age - возраст

school - вид школы

kids - количество детей

harm - вредность производства (1 - вредно, 0 - нет)

ill - количество дней, пропущенных по болезни

8. Список литературы:

- [1] Eugenia Chernina, Vladimir Gimpelson “Do wages grow with experience? Deciphering the Russian puzzle” // Journal of Comparative Economics. - 2023. - Volume 51, Issue 2.
- [2] Marco Paccagnella “Age, Ageing and Skills: Results from the Survey of Adult Skills” // OECD Education Working Papers. - 2017. - №132.
- [3] Wang Q, Shen JJ, Sotero M, Li CA, Hou Z (2018) Income, occupation and education: Are they related to smoking behaviours in China? PLoS ONE 13(2): e0192571. <https://doi.org/10.1371/journal.Pone.0192571>
- [4] Кадрева О.Н. “Влияние количества и возраста детей на заработка работающих женщин” // Прикладная эконометрика, 2016, т. 41, с. 62–77.
- [5] Shari J. Eli “What Determines Our Wage: The Econometric Analysis of Male-Female Wage Gap” // Economics 196, Special Topics in Economic Research. - 2011. https://eml.berkeley.edu/~webfac/eichengreen/e196_fall/econ196_paper4.pdf