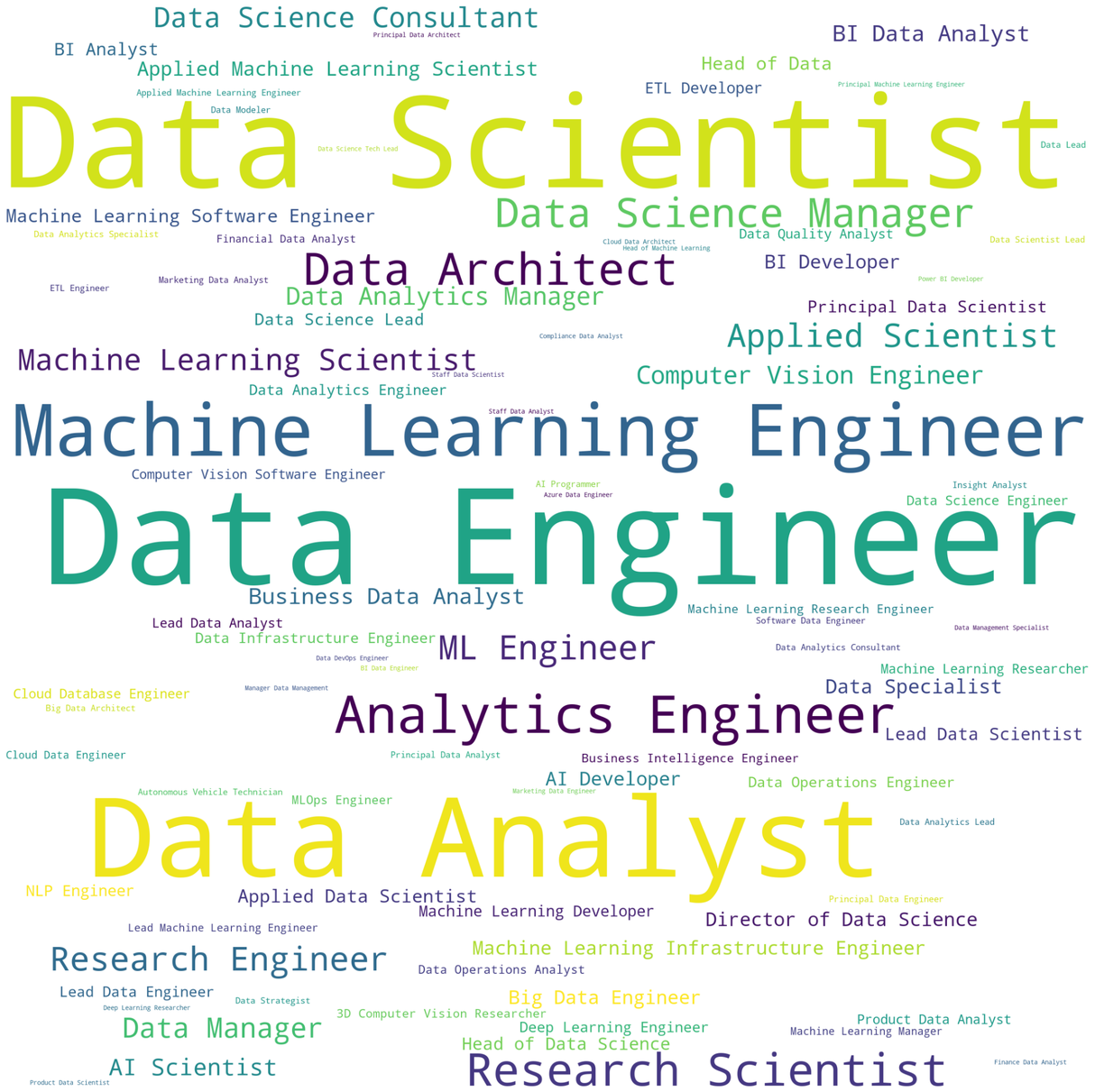
maths hub

международная школа анализа данных и разработки

SQL Project

Анализ влияния внешних и внутренних факторов на уровень заработной платы в Data-профессиях

Varlamov Nikita



DATASET

Data Science Salaries 2023

Для анализа данных был взят датасет с сайта Kaggle.com, содержащий информацию о заработной плате в различных областях Data-профессий на 2023 год.

Набор данных о зарплатах в сфере Data Science содержит 11 столбцов:

- work_year: год выплаты зарплаты;
- experience_level: уровень опыта работы на должности в течение года;
- employment_type: тип занятости на соответствующей должности;
- job_title: наименование должность;
- salary: общая сумма выплаченной заработной платы за год;
- salary_currency: валюта выплачиваемой зарплаты в виде кода валюты ISO 4217;
- salaryinusd: заработная плата в долларах США;
- employee_residence: основная страна проживания сотрудника в течение рабочего года в виде кода страны ISO 3166;
- remote_ratio: общий объем работы, выполненной удаленно;
- company_location: страна головного офиса или филиала работодателя;
- company_size: среднее количество людей, работавших в компании в течение года.

Overview | Alerts 5 | Reproduction

Dataset statistics

Number of variables	11
Number of observations	2584
Missing cells	0
Missing cells (%)	0.0%
Duplicate rows	0
Duplicate rows (%)	0.0%
Total size in memory	232.2 KiB
Average record size in memory	92.0 B

Variable types

Numeric	2
Categorical	6
Text	3

ЦЕЛЬ

Определение влияния внешних и внутренних факторов на уровень заработной платы в Data-профессиях

.01

Общий анализ

Определение состава выборки, процентных соотношений групп датасета

.02

Изменение з/п

Анализ изменения заработной платы г/г, определение средней зарплаты по Data-профессиям

.03

Формат занятости

Определение изменений в соотношении формата занятости (удаленной работы)

.04

Анализ заработной платы

Выявление зависимости заработной платы от различных факторов

.05

Корреляция

Определение статистической взаимосвязи

ПРЕДОБРАБОТКА ДАННЫХ

Python

```
# Основные
import numpy as np
import pandas as pd
from ydata_profiling import ProfileReport
import sqlite3
from sqlalchemy import create_engine

# Визуализация
import matplotlib.pyplot as plt
import seaborn as sns
import plotly.express as px
import plotly.figure_factory as ff
import plotly.graph_objects as go
import nltk
from wordcloud import WordCloud

# Статистика
from scipy import stats
from scipy.stats import norm
```

Библиотеки

Для работы с данными в основном использовался функционал pandas. Подключение к БД через sqlite3.

Визуализация данных: matplotlib, seaborn, wordcloud.

Статистика: scipy.

1

```
[349] df = pd.read_csv('ds_salaries.csv')
      df.head()
      ✓ 0.0s
```

	work_year	experience_level	employment_type	job_title	salary	salary_currency
0	2023	SE	FT	Principal Data Scientist	80000	EUR
1	2023	MI	CT	ML Engineer	30000	USD
2	2023	MI	CT	ML Engineer	25500	USD
3	2023	SE	FT	Data Scientist	175000	USD
4	2023	SE	FT	Data Scientist	120000	USD

```
[350] df.shape
      ✓ 0.0s
(3755, 11)
```

```
[352] # Просмотр типа данных
      df.dtypes
      ✓ 0.0s
work_year              int64
experience_level      object
employment_type       object
job_title             object
salary                 int64
salary_currency       object
salary_in_usd          int64
employee_residence    object
remote_ratio           int64
company_location      object
company_size          object
dtype: object
```

Формирование dataframe

Формирование датафрейм из формата csv.

Просмотр вида табличных данных,

определение формы таблицы и типа данных.

2

```
[353] # Проверка на пустые значения
      df.isnull().sum()
      ✓ 0.0s
work_year             0
experience_level      0
employment_type       0
job_title             0
salary                0
salary_currency       0
salary_in_usd         0
employee_residence    0
remote_ratio          0
company_location      0
company_size          0
dtype: int64
```

```
[354] #Проверка на дубликаты
      duplicate_rows_data = df[df.duplicated()]
      print(duplicate_rows_data.shape)
      ✓ 0.0s
(1171, 11)
```

```
[355] df = df.drop_duplicates()
      df.shape
      ✓ 0.0s
(2584, 11)
```

```
[356] # Првоерка уникальных значений
      for column in df.columns:
          num_distinct_values = len(df[column].unique())
          print(f"{column}: {num_distinct_values} уникальных значений")
      ✓ 0.0s
work_year: 4 уникальных значений
experience_level: 4 уникальных значений
employment_type: 4 уникальных значений
job_title: 93 уникальных значений
salary: 815 уникальных значений
salary_currency: 20 уникальных значений
salary_in_usd: 1035 уникальных значений
employee_residence: 78 уникальных значений
remote_ratio: 3 уникальных значений
company_location: 72 уникальных значений
company_size: 3 уникальных значений
```

Null и дубликаты

Проверка данных на пустые значения и дубликаты. Пустых значений не выявлено, дубликаты строк в количестве 1171 удалены. Промежуточное значение shape 2584/11.

3

```
# Наименование столбцов соответствует naming convention.
# Уточнение наименования значений данных.
df['experience_level'] = df['experience_level'].replace({
    'EN': 'Junior',
    'MI': 'Middle',
    'SE': 'Senior',
    'EX': 'Executive',
})

df['employment_type'] = df['employment_type'].replace({
    'FL': 'Freelancer',
    'CT': 'Contractor',
    'FT' : 'Full-time',
    'PT' : 'Part-time'
})

df['company_size'] = df['company_size'].replace({
    'S': 'SMALL',
    'M': 'MEDIUM',
    'L' : 'LARGE',
})
df['remote_ratio'] = df['remote_ratio'].astype(str)
df['remote_ratio'] = df['remote_ratio'].replace({
    '0': 'Office',
    '50': 'Half-Remote',
    '100' : 'Full-Remote',
```

Naming convention

Наименование столбцов таблицы соответствует Naming convention.

Однако значения данных в таблице требуют уточнения для более оперативного ориентирования в базе.


```
#Группировка профессий и удаление лишних колонок
def assign_broader_category(job_title):
    data_analyst = ['Data Analyst', ...                ]
    data_engineering = ['Analytics Engineer', ...      ]
    data_scientist = ['Principal Data Scientist', ...  ]
    machine_learning = ['ML Engineer', ...             ]
    data_architecture = ['Data Architect', ...         ]
    management = ['Data Analytics Manager', ...        ]

    if job_title in data_analyst:
        return "Data Analyst"
    elif job_title in data_engineering:
        return "Data Engineering"
    elif job_title in data_scientist:
        return "Data Science"
    elif job_title in machine_learning:
        return "Machine Learning"
    elif job_title in data_architecture:
        return "Data Architecture"
    elif job_title in management:
        return "Management"
    else:
        return "Other"

df['job_category'] = df['job_title'].apply(assign_broader_category)
df.insert(loc=0, column='id', value=(np.arange(1, 1 + len(df))))
df.drop(df[['salary','salary_currency']], axis = 1, inplace = True)
```

```
[ ]
```

```
df.shape
```

```
[360] ✓ 0.0s
```

```
(2584, 11)
```

Добавление/удаление столбцов

Формирование итогового вида таблицы.

Удаление данных о зарплате в местной валюте и виде валюты. Добавление колонки id, а также группировка Data-профессий по направлениям.

Итоговой значение shape 2584/11.


```
# Загрузка в БД SQLite
database_path = r'C:\sqlite\ds_salary'
engine = create_engine(f'sqlite:///{database_path}')
connection = engine.connect()
table_name = 'ds_salary'
df.to_sql(table_name, engine, if_exists='replace', index=False)
```

```
[362] ✓ 0.0s
```

```
2584
```

```
# Закрываем соединение
connection.close()
```

```
[363] ✓ 0.0s
```

Загрузка БД

Загрузка таблицы в базу данных.

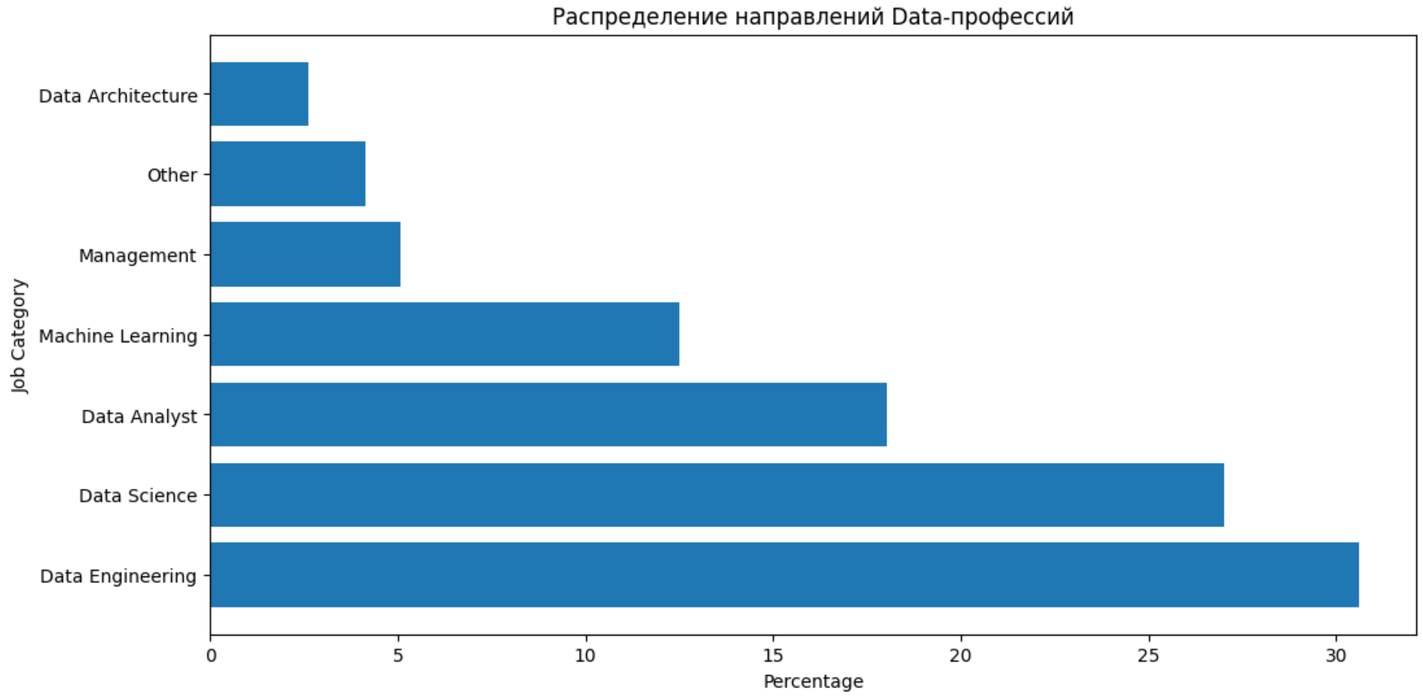
В качестве СУБД используем SQLite.

1. ОБЩИЙ АНАЛИЗ

Распределение направлений Data-профессий

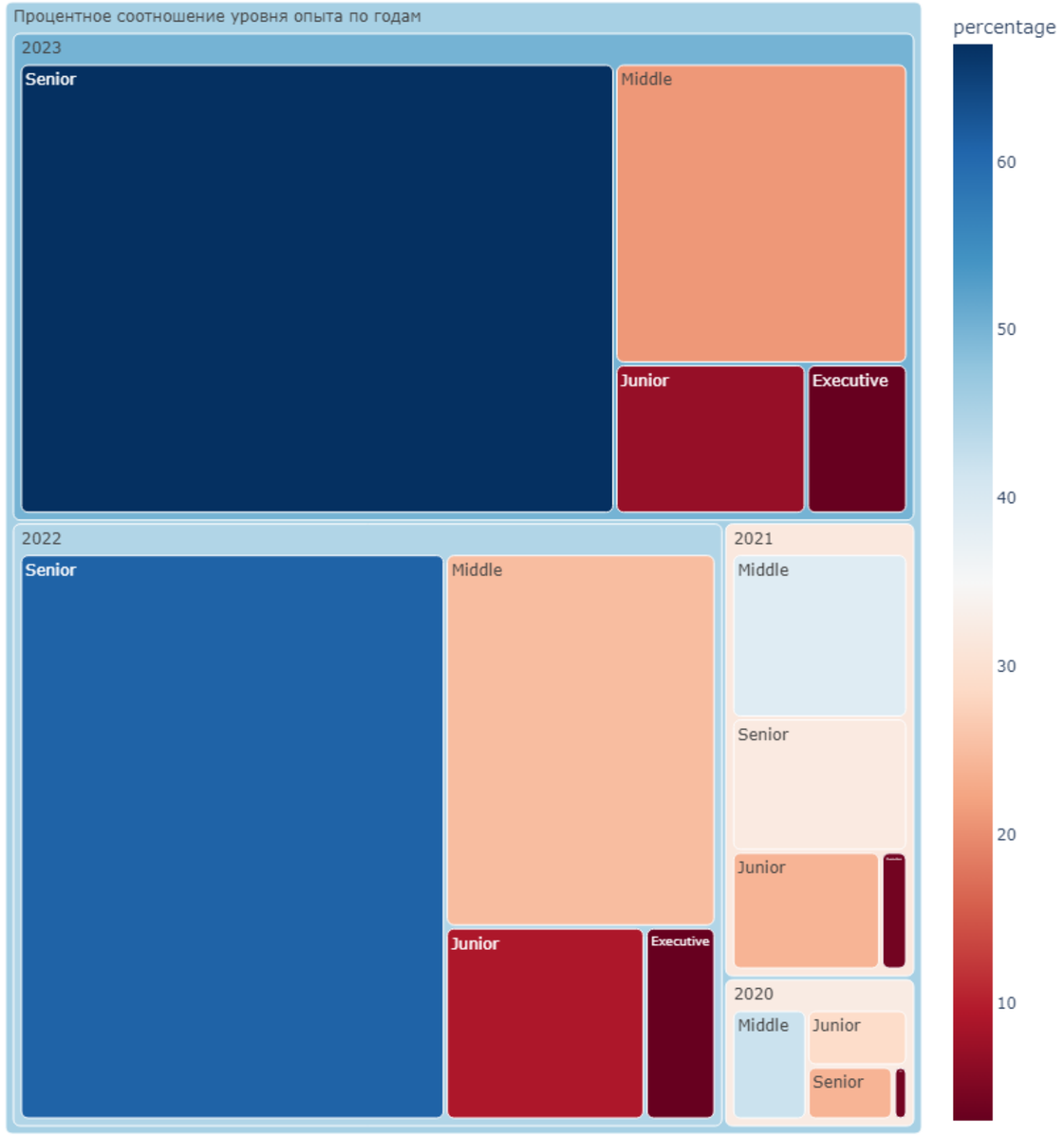
```sql
-- Процентное соотношение Data направлений по годам
SELECT
    ds.work_year,
    ds.job_category,
    COUNT(ds.job_category) AS job_category_cnt,
    COUNT(ds.job_category)*100/wyc.cnt AS percentage
FROM ds_salary ds
JOIN (SELECT
            work_year,
            count(work_year) AS cnt
        FROM ds_salary ds
        GROUP BY 1)
        AS wyc
ON ds.work_year = wyc.work_year
GROUP BY 1,2
```

```sql
--- за весь период
SELECT
    ds.job_category,
    COUNT(ds.job_category) AS job_category_cnt,
    COUNT(ds.job_category)*100/cnt_all.cnt_all AS percentage
FROM ds_salary ds
JOIN (SELECT
          COUNT(*) as cnt_all
       FROM ds_salary ds)
       AS cnt_all
GROUP BY 1
```



	work_year	job_category	job_category_cnt	percentage
1	2 020	Data Analyst	15	20
2	2 020	Data Engineering	19	25
3	2 020	Data Science	28	37
4	2 020	Machine Learning	7	9
5	2 020	Management	3	4
6	2 020	Other	3	4
7	2 021	Data Analyst	29	12
8	2 021	Data Architecture	5	2
9	2 021	Data Engineering	55	24
10	2 021	Data Science	68	29
11	2 021	Machine Learning	33	14
12	2 021	Management	19	8
13	2 021	Other	19	8
14	2 022	Data Analyst	206	18
15	2 022	Data Architecture	31	2
16	2 022	Data Engineering	348	30
17	2 022	Data Science	303	26
18	2 022	Machine Learning	142	12
19	2 022	Management	54	4
20	2 022	Other	41	3
21	2 023	Data Analyst	216	18
22	2 023	Data Architecture	32	2
23	2 023	Data Engineering	369	31
24	2 023	Data Science	299	25
25	2 023	Machine Learning	141	12
26	2 023	Management	55	4
27	2 023	Other	44	3

	job_category	job_category_cnt	percentage
1	Data Analyst	466	18
2	Data Architecture	68	2
3	Data Engineering	791	30
4	Data Science	698	27
5	Machine Learning	323	12
6	Management	131	5
7	Other	107	4



1. ОБЩИЙ АНАЛИЗ

Процентное соотношение уровня градаций специалистов

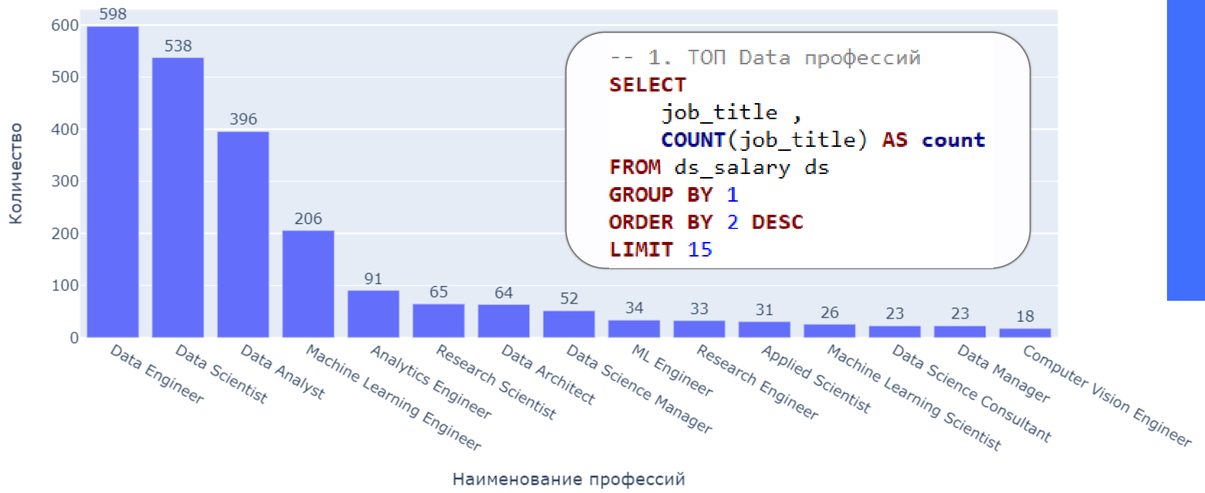
```sql
-- Процентное соотношение уровня опыта по годам
CREATE VIEW proc_exp_year AS
SELECT
    ds.work_year,
    ds.experience_level,
    COUNT(ds.experience_level) AS count_exp_lvl,
    COUNT(ds.experience_level)*100/wyc.cnt AS percentage
FROM ds_salary ds
JOIN (
    SELECT
        work_year,
        COUNT(work_year) AS cnt
    FROM ds_salary ds
    GROUP BY 1)
    AS wyc
ON ds.work_year = wyc.work_year
GROUP BY 1,2
```

	work_year	experience_level	count_exp_lvl	percentage
1	2 020	Executive	3	4
2	2 020	Junior	22	29
3	2 020	Middle	32	42
4	2 020	Senior	18	24
5	2 021	Executive	10	4
6	2 021	Junior	55	24
7	2 021	Middle	90	39
8	2 021	Senior	73	32
9	2 022	Executive	39	3
10	2 022	Junior	110	9
11	2 022	Middle	288	25
12	2 022	Senior	688	61
13	2 023	Executive	44	3
14	2 023	Junior	83	7
15	2 023	Middle	254	21
16	2 023	Senior	775	67

1. ОБЩИЙ АНАЛИЗ

Тенденции в Data-профессиях

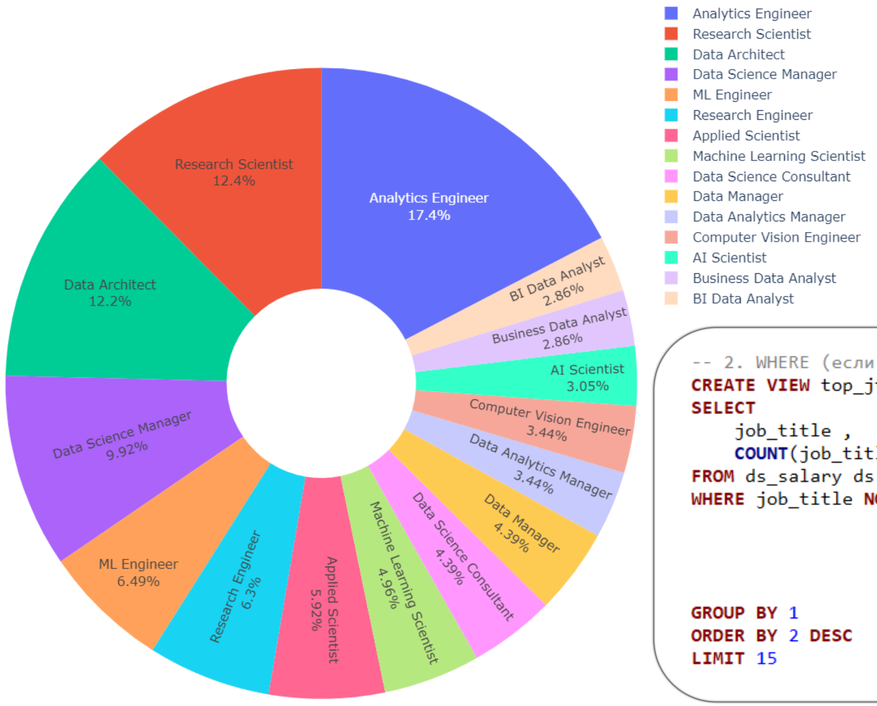
Тенденции в Data-профессиях




```sql
-- 1. ТОП Data профессий
SELECT
    job_title ,
    COUNT(job_title) AS count
FROM ds_salary ds
GROUP BY 1
ORDER BY 2 DESC
LIMIT 15
```

```sql
-- 4. Категории + профессии (с условием)
CREATE VIEW cat_title_2 AS
SELECT
    job_category,
    job_title,
    COUNT(job_title) AS count
FROM ds_salary ds
WHERE job_title NOT IN ('Data Engineer',
                        'Data Scientist',
                        'Data Analyst',
                        'Machine Learning Engineer',
                        'Data Architect')
GROUP BY 1,2
ORDER BY 3 DESC
```

Top 15 Data-профессий




```sql
-- 2. WHERE (если категория = профессия)
CREATE VIEW top_jt AS
SELECT
    job_title ,
    COUNT(job_title) AS count
FROM ds_salary ds
WHERE job_title NOT IN ('Data Engineer',
                        'Data Scientist',
                        'Data Analyst',
                        'Machine Learning Engineer',
                        'Data Architect')

GROUP BY 1
ORDER BY 2 DESC
LIMIT 15
```

	job_category	work_year	avarage_salary	y_y_changes_perc	changes_perc
1	Data Analyst	2 020	54 047	[NULL]	0
2	Data Analyst	2 021	77 373	43,16	43,16
3	Data Analyst	2 022	99 906	29,12	84,85
4	Data Analyst	2 023	109 859	9,96	103,27
5	Data Architecture	2 021	169 941	[NULL]	0
6	Data Architecture	2 022	165 752	-2,46	-2,46
7	Data Architecture	2 023	165 245	-0,31	-2,76
8	Data Engineering	2 020	84 025	[NULL]	0
9	Data Engineering	2 021	93 439	11,2	11,2
10	Data Engineering	2 022	136 759	46,36	62,76
11	Data Engineering	2 023	152 024	11,16	80,93
12	Data Science	2 020	102 120	[NULL]	0
13	Data Science	2 021	88 230	-13,6	-13,6
14	Data Science	2 022	132 341	50	29,59
15	Data Science	2 023	154 841	17	51,63
16	Machine Learning	2 020	129 966	[NULL]	0
17	Machine Learning	2 021	102 898	-20,83	-20,83
18	Machine Learning	2 022	137 146	33,28	5,52
19	Machine Learning	2 023	164 676	20,07	26,71
20	Management	2 020	210 768	[NULL]	0
21	Management	2 021	149 744	-28,95	-28,95
22	Management	2 022	163 471	9,17	-22,44
23	Management	2 023	162 247	-0,75	-23,02
24	Other	2 020	64 299	[NULL]	0
25	Other	2 021	46 119	-28,27	-28,27
26	Other	2 022	115 514	150,47	79,65
27	Other	2 023	144 892	25,43	125,34

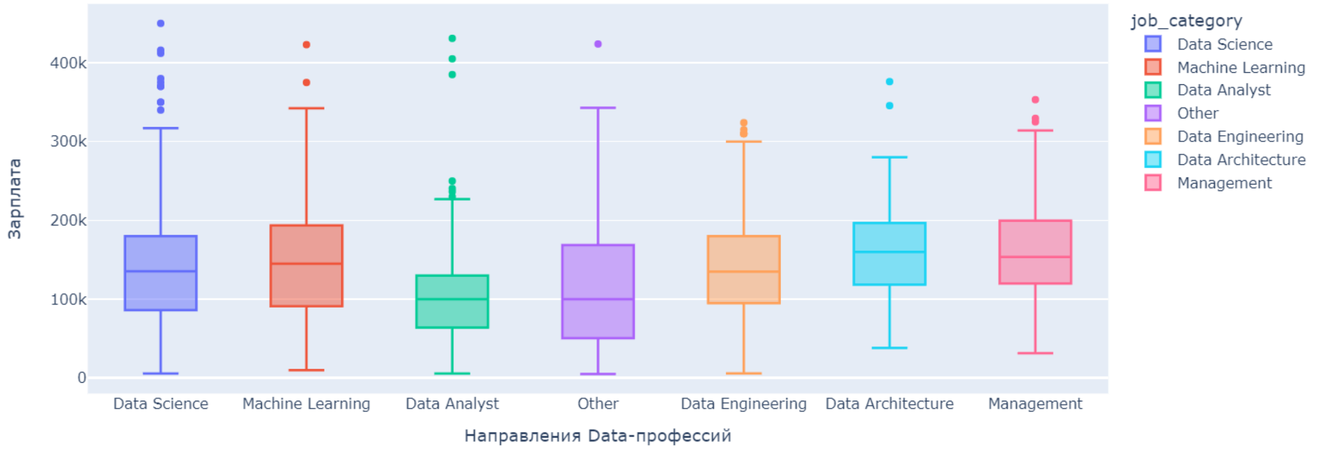
2. ИЗМЕНЕНИЕ З/П

Определение средней зарплаты по Data-профессиям. Анализ изменения з/п год к году

```sql
-- Максимальные зарплаты в направлениях по годам
SELECT
    ds.work_year,
    ds.job_category,
    ds.job_title,
    MAX(salary_in_usd) AS max_s
FROM ds_salary ds
GROUP BY 1,2
ORDER BY 1,4
```

```sql
-- Средние зарплаты в направлениях по годам
SELECT
    ds.work_year,
    ds.job_category,
    ds.job_title,
    ROUND(AVG(salary_in_usd)) AS avg_s
FROM ds_salary ds
GROUP BY 1,2
ORDER BY 1,4
```

Диаграмма размаха з/п по категориям Data-профессий

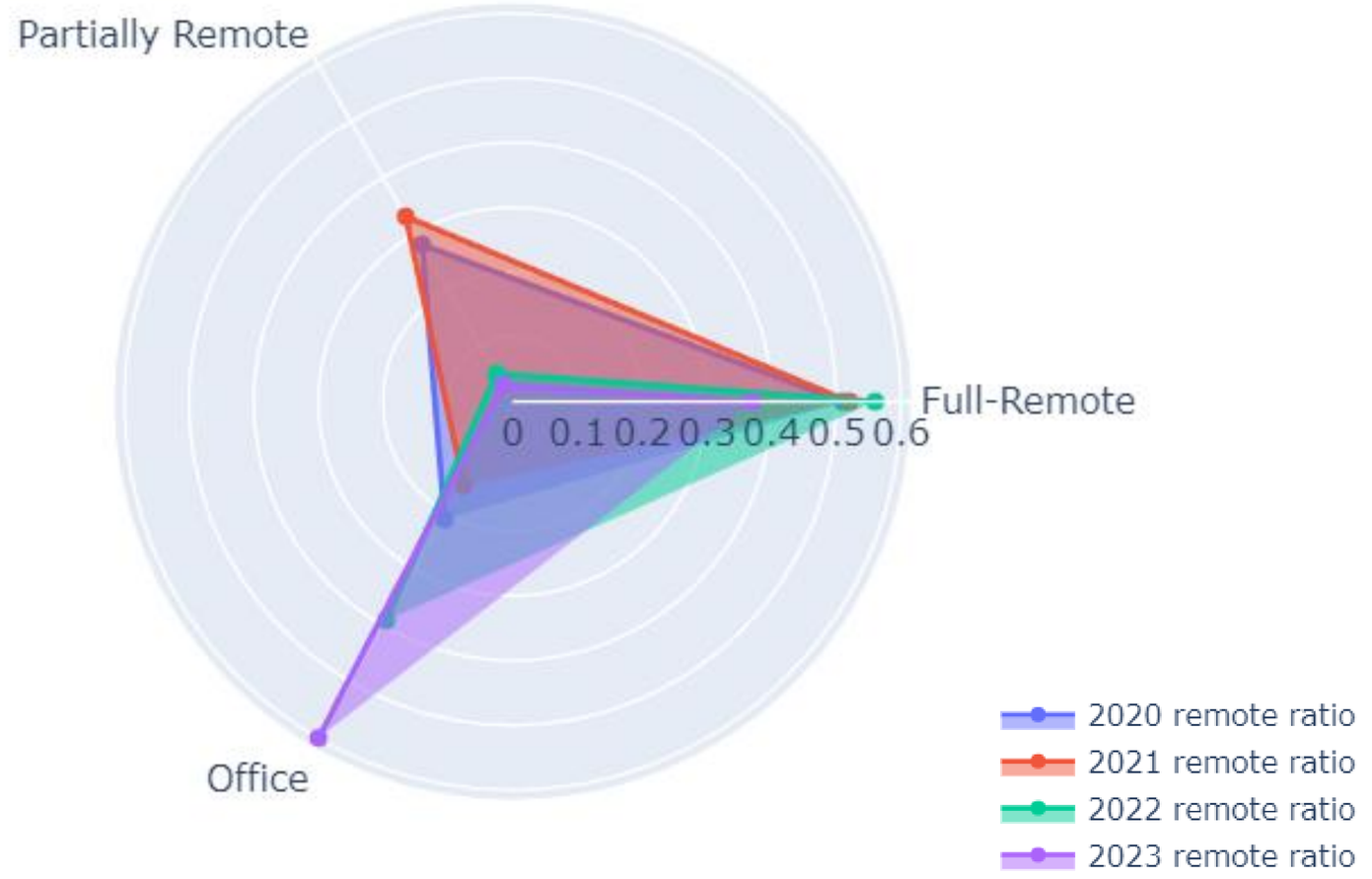



```sql
-- Изменения з/п год к году по направлениям Data профессий
SELECT
    ds.job_category,
    ds.work_year,
    ROUND(AVG(salary_in_usd)) AS avarage_salary,
    ROUND((ROUND(AVG(salary_in_usd)) * 100 / lag(ROUND(AVG(salary_in_usd))) OVER (PARTITION BY job_category ORDER BY job_category) - 100) , 2) AS y_y_changes_perc,
    ROUND((ROUND(AVG(salary_in_usd)) * 100 / first_value(ROUND(AVG(salary_in_usd))) OVER (PARTITION BY job_category ORDER BY job_category) - 100) , 2) AS changes_perc
    FROM ds_salary ds
GROUP BY 1,2
```

3. ФОРМАТ ЗАНЯТОСТИ

Определение изменений в соотношении формата занятости

remote_ratio	work_year	remote_perc	y_y_perc
Full-Remote	2 020	50	[NULL]
Full-Remote	2 021	52	4
Full-Remote	2 022	55	5
Full-Remote	2 023	36	-35
Half-Remote	2 020	28	[NULL]
Half-Remote	2 021	32	14
Half-Remote	2 022	5	-85
Half-Remote	2 023	2	-60
Office	2 020	21	[NULL]
Office	2 021	14	-34
Office	2 022	38	171
Office	2 023	60	57




```sql
-- Как менялось процентное соотношение "удаленки"
SELECT
    ds.remote_ratio,
    ds.work_year,
    COUNT(ds.remote_ratio)*100/wyc.cnt AS remote_perc,
    (COUNT(ds.remote_ratio)*100/wyc.cnt) * 100 / lag(COUNT(ds.remote_ratio)*100/wyc.cnt) OVER (PARTITION BY ds.remote_ratio ORDER BY ds.work_year) - 100 AS y_y_perc
FROM ds_salary ds
JOIN (SELECT work_year, count(work_year) AS cnt FROM ds_salary ds GROUP BY 1) AS wyc
ON ds.work_year = wyc.work_year
GROUP BY 1,2
```

	job_category	company_size	min_s	avg_s	max_s
1	Data Analyst	LARGE	6 072	76 659	405 000
2	Data Analyst	MEDIUM	8 050	108 449	430 967
3	Data Analyst	SMALL	5 723	64 201	200 000
4	Data Architecture	LARGE	38 154	158 345	250 000
5	Data Architecture	MEDIUM	63 000	166 818	376 080
6	Data Engineering	LARGE	5 882	107 358	276 000
7	Data Engineering	MEDIUM	7 500	145 433	324 000
8	Data Engineering	SMALL	12 608	81 135	160 000
9	Data Science	LARGE	8 000	116 625	412 000
10	Data Science	MEDIUM	5 707	146 739	450 000
11	Data Science	SMALL	5 679	76 211	416 000
12	Machine Learning	LARGE	12 000	127 818	423 000
13	Machine Learning	MEDIUM	12 000	159 022	375 000
14	Machine Learning	SMALL	10 000	88 053	260 000
15	Management	LARGE	54 094	143 376	325 000
16	Management	MEDIUM	45 600	169 303	353 200
17	Management	SMALL	31 520	103 173	168 000
18	Other	LARGE	5 409	126 085	423 834
19	Other	MEDIUM	5 132	121 172	342 810
20	Other	SMALL	6 304	76 273	275 000

```sql
-- Зависимость уровня з/п по направлениям Data профессий от:
-- 1. размера компании
SELECT
    ds.job_category,
    ds.company_size,
    MIN(ds.salary_in_usd) AS min_s,
    ROUND(AVG(ds.salary_in_usd)) AS avg_s,
    MAX(ds.salary_in_usd) AS max_s
FROM ds_salary ds
GROUP BY 1,2

-- в отрыве от направлений профессий
SELECT
    ds.company_size,
    MIN(ds.salary_in_usd) AS min_s,
    ROUND(AVG(ds.salary_in_usd)) AS avg_s,
    MAX(ds.salary_in_usd) AS max_s
FROM ds_salary ds
GROUP BY 1
```

4. АНАЛИЗ З/П

Выявление зависимости заработной платы от различных факторов

	remote_ratio	work_year	min_s	avg_s	max_s	y_y_perc	y_y_changes_perc
1	Full-Remote	2 020	6 072	102 033	412 000	[NULL]	[NULL]
2	Full-Remote	2 021	5 679	105 021	416 000	2,93	0,3
3	Full-Remote	2 022	5 132	131 690	405 000	25,39	39,02
4	Full-Remote	2 023	15 806	142 135	376 080	7,93	12,92
5	Half-Remote	2 020	5 707	77 591	250 000	[NULL]	[NULL]
6	Half-Remote	2 021	5 409	75 909	423 000	-2,17	0,3
7	Half-Remote	2 022	7 500	85 127	375 000	12,14	39,02
8	Half-Remote	2 023	12 767	72 054	220 000	-15,36	12,92
9	Office	2 020	6 072	93 426	450 000	[NULL]	[NULL]
10	Office	2 021	5 882	92 900	276 000	-0,56	0,3
11	Office	2 022	6 304	134 294	430 967	44,56	39,02
12	Office	2 023	7 000	153 186	423 834	14,07	12,92

```sql
-- Зависимость уровня з/п по направлениям Data профессий от:
-- 2. от типа занятости
-- в отрыве от направлений профессий, но с учетом года
SELECT
    ds.remote_ratio,
    ds.work_year,
    MIN(ds.salary_in_usd) AS min_s,
    ROUND(AVG(ds.salary_in_usd)) AS avg_s,
    MAX(ds.salary_in_usd) AS max_s,
    ROUND(((AVG(ds.salary_in_usd)) * 100 / lag(ROUND(AVG(ds.salary_in_usd)))
        OVER (PARTITION BY ds.remote_ratio ORDER BY ds.work_year) - 100), 2) AS y_y_perc,
    yy.y_y_changes_perc
FROM ds_salary ds
JOIN yy
ON yy.work_year = ds.work_year
GROUP BY 1,2
```

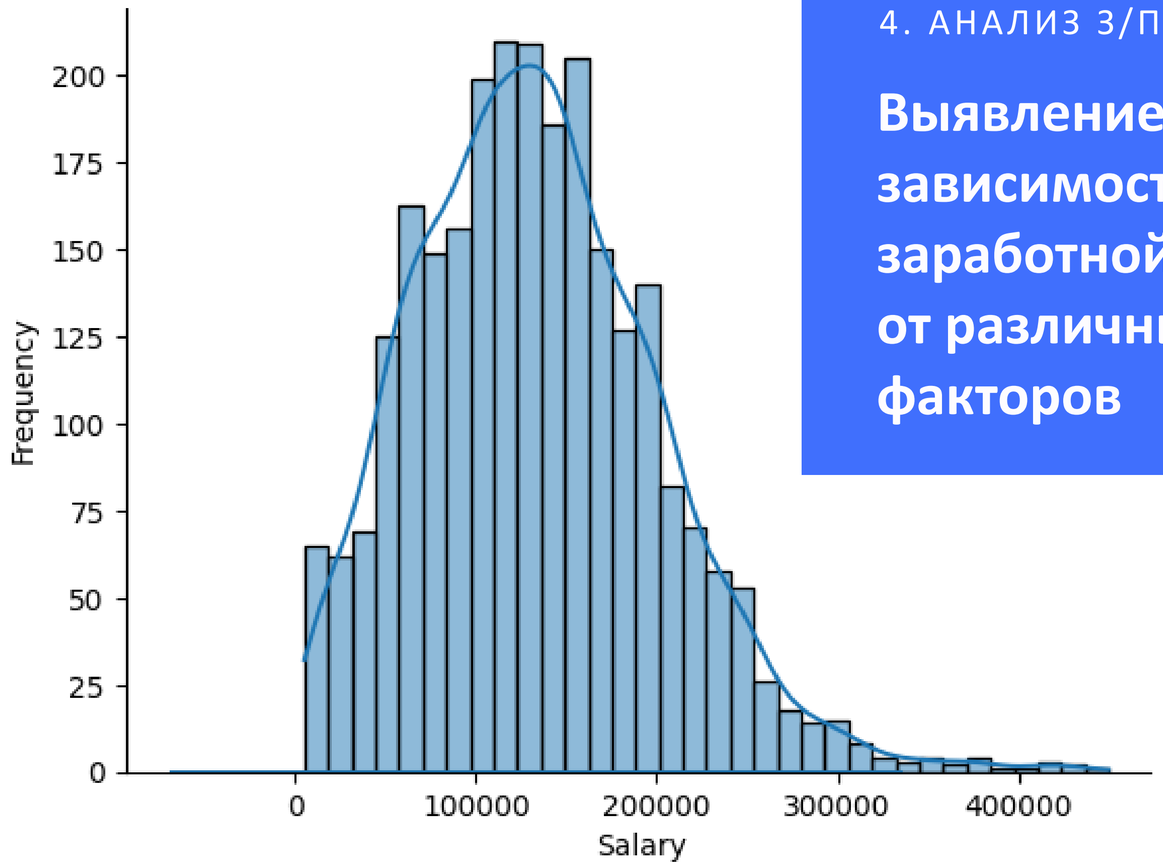
	experience_level	work_year	min_s	avg_s	max_s	y_y_perc	y_y_changes_perc
1	Executive	2 020	15 000	139 944	325 000	[NULL]	[NULL]
2	Executive	2 021	69 741	186 128	416 000	33	0,3
3	Executive	2 022	76 309	183 838	324 000	-1,23	39,02
4	Executive	2 023	100 000	202 107	353 200	9,94	12,92
5	Junior	2 020	5 707	59 512	250 000	[NULL]	[NULL]
6	Junior	2 021	5 409	54 905	225 000	-7,74	0,3
7	Junior	2 022	6 270	69 950	300 000	27,4	39,02
8	Junior	2 023	7 000	91 465	220 000	30,76	12,92
9	Middle	2 020	6 072	87 565	450 000	[NULL]	[NULL]
10	Middle	2 021	5 409	80 711	423 000	-7,83	0,3
11	Middle	2 022	5 132	99 579	430 967	23,38	39,02
12	Middle	2 023	16 414	113 660	340 000	14,14	12,92
13	Senior	2 020	33 511	137 241	412 000	[NULL]	[NULL]
14	Senior	2 021	18 907	126 085	276 000	-8,13	0,3
15	Senior	2 022	8 000	149 573	405 000	18,63	39,02
16	Senior	2 023	15 806	160 743	423 834	7,47	12,92

```sql
-- 3. от уровня профессий
SELECT
    ds.experience_level,
    ds.work_year,
    MIN(ds.salary_in_usd) as min_s,
    ROUND(AVG(ds.salary_in_usd)) as avg_s,
    MAX(ds.salary_in_usd) as max_s,
    ROUND(((AVG(ds.salary_in_usd)) * 100 / lag(ROUND(AVG(ds.salary_in_usd)))
        OVER (PARTITION BY ds.experience_level ORDER BY ds.work_year) - 100), 2) AS y_y_perc,
    yy.y_y_changes_perc
FROM ds_salary ds
JOIN yy
ON yy.work_year = ds.work_year
GROUP BY 1,2
```

```sql
-- Зависимость уровня з/п по направлениям Data профессий от:
-- 4. от локации компании/резиденции
CREATE VIEW loc AS
SELECT
    ROW_NUMBER() OVER (ORDER BY ROUND(AVG(ds.salary_in_usd)) DESC) as loc_id,
    ds.company_location,
    MIN(ds.salary_in_usd) AS min_s,
    ROUND(AVG(ds.salary_in_usd)) AS avg_s,
    MAX(ds.salary_in_usd) AS max_s,
    count(ds.company_location) AS loc_cnt
FROM ds_salary ds
GROUP BY 2
LIMIT 7

CREATE VIEW res AS
SELECT
    ROW_NUMBER() OVER (ORDER BY ROUND(AVG(ds.salary_in_usd)) DESC) as res_id,
    ds.employee_residence,
    MIN(ds.salary_in_usd) AS min_s,
    ROUND(AVG(ds.salary_in_usd)) AS avg_s,
    MAX(ds.salary_in_usd) AS max_s,
    count(ds.employee_residence) AS res_cnt
FROM ds_salary ds
GROUP BY 2
LIMIT 7

SELECT *
FROM loc
JOIN res
ON loc.loc_id = res.res_id
```



4. АНАЛИЗ З/П

Выявление зависимости заработной платы от различных факторов

	loc_id	company_location	min_s	avg_s	max_s	loc_cnt	res_id	employee_residence	min_s	avg_s	max_s	res_cnt
1	1	IL	119 059	271 447	423 834	2	1	IL	423 834	423 834	423 834	1
2	2	PR	135 000	167 500	200 000	4	2	MY	200 000	200 000	200 000	1
3	3	US	5 679	152 375	450 000	1 929	3	PR	135 000	166 000	200 000	5
4	4	RU	85 000	140 333	230 000	3	4	US	24 000	153 972	450 000	1 893
5	5	CA	15 000	130 573	275 000	83	5	CA	10 000	130 860	275 000	81
6	6	NZ	125 000	125 000	125 000	1	6	CN	125 404	125 404	125 404	1
7	7	BA	120 000	120 000	120 000	1	7	NZ	125 000	125 000	125 000	1

5. КОРРЕЛЯЦИЯ

Определение статистической взаимосвязи

ВЫВОДЫ

1. Год к году по данной выборке соотношение смещается в сторону увеличения уровня опыта.
2. Наиболее высокая зарплатная вилка у Менеджмента и Архитекторов данных.
3. Уровень удаленной формы работы стремительно уменьшается после 2021 года.
4. Значительный рост з/п Executive в 2021 году, на фоне снижения зарплаты у специалистов более низкого уровня. Vise versa в 2022 году.
5. Наиболее значительная сила корреляции с уровнем заработной платы наблюдается у уровня опыта Senior. Степень связи – умеренная.