



Stack Overflow Developer Survey 2019: Key Findings

Radmir Sagadiyev
March 2023

IBM Developer

SKILLS NETWORK 

OUTLINE



- Executive Summary
- Introduction
- Methodology
- Results
 - Visualization – Charts
- Conclusion
- Appendix

EXECUTIVE SUMMARY



- Key findings from an analysis of data collected as part of the **2019 Stack Overflow Developer Survey** summarized in this presentation:
- There is an increase in interest in non-relational databases, namely MongoDB, which likely reflects a growing need to handle non-relational and unstructured data. Current and aspiring data analysts should develop competence in NoSQL in addition to SQL database programs.
- Javascript and HTML/CSS remain the dominant languages for web development, however, TypeScript may show positive trend.
- Python's rising popularity likely reflects the growth of AI and ML work.

INTRODUCTION



- **Data used:**
 - Stack Overflow Annual Developer Survey.
 - <https://hh.kz/parsing>
- The primary objective of using Stack Overflow Annual Developer Survey is to analyze data regarding technology usage and trends among developers.
- **In appendix:**
 - Being a resident of Kazakhstan, I am aimed to find out what data science tools are popular in CIS countries by <https://hh.kz/parsing>

METHODOLOGY. [stack overflow Annual Developer Survey.](#)

Python was used for data wrangling. You can see some codes below. For the full process please refer to attached 'ipynb' files.

| | |
|-------------------------------|--|
| Finding duplicates | <pre># Find how many duplicate rows exist in the dataframe. survey_df['has_duplicates'] = survey_df.duplicated(['Respondent'], keep = False) survey_df.value_counts('has_duplicates') has_duplicates False 11305 True 247 dtype: int64</pre> |
| Finding Missing values | <pre># Find out how many rows are missing in the column 'Country' missing = survey_df['Country'].isnull() survey_df[missing].shape (0, 85)</pre> |

METHODOLOGY. stack overflow Annual Developer Survey.

Normalizing data

Create a new column named 'NormalizedAnnualCompensation'.

```
import numpy as np
# create a list of our conditions
conditions = [
    (survey_df['CompFreq'] == 'Yearly'),
    (survey_df['CompFreq'] == 'Monthly'),
    (survey_df['CompFreq'] == 'Weekly'),
]

# create a list of the values we want to assign for each condition
values = [survey_df['CompTotal'], survey_df['CompTotal']*12, survey_df['CompTotal']*52]

# create a new column and use np.select to assign values to it using our lists as arguments
survey_df['NormalizedAnnualCompensation'] = np.select(conditions, values)
```

METHODOLOGY. stack overflow Annual Developer Survey.

IQR

Find out the Inter Quartile Range for the column `NormalizedAnnualCompensation`.

```
Q1= survey_df['NormalizedAnnualCompensation'].quantile(q=0.25)
Q3= survey_df['NormalizedAnnualCompensation'].quantile(q=0.75)
IQR= Q3 - Q1
IQR
```

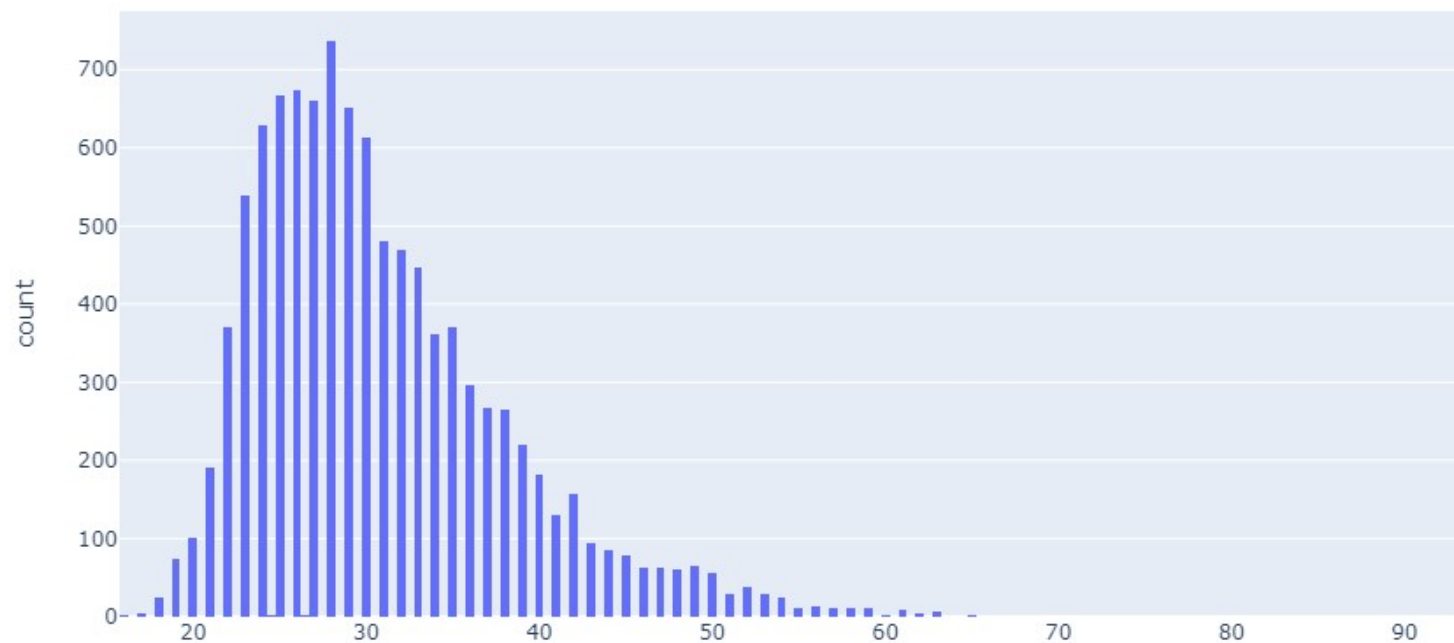
274000.0

```
# Find out the upper and lower bounds
upper_bound = Q3+IQR*1.5
lower_bound = Q1-IQR*1.5

print(upper_bound)
print(lower_bound)
```

RESULTS ■ stack Overflow Annual Developer Survey.

Distribution of Age



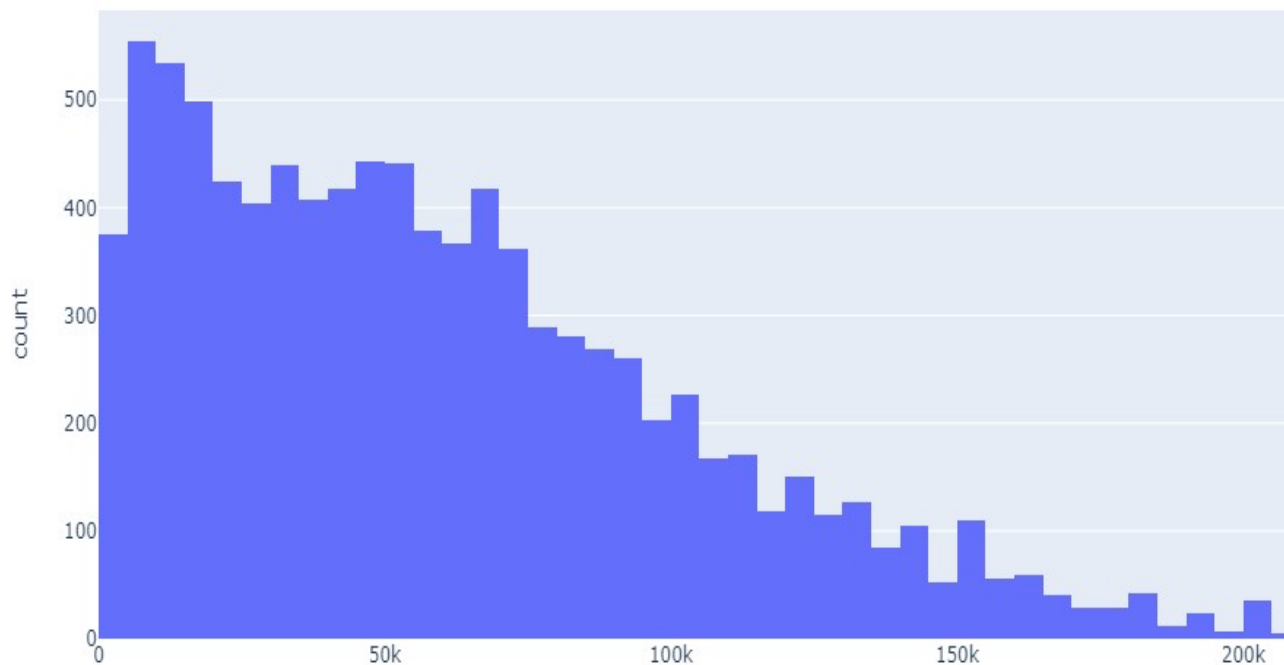
The median age of respondents is 29.

Max age is 99.

Min age is 16.

RESULTS ■ stack overflow Annual Developer Survey.

Distribution of Salary

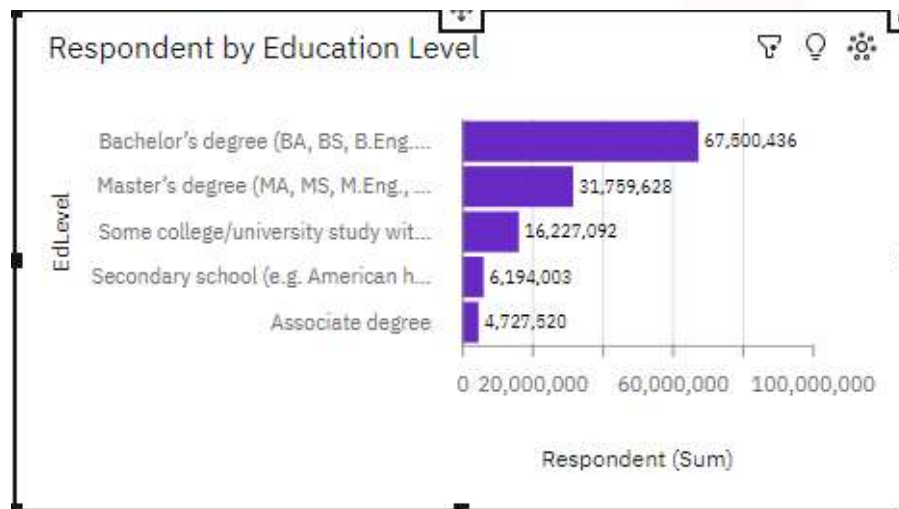


The median salary of respondents is 53K.

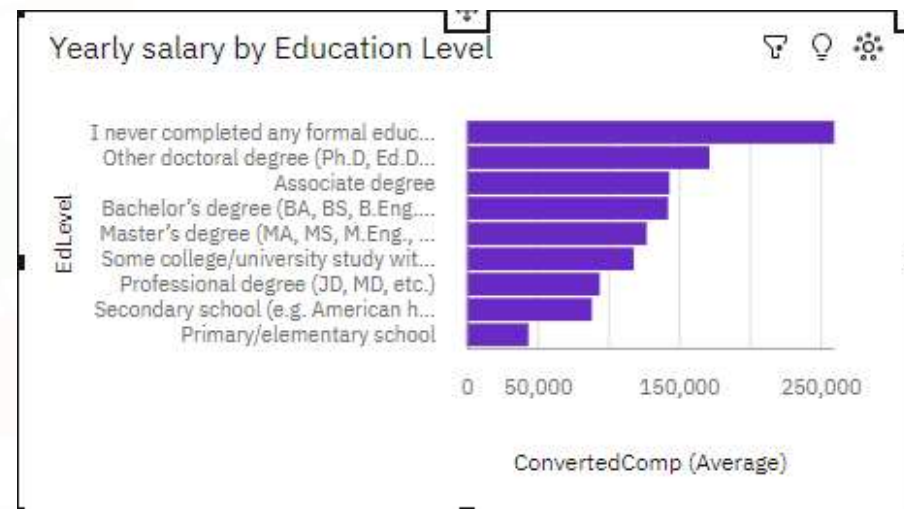
Max salary is 209K.

Min salary value is 4.

RESULTS ■ stack overflow Annual Developer Survey.



Majority of respondents have Bachelor's degree, followed by Master's degree holders

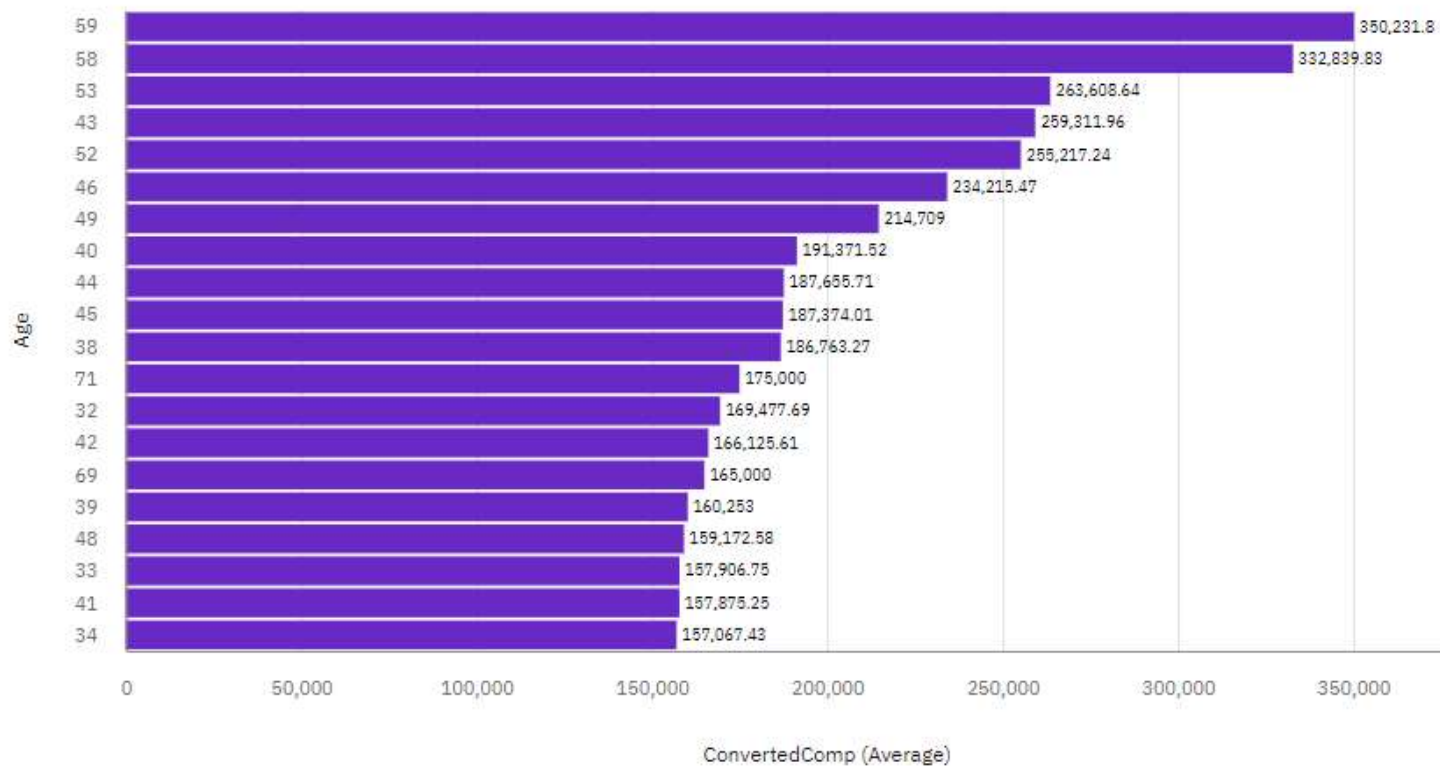


Maximum average amount of salary belongs to respondents without formal education (only 26 respondents), followed by doctoral degree holders (245 respondents).

The average salary of bachelors and masters ranks 4th and 5th respectively

RESULTS ■ stack overflow Annual Developer Survey.

Average salary by Age

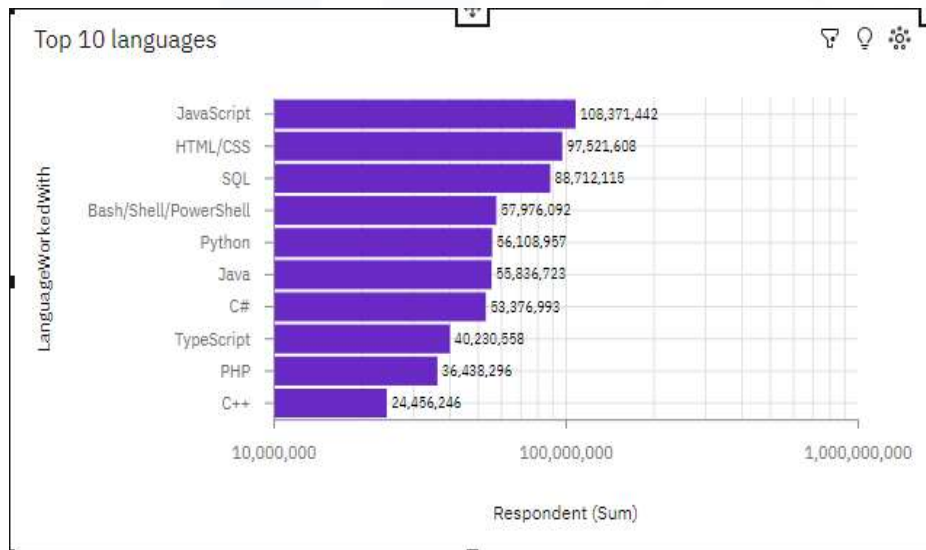


Top 3 average salaries belongs to respondents over 50 years old

Top 10 - belongs to respondents over 40 years old

PROGRAMMING LANGUAGE TRENDS

Fact



Next year expectations

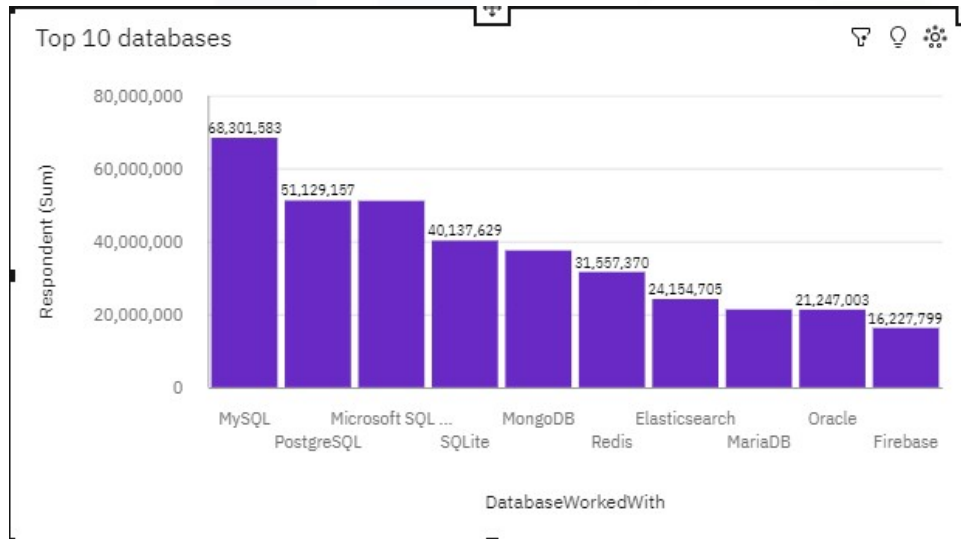


JavaScript and HTML/CSS continue to be ranked in top 3 most popular programming languages for 2019 year and the next.

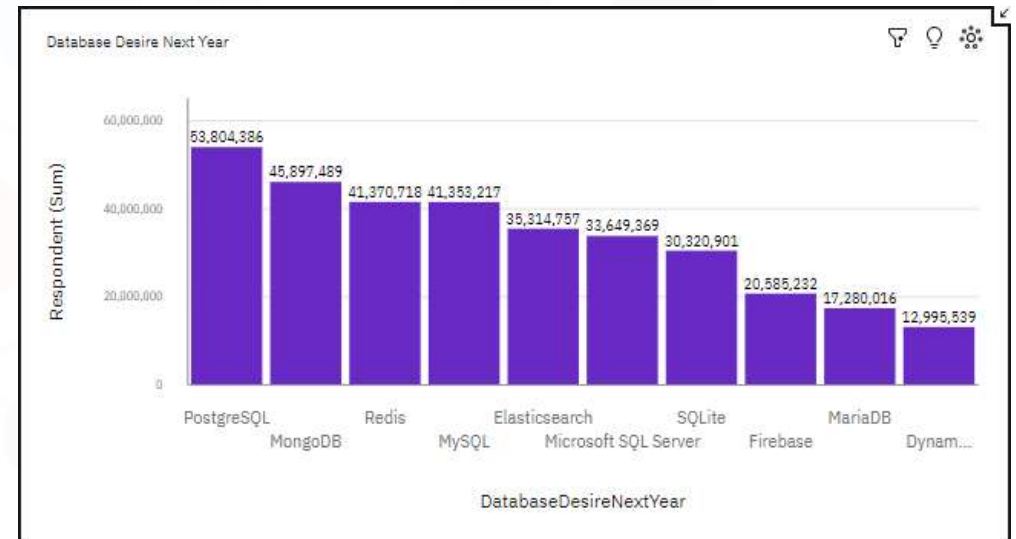
Python and TypeScript have gained more interest.

DATABASE TRENDS

Fact



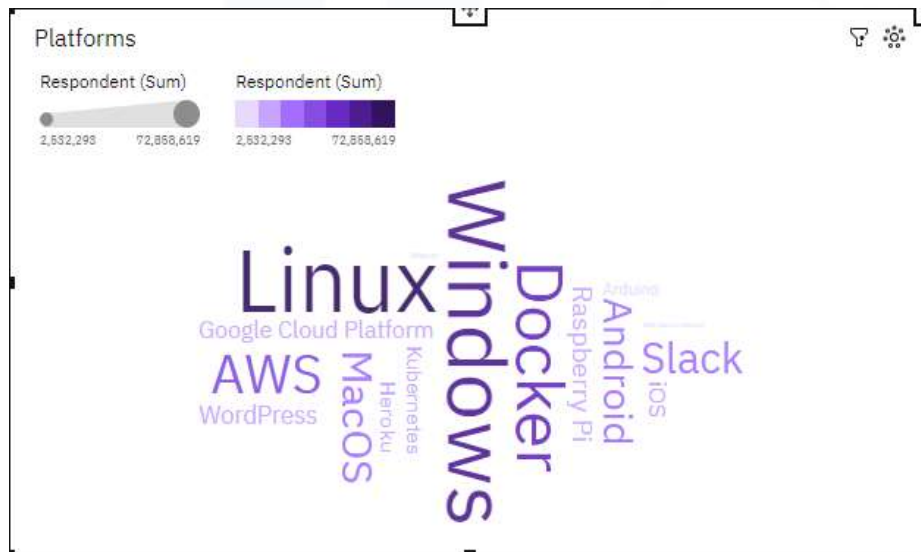
Next year expectations



MySQL, PostgreSQL, and Microsoft SQL are top 3 database in use.
There is an increase in interest in non-rational databases, namely MongoDB

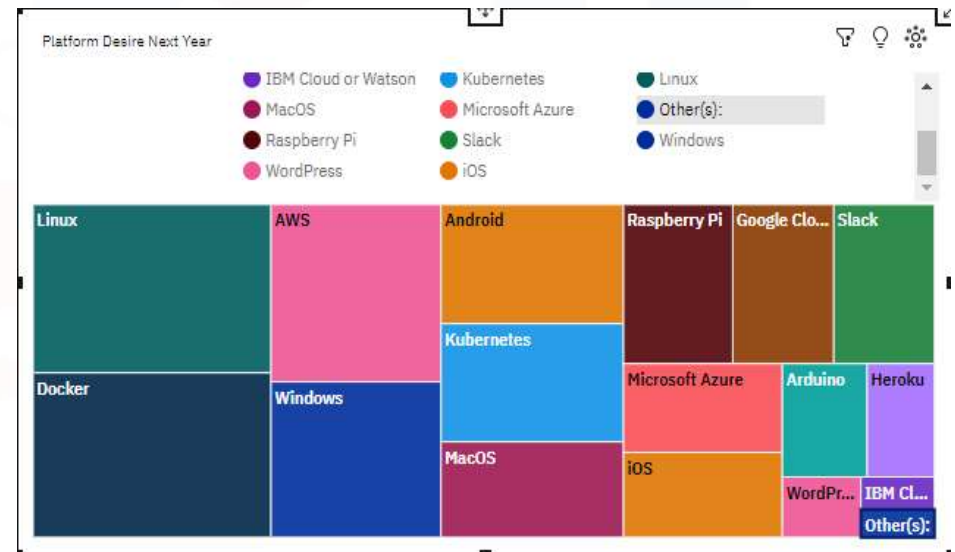
PLATFORMS TRENDS

Fact



Windows, Linux, and Docker are top 3 platforms in use.

Next year expectations



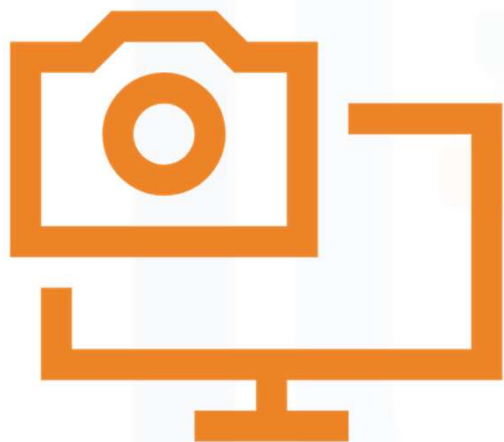
There is an increase in interest in AWS platform.

CONCLUSION. Stack Overflow Annual Developer Survey.



- There is an increase in interest in non-relational databases, namely MongoDB, which likely reflects a growing need to handle non-relational and unstructured data.
- Current and aspiring data analysts should develop competence in NoSQL in addition to SQL database programs.
- There is an increase in interest in Python, TypeScript AWS platform.
- The median age of respondents is 29, the median salary of respondents is 58K.
- Majority of respondents have Bachelor's degree, followed by Master's degree holders
- Maximum average amount of salary belongs to respondents without formal education (only 26 respondents), followed by doctoral degree holders (245 respondents).
- Top 3 average salaries belongs to respondents over 50 years old
- Top 10 average salaries belongs to respondents over 40 years old

APPENDIX



Brief analysis of data science tools in CIS countries.

METHODOLOGY in Appendix.

<https://hh.kz/> parsing.

You can see some Python codes below. For the full process please refer to attached 'ipynb' files.

Scraping

#это функция которая забирает со страницы по указанному url и параметрам html.

```
def get_html_ds(url, params=' '):  
    r = requests.get(url, headers=HEADERS, params=params)  
    return r
```

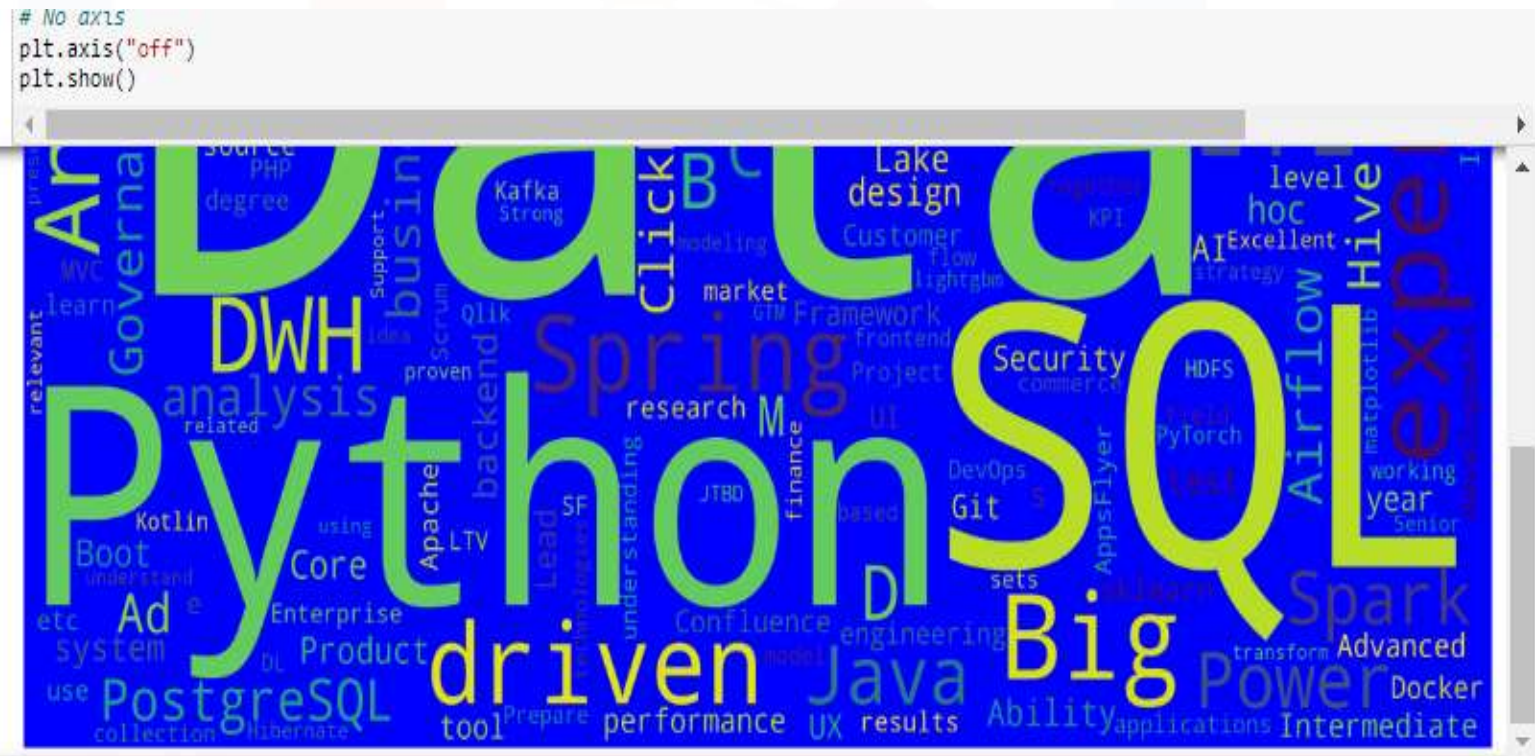
#теперь будем работать точно с html, будем вытягивать контент

```
def get_content_ds(html):  
    soup = BeautifulSoup(html, 'html.parser')  
    items = soup.find_all('div', class_="vacancy-serp-item__layout")  
    vacancies_ds = []  
  
    for item in items:  
        vacancies_ds.append({'job_title': item.find('div').find('a', class_='serp-item__title').get_text(strip=True),  
                             'city': item.find('div', class_='vacancy-serp-item-company').get_text(strip=True),  
                             'company': item.find('div', class_='vacancy-serp-item__info').find('div', class_='text').get_text(strip=True),  
                             'text': item.find('div', class_='g-user-content').get_text(strip=True)})  
  
    return vacancies_ds
```

METHODOLOGY in Appendix.

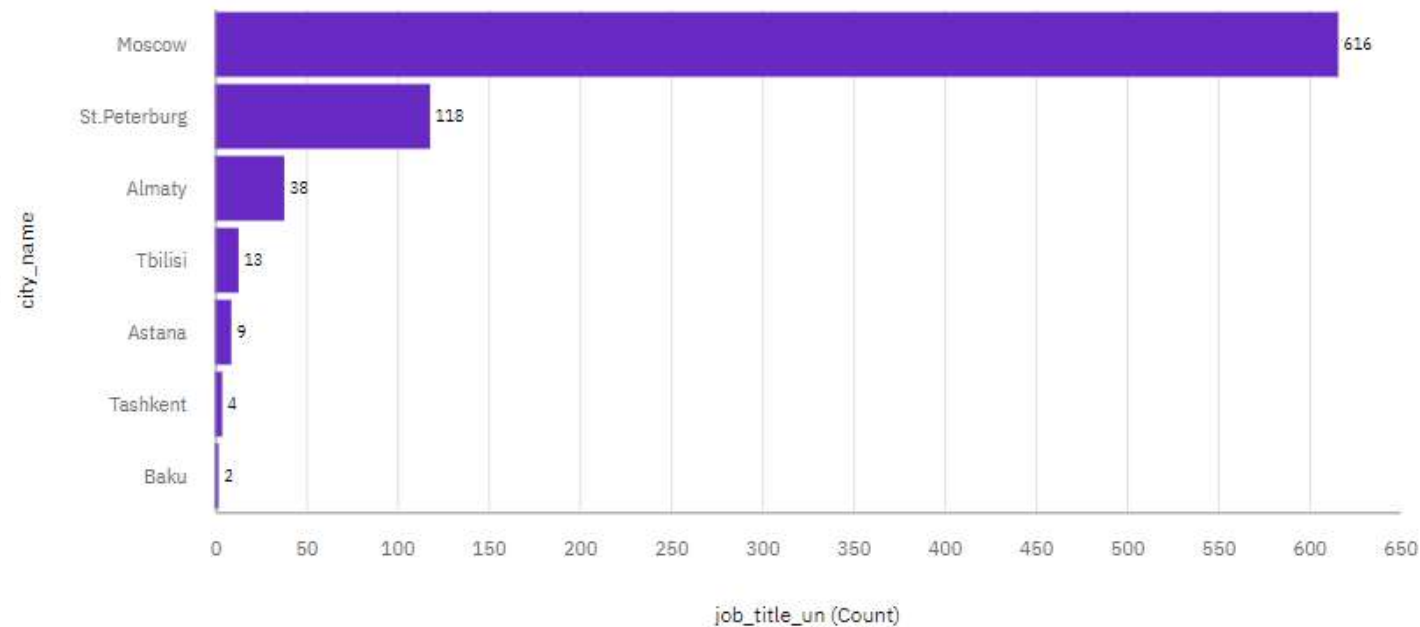
<https://hh.kz/> parsing.

Wordcloud viz



JOB POSTINGS IN CIS . <https://hh.kz/> parsing.

Number of vacabcies by cities



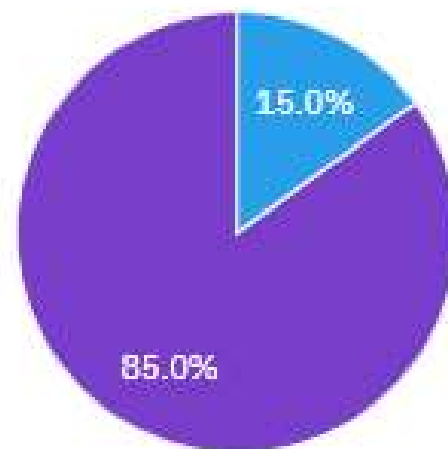
Most vacancies are concentrated in Moscow, which is not surprising, given the preponderance in population compared to Kazakhstan, Georgia and others.

JOB POSTINGS IN CIS . <https://hh.kz/> parsing.

Data analyst and data scientist positions

job_title_un

● Data Scientist ● Data Analyst



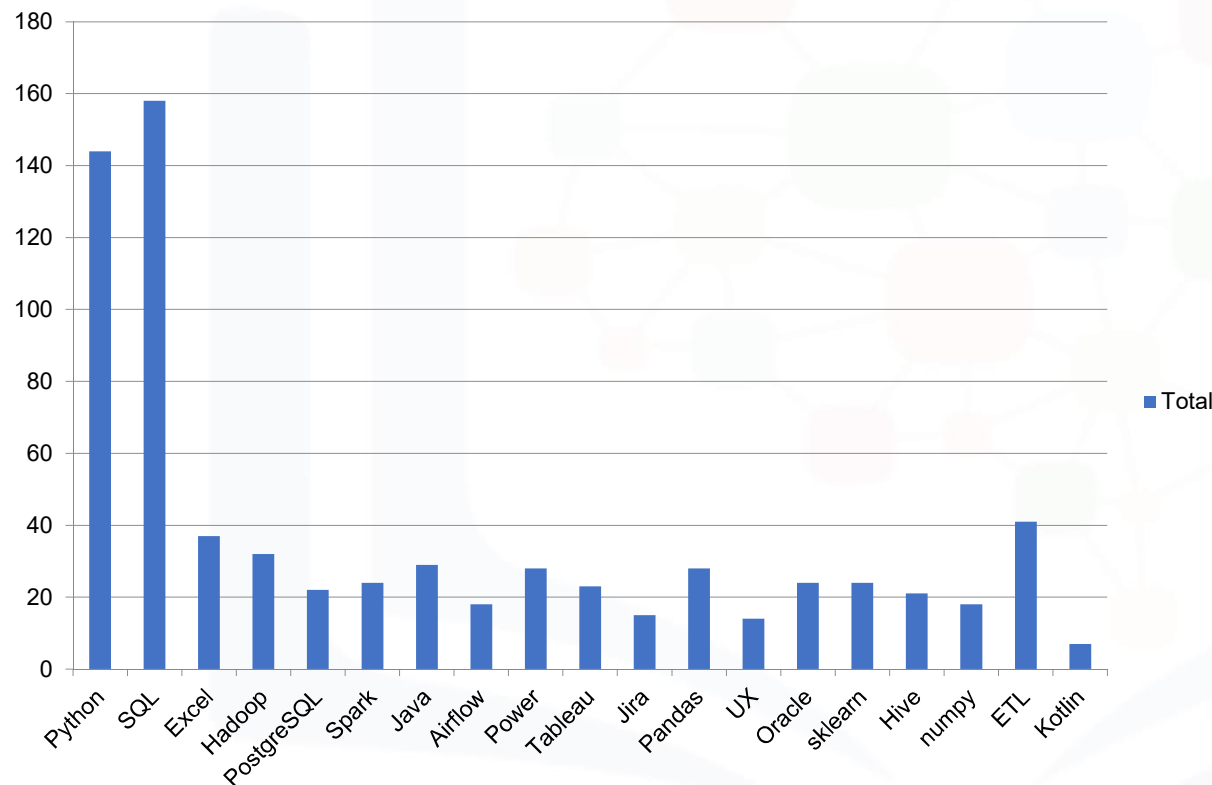
85% of scraped data relates to Data analyst position.

Data analyst are more in demand in CIS countries at the moment of parsing

JOB POSTINGS IN CIS .

<https://hh.kz/> parsing.

Tools

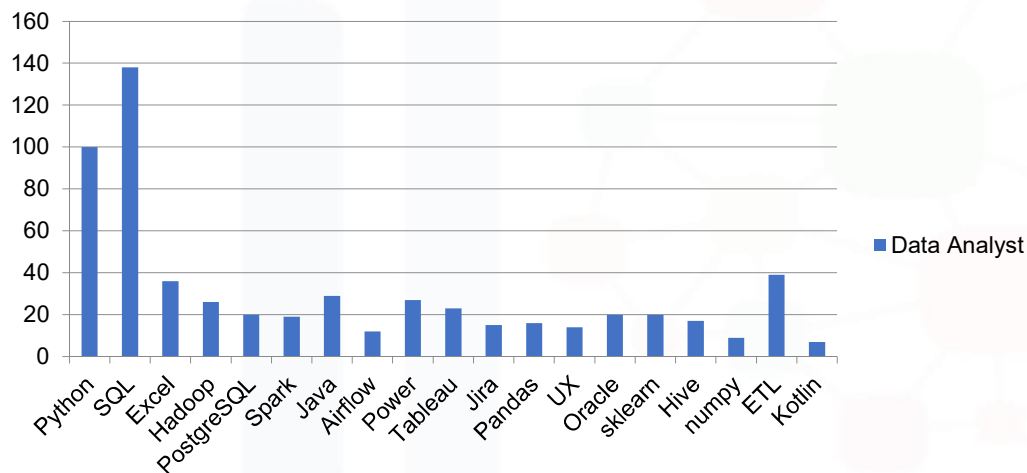


In CIS countries leading tools are SQL, then Python, experience and knowledge of the Extract-Transform-Load process is also important

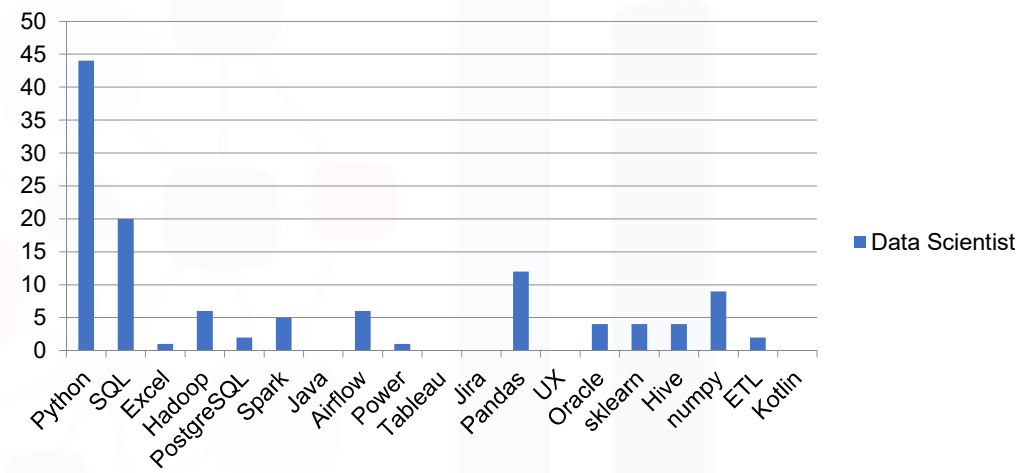
Vacancies with experience in Power BI slightly exceed the number of vacancies that require Tableau

JOB POSTINGS IN CIS . <https://hh.kz/> parsing.

Data Analyst



Data Scientist



You can see that for a Data Analyst job, SQL is the most in-demand tool, while for a Data scientist job, Python is more in demand.

It can also be noted that knowledge of Power BI and Tableau is not critical for Data scientist position.

THANK YOU



IBM Developer

SKILLS NETWORK 