

1. Скачали данные, посмотрели на их размерность — получили ответы на вопросы:

Сколько вопросов было в опросе? 154

Сколько разработчиков приняло участие в нем? 51392

2. Сосчитайте число участников опроса по странам (колонка Country). Выведите 10 стран с наибольшим числом респондентов. Покажите число респондентов и их долю в общем количестве. Прокомментируйте результаты.

	Number	Ratio
United States	11455	0.222895
India	5197	0.101125
United Kingdom	4395	0.085519
Germany	4143	0.080616
Canada	2233	0.043450
France	1740	0.033857
Poland	1290	0.025101
Australia	913	0.017765
Russian Federation	873	0.016987
Spain	864	0.016812

Большинство участников опроса из США (практически пятая часть). На эти 10 стран приходится 64% участников

3. Что если число участников определяется просто населением стран, в которых они живут? Сосчитайте и выведите 10 стран для которых отношение числа респондентов к населению страны наибольшее? Для определенности население стран мира возьмите на страничке

https://en.wikipedia.org/wiki/List_of_countries_and_dependencies_by_population. Для простоты при объединении наборов данных ограничьтесь странами, число участников в которых не меньше ста. Прокомментируйте результаты.

Скачали таблицу из википедии, обработали при помощи BeautifulSoup, смержили с таблицей в которой каждой стране сопоставлено число участников опроса.

Получили ответ:

	Responses	Population	Ratio
Slovenia	303	2065890.000000	0.000147
Ireland	345	4792500.000000	0.000072
Switzerland	595	8465234.000000	0.000070
United Kingdom	4395	65648000.000000	0.000067
Israel	575	8815980.000000	0.000065
Lithuania	176	2807495.000000	0.000063
Sweden	611	10112669.000000	0.000060
Canada	2233	37015700.000000	0.000060
Austria	477	8823054.000000	0.000054
Finland	287	5509984.000000	0.000052

Из лидеров первой таблицы есть только United Kingdom. Значит отношение числа респондентов к населению страны не константа.

4. Покажите, какими системами управления версиями (колонка VersionControl) пользуются участники опроса (упорядочите их по количеству ответов).

Используя команду `value_counts` получим нужный нам ответ

Git	21266
Subversion	2790
Team Foundation Server	2255
I don't use version control	1468
I use some other system	924
Zip file back-ups	609
Mercurial	591
Copying and pasting files to network shares	510
Visual Source Safe	196
Rational ClearCase	121

5. Теперь посмотрим какие языки программирования популярны. В колонке HaveWorkedLanguage перечислены языки, которыми респондент недавно пользовался. Проблема в том, что можно было отмечать несколько вариантов ответа (они разделены точкой с запятой):

- 1) Напишем функцию `get_uniq_words`, которая для каждого участника выдаст множество языков которое он использует в работе
- 2) Применив ее к датасету и объединив результаты получим все множество языков которые используют участники опроса

`['VBA', 'Java', 'Objective-C', 'Ruby', 'CoffeeScript', 'Erlang', 'Swift', 'Hack', 'Elixir', 'Go', 'F#', 'Julia', 'C', 'Haskell', 'Assembly', 'TypeScript', 'Lua', 'Groovy', 'R', 'Clojure', 'Visual Basic 6', 'Python', 'C++', 'Perl', 'PHP', 'VB.NET', 'Matlab', 'Scala', 'Dart', 'C#', 'Common Lisp', 'Smalltalk', 'SQL', 'JavaScript', 'Rust']`

6. Теперь сосчитайте, сколько респондентов пользуются каждым из указанных языков. Покажите 10 наиболее популярных. Есть ли среди них Python?

- 1) Для каждого языка посчитаем сколько раз он встречается в ответах
- 2) Создадим датафрейм с результатами и выведем топ языков

```
counts
languages
JavaScript 22875
SQL         18754
Java        14524
C#          12476
Python      11704
PHP         10290
C++         8155
C           6974
TypeScript  3488
Ruby        3324
```

Среди них есть и питон

7. Интересно, какие языки программирования популярны в разных странах? Для каждой страны, представленной в опросе найдите наиболее популярный язык программирования. Выведите его для 10 стран с наибольшим числом респондентов.

Вывели самые популярные языки в каждой стране,

United States	JavaScript
India	JavaScript
United Kingdom	JavaScript
Germany	JavaScript
Canada	JavaScript
France	JavaScript
Poland	JavaScript
Australia	JavaScript
Russian Federation	JavaScript
Spain	JavaScript

В 10 странах с наибольшим числом респондентов популярен JavaScript

Если посомтреть страны, где JavaScript не на первом месте, то получим список:

South Korea	Java
Albania	SQL
Morocco	SQL
Lebanon	SQL
Saudi Arabia	SQL
Malta	C#
Jordan	SQL
Ethiopia	PHP
Myanmar	Java
Uzbekistan	SQL

На первом месте **South Korea**

8. Придумайте интересный для вас вопрос, относящийся к исследуемому набору данных, и ответьте на него.

Вопрос: какова средняя, стандартное отклонение и медианная зарплаты в каждой стране? (учесть только те страны, из которых ответило более 100 респондентов)

Вывести топ:

	Counts	Mean_Salary	Std_Salary	Median_Salary
Country				
United States	3781	96195	34866	93000
Switzerland	150	91190	36940	90000
Australia	256	69216	25825	66856

Israel	151	62686	38556	70000
Canada	667	56925	21462	53030
United Kingdom	1386	52602	22839	48750
Germany	994	51616	21880	52903
Netherlands	216	49026	21747	48387
Sweden	195	48765	23598	49548
Austria	132	45299	17789	45161

Вывод самые большие зарплаты программисты получают в США и Швейцарии.