Homework 4
Pavlyuk Lyuba

Задача 1

Загрузила файл с данными опроса. Прочитала данные из файла внутри полученного архива:

```
['DeveloperSurvey2017QuestionaireCleaned.pdf', '__MACOSX/', '__MACOSX/._DeveloperSurvey2017QuestionaireCleaned.pdf',
'README_2017.txt', '__MACOSX/._README_2017.txt', 'survey_results_public.csv', 'survey_results_schema.csv']
```

survey_results_public.csv с ответами и survey_results_schema.csv с вопросами.

Всего 154 вопроса было в опросе и 51392 разработчиков приняло участие в нем.

Задача 2

Сосчитала число участников опроса по разным странам:

```
: survey.Country.value_counts()[0:10]
```

```
: United States          11455
  India                   5197
  United Kingdom          4395
  Germany                 4143
  Canada                  2233
  France                  1740
  Poland                  1290
  Australia                913
  Russian Federation       873
  Spain                    864
  Name: Country, dtype: int64
```

Сосчитала их долю в общем количестве:

```
survey.Country.value_counts()[0:10]/len(survey)
```

```
United States          0.22
India                  0.10
United Kingdom         0.09
Germany                0.08
Canada                 0.04
France                 0.03
Poland                 0.03
Australia              0.02
Russian Federation     0.02
Spain                  0.02
```

Большую часть опрошенных составляют жители США – 22%, также первые пять стран (США, Индия, Великобритания, Германия, Канада) занимают около 53% от всех опрошенных участников.

Задача 3

При помощи библиотеки BeautifulSoup скачала материал страницы из Википедии, далее по патерну td нашла таблицу и выделила в нее необходимые элементы: страну и население.

	Country	Population
0	China	1,389,190,000
1	India	1,327,810,000
2	United States	326,615,000
3	Indonesia	261,890,900
4	Pakistan	210,564,000

Далее соединала с таблицей survey_results_public.csv по стране и нашла отношение числа респондентов к населению страны:

	Country	People	Respondent	ratio
129	Slovenia	2065890.00000000	303	0.00014667
111	Ireland	4792500.00000000	345	0.00007199
92	Switzerland	8465234.00000000	595	0.00007029
19	United Kingdom	65648000.00000000	4395	0.00006695
91	Israel	8815980.00000000	575	0.00006522
121	Lithuania	2807495.00000000	176	0.00006269
84	Sweden	10112669.00000000	611	0.00006042
35	Canada	37015700.00000000	2233	0.00006033
90	Austria	8823054.00000000	477	0.00005406
106	Finland	5509984.00000000	287	0.00005209

Только Великобритания вошла в топ 10 относительно общего числа опрошенных и процентного соотношения опрошенных к жителям в стране.

Задача 4

Применили value_counts к нашей таблице.

```
Out[20]: 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
         Name: VersionControl, dtype: int64
```

***Задача* 5**.

Создадим множество, в которое будем дописывать языки всех участников поочереди

```
print(list(Language))
```

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

***Задача* 6.**

Для каждого языка поочереди сосчитаем сколько раз он встречался в колонке HaveWorkedLanguage и отсортируем полученные значения

	Language	counts
12	JavaScript	22875
16	SQL	18754
0	Java	14524
6	C#	12476
26	Python	11704
3	PHP	10290
27	C++	8155
28	C	6974
14	TypeScript	3488
18	Ruby	3324

Среди них есть и питон

***Задача* 7**.

Создадим колонки с названием языков. Для каждого языка создаем вектор, и берем поочереди респондентов. Если он указал этот язык, то добавляем в вектор 1, иначе 0. Потом добавляем эти значения в таблицу. Так мы получили таблицу с бинарными векторами.

Потом группируем по странам, суммируем и ищем в какой колонке был максимум. Так мы получем названия самых популярных языков в каждой стране.

```
Country
United States          JavaScript
India                  JavaScript
United Kingdom         JavaScript
Germany                JavaScript
Canada                 JavaScript
France                 JavaScript
Poland                 JavaScript
Australia              JavaScript
Spain                  JavaScript
Russian Federation     JavaScript
dtype: object
```

В топ-10 стран по числу респондентов самый популярный язык JavaScript. Страна в которой JavaScript не самый популярный язык это Южная Корея.

```
top_l [top_l !='JavaScript' ] [0:5]
```

```
Country
South Korea     Java
Morocco          SQL
Lebanon          SQL
Saudi Arabia     SQL
Malta             C#
dtype: object
```

***Задача* 8**.

Вопрос: как уровень образования влияет на зарплату программиста?

```
survey.groupby ('FormalEducation').mean( ).Salary
```

```
FormalEducation
Bachelor's degree                                                56914.36
Doctoral degree                                                  78527.93
I never completed any formal education                           44430.66
I prefer not to answer                                           38284.84
Master's degree                                                  58250.84
Primary/elementary school                                        62677.34
Professional degree                                              39503.66
Secondary school                                                 40395.15
Some college/university study without earning a bachelor's degree   55912.81
Name: Salary, dtype: float64
```

Наибольшую зарплату получают программисты с Doctoral degree
наименьшую с Professional degree (среди тех кто ответил)

Master's degree получают чуть больше Bachelor's degree
Primary/elementary school получают больше Master's degree и Bachelor's degree , как бы это не было странным.

Возможно это связано с малым объемом выборки:

```
survey.groupby ('FormalEducation').count( ).Salary
```

```
FormalEducation
Bachelor's degree                                                6407
Doctoral degree                                                   293
I never completed any formal education                             60
I prefer not to answer                                             45
Master's degree                                                  3077
Primary/elementary school                                          55
Professional degree                                               143
Secondary school                                                  761
Some college/university study without earning a bachelor's degree   2050
Name: Salary, dtype: int64
```