



18 обязательных инструментов Data Science для превращения данных в действия

Описание

Наука о данных – для тех, кто любит распутывать запутанные вещи и обнаруживать скрытые чудеса в кажущемся беспорядке. Это похоже на поиск иголок в стогах сена, только ученым, изучающим данные, совсем не нужно пачкать руки. Используя причудливые инструменты с красочными диаграммами и просматривая груды цифр, они просто погружаются в стог данных и находят ценные иголки в виде инсайтов, имеющих большое значение для бизнеса. Типичный набор инструментов для исследователя данных должен включать как минимум по одному элементу из каждой категории: реляционные базы данных, базы данных NoSQL, фреймворки для работы с большими данными, инструменты визуализации, инструменты скраппинга, языки программирования, среды разработки и инструменты глубокого обучения.

Реляционные базы данных

Реляционная база данных – это набор данных, структурированных в таблицы с атрибутами. Таблицы могут быть связаны друг с другом, определяя отношения и ограничения, и создавая так называемую модель данных. Для работы с реляционными базами данных обычно используется язык SQL (Structured Query Language). Приложения, которые управляют структурой и данными в реляционных базах данных, называются RDBMS (Relational DataBase Management Systems). Таких приложений очень много, и наиболее актуальные из них в последнее время стали ориентироваться на область науки о данных, добавляя функциональность для работы с хранилищами больших данных и применения таких методов, как

аналитика данных и машинное обучение.

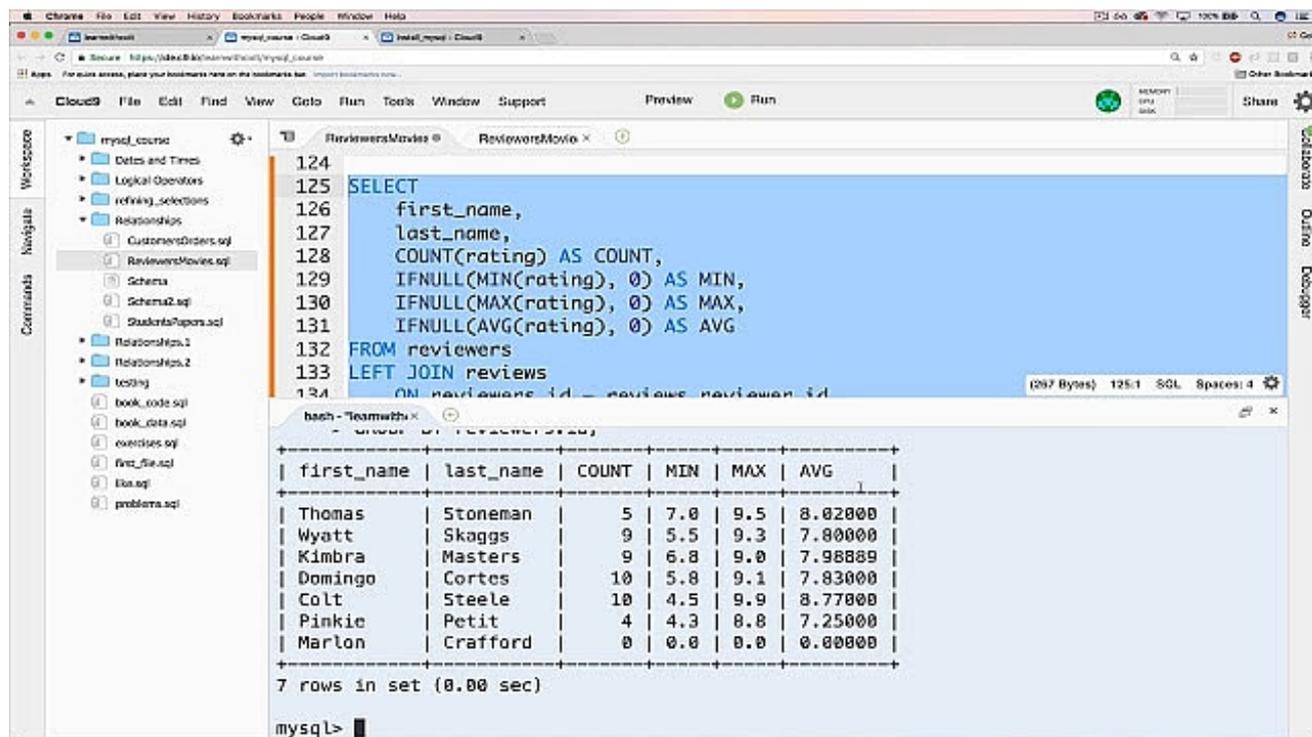
SQL Server

СУБД от Microsoft, которая развивается уже более 20 лет, последовательно расширяя свою корпоративную функциональность. Начиная с версии 2016 года, SQL Server предлагает портфель сервисов, включающих поддержку встроенного кода R. SQL Server 2017 повышает ставки, переименовывая свои службы R в службы машинного языка и добавляя поддержку языка Python (подробнее об этих двух языках ниже). Благодаря этим важным дополнениям SQL Server ориентирован на специалистов по обработке данных, которые могут не иметь опыта работы с Transact SQL, родным языком запросов Microsoft SQL Server.

SQL Server – далеко не бесплатный продукт. Вы можете приобрести лицензии на его установку на сервере Windows Server (цена зависит от количества одновременных пользователей) или использовать его как платную услугу через облако Microsoft Azure. Изучить Microsoft SQL Server очень просто.

MySQL

Если говорить о программном обеспечении с открытым исходным кодом, то корона популярности среди РСУБД принадлежит [MySQL](#). Несмотря на то, что в настоящее время она принадлежит компании Oracle, она по-прежнему свободна и распространяется на условиях лицензии GNU General Public License. Большинство веб-приложений используют MySQL в качестве базового хранилища данных, благодаря ее соответствию стандарту SQL.



The screenshot shows a web-based MySQL IDE interface. The top part displays a SQL query in a text editor:

```
124
125 SELECT
126     first_name,
127     last_name,
128     COUNT(rating) AS COUNT,
129     IFNULL(MIN(rating), 0) AS MIN,
130     IFNULL(MAX(rating), 0) AS MAX,
131     IFNULL(AVG(rating), 0) AS AVG
132 FROM reviewers
133 LEFT JOIN reviews
134     ON reviewers.id = reviews_reviewer_id
```

Below the query, the results are displayed in a table format:

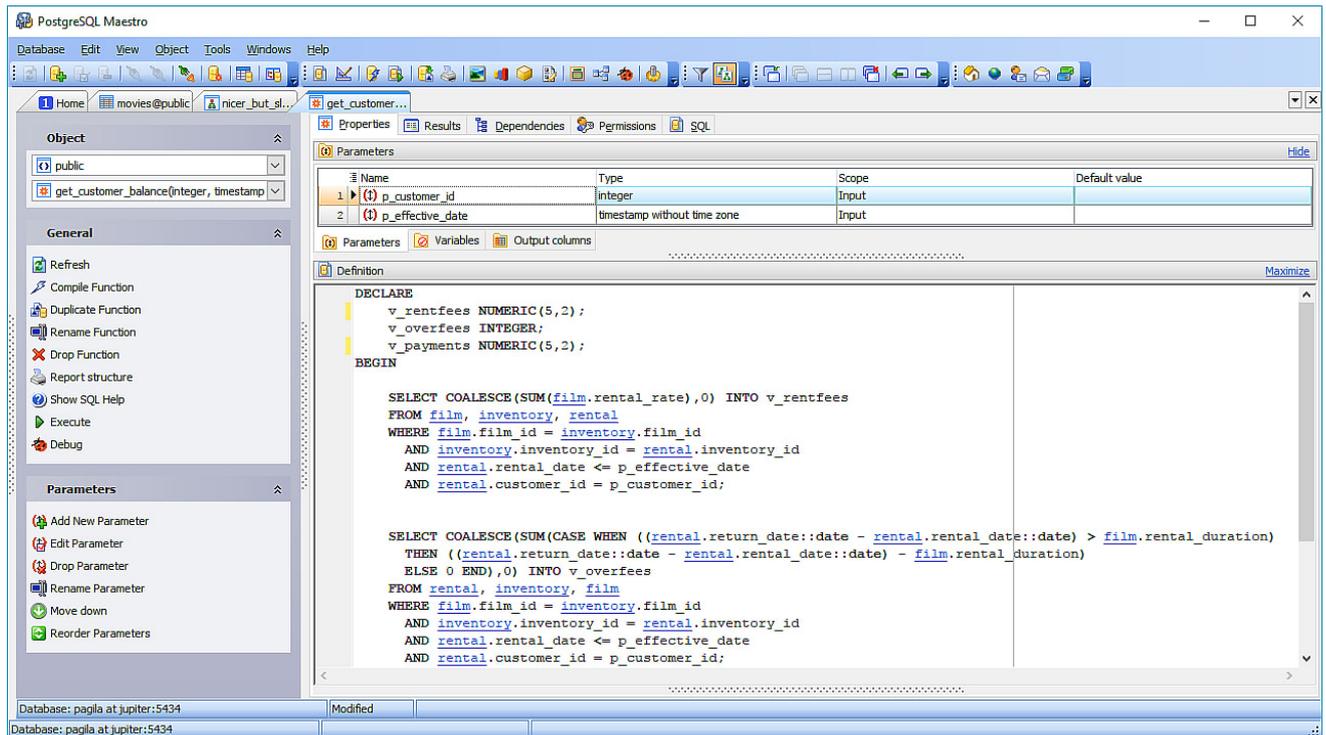
first_name	last_name	COUNT	MIN	MAX	AVG
Thomas	Stoneman	5	7.0	9.5	8.02000
Wyatt	Skaggs	9	5.5	9.3	7.80000
Kimbra	Masters	9	6.8	9.0	7.98889
Domingo	Cortes	10	5.8	9.1	7.83000
Colt	Steele	10	4.5	9.9	8.77000
Pinkie	Petit	4	4.3	8.8	7.25000
Marlon	Crafford	0	0.0	0.0	0.00000

Below the table, it indicates "7 rows in set (0.00 sec)". The MySQL prompt "mysql>" is visible at the bottom.

Популярности MySQL также способствуют простые процедуры установки, большое сообщество разработчиков, тонны исчерпывающей документации и сторонние инструменты, такие как phpMyAdmin, которые упрощают повседневную работу с системой. Хотя MySQL не имеет встроенных функций для анализа данных, его открытость позволяет интегрировать его практически с любым инструментом визуализации, отчетности и бизнес-аналитики, который вы можете выбрать.

PostgreSQL

Еще одна СУБД с открытым исходным кодом – [PostgreSQL](#). Хотя PostgreSQL не так популярна, как MySQL, она отличается гибкостью и расширяемостью, а также поддержкой сложных запросов, которые выходят за рамки базовых операторов, таких как SELECT, WHERE и GROUP BY. Эти возможности позволяют ему завоевать популярность среди специалистов по обработке данных. Еще одна интересная особенность – поддержка мультиокружения, что позволяет использовать его в облачных и локальных средах, а также в сочетании обоих вариантов, известных как гибридные облачные среды.



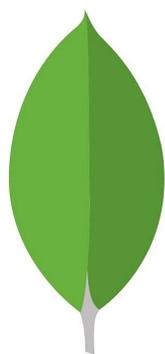
PostgreSQL способен сочетать аналитическую обработку в режиме онлайн (OLAP) с обработкой транзакций в режиме онлайн (OLTP), работая в режиме, называемом гибридной транзакционной/аналитической обработкой (HTAP). Он также хорошо подходит для работы с большими данными, благодаря добавлению PostGIS для географических данных и JSON-B для документов. PostgreSQL также поддерживает неструктурированные данные, что позволяет ему находиться в обеих категориях: SQL и NoSQL базы данных.

Базы данных NoSQL

Этот тип хранилищ данных, также известный как нереляционные базы данных, обеспечивает более быстрый доступ к нетабличным структурам данных. Примерами таких структур являются графы, документы, широкие колонки, ключевые значения и многое другое. Хранилища данных NoSQL позволяют отказаться от согласованности данных в пользу других преимуществ, таких как доступность, разбиение на разделы и скорость доступа. Поскольку в хранилищах данных NoSQL нет SQL, единственным способом запросить данные в такой базе данных является использование низкоуровневых языков, а такого языка, который был бы так же широко распространен, как SQL, не существует. Кроме того, для NoSQL не существует стандартных спецификаций. Поэтому, как ни странно, некоторые NoSQL-базы данных начинают добавлять поддержку SQL-скриптов.

MongoDB

MongoDB – это популярная система баз данных NoSQL, которая хранит данные в виде документов JSON. Основное внимание в ней уделяется масштабируемости и гибкости хранения данных в неструктурированном виде. Это означает, что не существует фиксированного списка полей, который должен соблюдаться во всех хранимых элементах. Кроме того, структура данных может быть изменена с течением времени, что в реляционной базе данных сопряжено с высоким риском нарушения работы приложений.

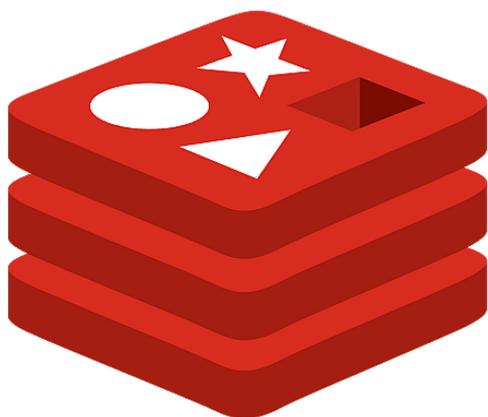


mongoDB

Технология MongoDB позволяет индексировать, выполнять специальные запросы и агрегировать данные, что создает прочную основу для анализа данных. Распределенная природа базы данных обеспечивает высокую доступность, масштабирование и географическое распределение без необходимости использования сложных инструментов.

Redis

[Это](#) еще один вариант NoSQL с открытым исходным кодом. По сути, это хранилище структур данных, работающее в памяти, и, помимо предоставления услуг базы данных, оно также работает как кэш-память и брокер сообщений.



redis

Он поддерживает множество нетрадиционных структур данных, включая хэши, геопространственные индексы, списки и отсортированные множества. Он хорошо подходит для работы с данными благодаря высокой производительности в задачах, требующих больших объемов данных, таких как вычисление пересечений множеств, сортировка длинных списков или составление сложных рейтингов. Причиной выдающейся производительности Redis является его работа в памяти. Его можно настроить на выборочное сохранение данных.

Фреймворки для работы с большими данными

Предположим, вам нужно проанализировать данные, которые пользователи Facebook генерируют в течение месяца. Речь идет о фотографиях, видео, сообщениях, обо всем этом. Учитывая, что каждый день пользователи добавляют в социальную сеть более 500 терабайт данных, трудно измерить объем, который представляет собой целый месяц ее данных. Чтобы эффективно манипулировать огромным объемом данных, необходим соответствующий фреймворк, способный вычислять статистику в распределенной архитектуре. На рынке лидируют два таких фреймворка: Hadoop и Spark.

Hadoop

Являясь платформой для работы с большими данными, Hadoop справляется со сложностями, связанными с поиском, обработкой и хранением огромных массивов данных. Hadoop работает в распределенной среде, состоящей из компьютерных кластеров, которые обрабатывают простые алгоритмы. Существует алгоритм MapReduce, который разделяет большие задачи на маленькие части, а затем распределяет эти маленькие задачи между доступными кластерами.



Hadoop рекомендуется для хранилищ данных корпоративного класса, требующих быстрого доступа и высокой доступности, и все это в недорогой схеме. Но вам нужен администратор Linux с глубокими знаниями Hadoop, чтобы поддерживать фреймворк в рабочем состоянии.

Spark

Hadoop – не единственный фреймворк для работы с большими данными. Еще одно громкое имя в этой области – Spark. Движок Spark был разработан для того, чтобы превзойти Hadoop по скорости аналитики и простоте использования. Судя по всему, он достиг этой цели: по некоторым сравнениям, Spark работает в 10 раз быстрее Hadoop при работе на диске и в 100 раз быстрее при работе в памяти. Кроме того, для обработки того же объема данных ему требуется меньшее количество машин.



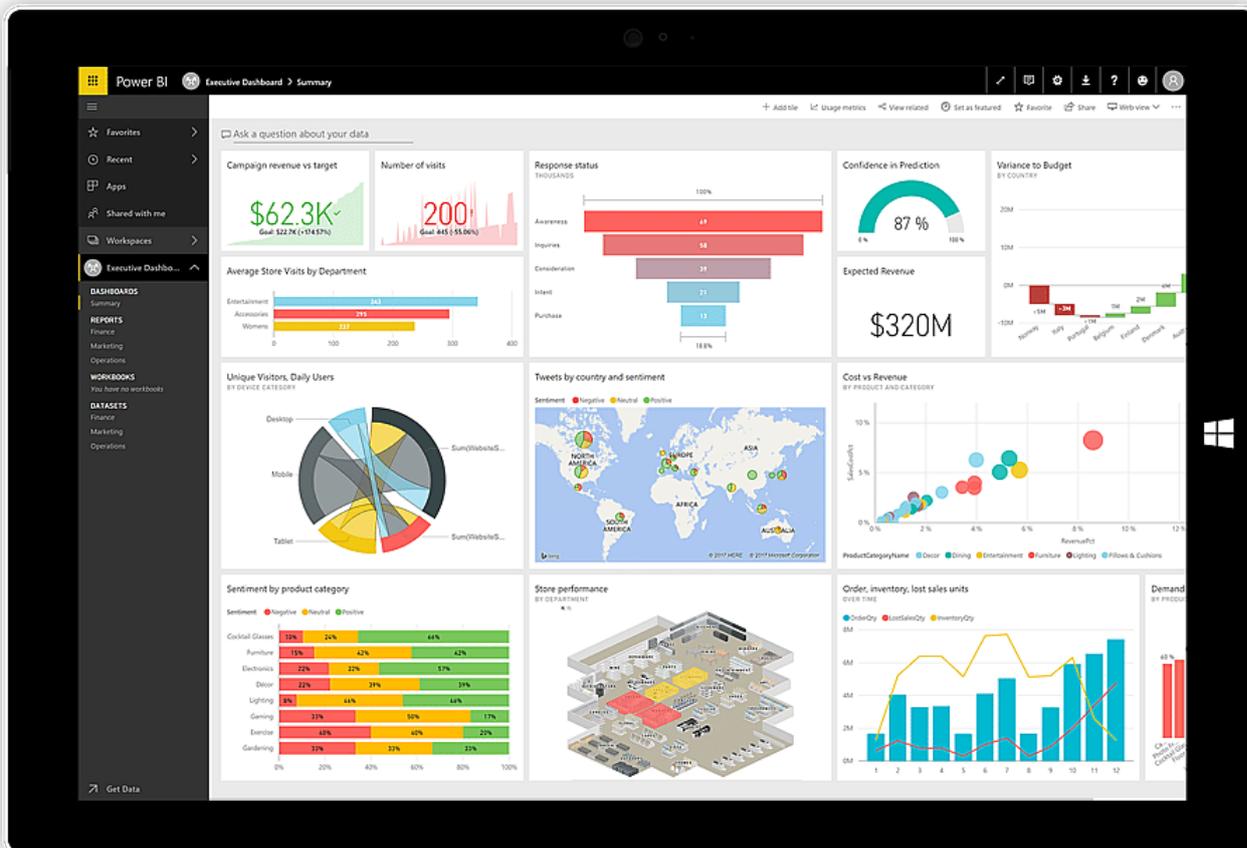
Помимо скорости, еще одним преимуществом Spark является поддержка потоковой обработки. Этот тип обработки данных, также называемый обработкой в реальном времени, предполагает непрерывный ввод и вывод данных.

Инструменты визуализации

Распространенная шутка среди специалистов по исследованию данных гласит, что если долго мучить данные, то они сами признаются в том, что вам нужно знать. В данном случае “пытать” означает манипулировать данными, преобразуя и фильтруя их, чтобы лучше их визуализировать. Именно здесь на сцену выходят инструменты визуализации данных. Эти инструменты берут предварительно обработанные данные из разных источников и показывают открывшиеся истины в графической, понятной форме. Существуют сотни инструментов, которые попадают в эту категорию. Нравится вам это или нет, но самым распространенным является Microsoft Excel и его инструменты построения диаграмм. Диаграммы Excel доступны каждому, кто использует Excel, но они имеют ограниченную функциональность. То же самое относится и к другим приложениям для работы с электронными таблицами, например Google Sheets и Libre Office. Но мы говорим о более специфических инструментах, специально предназначенных для бизнес-аналитики (BI) и анализа данных.

Power BI

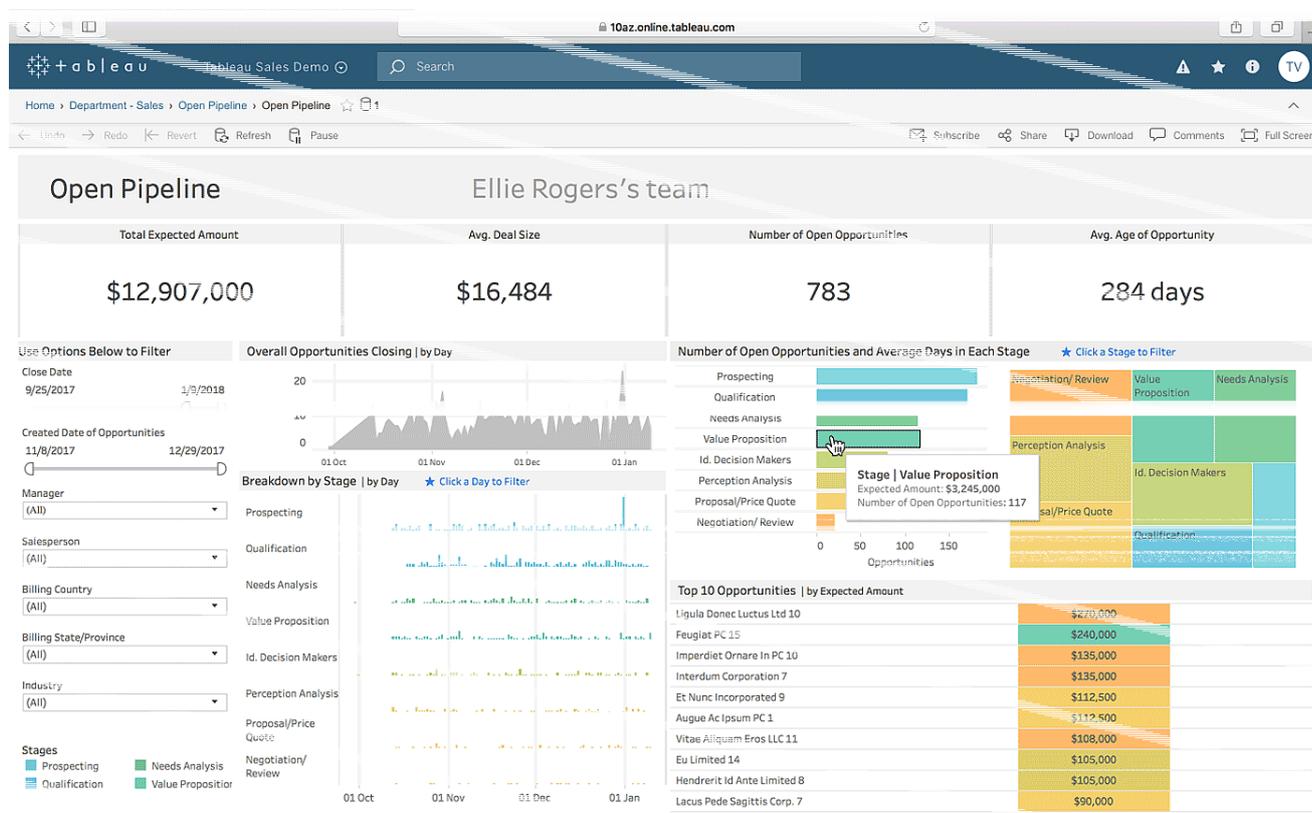
Не так давно компания Microsoft выпустила приложение для визуализации Power BI. Оно может получать данные из различных источников, таких как текстовые файлы, базы данных, электронные таблицы и многие онлайн-сервисы данных, включая Facebook и Twitter, и использовать их для создания приборных панелей, наполненных графиками, таблицами, картами и многими другими объектами визуализации. Объекты приборной панели интерактивны, то есть вы можете нажать на серию данных на графике, чтобы выбрать ее и использовать в качестве фильтра для других объектов на доске.



Power BI – это сочетание настольного приложения для Windows (часть пакета Office 365), веб-приложения и онлайн-сервиса для публикации информационных панелей в Интернете и их совместного использования с пользователями. Служба позволяет создавать и управлять разрешениями, чтобы предоставлять доступ к панелям только определенным людям.

Tableau

Tableau – еще один вариант для создания интерактивных информационных панелей на основе комбинации нескольких источников данных. Он также предлагает настольную версию, веб-версию и онлайн-сервис для обмена созданными панелями. Она работает естественно, “с учетом особенностей вашего мышления” (как утверждает компания), и проста в использовании для нетехнических специалистов, чему способствуют многочисленные учебники и онлайн-видео.



Среди наиболее выдающихся особенностей Tableau – неограниченное количество коннекторов для данных, живые данные и данные in-memory, а также оптимизированный для мобильных устройств дизайн.

QlikView

[QlikView](#) предлагает чистый и понятный пользовательский интерфейс, помогающий аналитикам извлекать новые знания из имеющихся данных с помощью визуальных элементов, понятных каждому.



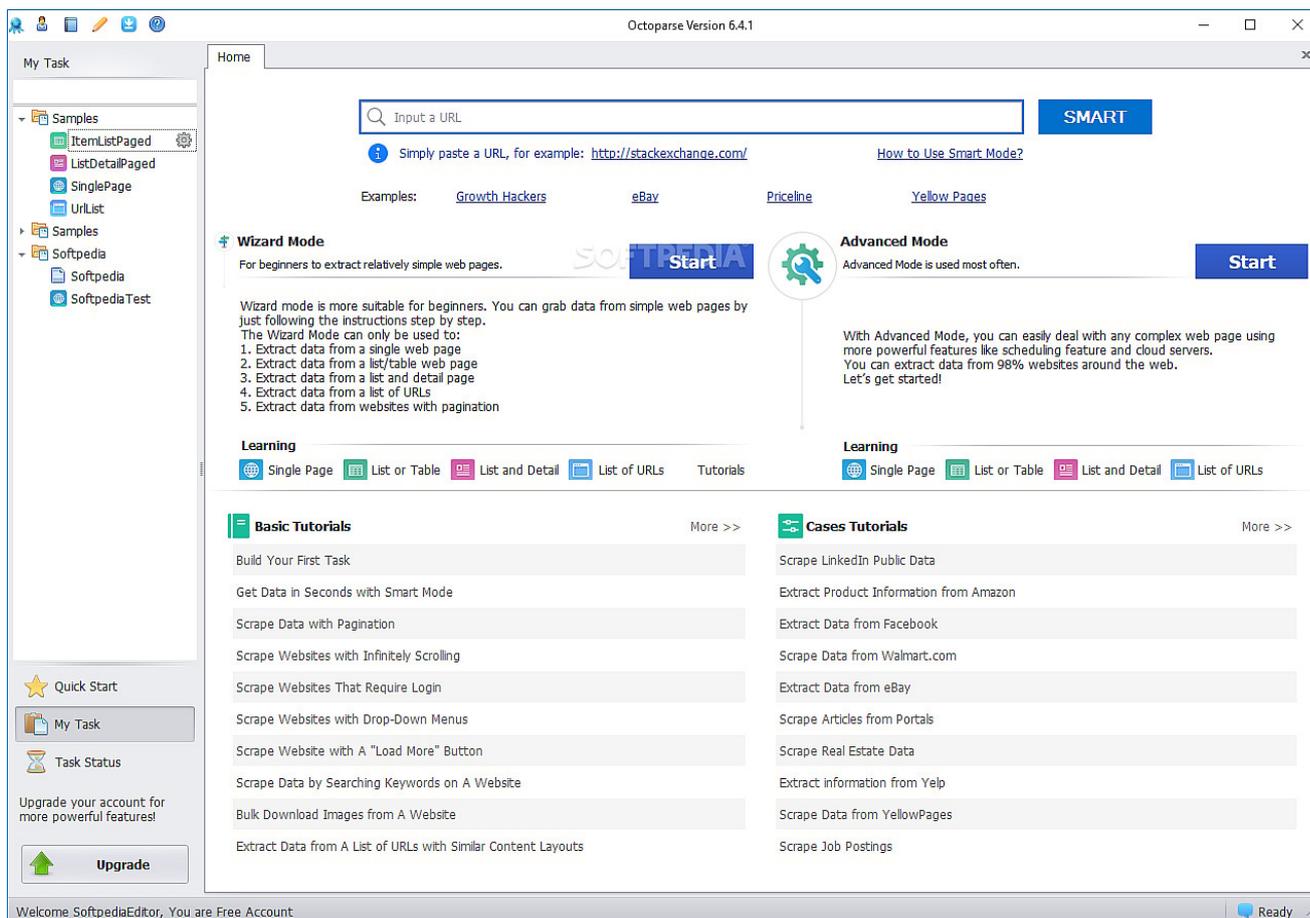
Этот инструмент известен как одна из самых гибких платформ для бизнес-аналитики. В нем предусмотрена функция ассоциативного поиска, которая помогает сосредоточиться на наиболее важных данных, экономя время на их самостоятельный поиск. С помощью QlikView вы можете сотрудничать с партнерами в режиме реального времени, проводя сравнительный анализ. Все необходимые данные можно объединить в одном приложении с функциями безопасности, ограничивающими доступ к данным.

Инструменты для скрапинга

Во времена, когда Интернет только зарождался, веб-краулеры начали путешествовать по сетям, собирая информацию на своем пути. По мере развития технологий термин web crawling сменился на web scraping, но смысл остался прежним: автоматическое извлечение информации с веб-сайтов. Для веб-скрапинга используются автоматизированные процессы, или боты, которые переходят с одной веб-страницы на другую, извлекают из них данные и экспортируют их в различные форматы или вставляют в базы данных для дальнейшего анализа. Ниже мы приводим характеристики трех наиболее популярных на сегодняшний день веб-скреперов.

Octoparse

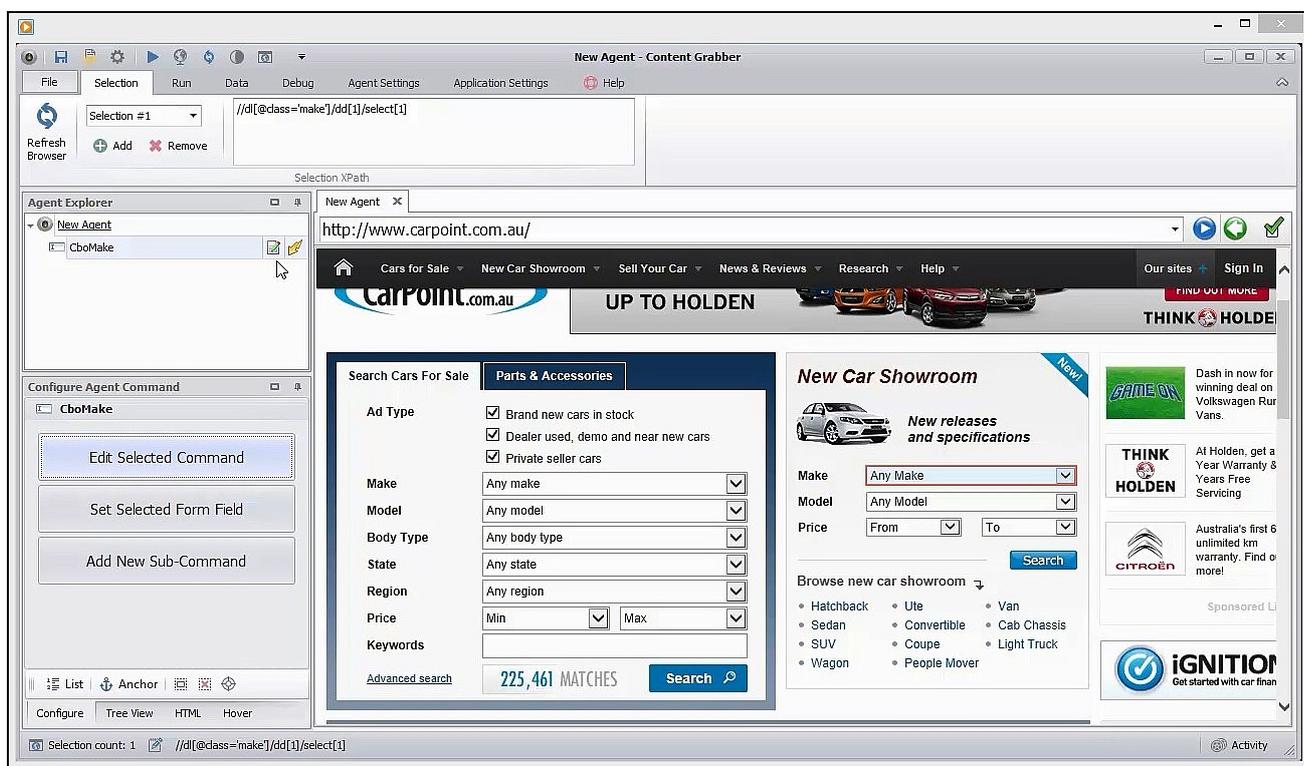
Веб-скрепер Octoparse обладает рядом интересных характеристик, включая встроенные инструменты для получения информации с сайтов, которые не позволяют ботам-скреперам легко выполнять свою работу. Это настольное приложение, не требующее кодирования, с удобным пользовательским интерфейсом, позволяющим визуализировать процесс извлечения информации с помощью графического конструктора рабочих процессов.



Вместе с отдельным приложением Octoparse предлагает облачный сервис для ускорения процесса извлечения данных. При использовании облачного сервиса вместо настольного приложения пользователи могут получить от 4 до 10-кратного прироста скорости. Если вы придерживаетесь настольной версии, то можете пользоваться Octoparse бесплатно. Но если вы предпочитаете использовать облачный сервис, вам придется выбрать один из его платных тарифных планов.

Content Grabber

Если вы ищете многофункциональный инструмент для скраппинга, вам стоит обратить внимание на Content Grabber. В отличие от Octoparse, для использования Content Grabber необходимо обладать продвинутыми навыками программирования. Взамен вы получаете редактирование сценариев, отладочные интерфейсы и другие расширенные возможности. С помощью Content Grabber вы можете использовать языки .Net для написания регулярных выражений. Таким образом, вам не придется генерировать выражения с помощью встроенного инструмента.



Инструмент предлагает API (интерфейс прикладного программирования), который можно использовать для добавления возможностей скраппинга в настольные и веб-приложения. Чтобы использовать этот API, разработчикам необходимо получить доступ к службе Content Grabber Windows.

ParseHub

Этот скрепер может работать с обширным списком различных типов контента, включая форумы, вложенные комментарии, календари и карты. Он также может работать со страницами, содержащими аутентификацию, Javascript, Ajax и многое

другое. ParseHub можно использовать как веб-приложение или настольное приложение, способное работать под Windows, macOS X и Linux.

The screenshot shows the ParseHub web interface. On the left, there's a 'Commands' panel with options like 'main_template', 'results', 'Select page', 'Begin new entry in selection1', 'Extract name', and 'Extract url'. Below that is a 'Selection Node' section with a '2nd div with class info' and a 'Wait up to 60 seconds' option. A 'Get Data' button is at the bottom left. The main area displays a browser view of a search for 'web developers' in 'New York City, NY'. It shows a list of search results with details like company name, address, phone, and services. On the right, there are advertisements for 'Matrix Technosystems LLC', 'Big Mouth Frog Development LLC', and 'Cheep Cheep Postcards - Direct Mail Advertising'. At the bottom, a JSON output window shows the extracted data for 'Falcon Graphics & Design Inc' and 'Matrix Technosystems LLC'. Buttons for 'CSV/Excel' and 'JSON' are visible.

Как и в случае с Content Grabber, для получения максимальной отдачи от ParseHub рекомендуется обладать некоторыми знаниями в области программирования. У него есть бесплатная версия, ограниченная 5 проектами и 200 страницами за прогон.

Языки программирования

Как и упомянутый ранее язык SQL, созданный специально для работы с реляционными базами данных, существуют и другие языки, созданные с четкой ориентацией на науку о данных. Эти языки позволяют разработчикам писать программы для анализа массивных данных, таких как статистика и машинное обучение. SQL также считается важным навыком, которым должны обладать разработчики, чтобы заниматься наукой о данных, но это связано с тем, что в большинстве организаций до сих пор хранится много данных в реляционных базах данных. "Настоящими" языками науки о данных являются R и Python.

Python



Python – это высокоуровневый интерпретируемый язык программирования общего назначения, хорошо подходящий для быстрой разработки приложений. Он обладает простым и легким в освоении синтаксисом, что позволяет сократить время обучения и снизить затраты на сопровождение программ. Существует множество причин, по которым этот язык является предпочтительным для науки о данных. Вот лишь некоторые из них: возможность написания сценариев, многословность, переносимость и производительность. Этот язык является хорошей отправной точкой для специалистов по изучению данных, которые планируют много экспериментировать, прежде чем приступить к реальной и сложной работе с данными, и которые хотят разрабатывать полноценные приложения.

R

Язык R в основном используется для статистической обработки данных и построения графиков. Хотя он не предназначен для разработки полноценных приложений, как, например, Python, в последние годы R стал очень популярен благодаря своему потенциалу для добычи и анализа данных.



Благодаря постоянно растущей библиотеке свободно распространяемых пакетов, расширяющих его функциональность, R способен выполнять все виды работ по обработке данных, включая линейное/нелинейное моделирование, классификацию, статистические тесты и т. д. Этот язык не так прост в изучении, но как только вы познакомитесь с его философией, вы будете заниматься статистическими вычислениями как профессионал.

IDE

Если вы всерьез решили посвятить себя науке о данных, то вам нужно будет тщательно выбрать интегрированную среду разработки (IDE), которая будет соответствовать вашим потребностям, потому что вы и ваша IDE будете проводить много времени, работая вместе. Идеальная IDE должна сочетать в себе все инструменты, необходимые вам в повседневной работе кодера: текстовый редактор с подсветкой синтаксиса и автозавершением, мощный отладчик, браузер объектов

и удобный доступ к внешним инструментам. Кроме того, она должна быть совместима с предпочитаемым вами языком, поэтому выбирать IDE следует, зная, какой язык вы будете использовать.

Spyder

[Эта](#) универсальная IDE предназначена в основном для ученых и аналитиков, которым также необходимо кодировать. Чтобы им было удобно, она не ограничивается функционалом IDE – в ней также есть инструменты для исследования/визуализации данных и интерактивного выполнения, как это может быть в научном пакете. Редактор в Spyder поддерживает множество языков, в нем есть браузер классов, разделение окон, переход к определению, автоматическое завершение кода и даже инструмент анализа кода.

The screenshot displays the Spyder Python IDE interface. The main editor window shows a Python script with the following code:

```

7 import pylab
8 from numpy import cos, linspace, pi, sin, random
9 from scipy.interpolate import splprep, splev
10
11 # %% Generate data for analysis
12
13 # Make ascending spiral in 3-space
14 t = linspace(0, 1.75 * 2 * pi, 100)
15
16 x = sin(t)
17 y = cos(t)
18 z = t
19
20 # Add noise
21 x += random.normal(scale=0.1, size=x.shape)
22 y += random.normal(scale=0.1, size=y.shape)
23 z += random.normal(scale=0.1, size=z.shape)
24
25
26 # %% Perform calculations
27
28 # Spline parameters
29 smoothness = 3.0 # Smoothness parameter
30 k_param = 2 # Spline order
31 nests = -1 # Estimate of number of knots needed (-1 = maximal)
32
33 # Find the knot points
34 knot_points, u = splprep([x, y, z], s=smoothness, k=k_param, nests=-1)
35
36 # Evaluate spline, including interpolated points
37 xnew, ynew, znew = splev(linspace(0, 1, 400), knot_points)
38
39
40 # %% Plot results
41
42 # TODO: Rewrite to avoid code smell
43 pylab.subplot(2, 2, 1)
44 data, = pylab.plot(x, y, 'bo-', label='Data with X-Y Cross Section')
45 fit, = pylab.plot(xnew, ynew, 'r-', label='Fit with X-Y Cross Section')
46 pylab.legend()
47 pylab.xlabel('x')
48 pylab.ylabel('y')
49
50 pylab.subplot(2, 2, 2)
51 data, = pylab.plot(x, z, 'bo-', label='Data with X-Z Cross Section')
52 fit, = pylab.plot(xnew, znew, 'r-', label='Fit with X-Z Cross Section')
53 pylab.legend()
54 pylab.xlabel('x')

```

The right sidebar shows the Variable explorer with the following data:

Name	Type	Size	Min:	Max:	Value
array_int8	int8	(2, 3)	-7	6	
array_uint32	uint32	(2, 2, 3)	1	7	
bars	container.BarContainer	20			BarContainer object of matplotlib.conta...
df	DataFrame	(3, 2)			Column names: bools, ints
filename	str	1			C:\ProgramData\Anaconda3\lib\site-packa...
list_test	list	2			[DataFrame, Numpy array]
nrows	int	1	344		
r	float64	1	7.611802589334796		
radii	float64	(20,)	Min: 0.4963083626525087	Max: 9.858484872642551	
region	tuple	2			(slice, slice)
rgb	float64	(45, 45, 4)	Min: 0.0	Max: 1.0	
series	Series	(1,)			Series object of pandas.core.series mod...
test_none	NoneType	1			NoneType object of builtins module

The Python console shows the following output:

```

...: ls = LightSource(270, 45)
...: # To use a custom hillshading mode, override the built-in shading
...: # in the rgb colors of the shaded surface calculated from "shade".
...: rgb = ls.shade(z, cmap=cm.gist_earth, vert_exag=0.1, blend_mode='soft')
...: surf = ax.plot_surface(x, y, z, rstride=1, cstride=1, facecolors=rgb,
...:                       linewidth=0, antialiased=False, shade=False)
...:
...: plt.show()

```

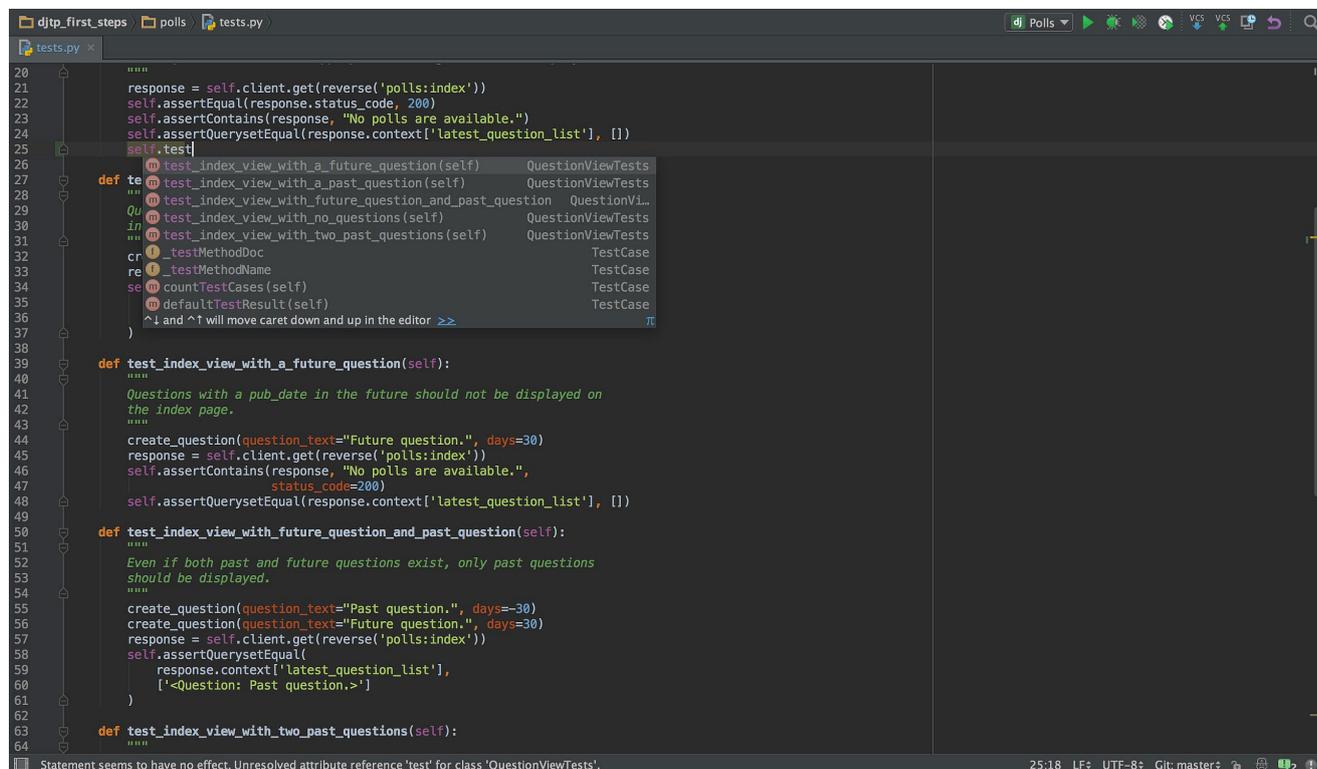
The console also displays a 3D surface plot and a 2D polar plot. The 3D plot shows a surface with a color scale from 0 to 700. The 2D plot shows a polar coordinate system with a color scale from 0 to 10.

Отладчик поможет вам в интерактивном режиме отследить каждую строку кода, а профилировщик – найти и устранить неэффективные ошибки.

PyCharm

Если вы программируете на Python, то, скорее всего, ваша IDE – PyCharm. В ней есть

интеллектуальный редактор кода с умным поиском, завершением кода, обнаружением и исправлением ошибок. Одним щелчком мыши вы можете перейти из редактора кода в любое контекстное окно, включая тест, суперметод, реализацию, декларацию и многое другое. PyCharm поддерживает Anaconda и многие научные пакеты, такие как NumPy и Matplotlib, и это только два из них.

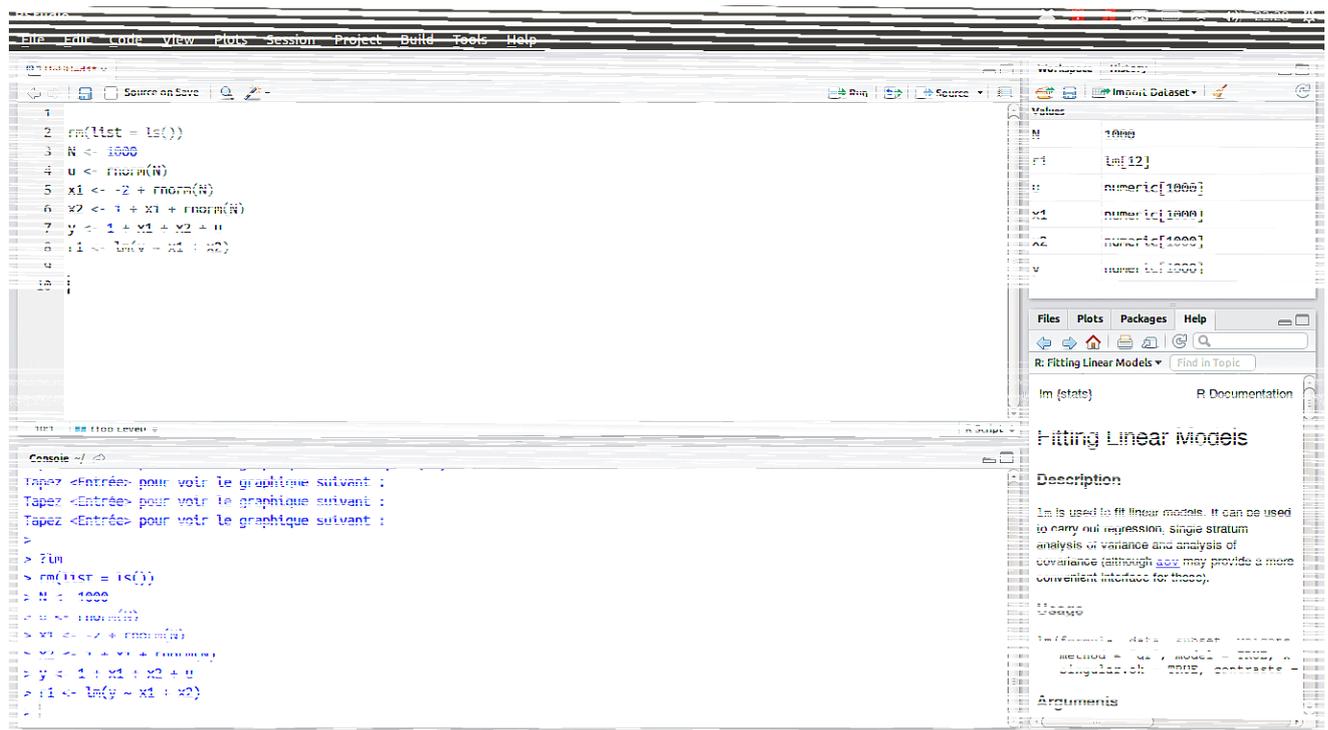


```
20 """
21 response = self.client.get(reverse('polls:index'))
22 self.assertEqual(response.status_code, 200)
23 self.assertContains(response, "No polls are available.")
24 self.assertQuerysetEqual(response.context['latest_question_list'], [])
25 self.test
26
27 def test_index_view_with_a_future_question(self)
28 """
29 test_index_view_with_future_question_and_past_question
30 """
31 in test_index_view_with_no_questions(self)
32 """
33 cr _testMethodDoc
34 re _testMethodName
35 se countTestCases(self)
36 defaultTestResult(self)
37 )
38
39 def test_index_view_with_a_future_question(self):
40 """
41 Questions with a pub_date in the future should not be displayed on
42 the index page.
43 """
44 create_question(question_text="Future question.", days=30)
45 response = self.client.get(reverse('polls:index'))
46 self.assertContains(response, "No polls are available.",
47 status_code=200)
48 self.assertQuerysetEqual(response.context['latest_question_list'], [])
49
50 def test_index_view_with_future_question_and_past_question(self):
51 """
52 Even if both past and future questions exist, only past questions
53 should be displayed.
54 """
55 create_question(question_text="Past question.", days=-30)
56 create_question(question_text="Future question.", days=30)
57 response = self.client.get(reverse('polls:index'))
58 self.assertQuerysetEqual(
59 response.context['latest_question_list'],
60 ['<Question: Past question.>']
61 )
62
63 def test_index_view_with_two_past_questions(self):
64 """
```

Он предлагает интеграцию с наиболее важными системами контроля версий, а также с прогонщиком тестов, профилировщиком и отладчиком. Кроме того, он интегрируется с Docker и Vagrant, обеспечивая кроссплатформенную разработку и контейнеризацию.

RStudio

Для тех специалистов по исследованию данных, которые предпочитают использовать команду R, наиболее подходящей IDE должна быть RStudio, поскольку она обладает множеством функций. Ее можно установить на настольный компьютер с Windows, macOS или Linux, а также запустить через веб-браузер, если вы не хотите устанавливать ее локально. Обе версии предлагают такие возможности, как подсветка синтаксиса, умный отступ и завершение кода. Есть встроенный просмотрщик данных, который пригодится, когда нужно просмотреть табличные данные.



Режим отладки позволяет наблюдать за динамическим обновлением данных при пошаговом выполнении программы или скрипта. Для контроля версий в RStudio интегрирована поддержка SVN и Git. Приятным плюсом является возможность создания интерактивной графики с помощью библиотек Shiny и gives.

Ваш личный ящик с инструментами

К этому моменту у вас должно быть полное представление об инструментах, которые необходимо знать, чтобы преуспеть в науке о данных. Кроме того, мы надеемся, что предоставили вам достаточно информации, чтобы решить, какой вариант из каждой категории инструментов является наиболее удобным. Теперь все зависит от вас. Наука о данных – это процветающая область, в которой можно сделать [карьеру](#). Но если вы хотите сделать это, вы должны следить за

изменениями в тенденциях и технологиях, поскольку они происходят практически ежедневно.

Дата Создания

25.02.2024