минобрнауки россии

федеральное государственное бюджетное образовательное учреждение

высшего образования

«ЧЕРЕПОВЕЦКИЙ ГОСУДАРСТВЕННЫЙ УНИВЕРСИТЕТ»

|  |  |
| --- | --- |
| Институт (факультет) | Институт Информационных Технологий |
| Кафедра | Математического и Программного Обеспечения ЭВМ |

**КУРСОВАЯ РАБОТА**

|  |
| --- |
| по дисциплине Прикладная статистика |

|  |  |
| --- | --- |
| на тему | Статистический анализ социально-экономических показателей |

|  |
| --- |
| Выполнил студент группы |
| 1ПИб-02-3оп-22 |
| направление подготовки (специальности) |
| 09.03.04, Программная инженерия |
| *шифр, наименование* |
| Маркелов Сергей Александрович |
| *фамилия, имя, отчество* |

|  |
| --- |
| Руководитель |
| Гонтарева Ирина Борисовна |
| *фамилия, имя, отчество* |
| доцент |
| *должность* |

|  |
| --- |
| Дата представления работы |
| «\_\_19\_\_»\_\_\_\_\_\_\_декабря\_\_\_\_\_\_\_\_2023 г. |
|  |
| Заключение о допуске к защите |
|  |
|  |
|  |
|  |
| Оценка \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_, \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_ |
| количество баллов |
| Подпись преподавателя\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_ |

Череповец, 2023

*Год*

Оглавление

[1. Введение 3](#_Toc153533691)

[2. Описание предметной области 5](#_Toc153533692)

[3. Группировка данных 9](#_Toc153533693)

[4. Изучение взаимосвязи 18](#_Toc153533694)

[4.1. Изучение взаимосвязи между двумя количественными признаками 19](#_Toc153533695)

[4.2. Изучение взаимосвязи между количественными и качественными признаками 21](#_Toc153533696)

[5. Ряды динамики 24](#_Toc153533697)

[Заключение 26](#_Toc153533698)

[Приложение 1 29](#_Toc153533699)

[Приложение 2 36](#_Toc153533700)

# Введение

Статистика – это наука, в которой излагаются общие вопросы сбора, измерения, мониторинга, анализа массовых статистических (количественных или качественных) данных и их сравнение. Статистика занимается количественным изучением различных явлений, а также позволяет делать выводы на основе полученных данных [2].

Прикладная статистика – это область статистики, которая фокусируется на использовании статистических методов и инструментов для решения конкретных практических задач в различных областях деятельности. Методы прикладной статистики активно применяются в технических исследованиях, экономике, менеджменте, социологии, медицине, геологии, истории и т. д. [3].

Основные задачи статистики:

* cбор данных – статистика начинается со сбора данных, которые могут быть получены из различных источников, таких как опросы, наблюдения, эксперименты и другие методы;
* обработка данных – этот этап включает в себя организацию и систематизацию собранных данных для удобства анализа, может включать в себя расчет различных статистических показателей, таких как среднее значение, медиана, стандартное отклонение и другие;
* анализ данных – статистика использует различные методы анализа для выявления закономерностей, трендов, корреляций и других характеристик данных;
* интерпретация данных – после анализа данные интерпретируются с целью сделать выводы, сформулировать гипотезы или принимать решения на основе полученной информации.

Сайты играют огромную роль в современном мире. По состоянию на август 2021 года, в интернете насчитывалось 1,88 млрд сайтов, и их количество с каждым днем только продолжает расти [4]. Легко можно найти сайты по самым различным темам – новостные, спортивные, музыкальные, развлекательные, познавательные и многие другие.

Целью данной курсовой работы является изучение посещаемости наиболее популярных сайтов, как в российском сегменте интернета, так и в зарубежном, а также нахождение зависимости популярности сайтов от каких-либо признаков.

Данные получены благодаря счетчикам аналитики – скриптам, встроенным в web-страницу, которые фиксируют посещения сайта и собирают информацию о посетителях.

Для проведения исследования необходимо собрать и обработать данные о посещаемости сайтов. Затем их необходимо проанализировать. Для этого нужно построить сводные таблицы и диаграммы, найти взаимосвязи между признаками, построить динамические ряды и т. д. На основании полученных данных можно будет сделать вывод о том, какие сайты являются наиболее популярными, и от каких признаков зависит их популярность.

# Описание предметной области

Генеральной совокупностью являются все сайты в интернете.

Единицей наблюдения являются сайты. В выборку вошли 75 популярных сайтов. В выборку вошли как российские, так и зарубежные сайты.

Сайт – это одна или несколько взаимосвязанных web-страниц, обычно доступных через интернет и имеющих общий домен или поддомен [5]. Сайт может содержать различные виды контента: текст, изображения, видео, аудио и интерактивные элементы. Сайты используются для различных целей, таких как предоставление информации, коммерции, образования, развлечений и многого другого.

Сайт обладает следующими характерными свойствами:

* состоит из web-страниц (как правило, из двух или более); реже встречаются одностраничные сайты;
* имеет уникальный домен или поддомен, т.е. для доступа к сайту используется уникальное доменное имя (например, www.example.com);
* создан с применением web-технологий, таких как HTML (язык разметки), CSS (каскадные таблицы стилей) и JavaScript (язык программирования);
* доступен через интернет, что позволяет зайти на сайт с любого устройства, имеющего доступ к сети;
* файлы сайта размещаются на хостинге – web-сервере, который обеспечивает доступ к содержимому сайта через интернет.

Если web-документ не соответствует какому-либо из данных свойств, то назвать его полноценным сайтом нельзя.

Наблюдение проводится по следующим признакам:

* тип сайта (качественный признак) – данный признак показывает, какой контент размещен на сайте (например, «Новости», «Фильмы и сериалы», «Музыка» и т. д.);
* страна (качественный признак) – этот признак указывает, в какой стране был разработан сайт;
* год создания (качественный признак) – данный признак указывает, в каком году начал работу данный сайт;
* посетители (количественный признак) – количество уникальных пользователей, хотя бы раз зашедших на сайт в течение определенного периода (в данной курсовой работе – в течение месяца);
* просмотры (количественный признак) – количество посещений сайта всеми пользователями в течение определенного периода (в данной курсовой работе – в течение месяца); данный признак будет иметь большее значение, чем количество посетителей, т. к. один пользователь может зайти на сайт несколько раз за этот период;
* среднее время посещения (количественный признак) – время, которое в среднем проводит пользователь на сайте (кто-то может зайти и сразу выйти, а кто-то задержаться на сайте на несколько часов); вычисляется, как среднее арифметическое суммы времени посещения всех пользователей (1):

где n – количество пользователей, зашедших на сайт, t1, t2, … tn – время, которое провели на сайте 1-ый, 2-ой, …, n-ный пользователи;

* рейтинг сайта (количественный признак) – процент положительных отзывов от пользователей сайта на сервисе «Яндекс.Вебмастер»;
* доступность в России (качественный признак) – данный признак показывает, можно ли воспользоваться данным сайтом на территории РФ без использования VPN или иных способов обхода блокировки (сайт может быть заблокирован Роскомнадзором за нарушение законов РФ, или же владельцы сайта могут ограничить частично или полностью закрыть доступ с российских IP-адресов из-за санкций).

Данные для проведения исследования были получены благодаря счетчикам аналитики – скриптам, встроенным в web-страницу, которые фиксируют посещения сайта и собирают информацию о посетителях. Эта информация отправляется на сервер, который обрабатывает эти данные и рисует отчёт: кто был на сайте и что делал.

В качестве источников данных были использованы следующие интернет-сервисы web-аналитики:

* «SimilarWeb» – платформа, занимающаяся web-аналитикой и предоставляющая владельцам сайтов данные о сетевом трафике и активности пользователей на сайте [6]. На платформе также формируется рейтинг сайтов на основании показателей трафика и действий пользователей. Рейтинг рассчитывается с учётом собранных наборов данных и обновляется ежемесячно;
* «LiveInternet» – сервис для сбора и анализа метрик эффективности сайтов [7]. Он отличается простотой использования, интуитивно понятным интерфейсом и наличием уникальных для служб web-аналитики функций. Счетчики «Liveinternet» установлены на более чем 1,5 млн сайтов;
* «Рейтинг@Mail.Ru» – система web-аналитики, собирающая данные о посетителях web-сайтов и их устройствах [8]. Установлена на 850 тыс. сайтов, ежедневно замеряет статистику посещений более 90 млн уникальных посетителей;
* «Рамблер / Топ-100» – счётчик посещаемости и система web-аналитики [9]. Собранные счётчиком данные доступны владельцу ресурса за весь период, когда счётчик был установлен на сайте;
* «Яндекс.Метрика» – бесплатный интернет-сервис от компании «Яндекс», предназначенный для оценки посещаемости web-сайтов и анализа поведения пользователей на сайте [10];
* «Яндекс.Вебмастер» – сервис от компании «Яндекс» для web-мастеров, представляющий собой панель инструментов для оценки индексации сайта и настройки описания сайта в результатах поиска «Яндекса» [11].

Статистический анализ выполнялся с помощью программы Microsoft Excel. Excel – это программа для работы с электронными таблицами, разработанная компанией Microsoft. Она входит в состав офисного пакета Microsoft Office. Excel предоставляет пользователю мощные инструменты для создания, форматирования, анализа и визуализации данных в виде таблиц [12].

Основные возможности Excel включают в себя:

* электронные таблицы – Excel представляет собой удобное средство для организации и визуализации статистических данных в виде таблиц с ясной структурой;
* формулы и функции – Excel предоставляет широкий набор встроенных формул и функций для выполнения различных математических и статистических операций;
* графики и диаграммы – Excel позволяет строить различные виды графиков и диаграмм для визуализации данных, что облегчает визуализацию статистических тенденций и зависимостей;
* анализ данных – пользователи могут проводить анализ статистических данных с использованием фильтров, сортировки, условного форматирования и других инструментов;
* работа с большим объемом данных – Excel обладает высокой производительностью и может обрабатывать обширные наборы статистических данных, что важно для статистических исследований.

# Группировка данных

Статистическая сводка – это комплекс последовательных операций по первичной обработке данных с целью выявления характерных черт и закономерностей изучаемого явления. Сводка включает в себя подсчет групповых и общих итогов, получение производных показателей (средних, относительных величин), систематизацию, группировку данных и составление таблиц и графиков. Задача сводки заключается в предоставлении характеристики объекта исследования с помощью системы статистических показателей, а также в выявлении и измерении его существенных черт и особенностей [13].

Процесс сводки статистических данных включает в себя 3 этапа:

1. Систематизация и группировка данных: на данном этапе проводится организация и классификация данных, полученных в результате наблюдений, с целью создания систематизированной структуры;
2. Уточнение системы показателей: этот этап включает в себя доработку системы показателей, которые будут использоваться для количественной характеристики свойств и особенностей изучаемого объекта, с целью более точного и полного отражения характеристик;
3. Исчисление показателей и представление данных: на данном этапе производится расчет статистических показателей, а также их представление в виде таблиц, статистических рядов, графиков и/или диаграмм для наглядного и понятного восприятия обобщенных данных.

Группировка – это статистический метод, при котором вся исследуемая совокупность разделяется на несколько групп по какому-либо признаку [13]. К задачам группировки относятся:

* выделение социально-экономических типов явлений;
* изучение структуры и структурных сдвигов, происходящих в социально-экономических явлениях;
* анализ взаимосвязей между явлениями;
* выявление влияния отдельных единиц на средние итоговые показатели.

Группировочный признак (основание группировки) – это признак, по которому осуществляется группировка. По количеству группировочных признаков выделяют простую группировку (если для построения группировки используется один группировочный признак) и сложную, многомерную группировку (если для построения группировки используется несколько группировочных признаков).

В зависимости от решаемых задач различают следующие виды группировки:

* типологическая – разделение совокупности на однородные группы;
* структурная – разделение однородной совокупности на группы, характеризующие ее структуру по какому-либо варьирующему признаку;
* аналитическая (факторная) – выявление взаимосвязи между изучаемыми явлениями и их признаками.

Интервал группировки – это значение варьирующего признака, которое лежит между нижней и верхней границами (наименьшим и наибольшим значениями группировки). Величина (ширина) интервала – это разница между верхней и нижней границами.

Интервалы могут быть равными (когда разность между максимальным и минимальным значениями в каждом из интервалов одинакова или значение количественного признака изменяется равномерно) или неравными (когда значения признака варьируются неравномерно и в значительных размерах); открытыми (когда у интервала есть только одна граница) или закрытыми (когда имеются обе границы).

Исходные данные были сгруппированы и представлены в виде таблицы со столбцами, соответствующими признакам: «Тип сайта», «Страна», «Год создания», «Посетители», «Просмотры», «Среднее время посещения», «Рейтинг сайта», «Доступность в России». В строках таблицы представлены сайты. Таблица с исходными данными представлена в приложении 2.

На основании исходных данных с помощью Мастера сводных таблиц были построены сводные таблицы по различным признакам (табл. 1-6). Затем полученные данные были представлены с помощью гистограмм (рис. 1-6).

Таблица 1

Сводная таблица: количество посетителей по типам сайта

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **Названия строк** | **Сумма по столбцу Посети-**  **тели, млн чел/мес** | **Среднее по столбцу Посети-**  **тели, млн чел/мес** | **Минимум в столбце Посети-**  **тели, млн чел/мес** | **Максимум в столбце Посети-тели, млн чел/мес** | **Число элементов в столбце Посети-тели** |
| Поисковик | 10940 | 1367,5 | 28 | 5560 | 8 |
| Энциклопедия | 5986 | 2993 | 786 | 5200 | 2 |
| Видеохостинг | 5607 | 1869 | 39 | 5442 | 3 |
| Социальная сеть | 5235 | 523,5 | 14 | 2432 | 10 |
| Новости | 3994 | 234,9411765 | 4 | 3440 | 17 |
| Почта | 1929 | 643 | 18 | 1823 | 3 |
| Трансляции | 1553 | 776,5 | 610 | 943 | 2 |
| Маркетплейс | 1508 | 301,6 | 7 | 650 | 5 |
| ПО | 1402 | 701 | 237 | 1165 | 2 |
| Карты | 1312 | 437,3333333 | 34 | 1234 | 3 |
| Мессенджер | 1240 | 413,3333333 | 18 | 1087 | 3 |
| Вопросы и ответы | 567 | 283,5 | 15 | 552 | 2 |
| Музыка | 496 | 248 | 16 | 480 | 2 |
| Фотохостинг | 493 | 246,5 | 52 | 441 | 2 |
| Фильмы, сериалы | 438 | 219 | 22 | 416 | 2 |
| Прогноз погоды | 295 | 147,5 | 40 | 255 | 2 |
| Видеотелефония | 238 | 79,33333333 | 47 | 115 | 3 |
| Спорт | 87 | 43,5 | 18 | 69 | 2 |
| Поиск работы | 37 | 18,5 | 6 | 31 | 2 |
| **Общий итог** | **43357** | **578,0933** | **4** | **5560** | **75** |

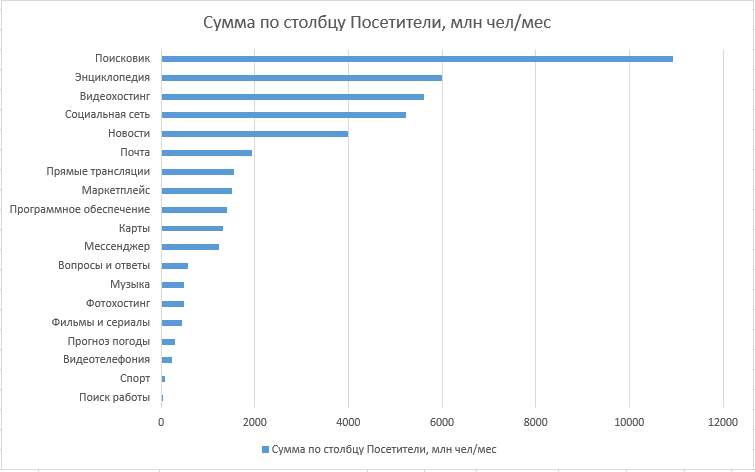


Рис. 1. Гистограмма: количество посетителей по типам сайта

По данной сводной таблице и гистограмме можно определить, какие тематики сайтов являются наиболее популярными в данной выборке (энциклопедии, поисковики, видеохостинги), а какие наименее популярными (видеотелефония, спортивные сайты, сервисы по поиску работы).

Таблица 2

Сводная таблица: количество просмотров по странам

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **Названия строк** | **Сумма по столбцу Про-смотры, млн просм./мес** | **Среднее по столбцу Про-смотры, млн просм./мес** | **Минимум по столбцу Про-**  **смотры, млн просм./мес** | **Максимум по столбцу Про-**  **смотры, млн просм./мес** | **Число элементовв столбце Про-смотры** |
| США | 143221 | 5115,035714 | 200 | 23237 | 28 |
| Китай | 23809 | 11904,5 | 68 | 23741 | 2 |
| Россия | 18512 | 440,7619048 | 9 | 5123 | 42 |
| Швеция | 3156 | 3156 | 3156 | 3156 | 1 |
| Великобритания | 576 | 576 | 576 | 576 | 1 |
| Израиль | 33 | 33 | 33 | 33 | 1 |
| **Общий итог** | **189307** | **2524,093** | **9** | **23741** | **75** |

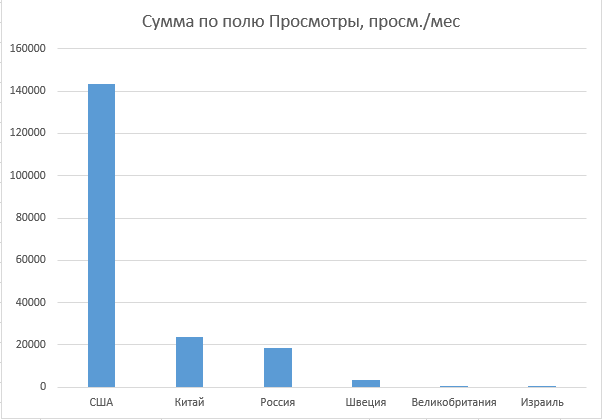


Рис. 2. Гистограмма: количество просмотров по странам

Здесь мы можем увидеть, что наиболее посещаемыми в данной выборке являются сайты, созданные в США. Многие американские сайты популярны во всем мире, в том числе и в России. Именно этим объясняются такие результаты. На 2 месте – Китай, самая густонаселённая страна мира. Россия расположилась на 3 месте. Посещаемость сайтов из остальных стран в данной выборке незначительна.

Таблица 3

Сводная таблица: рейтинг по годам создания

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **Названия строк** | **Сумма по столбцу Рейтинг сайта, баллов** | **Среднее по столбцу Рейтинг сайта, баллов** | **Минимум по столбцу Рейтинг сайта, баллов** | **Максимум по столбцу Рейтинг сайта, баллов** | **Число элементов в столбце Рейтинг сайта** |
| 1975 | 65 | 65 | 65 | 65 | 1 |
| 1994 | 69 | 69 | 69 | 69 | 1 |
| 1995 | 161 | 80,5 | 76 | 85 | 2 |
| 1996 | 143 | 71,5 | 65 | 78 | 2 |
| 1997 | 336 | 84 | 77 | 91 | 4 |
| 1998 | 473 | 78,83333333 | 66 | 86 | 6 |
| 1999 | 689 | 76,55555556 | 65 | 94 | 9 |
| 2000 | 403 | 80,6 | 68 | 90 | 5 |
| 2001 | 161 | 80,5 | 71 | 90 | 2 |
| 2002 | 83 | 83 | 83 | 83 | 1 |
| 2003 | 173 | 86,5 | 86 | 87 | 2 |
| 2004 | 516 | 86 | 65 | 95 | 6 |
| 2005 | 505 | 84,16666667 | 78 | 89 | 6 |
| 2006 | 504 | 72 | 55 | 84 | 7 |
| 2007 | 75 | 75 | 75 | 75 | 1 |
| 2008 | 236 | 78,66666667 | 56 | 97 | 3 |
| 2009 | 199 | 66,33333333 | 47 | 92 | 3 |
| 2010 | 342 | 85,5 | 79 | 93 | 4 |
| 2011 | 71 | 71 | 71 | 71 | 1 |
| 2013 | 226 | 75,33333333 | 67 | 91 | 3 |
| 2015 | 123 | 61,5 | 32 | 91 | 2 |
| 2016 | 137 | 68,5 | 65 | 72 | 2 |
| 2019 | 58 | 58 | 58 | 58 | 1 |
| 2021 | 72 | 72 | 72 | 72 | 1 |
| **Общий итог** | **5820** | **77,6** | **32** | **97** | **75** |

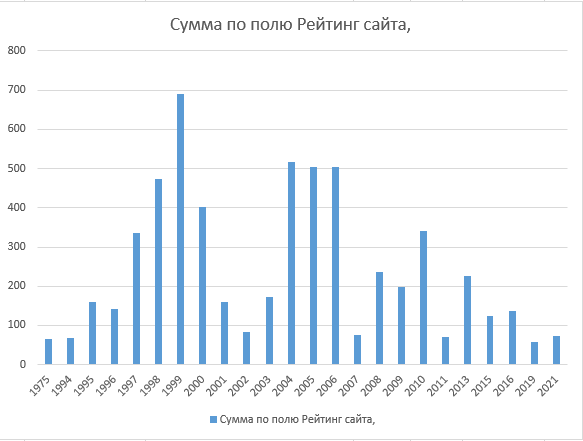


Рис. 3. Гистограмма: рейтинг по годам создания

В данной сводке на гистограмме отчетливо виден всплеск рейтинга у сайтов, созданных в 1999 году. Следом идут сайты, созданные с 2004 по 2006 год. При этом, у сайтов, созданных в 2010-х и 2020-х годах, рейтинг существенно снизился по сравнению с сайтами конца 1990-х и 2000-х годов.

Таблица 4

Сводная таблица: количество посетителей и просмотров по рейтингу

|  |  |  |
| --- | --- | --- |
| **Названия строк** | **Среднее по столбцу Посетители, млн чел/мес** | **Среднее по столбцу Просмотры, млн просм./мес** |
| 30-39 | 120 | 512 |
| 40-49 | 552 | 2357 |
| 50-59 | 147 | 627 |
| 60-69 | 393,8333 | 1680,833 |
| 70-79 | 419,1333 | 1802,733 |
| 80-89 | 808,0455 | 3451,136 |
| 90-100 | 746,6364 | 3187,727 |
| **Общий итог** | **593,2656** | **2536,313** |

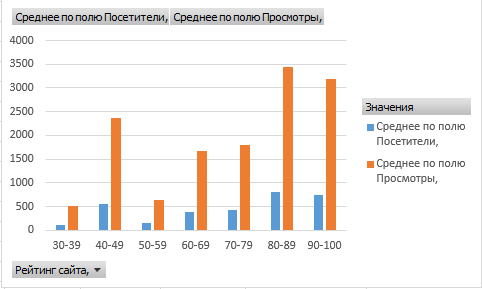


Рис. 4. Гистограмма: количество посетителей и просмотров по рейтингу

В этой сводке данных можно заметить рост количества просмотров и посетителей при росте рейтинга. Однако есть отклонение: сайты с рейтингом 40-49 гораздо популярнее сайтов, у которых рейтинг составляет от 50 до 79 баллов. Причина такого скачка может заключаться в нестабильной работе какого-либо популярного сайта, или в выходе на таком сайте обновлений, которые не понравились пользователям, из-за чего они «обрушили» рейтинг сайта.

Также наблюдается небольшой спад посещаемости у сайтов с рейтингом 90-100 по сравнению с сайтами рейтинга 80-89. Это можно объяснить тем, что у какого-либо сайта из-за низкой посещаемости может быть мало отзывов. Однако, если большинство этих отзывов положительные, то рейтинг у сайта будет выше 90.

Таблица 5

Сводная таблица: рейтинг по количеству просмотров

|  |  |
| --- | --- |
| **Названия строк** | **Среднее по столбцу Рейтинг сайта, баллов** |
| 1-100 | 77,69444 |
| 101-200 | 77,66667 |
| 201-300 | 70,5 |
| 401-500 | 87 |
| 501-600 | 47 |
| 601-700 | 69 |
| 701-800 | 85 |
| 901-1000 | 83,33333 |
| 1001-1100 | 76 |
| 1301-1400 | 73 |
| 2401-2500 | 65 |
| 2901-3000 | 87 |
| 3401-3500 | 76 |
| 5101-5200 | 90 |
| 5401-5500 | 89 |
| 5501-5600 | 84 |
| **Общий итог** | **77,67188** |

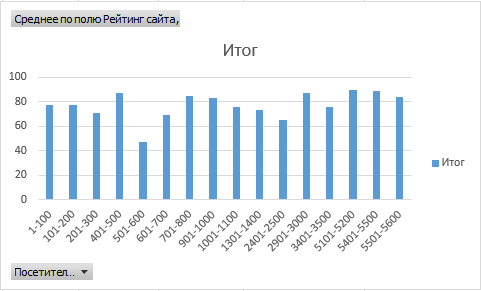


Рис. 5. Гистограмма: рейтинг по количеству просмотров

В данной сводке зависимость рейтинга сайта от количества просмотров не наблюдается.

Таблица 6

Сводная таблица: количество посетителей по количеству просмотров

|  |  |
| --- | --- |
| **Названия строк** | **Среднее по столбцу Посетители,** |
| 1-1000 | 56,15909 |
| 1001-2000 | 337,25 |
| 2001-3000 | 601 |
| 3001-4000 | 771 |
| 4001-5000 | 1002,4 |
| 5001-6000 | 1330 |
| 10001-11000 | 2432 |
| 12001-13000 | 2989 |
| 14001-15000 | 3440 |
| 22001-23000 | 5200 |
| 23001-24000 | 5501 |
| **Общий итог** | **593,2656** |

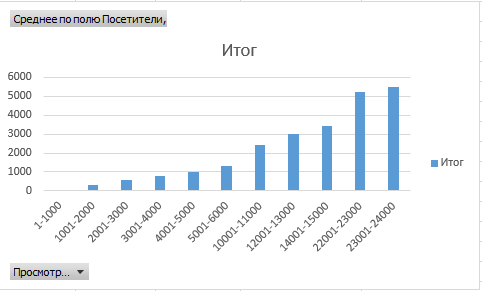


Рис. 6. Гистограмма: рейтинг по количеству просмотров

Из данной сводки отчетливо видно, что с ростом числа посетителей растет и число просмотров. Это очевидная взаимосвязь, однако построить данную сводку необходимо с целью проверки правильности данных.

# Изучение взаимосвязи

Задачей статистики является изучение статистических закономерностей для предсказания и управления социально экономическими процессами.

Прежде чем приступить к изучению связи между явлениями необходимо выяснить вид связи между факторным и результативным признаками.

По степени причинно-следственной определенности различают два вида связи:

1. функциональная связь – значение результативного признака (у) полностью определяется значением факторного признака (х);
2. корреляционная связь – значение результативного признака (у) определяется значением факторного признака (х), при этом имеет место влияние прочих, неучтенных, факторов [14].

По направлению связь может быть прямой и обратной. По аналитическому выражению – линейной и нелинейной. По количеству факторных признаков – линейной и нелинейной.

При исследовании корреляционных связей одной из главных задач является определение формы корреляционной связи. Форма связи – это тенденция, которая проявляется в изменении результативного признака в связи с изменением признака-фактора.

Построение и анализ корреляционной модели связи осуществляются с помощью корреляционно-регрессионного анализа. Он заключается в построении и анализе уравнения регрессии (корреляционной связи), приближенно выражающей зависимость результативного признака от признаков-факторов.

Дисперсия – это один из основных показателей в статистике. Она отражает меру разброса значений в наборе данных. Дисперсия измеряет, насколько далеко каждое значение в наборе данных находится от среднего значения. Чем больше дисперсия, тем больше изменчивость данных.

Существуют различные виды дисперсии, например, несмещенная, средняя внутригрупповая, межгрупповая, общая и др.

Коэффициент корреляции – это статистическая мера, которая измеряет степень линейной зависимости между двумя переменными. Вычисляется по формуле (2):

где – среднее значение факторного признака x, – среднее значение результативного признака у, – дисперсии признаков x и y.

По значению коэффициента можно определить наличие связи и ее степень:

* менее 0,3 – очень слабая связь, практически отсутствует;
* 0,3 – 0,5 – слабая связь;
* 0,5 – 0,7 – средняя связь;
* более 0,7 – тесная связь.

Коэффициент детерминации (R²) является статистическим показателем, используемым в регрессионном анализе, чтобы измерить, насколько хорошо независимая переменная объясняет изменения зависимой переменной. Он предоставляет меру того, насколько точно модель соответствует данным. Вычисляется по формуле (3):

где – межгрупповая дисперсия, – общая дисперсия.

## Изучение взаимосвязи между двумя количественными признаками

Найдем взаимосвязи между количественными признаками по сводным таблицам 4-6. Для этого необходимо построить график зависимости среднего значения одного признака от значения другого. Затем по полученным значениям нужно провести линию тренда (рис. 7-9).

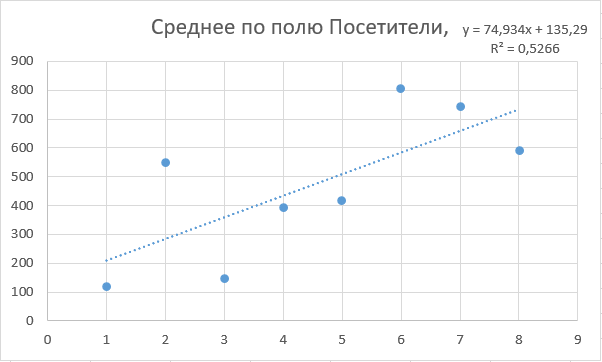


Рис. 7. График зависимости среднего количества посетителей от рейтинга

Найдем коэффициент детерминации: R2 = 0,5266 > 0,5, т.е. взаимосвязь между средним количеством посетителей и рейтингом сайта есть.

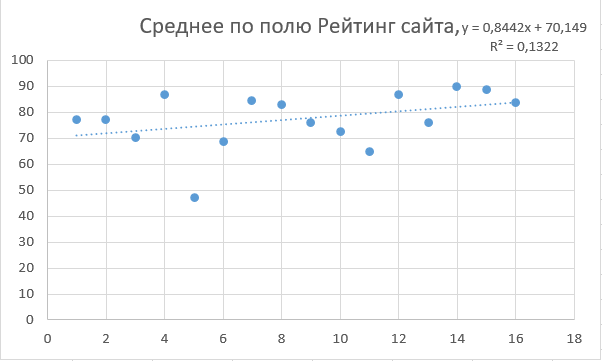


Рис. 8. График зависимости среднего рейтинга сайта от количества посетителей

Найдем коэффициент детерминации: R2 = 0,1322 < 0,5, т.е. взаимосвязи между средним рейтингом сайта и количеством посетителей нет.

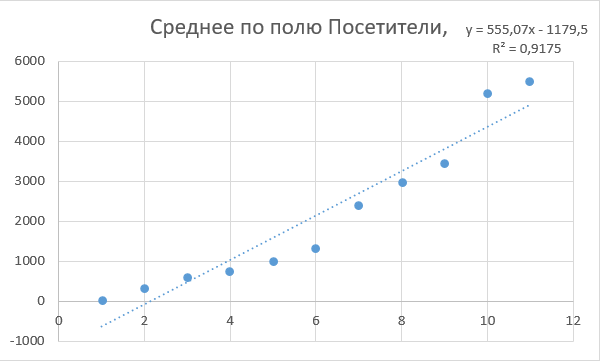


Рис. 9. График зависимости среднего количества посетителей от количества просмотров

Найдем коэффициент детерминации: R2 = 0,9175 > 0,5, т.е. взаимосвязь между средним количеством посетителей и количеством просмотров есть.

## Изучение взаимосвязи между количественными и качественными признаками

Для изучения взаимосвязей между количественным и качественным признаком необходимо рассчитать коэффициент детерминации вручную. Для этого понадобится вычислить значения несмещенной, средней внутригрупповой, межгрупповой и общей дисперсий (рис. 10-12).



Рис. 10. Расчет коэффициента детерминации между типом сайта и количеством посетителей

Проанализируем полученный коэффициент детерминации: R2 = 0,2624 < 0,5, т.е. взаимосвязи между типом сайта и количеством посетителей нет;

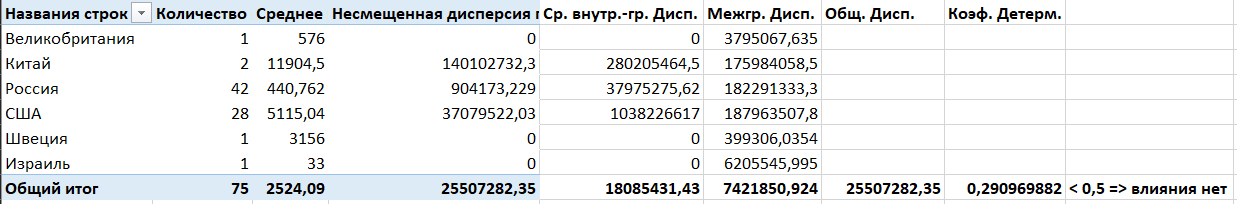


Рис. 11. Расчет коэффициента детерминации между страной и количеством просмотров

Проанализируем полученный коэффициент детерминации: R2 = 0,291 < 0,5, т.е. взаимосвязи между страной сайта и количеством просмотров нет;

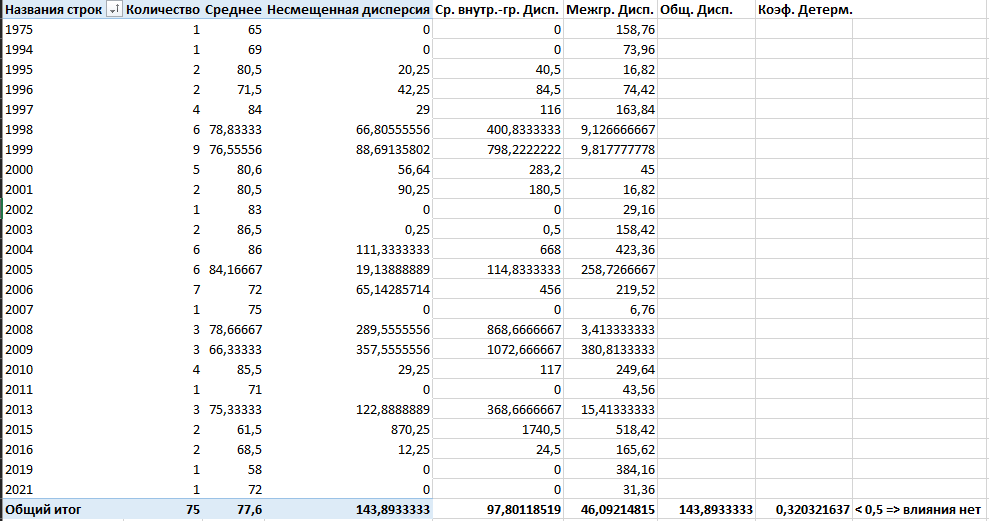


Рис. 12. Расчет коэффициента детерминации между годом и рейтингом сайта

Проанализируем полученный коэффициент детерминации: R2 = 0,3264 < 0,5, т.е. взаимосвязи между годом создания сайта и рейтингом нет.

# Ряды динамики

Ряд динамики – это совокупность значений какого-либо показателя за несколько последовательных моментов (периодов) времени.

Ряды динамики могут быть стационарными или нестационарными (содержащими тренд либо сезонную составляющую).

Основной задачей исследования динамического ряда является выявление тенденций, периодичности, случайных компонент в целях использования будущих значений ряда [15].

Базисный темп роста – это изменение значения величины по сравнению с базовым показателем.

Цепной темп роста – это изменение значения величины по сравнению с предыдущим значением этой величины.

Построим ряд динамики количества пользователей интернета по годам. Для этого рассчитаем базисный и цепной абсолютный прирост, а также базисный и цепной темпы роста и прироста (табл. 7).

Таблица 7

Расчёт темпов роста количества пользователей интернета

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| Год | Кол-во пользователей,  млрд чел | абсолютный прирост | | темпы роста | | темпы прироста | |
| базисный | цепной | базисный | цепной | базисный | цепной |
| 2012 | 2,177 | – | – | – | – | – | – |
| 2013 | 2,431 | 0,254 | 0,254 | 1,116674 | 1,116674 | 0,116674 | 0,116674 |
| 2014 | 2,692 | 0,515 | 0,261 | 1,236564 | 1,107363 | 0,236564 | 0,107363 |
| 2015 | 2,916 | 0,739 | 0,224 | 1,339458 | 1,08321 | 0,339458 | 0,08321 |
| 2016 | 3,282 | 1,105 | 0,366 | 1,507579 | 1,125514 | 0,507579 | 0,125514 |
| 2017 | 3,64 | 1,463 | 0,358 | 1,672026 | 1,10908 | 0,672026 | 0,10908 |
| 2018 | 3,95 | 1,773 | 0,31 | 1,814424 | 1,085165 | 0,814424 | 0,085165 |
| 2019 | 4,212 | 2,035 | 0,262 | 1,934773 | 1,066329 | 0,934773 | 0,066329 |
| 2020 | 4,418 | 2,241 | 0,206 | 2,029398 | 1,048908 | 1,029398 | 0,048908 |
| 2021 | 4,758 | 2,581 | 0,34 | 2,185576 | 1,076958 | 1,185576 | 0,076958 |
| 2022 | 4,95 | 2,773 | 0,192 | 2,273771 | 1,040353 | 1,273771 | 0,040353 |
|  | | | | | | | |
| Средние уровни ряда | | Ср. абс. прирост | | Сред. темп роста | | Ср. темп прироста | |
| 3,584 | | 0,252 | | 1,086 | | 8,56% | |

На основании полученных данных построим график динамического ряда (рис. 13).

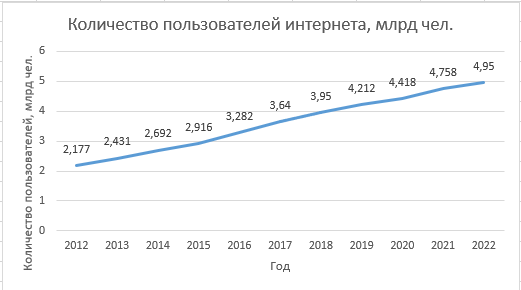


Рис. 13. Ряд динамики: количество пользователей интернета по годам

По полученному динамическому ряду видно, что количество пользователей интернета с каждым годом растет, в среднем на 8,56% в год.

# Заключение

В ходе выполнения курсовой работы была исследована зависимость посещаемости сайтов в интернете от различных признаков. В рамках исследования была проведена сводка и группировка данных, изучены взаимосвязи, построен динамический ряд.

В результате статистического анализа была выявлена взаимосвязь между средним количеством посетителей и рейтингом сайта, а также взаимосвязь между средним количеством посетителей и количеством просмотров.

Остальные признаки сайта (тема, год создания, страна и др.) на посещаемость значительно не влияют.

Исходя из результатов курсовой работы, можно порекомендовать создателям сайта делать упор на качество размещаемого контента на сайтах, что повысит рейтинг сайта, а также на привлечение внимания посетителей, например, с помощью рекламы, что поможет повысить посещаемость сайта.

# Список литературы

1. Методика и организация самостоятельной работы студентов: учебно-методическое пособие / Е.В. Ершов, Л.Н. Виноградова, В.В. Селивановских [и др.]. – Череповец: ЧГУ, 2015.
2. Статистика — Википедия [электр.ресурс] https://ru.wikipedia.org/wiki/Статистика. Дата обращения: 15.12.2023.
3. Прикладная статистика — Википедия [электр.ресурс] https://ru.wikipedia.org/wiki/Прикладная\_статистика. Дата обращения: 15.12.2023.
4. Инфографика: сколько в мире сайтов — Вокруг света [электр.ресурс] https://www.vokrugsveta.ru/articles/infografika-skolko-v-mire-saitov-id672700. Дата обращения: 15.12.2023.
5. Сайт — Википедия [электр.ресурс] https://ru.wikipedia.org/wiki/Сайт. Дата обращения: 15.12.2023.
6. SimilarWeb [электр.ресурс] https://www.similarweb.com/ru. Дата обращения: 15.12.2023.
7. LiveInternet [электр.ресурс] https://www.liveinternet.ru. Дата обращения: 15.12.2023.
8. Рейтинг@Mail.Ru [электр.ресурс] https://top.mail.ru. Дата обращения: 15.12.2023.
9. Рамблер / Топ-100 [электр.ресурс] https://top100.rambler.ru. Дата обращения: 15.12.2023.
10. Яндекс.Метрика [электр.ресурс] https://metrika.yandex.ru. Дата обращения: 15.12.2023.
11. Яндекс.Вебмастер [электр.ресурс] https://webmaster.yandex.ru. Дата обращения: 15.12.2023.
12. Microsoft Excel — Википедия [электр.ресурс] https://ru.wikipedia.org/wiki/Microsoft\_Excel. Дата обращения: 15.12.2023.
13. Сводка и группировка статистических данных — PPT-Online [электр.ресурс] https://ppt-online.org/453048. Дата обращения: 15.12.2023.
14. Статистическое изучение взаимосвязи социально-экономических явлений — PPT-Online [электр.ресурс] https://ppt-online.org/63442. Дата обращения: 15.12.2023.
15. Временные ряды — PPT-Online [электр.ресурс] https://ppt-online.org/303672. Дата обращения: 15.12.2023.

# Приложение 1

МИНОБРНАУКИ РОССИИ

Федеральное государственное бюджетное образовательное учреждение высшего профессионального образования

«ЧЕРЕПОВЕЦКИЙ ГОСУДАРСТВЕННЫЙ УНИВЕРСИТЕТ»

Институт информационных технологий

наименование института (факультета)

Математическое и программное обеспечение ЭВМ

наименование кафедры

Прикладная статистика

наименование дисциплины в соответствии с учебным планом

УТВЕРЖДАЮ

Зав. кафедрой МПО ЭВМ ,

д.т.н., профессор Ершов Е.В.

« » 2023 г.

Статистический анализ социально-экономических показателей

Техническое задание на курсовую работу

Листов 7

Руководитель: Гонтарева И. Б.

Ф.И.О. преподавателя

Исполнитель:

студент 1ПИб-02-3оп-22

группа

Маркелов С.А.

Фамилия, имя, отчество

2023 год

Введение

Данная курсовая работа посвящена статистическому изучению посещаемости наиболее популярных сайтов в интернете и нахождению взаимосвязей между популярностью сайтов и различными признакам.

1. Основания для разработки

Основанием для разработки является задание на курсовую работу по дисциплине «Прикладная статистика», выданное на кафедре МПО ЭВМ ИИТ ЧГУ.

Дата утверждения: 3 октября 2023 года.

Наименование темы разработки: Изучение влияния различных признаков на популярность сайтов.

2. Назначение разработки

Данная курсовая работа поможет выяснить, существуют ли какие-либо факторы, которые могут влиять на популярность разрабатываемого сайта.

3. Требования к разработке

* 1. Требования к функциональным характеристикам

Все таблицы, диаграммы и графики должны быть правильно оформлены и понятны логически. Исходные данные должны быть взяты только из авторитетных, проверенных источников.

* 1. Требования к надёжности

Во избежание потери или повреждения данных необходимо хранить копии файла Excel с данными на резервном носителе информации.

* 1. Условия эксплуатации

Для корректной работы разработки необходимо:

* наличие любого устройства, на котором возможно запускать файлы формата .xls (компьютер, ноутбук);
* наличие на устройстве современной версии любой операционной системы.
  1. Требования к составу и параметрам технических средств

Для корректной работы с данными необходимо:

* оперативная память: не менее 1 Гб для 32-разрядной ОС, 2 Гб для 64-разрядной ОС;
* процессор не менее чем с 2 ядрами и тактовой частотой не ниже 1 ГГц;
* разрешение экрана не менее 800 х 600;
* видеокарта с видеопамятью не менее 1 Гб;
* наличие мыши и клавиатуры.
  1. Требования к информационной и программной совместимости

Для корректной работы с данными необходимо:

* установленная современная операционная система (Windows 7, 8.1, 10, 11, macOS, Linux);
* установленная на устройстве программа Microsoft Excel из пакета Microsoft Office версии 2016 года или младше.
  1. Требования к маркировке и упаковке

Требования к маркировке и упаковке не предъявляются.

* 1. Требования к транспортированию и хранению

Файлы с данными должны храниться и транспортироваться на носителе информации (флешка, CD-диск).

* 1. Специальные требования

Специальные требования к разработке не предъявляются.

4. Требование к программной документации

* 1. Содержание расчетно-пояснительной записки

Программная документация должна содержать расчётно-пояснительную записку с содержанием:

Титульный лист;

Оглавление;

1. Введение;

2. Описание предметной области;

3. Группировка данных;

4. Изучение взаимосвязи;

5. Ряды динамики;

Заключение;

Список литературы;

Приложение 1. Техническое задание;

Приложение 2. Исходные данные.

* 1. Требования к оформлению

В данном пункте представлены требования к оформлению документации в соответствии с ГОСТ (табл. П1.1).

Таблица П1.1

Требования к оформлению документации

|  |  |
| --- | --- |
| Документ | Печать на отдельных листах формата А4 (210х297 мм); оборотная сторона не заполняется; листы нумеруются. Печать возможна ч/б. Листы вложены в пластиковую папку скоросшивателя. |
| Страницы | Ориентация – книжная; отдельные страницы, при необходимости, альбомная. Поля: верхнее, нижнее – по 2 см, левое – 3 см , правое – 1 см. |
| Абзацы | Межстрочный интервал – 1,5, перед и после абзаца – 0. |
| Шрифты | Кегль – 14. В таблицах шрифт 12 |
| Рисунки | Подписывается под ним по центру: Рис.Х. Название. В приложениях: Рис.П1.3. Название |
| Таблицы | Подписывается: над таблицей, выравнивание по правому: «Таблица Х». В следующей строке по центру Название Надписи в «шапке» (имена столбцов, полей) – по центру. В теле таблицы (записи) текстовые значения – выровнены по левому краю, числа, даты – по правому. |

5. Технико-экономические показатели

Требования не предъявляются.

6. Стадии и этапы разработки

В данном пункте описаны стадии и этапы проведения статистического исследования (табл. П1.2).

Таблица П1.2

Стадии и этапы разработки

|  |  |  |  |
| --- | --- | --- | --- |
| Наименование этапа разработки | Сроки разработки | Результат выполнения | Отметка о выполнении |
| Определение темы для курсовой работы | 03.10.2023 | Утверждена тема для разработки |  |
| Оформление технического задания | 03.10.2023 | Оформленное техническое задание |  |
| Сбор данных | 03.10.2023 | Собраны исходные данные для проведения статистического анализа |  |
| Сводка и группировка данных | 17.10.2023 | Подготовлены сводные таблицы и диаграммы |  |
| Нахождение взаимосвязей между количественными и качественными признаками | 31.10.2023 | Найденные взаимосвязи между признаками |  |
| Нахождение взаимосвязей между количественными признаками | 14.11.2023 | Найденные взаимосвязи между признаками |  |
| Построение динамических рядов | 28.11.2023 | Построенные динамические ряды |  |
| Оформление сопроводительной документации | 12.12.2023 | Оформленная сопроводительная документация |  |

7. Порядок контроля и приемки

В данном пункте описан порядок контроля и приемки курсовой работы (табл. П1.3).

Таблица П1.3

Порядок контроля и приемки

|  |  |  |  |
| --- | --- | --- | --- |
| Наименование контрольного этапа выполнения курсовой работы | Сроки контроля | Результат выполнения | Отметка о приемке результата контрольного этапа |
| Сдача технического задания | 17.10.2023 | Согласованное техническое задания |  |
| Сдача расчетно-пояснительной записки | 15.12.2023 | Согласованная расчетно-пояснительная записка |  |
| Сдача курсовой работы | 19.12.2023 | Получение оценки за выполненную работу |  |

# Приложение 2

Таблица П2.1

Исходные данные: информация о сайтах

36

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **Название сайта** | **Тип сайта** | **Страна** | **Год создания** | **Посетители, млн чел/мес** | **Просмотры, млн просм./мес** | **Среднее время посещения, мин** | **Рейтинг сайта, баллов** | **Доступность в России** |
| **1** | **2** | **3** | **4** | **5** | **6** | **7** | **8** | **9** |
| 2ГИС | Карты | Россия | 1999 | 34 | 153 | 01:17 | 94 | да |
| Amazon | Маркетплейс | США | 1994 | 650 | 2775 | 07:08 | 69 | да |
| Baidu | Поисковик | Китай | 2000 | 5560 | 23741 | 05:06 | 84 | да |
| Bing | Поисковик | США | 2009 | 1070 | 4568 | 06:42 | 60 | да |
| Discord | Видеотелефония | США | 2015 | 76 | 324 | 06:57 | 91 | да |
| DuckDuckGo | Поисковик | США | 2008 | 952 | 4065 | 08:41 | 97 | да |
| Ebay | Маркетплейс | США | 1995 | 124 | 529 | 06:50 | 85 | да |
| Facebook\* | Социальная сеть | США | 2004 | 2432 | 10384 | 10:31 | 65 | нет |
| Fandom | Энциклопедия | США | 2004 | 786 | 3356 | 05:11 | 88 | да |
| Gmail | Почта | США | 2004 | 1823 | 7854 | 02:16 | 92 | да |
| Google | Поисковик | США | 1997 | 2989 | 12763 | 10:38 | 87 | да |
| Google Maps | Карты | США | 2005 | 1234 | 5348 | 02:02 | 87 | да |
| Google Play | ПО | США | 2008 | 1165 | 2756 | 02:49 | 83 | да |
| HeadHunter | Поиск работы | Россия | 2000 | 31 | 489 | 05:18 | 77 | да |
| Instagram\* | Социальная сеть | США | 2010 | 756 | 3228 | 08:15 | 82 | нет |
| Lenta.ru | Новости | Россия | 1999 | 51 | 271 | 01:06 | 65 | да |
| Life | Новости | Россия | 2013 | 17 | 48 | 01:03 | 67 | да |
| LinkedIn | Социальная сеть | США | 2002 | 180 | 768 | 07:32 | 83 | нет |

\* Facebook и Instagram принадлежат компании Meta, признанной экстремистской организацией и запрещенной в РФ.

Продолжение табл. П2.1

37

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **1** | **2** | **3** | **4** | **5** | **6** | **7** | **8** | **9** |
| Mail.ru | Поисковик | Россия | 1998 | 101 | 431 | 08:16 | 84 | да |
| Microsoft | ПО | США | 1975 | 237 | 1011 | 04:13 | 65 | частично |
| Netflix | Фильмы, сериалы | США | 1997 | 416 | 1776 | 05:41 | 81 | нет |
| Ozon | Маркетплейс | Россия | 1998 | 311 | 3542 | 04:14 | 86 | да |
| Pinterest | Фотохостинг | США | 2010 | 441 | 1883 | 06:05 | 93 | да |
| Quora | Вопросы-ответы | США | 2009 | 552 | 2357 | 02:55 | 47 | нет |
| Rambler | Поисковик | Россия | 1996 | 28 | 119 | 01:17 | 65 | да |
| Reddit | Социальная сеть | США | 2005 | 225 | 960 | 08:31 | 89 | да |
| Russia Today | Новости | Россия | 2005 | 21 | 52 | 01:10 | 82 | да |
| Rutube | Видеохостинг | Россия | 2006 | 39 | 166 | 05:15 | 55 | да |
| Shutterstock | Фотохостинг | США | 2003 | 52 | 241 | 02:39 | 86 | да |
| Sports | Спорт | Россия | 1998 | 69 | 309 | 03:27 | 82 | да |
| Spotify | Музыка | Швеция | 2006 | 480 | 3156 | 03:13 | 75 | нет |
| SuperJob | Поиск работы | Россия | 2000 | 6 | 24 | 02:59 | 68 | да |
| Teams | Видеотелефония | США | 2016 | 115 | 512 | 07:26 | 65 | да |
| Telegram | Мессенджер | Велико-британия | 2013 | 135 | 576 | 03:13 | 91 | да |
| TikTok | Социальная сеть | Китай | 2016 | 16 | 68 | 03:51 | 72 | частично |
| Twitch | Трансляции | США | 2011 | 943 | 4026 | 09:14 | 71 | да |
| Twitter | Социальная сеть | США | 2006 | 1330 | 5679 | 10:35 | 73 | нет |
| Ura.ru | Новости | Россия | 2006 | 25 | 134 | 01:24 | 70 | да |
| Viber | Мессенджер | Израиль | 2010 | 18 | 33 | 47:00 | 79 | да |
| VK | Социальная сеть | Россия | 2006 | 202 | 862 | 12:00 | 72 | да |
| VK Видео | Видеохостинг | Россия | 2021 | 126 | 423 | 06:05 | 72 | да |
| WASD | Трансляции | Россия | 2019 | 610 | 2124 | 12:00 | 58 | да |
| Weather | Прогноз погоды | США | 2008 | 255 | 1088 | 01:00 | 56 | нет |
| WhatsApp | Мессенджер | США | 2009 | 1087 | 4641 | 19:49 | 92 | да |

Продолжение табл. П2.1

38

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **1** | **2** | **3** | **4** | **5** | **6** | **7** | **8** | **9** |
| Wildberries | Маркетплейс | Россия | 2004 | 416 | 5123 | 12:00 | 81 | да |
| Yahoo! | Новости | США | 1995 | 3440 | 14688 | 10:01 | 76 | да |
| YouTube | Видеохостинг | США | 2005 | 5442 | 23237 | 20:19 | 89 | да |
| Zoom | Видеотелефония | США | 2013 | 47 | 200 | 03:30 | 68 | да |
| Аргументы и факты | Новости | Россия | 1997 | 34 | 78 | 00:56 | 77 | да |
| Википедия | Энциклопедия | США | 2001 | 5200 | 22204 | 03:55 | 90 | да |
| Газета.Ru | Новости | Россия | 1998 | 45 | 164 | 01:24 | 69 | да |
| Дзен | Поисковик | Россия | 2015 | 120 | 512 | 09:03 | 32 | да |
| Живой Журнал | Социальная сеть | Россия | 1999 | 22 | 95 | 01:17 | 78 | да |
| Известия | Новости | Россия | 2005 | 48 | 204 | 01:24 | 78 | да |
| Кинопоиск | Фильмы, сериалы | Россия | 2003 | 22 | 93 | 01:34 | 87 | да |
| Коммерсантъ | Новости | Россия | 1999 | 19 | 72 | 01:24 | 83 | да |
| Комсомольская правда | Новости | Россия | 1998 | 94 | 431 | 01:14 | 66 | да |
| Мой мир | Социальная сеть | Россия | 2007 | 14 | 41 | 01:07 | 75 | да |
| Московский комсомолец | Новости | Россия | 1999 | 43 | 170 | 01:24 | 65 | да |
| НТВ | Новости | Россия | 1999 | 12 | 29 | 01:15 | 65 | да |
| Одноклассники | Социальная сеть | Россия | 2006 | 58 | 247 | 01:23 | 75 | да |
| Ответы Mail.ru | Вопросы-ответы | Россия | 2006 | 15 | 39 | 01:56 | 84 | да |
| Почта Mail.ru | Почта | Россия | 1998 | 88 | 375 | 02:02 | 86 | да |
| РБК | Новости | Россия | 1996 | 46 | 191 | 01:36 | 78 | да |
| РЕН ТВ | Новости | Россия | 2001 | 14 | 30 | 00:59 | 71 | да |
| РИА Новости | Новости | Россия | 1999 | 64 | 209 | 02:00 | 76 | да |
| Российская газета | Новости | Россия | 1999 | 4 | 9 | 01:44 | 82 | да |
| ТАСС | Новости | Россия | 1999 | 17 | 46 | 00:59 | 81 | да |

Продолжение табл. П2.1

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **1** | **2** | **3** | **4** | **5** | **6** | **7** | **8** | **9** |
| Чемпионат.com | Спорт | Россия | 2005 | 18 | 166 | 02:54 | 80 | да |
| Яндекс | Поисковик | Россия | 1997 | 120 | 512 | 08:47 | 91 | да |
| Яндекс.Карты | Карты | Россия | 2004 | 44 | 187 | 01:36 | 95 | да |
| Яндекс.Маркет | Маркетплейс | Россия | 2000 | 7 | 28 | 03:46 | 84 | да |
| Яндекс.Музыка | Музыка | Россия | 2010 | 16 | 68 | 01:51 | 88 | да |
| Яндекс.Погода | Прогноз погоды | Россия | 2004 | 40 | 170 | 01:09 | 95 | да |
| Яндекс.Почта | Почта | Россия | 2000 | 18 | 76 | 01:15 | 90 | да |

39