Статистические данные за период с 2010 по 2020 год

Автор Фамилия

04.03.2023

Мини-исследование по основным характеристикам объекта.

# Аннотация

Мы исследуем в статье средние показатели за 10 лет: с 2010 по 2020 год. В работе дана общая краткая характеристика величин, рассмотрено распределение … с учетом региональных особенностей и более подробно сделано исследование … с учетом географического положения. В конце работы приведен пример построения простейшей модели машинного обучения для возможного прогнозирования.

# Введение

В настоящее время большое внимание уделяется повышению эффективности деятельности … на основе современных алгоритмов науки о данных, новых научно-методических подходах, использования геоинформационных систем для решения сложных задач [1, с. 560–570]. Одной из задач такого рода является расчет …, исходя из требований статьи 11 123-ФЗ [2] на основе современных технологий. Напомним, что …

Теоретическая основа и практические подходы для решения задачи размещения решались в НИР «…», выполнявшейся в 2022 году [3]. Ключевым подходом являлась реализация алгоритмов оптимизации на основе теории … на языке программирования Python с последующей визуализацией результата в QGIS – свободной кроссплатформенной геоинформационной системе.

В качестве исходных данных указанной выше НИР для реализации алгоритмов выступают средние показателей на основе практических наблюдений, как, например, в [4] и статистических данных.

# 1. Методология исследования

## 1.1 Исходные данные

Данная статья посвящена обобщению информации по … на основе почти 10-летней статистики. В настоящее время …, однако настоящее исследование не потеряло актуальности.

## 1.2 Методы анализа

Анализ исходных данных проводился с использованием с помощью языка программирования R [5–8]. Основным инструментом для работы в R в данной работе является набор библиотек, основанный на tidyverse – коллекции пакетов, объединенных общей базовой философией проектирования, грамматикой и структурой данных. В первую очередь мы используем библиотеку dplyr для работы с табличными данными, ggplot2 для визуализации данных путем добавления новых слоев. Для диагностики данных на основе основных статистических тестов применялась библиотека ggstatsplot. Кроме того, работа с географическими данными с учетом CRS-проекций осуществлялась в библиотеке sf, а картографическая подложка дорожной сети была сделана на основе *OpenStreetMap*-данных. Моделирование осуществлялось в современной библиотеке tidymodels, учет географических особенностей данных для кросс-валидации основан на работе в библиотеке spatialsample.

# 2. Результаты

## 2.1 Средние показатели

Рассмотрим таблицы и графики, отражающие различные распределения показателей. Отметим, что за исследуемый промежуток времени общее суммарное значение показателя составило 33 286 424. Найдем суммарные показатели по различным категориям. Как видно из таблицы ниже, более половины значений приходится на сельские населенные пункты.

Таблица 1: Суммарные показатели по типам населенных пунктов

|  | тип населенного пункта | показатель | процент |
| --- | --- | --- | --- |
| 1 | Сельский населенный пункт | 17 756 913 | 53,35% |
| 2 | Город | 11 240 888 | 33,77% |
| 3 | Вне территории населенного пункта | 2 835 312 | 8,52% |
| 4 | Населенный пункт городского типа | 1 346 053 | 4,04% |
| 5 | Станция | 68 661 | 0,21% |
| 6 | Вахтовый поселок | 15 773 | 0,05% |
| 7 | Жилой поселок при станции | 11 391 | 0,03% |
| 8 | Разъезд, перегон | 10 580 | 0,03% |
| 9 | не указан | 853 | 0,00% |

|  |
| --- |
| Рисунок 1: Карта: значения показателя в Российской Федерации с 2010 по 2020 год по регионам; разделение цвета соответствует средней скорости, равной 31,65 км/ч |

Пример кода:

linear\_reg() |>  
 set\_engine("keras")  
#> Linear Regression Model Specification (regression)  
#>   
#> Computational engine: keras

# 3. Обсуждение

Из полученных результатов можно сделать следующие выводы: …

# Заключение

В работе были кратко рассмотрены статистические данные по некоторым основным параметрам, характеризующим …

|  |
| --- |
| Замечание |
| Обратите внимание, что в нижней части содержания HTML-страницы имеется ссылка на MS Word документ в разделе «Другие форматы». MS Word-документ рендерится автоматически благодаря [шаблону с соответствующими стилями](https://quarto.org/docs/output-formats/ms-word-templates.html) и соответствующему указанию в YAML-преамбуле в Quarto версии >1.3, как это описано [на странице](https://quarto.org/docs/prerelease/1.3/multi-format.html). |

# Список литературы

1. Liu S. и др. [A Generalized Framework for Measuring Pedestrian Accessibility around the World Using Open Data](https://doi.org/10.1111/gean.12290) // Geographical Analysis. 2022. Т. 54, № 3. С. 559–582.

2. Федеральный закон от 11 мая 2008 г. № 123-ФЗ «Технический регламент» (с изменениями и дополнениями). 2008.

3. Баталов А.А., Маркин О.О., Исполнитель И.И. Отчет о НИР «Разработка научно-практических подходов и методов». Наукоград, 2022.

4. Иванов А.Н. Диссертация на соискание ученой степени кандидат технических наук «Информационно-аналитическая поддержка управления». Москва, 2018.

5. Wickham H. R and S // Encyclopedia of Environmetrics. 2013.

6. Wickham H. Statistical graphics // Encyclopedia of Environmetrics. 2013.

7. Wickham H., Cook D., Hofmann H. [Visualizing statistical models: Removing the blindfold](http://onlinelibrary.wiley.com/doi/10.1002/sam.11271/full) // Statistical Analysis and Data Mining: The ASA Data Science Journal. 2015. Т. 8, № 4. С. 203–225.

8. Grolemund G., Wickham H. Visualizing Complex Data with Embedded Plots // Journal of Computational and Graphical Statistics. 2015. Т. 24, № 1. С. 26–43.