

Анализ данных и визуализация результатов с помощью low-code платформы Loginom

Языков Кирилл Александрович

Приамурский государственный университет имени Шолом-Алейхема

Студент

Аннотация

В данной статье рассмотрен проведенный анализ данных по продажам различных товаров в разных городах за 2020–2022 годы, а также показана возможная визуализация полученных результатов. Весь анализ проведен и визуализирован с помощью low-code платформы Loginom. Данные же были собраны для удобства в таблицу MS Excel специалистами Finalytics.pro.

Ключевые слова: Loginom, анализ данных, визуализаторы,

Data analysis and visualization of results using the low-code Loginom platform

Yazykov Kirill Aleksandrovich

Sholom Aleichem Priamursky State University

Student

Abstract

This article examines the analysis of data on sales of various goods in different cities for the years 2020–2022, and also shows a possible visualization of the results obtained. The entire analysis was carried out and visualized using the low-code Loginom platform. The data was collected for convenience in an MS Excel spreadsheet by specialists Finalytics.pro.

Keywords: Loginom, data analysis, visualizers,

1 Введение

1.1 Актуальность

Анализ данных на сегодняшний день развивается достаточно быстро. С каждым годом появляется все больше способов проанализировать большие массивы данных, а также упростить весь процесс для рядового пользователя. Сейчас каждый может научиться, пройдя нужные курсы, но также нужно правильно выбрать платформу для проведения анализа. В данной статье будет рассмотрен анализ данных с помощью платформы Loginom, а также проведена визуализация полученных результатов с помощью встроенных возможностей low-code платформы.

1.2 Обзор исследований

В работе С. В. Рубакова. «Современные методы анализа данных» были подробно рассмотрены некоторые виды анализа данных, а именно

кластерный анализ, факторный анализ, нейронные сети, деревья решений и регрессионный анализ [1]. М. А. Ковалева и С. Б. Волошин в своем учебном пособии «Анализ данных» рассмотрели точечное оценивание параметров регрессионного и корреляционного анализа, а также связь между случайными величинами [2]. Сбор и подготовка данных для анализа с приведением примеров были рассмотрены в учебном пособии М. А. Поручикова – «Анализ данных» [3].

1.3 Цель исследования

Целью данной работы является проведение анализа данных по продажам различных товаров в разных городах за 2020–2022 годы и визуализация полученных результатов в формате графиков и таблиц с помощью low-code платформы Loginom.

2 Материалы и методы

Для реализации поставленных задач использовались данные, собранные в таблицу MS Excel специалистами Finalytics.pro. Весь анализ проводился с помощью low-code платформы Loginom.

3 Результаты исследования

Для проведения анализа были взяты данные, собранные командой Finalytics.pro по продажам различных товаров в нескольких городах за 2020–2022 годы. Данные собраны в таблицы MS Excel. Всего в таблице представлено 23330 строк и 9 столбцов данных (Рис. 1).

2	город	регион	контрагент	товар	дата	№ мес	продажи	себестоимость	прибыль
3	Абакан	Республика Хакасия	Новакс	Adilac	01.01.2020	01	284	133	151
4	Абакан	Республика Хакасия	Новакс	Adilac	01.02.2020	02	472	212	260
5	Абакан	Республика Хакасия	Новакс	Adilac	01.03.2020	03	484	198	286
6	Абакан	Республика Хакасия	Новакс	Adilac	01.04.2020	04	540	212	328
7	Абакан	Республика Хакасия	Новакс	Adilac	01.05.2020	05	426	180	246
8	Абакан	Республика Хакасия	Новакс	Adilac	01.06.2020	06	341	136	204
9	Абакан	Республика Хакасия	Новакс	Adilac	01.07.2020	07	318	132	186
10	Абакан	Республика Хакасия	Новакс	Adilac	01.08.2020	08	222	107	115
11	Абакан	Республика Хакасия	Новакс	Adilac	01.09.2020	09	309	140	169
12	Абакан	Республика Хакасия	Новакс	Adilac	01.10.2020	10	480	185	295
13	Абакан	Республика Хакасия	Новакс	Adilac	01.11.2020	11	437	169	267
14	Абакан	Республика Хакасия	Новакс	Adilac	01.12.2020	12	454	136	319
15	Абакан	Республика Хакасия	Новакс	Babidas	01.01.2020	01	192	144	47
16	Абакан	Республика Хакасия	Новакс	Babidas	01.02.2020	02	263	219	45
17	Абакан	Республика Хакасия	Новакс	Babidas	01.03.2020	03	286	270	16

Рисунок 1– Данные, собранные в таблицы

Весь анализ с визуализацией в итоге представляет собой небольшую схему, внутри которой скрыты данные и результаты (Рис. 2).

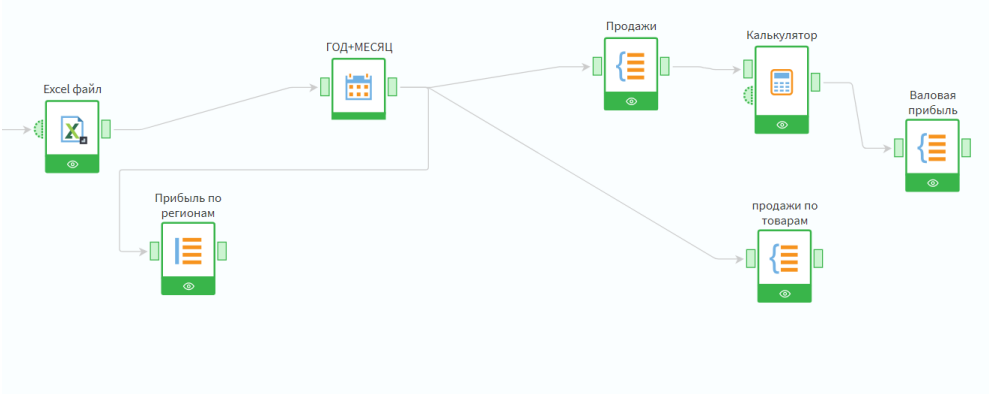


Рисунок 2 – Итог анализа данных на платформе Loginom

Первым этапом анализа было – загрузить данные в платформу Loginom. Для этого в программе имеется импорт Excel файлов. Указав путь к документу, платформа сама загружает его и ставит тип данных, которому те соответствуют (Рис. 3).

Настройка соответствия между столбцами

☒ Таблица ☐ Связи

Входные	Выходные	Имя	Вид данных	Назначение
ab город	ab город	gorod	Дискретный	Не задано
ab регион	ab регион	region	Дискретный	Не задано
ab контрагент	ab контрагент	kontragent	Дискретный	Не задано
ab товар	ab товар	tovar	Дискретный	Не задано
31 дата	31 дата	data	Непрерывный	Не задано
ab № мес	ab № мес	__mes	Дискретный	Не задано
9.0 продажи_	9.0 продажи_	prodazhi_	Непрерывный	Не задано
9.0 себестоимость_	9.0 себестоимость_	sebestoimost_	Непрерывный	Не задано
9.0 прибыль_	9.0 прибыль_	pribyl_	Непрерывный	Не задано
COL10	COL10	COL10	Неопределе...	Не задано
COL11	COL11	COL11	Неопределе...	Не задано
9.0 % новых покупателей	9.0 % новых покупателей	__novykh_pokupateley	Непрерывный	Не задано
9.0 COL13	9.0 COL13	COL13	Непрерывный	Не задано

Рисунок 3 – Загрузка данных в платформу Loginom

Часто результат различных событий наблюдают в динамике по месяцам и по годам, поэтому с помощью узла «Дата и время» есть возможность выделить нужные временные промежутки. В нашем случае требуется выделить «Год+месяц», «Год» и «Месяц» (Рис. 4).

Разбиение	31 Дата начала	31 Дата конца	12 Число	ab Строка	
Обычный					^
Год + Квартал	☐	☐		☐	
Год + Месяц	☐	☐		☒ %Y-M%m	
Год + Неделя	☐	☐		☐	
Год + День	☐			☐	
Год	☐	☐	☐	☒ %Y	
Квартал			☐	☐	
Месяц			☐	☒ %m %F	
Неделя			☐	☐	
День года			☐	☐	
День квартала			☐	☐	
День месяца			☐	☐	
День недели			☐	☐	
Часы			☐	☐	
Минуты			☐	☐	
Секунды			☐	☐	
Миллисекунды			☐	☐	
Дата	☐			☐	
Время	☐			☐	
Свой формат				☐	
ISO					^
Год + Квартал	☐	☐		☐	
Год + Месяц	☐	☐		☐	
Год + Неделя	☐	☐		☐	

Рисунок 4 – Выделение требуемых временных промежутков

Следующим этапом является группировка данных с помощью узла «Группировка». Данный узел помогает выбрать и сгруппировать нужные данные из всего массива для дальнейшего анализа (Рис. 5).

Кросс-таблица

Фильтрация

Доступные поля

31 дата

ab город

ab контрагент

ab товар

ab дата (Месяц)

ab № мес

9.0 себестоимость_

COL10

COL11

9.0 COL13

ab дата (Год + Месяц)

12 Количество

Выбранные поля

Колонки

Строки

ab регион

ab дата (Год)

Факты

9.0 прибыль_ (Сумма)

9.0 % новых покупателей (Первый, Последний)

9.0 продажи_ (Сумма)

Рисунок 5 – Группировка данных

Выделив нужные параметры, можно все визуализировать в таблицу. Платформа позволяет превратить данные в различные визуальные итоги – диаграммы, статистические измерения (медиана, среднее, мода и т.д.), таблицы и кубы. В данном случае полученную группировку данных возможно посмотреть с помощью визуализатора «Таблица» (Рис. 6).

#	ab регион	ab дата (Год)	98 прибыль_Сумма	98 продажи_Сумма
1	Иркутская область	2020	35483,60895	82702,166
2	Иркутская область	2021	32067,2527	74470,9884
3	Иркутская область	2022	23544,60815	56151,6987
4	Кемеровская область	2020	30050,46655	69360,7939
5	Кемеровская область	2021	26368,05815	61203,43685
6	Кемеровская область	2022	19916,42335	47446,5091
7	Новосибирская область	2020	46517,3948	109483,4291
8	Новосибирская область	2021	41827,9145	97825,94695
9	Новосибирская область	2022	32680,00915	75181,6897
10	Республика Тыва	2020	25127,1546	57252,9279
11	Республика Тыва	2021	21804,4792	50911,88935
12	Республика Тыва	2022	16246,62375	38268,26445
13	Республика Хакасия	2020	54019,2251	124624,6568
14	Республика Хакасия	2021	46812,90895	110483,9921
15	Республика Хакасия	2022	35557,126	84041,20115
16	Томская область	2020	22824,83215	52820,77265
17	Томская область	2021	20120,3495	47118,36395
18	Томская область	2022	15098,342	35970,30645

Рисунок 6 – Часть массива данных после Группировки

В данном случае видно, что таблица собралась по таким параметрам, как регион, год, прибыль, продажи. С помощью таблицы можно проверить правильность группировки данных, а также достаточность параметров.

С помощью визуализатора «Куб» можно получить сумму прибыли по регионам за 2020 – 2022 годы с подсчитанными итогами, а также столбчатые диаграммы для более легкого зрительного восприятия (Рис. 7, Рис. 8).

+					
		дата (Год)	+	Σ Факты	
регион		2020	2021	2022	Итого:
Иркутская ...		35 483,61	32 067,25	23 544,61	91 095,47
Кемеровск...		30 050,47	26 368,06	19 916,42	76 334,95
Новосиби...		46 517,39	41 827,91	32 680,01	121 025,32
Республ...		25 127,15	21 804,48	16 246,62	63 178,26
Республ...		54 019,23	46 812,91	35 557,13	136 389,26
Томская о...		22 824,83	20 120,35	15 098,34	58 043,52
Итого:		214 022,68	189 000,96	143 043,13	546 066,78

Рисунок 7 – Прибыль по регионам за 2020 – 2022 годы

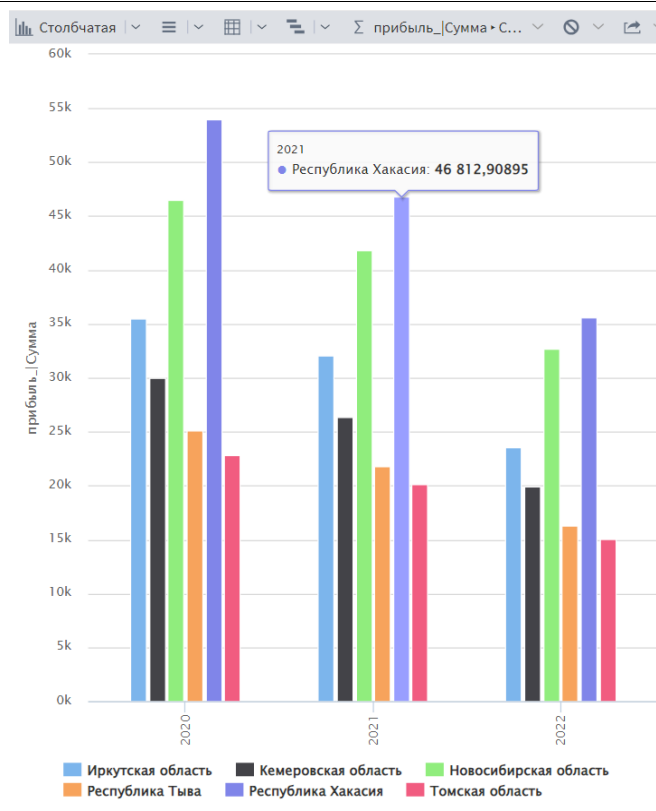


Рисунок 8 – Гистограмма прибыли по регионам за 2020 – 2022 годы

Собрав следующую группировку уже с немного другими параметрами, получим иной результат. В данном случае были собраны следующие параметры: город, регион, год, месяц, контрагент, продажи и себестоимость (Рис. 9).

Группировка

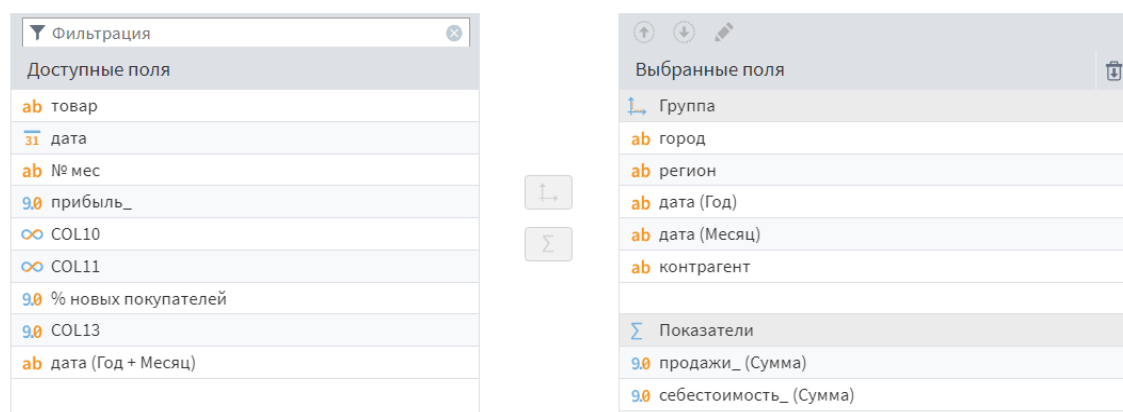


Рисунок 9 – Группировка данных

С помощью уже известного визуализатора «Куб» выведем продажи по городам за 2020 – 2022 годы, а также гистограмму для визуального восприятия. Также можно отфильтровать города отдельно по регионам для более удобного изучения (Рис. 10).

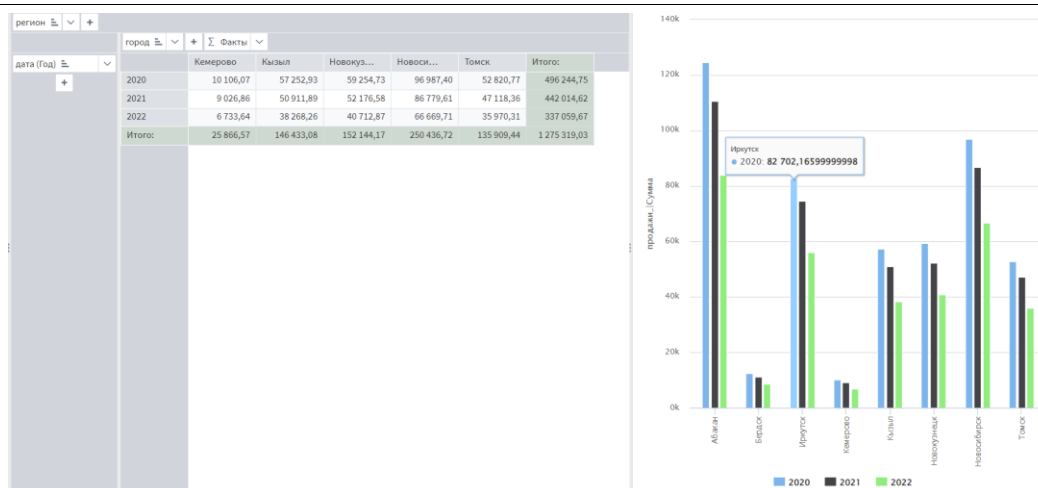


Рисунок 10 – Продажи по городам за 2020 – 2022 г.

Также с помощью узла «Группировка» и визуализатора «Куб» можно получить общие итоги по продажам за три года в период с 2020 – 2022 года.

С помощью узла калькулятор введем новый параметр «Валовая прибыль», вычтя из продаж себестоимость товаров (Рис. 11).

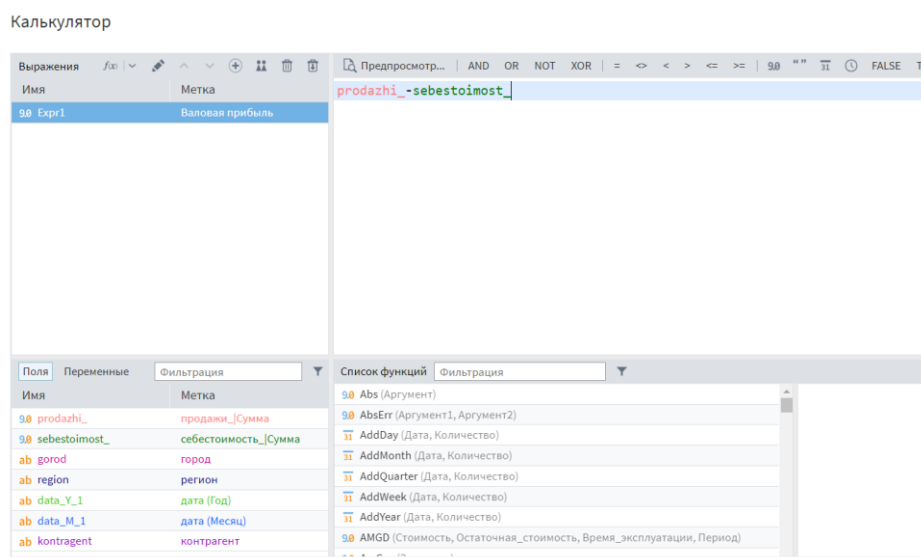


Рисунок 11 – Валовая прибыль

После калькулятора новый параметр сразу попадает общую таблицу с данными, поэтому дальше можем сгруппировать данный параметр с еще несколькими и вывести итог через «Куб» и гистограмму (Рис. 12).

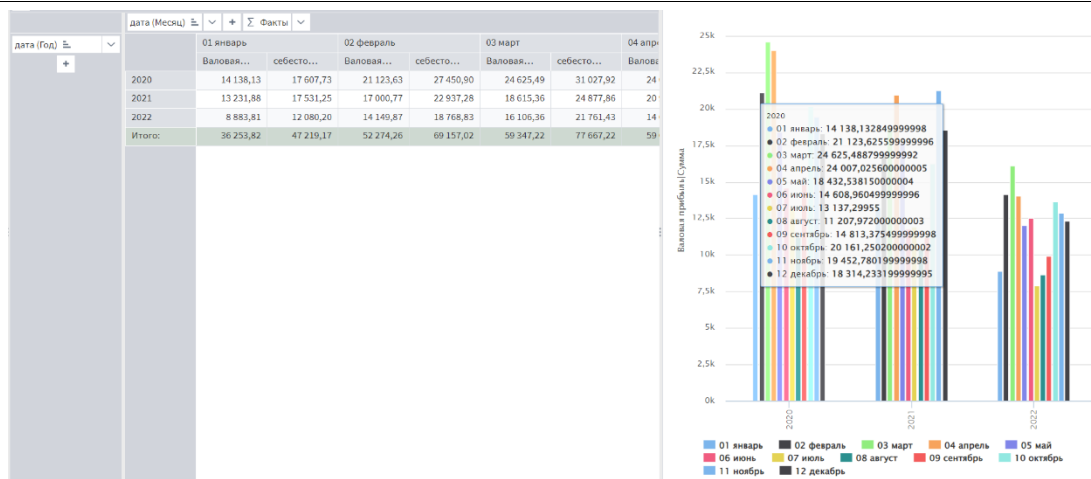


Рисунок 12 – Валовая прибыль и себестоимость по месяцам

Также платформа Logiplot позволяет отдельно составлять различные диаграммы для изучения и анализа.

Выбрав вид диаграммы и данные для нее, можно получить график динамики валовой прибыли за 2020 – 2022 годы (Рис. 13).

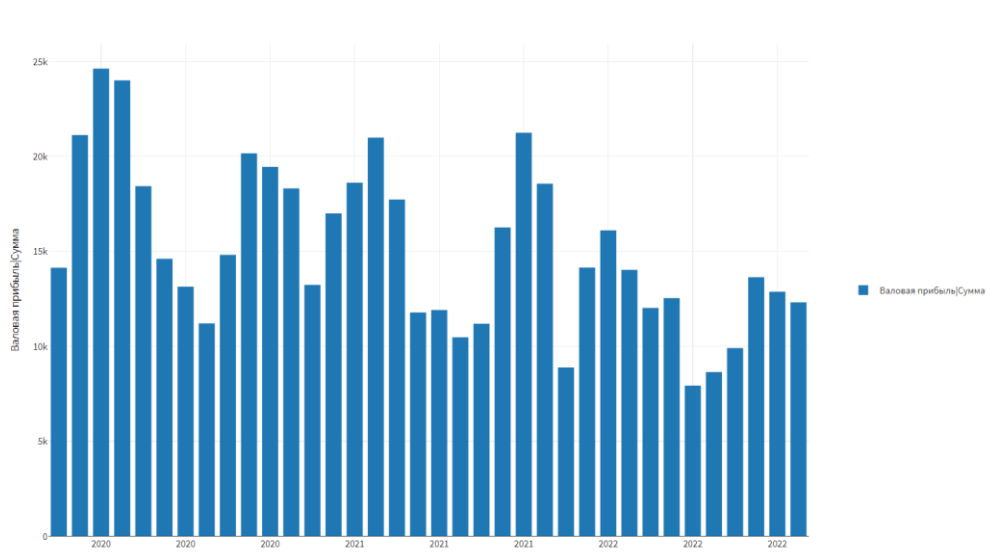


Рисунок 13 – Динамика Валовой прибыли 2020 – 2022 год

С помощью тех же манипуляций с группировкой и «Кубом» получим итоги по продажам определенных товаров за 2020 – 2022 годы, а также гистограмму (Рис. 14).

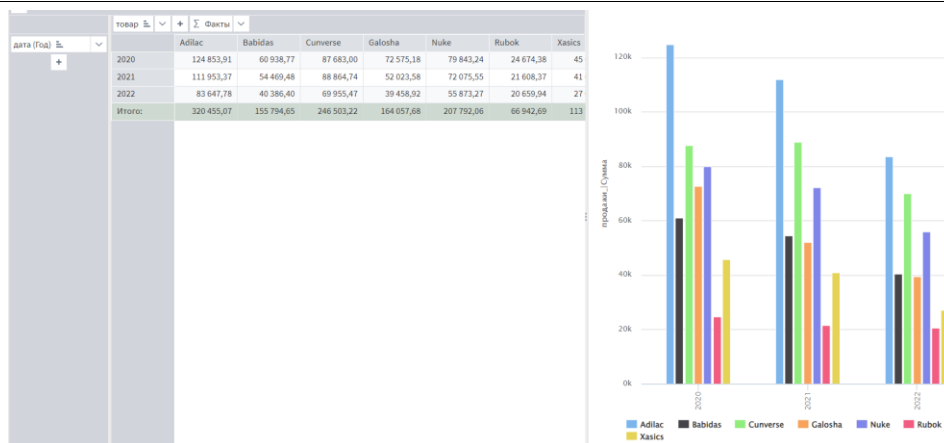


Рисунок 14 – Продажи товаров 2020 – 2022 г

4 Выводы

Таким образом даже не имея особенных знаний в сфере анализа данных с помощью платформы Loginot становится возможным провести несложный анализ с большим массивом данных и показать это в доступном визуальном виде. Интерфейс максимально разнообразен и удобен простому пользователю, что делает процесс анализа еще понятнее. Загрузив правильно данные, не составит труда провести и более сложные анализы вроде регрессионного или дерева данных.

Библиографический список

1. Рубаков С. В. Современные методы анализа данных // Управление наукой и наукометрия. 2008. №. 7. С. 165-176.
2. Ковалева М. А., Волошин С. Б. Анализ данных. Учебное пособие. М.: Мир науки, 2019. URL: <https://izd-mn.com/PDF/32MNNPU19.pdf>
3. Поручиков М. А. и др. Анализ данных. Самара: Изд-во Самар. ун-та., 2016.
4. Орлов Г. М. и др. Современные методы обработки и анализа данных. СПб.: Санкт-Петербургский национальный исследовательский университет информационных технологий, механики и оптики, 2021. 147 с.