

# Data analysis in the educational activities of the Dubna University



Author: Kristina Zhatkina

Co-authors: Oksana Streltsova, Mikhail Belov

The aim of the work was to study and analyze data in the educational activities of the Dubna University for the subsequent identification of data patterns and the construction of new hypotheses for the organization of the educational process. To study the existing data at the Dubna University and possible algorithms for their processing.

## Types of data at the Dubna University

- Admission details:
  - Applicant scores
  - Personal data students
  - Regions in which applicants studied, etc.
- Student performance data (session results during training)
- Passing scores (dynamics of passing scores by directions)
- The ratio of the number of students per teacher (the number of students by year and the number of employed by year)
- Availability of graduates with honors (number of graduates who received a diploma with honors)
- Number of students who received a degree (those who defended their post-graduate degree)
- Percentage of international students
- Data on the rating of students (accounting for achievements and academic performance)
- Data from video cameras about the behavior of students in the classroom
- Online class recordings

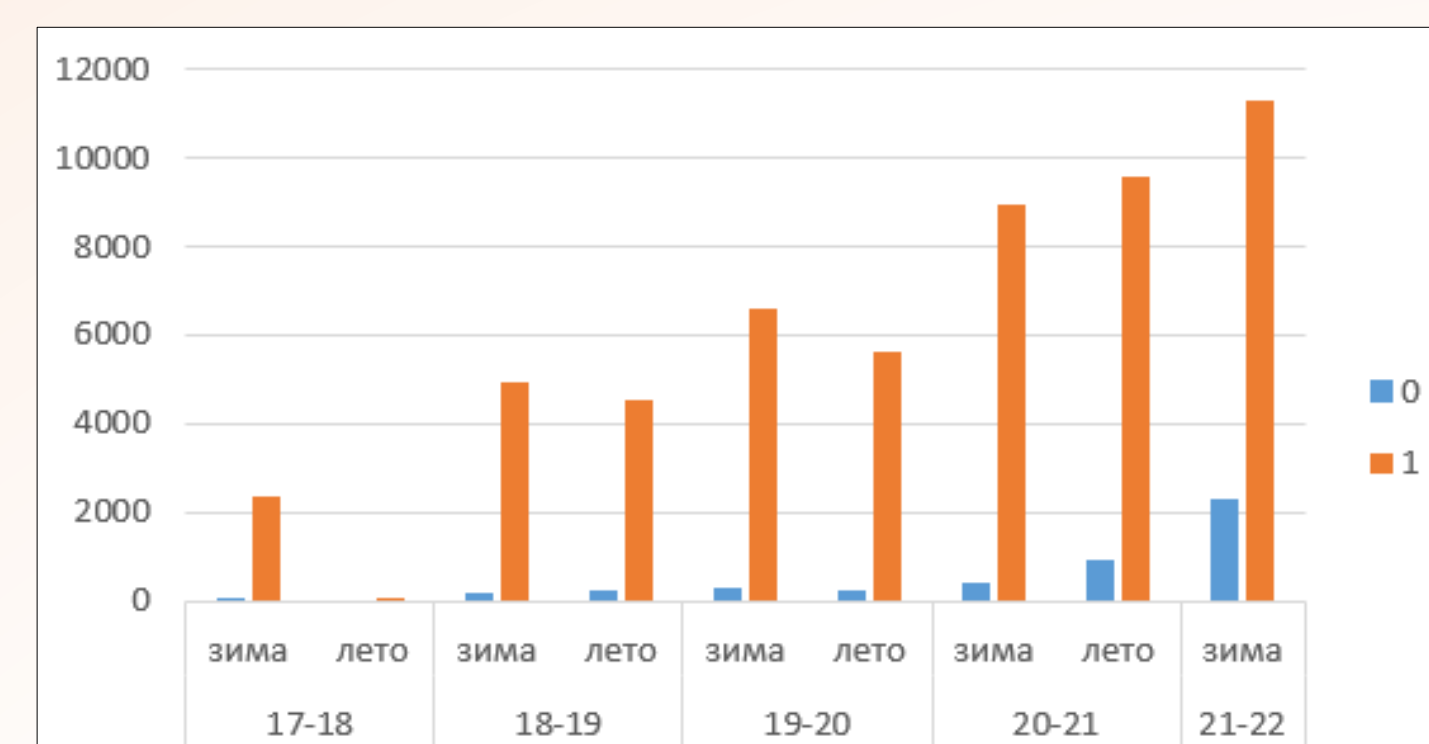
## Student Performance Data Structure

#	Column	Non-Null	Count	Dtype
0	Группа	58469	non-null	int64
1	Курс	58469	non-null	int64
2	Физическое лицо	58469	non-null	object
3	Дисциплина	58469	non-null	object
4	Вид контроля	58469	non-null	object
5	Отметка	58065	non-null	object
6	Период контроля	58469	non-null	object
7	Преподаватель	58469	non-null	object
8	Сдано	58469	non-null	int64
9	Дата занятия	58469	non-null	object
10	Год	58469	non-null	object
11	Сессия	58469	non-null	object

Группа	130
Курс	6
Физическое лицо	1346
Дисциплина	357
Вид контроля	6
Отметка	10
Период контроля	8
Преподаватель	491
Сдано	2
Дата занятия	617
Год	5
Сессия	2

	Группа	Курс	Сдано
count	58469.000000	58469.000000	58469.000000
mean	3840.004943	3.714242	0.922318
std	1379.907905	1.419076	0.267673
min	1011.000000	1.000000	0.000000
25%	3014.000000	3.000000	1.000000
50%	4031.000000	4.000000	1.000000
75%	5013.000000	5.000000	1.000000
max	6181.000000	6.000000	1.000000

After primary processing (~ 70 thousand lines) and analysis of data from 1C, a positive trend in the progress of students of the Institute of ACS of the Dubna University was established during distance learning. However, it should be noted that the data is not evenly distributed over the years, this is primarily due to the digitalization of education, which began to gain momentum only a few years ago.



## Acquisition, processing and analysis of data

```
df = pd.read_csv('/content/drive/MyDrive/data_academic.csv')
```

Группа	Курс	Физическое лицо	Дисциплина	Вид контроля	Отметка	Период контроля	Преподаватель	Сдано	Дата занятия	Год	Сессия
0	1011	1	Безопасность жизнедеятельности	Зачет	Зачтено	Первый семестр	Томашова Елена Юрьевна (ин. язык)	1	25.12.2021	21-22	зима
1	1011	1	Введение в программирование	Экзамен	Отлично	Первый семестр	Сиднев Петр Павлович	1	25.12.2021	21-22	зима
2	1011	1	Волейбол	Зачет	Зачтено	Первый семестр	Белова Елена Валерьевна	1	25.12.2021	21-22	зима
3	1011	1	Волейбол	Зачет	Зачтено	Первый семестр	Савдолова Татьяна Валерьевна	1	25.12.2021	21-22	зима
4	1011	1	Иностранный язык	Зачет	Зачтено	Первый семестр	Харитонюк Татьяна Александровна	1	25.12.2021	21-22	зима
...	...	...	...	...	...	...	...	...	...	...	...
58464	6181	6	Основы менеджмента и маркетинга	Зачет	Зачтено	Третий семестр	Николаев Николай Валерьевич	1	25.12.2017	17-18	зима
58465	6181	6	Социология	Зачет	Зачтено	Третий семестр	Шамон Иван Ярославич	1	25.12.2017	17-18	зима
58466	6181	6	ТЕОРИЯ ВЕРОЯТНОСТЕЙ И МАТЕМАТИЧЕСКАЯ СТАТИСТИКА	Курсовая работа	Удовлетворительно	Третий семестр	Тарасов Владимир Александрович (ин. язык)	1	25.12.2017	17-18	зима
58467	6181	6	ТЕОРИЯ ВЕРОЯТНОСТЕЙ И МАТЕМАТИЧЕСКАЯ СТАТИСТИКА	Зачет с оценкой	Удовлетворительно	Третий семестр	Белова Елена Валерьевна	1	25.12.2017	17-18	зима

The increase in the number of debts in various disciplines is primarily associated with an increase in the total amount of data entered into the system for recording the progress of students at the Dubna University 1C:University.

The result of the work is a prepared set of data on the progress of students of the Institute of ACS of the Dubna University. Machine learning algorithms were selected for subsequent testing on data.