

Учреждение образования
«БЕЛОРУССКИЙ ГОСУДАРСТВЕННЫЙ УНИВЕРСИТЕТ
ИНФОРМАТИКИ И РАДИОЭЛЕКТРОНИКИ»

**Отчет по лабораторной работе №1
по курсу «СтатОИВ»
на тему:
«Введение в Python»**

Выполнил студент группы
121703:

Якимович И.В.

Проверила:

Русина Н.В.

МИНСК, 2023

Цели: Приобрести навыки работы с наборами данных с помощью языка Python. Освоить среду разработки Jupyter.

Задачи: Установить на устройство среду разработки. Подготовить набор данных к дальнейшей обработке

Задание: Выбрать с сайта [kaggle.com](https://www.kaggle.com) набор данных в формате .csv, загрузить и подготовить его к дальнейшей обработке с помощью описанных ранее методов. Наборы данных не должны повторяться внутри группы. Задание индивидуальное. Требования:

- Значения подразумевающие числа должны иметь числовой тип
- Строковые значения должны заменены на числовые значения и иметь числовой тип
- Индексы строк должны идти последовательно
- Незначимые данные должны быть удалены
- Указать какие знания можно получить из набора
- Сохранить обработанный набор
- Сохранить IPython Notebook

Ход выполнения:

1. Подключаем нужные модули

```
import numpy as np
import pandas as pd
```

2. Считываем с файла, выбранного на [kaggle.com](https://www.kaggle.com), и присваиваем переменной dataset, выводим

```
dataset = pd.read_csv("chennai_house_price.csv")
✓ 0.7s
```

dataset

✓ 0.8s

	price	area	status	bhk	bathroom	age	location	builder
0	37.49	872	Ready to move	2	NaN	1.0	Sembakkam	MP Developers
1	93.54	1346	Under Construction	3	2.0	NaN	Selaiyur	DAC Promoters
2	151.00	2225	Under Construction	3	NaN	0.0	Mogappair	Casagrand Builder Private Limited
3	49.00	1028	Ready to move	2	2.0	3.0	Ambattur	Dugar Housing Builders
4	42.28	588	Under Construction	2	1.0	0.0	Pallavaram	Radiance Realty Developers India Ltd
...
2615	27.50	750	Ready to move	2	2.0	3.0	Veppampattu	Velan Housing Properties
2616	27.50	750	Ready to move	2	2.0	3.0	Veppampattu	Velan Housing Properties
2617	30.00	800	Ready to move	2	2.0	3.0	Veppampattu	Velan Housing Properties
2618	31.50	850	Ready to move	2	2.0	3.0	Veppampattu	Velan Housing Properties
2619	34.00	900	Ready to move	2	2.0	3.0	Veppampattu	Velan Housing Properties

2620 rows x 8 columns

3. Создаем копию с важными столбцами

```
important_columns = ["price", "area", "status", "bhk", "bathroom", "age"]
dataset = dataset[important_columns].copy()
dataset
```

✓ 0.8s

	price	area	status	bhk	bathroom	age
0	37.49	872	Ready to move	2	NaN	1.0
1	93.54	1346	Under Construction	3	2.0	NaN
2	151.00	2225	Under Construction	3	NaN	0.0
3	49.00	1028	Ready to move	2	2.0	3.0
4	42.28	588	Under Construction	2	1.0	0.0
...
2615	27.50	750	Ready to move	2	2.0	3.0
2616	27.50	750	Ready to move	2	2.0	3.0
2617	30.00	800	Ready to move	2	2.0	3.0
2618	31.50	850	Ready to move	2	2.0	3.0
2619	34.00	900	Ready to move	2	2.0	3.0

2620 rows × 6 columns

4. Подготавливаем данные к дальнейшей обработке, совершаем такие изменения:

- Делаем статусы целочисленными
- Удаляем строки со значением Nan
- Пронумеруем id заново
- Изменим название столбца

```
dataset["status"] = dataset["status"].replace({"Ready to move": 1, "Under Construction": 0}) # Делаем статусы целочисленными
dataset.dropna(inplace=True) # Удалили строки со значением Nan
dataset.reset_index(drop=True, inplace=True) # Пронумеровал id заново
dataset.rename(columns={"bhk": "bedrooms"}, inplace=True) # Изменить название столбца
dataset
```

✓ 0.9s

	price	area	status	bedrooms	bathroom	age
0	49.00	1028	1	2	2.0	3.0
1	42.28	588	0	2	1.0	0.0
2	188.00	2221	0	3	3.0	0.0
3	240.00	1500	1	3	3.0	1.0
4	113.00	2343	1	4	4.0	5.0
...
575	27.50	750	1	2	2.0	3.0
576	27.50	750	1	2	2.0	3.0
577	30.00	800	1	2	2.0	3.0
578	31.50	850	1	2	2.0	3.0
579	34.00	900	1	2	2.0	3.0

580 rows × 6 columns

5. Сохраняем обработанный набор

```
dataset.to_csv("filtered_chennai_house_price.csv")
```

✓ 0.2s