



# РАНХиГС

РОССИЙСКАЯ АКАДЕМИЯ НАРОДНОГО ХОЗЯЙСТВА  
И ГОСУДАРСТВЕННОЙ СЛУЖБЫ  
ПРИ ПРЕЗИДЕНТЕ РОССИЙСКОЙ ФЕДЕРАЦИИ



## Программирование на Python

Библиотека pandas - обзор

22.06.2023

Библиотека pandas базируется на библиотеке numpy и имеет огромное количество схожих методов и почти всегда используются вместе.

Обе библиотеки используются для анализа данных.

- Ссылка на документацию pandas: <https://pandas.pydata.org/docs/index.html>



- Ссылка на документацию numpy: <https://numpy.org/doc/stable/index.html>



```
s = pd.Series(np.random.randn(5), index = ['a', 'b', 'c', 'd', 'e'])
```

Индекс	a	b	c	d	e
Серия					

```
print(s)
```

```
a -0.541720  
b -0.547071  
c -0.942831  
d  1.952850  
e -0.400304  
dtype: float64
```

```
print (s['e'])
```

```
-0.4003044524065795
```

```
df = pd.DataFrame(np.random.randn(5,3),
index = pd.date_range ('1/1/2022', periods = 5),
columns = ['A', 'B', 'C'])
```

```
print (df)
```

	A	B	C
2022-01-01	0.387881	-1.531001	-0.637188
2022-01-02	-1.482243	0.767467	-1.541563
2022-01-03	-0.638061	-0.506002	-0.527962
2022-01-04	-0.651714	-0.381246	0.603340
2022-01-05	0.168934	-1.819145	-0.671056

```
print (df.columns) – Узнать наименования колонок
```

	Колонки		
	A	B	C
Индекс	Серия	Серия	Серия

```
import pandas as pd  
print(pd.__version__) # Узнать версию pandas  
xl = pd.read_excel("data.xlsx", sheet_name = 'Sheet1')  
# Открыть файл Excel  
print (df) # Отобразить полученный датафрейм  
df.describe(include=[object]) # Вывести описание  
датафрейма
```

### Серия

```
s = pd.Series([1, 2, 3, 4, 5])  
s.replace(1, 5)
```

```
0    5
```

```
1    2
```

```
2    3
```

```
3    4
```

```
4    5
```

```
dtype: int64
```

### Датафрейм

```
df = pd.DataFrame({'A': [0, 1, 2, 3, 4],  
                   'B': [5, 6, 7, 8, 9],  
                   'C': ['a', 'b', 'c', 'd', 'e']})
```

```
df.replace(0, 5)
```

```
   A  B  C
```

```
0  5  5  a
```

```
1  1  6  b
```

```
2  2  7  c
```

```
3  3  8  d
```

```
4  4  9  e
```



Операторы аналогичны Python

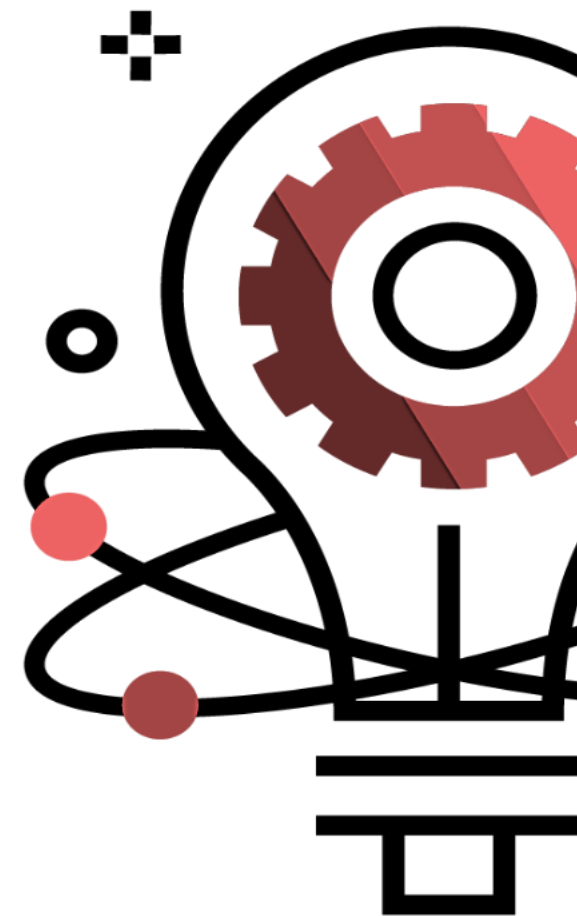
`df[df['columnname'] == 'value']`

```
df.to_excel('test.xlsx')
```

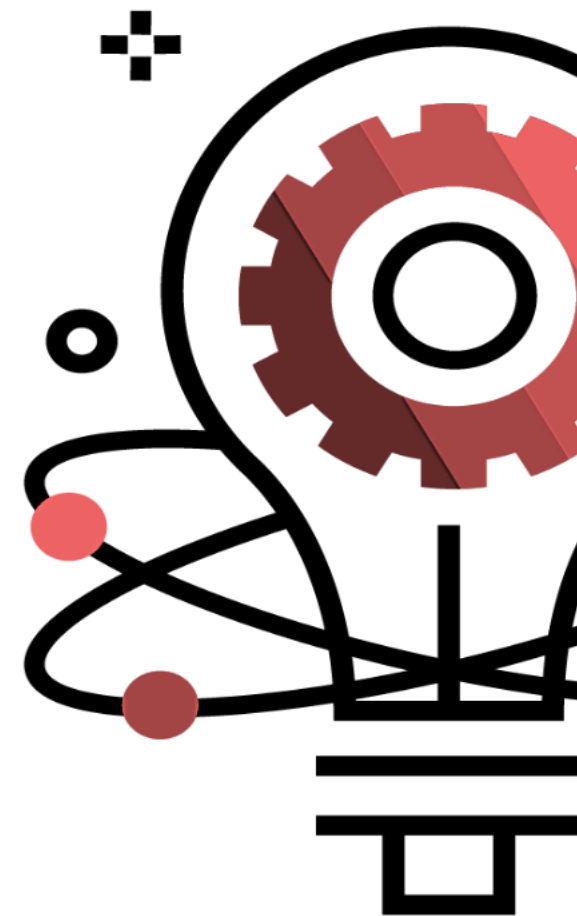
```
df.to_csv('test.csv')
```



- Скачайте файл data.xlsx
- Откройте его с помощью библиотеки pandas
- Выберите все значения, где sku = капуста
- Запишите датафрейм в файл task1 в формате csv



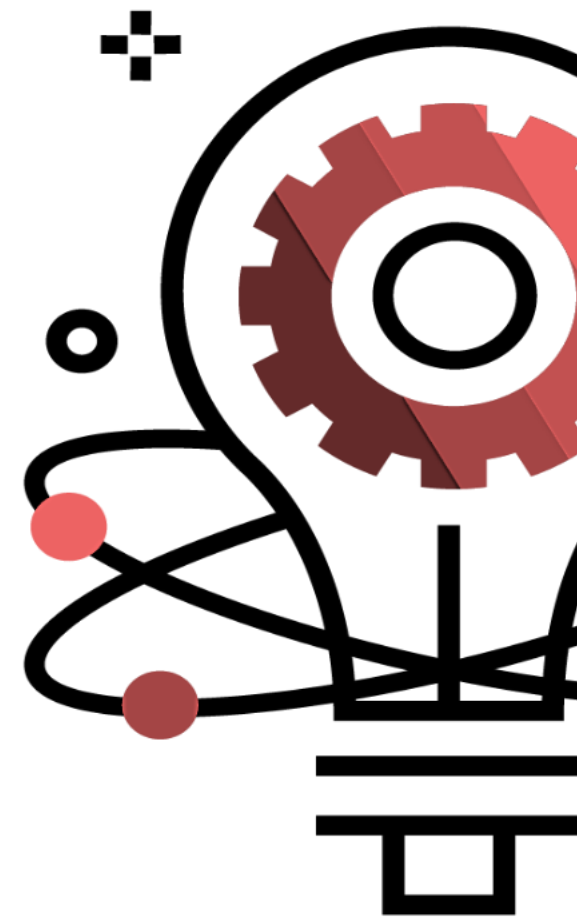
- Скачайте файл data.xlsx
- Выберите из всего датафрейма (data.xlsx) sku со значением Киви и ценой (priceoforder) больше 1000 рублей
- Запишите датафрейм в файл task2 в формате.xlsx



## Задача 3

---

- Скачайте файл data.xlsx
- Выбрать все заказы где **volume** > **15** и вывести на экран полученный датафрейм
- Выбрать из полученного датафрейма заказы где **volume** < **50** и записать данный датафрейм в excel файл под названием task3
- Запишите датафрейм в файл task3 в формате xlsx



С помощью библиотеки `pandas` и библиотеки `sqlite3` мы можем выгружать и загружать датафреймы в/из базы. Для того, чтобы произвести экспорт необходимо:

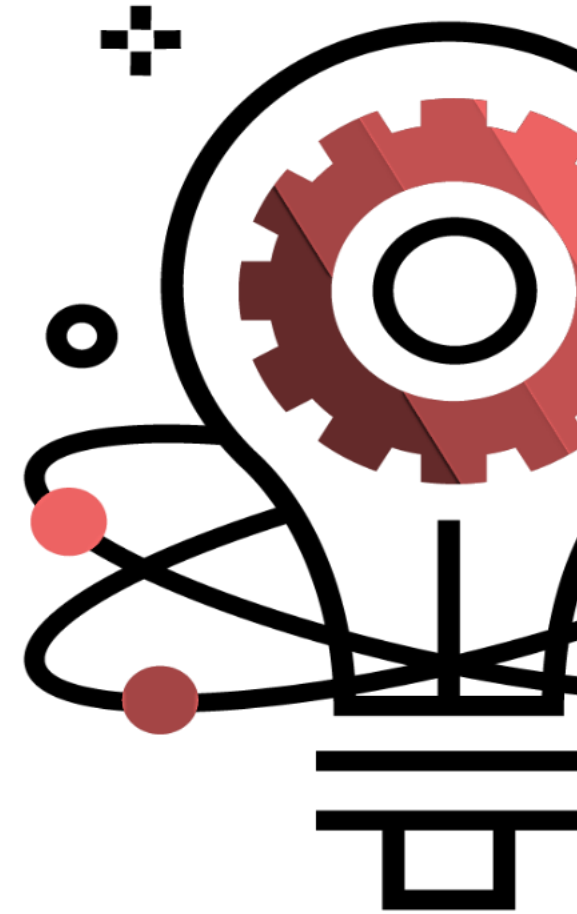
- 1) Импортировать названные библиотеки.
- 2) Создать соединение с БД.
- 3) Воспользоваться методом `read_sql` и считать необходимые данные из БД.
- 4) Полученный датафрейм можно записать в тот формат, который Вам необходим.

Также с ним можно производить все операции, которые производятся с обычным датафреймом.

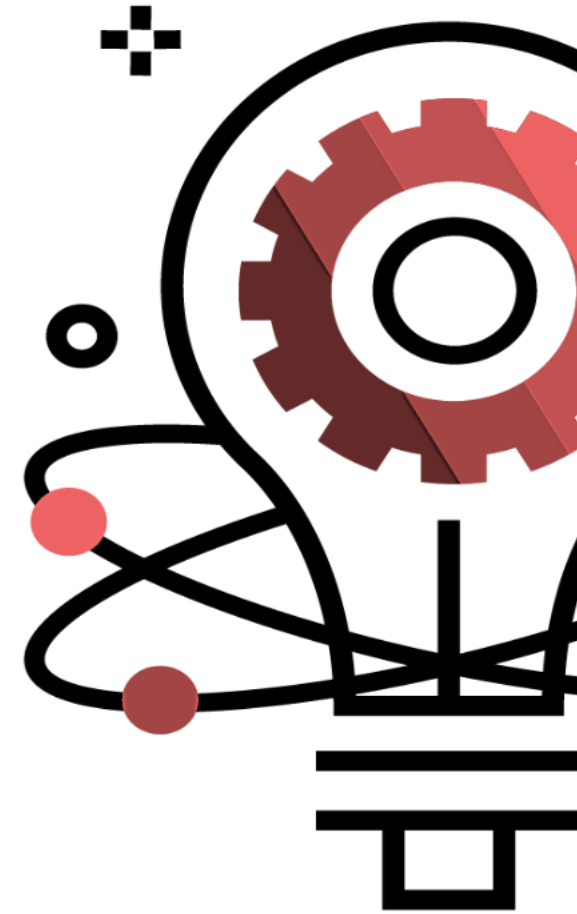
- 1) Импортировать названные библиотеки.
- 2) Создать соединение с БД.
- 3) Создать необходимый для загрузки датафрейм
- 4) Воспользоваться методом **датафрейм.to\_sql("tablename", con=connection, if\_exists="append", index=False)**
- 5) Выгрузить обновленную таблицу в датафрейм

- 1) Импортировать названные библиотеки.
- 2) Создать соединение с БД.
- 3) Создать необходимый для создания датафрейм
- 4) Воспользоваться методом **датафрейм.to\_sql("название таблицы", connection)**
- 5) Вывести полученные результаты на экран с помощью SQL запроса или датафрейма.

- Скачайте файл Students.db
- Создайте соединение с БД
- Выгрузите содержимое таблицы Students в датафрейм
- Полученный датафрейм запишите в формат.xlsx под названием Task4.xlsx



- Скачайте файл Students.db
- Создайте соединение с БД
- Добавьте к таблице Students датафрейм, содержащий 3-х новых студентов
- Выгрузите полученный датафрейм запишите в формат xlsx под названием Task5.xlsx

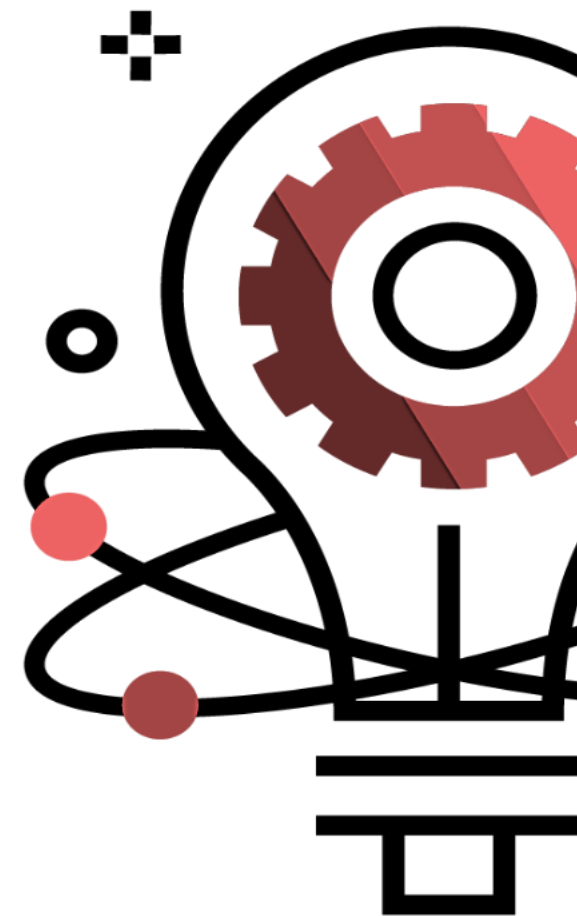




## Задача 6

---

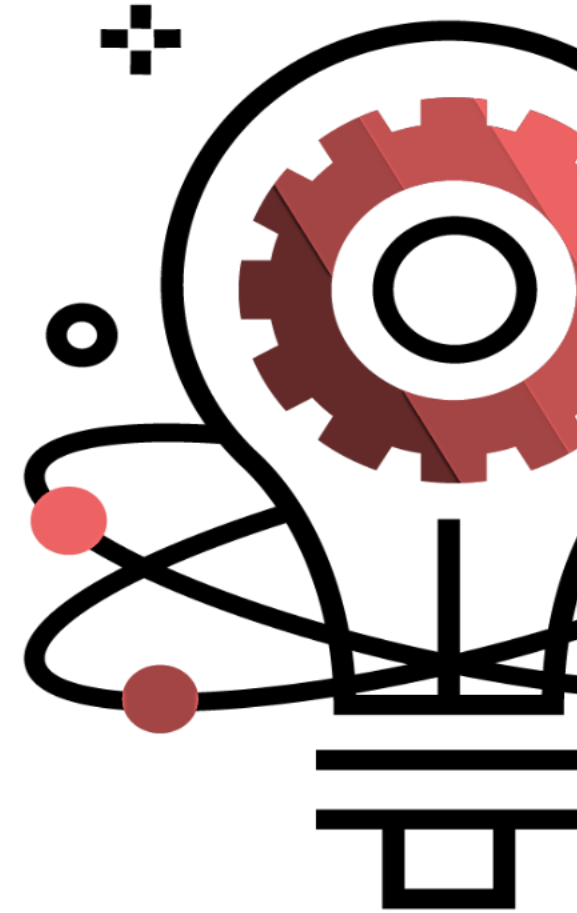
- Скачайте файл Students.db
- Создайте соединение с БД
- Создайте новую таблицу Cars, в таблице Cars должно быть 2 колонки: Car, Student\_Id (Student\_Id должно соответствовать ID добавленным вами ранее студентам) – добавьте вашим студентам любые авто
- Выведите и запишите датафрейм с Student\_Name, Car в формат xlsx под названием Task6.xlsx



## Задача 7

---

- Скачайте БД Schools.db
- Создайте подключение к БД
- Добавьте 3 новые школы к существующим. Названия: Ромашка, Василек, Победитель.
- Выгрузите общий перечень школ в формат `xlsx` под название `task7.xlsx`



## Задача 8

---

- Скачайте БД Schools.db, Students.db
- Создайте подключения к БД
- Выгрузите таблицу Students из БД Students.db и загрузите ее в БД Schools.db
- Из БД Schools.db сделайте выборку Student\_name, School\_Id, School\_name. Студент должен соответствовать школе. Запишите полученный результат в файл task8.xlsx

