МОСКОВСКИЙ ГОСУДАРСТВЕННЫЙ ТЕХНИЧЕСКИЙ УНИВЕРСИТЕТ

им. Н.Э. Баумана

Кафедра «Систем обработки информации и управления»

ОТЧЕТ

**Лабораторная работа №2**

по курсу Методы машинного обучения

«Изучение библиотек обработки данных Pandas и PandaSQL»

10

(количество листов)

|  |  |
| --- | --- |
| ИСПОЛНИТЕЛЬ: |  |
| студент группы ИУ5-24М | \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_ |
| Савельева М. А. | "\_\_"\_\_\_\_\_\_\_\_\_2019 г. |

|  |  |
| --- | --- |
| ПРЕПОДАВАТЕЛЬ: |  |
| Гапанюк Ю. Е.  к.т.н., доцент | \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_ |
|  | "\_\_"\_\_\_\_\_\_\_\_\_2019 г. |

Москва – 2019

**Задание**

* **Часть 1**

Выполните первое демонстрационное задание "demo assignment" под названием "Exploratory data analysis with Pandas" со страницы курса <https://mlcourse.ai/assignments>

Условие задания - <https://nbviewer.jupyter.org/github/Yorko/mlcourse_open/blob/master/jupyter_english/assignments_demo/assignment01_pandas_uci_adult.ipynb?flush_cache=true>

Набор данных можно скачать здесь - <https://archive.ics.uci.edu/ml/datasets/Adult>

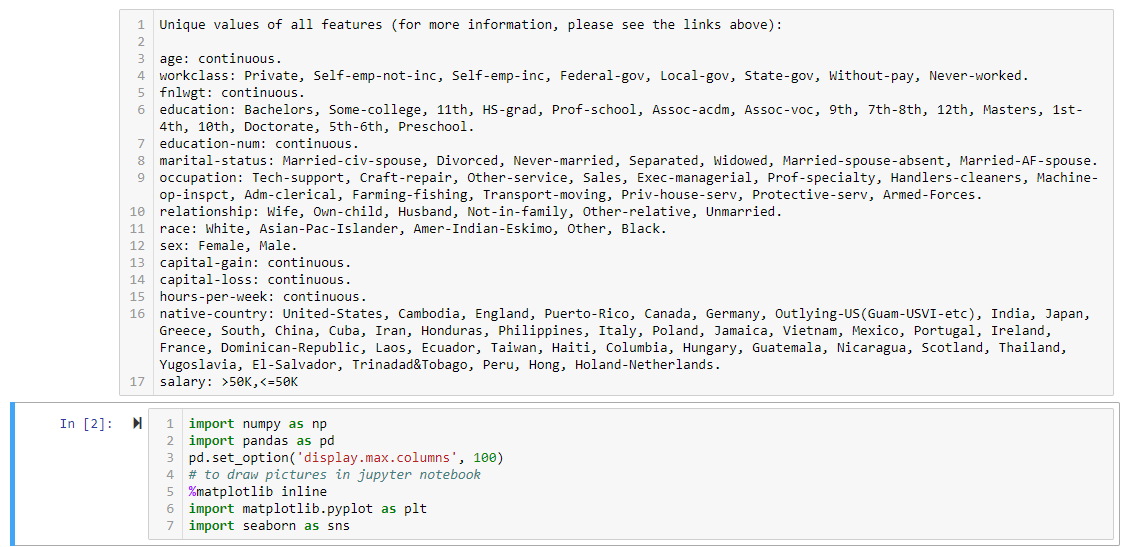
* **Часть 2**

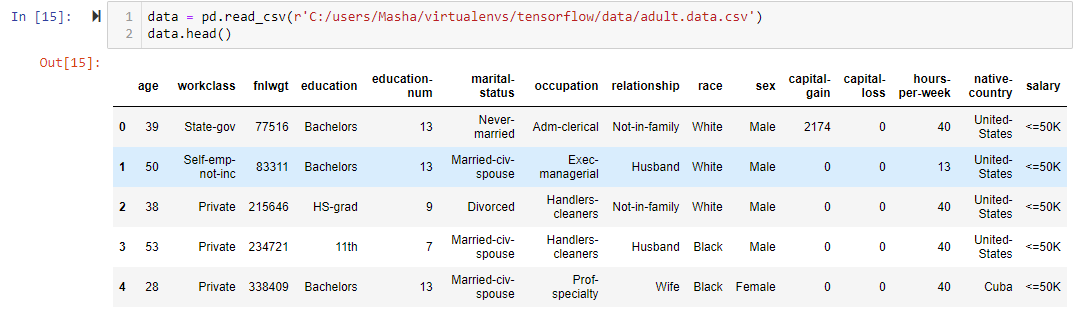
Выполните следующие запросы с использованием двух различных библиотек - [Pandas](https://pandas.pydata.org/) и [PandaSQL](https://github.com/yhat/pandasql):

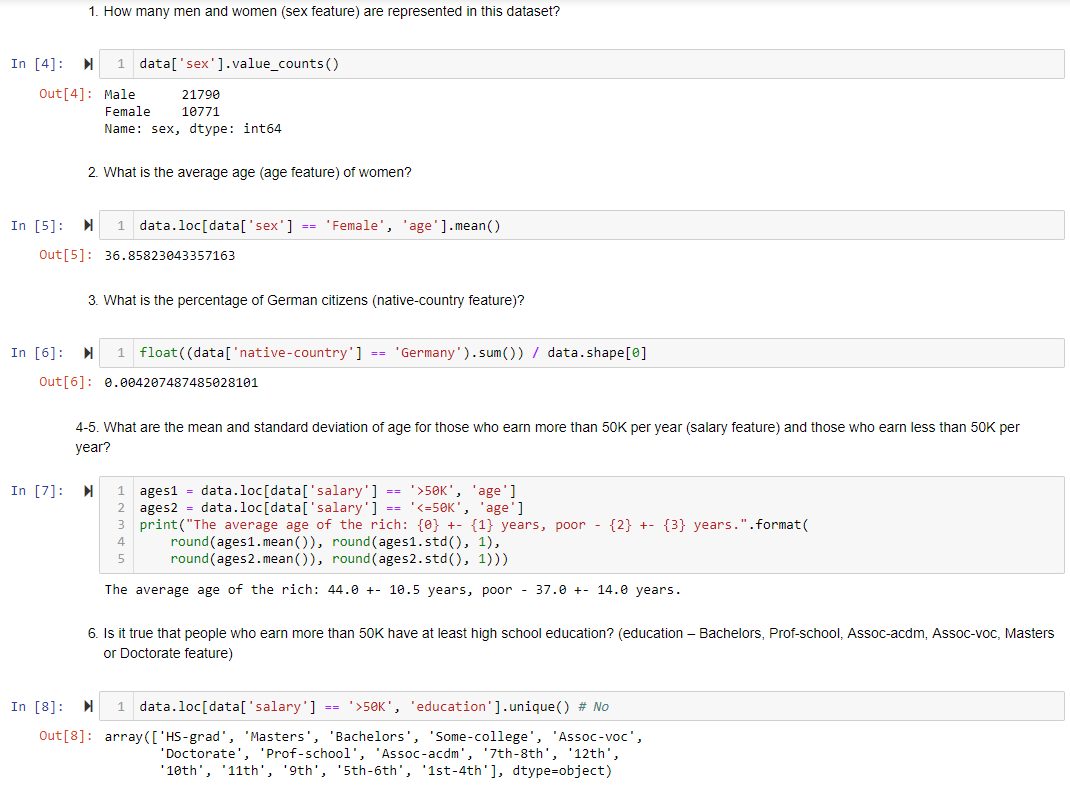
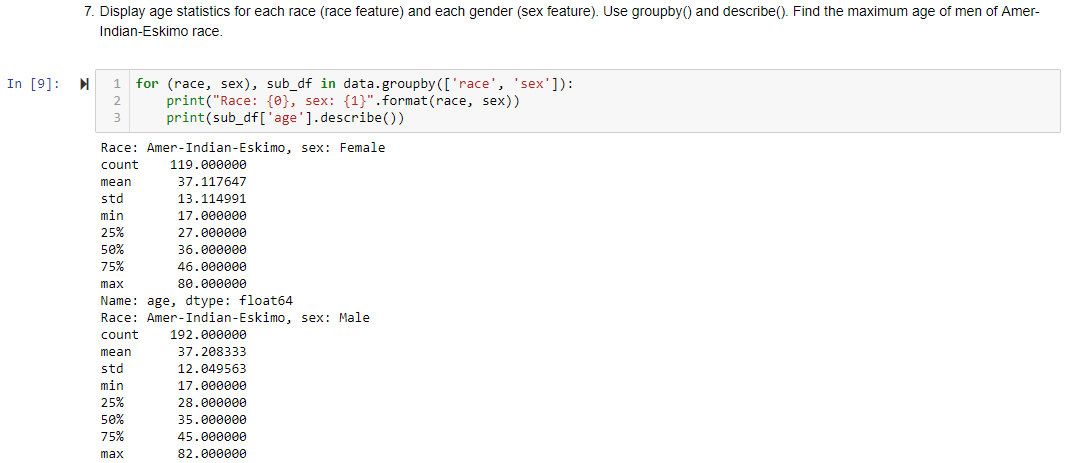
* один произвольный запрос на соединение двух наборов данных;
* один произвольный запрос на группировку набора данных с использованием функций агрегирования.

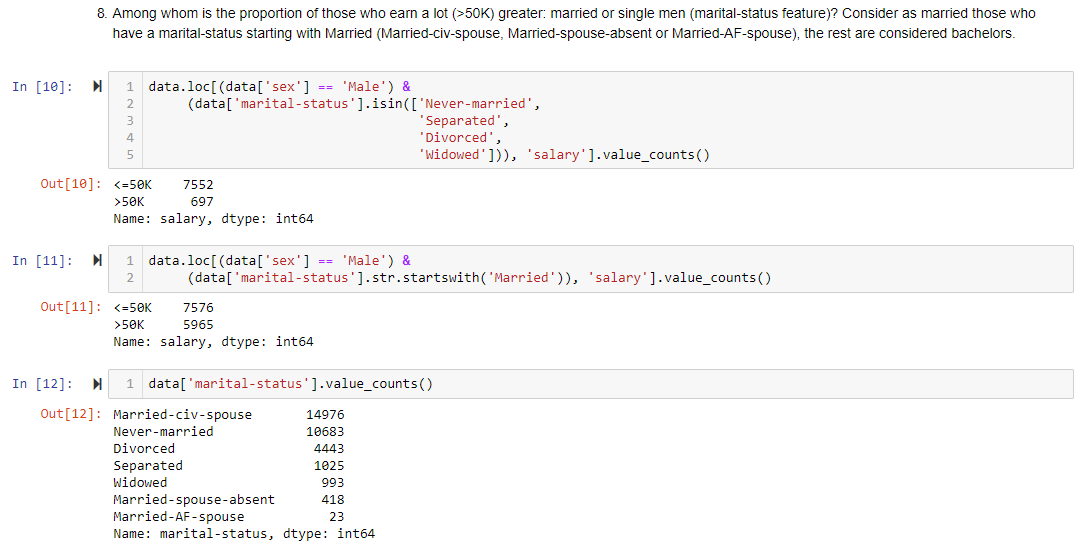
Сравните время выполнения каждого запроса в Pandas и PandaSQL.

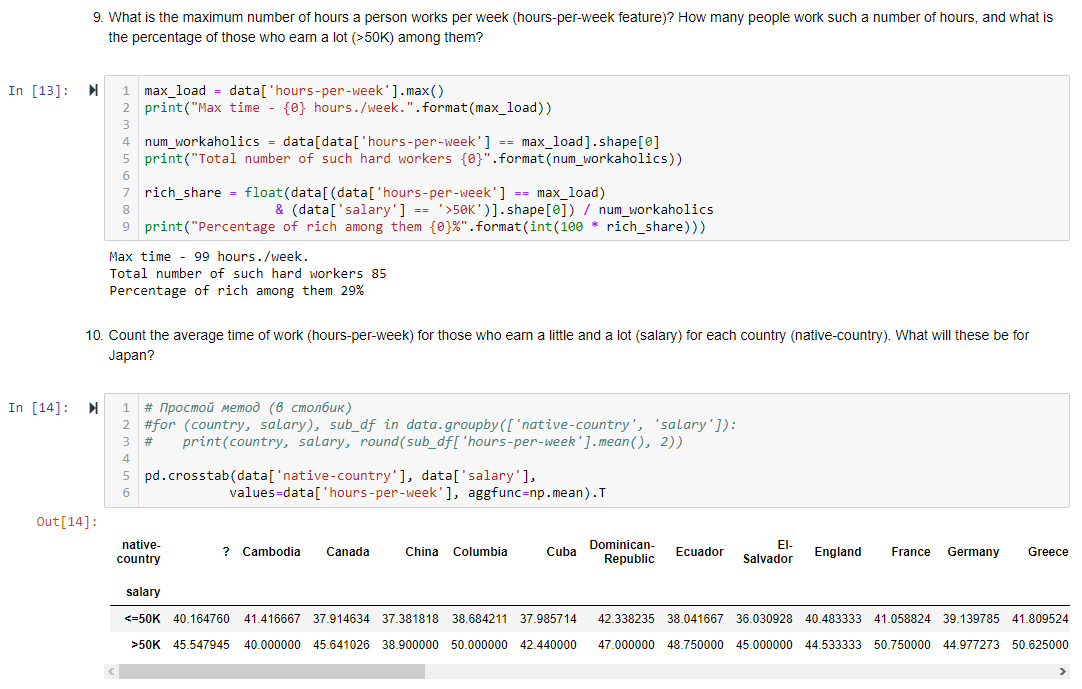
**Текст программы. Часть 1**



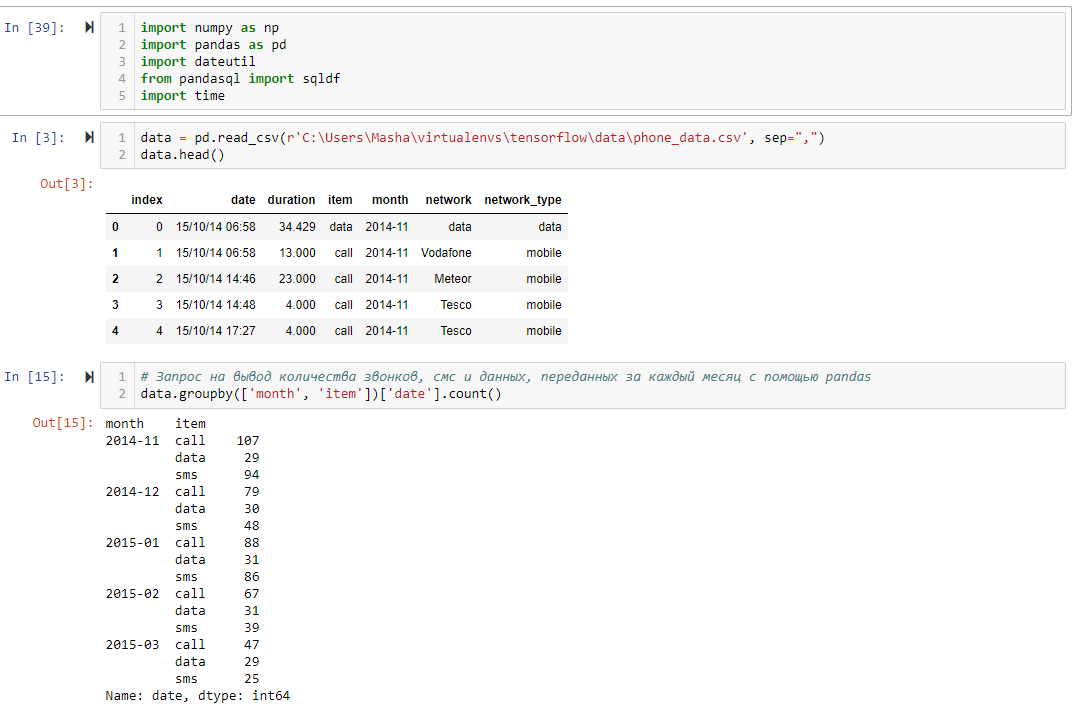


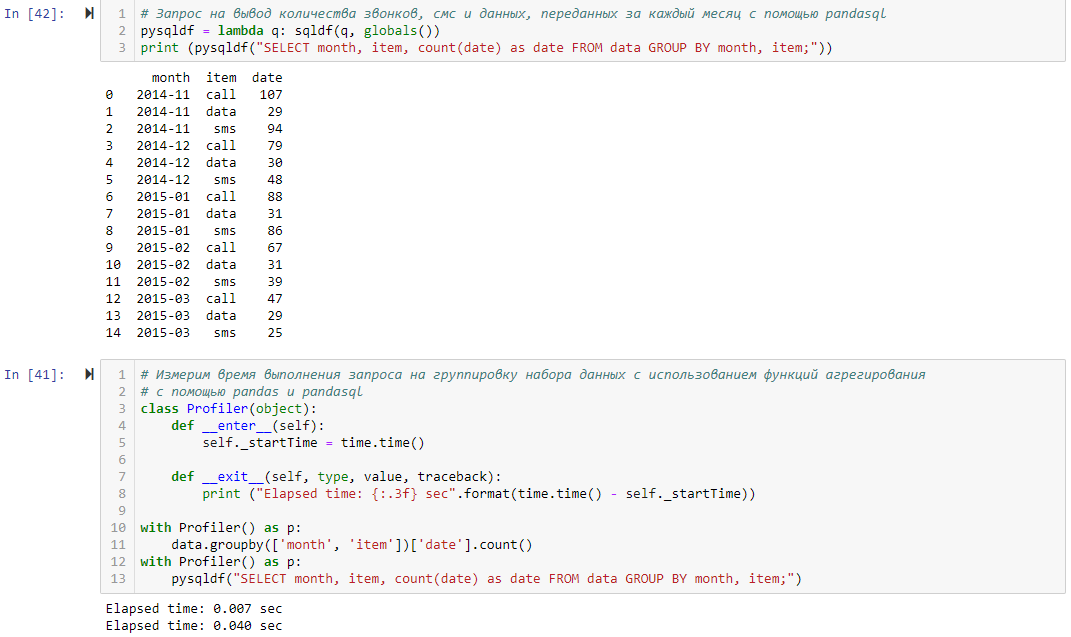
 

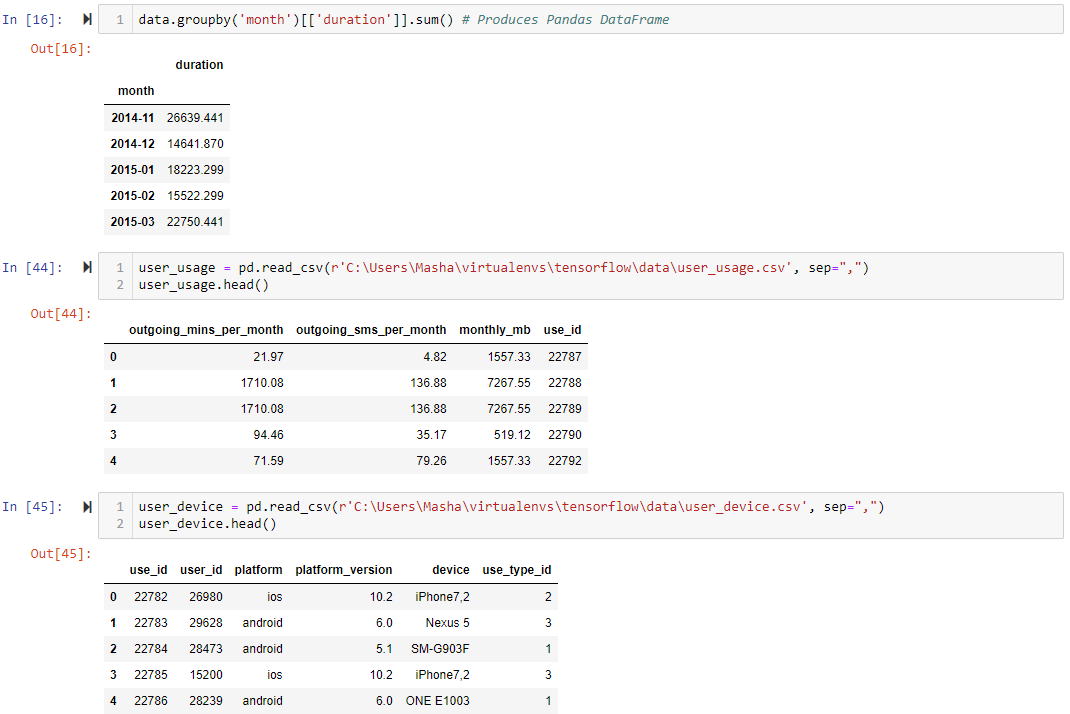


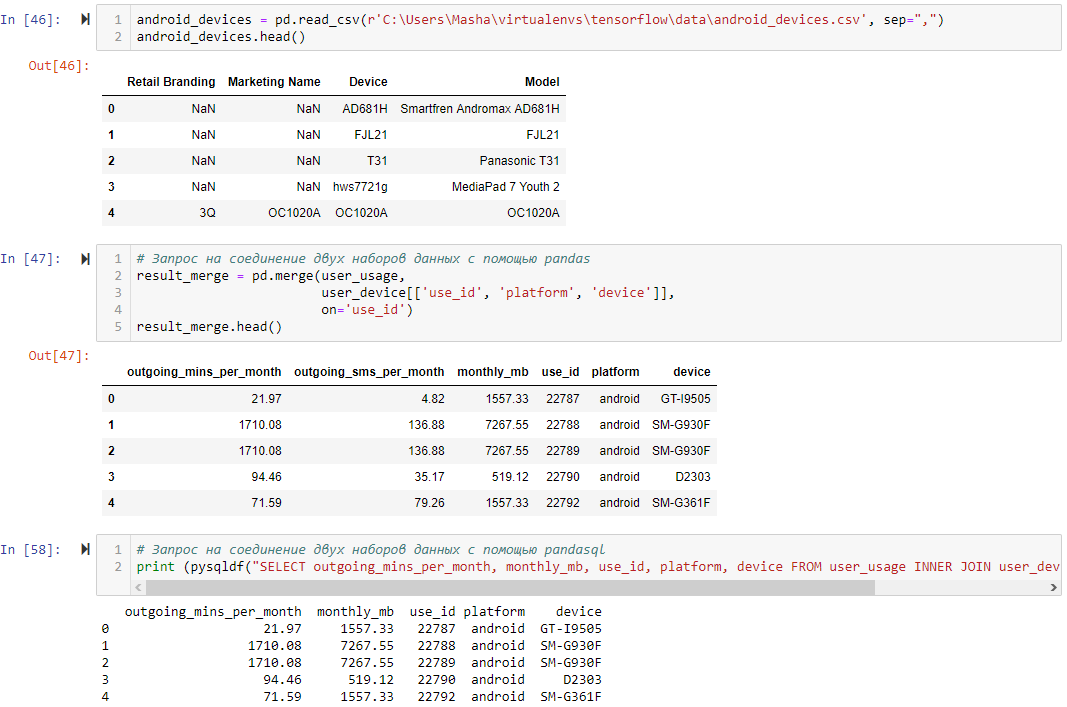


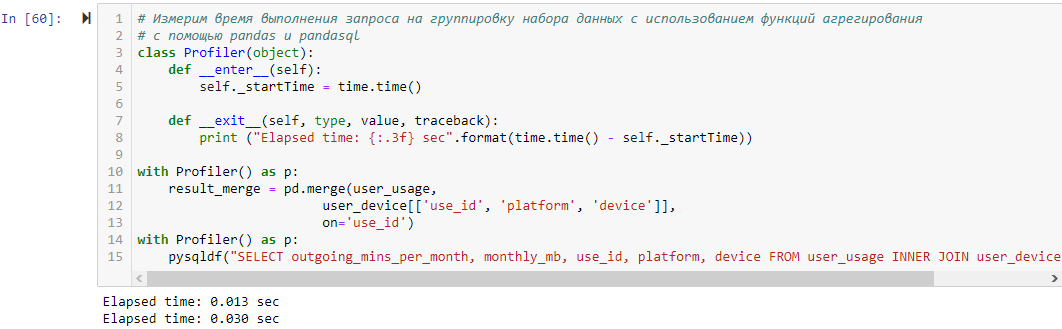
**Текст программы. Часть 2**









**Вывод**

После выполненных запросов можно говорить о том, что запросы с помощью библиотеки Pandas выполняются быстрее, чем запросы с помощью PandaSQL.