МГТУ им. Н. Э. Баумана, кафедра ИУ5

курс “Технология машинного обучения”

Лабораторная работа №2

# «[Изучение](https://github.com/ugapanyuk/ml_course/wiki/LAB_EDA_VISUALIZATION) библиотек обработки данных»

ВЫПОЛНИЛ:

Мелисов А.Е.

Группа: ИУ5-61Б

ПРОВЕРИЛ:

Гапанюк Ю.Е.

Москва 2020

**Цель:**

изучение библиотек обработки данных Pandas и PandaSQL.

**Задание:**

#### Часть 1.

Условие задания

* **In this task you should use Pandas to answer a few questions about the [Adult](https://archive.ics.uci.edu/ml/datasets/Adult) dataset.** <https://nbviewer.jupyter.org/github/Yorko/mlcourse_open/blob/master/jupyter_english/assignments_demo/assignment01_pandas_uci_adult.ipynb?flush_cache=true>

Набор данных

* <https://archive.ics.uci.edu/ml/datasets/Adult>

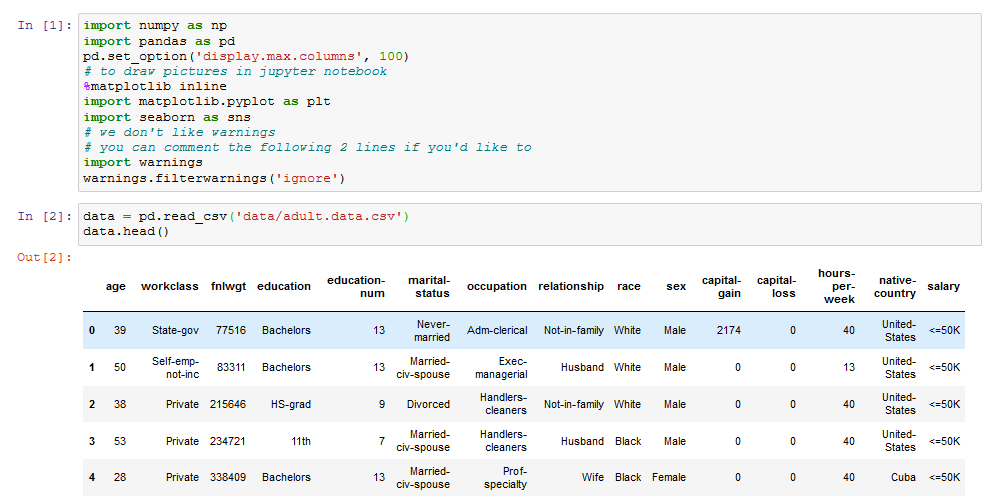
#### Часть 2.

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.

### **Выполненная работа**

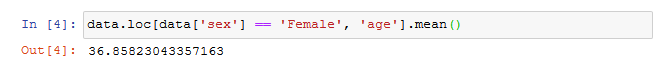


#### Часть 1.

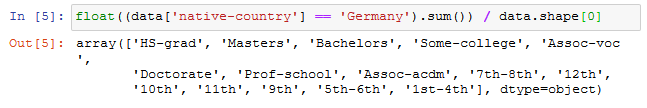
1. **How many men and women (**sex**feature) are represented in this dataset?**



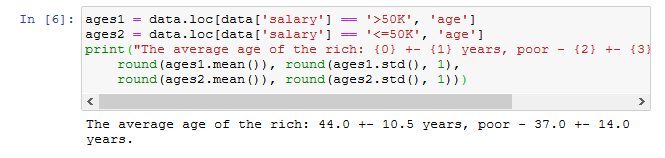
1. **What is the average age (**age**feature) of women?**



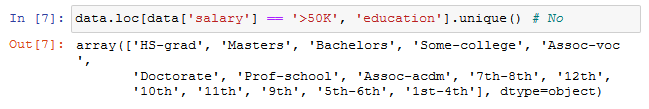
1. **What is the percentage of German citizens (**native-country**feature)?**



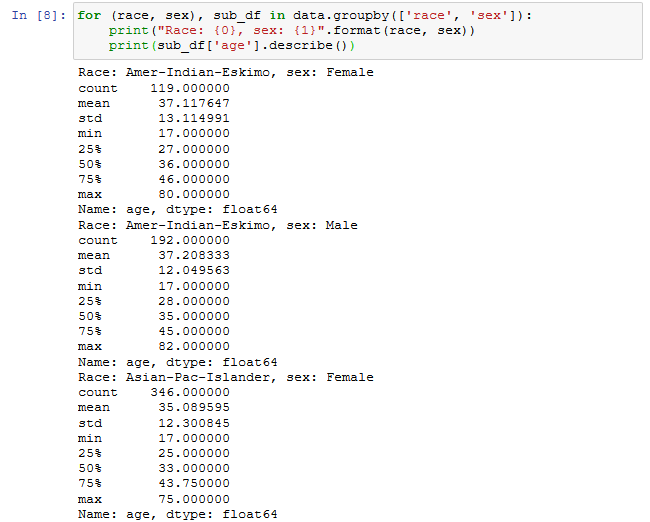
1. **What are the mean and standard deviation of age for those who earn more than 50K per year (**salary**feature) and those who earn less than 50K per year?**



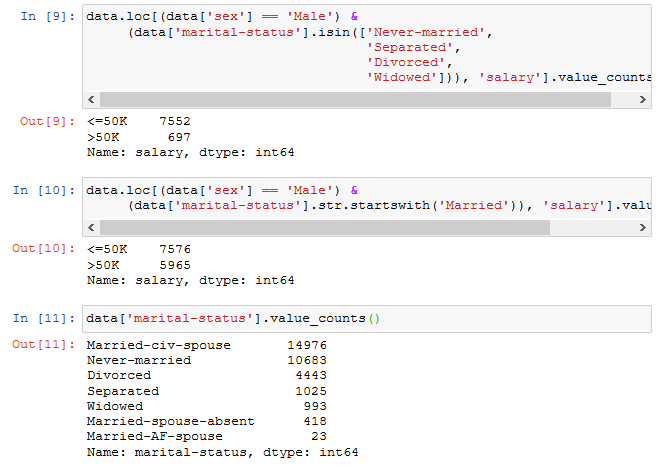
1. **Is it true that people who earn more than 50K have at least high school education? (**education – Bachelors, Prof-school, Assoc-acdm, Assoc-voc, Masters**or**Doctorate**feature)**



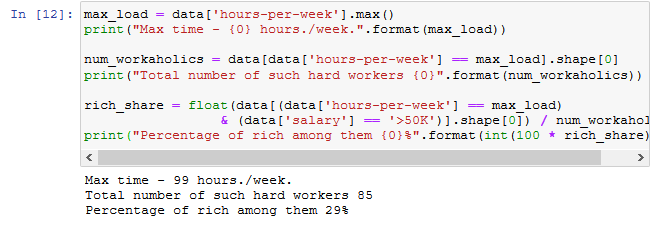
1. **Display statistics of age for each race (**race**feature) and each gender. Use**groupby()**and**describe()**. Find the maximum age of men of**Amer-Indian-Eskimo**race.**



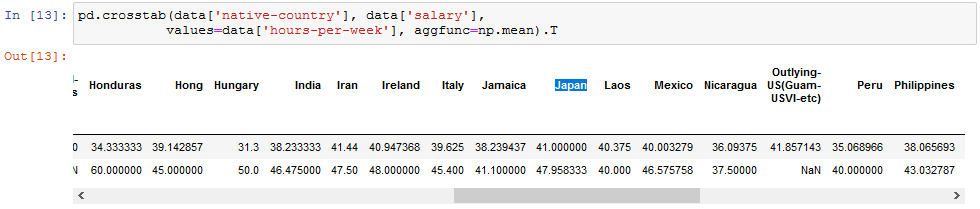
1. **Among whom is the proportion of those who earn a lot (>50K) greater: married or single men (**marital-status**feature)? Consider as married those who have a**marital-status**starting with**Married**(Married-civ-spouse, Married-spouse-absent or Married-AF-spouse), the rest are considered bachelors.**



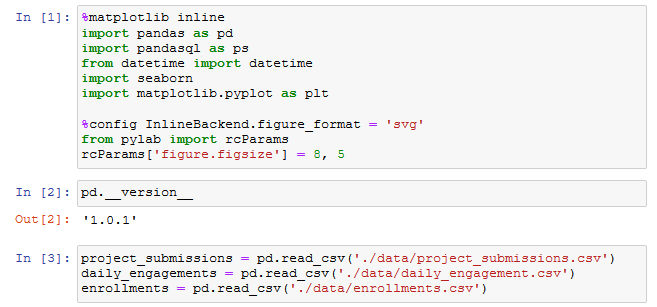
1. **What is the maximum number of hours a person works per week (**hours-per-week**feature)? How many people work such a number of hours, and what is the percentage of those who earn a lot (>50K) among them?**



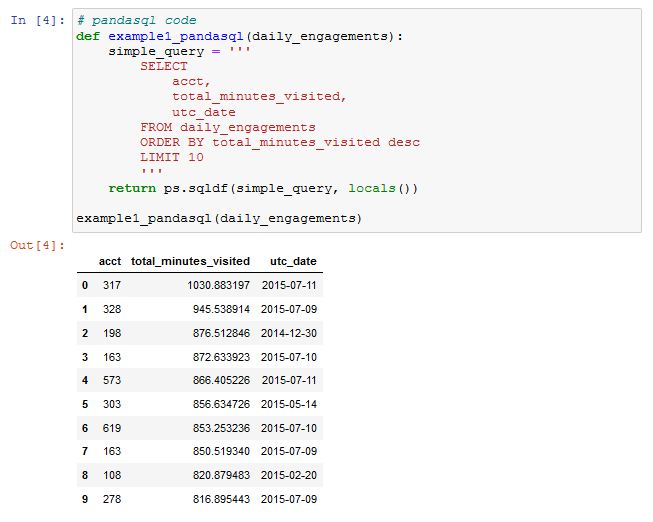
1. **Count the average time of work (**hours-per-week**) for those who earn a little and a lot (**salary**) for each country (**native-country**). What will these be for Japan?**

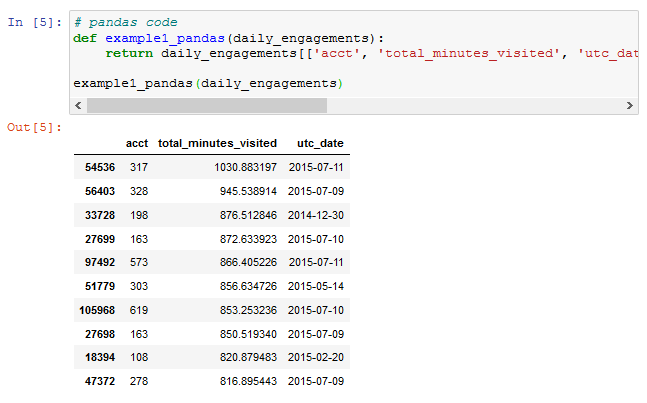


#### Часть 2.

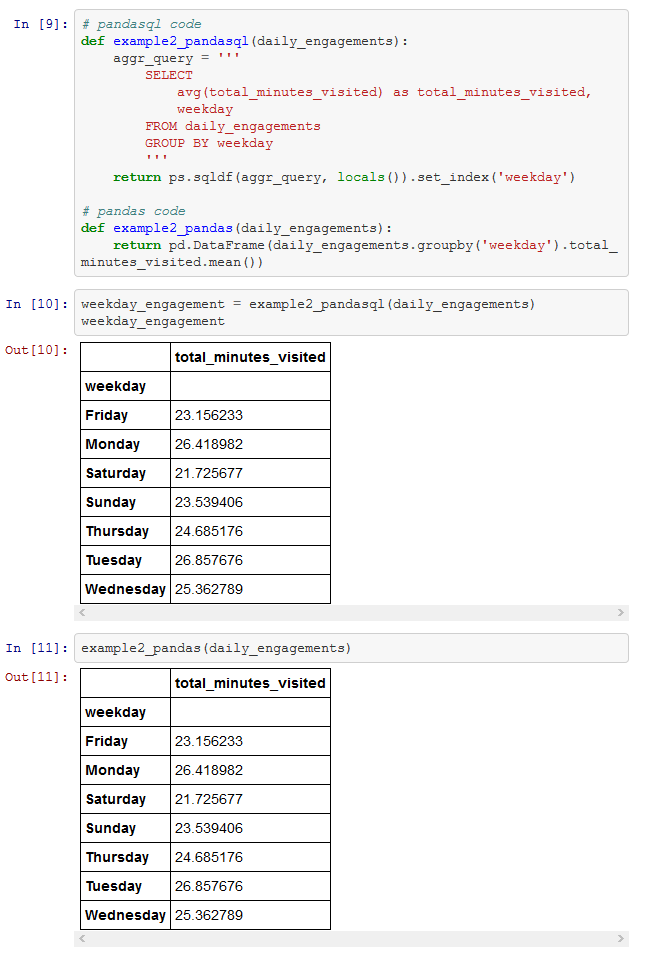


1. Один произвольный запрос на соединение двух наборов данных:





1. Один произвольный запрос на группировку набора данных с использованием функций агрегирования:



Гитхаб: <https://github.com/adammemes/TMO>