|  |  |
| --- | --- |
| Gerb-BMSTU_01 | **Министерство науки и высшего образования Российской Федерации**  **Федеральное государственное бюджетное образовательное учреждение**  **высшего образования**  **«Московский государственный технический университет**  **имени Н.Э. Баумана**  **(национальный исследовательский университет)»**  **(МГТУ им. Н.Э. Баумана)** |

**Факультет «Информатика и системы управления»**

**Кафедра ИУ5 «Системы обработки информации и управления»**

Лабораторная работа№2

по дисциплине «Технологии машинного обучения»

по теме «Изучение библиотек обработки данных»

Выполнил:

студентка группы ИУ5-64

Изъюрова Маргарита

\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

подпись, дата

Проверил:

Гапанюк Юрий Евгеньевич

\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

подпись, дата

2019 г.

Содержание:

1. [Задание 3](#_Toc6264189)
2. [Текст программы с примерами выполнения 3](#_Toc6264194)

## 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>

Пример решения задания - <https://www.kaggle.com/kashnitsky/a1-demo-pandas-and-uci-adult-dataset-solution>

#### 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.

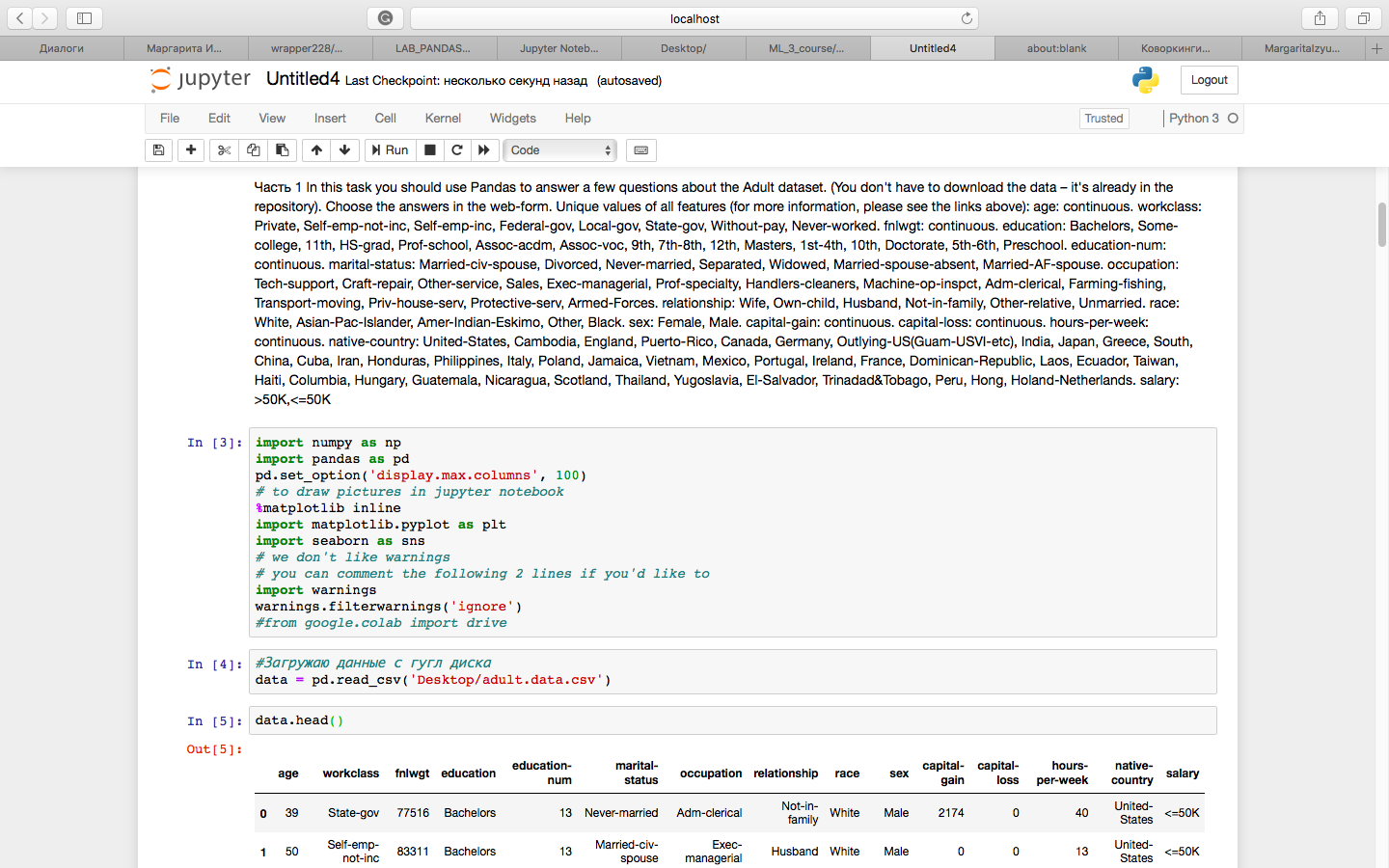
В качестве примеров можно использовать следующие статьи:

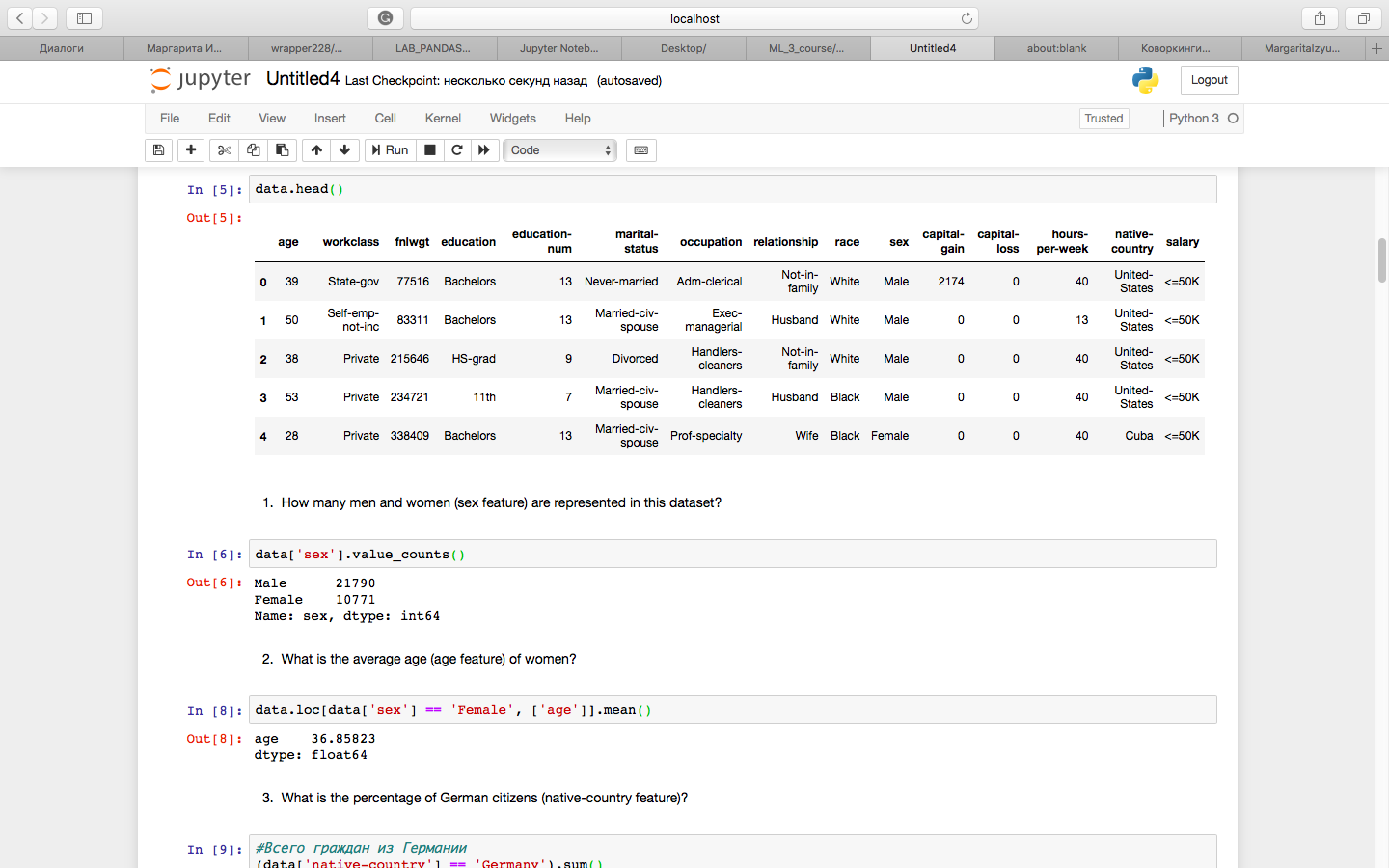
* <https://www.shanelynn.ie/summarising-aggregation-and-grouping-data-in-python-pandas/>
* <https://www.shanelynn.ie/merge-join-dataframes-python-pandas-index-1/> (в разделе "Example data" данной статьи содержится рекомендуемый набор данных для проведения экспериментов).

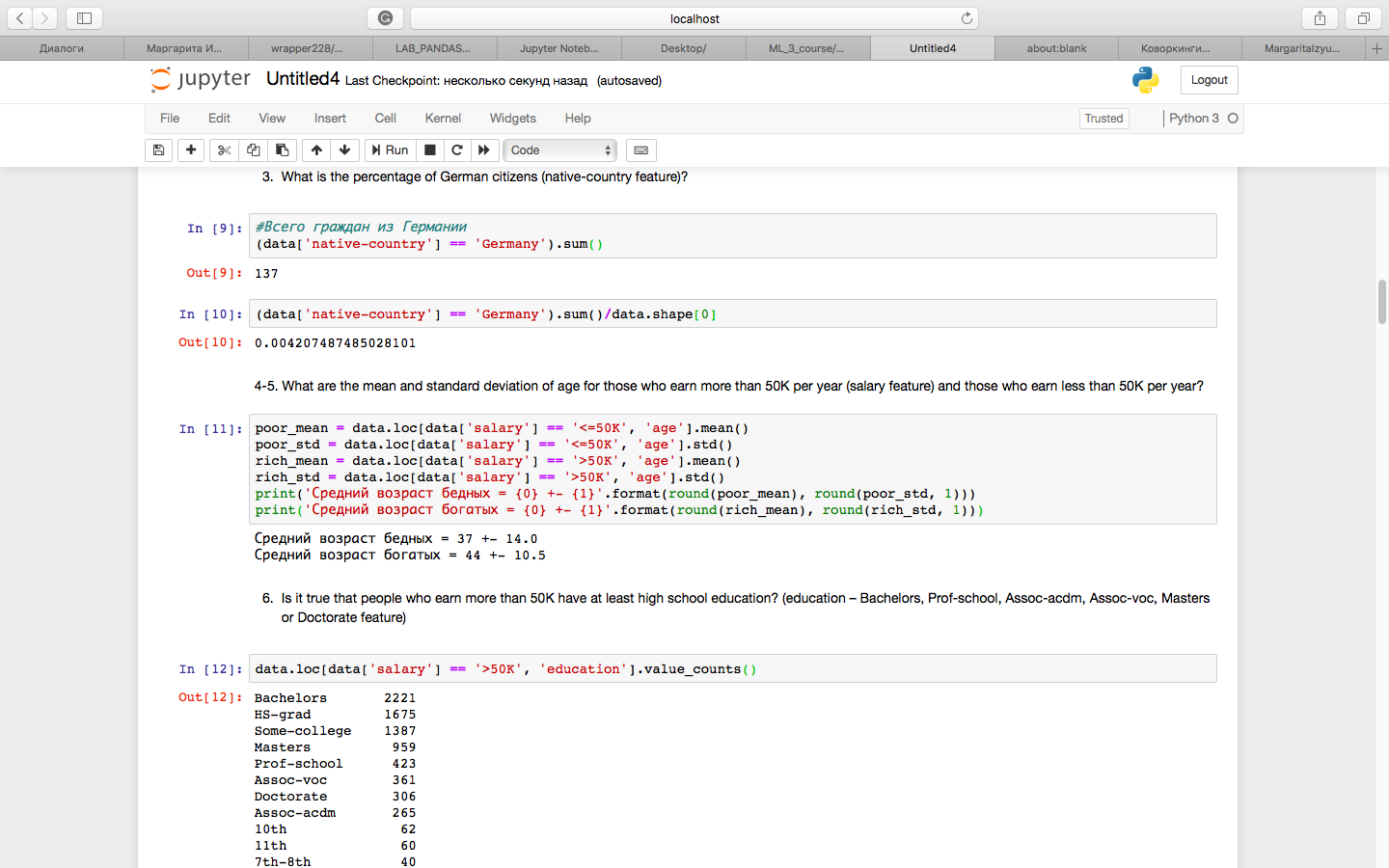
Пример сравнения Pandas и PandaSQL - <https://github.com/miptgirl/udacity_engagement_analysis/blob/master/pandasql_example.ipynb>

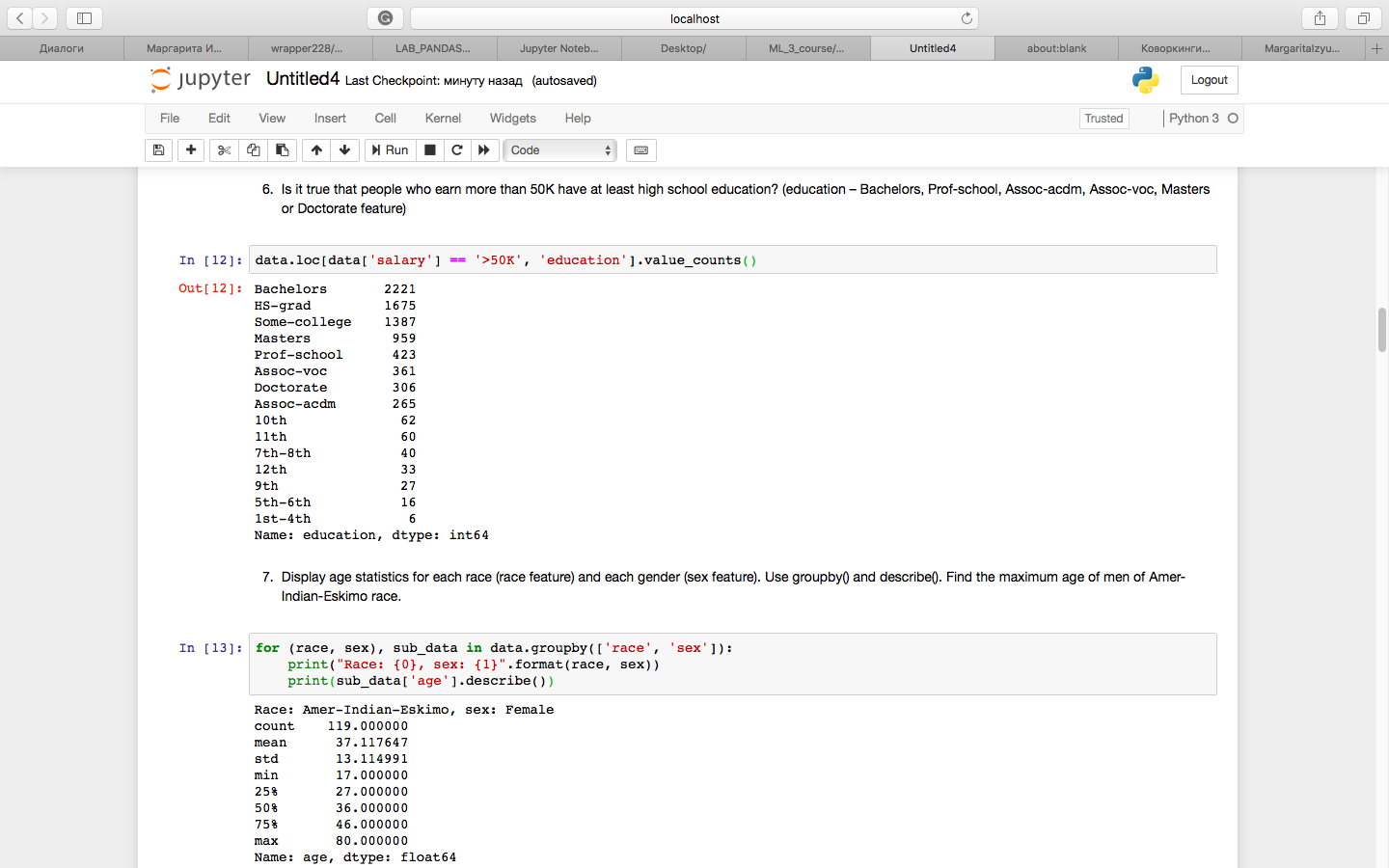
Набор упражнений по Pandas с решениями - <https://github.com/guipsamora/pandas_exercises>

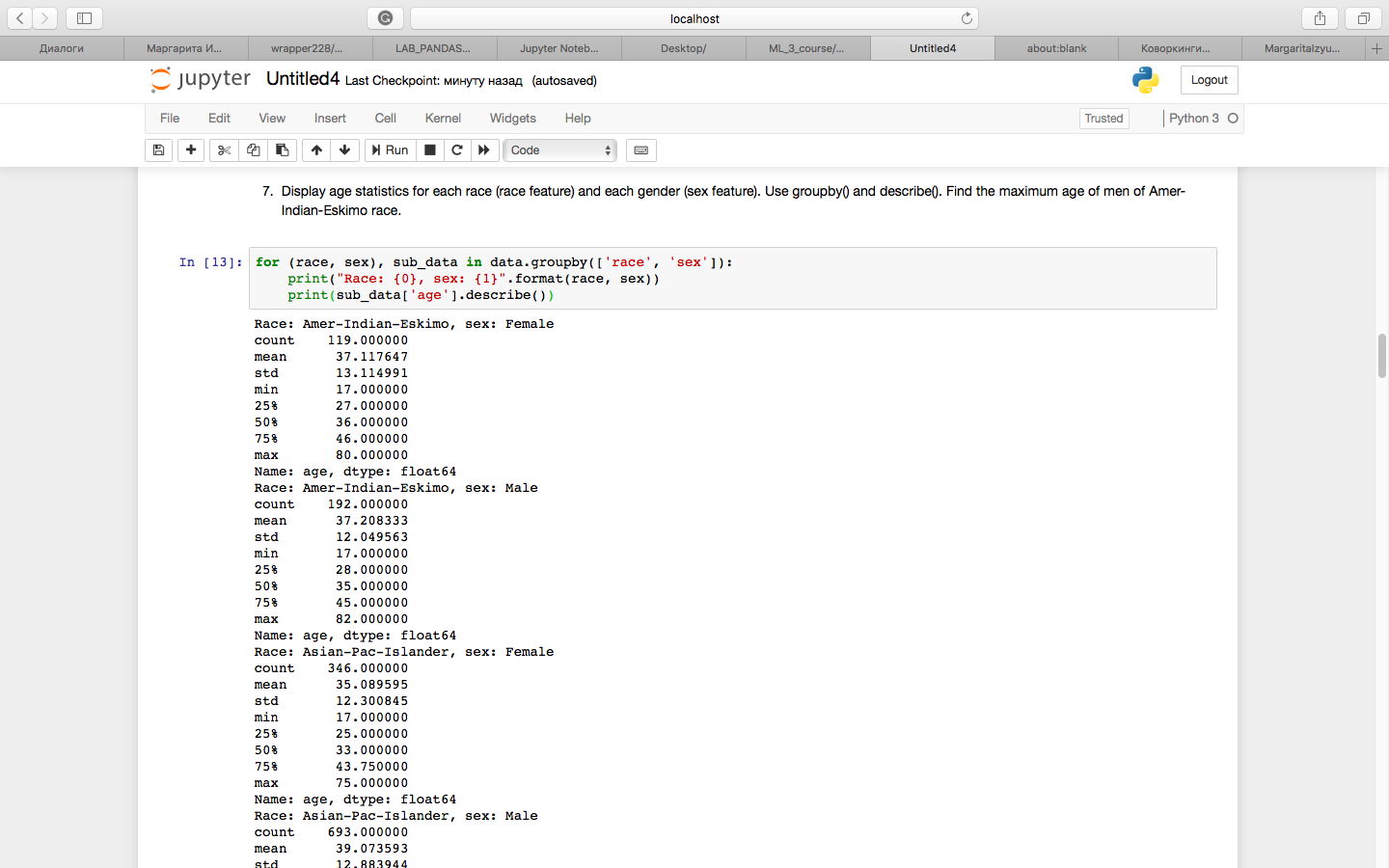
**Текст программы с примерами выполнения**

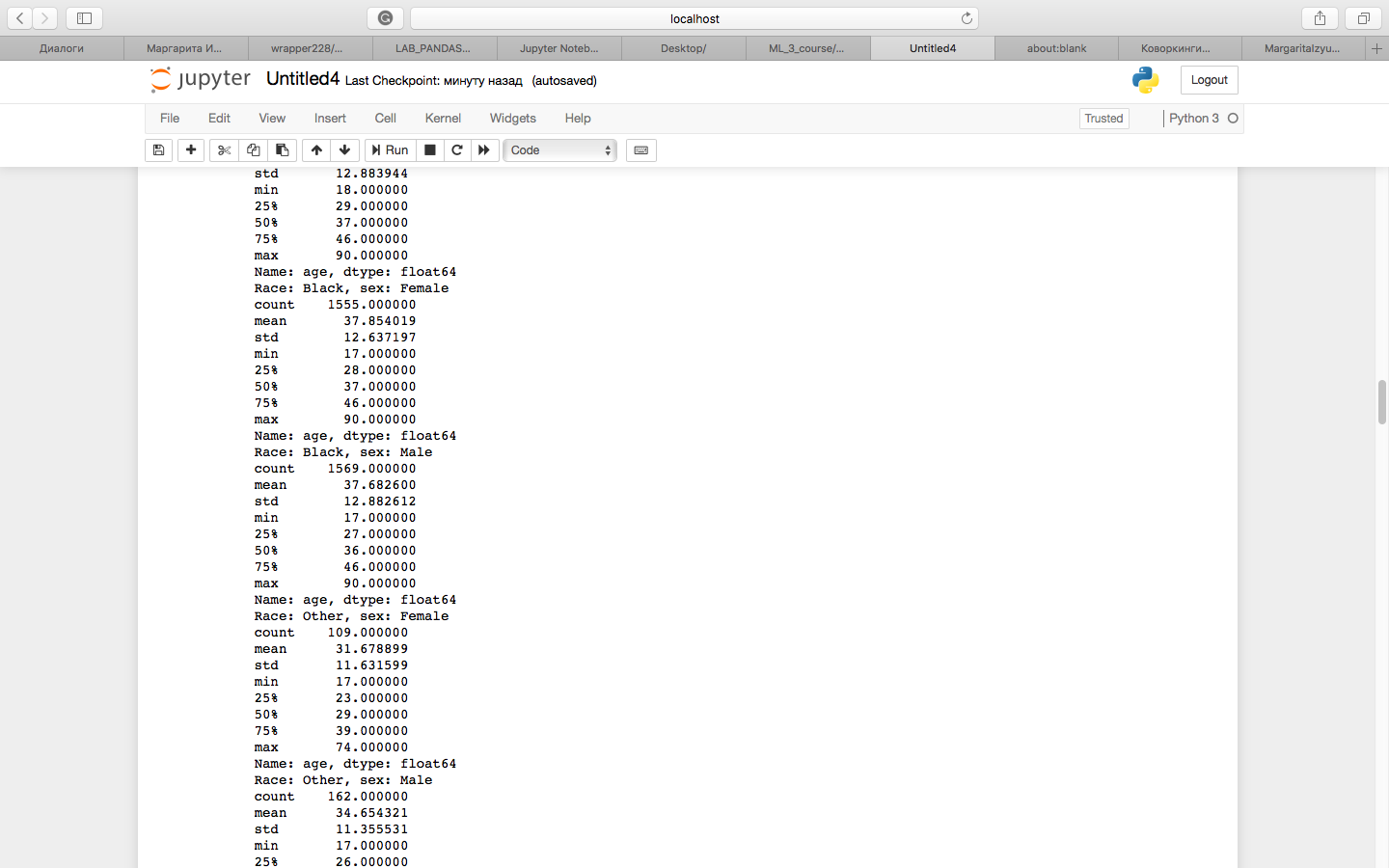
****

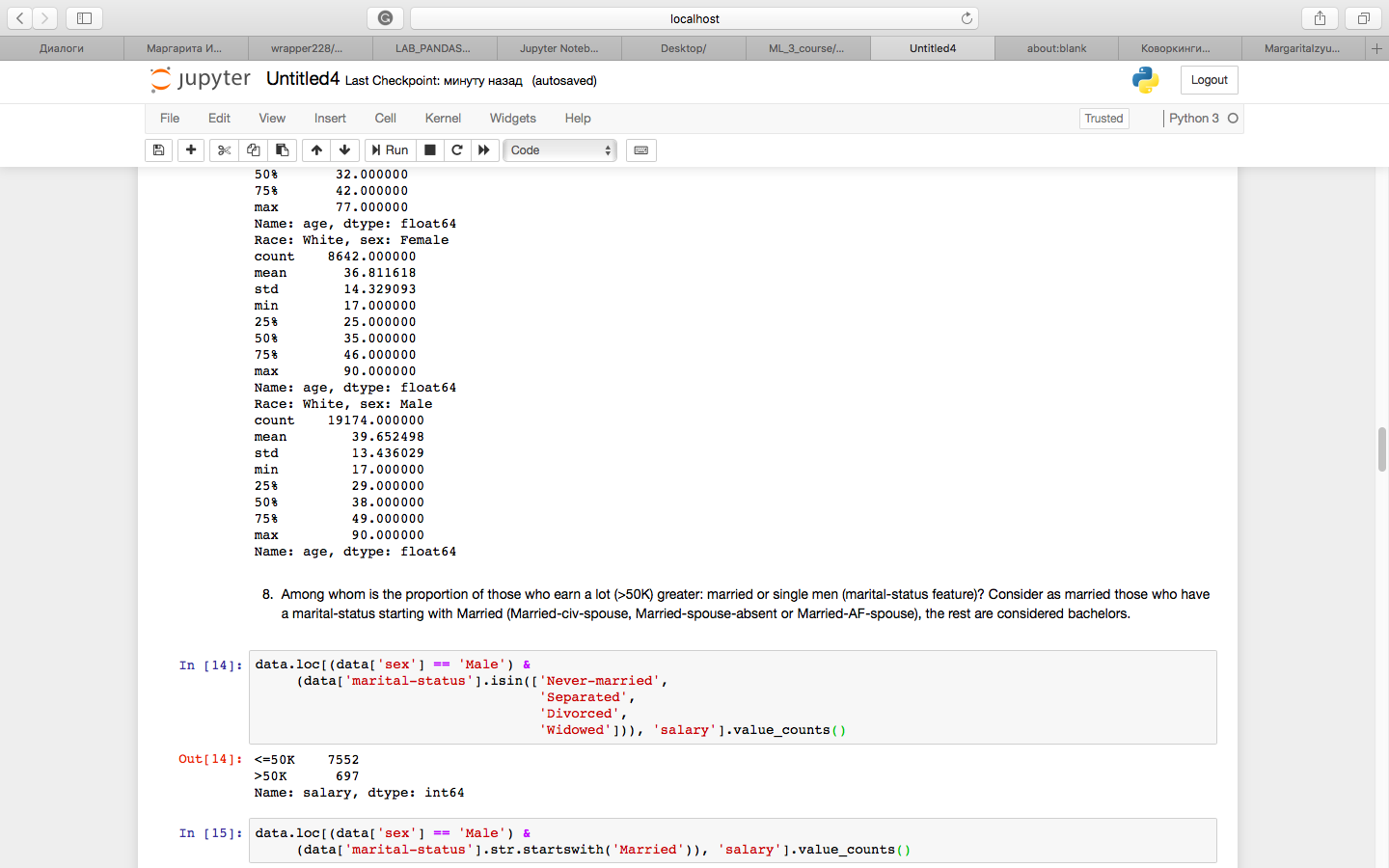
****

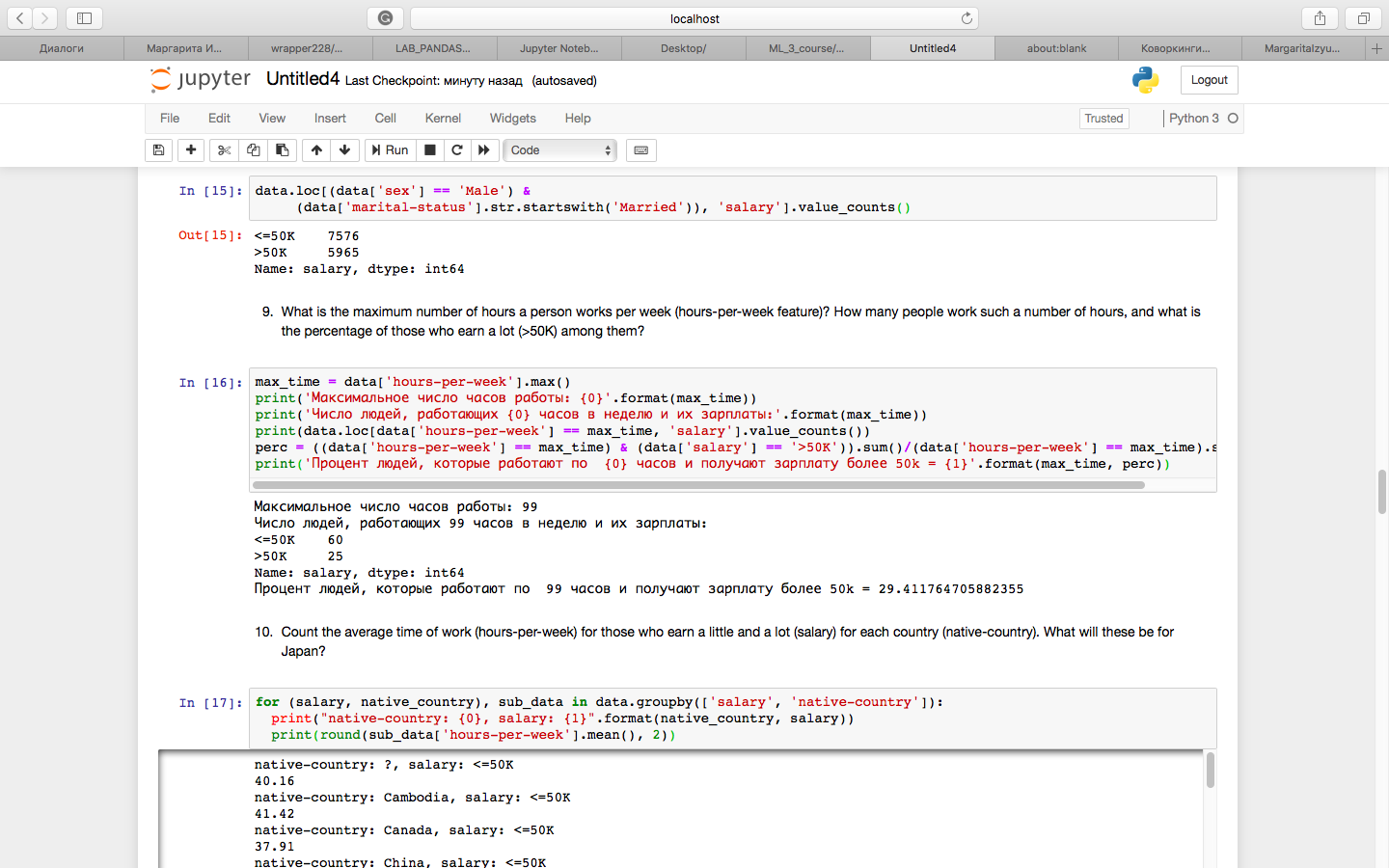
****

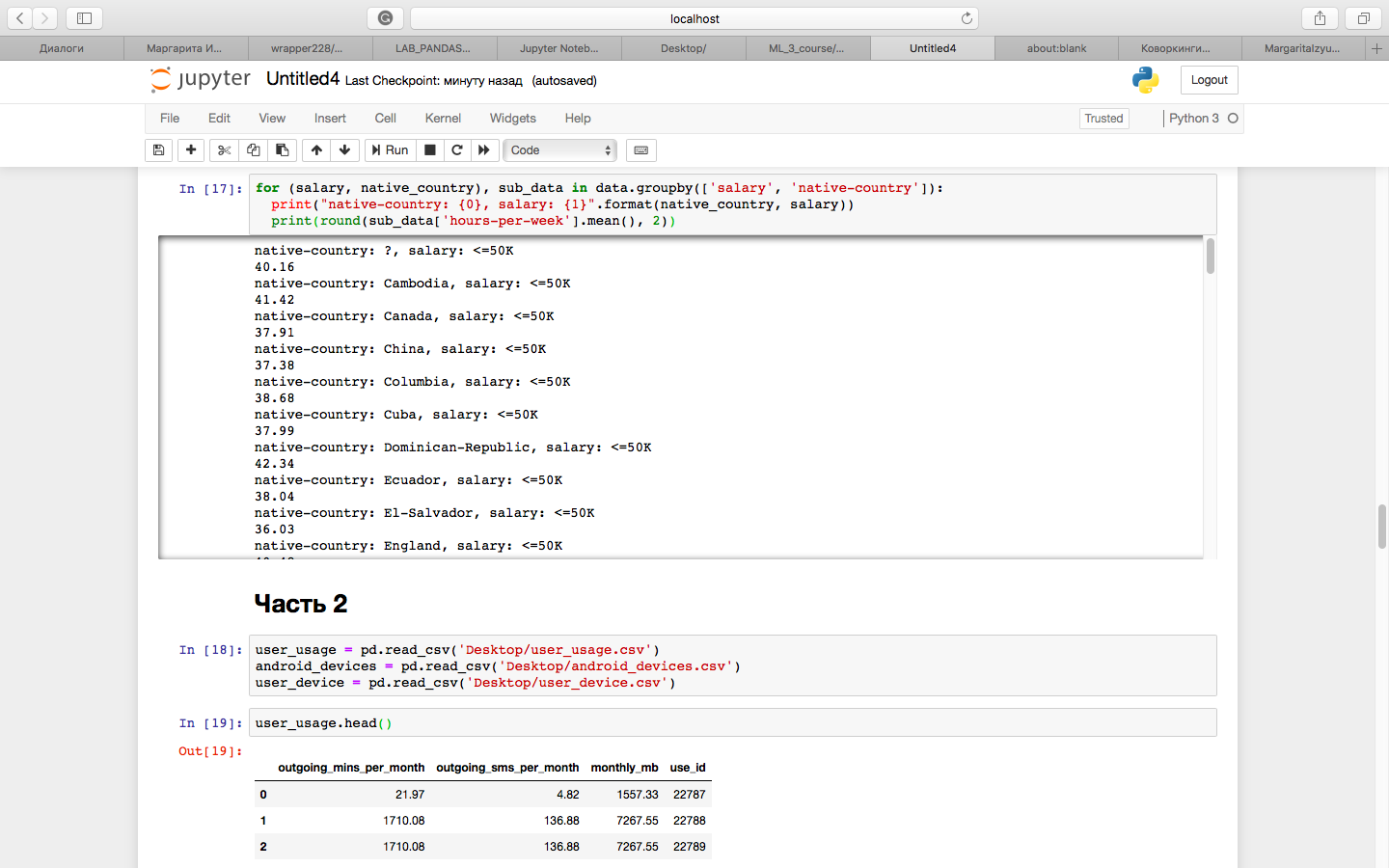
****

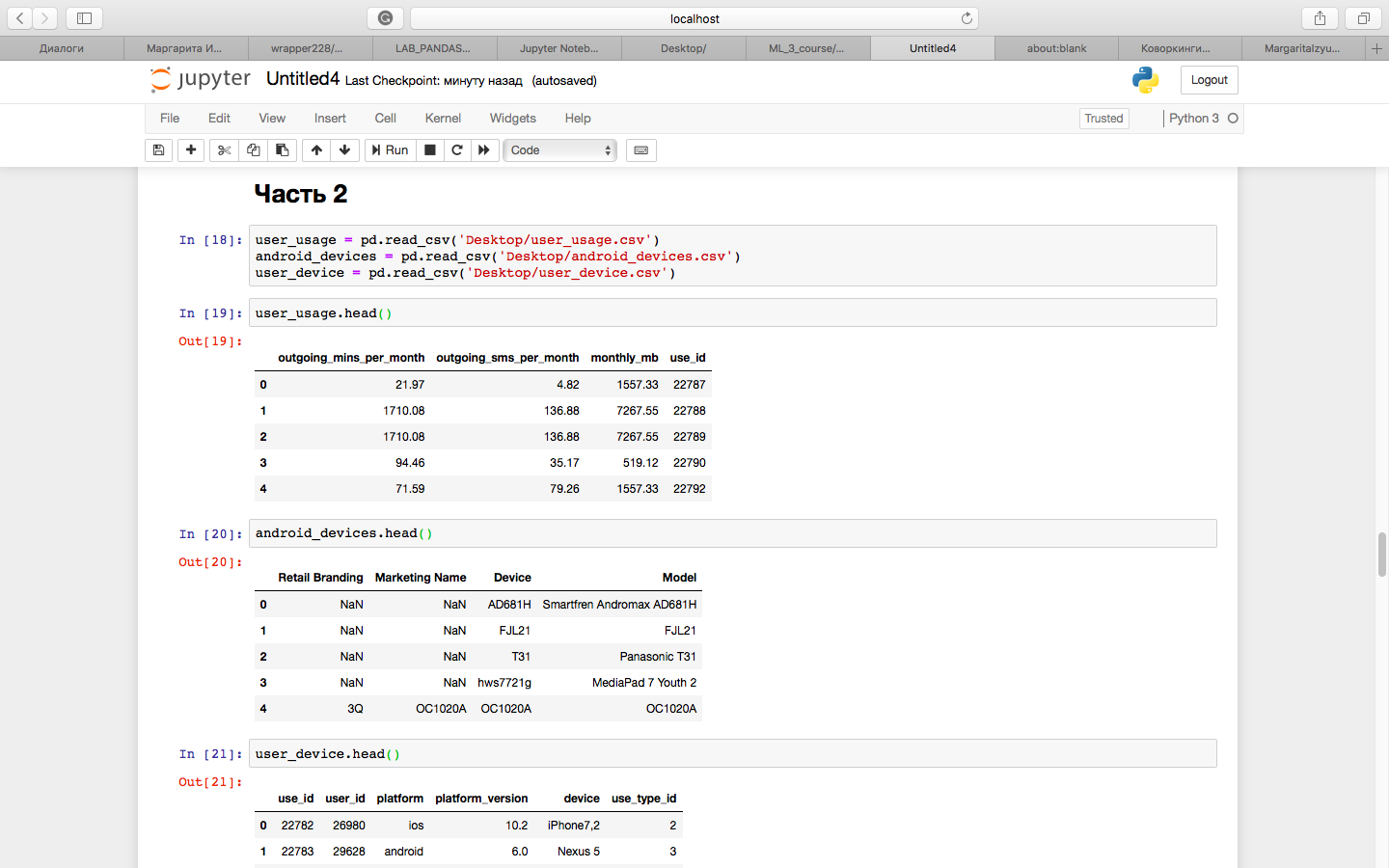
****

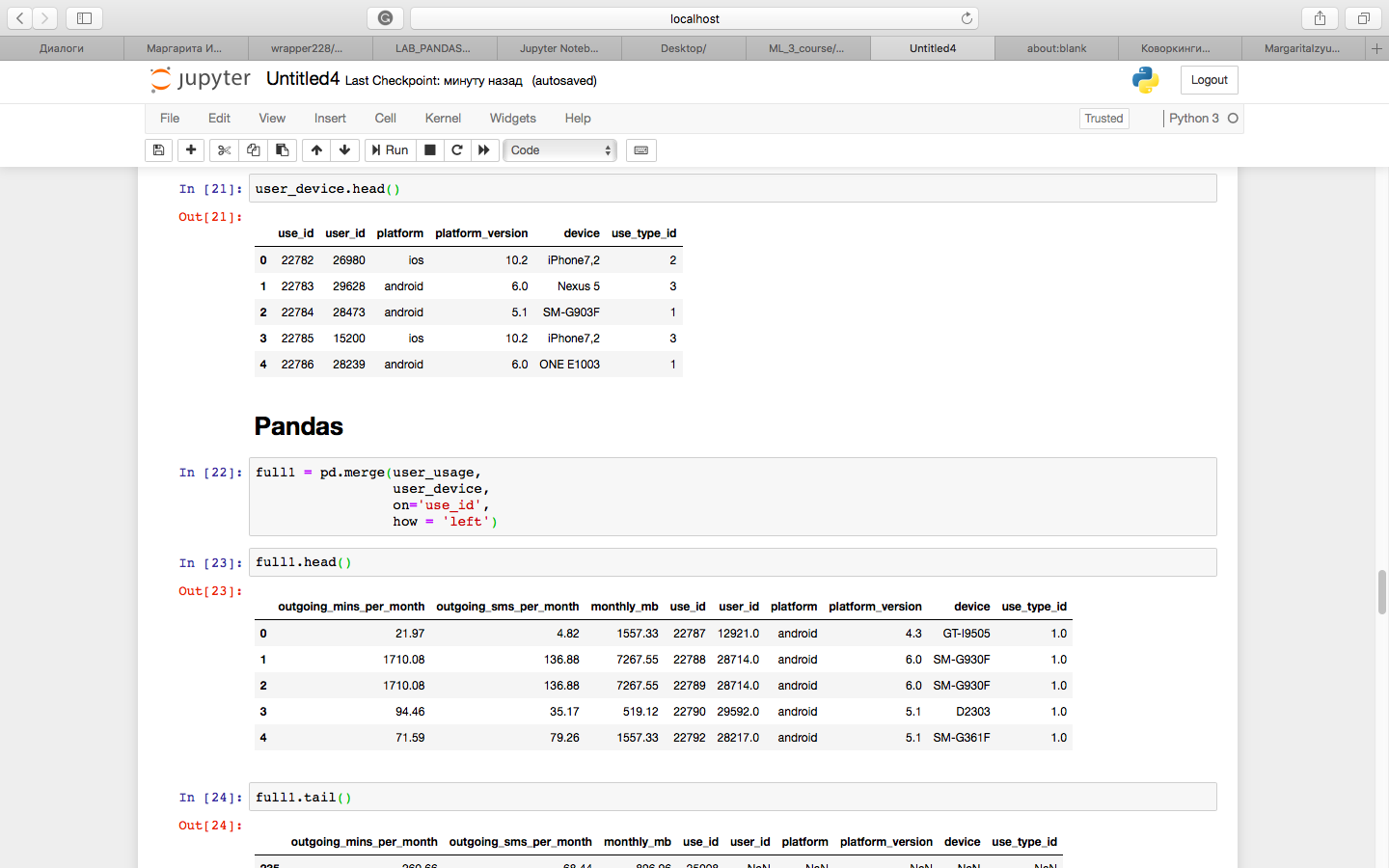
****

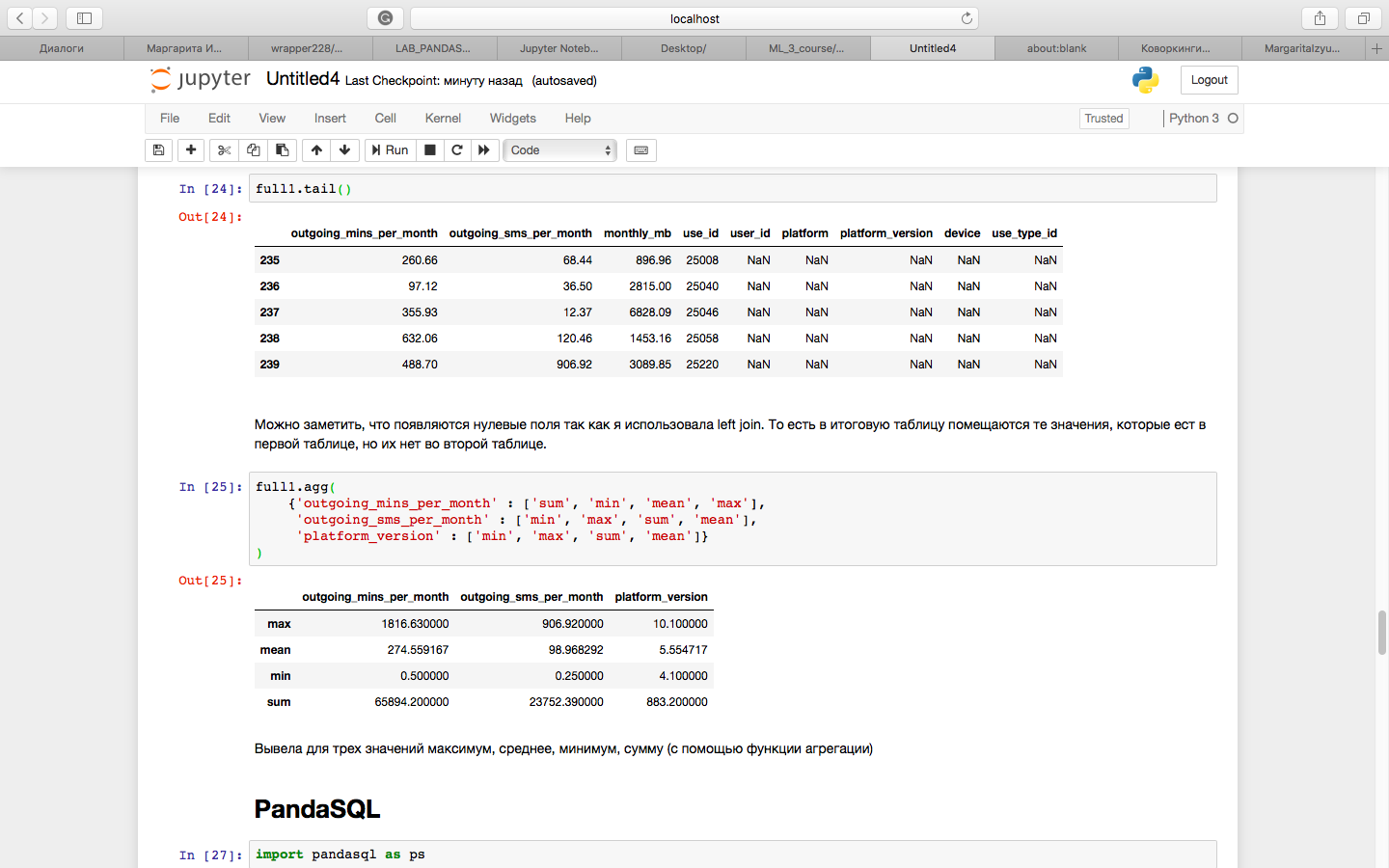
****

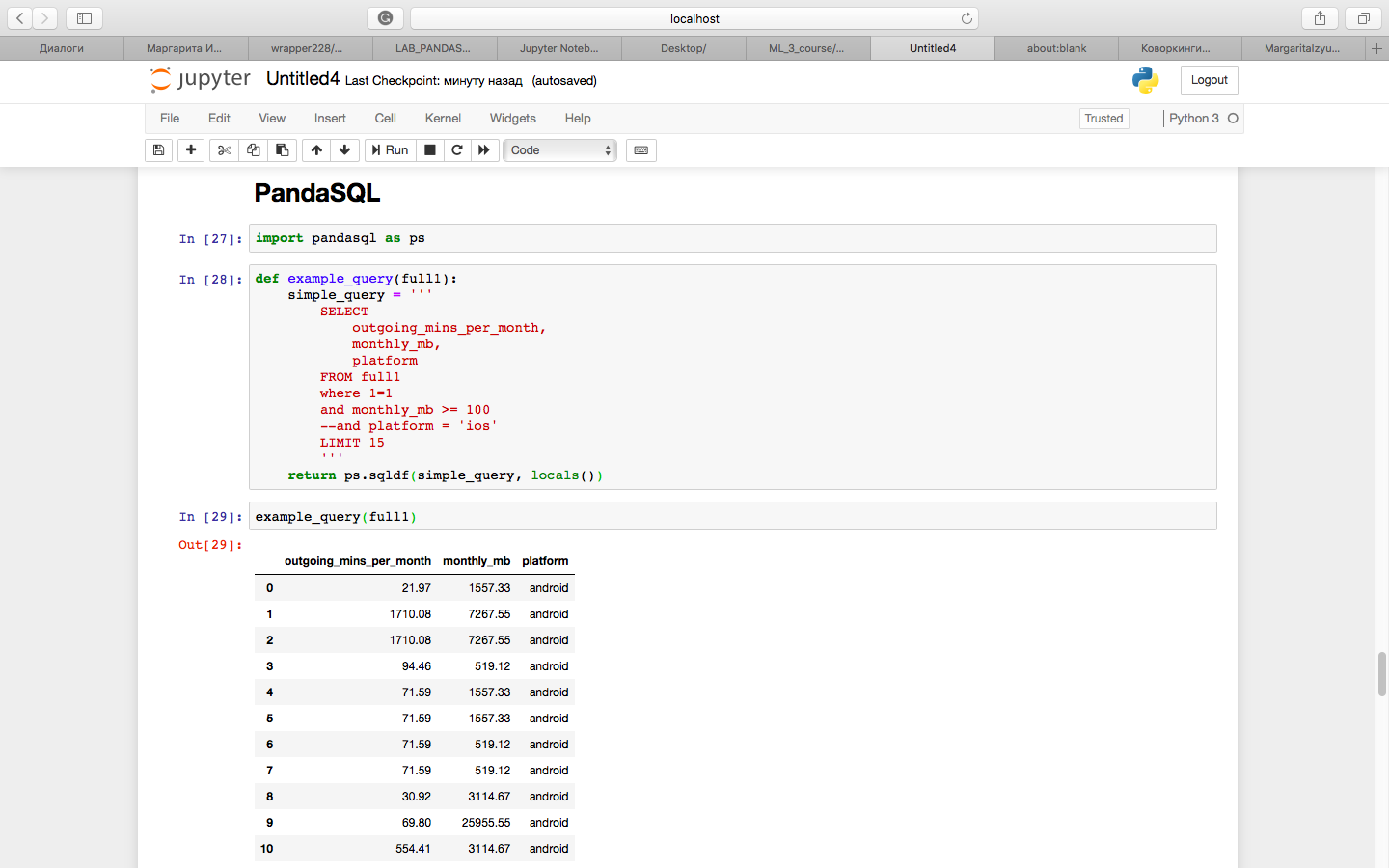
****

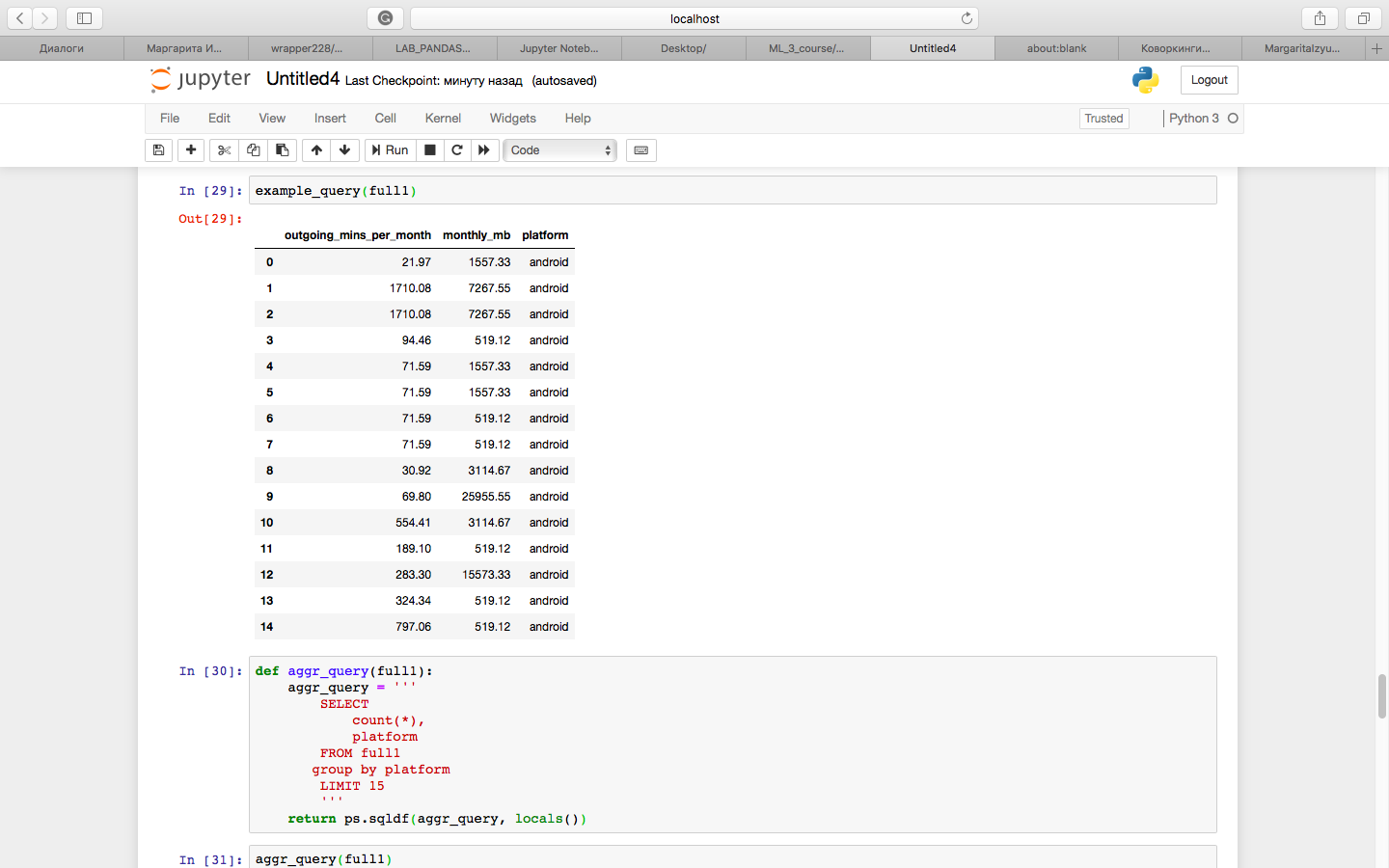
****

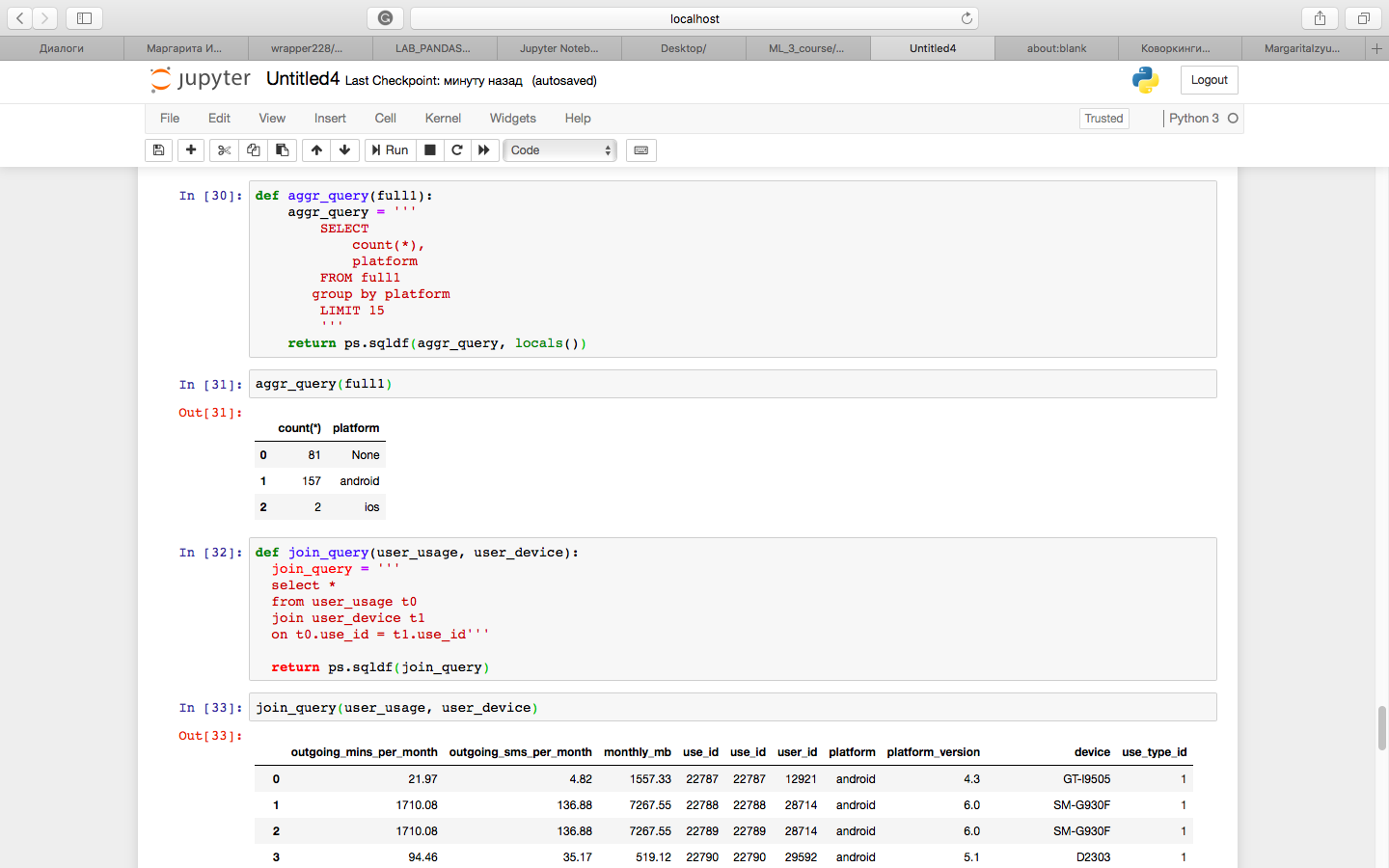
****

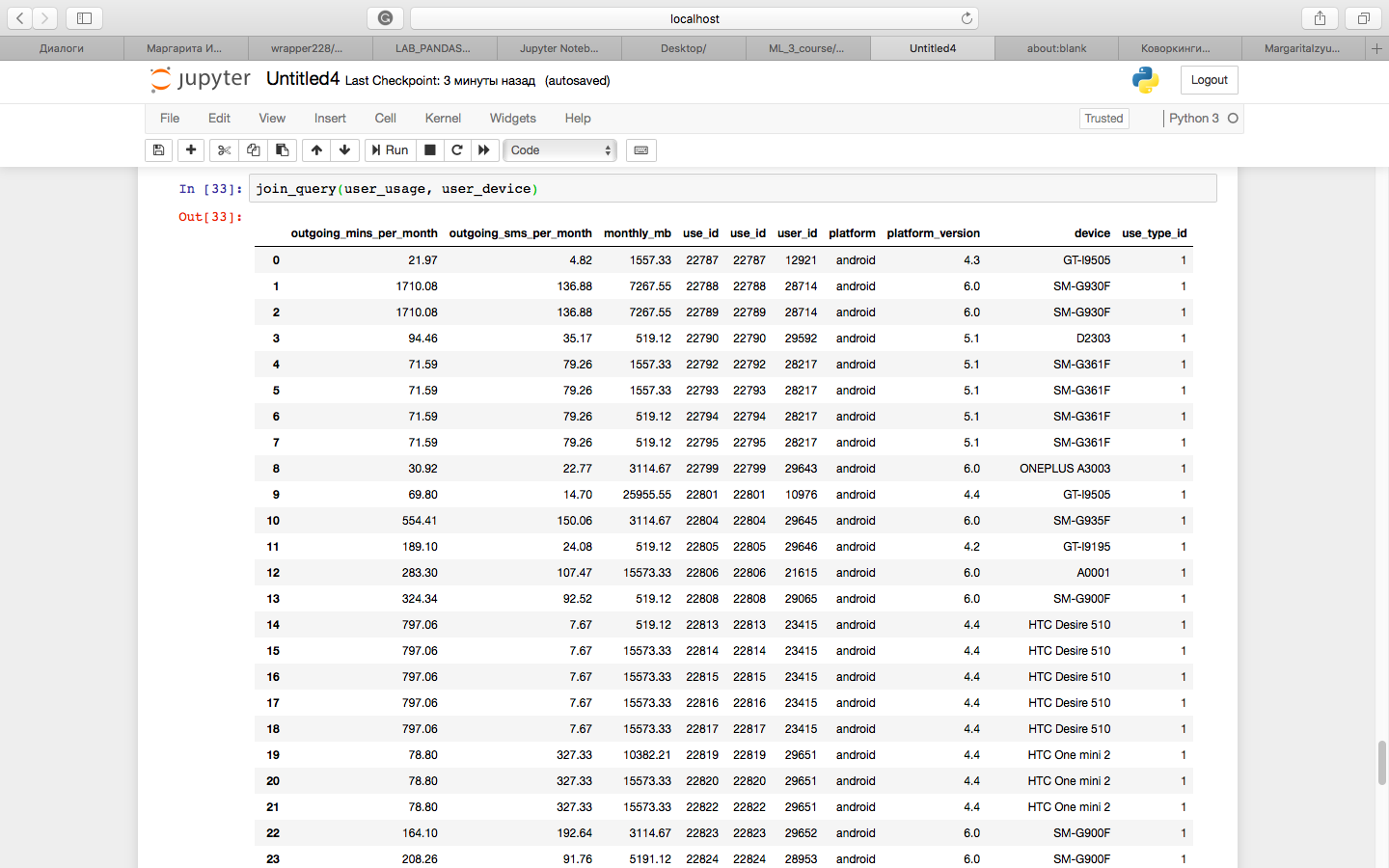
****

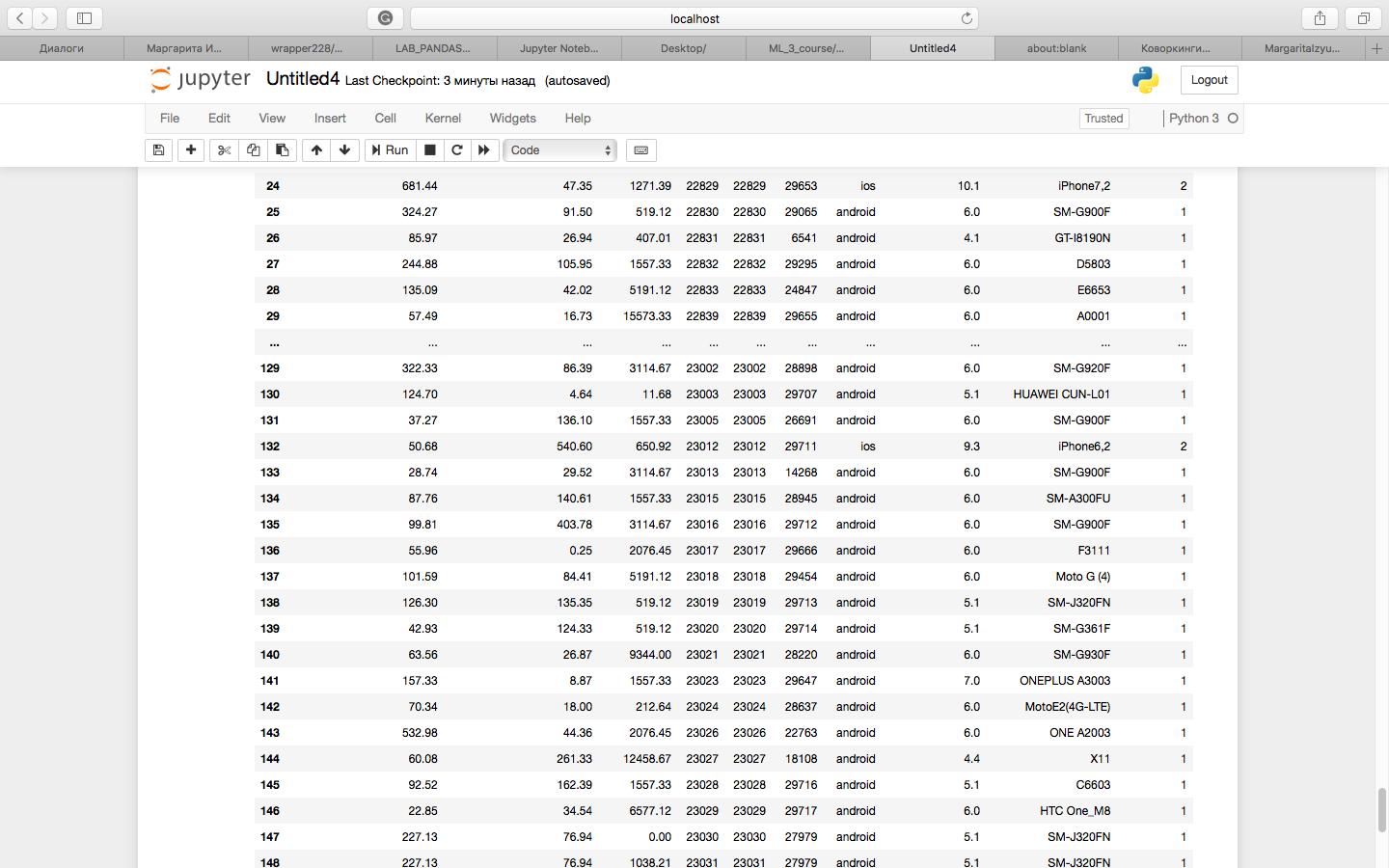
****

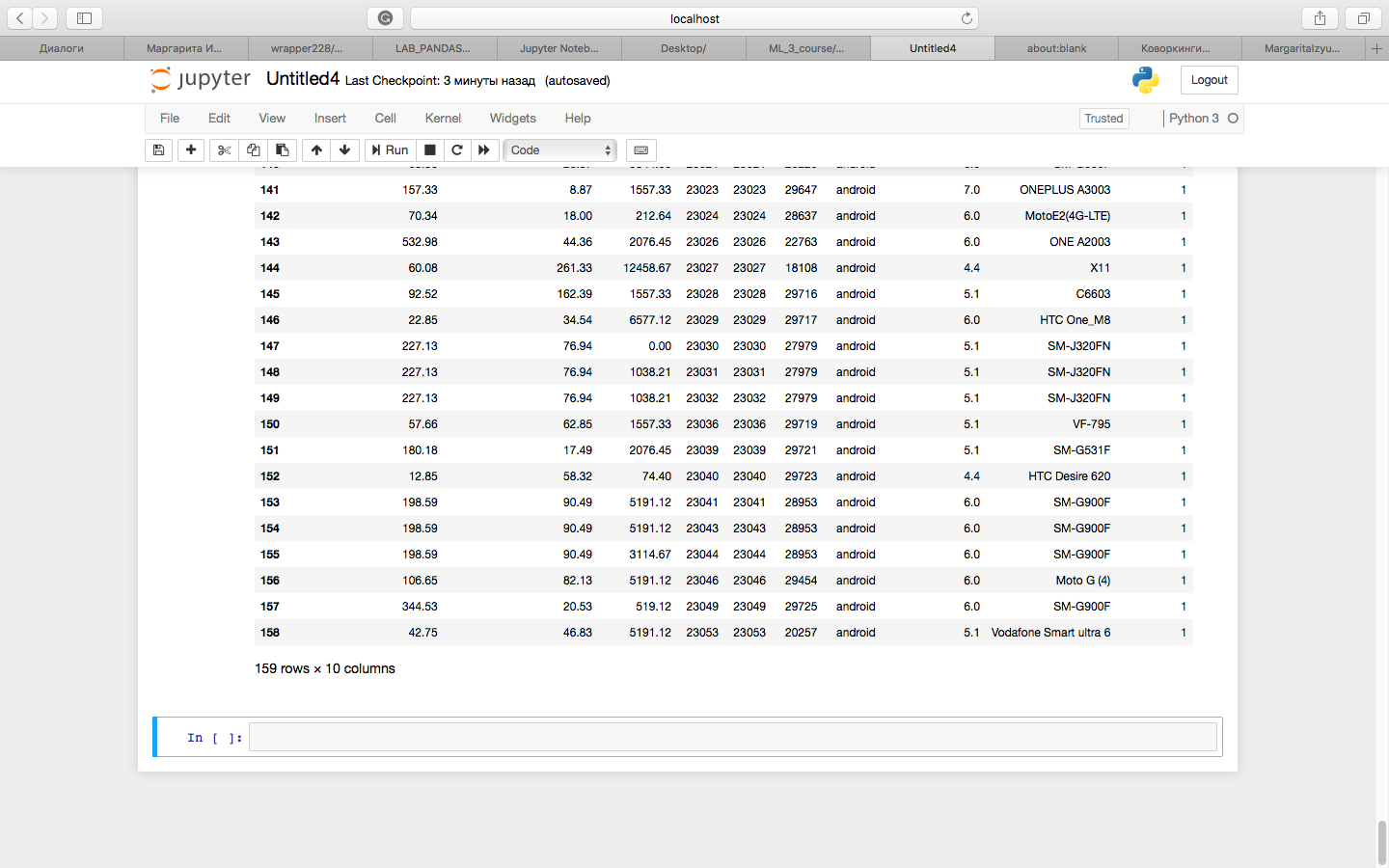
****

****

****

****

****

****