|  |  |
| --- | --- |
| 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»**

Курсовая работа

по дисциплине «Технологии машинного обучения»

на тему:

Исследование датасета. Решение задачи регрессии

Выполнил:

студент группы № ИУ5-62

Чернышев Павел

Проверил:

Гапанюк Ю.Е.

2020 г.

**Задание**

1. Поиск и выбор набора данных для построения моделей машинного обучения. Построение модели машинного обучения и решение задачи регрессии.
2. Проведение разведочного анализа данных. Построение графиков, необходимых для понимания структуры данных. Анализ и заполнение пропусков в данных.
3. Выбор признаков, подходящих для построения моделей. Кодирование категориальных признаков Масштабирование данных. Формирование вспомогательных признаков, улучшающих качество моделей.
4. Проведение корреляционного анализа данных. Формирование промежуточных выводов о возможности построения моделей машинного обучения.
5. Выбор метрик для последующей оценки качества моделей.
6. Выбор наиболее подходящих моделей для решения задачи регрессии.
7. Построение базового решения (baseline) для выбранных моделей без подбора гиперпараметров. Производится обучение моделей на основе обучающей выборки и оценка качества моделей на основе тестовой выборки.
8. Подбор гиперпараметров для выбранных моделей.
9. Формирование выводов о качестве построенных моделей на основе выбранных метрик

**Оглавление**

[Введение 4](#_Toc10193111)

[Основная часть 5](#_Toc10193112)

[Заключение 12](#_Toc10193113)

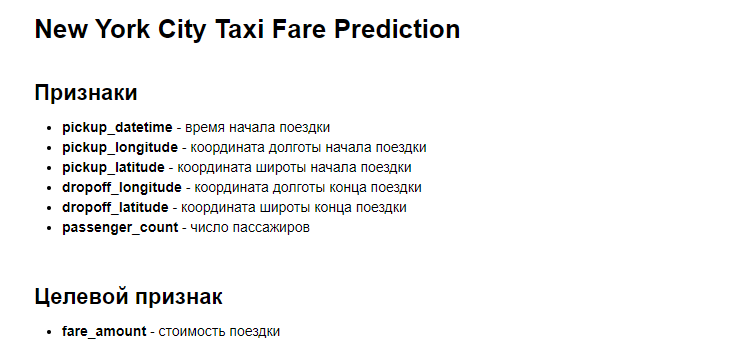
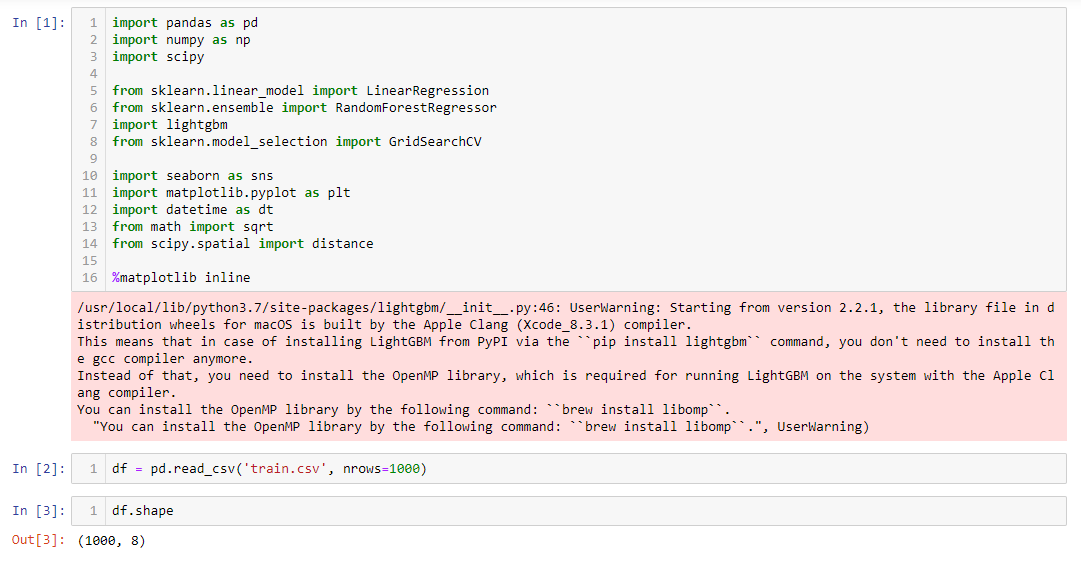
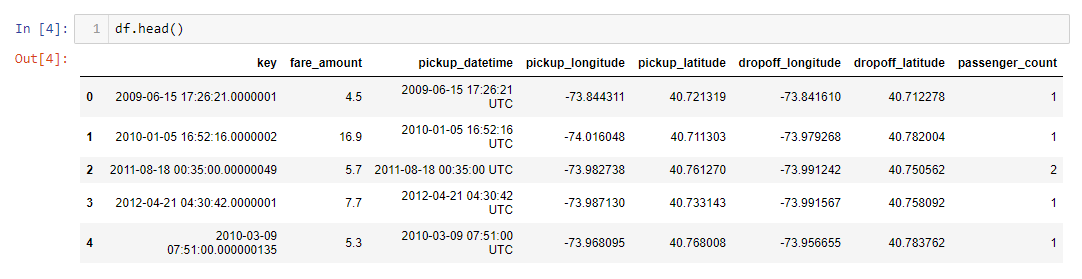
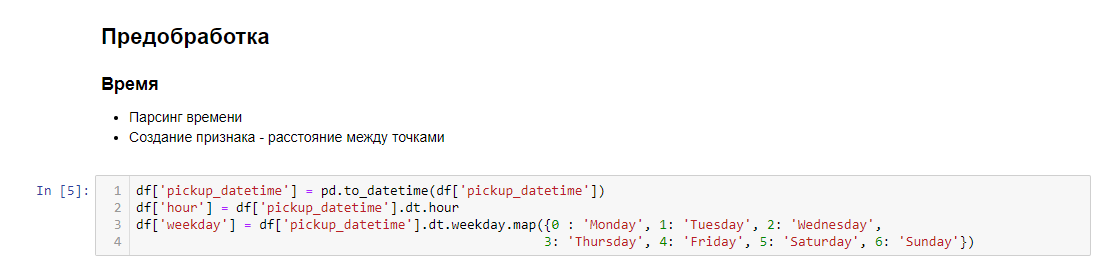
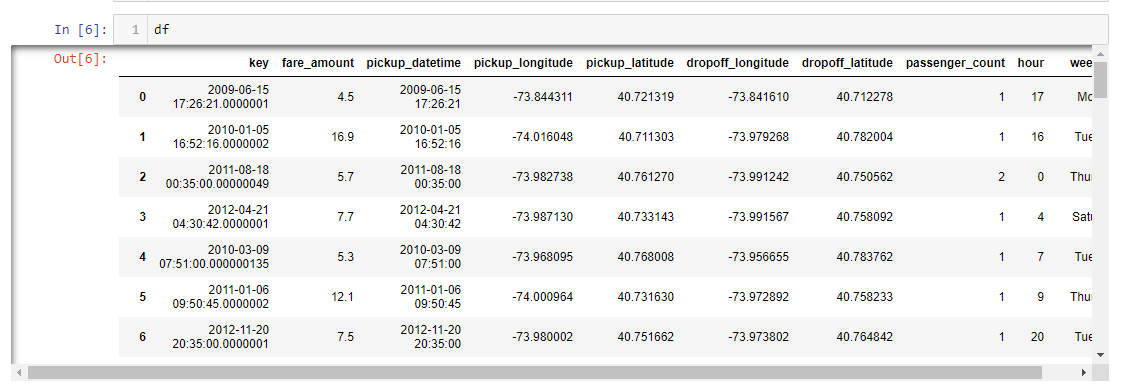
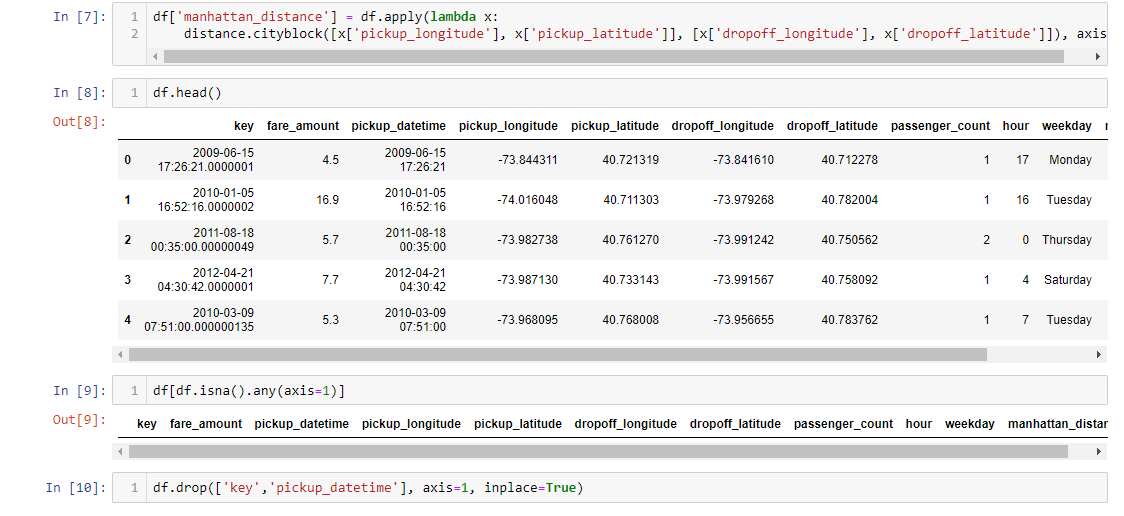
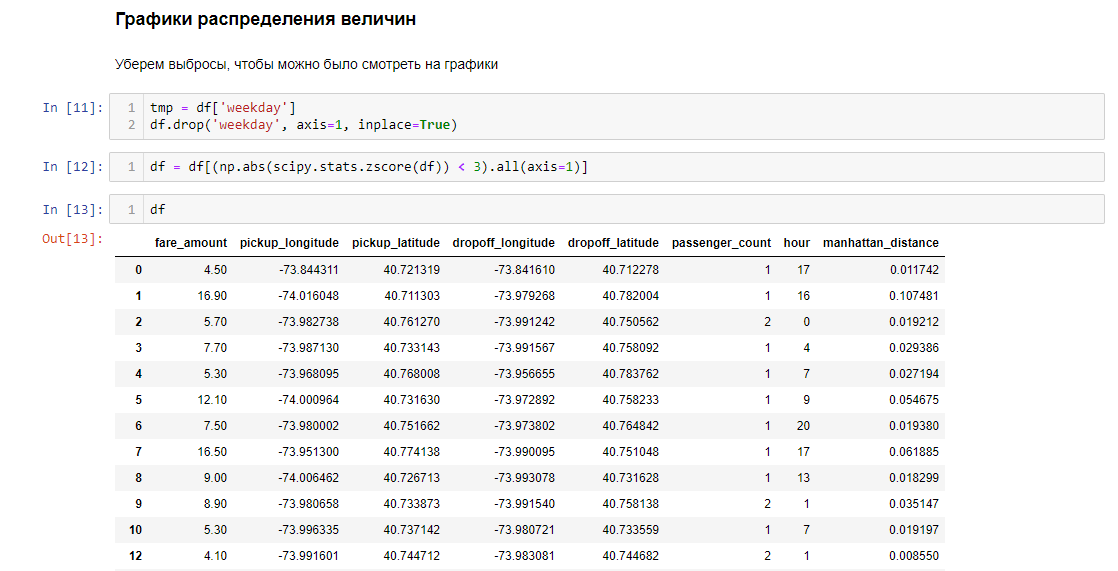
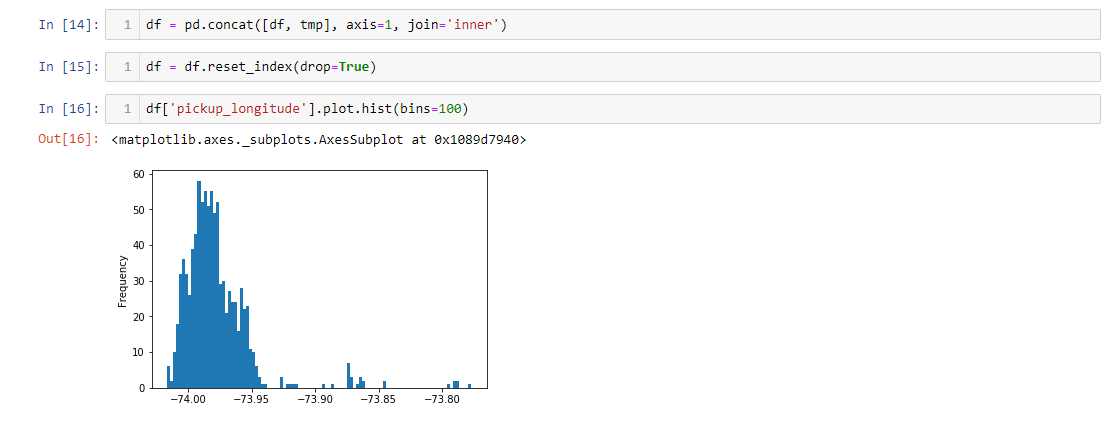
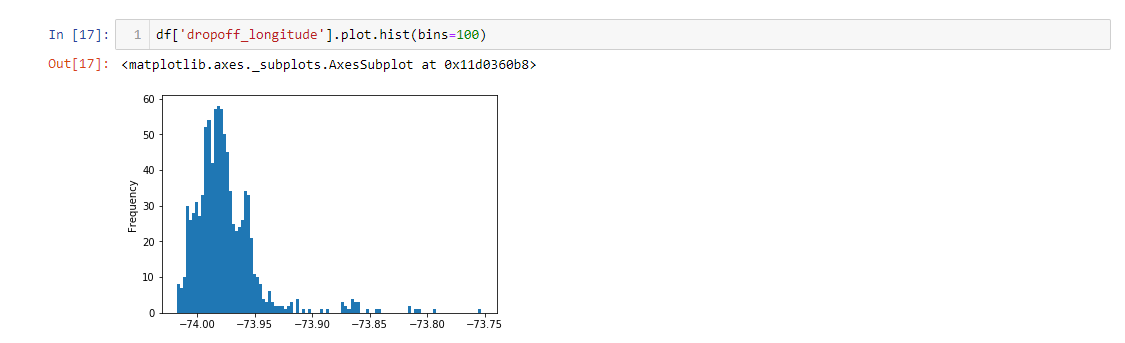
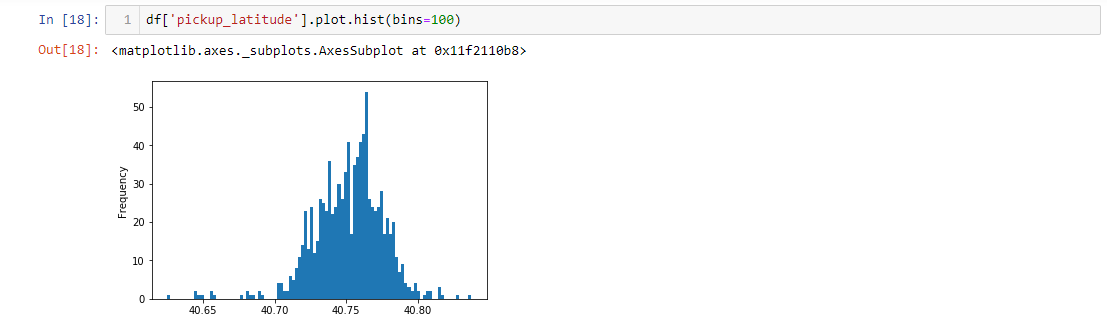
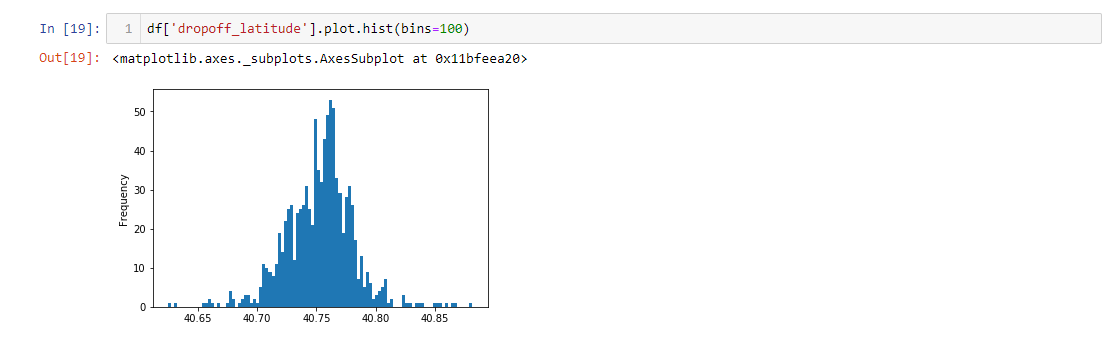
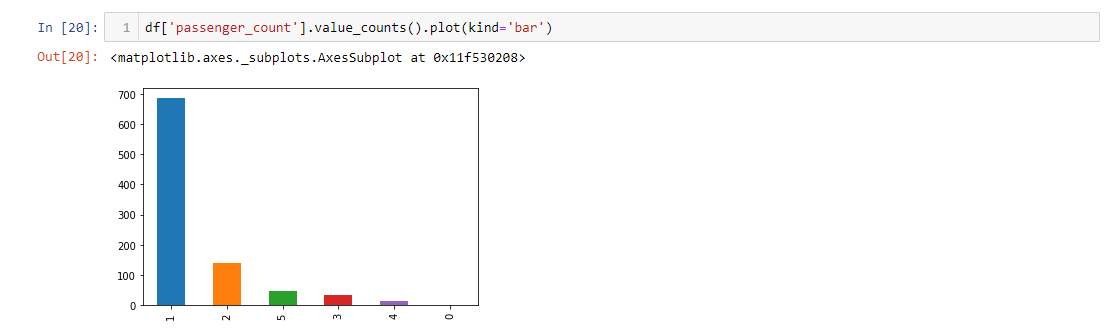
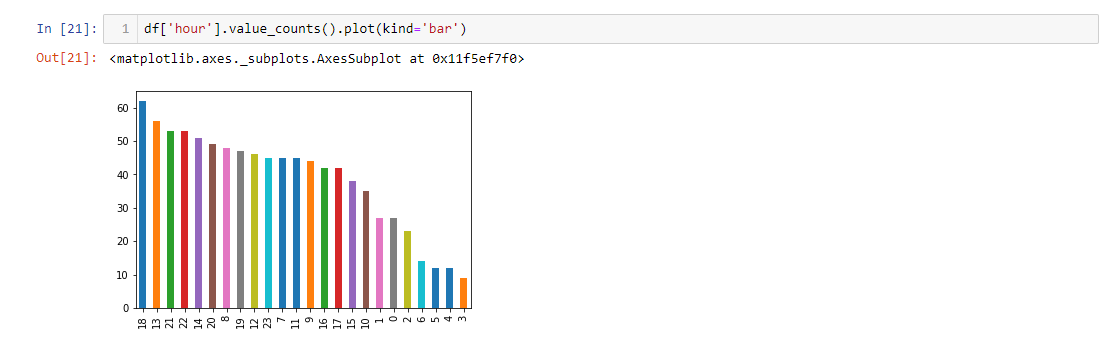
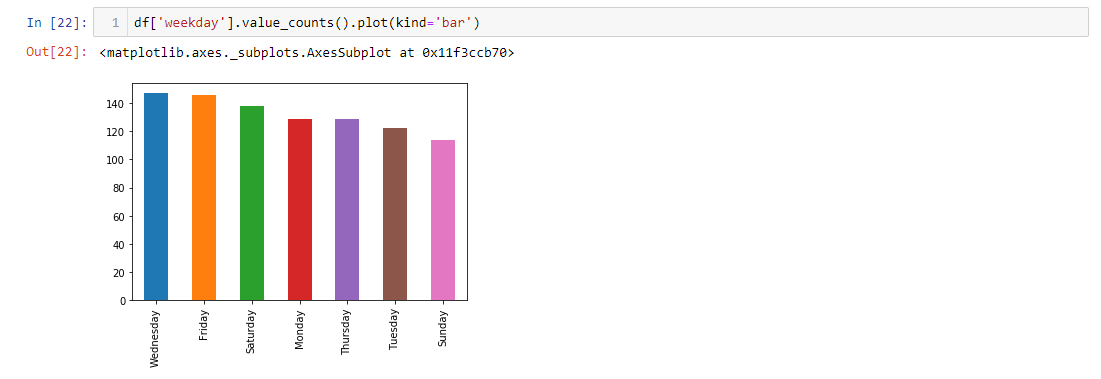
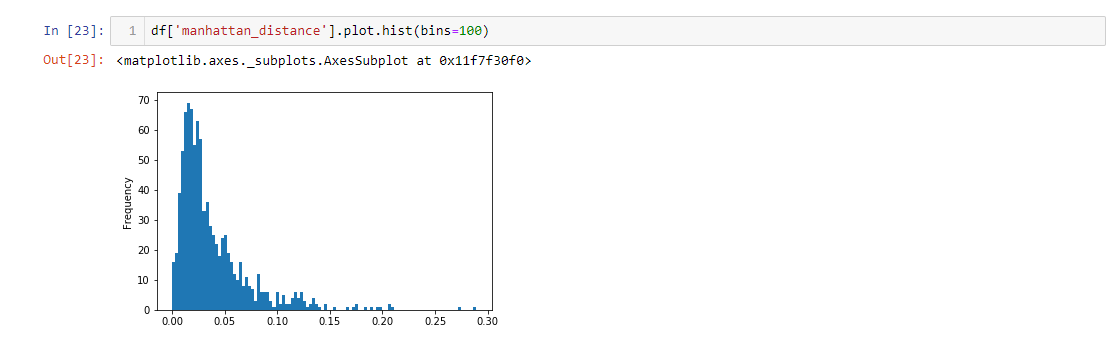
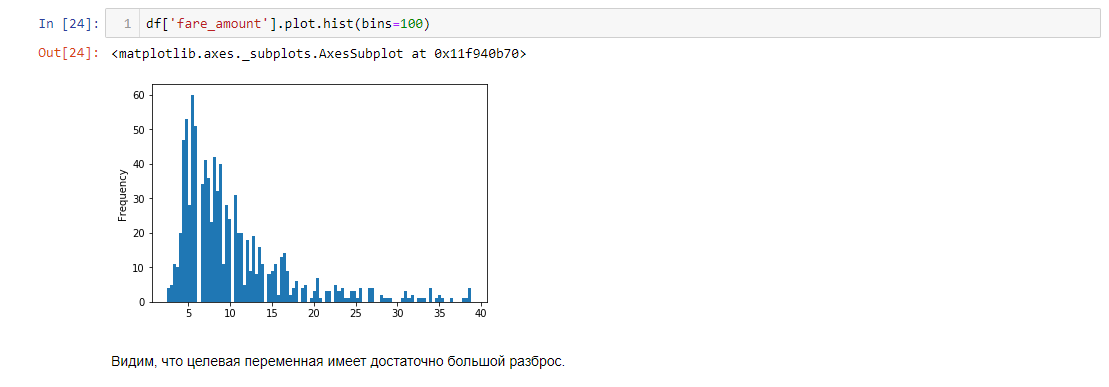
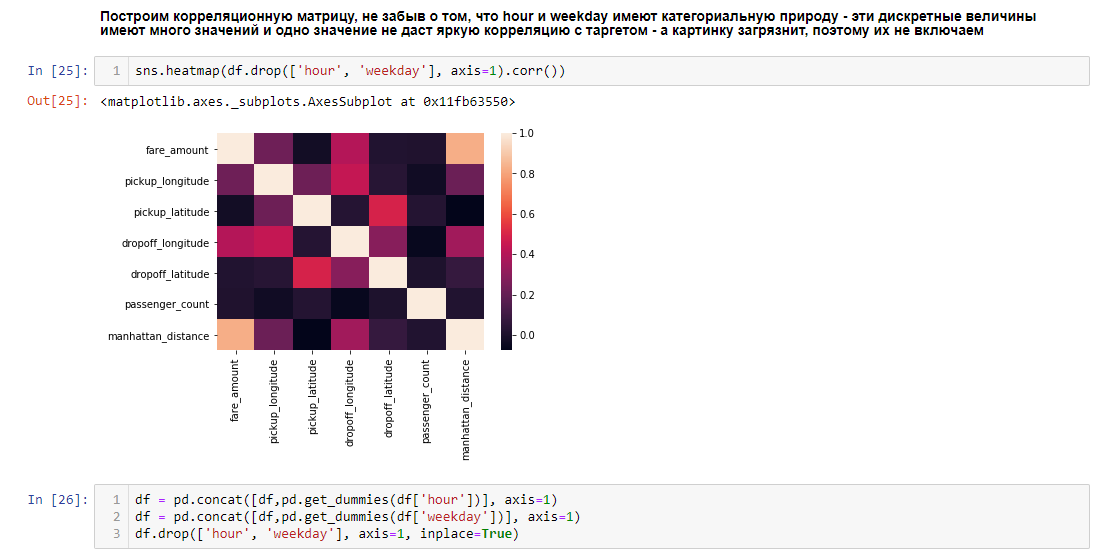
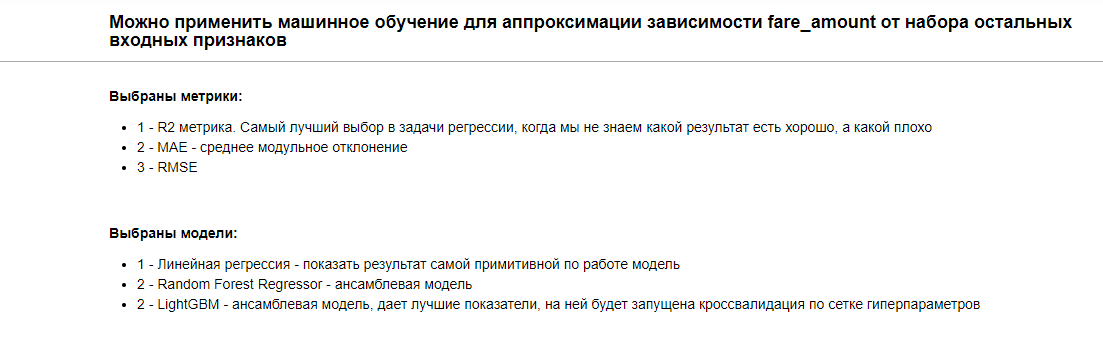
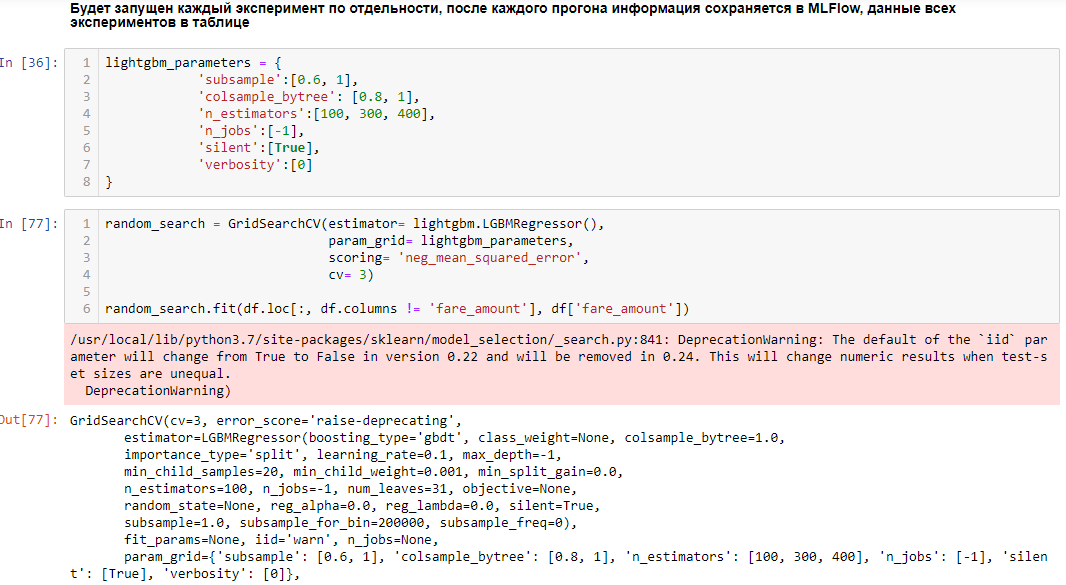
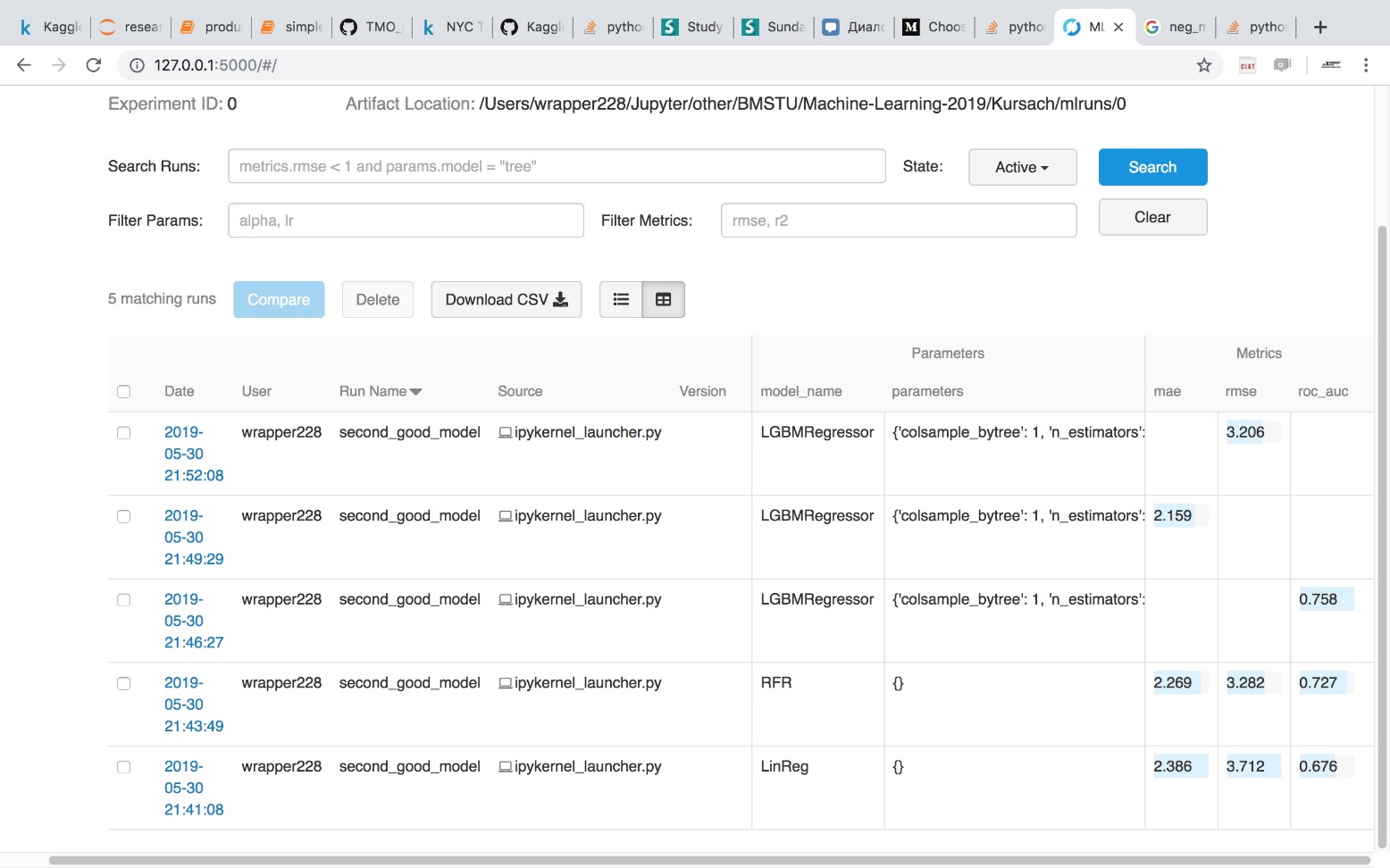
[Список использованной литературы 13](#_Toc10193114)

# 

# Введение

Данная работа предназначена для усвоения и закрепления знаний по дисциплине «Технологии машинного обучения». Здесь закрепляются навыки проведения разведочного анализа данных, выбора признаков для построения модели, проведения корреляционного анализа, подбора метрик, решения задачи регрессии, построения базового решения и подбора гиперпараметров.

# Основная часть

# Заключение

В данной работе были закреплены навыки проведения разведочного анализа данных, выбора признаков для построения модели, проведения корреляционного анализа, подбора метрик, решения задачи регрессии, построения базового решения и подбора гиперпараметров, сравнены качества моделей.

# Список использованной литературы

# <https://github.com/ugapanyuk/ml_course/wiki/COURSE_TMO>

# <https://lightgbm.readthedocs.io/en/latest/>

# <https://www.mlflow.org/>

# <https://scikit-learn.org/stable/modules/generated/sklearn.ensemble.RandomForestRegressor.html>

# <https://scikit-learn.org/stable/modules/generated/sklearn.metrics.mean_squared_error.html>