Федеральное государственное бюджетное   
образовательное учреждение высшего   
профессионального образования

Московский государственный технический   
университет имени Н.Э. Баумана   
(МГТУ им. Н.Э. Баумана)

**РУБЕЖНЫЙ КОНТРОЛЬ 2  
«Технологии использования и оценки моделей машинного обучения»**

**ПО ДИСЦИПЛИНЕ «Технологии машинного обучения»**

**Вариант 12**

Выполнил(а): Орлова С. М.   
 (Фамилия И.О. студента)

РТ5-61   
 (Индекс группы)

Проверил(а): Гапанюк Ю. Е.   
 (Фамилия И.О. преподавателя)

«\_\_\_» \_\_\_\_\_\_\_\_\_\_\_\_ 2020 г.

ЗАЧТЕНО / НЕ ЗАЧТЕНО \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_  
 (нужное выделить) (подпись)

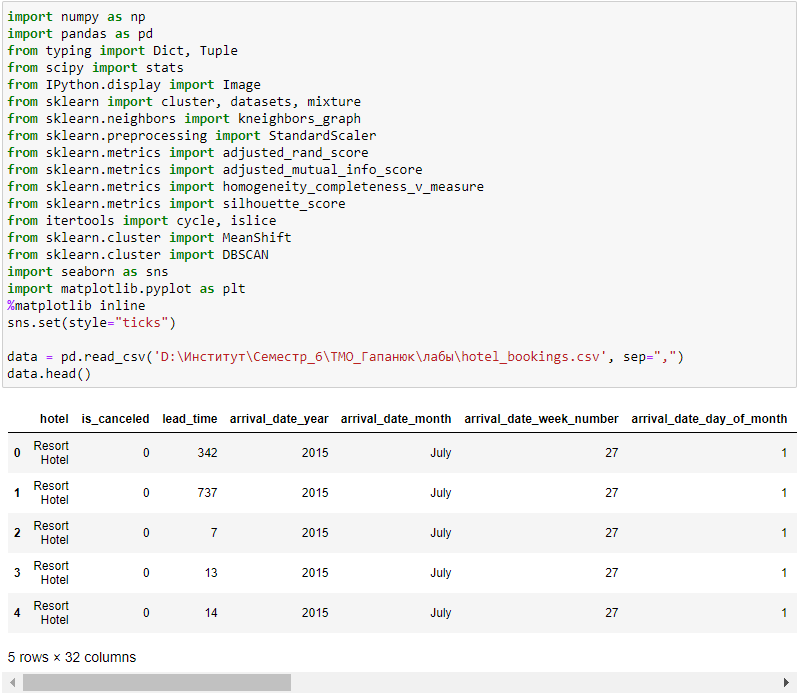
Москва, 2020

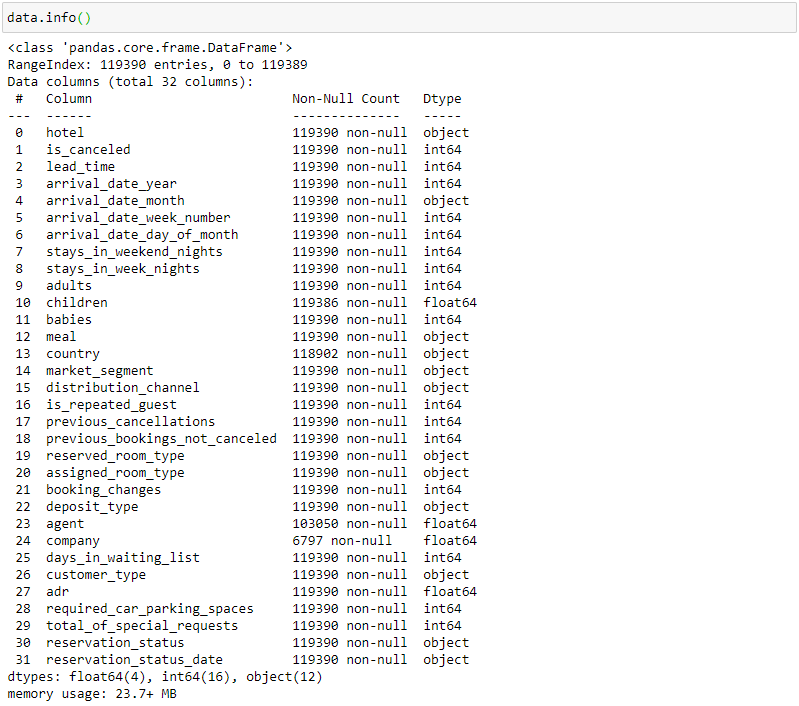
1. Описание задания

Задача №2:  
Кластеризируйте данные (10) с помощью двух алгоритмов кластеризации: MeanShift и DBSCAN. Сравните качество кластеризации с помощью следующих метрик качества кластеризации (если это возможно для Вашего набора данных): Adjusted Rand index; Adjusted Mutual Information; Homogeneity, completeness, V-measure; Коэффициент силуэта. Сделайте выводы о том, какой алгоритм осуществляет более качественную кластеризацию на Вашем наборе данных.

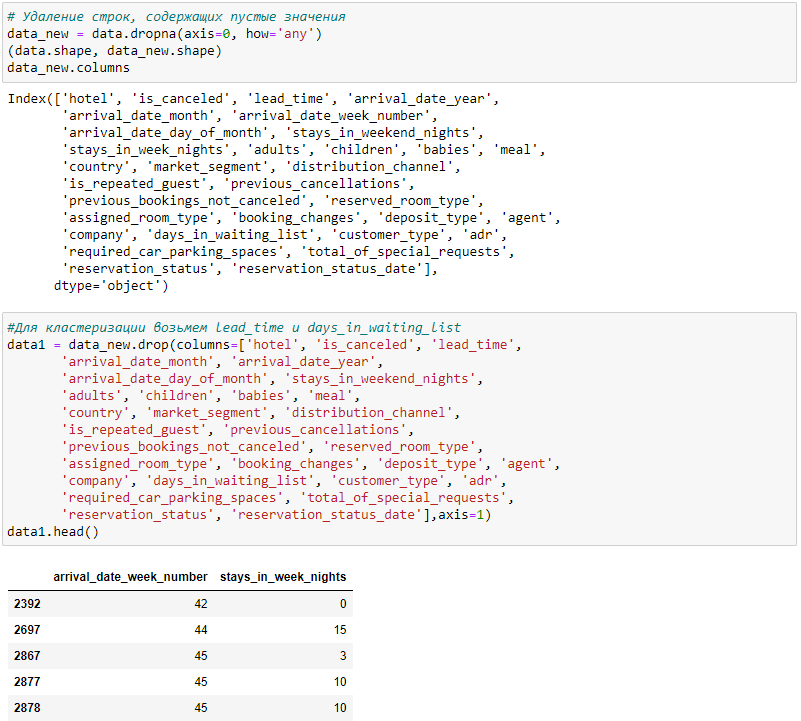
1. Текст программы и экранные формы с примерами выполнения программы

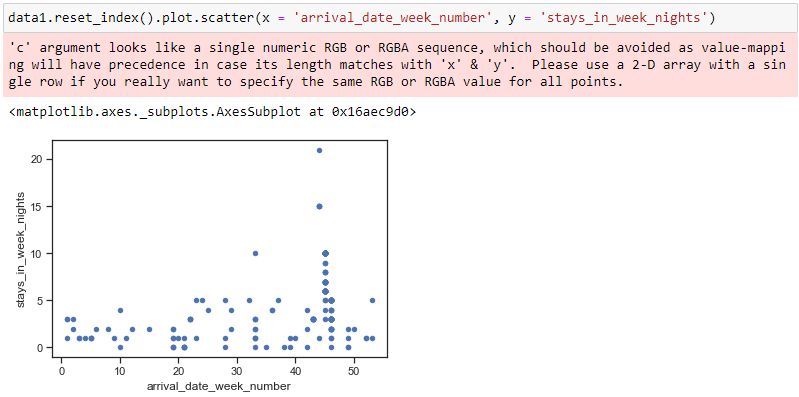
Загрузка и первичный анализ данных:



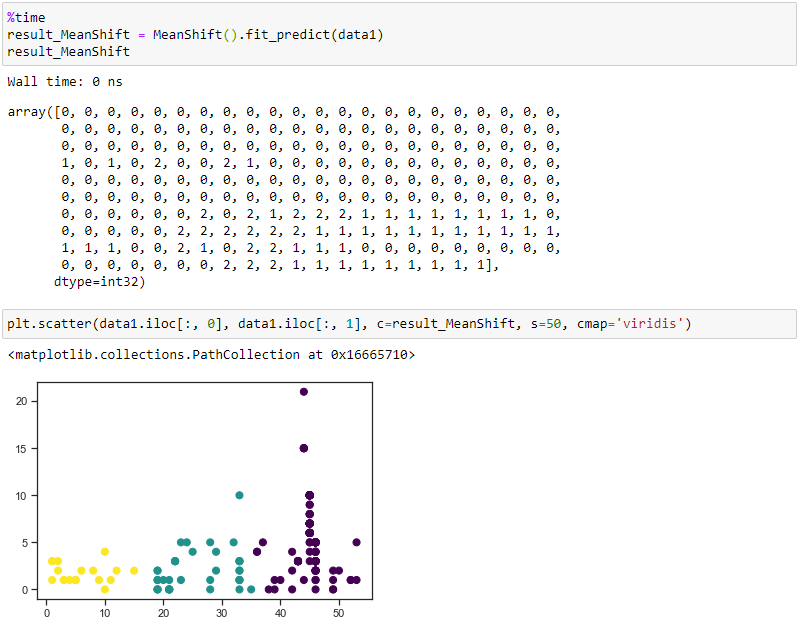


Обработка пропусков в данных:

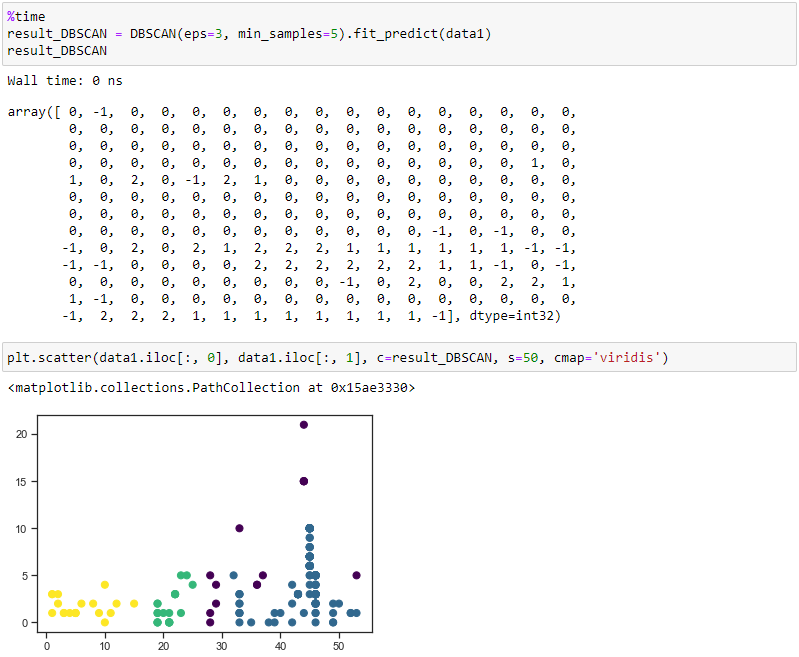




MeanShift:



DBSCAN:



Сравнение качества кластеризации:

Поскольку в наборе данных истинные значения меток кластеров не известны, то для оценки качества моделей используем только метрику Коэффициента силуэта.



Выводы:

В данной работе после сравнения двух метрик можно сделать вывод, что для данных гиперпараметров лучшим будет алгоритм MeanShift.