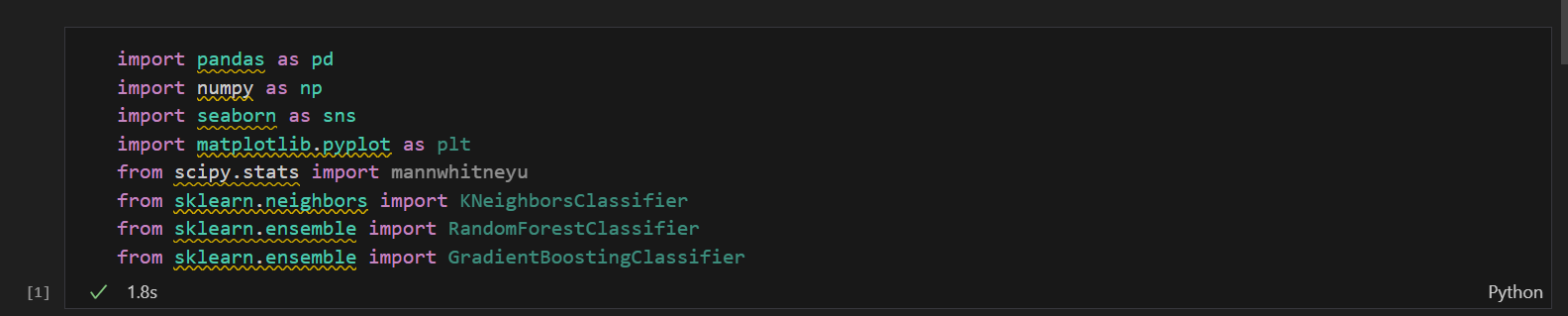
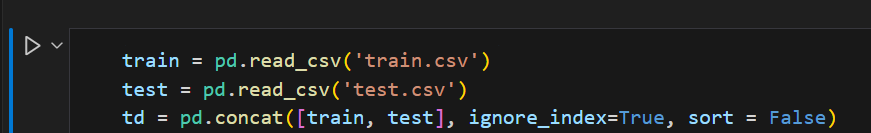
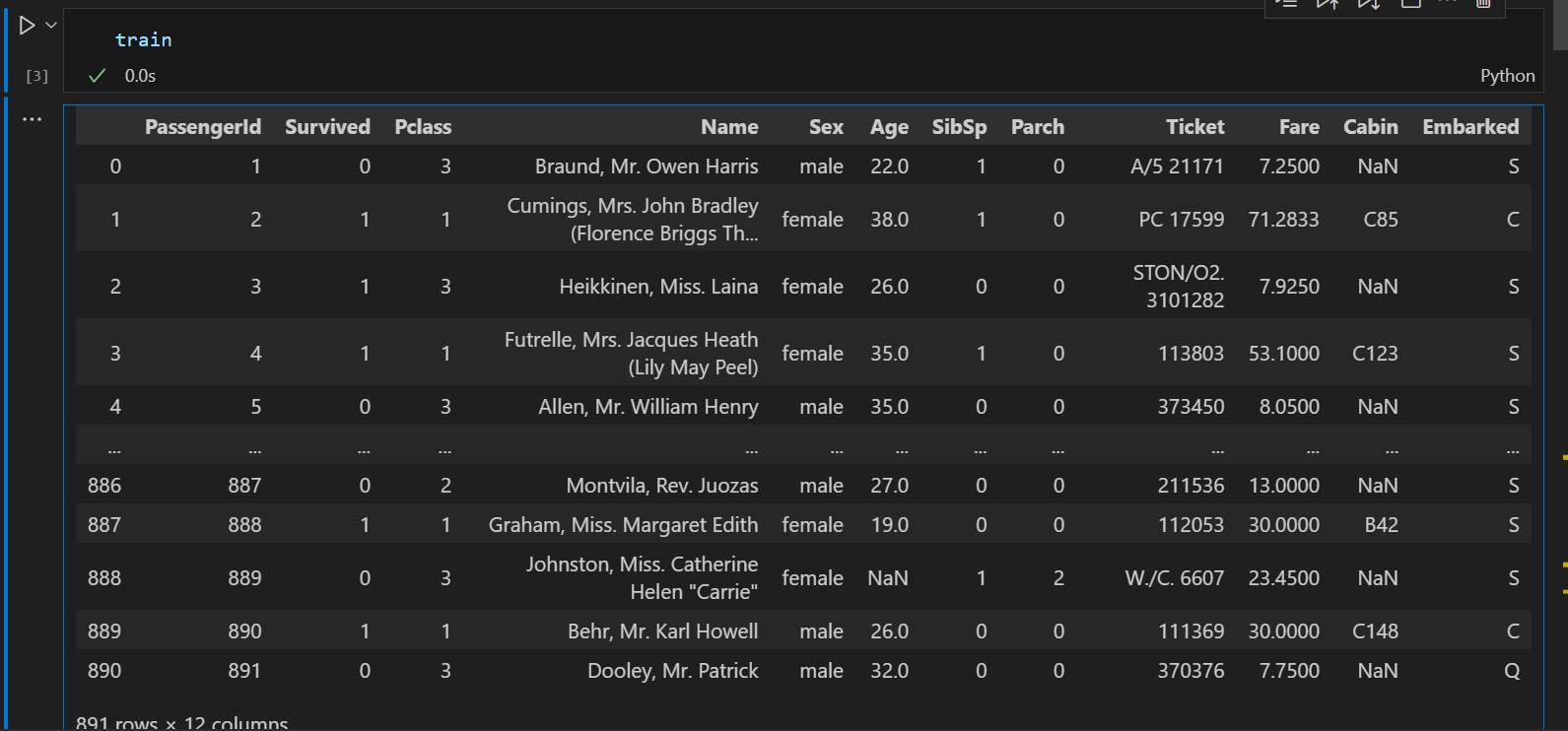
Отчет Стринковского Егора по лабораторной работе 1

1)Добавляем библиотеки



2) Загрузка данных

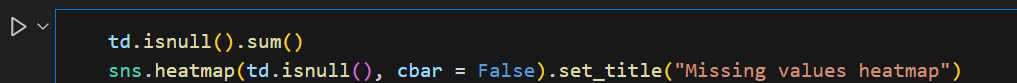
3) Отображаем train

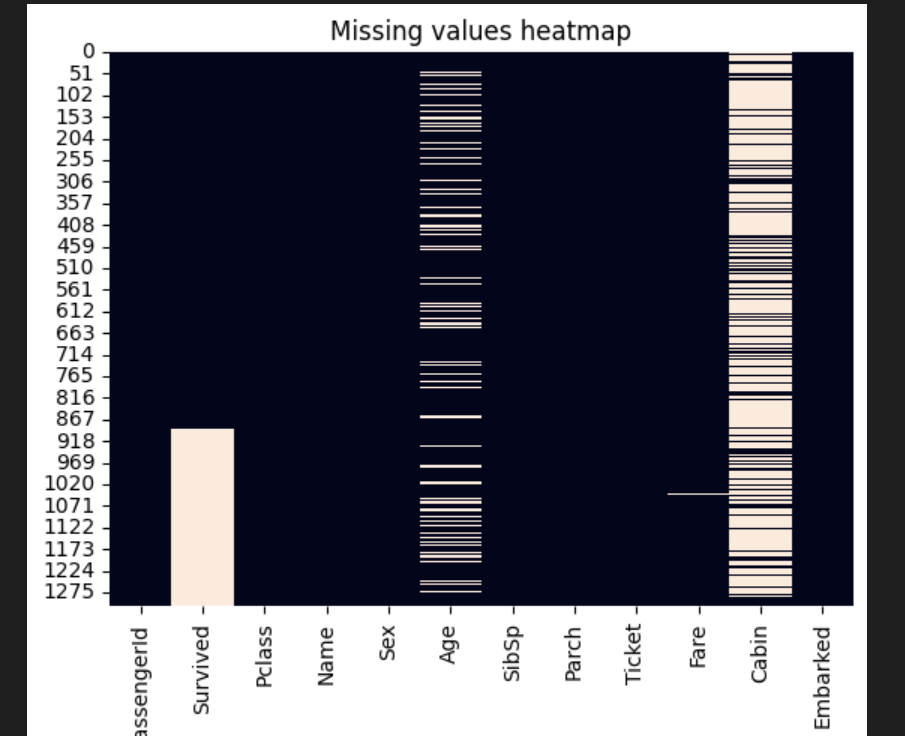


4) Отображаем test

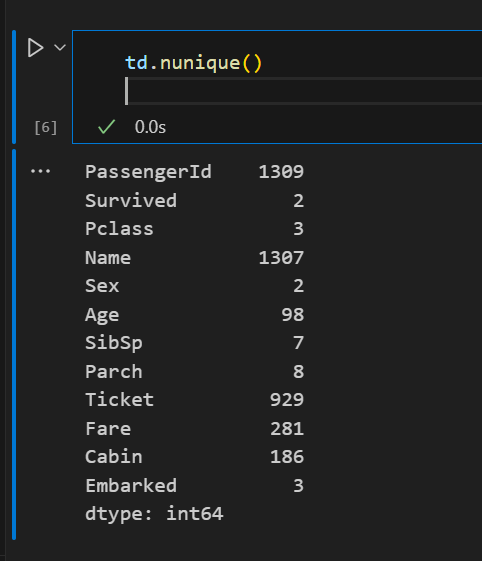


5) Пропущенные значения

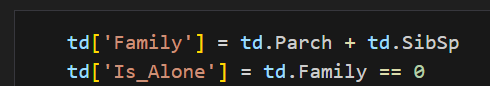


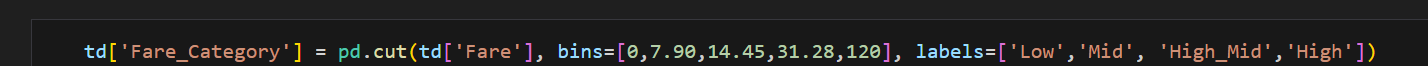


6) Категориальные признаки

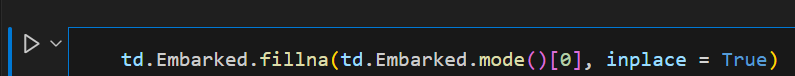


7) Для хранения суммарных значений «Parch» и «SibSp» - столбец «Family». Более того, шансы на выживание взлетели до небес, когда путешественник путешествовал один. Создал другой столбец Is\_Alone и присвоил значение True, если значение в столбце «Family» было 0.

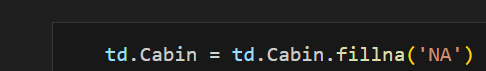


8) Чем больше заплатит пассажир, тем выше будут его шансы на выживание.

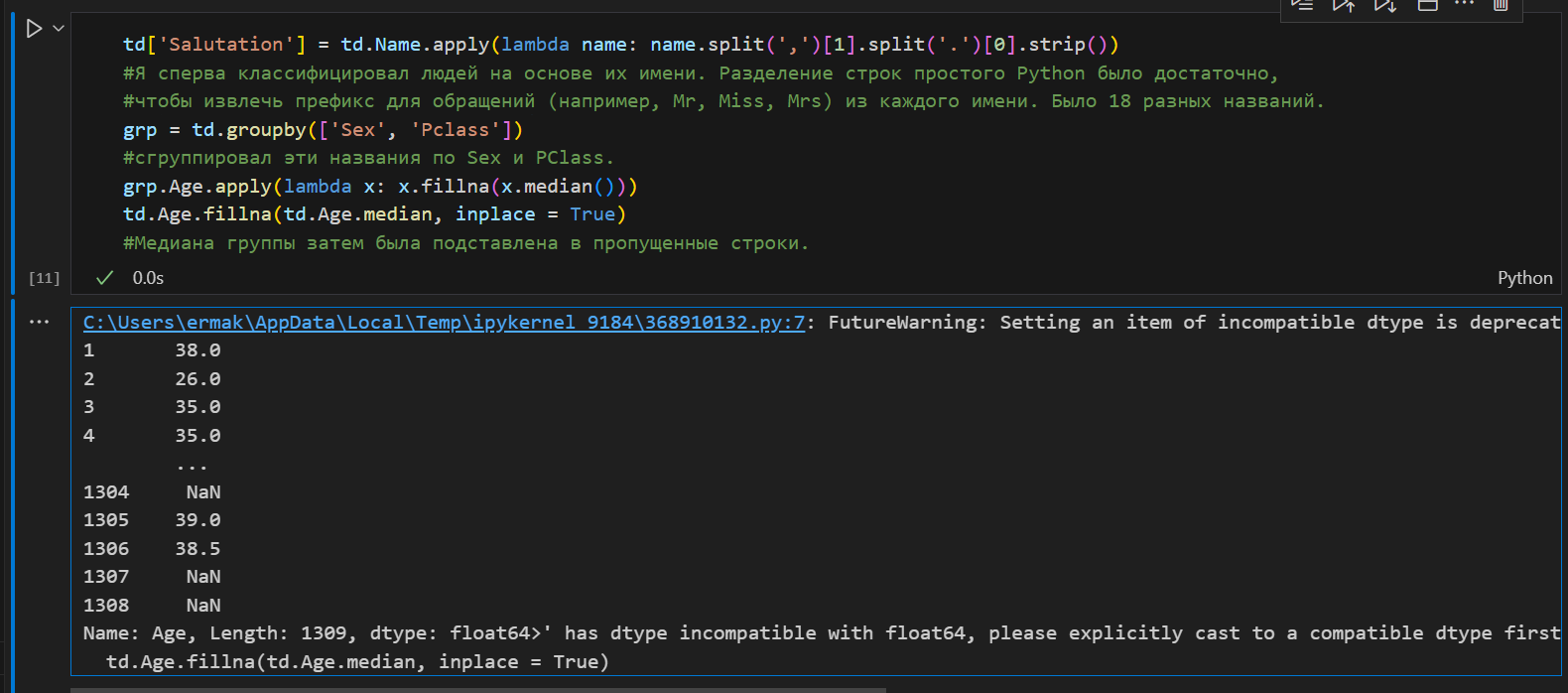
9) Поскольку у «Embarked» было только два пропущенных значения и наибольшее количество пассажиров отправлялось из Саутгемптона, вероятность посадки в Саутгемптоне выше.

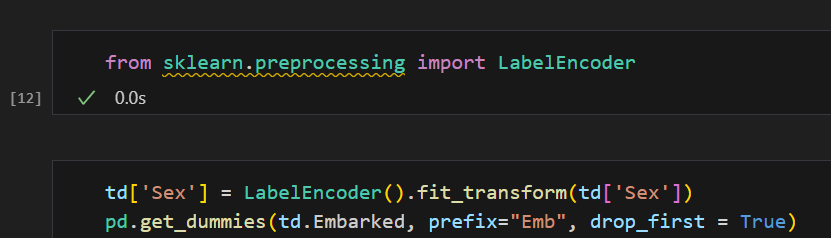


10) В колонке «Cabin» было много пропущенных данных. Я решил определить все отсутствующие данные в отдельный класс. Я назвал его NA и заполнил все пропущенные значения этим значением.



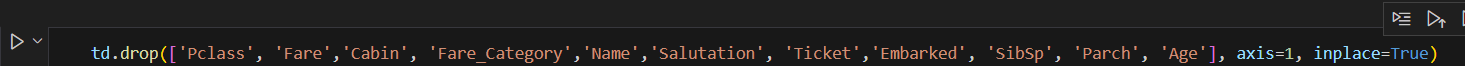
11) Я сперва классифицировал людей на основе их имени. Разделение строк простого Python было достаточно, чтобы извлечь префикс для обращений (например, Mr, Miss, Mrs) из каждого имени. Было 18 разных названий. Сгруппировал эти названия по Sex и PClass. Медиана группы затем была подставлена ​​в пропущенные строки.



12) Добавляем LabelEncoder библиотеку. Поскольку текстовые данные плохо сочетаются с алгоритмами машинного обучения, мне нужно было преобразовать нечисловые данные в числовые. Я использовал LabelEncoder для кодирования столбца «Sex». LabelEncoder будет заменять «мужские» значения одним числом, а «женские» значения — другим числом.



13) Я отбросил столбцы, которые мне не нужны для прогнозирования, и столбцы, которые я кодировал функцией get\_dummies.



14) #Это был случай задачи классификации, и я попытался сделать предсказания с помощью двух алгоритмов Случайный лес и Гауссовский Наивный Байесовский классификатор. 