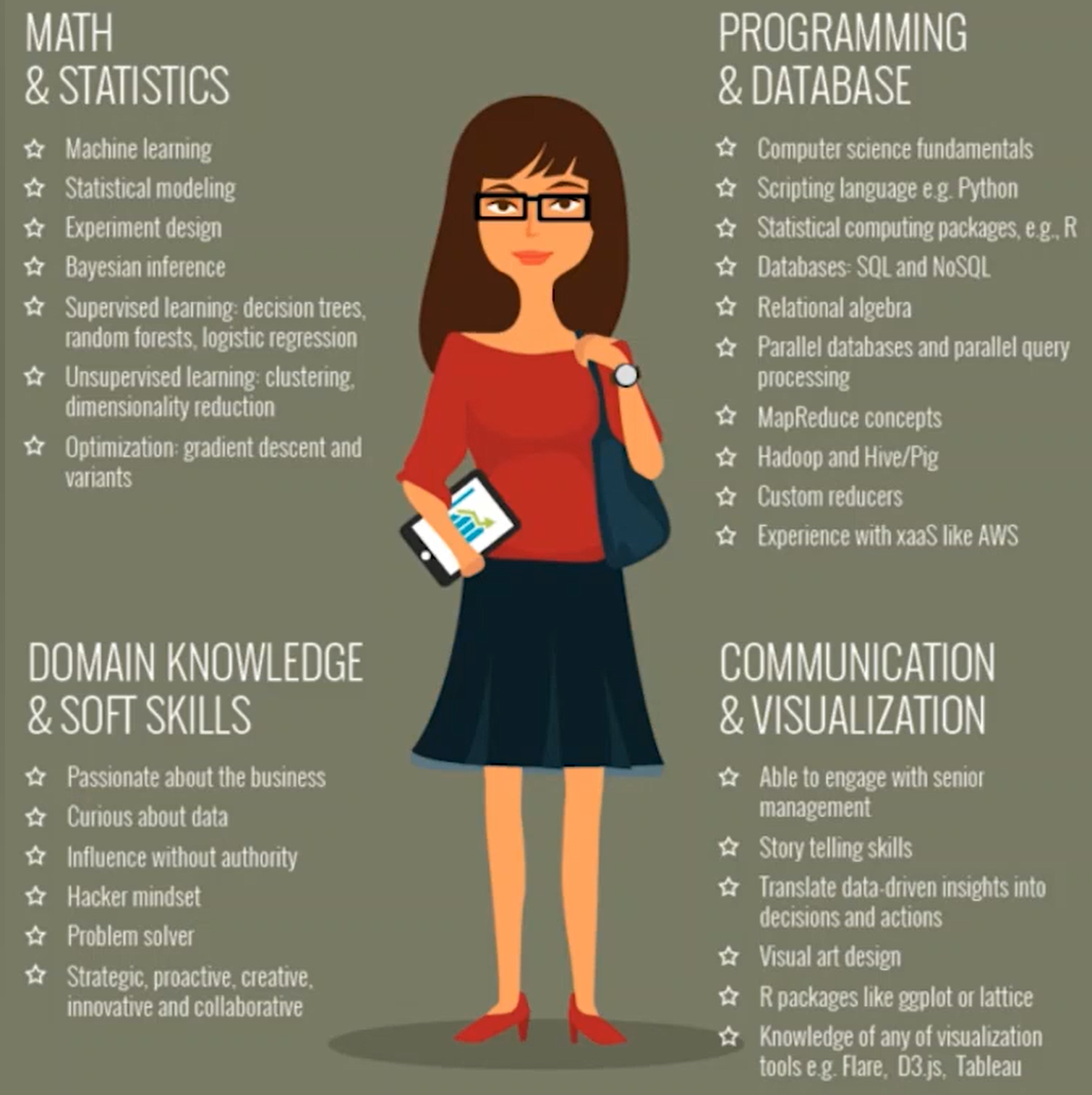


При написании данной дипломной работы я ставил перед собой следующие задачи: как устроены модели машинного обучения, как происходит процесс обучения моделей, зачем это нужно, какие ключевые концепции лежат в основании процессов, связанных с машинным обучением и какие экономические задачи можно решать, используя машинное обучение.

Методы помогают нам обрабатывать и преобразовывать данные. Атрибуты помогают нам увидеть, что хранится в данных.

DataFrame в Pandas – это не просто таблицы, но и множество атрибутов и методов к ним, состоящие из множества Series (векторов-столбцов).

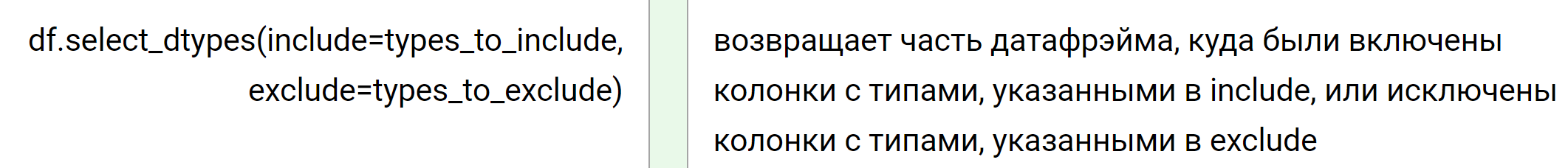


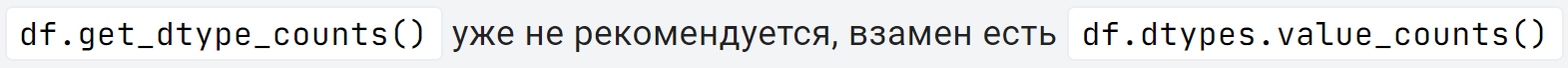
Бизнес-аналитик формируем проблемный вопрос, который нужно решить и собирает требования. Дата инженер собирает и обрабатывает данные. Дата Сайентист анализирует и подбирает модель. Разработчик машинного обучения внедряет модель в программу.

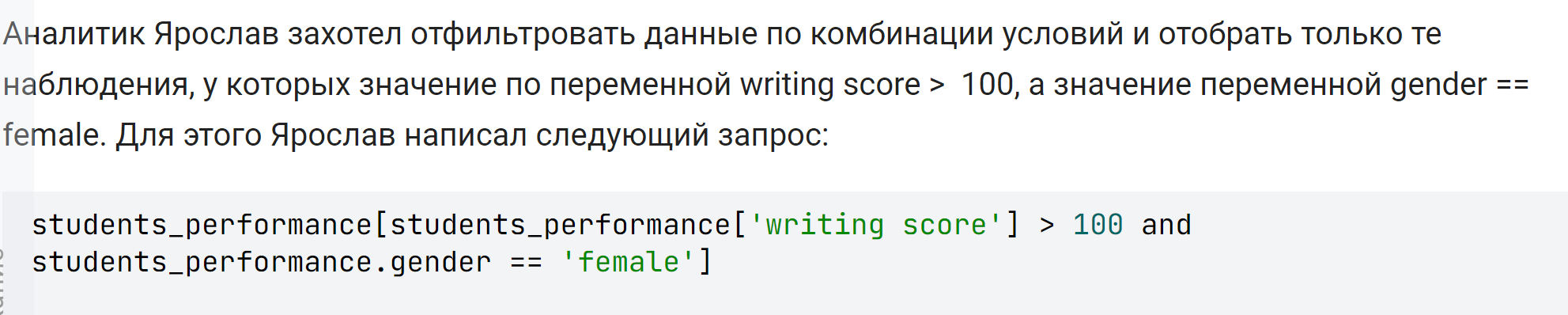
Для начала данные нужно подготовить (чистка пустых данных), формируем признаки (фичи) датафрейма, агрегируем и т.д. Далее с помощью методов машинного обучения обучаем нашу модель и прогнозируем целевую переменную. Затем мы валидируем нашу модель, проверяем, что она предсказывает верно при работе с тестовыми, а не тренировочными данными.

Подбор коэффициентов b0 и b1 – это задача машинного обучения.

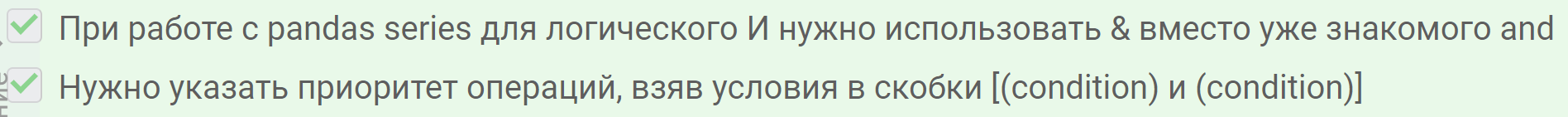
Задача всего обучения – научить модель находить общую ЗАКОНОМЕРНОСТЬ в данных, обучивших на тренировочной выборке, в которой есть некоторая частная закономерность.







<https://stackoverflow.com/questions/36921951/truth-value-of-a-series-is-ambiguous-use-a-empty-a-bool-a-item-a-any-o>



Надо просто помнить, что выражения вида `students\_performance['writing score'] > 100` выдают не логические значения, а булевы вектора; поэтому операции `and` и `or` для них не подходят. Нужна какая-то операция `and\_bool\_vector` (и `or\_bool\_vector`, соответственно), и она уже есть: `&` (и '|')

У какой доли студентов из датасэта в колонке lunch указано free/reduced?

stud\_perf['lunch'].value\_counts(normalize=True)

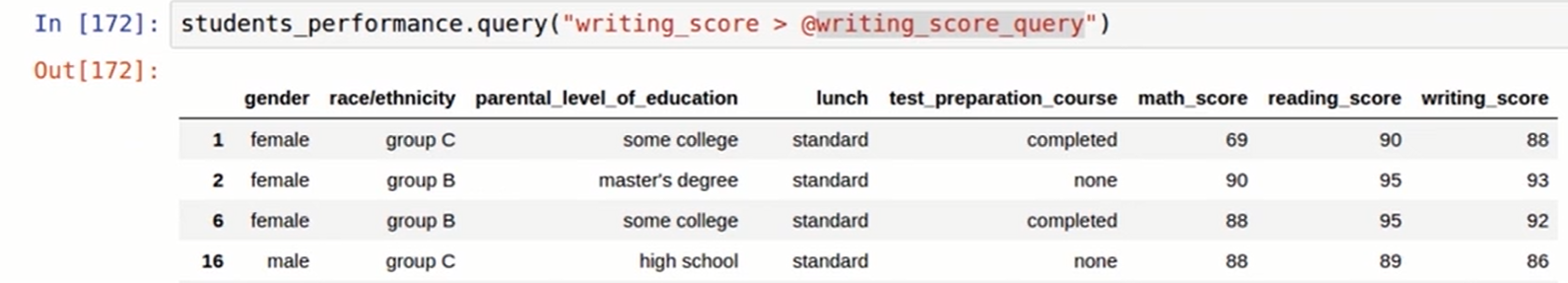
Как различается среднее и дисперсия оценок по предметам у групп студентов со стандартным или урезанным ланчем?

lunch\_std = stud\_perf[stud\_perf['lunch'] == 'standard'].describe()

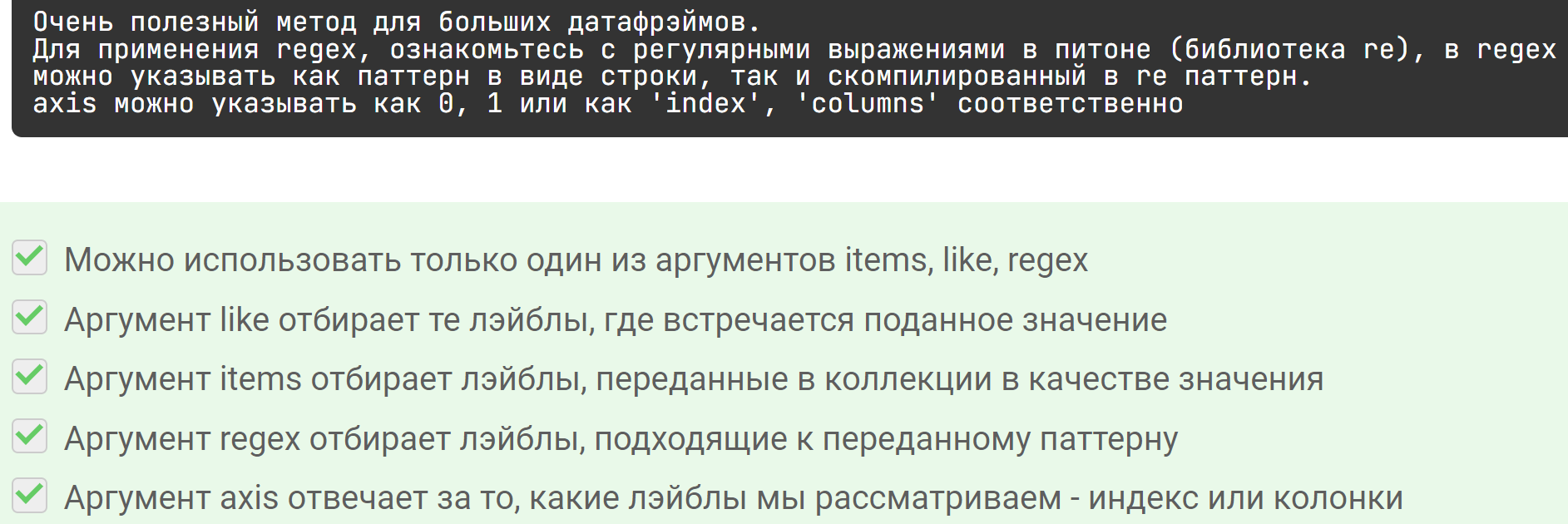
lunch\_free = stud\_perf[stud\_perf['lunch'] == 'free/reduced'].describe()

lunch\_std.compare(lunch\_free)

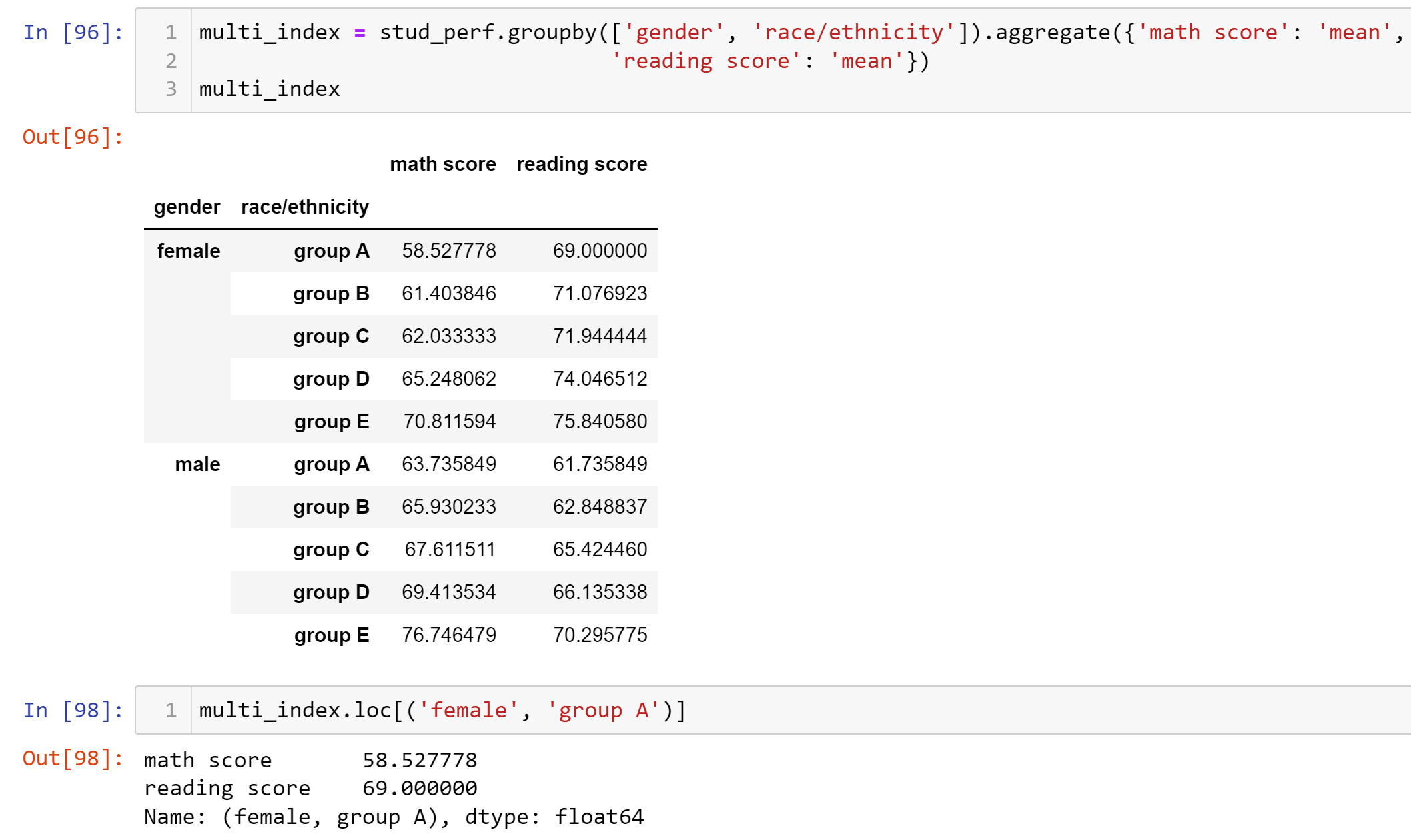
<https://pandas.pydata.org/docs/reference/api/pandas.DataFrame.compare.html>

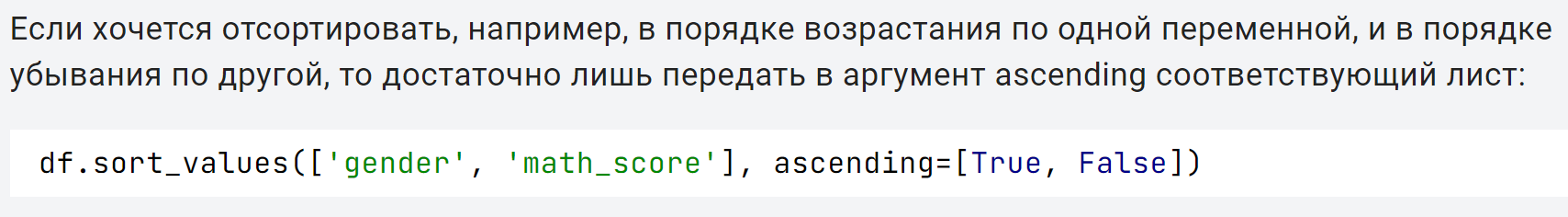


Изучите документацию метода filter и укажите правильные утверждения



**Также небольшой совет по решению задач в целом, если вы понимаете, что не можете решить задачу, отложите ее, посмотрите несколько следующих уроков, порешайте другие задачи и потом вернитесь к проблемной задаче позже.**





К нам поступили данные из бухгалтерии о заработках Лупы и Пупы за разные задачи! Посмотрите у кого из них больше средний заработок в различных категориях (колонка Type) и заполните таблицу, указывая исполнителя с большим заработком в каждой из категорий.

<https://stepik.org/media/attachments/course/4852/accountancy.csv>

