Кейс от ООО «Газпромнефть НТЦ»

для студенческого чемпионата ГПН по data science

Со дня учреждения гильдии искателей приключений члены гильдии взяли бесчисленное множество поручений, для выполнения которых им приходилось сражаться с кровожадными монстрами, исследовать древние руины, искать сокровища, добывать редкие минералы, спасать нерадивых путешественников и делать многие другие вещи. Любой житель волшебного государства мог, предложив разумное вознаграждение, разместить поручение в Гильдии. Многие герои в ходе выполнения поручений попробовали себя в новых для себя ролях, использовали уникальные виды оружия. Гильдия всегда ответственно подходила к работе со своими клиентами, собирая отзывы о выполненных заказах: насколько заказчик удовлетворён сроками, в которые герой смог выполнить поручение, качеством работы героя и даже вежливостью героя.



Поток клиентов не иссякал, и со временем поручений стало так много, что герои перестали вдумчиво выбирать, за какое из них взяться, а просто брали первый попавшийся листок с заданием и шли выполнять его. Это привело к тому, что стали случаться первые провалы миссий: некоторые герои не могли справиться со слишком могучим монстром, другие не были достаточно опытными следопытами, чтобы найти пропавшего путешественника. Появились негативные отзывы от клиентов, некоторые из них даже отказывались платить!

Вскоре у членов гильдии появилась идея использовать волшебный древесный компьютер лесной академии – Аркашу для определения того, кто из героев наилучшим образом подойдёт для выполнения каждой из миссий. Кроме того, было решено, что для выполнения особенно сложных поручений можно привлекать не одного, а сразу нескольких героев. Например, кто-то из членов команды будет отвечать за поиск убежища монстра и отвлечение его внимания, а остальные будут заниматься нанесением урона и спасением похищенных монстром жителей. Для реализации этого амбициозного плана в Аркашу были загружены все когда-либо выданные поручения, а также послужной список каждого члена гильдии с отзывами от клиентов.



Помогите учёным лесной академии обучить Аркашу выдавать правильные рекомендации: научите его на основе истории выполнения поручений и владения героя теми или иными навыками формировать команды для выполнения поручений.

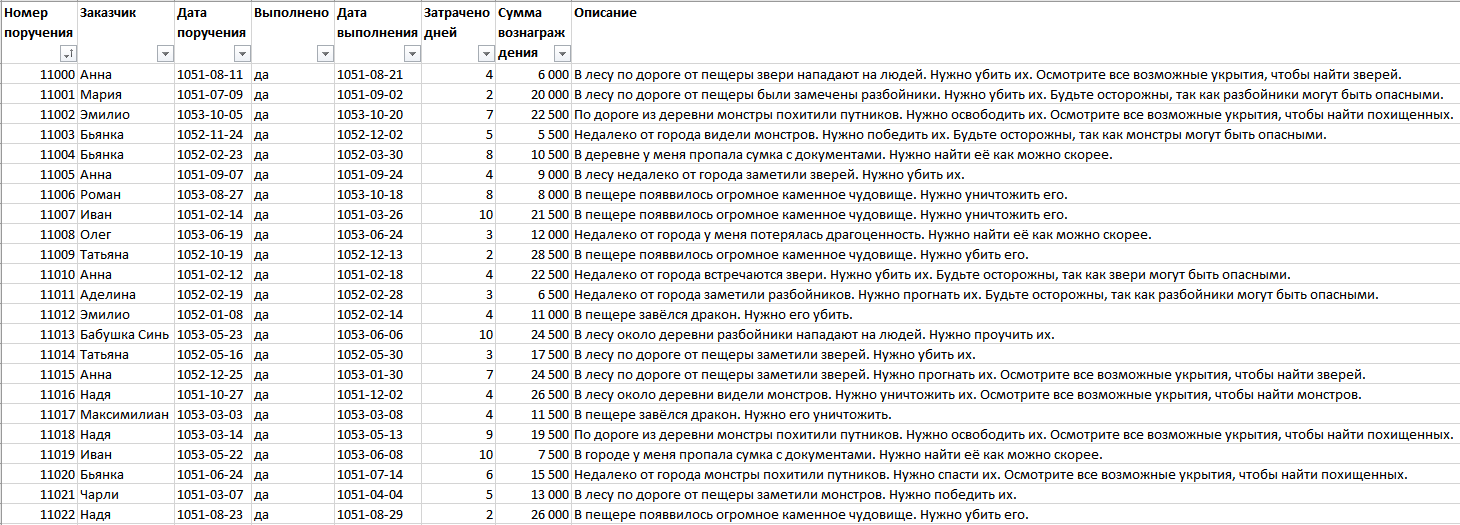
Поручения написаны обычными жителями, которые не слишком разбираются в монстрах и пещерах, поэтому описания похожих поручений могут очень сильно отличаться друг от друга: где-то каменный дракон назван ходящей горой, а кто-то назвал колодец, в котором видел кикимору, дыркой в земле. Возможно, для более качественного анализа Аркашу нужно научить объединять синонимичные формулировки.

В ходе выполнения поручений все члены гильдии в обязательном порядке ведут дневник, где записывают, какие действия им пришлось совершить для выполнения поручений, сколько времени это заняло и насколько сложно для них это было.

В команде, которую Аркаша может посоветовать для выполнения поручения, должно быть от одного до четверых героев. При этом задача Аркаши – обеспечить максимальную суммарную выручку гильдии. Если на каждое поручение отправлять по четыре героя, то героев на все поручения точно не хватит. С другой стороны, если команды будут слишком долго выполнять поручения, или вообще не справятся с ними, то приток денег в гильдию также сократится.

# Используемые данные:

1. Список всех когда-либо выданных гильдии поручений – как выполненных, так и новых. В списке для каждого поручения приведено его описание, размер денежного вознаграждения за выполнение, затраченное на выполнение время.



1. Дневники всех героев. В каждой записи указано, к какому из поручений она относится, а также в какой роли герой выполнял соответствующее действие, и сколько времени заняло выполнение.



1. Оценки, выставленные заказчиками каждому герою за каждое выполненное поручение по шкале от 2 до 5



# Формат решения:

Ссылка на GitHub, содержащая:

1. Файл Jupyter notebook с реализацией решения
2. Текстовое описание процесса выполнения задания с обязательным указанием принятых допущений (можно отдельным файлом, но лучше – как часть Jupyter notebook, чтобы текстовые описания шли перед каждым крупным блоком кода)
3. Таблица с командами героев, рекомендуемых для каждого из не выполненных поручений (отдельным файлом)
4. Любые другие таблицы, которые участник посчитает нужным предоставить (например, профили навыков героев)
5. Сравнение рекомендуемых Аркашей команд с несколькими вариантами случайно распределённых команд по полученному суммарному денежному вознаграждению и затраченному времени (отдельным файлом или как часть Jupyter notebook)

# Что мы будем искать, при проверке решений:

1. внимательность к данным
2. методичность и последовательность выводов
3. подходы к решению задачи и их реализация
4. аккуратность и читаемость кода
5. информативность отчёта