1) **Аффективные_расстройства**. Много пропусков: так как столбец подразумевает наличие или отсутствие расстройства, которое достаточно редко встречается, то пропуски можно интерпретировать как "отсутствие". От матери психические расстройства передаются сильнее, чем от отца, будем это учитывать. Фичу можно закодировать исходя из того, насколько сильно это может повлиять на человека (которого опрашивают), таким образом: отсутствие, отец, дед по линии матери, бабка по линии отца - 0; бабка по линии матери, дядя/тётя по линии матери - 1; мать - 2.
2) **ПсихПоздВозр**. Также много пропусков: так как столбец подразумевает наличие или отсутствие расстройства, то пропуски можно интерпретировать как "отсутствие". Аналогично, как с **Аффективными_расстройствами**: отсутствие, дед по линии матери, бабка по линии отца - 0; бабка по линии матери - 1.
3) **НеврСтресс**. Аналогично, как с **ПсихПоздВозр**: отсутствие, отец - 0; дядя/тётя по линии матери, бабка по линии матери - 1; мать, сиблинги - 2. Сиблинги стоят на уровне матери, так как если брат/сестра имеют расстройства, то и у другого брата/сестры они могут быть с высокой вероятностью.
4) **УО**. Есть только 1 непустое значение - сиблинги. Этот столбец в принципе можно исключить, так как он является практически константным, но если его нужно оставить, то отсутствие значения - это 0, а сиблинги - 1.
5) **Эпилепсия**. Аналогично как, с **ПсихПоздВозр** и **НеврСтресс**: отсутствие, отец, Дядя/Тетя по линии отца, дед по линии матери - 0; дядя/тётя по линии матери - 1; мать, сиблинги - 2.
6) **ОргПорЦНС**. Для начала, необходимо в столбец разбить значения по запятой, чтобы для людей с несколькими проблемными родственниками значение было выше. Дальше аналогично, как с **Эпилепсией**: отсутствие, отец - 0; бабка по линии матери, дядя/тётя по линии матери - 1; мать, сиблинги, дети - 2. Дети это 2, если рассматриваемый человек - женщина, иначе - 0.
7) **Химическая зависимость**. Всё аналогично, как с **ОргПорЦНС**: отсутствие, отец, дед по линии матери, дед по линии отца, бабка по линии отца - 0; бабка по линии матери, дядя/тётя по линии матери - 1; мать, сиблинги - 2.
8) **БерРодыМать**. Это фича на 99% имеет значения, поэтому по поводу отсутствующих значений можно не париться (их можно считать просто за

0). А вот по поводу имеющихся значений всё не так однозначно. Во-первых, как в случае с **ОргПорЦНС**, необходимо разбить значения в столбце по запятой, чтобы несколько проблем суммировались по значению. После разделения на значения было решено закодировать их по степени возможного вредоносного влияния на ребёнка, то есть на рассматриваемого человека, исходя из имеющихся знаний и найденных фактов в Интернете (не факт, что так и есть): не оценивались (в т.ч. не известно), без патологий - 0; токсикоз первой половины беременности - 1; токсикоз второй половины беременности - 2; психотравмирующие события, инфекционные заболевания - 3; патология беременности, другие тяжелые заболевания - 4; невынашивание беременности в анамнезе - 5. Для одного человека может быть несколько проблем, в таком случае результатом будет сумма значений этих проблем.
9) **ПатРодов** и **ПатРодовДруг**. Сначала рассмотрим эти 2 столбца по отдельности. В **ПатРодов** практически нет пропусков, а вот в **ПатРодовДруг** - их более 95%. Так как, опять же, эти столбцы подразумевают наличие/отсутствие достаточно редких проблем при родах, то отсутствие можно преобразовать как 0. Также значения как “нет” и “нет данных” будем считать также 0. Другие же патологии надо преобразовать в значения 1, 2, 3 и так далее исходя из их тяжести, однако конкретно это может сделать только специалист в этой области. После преобразования этих 2 столбцов необходимо их объединить в 1, просуммировав значения в соответствующих строках, так как оба эти столбца описывают одну и ту же проблему.
10) **РанРазы**. Столбец имеет только 3 значения: не оценивалось, в соответствии с возрастом и наличие отклонений. Также есть 15-20% пропусков. В контексте этого столбца, его значения можно преобразовать по “нормальности” развития следующим образом: наличие отклонений - это -1; не оценивалось, пропуск в данных - 0; в соответствии с возрастом - 1.
11) **Ут2**. Очень необычный столбец, так как в нем практически все значения уникальные, и более 90% - пропуски. В таком случае пропуски надо заменить, а для других изменений 100% необходимо участие специалиста, который сможет оценить по некой шкале эти отклонения.
12) **НевротРасстрДетство**. Много пропусков в данных, опять же, которые можно заменить на 0, так как расстройства проявляются далеко не у всех. Далее необходимо разделить по запятой разные расстройства для 1

человека, а потом, опять же, с помощью специалиста преобразовать расстройства по некой шкале, при этом просуммировав значения нескольких расстройств в одном столбце.
13) **НевротРасстрШкола** и **НевротРасстрШколаДоп**. Идея преобразования схожа со столбцами **ПатРодов** и **ПатРодовДруг**. Рассмотрим эти 2 столбца сначала по отдельности. В столбце **НевротРасстрШкола** вместо пропуска стоит “не оценивалось”, что можно интерпретировать как 0, так как в оставшихся данных более 70% - это “в соответствии с возрастом” и можно предположить, что и у тех, кого не оценивали, с большой вероятностью тоже будет всё хорошо. Далее аналогично, как с **НевротРасстрДетство**, необходим специалист для ранжирования оставшихся расстройств от 1 до N. В столбце **НевротРасстрШколаДоп** все значения являются некоторым набором расстройств (иногда только одним), каждый из которых может оценить только специалист. В итоге после преобразования этих 2-ух столбцов их можно объединить в 1, просуммировав значения в соответствующих строках, так как оба эти столбца описывают одну и ту же проблему.
14) **Симптомы_е3**. Непростой столбец, так как непонятно, что означает приписка “статус” в некоторых случаях. После поиска в Интернете однозначного ответа не нашлось, однако есть предположение, что это может означать “в данный момент”, однако непонятно в таком случае, что означает отдельно “статус”. По-хорошему, надо обратиться к специалисту в таком случае. Если же попробовать использовать свои силы, то в таком случае можно сделать следующее: статус, статус до лечения, пропуск - 0; Анамнез, “Анамнез, Статус до лечения” - 1; “Анамнез, Статус до лечения, Статус после лечения” - 2.
15) **Симптомы_з2**. Аналогично как с **Симптомы_е3**: статус, статус до лечения, пропуск - 0; Анамнез, “Анамнез, Статус до лечения” - 1; “Анамнез, Статус до лечения, Статус после лечения” - 2.