Необходимо построить модель прогнозирования шанса выхода из строя трубы в тепловых сетях по имеющимся признакам. Список признаков представлен в рис 1.1.

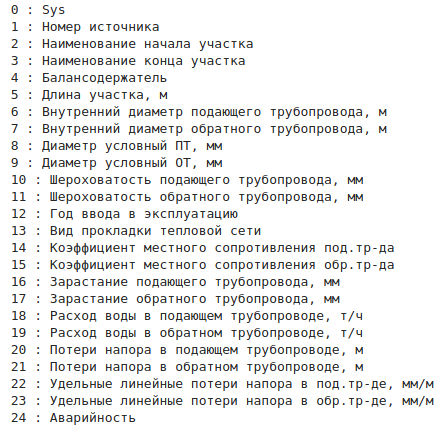


Рис. 1.1 — Список признаков

Перед тем как анализировать данные, нужно избавиться от лишних признаков. Убрав все признаки типа string, оставив и закодировав только «Вид прокладки тепловой сети» получим следующий список признаков.

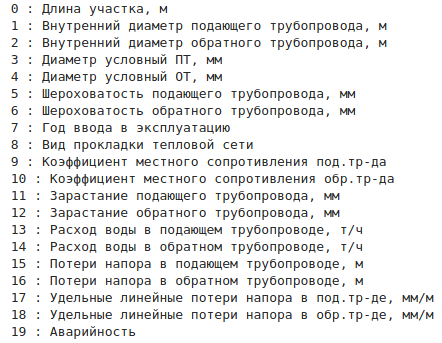


Рис. 1.2. - Числовые признаки

Далее была построена матрица содержащая коэффициенты корреляции между всеми признаками.

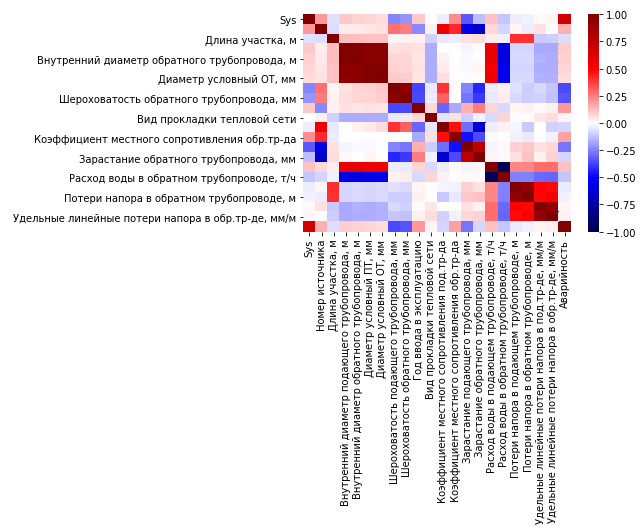


Рис 1.3 - Матрица корреляций

На матрице корреляции видно, что некоторые признаки имеют слабое влияние на аварийность. В финальный список признаков они не войдут.

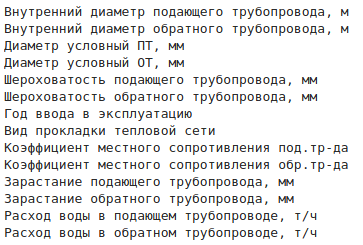


Рис. 1.4 — Финальный список признаков

Построив диаграмму аварийности по признаку «Год ввода в эксплуатацию» можно заметить огромное количество аварий у труб установленных в 1980~ гг.

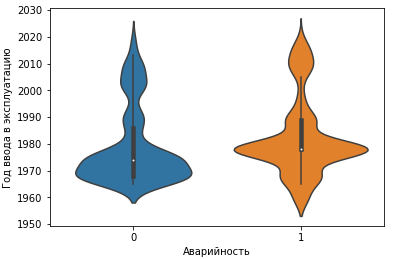


Рис. 1.5 — Скрипичный график «Год ввода в эксплуатацию»

Первое предположение, что трубы довольно старые и неудивительно что они вышли из строя. Однако та же диаграмма показывает, что у труб установленных в 1968 году аварийность намного меньше.

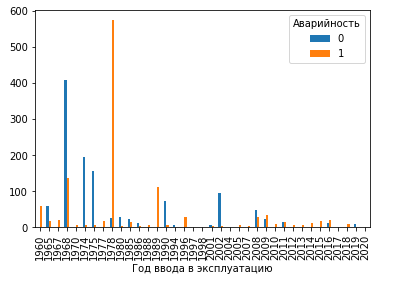


Рис. 1.6 — Сводная таблица «Год ввода в эксплуатацию»

Тем более, что количество труб введенных в 1968 и 1978 гг. в выборке примерно одинаковое.

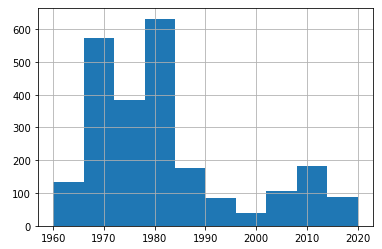


Рис. 1.7 — Гистограмма «Год ввода в эксплуатацию»

В матрице корреляции явно выделяется признак «Шероховатость подающего трубопровода, мм». На рис. 1.8 можно заметить, что практически во всех трубах со значениями от 0.5 до 2 мм случилась авария.

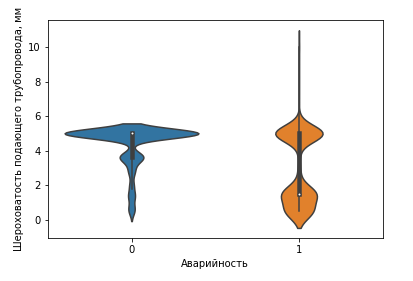


Рис. 1.8 — Скрипичный график «Шероховатость подающего трубопровода, мм»

На рис. 1.9 заметен большой всплеск на значениях 0.5 и 1.5 мм.

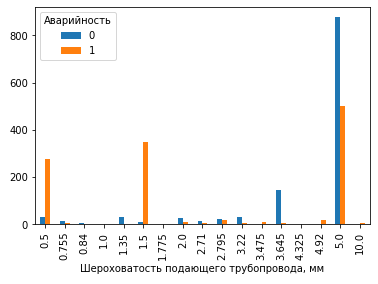


Рис. 1.9 — Сводная таблица «Шероховатость подающего трубопровода, мм»