

Кейс анализа станционных операций с ж/д вагонами

Аван Бувонита
21/05/2022



Цели и задачи исследования

Для управления железнодорожным движением необходимо ежедневно прогнозировать поток вновь отправленных вагонов (включенных в состав к отправлению со станции) на ближайшие 36-48 часов.

Необходимо проверить предположение о возможности построения прогностической модели количества вновь отправленных (разгруженных и/или погруженных) вагонов на основании статистических данных без проведения интервью на месте.

Необходимо ответить на вопрос, возможно ли по имеющимся статданным, сделать прогноз включения в поезда вновь отправленных вагонов на период 36 часов, просчитывая «судьбы» вагонов по техпроцессам.

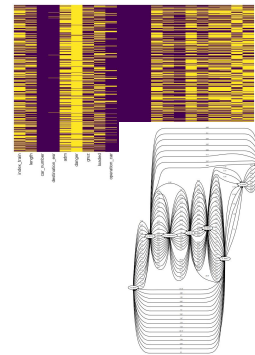
Ход исследования

Проведен ознакомительный анализ данных.

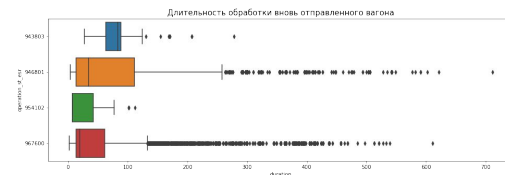
Построены графы по нескольким парам станция-вагон. Графы имеют устойчивую структуру, возможно эта структура будет повторяться для всех вагонов одного типа на данной станции.

Построено MVP таблицы для анализа длительности цепочки операций вновь отправленного вагона (исключение → подача → разгрузка → погрузка → уборка → включение в состав).

Проанализированы выборочные 4 станции и затем весь набор цепочек формирования вновь отправленных вагонов.



	operation_at_sar	car_number	start3_date	end4_date	1Xduration	2Xduration	duration
0	943803	24562753	2020-07-16 17:11:00	2020-07-21 19:56:00	35.566667	135.183333	170.750000
1	943803	29196490	2020-07-16 05:52:00	2020-07-25 07:03:00	0.000000	169.183333	169.183333
2	943803	29501491	2020-07-16 15:55:00	2020-07-25 07:03:00	0.000000	207.133333	207.133333
3	943803	29649290	2020-07-16 05:52:00	2020-07-25 07:03:00	0.000000	169.183333	169.183333
4	943803	29649399	2020-07-13 16:37:00	2020-07-16 07:58:00	42.383333	20.966667	63.350000



Результаты исследования

Выявлены значимые пропуски в данных и дублирования владельцев вагона.

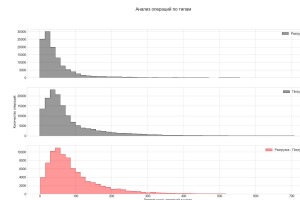
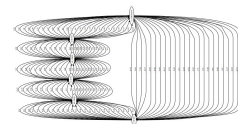
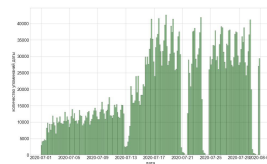
Выявлены повторяющиеся графы операций. Предположение о возможности построения прогностической модели без выезда на место подтверждено.

Выявлены статистические зависимости средней длительности операций от загруженности станции.

Небольшое количество операций имеет существенную длительность до 700 часов. Распределение длительности операций имеет длинные правые хвосты.

23% операций имеют длительность менее 24 часов. Они будут конфликтовать с ежедневным прогнозированием. Возможно прогноз надо строить каждые 8 или 12 часов.

Имеет смысл разработать SML модель прогнозирования времени обработки вагона на станции: исключение → разгрузка → погрузка → включение в состав.



Поле	Пояснение
Код станции	
Код вагона	
Тип вагона	
Тип груза	
Масса брутто	
Масса нетто	
Месяц	Календарный месяц начала операции
Момент исключения из состава	Дата, время исключения вагона (опер. 3)
Длительность подачи	
Код разгрузки	
Длительность разгрузки	Разница между моментами разгрузки и исключения вагона
Код погрузки	
Длительность погрузки	Разница между моментами погрузки и исключения вагона
Длительность уборки	
Момент включения в состав	Дата, время включения (опер. 4)
Длительность пребывания	Разница между операциями 4 и 3



Полная презентация

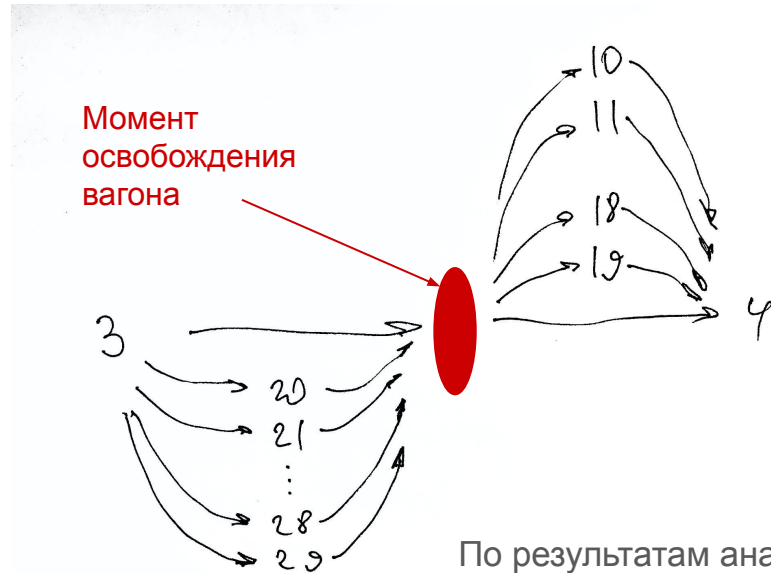
Содержание

- Постановка задачи, гипотеза
- Анализ данных, замечания по данным
- Анализ типовых цепочек операций с вагонами
- MVP таблицы для статистического анализа длительности
- Результаты анализа 4х выборочных станций
- Результаты анализа всех станций
- Подтверждения гипотез

Понимание задачи

Для управления железнодорожным движением необходимо прогнозировать поток вновь отправленных вагонов (включенных в состав к отправлению со станции). Прогноз необходимо формировать на ближайшие 36-48 часов. Предоставлен набор статистических данных об операциях с вагонами на станции и существует предположение о стационарности процессов. Предложен подход по анализу продолжительности технологических операций по станциям выгрузки-погрузки в цепочках от исключения из состава до включения в состав при наличии операций выгрузки и/или погрузки между ними. **Необходимо проверить предположение о возможности построения прогностической модели количества вновь отправленных (разгруженных и/или погруженных) вагонов на основании статистических данных без проведения интервью на месте.**

Понятие “вновь отправленный вагон”



Вновь отправленный вагон - вагон, который

- Исключен из состава (операция 3),
- Разгружен (операция 2X) и/или
- Погружен (операция 1X),
- Включен в состав (операция 4)

Между операциями 3 и 4 обязательно должна быть операция 2X и/или операция 1X.

Другие операции типа подачи на ПП или в МОП ... не имеют значения на данном этапе исследования.

По результатам анализа MVP таблицы предлагается следующая типовая цепочка операций обработки вновь отправленного вагона (**исключение** → **подача** → **разгрузка** → **погрузка** → **уборка** → **включение в состав**).

Формулировка гипотезы

Цепочки операций на станциях стабильны и могут быть изучены без выезда и проведения интервью на месте. Длительность операции может варьироваться в пределах, которые необходимо определить.

Предоставленные данные могут быть преобразованы к формату записей о длительности времени обработки вновь отправляемых вагонов. Такие записи могут быть использованы в качестве входных данных для задачи машинного обучения с учителем. Для прогнозирования потока вновь отправляемых ж/д вагонов необходимо решить одну или несколько ML задач типов классификации и регрессии.

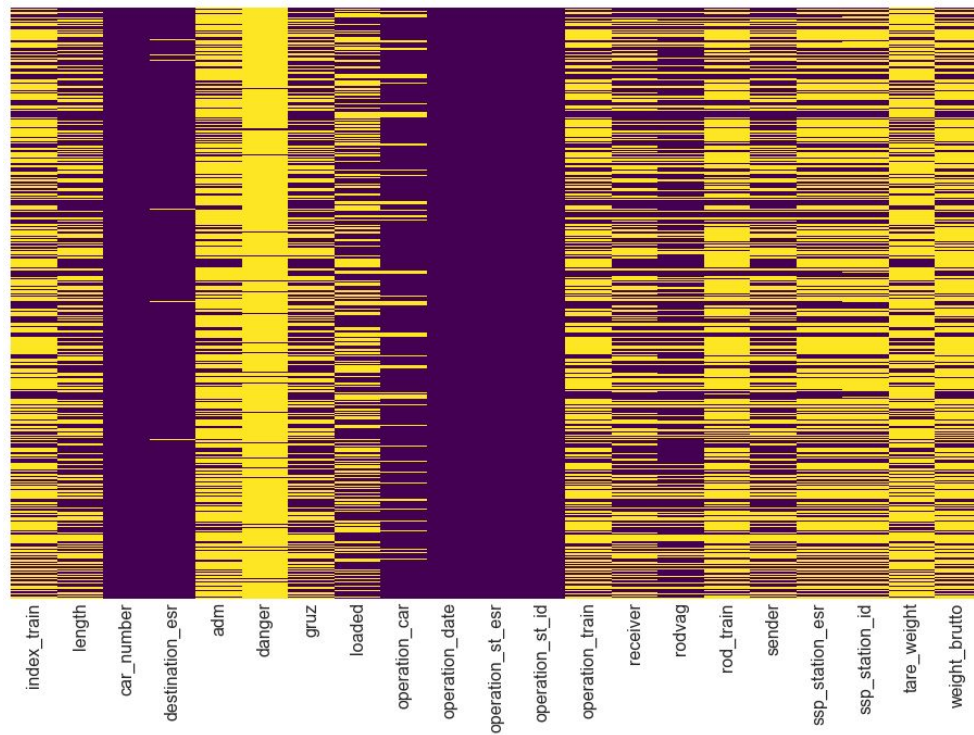
Тестирование гипотезы

- Проверка цепочек операций обработки вагонов на станции
- Разработка конвертера для формирования статистического набора данных по времени обработки вновь отправляемых вагонов.
- Статистический анализ времени обработки вновь отправляемых вагонов по одной станции.
- Сравнительный статистический анализ времени обработки по нескольким станциям.
- Презентация результатов статистического анализа.
- Разработка простейшей модели предсказания времени обработки вновь отправляемых вагонов.
- Презентация результатов моделирования.

Выявленные пропуски в базе данных

Данная тепловая карта показывает пропуски в данных. Как видим из всех столбцов у нас без пропусков только 2 и еще 2 еще имеют пропуски меньше 0.5%. В остальных столбцах пропуски достигают до 95%.

Согласно комментария ментора - в основном пропуски связаны с неполным переносом данных из таблицы с операциями по поездам.

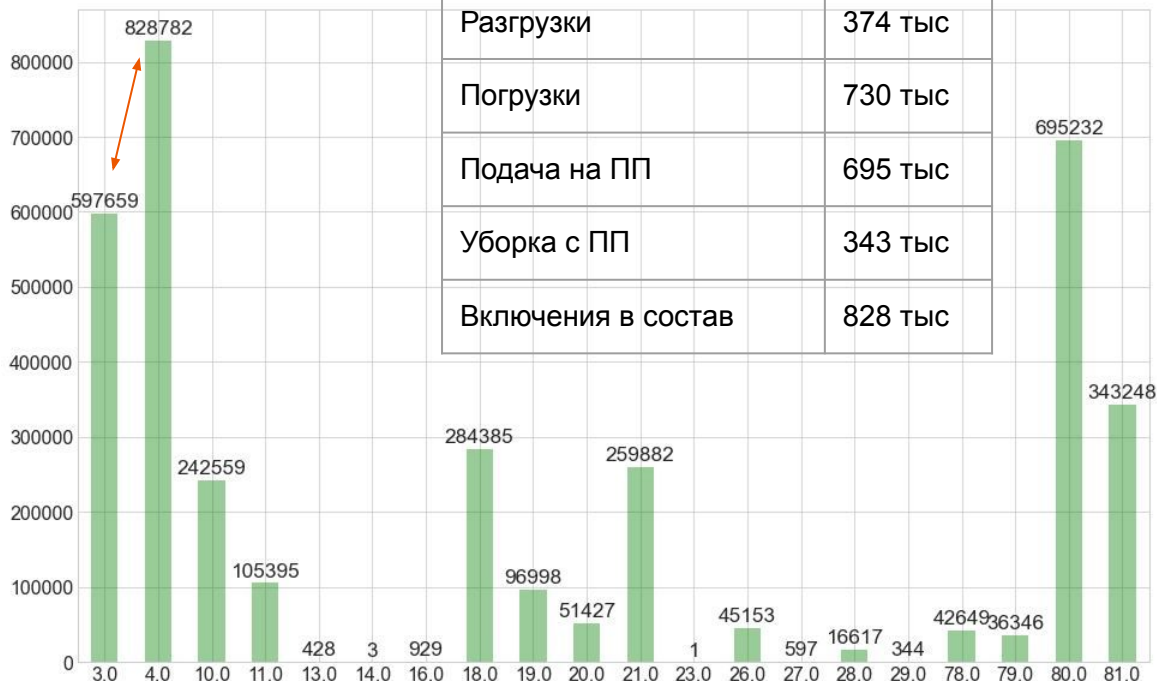


Общее количество операций

Данные включают операции с ж/д вагонами за июль 2020

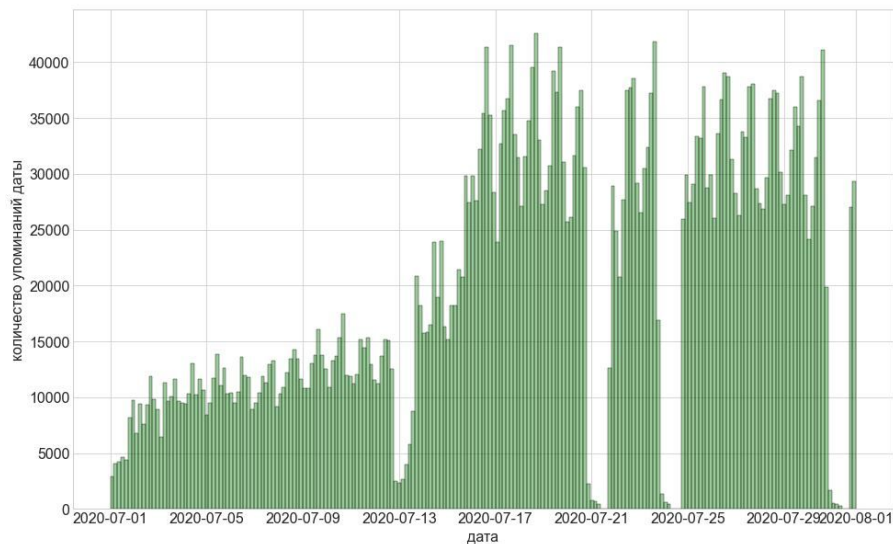
Заметна несбалансированность набора. За период данных отправлено дополнительно 230 тыс вагонов (здесь возможно 3 операция не попала в график ?).

Операции перемещения по путям занимают значительную долю по количеству и длительности(времени после момента начала этой операций до момента начала следующей).



Исключения из состава	597 тыс
Разгрузки	374 тыс
Погрузки	730 тыс
Подача на ПП	695 тыс
Уборка с ПП	343 тыс
Включения в состав	828 тыс

Количество записей по операциям



Данные в наборе распределены неравномерно. Есть странные пропуски по дням, например Аномальный провал в количестве данных в период с 18.00 20.07.2020 по 18.00 21.07.2020. В результате - нелогичное поведение вагона (разрывы в цепочках обработки).

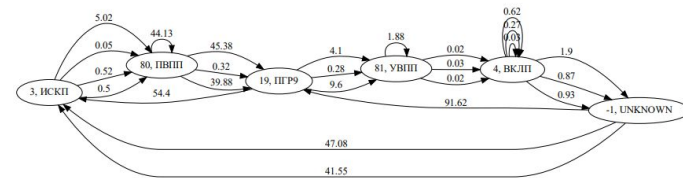
operation_car	operation_date
3	2020-07-17 15:35:00
80	2020-07-17 15:36:00
21	2020-07-20 12:34:00
19	2020-07-20 14:42:00
81	2020-07-20 15:30:00
3	2020-07-27 02:59:00
80	2020-07-27 03:00:00
21	2020-07-28 10:53:00
19	2020-07-28 12:17:00
81	2020-07-28 14:30:00
81	2020-07-28 19:20:00
4	2020-07-28 20:17:00

Нарушение логики нумерации вагонов.

Найдено 44 вагона, у которых по одному car_number значится 2 разных собственника adm.

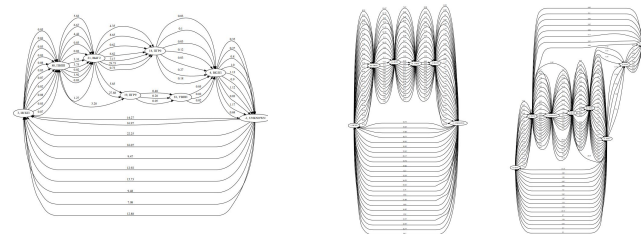
При выборочной проверке нескольких вагонов, поведение одного (№58680059) показалось странным, т. к. примерно в одно и то же время находился в разных станциях. В ходе дальнейшей проверки выяснилось, что это 2 соседние станции, в 10 км друг от друга. Этот вагон реально уехал из одной, через 30 минут засветился на другой и еще через 40 минут вернулся в исходную точку. Описанный случай является не отлавливаемым выбросом для ежедневного формирования прогноза количества вновь отправленных поездов. Если такие выбросы значимы - необходимо повышать частоту уточнения прогнозов.

Анализ графов обслуживания вагонов на станции



1. Цепочки обработки вагона однотипны и стабильны. Их можно выделить. На станциях, которые встречаются чаще, просто увеличивается количество повторов одинаковых цепочек. Можно предположить что все вагоны того же типа будут обрабатываться по сходной цепочке.
2. Можно подтвердить гипотезу о том, что специалист может выявить шаблоны движения вагонов в ходе выгрузки-погрузки без выезда на место.
3. При обработке вагона для его нового отправления могут занять значительное время операции не связанные с выгрузкой (2X) и/или погрузкой (1X). Это в первую очередь 80,81 операции - подача/уборка вагона на/с подъездных путей. Для прогнозирования времени обработки вагона они могут иметь решающее значение.
4. Время обработки одного вагона может быть меньше 24 часов (например, в случае небольшой дистанции пробега вагона с грузом). Такие обработки могут создавать сложности при формировании ежедневного прогноза отправки на ближайшие 36-48 часов. Необходимо будет формировать прогноз чаще, например каждые 12 часов - если это необходимо.

Подробности приведены в приложении.



Анализ цепочек вновь отправленных вагонов (MVP)

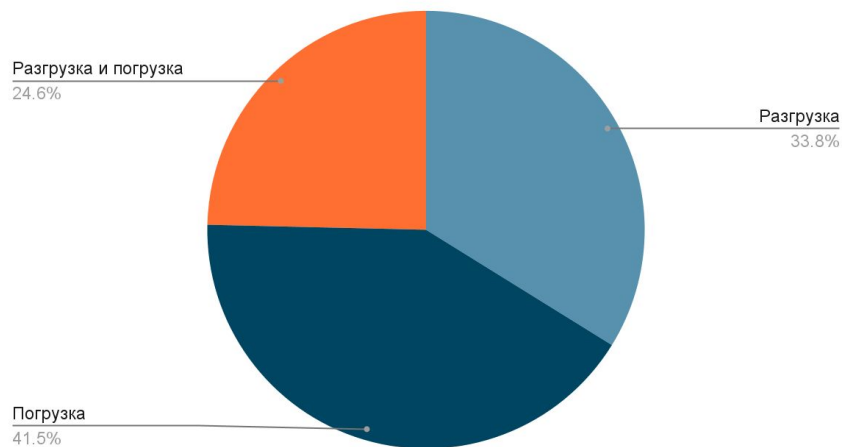
Из представленных данных сформирована простая таблица (MVP) вида:

operation_st_esr	уник номер станции
car_number	уник номер вагона
start3_date	дата/время начала операции исключения вагона из ж/д состава
end4_date	дата/время начала операции включения вагона в ж/д состав
1Xduration	[ч], длительность погрузки, обязательно включая операцию 1X и часть операций подачи-уборки вагона
2Xduration	[ч], длительность разгрузки аналогично для операций 2X
duration	длительность между 3 и 4 операциями=сумма 1X+2X

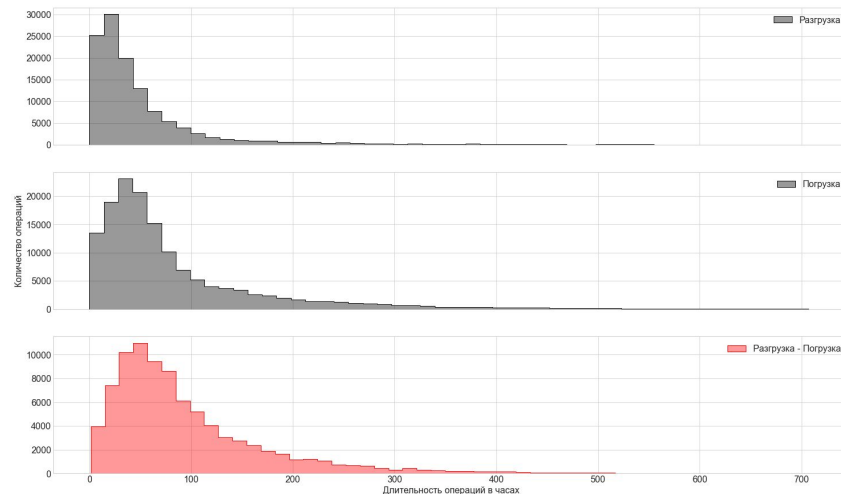
Таблица может быть использована для анализа статистики длительностей разгрузки - погрузки. Для прогнозирования нужны дополнительные признаки (тип вагона, характер груза, масса груза,...).

Анализ операций по типам

Доля операций в данных



Анализ операций по типам



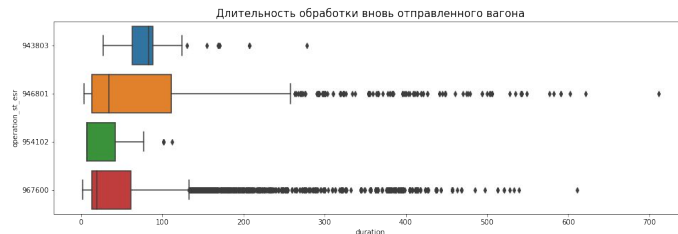
Данные в полном наборе распределены равномерно и особых отличий по типу не заметно. Заметны длинные хвосты распределений длительности операций по типам.

Выборочный анализ обслуживания вагонов

Первоначально был сформирован выборочный набор для анализа по 4м станциям:

1. На основе предоставленных данных могут быть сформированы записи по операциям новой отправки вагонов в одной строке. Статистический анализ и последующая реализация ML задачи лучше производятся на таких строках.
2. Необходимо согласовать формат представления таких записей (включение отдельных столбцов по операциям движения вагона по станции (80,81)и прочим операциям.
3. Вызывают опасения возможность ежедневного планирования при наличии операций новой отправки длительностью менее 24 часов.
4. Различия в средней длительности обслуживания вагона заметны, однако не понятно - общая ли это тенденция или индивидуальные особенности станций.

Детали приведены в Приложении.



Анализ обслуживания вагонов на полном наборе

1. Формирование записей по операциям новой отправки вагонов в одной строке занимает на полном наборе значительное время (порядка 10 часов). Это связано с проверками цепочек операций с вагонами на станции на интерпретируемом языке. Возможно это время удастся сократить при переходе на SQL или к векторным операциям pandas.
2. Операции обработки длительностью менее 24 часов составляют существенную часть набора. Вызывают опасения возможность ежедневного планирования при наличии операций новой отправки длительностью менее 24 часов.
3. Все операции характеризуются длинными хвостами распределения. При разработке решения надо понять и принять решение по обрубанию хвостов распределений.
4. Возможно, тенденция стабильная на уровне одной станции и необходимо будет разработать простые модели по количеству станций.

Подробности в приложении.

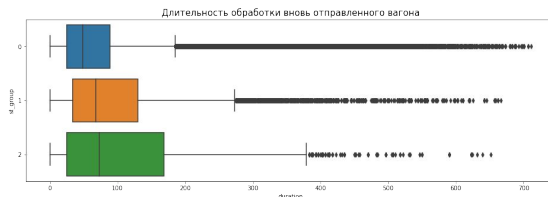


Таблица прогнозирования времени обработки

Поле	Пояснение
Код станции	
Код вагона	
Тип вагона	
Тип груза	
Масса брутто	
Масса нетто	
Месяц	Календарный месяц начала операции
.....	
Момент исключения из состава	Дата, время исключения вагона (опер.3)
Длительность подачи	
Код разгрузки	Код операции разгрузки, если есть
Длительность разгрузки	Разница между моментами разгрузки и исключения вагона
Код погрузки	Код операции погрузки, если есть
Длительность погрузки	Разница между моментами включения и освобождения
Длительность уборки	
Момент включения в состав	Дата, время включения (опер 4)
Длительность пребывания	Разница между операциями 4 и 3

Таблица предполагается состоящей из двух частей:

- Коды и признаки из исходной таблицы (состав надо уточнить)
- Укрупненная цепочка анализа времени подачи-разгрузки-погрузки-уборки вагона

ML задача будет относиться к типу обычных регрессий (или классификаций)

Подтверждение гипотез

Цепочки операций на станциях стабильны и могут быть изучены без выезда и проведения интервью на месте. Длительность операции может варьироваться в пределах, которые необходимо определить.	Подтверждено
Предоставленные данные могут быть преобразованы к формату записей о длительности времени обработки вновь отправляемых вагонов.	Подтверждено
Такие записи могут быть использованы в качестве входных данных для задачи машинного обучения с учителем.	Подтверждено
Такие записи могут быть использованы в качестве входных данных для задачи машинного обучения с учителем.	Не хватило времени



Приложения

1. Анализ графов обслуживания на станции

Пример анализа случайной пары

Для анализа использован ноутбук XXX

В таблице перечислены даты и номера операций. В каждой строке начальная операция -> следующая за ней операция -> разница в часах между ними.

В таблице встречается код операции NaN. Пропуск в поле operation_st_esr почти всегда идет в цепочках после операции 4 за короткое время, а после NaN через продолжительное время идет операция 3. Вероятнее всего NaN означает конец жизненного цикла вагона на станции (например, отправление поезда со станции).

	start_oper_date	start_oper	end_oper_date	end_oper	duration
0	2020-07-15 14:46:00	3.00	2020-07-17 04:32:00	78.00	37.77
1	2020-07-17 04:32:00	78.00	2020-07-17 06:10:00	20.00	1.63
2	2020-07-17 06:10:00	20.00	2020-07-17 12:01:00	80.00	5.85
3	2020-07-17 12:01:00	80.00	2020-07-17 21:51:00	80.00	9.83
4	2020-07-17 21:51:00	80.00	2020-07-18 08:16:00	78.00	10.42
5	2020-07-18 08:16:00	78.00	2020-07-18 13:05:00	10.00	4.82
6	2020-07-18 13:05:00	10.00	2020-07-18 14:00:00	79.00	0.92
7	2020-07-18 14:00:00	79.00	2020-07-18 14:08:00	4.00	0.13
8	2020-07-18 14:08:00	4.00	2020-07-19 00:18:00	-1.00	10.17
9	2020-07-19 00:18:00	-1.00	2020-07-23 15:58:00	3.00	111.67
10	2020-07-23 15:58:00	3.00	2020-07-23 16:40:00	80.00	0.70
11	2020-07-23 16:40:00	80.00	2020-07-25 09:16:00	78.00	40.60
12	2020-07-25 09:16:00	78.00	2020-07-25 11:39:00	10.00	2.38
13	2020-07-25 11:39:00	10.00	2020-07-25 14:40:00	79.00	3.02
14	2020-07-25 14:40:00	79.00	2020-07-25 15:26:00	4.00	0.77

Вагон №: 94264926, станция №: 988386.8

Расшифровка операций:

3.0 ИСКЛЮЧЕНИЕ ВАГОНА ИЗ ПОЕЗДА

4.0 ВКЛЮЧЕНИЕ ВАГОНА В ПОЕЗД

10.0 ПОГРУЗКА НА МЕСТАХ ОБЩ. ПОЛЬЗОВАНИЯ

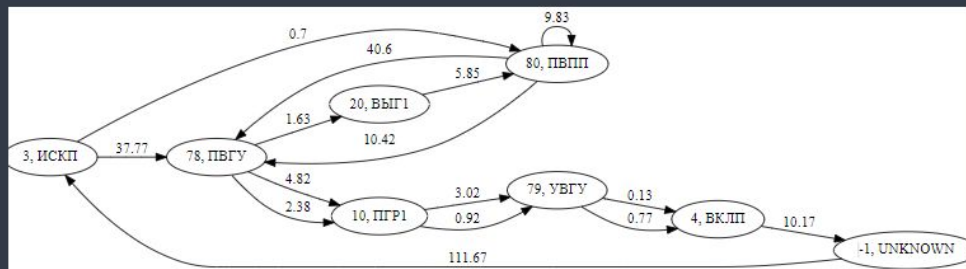
78.0 ПРОЧИЕ ПОДАЧИ ВАГОНА ГУ-45М

79.0 ПРОЧИЕ УБОРКИ ВАГОНА ГУ-45М

80.0 ПОДАЧА ВАГОНА НА ПП

20.0 ВЫГРУЗКА НА МЕСТАХ ОБЩ. ПОЛЬЗОВАНИЯ

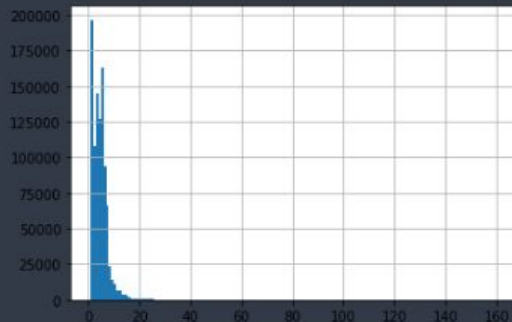
-1.0 UNKNOWN



Анализ пар “станция-вагон”

Подавляющее большинство встречается менее 10 раз. Для дальнейшего анализа будут использованы самые популярные пары.

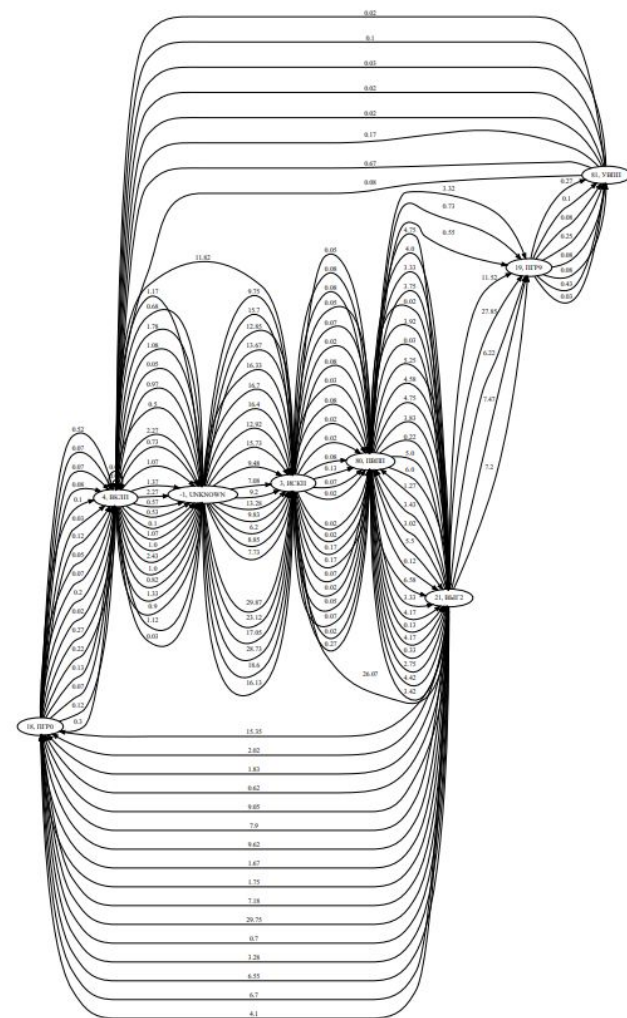
```
car_number  operation_st_esr
55864821    925701.00      162
55822928    925701.00      151
55927537    925701.00      148
55864821    926206.00      148
55822944    925701.00      142
...
61778155    986103.00       1
61778056    986103.00       1
61778007    982600.00       1
61777884    967808.00       1
98099997    936903.00       1
Length: 981058, dtype: int64
```



Кол-во операций	1	...	5	...	10	...	15	...	20	...	40	...	60	...	80	...	100	...	162
Частота	196511	...	163086	...	11512	...	2192	...	856	...	39	...	24	...	4	...	13	...	1

Вагон № 55864821
станция № 925701.0

3.0 ИСКЛЮЧЕНИЕ ВАГОНА ИЗ ПОЕЗДА
4.0 ВКЛЮЧЕНИЕ ВАГОНА В ПОЕЗД
80.0 ПОДАЧА ВАГОНА НА ПП
81.0 УБОРКА ВАГОНА С ПП
18.0 ПОГРУЗКА БЕЗ ЗАЧЕТА В ПОГРУЗКУ
19.0 ПОГРУЗКА БЕЗ ЗАЧЕТА В ПОГРУЗКУ НА ПП
21.0 ВЫГРУЗКА НА ПП
-1.0 UNKNOWN



Вагон N° 55864821 станция N° 926206.0

Расшифровка операций:

3.0 ИСКЛЮЧЕНИЕ ВАГОНА ИЗ ПОЕЗДА

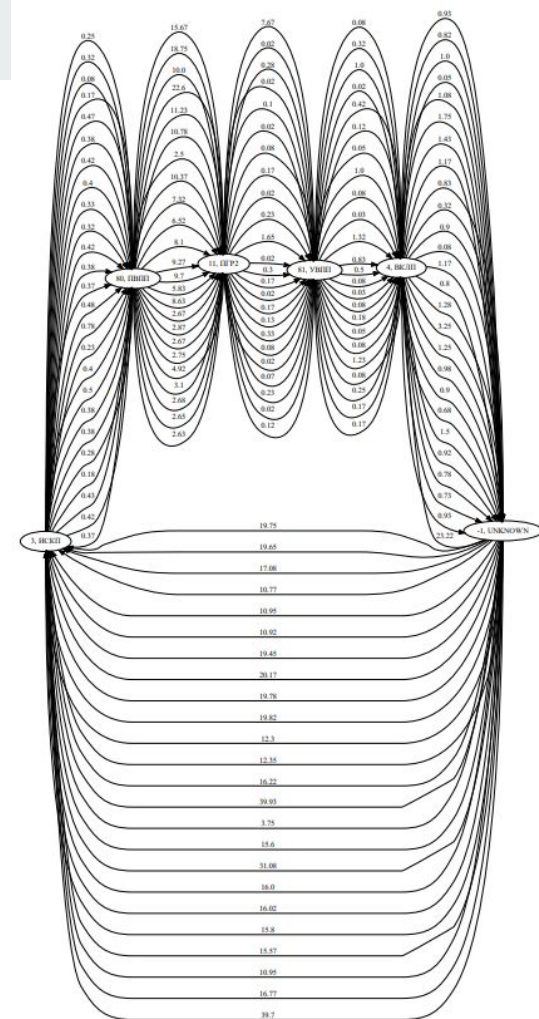
4.0 ВКЛЮЧЕНИЕ ВАГОНА В ПОЕЗД

11.0 ПОГРУЗКА НА ПП

80.0 ПОДАЧА ВАГОНА НА ПП

81.0 УБОРКА ВАГОНА С ПП

-1.0 UNKNOWN



Вагон № 55514384
станция № 925701

Количество повторов 66

3.0 ИСКЛЮЧЕНИЕ ВАГОНА ИЗ ПОЕЗДА

4.0 ВКЛЮЧЕНИЕ ВАГОНА В ПОЕЗД

80.0 ПОДАЧА ВАГОНА НА ПП

81.0 УБОРКА ВАГОНА С ПП

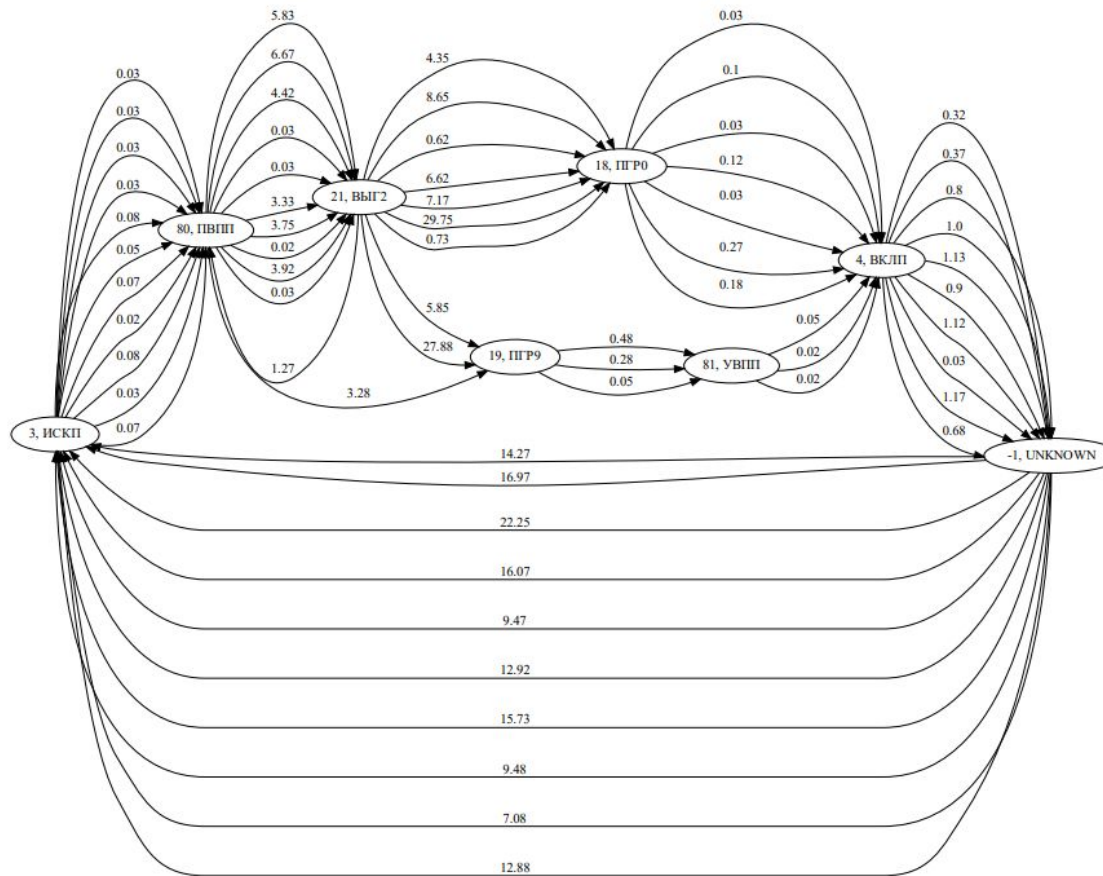
18.0 ПОГРУЗКА БЕЗ ЗАЧЕТА В ПОГРУЗКУ

19.0 ПОГРУЗКА БЕЗ ЗАЧЕТА В ПОГРУЗКУ

НА ПО

21.0 ВЫГРУЗКА НА ПП

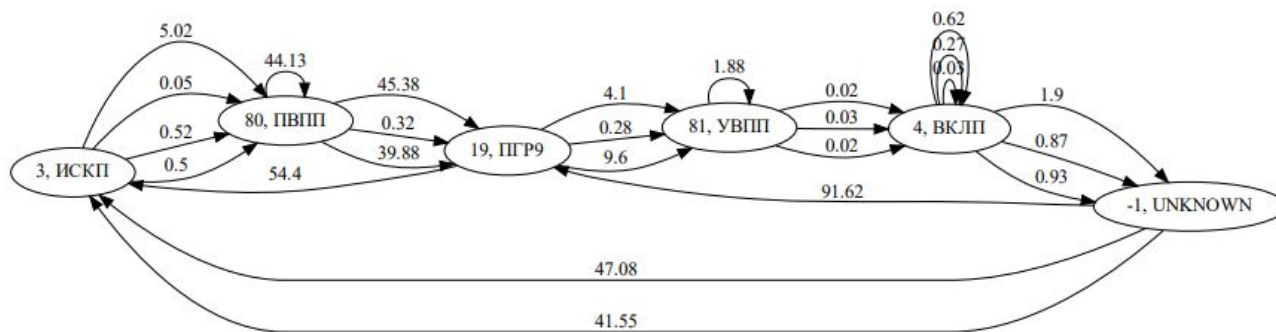
-1.0 UNKNOWN



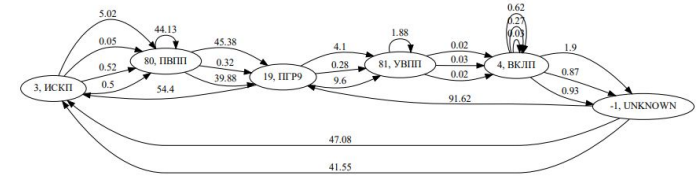
Вагон № 30027791

станция № 903204

3.0 ИСКЛЮЧЕНИЕ ВАГОНА ИЗ ПОЕЗДА
4.0 ВКЛЮЧЕНИЕ ВАГОНА В ПОЕЗД
80.0 ПОДАЧА ВАГОНА НА ПП
81.0 УБОРКА ВАГОНА С ПП
19.0 ПОГРУЗКА БЕЗ ЗАЧЕТА В ПОГРУЗКУ НА ПП
-1.0 UNKNOWN



Анализ графов обслуживания вагонов на станции



1. Цепочки обработки вагона однотипны и стабильны. Их можно выделить. На станциях, которые встречаются чаще, просто увеличивается количество повторов одинаковых цепочек. Можно предположить что все вагоны того же типа будут обрабатываться по сходной цепочке.
2. Можно подтвердить гипотезу о том, что специалист может выявить шаблоны движения вагонов в ходе выгрузки-погрузки без выезда на место.
3. При обработке вагона для его нового отправления могут занять значительное время операции не связанные с выгрузкой (2X) и/или погрузкой (1X). Это в первую очередь 80,81 операции - подача/уборка вагона на/с подъездных путей. Для прогнозирования времени обработки вагона они могут иметь решающее значение.
4. Время обработки одного вагона может быть меньше 24 часов (например, в случае небольшой дистанции пробега вагона с грузом). Такие обработки могут создавать сложности при формировании ежедневного прогноза отправки на ближайшие 36-48 часов. Необходимо будет формировать прогноз чаще, например каждые 12 часов - если это необходимо.

Подробности приведены в приложении _

2. Выборочный анализ 4х станций

MVP таблицы анализа длительности

	operation_st_esr	car_number	start3_date	end4_date	1Xduration	2Xduration	duration
0	943803	24562753	2020-07-14 17:11:00	2020-07-21 19:56:00	35.566667	135.183333	170.750000
1	943803	29196490	2020-07-18 05:52:00	2020-07-25 07:03:00	0.000000	169.183333	169.183333
2	943803	29501491	2020-07-16 15:55:00	2020-07-25 07:03:00	0.000000	207.133333	207.133333
3	943803	29649290	2020-07-18 05:52:00	2020-07-25 07:03:00	0.000000	169.183333	169.183333
4	943803	29649399	2020-07-13 16:37:00	2020-07-16 07:58:00	42.383333	20.966667	63.350000

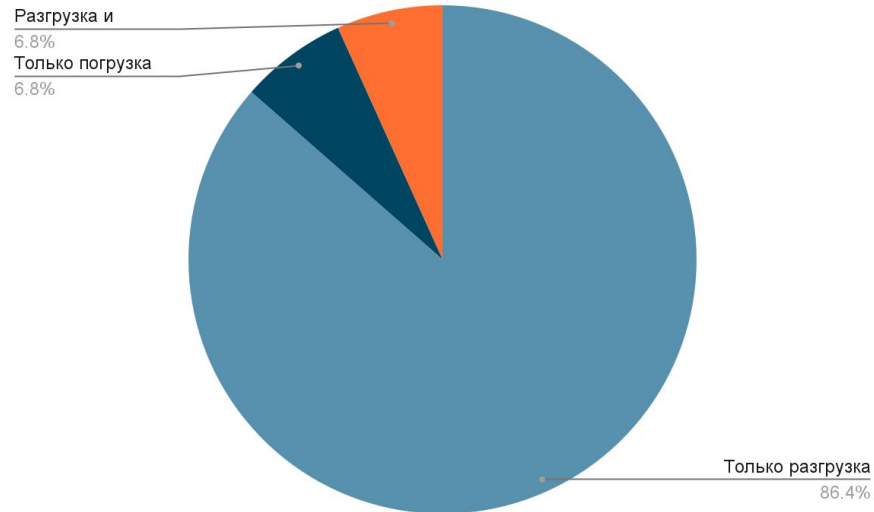
№ станции	Кол-во цепочек
967600	19332
946801	2003
943803	116
954102	111

Формирование таблицы занимает длительное время - порядка 10 часов. Для формирования требуются сложные логические проверки цепочек операций с вагонам, реализованные в интерпретируемом коде.

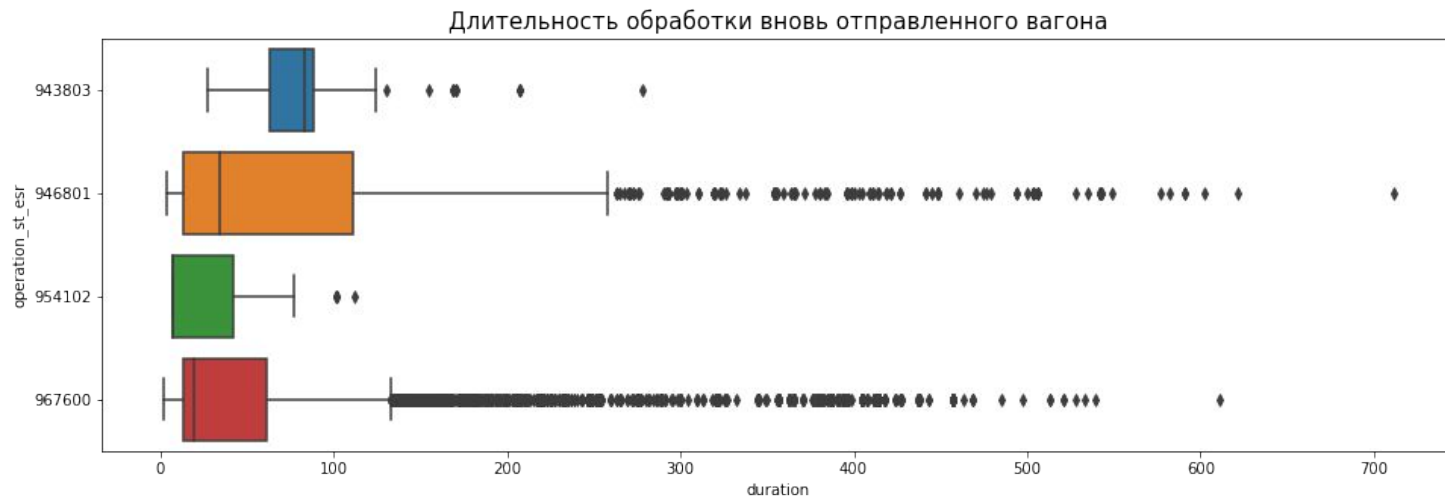
Ограниченный набор данных сформирован для 4х станций, выделенных по разной интенсивности новой отправки вагонов.

Репрезентативность выборки

Пробная выборка не совсем сбалансирована.
В нее попали станции в основном с разгрузкой
(подача, разгрузка, уборка).



Общая длительность операций новой отправки

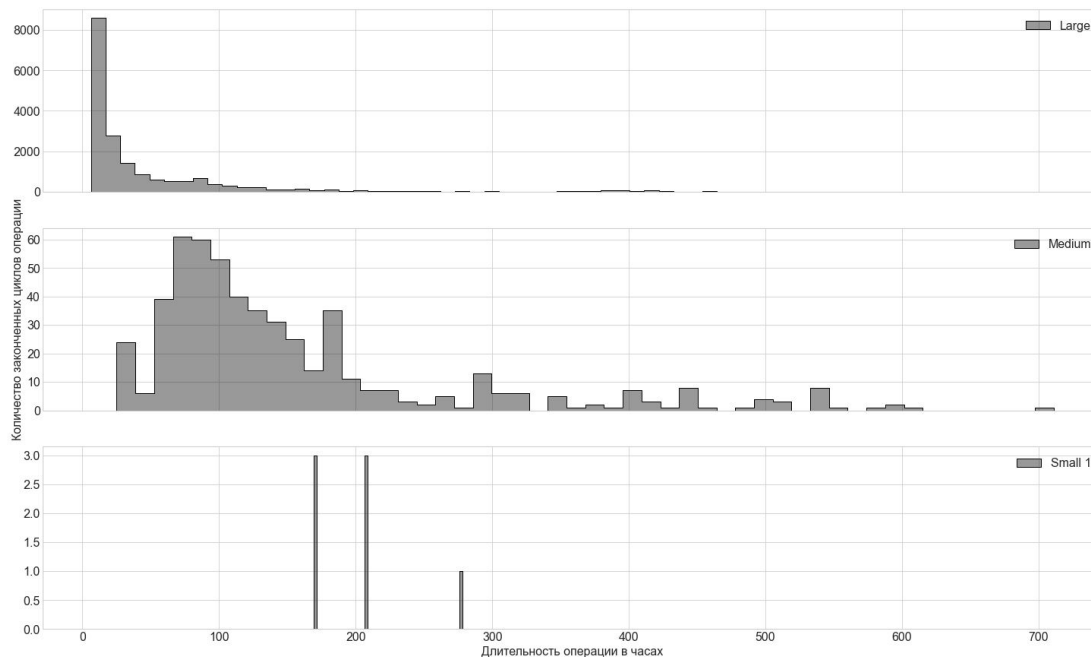


Заметна общая тенденция снижения средней длительности операции новой отправки с ростом интенсивности грузооборота. Однако вагон может находиться на станции очень долго - выбросы могут быть значительны.

Длительность операций “освобождения” вагона

Анализируются только операции освобождения (подача, разгрузка, уборка) загруженного вагона (N=167751). Вагон вновь отправлен со станции пустым.

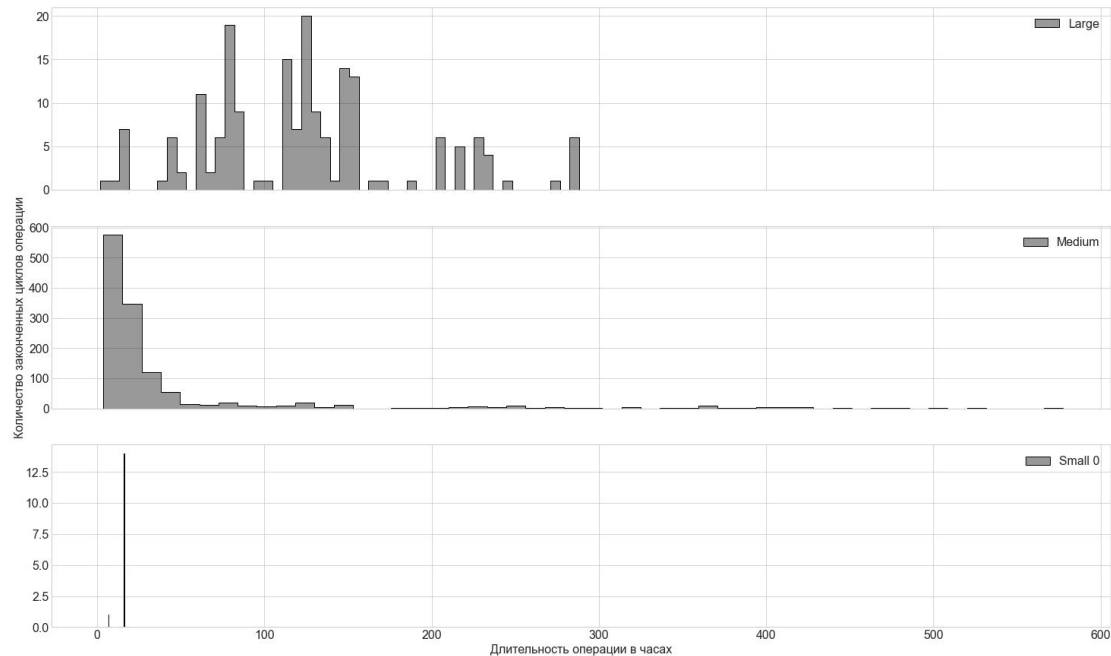
Длительность "освобождения" вагона



Длительность операций “заполнения” вагона

Длительность “заполнения” вагона

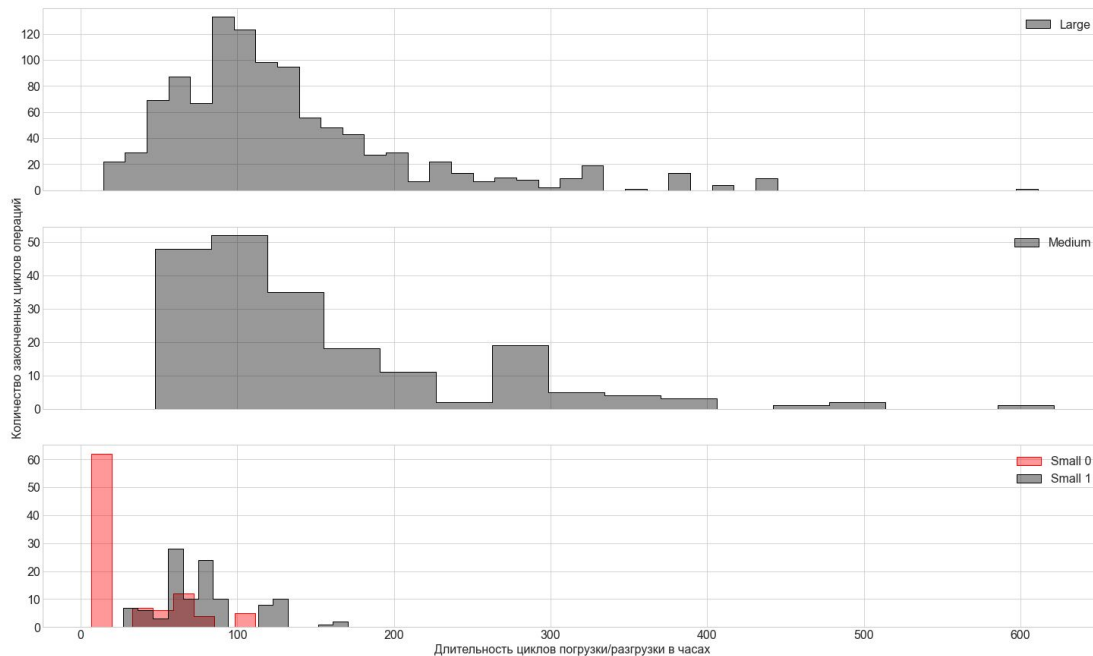
В данном случае анализируются только операции “заполнения” (подача, погрузка, уборка) пустого вагона. N=13194



Длительность разгрузки-погрузки вагонов

Длительность и количество разгрузки/погрузки

В выборке анализируются только операции (N=13113), связанные с разгрузкой и последующей погрузкой вновь отправляемых вагонов.



Выборочный анализ обслуживания вагонов

1. На основе предоставленных данных могут быть сформированы записи по операциям новой отправки вагонов в одной строке. Статистический анализ и последующая реализация ML задачи лучше производятся на таких строках.
2. Необходимо согласовать формат представления таких записей (включение отдельных столбцов по операциям движения вагона по станции (80,81)и прочим операциям.
3. Вызывают опасения возможность ежедневного планирования при наличии операций новой отправки длительностью менее 24 часов.
4. Различия в средней длительности обслуживания вагона заметны, однако не понятно - общая ли это тенденция или индивидуальные особенности станций.

3. Анализ на полном наборе станции

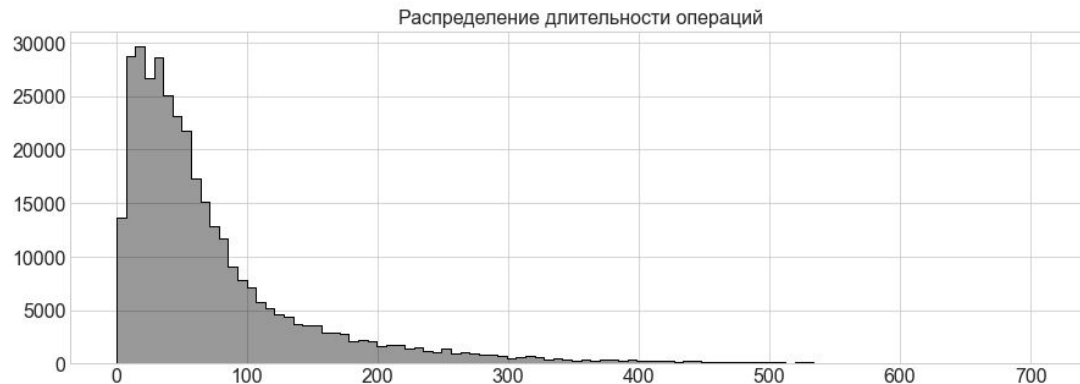
Анализ всех операций

Набор всех операций имеет ту же структуру полей, что и набор операций по 4м станциям. N=2 456 398

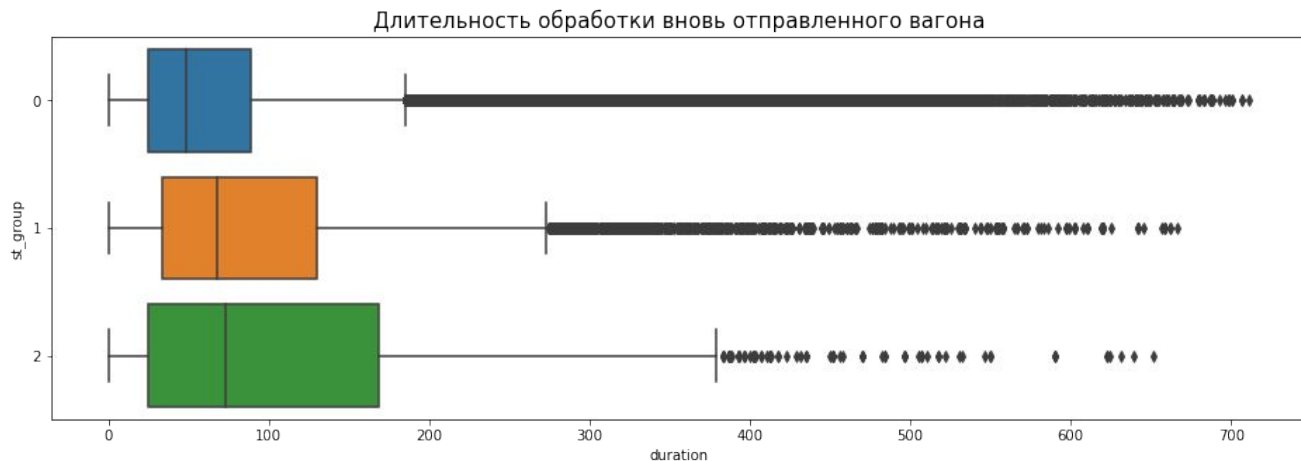
Операции распределены неравномерно по станциям и в основном представлены слабо загруженными станциями. Для целей дальнейшего анализа выделены три группы: до 40 операций, до 250 операций и свыше 250 операций

Операций длительностью менее 24 часов - 570 437 (23%).

Квантиль	Кол-во операций
25%	20
50%	96
75%	394
95%	2368



Общая длительность операций по группам станций

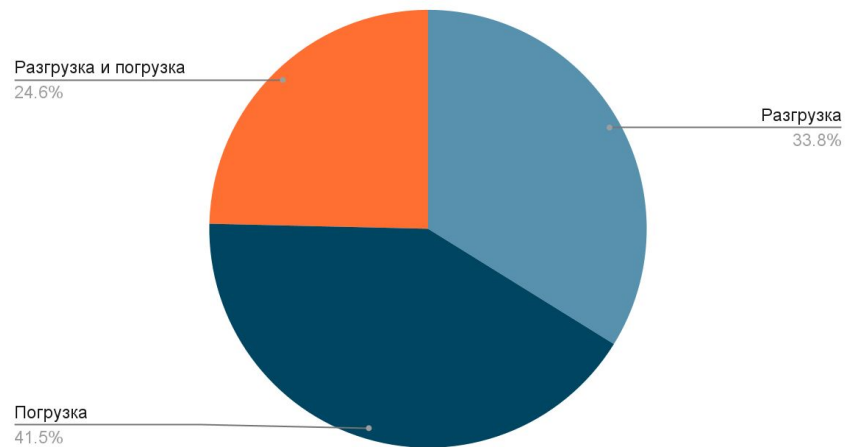


0	> 250 операций
1	40 < операций < 250
2	< 40 операций

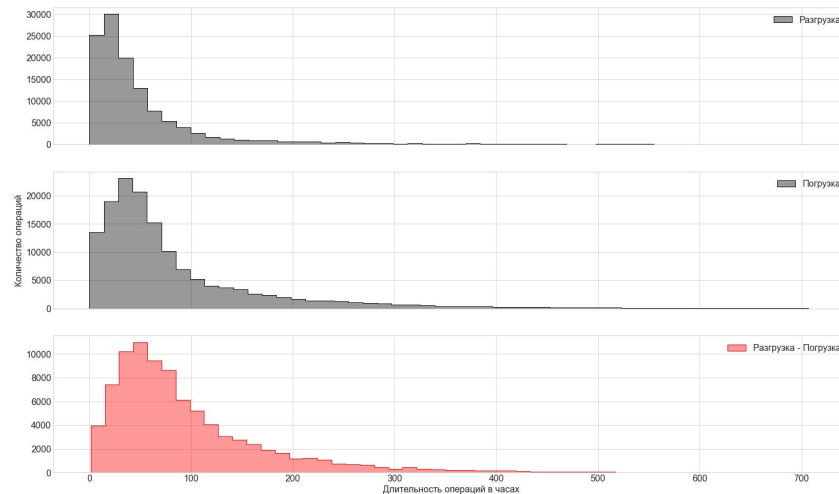
Подтверждается общая тенденция снижения средней длительности операции новой отправки с ростом интенсивности грузооборота на станции. Однако вагон может находиться на станции очень долго - выбросы могут быть значительны.

Анализ операций по типам

Доля операций в данных



Анализ операций по типам

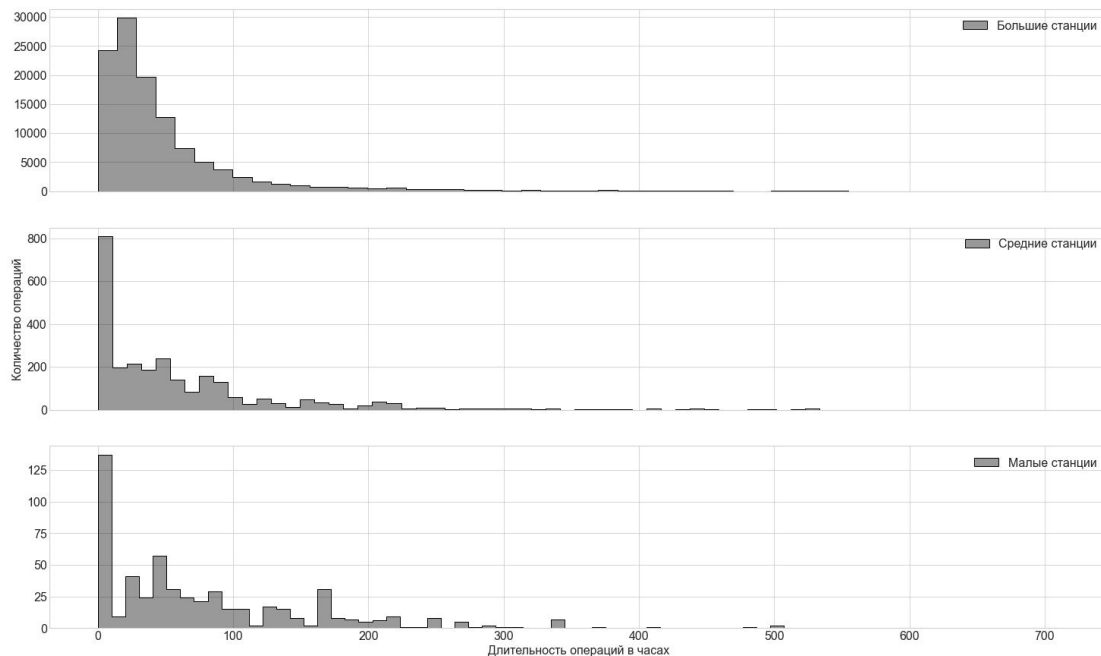


Данные в полном наборе распределены равномерно и особых отличий по типу не заметно.

Длительность операций “освобождения” вагона

В **полном** наборе анализируются только операции освобождения (подача, разгрузка, уборка) загруженного вагона (N= 1 187 970). Вагон вновь отправлен со станции пустым.

Длительность операции выгрузки вагона

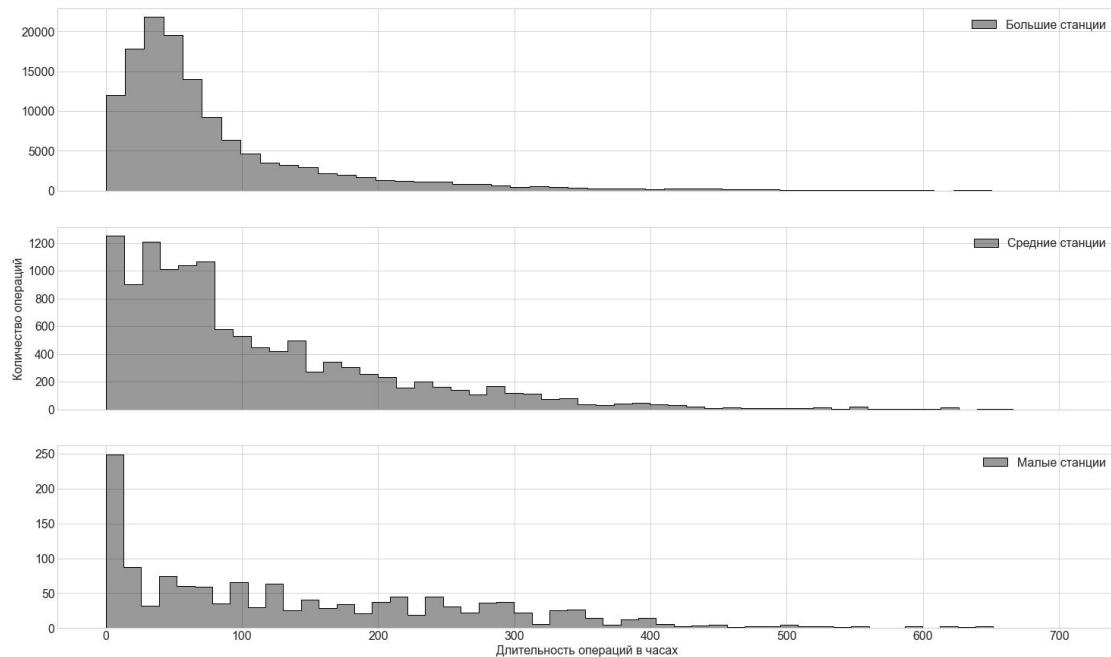


Длительность операций “заполнения” вагона

Длительность операции загрузки вагона

В **полном** наборе анализируются только операции “заполнения” (подача, погрузка, уборка) пустого вагона. N=1458720

В данном случае быстрее грузятся вагоны на станциях со слабой загрузкой. Но выбросы все равно есть.

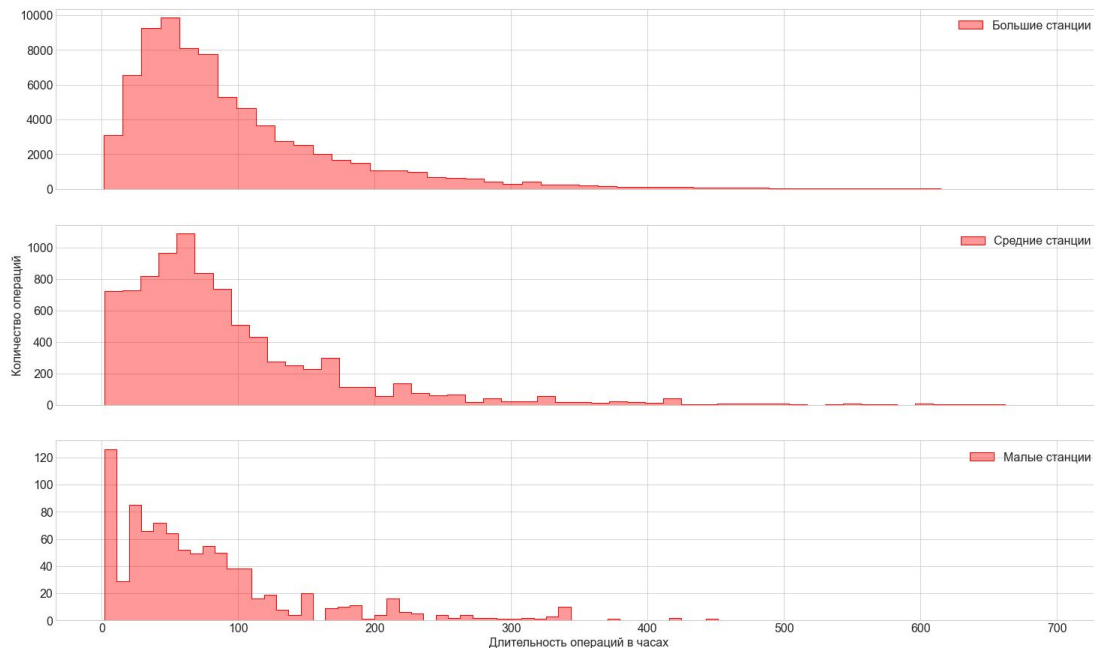


Длительность разгрузки-погрузки вагонов

Длительность операции разгрузки/погрузки вагона

В выборке из полного набора анализируются только операции (N=865190), связанные с разгрузкой и последующей погрузкой вновь отправляемых вагонов.

Здесь зависимость имеет другой характер.



Анализ обслуживания вагонов на полном наборе

1. Формирование записей по операциям новой отправки вагонов в одной строке занимает на полном наборе значительное время (порядка 10 часов). Это связано с проверками цепочек операций с вагонами на станции на интерпретируемом языке. Возможно это время удастся сократить при переходе на SQL или к векторным операциям pandas.
2. Операции обработки длительностью менее 24 часов составляют существенную часть набора. Вызывают опасения возможность ежедневного планирования при наличии операций новой отправки длительностью менее 24 часов.
3. Все операции характеризуются длинными хвостами распределения. При разработке решения надо понять и принять решение по обрубанию хвостов распределений.
4. Возможно, тенденция стабильная на уровне одной станции и необходимо будет разработать простые модели по количеству станций.

4. Литература, ноутбуки, команда

Источники.

справочники ржд - [Справочники](#)

[Технология и управление работой станций и узлов](#)

Организация работы сортировочной станции -

https://studbooks.net/2455347/tehnika/organizatsiya_raboty_sortirovochnoy_stantsii

[Единая сетевая разметка — Википедия](#)

[Приказ МПС РФ от 29.09.2003 N 67 "Об утверждении Порядка разработки и определения технологических сроков оборота вагонов и технологических норм погрузки грузов в вагоны и выгрузки грузов из вагонов" | ГАРАНТ](#)

Ноутбуки с кодом для отдельных подзадач

№	Название файла ipynb	Краткое описание выполняемой подзадачи
1	op34_time_calc.ipynb	Из всех событий в исходной базе(statistics-07-20.csv) сформировать записи, содержащие информацию о длительности каждого законченного цикла обмена грузами между вагоном и станцией. Необходимое время для обработки базы составляет 10часов.
2	op34_firstbase.ipynb	Скрипт для обработки первой базы "VIVSD_ASOUN_V_OPER_V2.3DAYS.600000-616000.hashcd.1.csv"; назначение аналогично скрипту op34_time_calc.ipynb , но реализовано значительно меньше проверок. Необходимое время работы - несколько минут.
3	DS-1.ipynb	Статистический анализ (гистограммы) анализа длительности операций по 4м выборочным станциям
4	DS-2.ipynb	Статистический анализ (гистограммы) анализа длительности операций по полному набору данных
5	Хакатон(static).ipynb	Анализ одномерных данных и графическое представление. Сделана попытка реализаций подсчета времени м/д операциями, с применением встроенных средств Pandas(alfa-verion)
6	Probability algorithm Max Andriichuk.ipynb	Реализован алгоритм вычисляющий самую популярную операцию на станции после текущей. После этого переходит к следующей самой популярной операции прибавляя ее среднюю длительность. И так далее, пока не наткнется на операцию 4 "ВКЛЮЧЕНИЕ В ПОЕЗД" или NAN. В итоге суммирует вероятность этой цепочки и общую предполагаемую длительность.
7	RZD_graphs.ipynb	Содержит функцию для построения графа последовательности операций для заданного вагона на заданной станции
8	RZD_chains.ipynb	Содержит функцию определения последовательностей операций (цепочки) вагонов на всех станциях. Жизненный цикл вагона на станции заканчивается, когда признак Код операции BM АСОУП == nan (эти циклы хорошо видны на графах). По сформированным цепочкам (топ 5 по популярности) построены графики распределения продолжительности цепочки в разрезе рода вагона и в разрезе рейтинга станции (частые-редкие)

Состав команды

					
Аверкин Степан	Андрийчук Максим	Бут Геннадий	Волков Всеволод	Николаев Илья	Табольжин Владимир
Санкт-Петербург	Самара	Москва	Москва	Санкт-Петербург	Томск
DS, ML	DS, ML	NLP, TSF, GAI	DS,ML,DL	DS, ML, CV	DS, ML