

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.

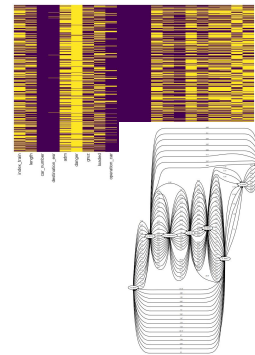
Ход исследования

Проведен ознакомительный анализ данных.

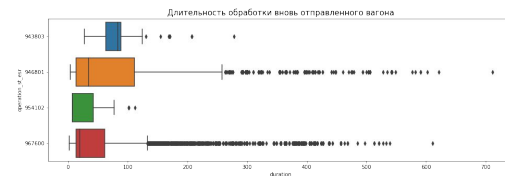
Построены графы по нескольким парам станция-вагон. Графы имеют устойчивую структуру, возможно эта структура будет повторяться для всех вагонов одного типа на данной станции.

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.

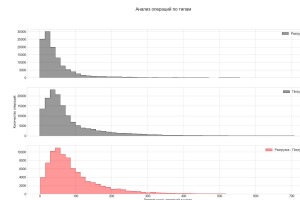
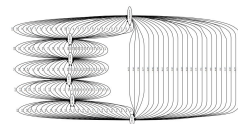
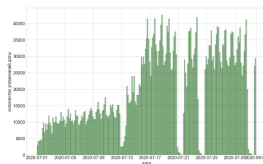
Выявлены повторяющиеся графы операций. Предположение о возможности построения прогностической модели без выезда на место подтверждено.

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

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