

Transformation plan

BOLD - нове поле

CROSSED - видалене поле

Номер	Назва	Трансформація	Коментар
1	trip_id	Додати унікальний ідентифікатор поїздки	
2	vendor_id	винести в діменшн, ID буде 1, 2, 6	6 - незадокументований вендор
	PickupDatetime		Винесли в діменшн
3	pickup_datetime_id	Форматування id	Формат id: 08012024090159
	DropoffDatetime		Винесли в діменшн
4	dropoff_datetime_id	Форматування id	Формат id: 08012024090159
5	passenger_count	Форматувати в tinyint, NULL значення замінити на 0	Можливо таксисты в NYC включають свій трекер коли їдуть собі за колом
6	trip_distance	Перевіряти на від'ємність	Є значення де відстань 0, а локація змінювалася і було оплачено, з одного боку можна це рахувати як простий, а як щодо зміни локації?
7	rate_code_id	винести в діменшн, id 1-6 tinyint	Є записи де rate code пустий і де 99 Такі дані можна видалити з офіційної факт таблиці - вони можуть бути тестовими і проаналізовані окремо.
	StoreAndFwdFlag		

8	flag_id	винести в діменшн, id 1,0 tinyint	1 - Y, 0 - N
9	pickup_loc_id	додати діменшн з описом локації	
10	dropoff_loc_id	додати діменшн з описом локації	
	PaymentType		
11	payment_type_id	винести в діменшн, id 1-6	PaymentType 0 тоді коли і rate_code_id пустий, видалиться коли почистимо rate_code_id
12	fare_amount		є негативні значення, які мають бути видалені
13	extra		є негативні значення, які мають бути видалені
14	mta_tax		є негативні значення, які мають бути видалені
15	tip_amount		є негативні значення, які мають бути видалені
16	tolls_amount		є негативні значення, які мають бути видалені
17	improvement_supercharge	є значення 1. Вони можуть бути введені помилково як відмітка що додатковий збір відбувся = це треба змінити на 0.30	є негативні значення, які мають бути видалені
18	total_amount		є негативні значення, які мають бути видалені

19	congestion_supercharge		<p>є негативні значення, які мають бути видалені</p> <p>є NULL значення які видаляться після очищення rate code</p>
20	airport_fee		<p>є NULL значення які видаляться після очищення rate code</p>