

Table	Field Name	Data Type	Description	Source Table	Source Column					
stg_yellow_tripdata	vendorid	DOUBLE PRECISION	Raw TLC vendor identifier from source CSV	file: yellow_tripdata_2023-01.csv	vendorid					
stg_yellow_tripdata	tpep_pickup_datetime	TIMESTAMP	Raw pickup timestamp from source CSV	file: yellow_tripdata_2023-01.csv	tpep_pickup_datetime					
stg_yellow_tripdata	tpep_dropoff_datetime	TIMESTAMP	Raw dropoff timestamp from source CSV	file: yellow_tripdata_2023-01.csv	tpep_dropoff_datetime					
stg_yellow_tripdata	passenger_count	DOUBLE PRECISION	Number of passengers reported by driver/file	device	file: yellow_tripdata_2023-01.csv	passenger_count				
stg_yellow_tripdata	trip_distance	DOUBLE PRECISION	Trip distance in miles	file: yellow_tripdata_2023-01.csv	trip_distance					
stg_yellow_tripdata	ratecodeid	DOUBLE PRECISION	Taxi rate code ID	file: yellow_tripdata_2023-01.csv	RatecodeID					
stg_yellow_tripdata	store_and_fwd_flag	VARCHAR(10)	Flag indicating if trip record was stored and forwarded (Y/N)	file: yellow_tripdata_2023-01.csv	store_and_fwd_flag					
stg_yellow_tripdata	pulocationid	DOUBLE PRECISION	Pickup zone location ID	file: yellow_tripdata_2023-01.csv	PULocationID					
stg_yellow_tripdata	dolocationid	DOUBLE PRECISION	Dropoff zone location ID	file: yellow_tripdata_2023-01.csv	DOLocationID					
stg_yellow_tripdata	payment_type	DOUBLE PRECISION	Payment type code	file: yellow_tripdata_2023-01.csv	payment_type					
stg_yellow_tripdata	fare_amount	DOUBLE PRECISION	Base fare amount in USD	file: yellow_tripdata_2023-01.csv	fare_amount					
stg_yellow_tripdata	extra	DOUBLE PRECISION	Extra charges (e.g. night surcharge) in USD	file: yellow_tripdata_2023-01.csv	extra					
stg_yellow_tripdata	mta_tax	DOUBLE PRECISION	MTA tax in USD	file: yellow_tripdata_2023-01.csv	mta_tax					
stg_yellow_tripdata	tip_amount	DOUBLE PRECISION	Tip amount in USD	file: yellow_tripdata_2023-01.csv	tip_amount					
stg_yellow_tripdata	tolls_amount	DOUBLE PRECISION	Tolls amount in USD	file: yellow_tripdata_2023-01.csv	tolls_amount					
stg_yellow_tripdata	improvement_surcharge	DOUBLE PRECISION	Improvement surcharge in USD	file: yellow_tripdata_2023-01.csv	improvement_surcharge					
stg_yellow_tripdata	total_amount	DOUBLE PRECISION	Total trip amount in USD	file: yellow_tripdata_2023-01.csv	total_amount					
stg_yellow_tripdata	congestion_surcharge	DOUBLE PRECISION	Congestion surcharge in USD	file: yellow_tripdata_2023-01.csv	congestion_surcharge					
stg_yellow_tripdata	airport_fee	DOUBLE PRECISION	Airport fee in USD	file: yellow_tripdata_2023-01.csv	airport_fee					
taxi_trips_clean	vendorid	DOUBLE PRECISION	TLC vendor identifier after cleaning	stg_yellow_tripdata	vendorid					
taxi_trips_clean	ratecodeid	DOUBLE PRECISION	Rate code ID after cleaning	stg_yellow_tripdata	ratecodeid					
taxi_trips_clean	payment_type	DOUBLE PRECISION	Payment type after cleaning	stg_yellow_tripdata	payment_type					
taxi_trips_clean	pulocationid	DOUBLE PRECISION	Pickup location zone ID	stg_yellow_tripdata	pulocationid					
taxi_trips_clean	dolocationid	DOUBLE PRECISION	Dropoff location zone ID	stg_yellow_tripdata	dolocationid					
taxi_trips_clean	tpep_pickup_datetime	TIMESTAMP	Clean pickup timestamp (non-null)	stg_yellow_tripdata	tpep_pickup_datetime					
taxi_trips_clean	tpep_dropoff_datetime	TIMESTAMP	Clean dropoff timestamp (non-null)	stg_yellow_tripdata	tpep_dropoff_datetime					
taxi_trips_clean	trip_date	DATE	Trip pickup calendar date	Derived from tpep_pickup_datetime	CAST(tpep_pickup_datetime AS DATE)					
taxi_trips_clean	trip_year	INT	Trip pickup year	Derived from tpep_pickup_datetime	EXTRACT(YEAR FROM tpep_pickup_datetime)					
taxi_trips_clean	trip_quarter	INT	Trip pickup quarter (1-4)	Derived from tpep_pickup_datetime	EXTRACT(QUARTER FROM tpep_pickup_datetime)					
taxi_trips_clean	trip_month	INT	Trip pickup month (1-12)	Derived from tpep_pickup_datetime	EXTRACT(MONTH FROM tpep_pickup_datetime)					
taxi_trips_clean	trip_day	INT	Trip pickup day of month	Derived from tpep_pickup_datetime	EXTRACT(DAY FROM tpep_pickup_datetime)					
taxi_trips_clean	trip_hour	INT	Trip pickup hour (0-23)	Derived from tpep_pickup_datetime	EXTRACT(HOUR FROM tpep_pickup_datetime)					
taxi_trips_clean	trip_dow	INT	Trip pickup day of week (0=Sun)	Derived from tpep_pickup_datetime	EXTRACT(DOW FROM tpep_pickup_datetime)					
taxi_trips_clean	passenger_count	DOUBLE PRECISION	Number of passengers (clean, non-negative)	stg_yellow_tripdata	passenger_count					
taxi_trips_clean	trip_distance	DOUBLE PRECISION	Trip distance in miles (clean, non-negative)	stg_yellow_tripdata	trip_distance					
taxi_trips_clean	fare_amount	DOUBLE PRECISION	Fare amount (clean, non-negative)	stg_yellow_tripdata	fare_amount					
taxi_trips_clean	extra	DOUBLE PRECISION	Extra charges (clean)	stg_yellow_tripdata	extra					
taxi_trips_clean	mta_tax	DOUBLE PRECISION	MTA tax (clean)	stg_yellow_tripdata	mta_tax					
taxi_trips_clean	tip_amount	DOUBLE PRECISION	Tip amount (clean)	stg_yellow_tripdata	tip_amount					
taxi_trips_clean	tolls_amount	DOUBLE PRECISION	Tolls amount (clean)	stg_yellow_tripdata	tolls_amount					
taxi_trips_clean	improvement_surcharge	DOUBLE PRECISION	Improvement surcharge (clean)	stg_yellow_tripdata	improvement_surcharge					
taxi_trips_clean	congestion_surcharge	DOUBLE PRECISION	Congestion surcharge (clean)	stg_yellow_tripdata	congestion_surcharge					
taxi_trips_clean	airport_fee	DOUBLE PRECISION	Airport fee (clean)	stg_yellow_tripdata	airport_fee					
taxi_trips_clean	total_amount	DOUBLE PRECISION	Total amount (clean, non-negative)	stg_yellow_tripdata	total_amount					
taxi_trips_clean	trip_duration_minutes	INT	Trip duration in minutes	Derived from timestamps	DATEDIFF(minute,tpep_pickup_datetime,tpep_dropoff_datetime)					
taxi_trips_clean	total_surcharges	DOUBLE PRECISION	Total surcharges (extra + taxes + fees)	Derived from surcharge components	COALESCE(extra,0)+COALESCE(mta_tax,0)+COALESCE(improvement_surcharge,0)+COALESCE(congestion_surcharge,0)+COALESCE(airport_fee,0)					
dim_vendor	vendor_key	BIGINT IDENTITY(1,1)	Surrogate key for vendor dimension	taxi_trips_clean	vendorid (distinct)					
dim_vendor	vendorid	BIGINT	Business vendor ID used in TLC data	taxi_trips_clean	vendorid					
dim_vendor	vendor_name	VARCHAR(100)	Human-readable vendor name (CMT, Verifone, Other)	Derived from vendorid	CASE vendorid WHEN 1 THEN 'Creative Mobile Technologies' ... END					
dim_rate_code	ratecode_key	BIGINT IDENTITY(1,1)	Surrogate key for rate code dimension	taxi_trips_clean	ratecodeid (distinct)					
dim_rate_code	ratecodeid	BIGINT	Business rate code ID	taxi_trips_clean	ratecodeid					
dim_rate_code	ratecode_desc	VARCHAR(100)	Description of rate code (Standard, JFK, etc.)	Derived from ratecodeid	CASE ratecodeid WHEN 1 THEN 'Standard rate' ... END					
dim_payment_type	payment_type_key	BIGINT IDENTITY(1,1)	Surrogate key for payment type dimension	taxi_trips_clean	payment_type (distinct)					
dim_payment_type	payment_type	BIGINT	Payment type code (1=credit,2=cash,...)	taxi_trips_clean	payment_type					
dim_payment_type	payment_desc	VARCHAR(100)	Description of payment type (Credit card, Cash, etc.)	Derived from payment_type	CASE payment_type WHEN 1 THEN 'Credit card' ... END					
dim_datetime	datetime_key	BIGINT IDENTITY(1,1)	Surrogate key for pickup datetime	taxi_trips_clean	trip_date + time parts (distinct)					
dim_datetime	pickup_date	DATE	Pickup calendar date	taxi_trips_clean	trip_date					
dim_datetime	pickup_year	INT	Pickup year	taxi_trips_clean	trip_year					
dim_datetime	pickup_quarter	INT	Pickup quarter (1-4)	taxi_trips_clean	trip_quarter					
dim_datetime	pickup_month	INT	Pickup month (1-12)	taxi_trips_clean	trip_month					
dim_datetime	pickup_day	INT	Pickup day of month	taxi_trips_clean	trip_day					
dim_datetime	pickup_hour	INT	Pickup hour of day (0-23)	taxi_trips_clean	trip_hour					
dim_datetime	pickup_dow	INT	Pickup day of week (0=Sun)	taxi_trips_clean	trip_dow					

Source Table	Source Column	Transform / Business Rule	Destination Table	Destination Column
yellow_tripdata_2023-01.csv	vendorid	Direct copy from raw CSV	stg_yellow_tripdata	vendorid
yellow_tripdata_2023-01.csv	tpep_pickup_datetime	Direct copy from raw CSV	stg_yellow_tripdata	tpep_pickup_datetime
yellow_tripdata_2023-01.csv	tpep_dropoff_datetime	Direct copy from raw CSV	stg_yellow_tripdata	tpep_dropoff_datetime
yellow_tripdata_2023-01.csv	passenger_count	Direct copy from raw CSV	stg_yellow_tripdata	passenger_count
yellow_tripdata_2023-01.csv	trip_distance	Direct copy from raw CSV	stg_yellow_tripdata	trip_distance
yellow_tripdata_2023-01.csv	RatecodeID	Direct copy from raw CSV as numeric	stg_yellow_tripdata	ratecodeid
yellow_tripdata_2023-01.csv	store_and_fwd_flag	Direct copy from raw CSV	stg_yellow_tripdata	store_and_fwd_flag
yellow_tripdata_2023-01.csv	PULocationID	Direct copy from raw CSV	stg_yellow_tripdata	pulocationid
yellow_tripdata_2023-01.csv	DOLocationID	Direct copy from raw CSV	stg_yellow_tripdata	dolocationid
yellow_tripdata_2023-01.csv	payment_type	Direct copy from raw CSV	stg_yellow_tripdata	payment_type
yellow_tripdata_2023-01.csv	fare_amount	Direct copy from raw CSV	stg_yellow_tripdata	fare_amount
yellow_tripdata_2023-01.csv	extra	Direct copy from raw CSV	stg_yellow_tripdata	extra
yellow_tripdata_2023-01.csv	mta_tax	Direct copy from raw CSV	stg_yellow_tripdata	mta_tax
yellow_tripdata_2023-01.csv	tip_amount	Direct copy from raw CSV	stg_yellow_tripdata	tip_amount
yellow_tripdata_2023-01.csv	tolls_amount	Direct copy from raw CSV	stg_yellow_tripdata	tolls_amount
yellow_tripdata_2023-01.csv	improvement_surcharge	Direct copy from raw CSV	stg_yellow_tripdata	improvement_surcharge
yellow_tripdata_2023-01.csv	total_amount	Direct copy from raw CSV	stg_yellow_tripdata	total_amount
yellow_tripdata_2023-01.csv	congestion_surcharge	Direct copy from raw CSV	stg_yellow_tripdata	congestion_surcharge
yellow_tripdata_2023-01.csv	airport_fee	Direct copy from raw CSV	stg_yellow_tripdata	airport_fee
stg_yellow_tripdata	*	Filter and clean records (non-null timestamps, non-negative distance and total_amount)	taxi_trips_clean	All copied fields (vendorid, ratecodeid, payment_type, pulocationid, dolocationid, timestamps, measures)
stg_yellow_tripdata	tpep_pickup_datetime	CAST to DATE for unified date	taxi_trips_clean	trip_date
stg_yellow_tripdata	tpep_pickup_datetime	EXTRACT(YEAR) for year	taxi_trips_clean	trip_year
stg_yellow_tripdata	tpep_pickup_datetime	EXTRACT(QUARTER) for quarter	taxi_trips_clean	trip_quarter
stg_yellow_tripdata	tpep_pickup_datetime	EXTRACT(MONTH) for month	taxi_trips_clean	trip_month
stg_yellow_tripdata	tpep_pickup_datetime	EXTRACT(DAY) for day of month	taxi_trips_clean	trip_day
stg_yellow_tripdata	tpep_pickup_datetime	EXTRACT(HOUR) for hour of day	taxi_trips_clean	trip_hour
stg_yellow_tripdata	tpep_pickup_datetime	EXTRACT(DOW) for day of week	taxi_trips_clean	trip_dow
stg_yellow_tripdata	tpep_pickup_datetime & tpep_dropoff_datetime	DATEDIFF(minute, pickup, dropoff)	taxi_trips_clean	trip_duration_minutes
stg_yellow_tripdata	extra, mta_tax, improvement_surcharge, congestion_surcharge, airport_fee	COALESCE and sum into one total	taxi_trips_clean	total_surcharges
taxi_trips_clean	vendorid	DISTINCT group to build dimension members	dim_vendor	vendorid
taxi_trips_clean	vendorid	CASE statement mapping 1,2,... to names	dim_vendor	vendor_name
taxi_trips_clean	ratecodeid	DISTINCT group to build dimension members	dim_rate_code	ratecodeid
taxi_trips_clean	ratecodeid	CASE statement mapping to rate descriptions	dim_rate_code	ratecode_desc
taxi_trips_clean	payment_type	DISTINCT group to build dimension members	dim_payment_type	payment_type
taxi_trips_clean	payment_type	CASE statement mapping to payment descriptions	dim_payment_type	payment_desc
taxi_trips_clean	trip_date, trip_year, trip_quarter, trip_month, trip_day, trip_hour, trip_dow	DISTINCT combinations to build datetime dimension	dim_datetime	All fields (pickup_date, pickup_year, pickup_quarter, pickup_month, pickup_day, pickup_hour, pickup_dow)
taxi_trips_clean	vendorid	Join to dim_vendor on vendorid to get surrogate key	fact_taxi_trips	vendor_key
taxi_trips_clean	ratecodeid	Join to dim_rate_code on ratecodeid to get surrogate key	fact_taxi_trips	ratecode_key
taxi_trips_clean	payment_type	Join to dim_payment_type on payment_type to get surrogate key	fact_taxi_trips	payment_type_key
taxi_trips_clean	trip_date, trip_year, trip_month, trip_day, trip_hour	Join to dim_datetime on date + time parts	fact_taxi_trips	datetime_key
taxi_trips_clean	passenger_count	Direct copy into fact table	fact_taxi_trips	passenger_count
taxi_trips_clean	trip_distance	Direct copy into fact table	fact_taxi_trips	trip_distance
taxi_trips_clean	fare_amount	Direct copy into fact table	fact_taxi_trips	fare_amount
taxi_trips_clean	extra	Direct copy into fact table	fact_taxi_trips	extra
taxi_trips_clean	mta_tax	Direct copy into fact table	fact_taxi_trips	mta_tax
taxi_trips_clean	tip_amount	Direct copy into fact table	fact_taxi_trips	tip_amount
taxi_trips_clean	tolls_amount	Direct copy into fact table	fact_taxi_trips	tolls_amount
taxi_trips_clean	improvement_surcharge	Direct copy into fact table	fact_taxi_trips	improvement_surcharge
taxi_trips_clean	congestion_surcharge	Direct copy into fact table	fact_taxi_trips	congestion_surcharge
taxi_trips_clean	airport_fee	Direct copy into fact table	fact_taxi_trips	airport_fee
taxi_trips_clean	total_surcharges	Direct copy into fact table	fact_taxi_trips	total_surcharges
taxi_trips_clean	total_amount	Direct copy into fact table	fact_taxi_trips	total_amount
taxi_trips_clean	trip_duration_minutes	Direct copy into fact table	fact_taxi_trips	trip_duration_minutes
taxi_trips_clean	pulocationid	Direct copy into fact table for mapping	fact_taxi_trips	pulocationid
taxi_trips_clean	dolocationid	Direct copy into fact table for mapping	fact_taxi_trips	dolocationid
fact_taxi_trips + dim_*	Star join of fact and dimensions	for BI/serving layer	v_taxi_trips_star	All fields exposed for Tableau
taxi_trips_clean	pulocationid	Cast to INT and rename to locationid for map view	v_taxi_trips_for_map	locationid
taxi_trips_clean	tpep_pickup_datetime	Direct copy for mapping view	v_taxi_trips_for_map	pickup_datetime
taxi_trips_clean	tpep_dropoff_datetime	Direct copy for mapping view	v_taxi_trips_for_map	dropoff_datetime
taxi_trips_clean	passenger_count	Direct copy for mapping view	v_taxi_trips_for_map	passenger_count
taxi_trips_clean	trip_distance	Direct copy for mapping view	v_taxi_trips_for_map	trip_distance
taxi_trips_clean	total_amount	Direct copy for mapping view	v_taxi_trips_for_map	total_amount