

✔ Typical Relationships in Taxi Data

Dimension	Fact FK	Relationship	Cardinality	Ordinality	
dim_time	pickup_time_id	1 time → many trips	One-to-Many	Mandatory	
dim_time	dropoff_time_id	1 time → many trips	One-to-Many	Mandatory	
dim_location	pickup_location_id	1 location → many trips	One-to-Many	Mandatory	
dim_location	dropoff_location_id	1 location → many trips	One-to-Many	Mandatory	
dim_payment	payment_type_id	1 payment type → many trips	One-to-Many	Mandatory	
dim_vendor	vendor_id	1 vendor → many trips	One-to-Many	Mandatory	
dim_rate_code	ratecode_id	1 rate code → many trips	One-to-Many	Mandatory	

dim_vendor	
PK	<u>vendor_id</u>
	vendor_name

dim_datetime	
PK	<u>datetime_id</u>
	full_datetime
	year
	month
	day
	hour
	dow

dim_pickup_location	
PK	<u>pu_location_id</u>
	pu_borough
	pu_zone

dim_dropoff_location	
PK	<u>do_location_id</u>
	do_borough
	do_zone

for dim_location, zone info need to
lookup in another CSV 'NYC Taxi
Zone Lookup'

Basically our table
dim_location would be sthe
NYC Taxi Zone Lookup.csv

Fact Table consist of r
they are
1. Numeric V.
2. Vary for eac
3. Used in AGG (Su
4. And it's a fact, it wi
change or transitio

fact_taxi_trip	
PK	<u>trip_id</u>
FK	vendor_id
FK	pickup_datetime
FK	dropff_datetime
	passenger_count
	trip_distance
FK	ratecode_id
FK	pickup_location_id
FK	dropoff_location_id
FK	payment_type_id
	fare_amount
	tip_amount

dim_ra	
PK	<u>ratecode</u>
	ratecode_

dim_pay	
PK	<u>payment</u>
	payment_

measures and
:
'alues
ch row
im, Avg, etc)
ill not going to
n like SCD

itecode
<u>_id</u>
_description

yment_type
<u>_type_id</u>
.type_description