

```
In [0]: import pandas as pd
import time
from pyspark.sql.functions import col, count, avg, sum as spark_sum, d
```

```
In [0]: df_base = spark.table("default.yellow_tripdata_2015_01")

# Create a larger dataset
df_large = df_base
for i in range(9):
    df_large = df_large.union(df_base)
```

```
In [0]: df_large.sample(withReplacement=False, fraction=0.1)
```

```
Out[0]: DataFrame[VendorID: int, tpep_pickup_datetime: timestamp, tpep_dropoff_datetime: timestamp, passenger_count: int, trip_distance: double, pickup_longitude: double, pickup_latitude: double, RateCodeID: int, store_and_fwd_flag: string, dropoff_longitude: double, dropoff_latitude: double, payment_type: int, fare_amount: double, extra: double, mta_tax: double, tip_amount: double, tolls_amount: double, improvement_surcharge: double, total_amount: double]
```

```
In [0]: df_filtered = df_large.select(
    "VendorID", "tpep_pickup_datetime", "tpep_dropoff_datetime",
    "passenger_count", "trip_distance", "fare_amount", "tip_amount",
    "total_amount", "pickup_latitude", "pickup_longitude")
```

```
In [0]: from pyspark.sql.functions import col

# Filter trips with valid passenger count and reasonable trip distance
df_filtered = df_filtered.filter(
    (col("passenger_count") > 0) &
    (col("trip_distance") > 0) &
    (col("trip_distance") < 100)
)
```

```
In [0]: from pyspark.sql.functions import col, mean, when, unix_timestamp

# -----
# 1. Remove rows with invalid critical values
# -----
critical_cols = ["passenger_count", "trip_distance", "fare_amount", "tip_amount"]

df_filtered = df_large
for c in critical_cols:
    df_filtered = df_filtered.filter(col(c) > 0)

# -----
# 2. Fill zeros or nulls in optional columns with column average
# -----
replace_cols = ["tip_amount", "tolls_amount", "improvement_surcharge",
```

```
for c in replace_cols:
    avg_val = df_filtered.select(mean(col(c))).collect()[0][0]
    df_filtered = df_filtered.withColumn(
        c,
        when(col(c).isNull() | (col(c) == 0), avg_val).otherwise(col(c))
    )

# -----
# 3. Ensure datetime columns are valid
# -----
df_filtered = df_filtered.filter(
    (col("tpep_pickup_datetime").isNotNull()) &
    (col("tpep_dropoff_datetime").isNotNull())
)

# -----
# 4. Calculate trip_duration safely
# -----
df_filtered = df_filtered.withColumn(
    "trip_duration",
    (unix_timestamp(col("tpep_dropoff_datetime")) - unix_timestamp(col("tpep_pickup_datetime"))).filter(col("trip_duration") > 0)

# -----
# 5. filter trips within NYC coordinates
# -----
df_filtered = df_filtered.filter(
    (col("pickup_latitude").between(40.5, 41)) &
    (col("pickup_longitude").between(-74.5, -73.5))
)
```

```
In [0]: from pyspark.sql.functions import unix_timestamp, round

# Trip duration in minutes
df_filtered = df_filtered.withColumn(
    "trip_duration",
    (unix_timestamp(col("tpep_dropoff_datetime")) - unix_timestamp(col("tpep_pickup_datetime")))
)

# Tip percentage
df_filtered = df_filtered.withColumn(
    "tip_percentage",
    round((col("tip_amount") / col("fare_amount")) * 100, 2)
)

df_filtered.show()
```

```
+-----+-----+-----+-----+
-----+-----+-----+-----+
-----+-----+-----+-----+
-----+-----+-----+-----+
-----+-----+-----+-----+
```

VendorID	tpep_pickup_datetime	tpep_dropoff_datetime	passenger_count	trip_distance	pickup_longitude	pickup_latitude	RateCodeID	store_and_fwd_flag	dropoff_longitude	dropoff_latitude	payment_type	fare_amount	extra	mta_tax	tip_amount	tolls_amount	improvement_surcharge	total_amount	trip_duration	tip_percentage
1	2015-01-07 12:26:05	2015-01-07 12:29:08	1	0.6	-73.99744415283203	40.74155044555664	1	N												1
0.6	-73.99744415283203	40.74155044555664	1	-74.00391387939453	40.73738098144531															4.0   0.30903
683757945827	0.5			0.96	0.2416871612181792															
0.3	5.76	3.05	24.0																	
1	2015-01-07 12:26:05	2015-01-07 12:40:01	1	1.5	-73.9557113647461	40.7642936706543	1	N												
1.5	-73.9557113647461	40.7642936706543	1	-73.97731018066406	40.76102066040039															10.0   0.30903
683757945827	0.5	2.0	0.2416871612181792																	
0.3	12.8	13.93333333333334	20.0																	
1	2015-01-07 12:26:05	2015-01-07 12:57:40	1	10.1	-73.86416625976562	40.77025604248047	1													1
10.1	-73.86416625976562	40.77025604248047	1	N	-73.9781265258789	40.78351593017578	1													33.5   0.3090
3683757945827	0.5	7.9	5.33																	
0.3	47.53	31.58333333333332	23.58																	
1	2015-01-07 12:26:05	2015-01-07 12:32:22	3	2.3	-73.97956085205078	40.78364562988281	1	N												
2.3	-73.97956085205078	40.78364562988281	1	-74.0027084350586	40.760684967041016															9.0   0.30903
683757945827	0.5	2.0	0.2416871612181792																	
0.3	11.8	6.28333333333333	22.22																	
1	2015-01-07 12:26:05	2015-01-07 12:30:30	1	0.8	-73.97408294677734	40.79190444946289	1	N												
0.8	-73.97408294677734	40.79190444946289	1	-73.98139953613281	40.78154754638672															5.5   0.30903
683757945827	0.5	0.5	0.2416871612181792																	
0.3	6.8	4.4166666666666667	9.09																	
1	2015-01-07 12:26:05	2015-01-07 12:35:36	1	1.8	-73.95401763916016	40.76653289794922	1	N												
1.8	-73.95401763916016	40.76653289794922	1	-73.97132110595703	40.75547790527344															9.0   0.30903
683757945827	0.5	1.851125934112553	0.2416871612181792																	
0.3	9.8	9.5166666666666667	20.57																	
1	2015-01-07 12:26:05	2015-01-07 13:10:00	1	21.0	-73.9776611328125	40.72565841674805	2													
21.0	-73.9776611328125	40.72565841674805	2	N	-73.78299713134766	40.648948669433594	2													52.0   0.3090
3683757945827	0.5	1.851125934112553	0.2416871612181792																	
0.3	52.8	43.916666666666664	3.56																	
1	2015-01-07 12:26:06	2015-01-07 12:43:20	1	1.7	-73.96404266357422	40.76774597167969	1	N												
1.7	-73.96404266357422	40.76774597167969	1	-73.97071838378906	40.751747131347656															11.5   0.30903
683757945827	0.5	2.46	0.2416871612181792																	
0.3	14.76	17.23333333333334	21.39																	
1	2015-01-07 12:26:06	2015-01-07 12:33:54	1	1.0	-73.975341796875	40.757713317871094	1	N												

-73.98509216308594	40.7513427734375	2	7.0 0.30903
683757945827	0.5 1.851125934112553 0.2416871612181792		
0.3	7.8	7.8	26.44
1  2015-01-07 12:26:07	2015-01-07 12:32:10	1	
0.7 -73.99703979492188	40.736690521240234	1	N
-74.00613403320312	40.733184814453125	1	5.5 0.30903
683757945827	0.5	1.85 0.2416871612181792	
0.3	8.15	6.05	33.64
1  2015-01-07 12:26:07	2015-01-07 12:35:57	1	
2.0 -73.98589324951172	40.740684509277344	1	N
-73.96318054199219	40.75823211669922	1	9.5 0.30903
683757945827	0.5	1.03 0.2416871612181792	
0.3	11.33	9.83333333333334	10.84
1  2015-01-07 12:26:07	2015-01-07 12:28:45	3	
0.6 -73.97511291503906	40.792232513427734	1	N
-73.98124694824219	40.7867431640625	1	4.0 0.30903
683757945827	0.5	0.95 0.2416871612181792	
0.3	5.75 2.63333333333333	23.75	
1  2015-01-07 12:26:08	2015-01-07 12:38:15	1	
1.3 -74.0051498413086	40.74177932739258	1	N
-74.00894165039062	40.73092269897461	2	9.0 0.30903
683757945827	0.5 1.851125934112553 0.2416871612181792		
0.3	9.8 12.11666666666667	20.57	
2  2015-01-26 11:01:41	2015-01-26 11:35:10	1	
10.6 -73.9811782836914	40.76311111450195	1	
N -73.87049865722656	40.77389144897461	1	36.0 0.3090
3683757945827	0.5	10.53	5.33
0.3	52.66 33.4833333333334	29.25	
2  2015-01-26 11:01:41	2015-01-26 11:21:49	1	
3.24 -73.98988342285156	40.73561096191406	1	
N -73.96280670166016	40.77528381347656	1	15.5 0.3090
3683757945827	0.5	4.08 0.2416871612181792	
0.3	20.38 20.133333333333	26.32	
2  2015-01-26 11:01:41	2015-01-26 11:08:17	3	
1.21 -73.9814453125	40.78089904785156	1	
N -73.98998260498047	40.76873779296875	1	6.5 0.3090
3683757945827	0.5	1.46 0.2416871612181792	
0.3	8.76	6.6	22.46
2  2015-01-26 11:01:42	2015-01-26 11:26:09	1	
3.07 -73.99935150146484	40.72515106201172	1	
N -73.96906280517578	40.758399963378906	1	17.0 0.3090
3683757945827	0.5	4.45 0.2416871612181792	
0.3	22.25	24.45	26.18
2  2015-01-26 11:01:43	2015-01-26 11:12:42	6	
1.27 -73.95466613769531	40.76416778564453	1	
N -73.97111511230469	40.761268615722656	1	8.5 0.3090
3683757945827	0.5	1.86 0.2416871612181792	
0.3	11.16 10.983333333333	21.88	
2  2015-01-26 11:01:43	2015-01-26 11:10:09	1	
1.39 -73.99331665039062	40.722564697265625	1	
N -74.00322723388672	40.73514175415039	1	7.5 0.3090
3683757945827	0.5 1.851125934112553 0.2416871612181792		

```
0.3|          8.3| 8.43333333333334|          24.68|
|          2| 2015-01-26 11:01:44| 2015-01-26 11:07:43|          1|
0.87| -73.9914321899414| 40.74974822998047|          1|
N|-73.98548126220703| 40.74357986450195|          1| 5.5|0.3090
3683757945827| 0.5|          1.65|0.2416871612181792|
0.3|          7.95| 5.98333333333333|          30.0|
+-----+-----+-----+-----+
-----+-----+-----+-----+
-----+-----+-----+-----+
-----+-----+-----+-----+
-----+-----+-----+-----+
only showing top 20 rows
```

In [0]: df\_filtered.take(5)

```
Out[0]: [Row(VendorID=2, tpep_pickup_datetime=datetime.datetime(2015, 1, 3, 1
4, 7, 26), tpep_dropoff_datetime=datetime.datetime(2015, 1, 3, 14, 1
2, 56), passenger_count=1, trip_distance=0.64, pickup_longitude=-73.9
4904327392578, pickup_latitude=40.78160858154297, RateCodeID=1, store
_and_fwd_flag='N', dropoff_longitude=-73.95654296875, dropoff_latitude=40.77803039550781, payment_type=1, fare_amount=5.5, extra=0.3090368
3757945827, mta_tax=0.5, tip_amount=1.1, tolls_amount=0.2416871612181
792, improvement_surcharge=0.3, total_amount=7.4, trip_duration=5.5,
tip_percentage=20.0),
Row(VendorID=2, tpep_pickup_datetime=datetime.datetime(2015, 1, 3, 1
4, 7, 26), tpep_dropoff_datetime=datetime.datetime(2015, 1, 3, 14, 1
4, 49), passenger_count=1, trip_distance=1.42, pickup_longitude=-73.9
6768951416016, pickup_latitude=40.76264190673828, RateCodeID=1, store
_and_fwd_flag='N', dropoff_longitude=-73.97284698486328, dropoff_latitude=40.74626541137695, payment_type=2, fare_amount=7.0, extra=0.3090
3683757945827, mta_tax=0.5, tip_amount=1.851125934112553, tolls_amount=0.2416871612181792, improvement_surcharge=0.3, total_amount=7.8, trip_duration=7.383333333333334, tip_percentage=26.44),
Row(VendorID=2, tpep_pickup_datetime=datetime.datetime(2015, 1, 3, 1
4, 7, 26), tpep_dropoff_datetime=datetime.datetime(2015, 1, 3, 14, 4
7, 7), passenger_count=1, trip_distance=18.88, pickup_longitude=-73.9
458999633789, pickup_latitude=40.77858352661133, RateCodeID=2, store
_and_fwd_flag='N', dropoff_longitude=-73.78302764892578, dropoff_latitude=40.64399337768555, payment_type=2, fare_amount=52.0, extra=0.3090
3683757945827, mta_tax=0.5, tip_amount=1.851125934112553, tolls_amount=5.33, improvement_surcharge=0.3, total_amount=58.13, trip_duration=39.6833333333333, tip_percentage=3.56),
Row(VendorID=1, tpep_pickup_datetime=datetime.datetime(2015, 1, 3, 1
4, 7, 27), tpep_dropoff_datetime=datetime.datetime(2015, 1, 3, 14, 4
3, 29), passenger_count=2, trip_distance=18.4, pickup_longitude=-73.9
7296142578125, pickup_latitude=40.75957107543945, RateCodeID=3, store
_and_fwd_flag='N', dropoff_longitude=-74.18338775634766, dropoff_latitude=40.68806457519531, payment_type=2, fare_amount=68.0, extra=0.3090
03683757945827, mta_tax=0.4987789140890973, tip_amount=1.851125934112
553, tolls_amount=16.65, improvement_surcharge=0.2832208678472794, total_amount=84.95, trip_duration=36.0333333333333, tip_percentage=2.7
2),
Row(VendorID=1, tpep_pickup_datetime=datetime.datetime(2015, 1, 3, 1
4, 7, 27), tpep_dropoff_datetime=datetime.datetime(2015, 1, 3, 14, 3
2, 59), passenger_count=2, trip_distance=4.5, pickup_longitude=-73.99
04556274414, pickup_latitude=40.73509979248047, RateCodeID=1, store
_and_fwd_flag='N', dropoff_longitude=-73.94779968261719, dropoff_latitude=40.78046798706055, payment_type=2, fare_amount=19.5, extra=0.30903
683757945827, mta_tax=0.5, tip_amount=1.851125934112553, tolls_amount=0.2416871612181792, improvement_surcharge=0.2832208678472794, total_amount=20.3, trip_duration=25.53333333333335, tip_percentage=9.49)]
```

```
In [0]: from pyspark.sql.functions import avg, sum, round

df_agg = df_filtered.groupBy("VendorID").agg(
    round(avg("trip_distance"), 2).alias("avg_trip_distance"),
    round(sum("fare_amount"), 2).alias("total_fare"),
```

```
    round(avg("tip_percentage"), 2).alias("avg_tip_pct"),
    round(avg("trip_duration"), 2).alias("avg_trip_duration")
)

df_agg.show()
```

```
+-----+-----+-----+-----+
+|VendorID|avg_trip_distance|  total_fare|avg_tip_pct|avg_trip_duration
+|        |              |              |          |              |
+-----+-----+-----+-----+
+|      1|          25.97|6.903424952E8|      25.54|      12.36
+|      2|          2.85|7.848906811E8|      25.18|      14.24
+-----+-----+-----+-----+
+
```

```
In [0]: df_filtered.createOrReplaceTempView("taxi_trips")
```

```
In [0]: from pyspark.sql.functions import when

df_filtered = df_filtered.withColumn(
    "tip_percentage",
    round(when(col("fare_amount") != 0, col("tip_amount") / col("fare_")
)
```

```
In [0]: %sql
SELECT *
FROM taxi_trips
ORDER BY fare_amount DESC
LIMIT 5
```

VendorID	tpep_pickup_datetime	tpep_dropoff_datetime	passenger_count	trip_type
1	2015-01-22T21:12:26.000Z	2015-01-22T21:20:36.000Z	1	
1	2015-01-22T21:12:26.000Z	2015-01-22T21:20:36.000Z	1	
1	2015-01-22T21:12:26.000Z	2015-01-22T21:20:36.000Z	1	
1	2015-01-22T21:12:26.000Z	2015-01-22T21:20:36.000Z	1	
1	2015-01-22T21:12:26.000Z	2015-01-22T21:20:36.000Z	1	

```
In [0]: %sql
```

```
SELECT passenger_count, AVG(fare_amount) AS avg_fare
FROM taxi_trips
GROUP BY passenger_count
ORDER BY avg_fare DESC
```

passenger_count	avg_fare
9	74.41666666666667
8	33.5
7	13.5
2	12.353915909078612
4	12.12988564700566
3	12.06589180693754
5	11.965361292675723
6	11.805331844890977
1	11.74596665091072

```
In [0]: %sql
SELECT VendorID, AVG(try_divide(tip_amount, fare_amount) * 100) AS avg
FROM taxi_trips
GROUP BY VendorID
ORDER BY avg_tip_pct DESC;
```

VendorID	avg_tip_pct
1	25.537875549198585
2	25.179177727395285

```
In [0]: %sql
SELECT VendorID, COUNT(*) AS long_trip_count, AVG(fare_amount) AS avg_
AVG(try_divide(tip_amount, fare_amount) * 100) AS avg_tip_pct_l
FROM taxi_trips
WHERE trip_distance > 10
GROUP BY VendorID
```

VendorID	long_trip_count	avg_fare_long_trip	avg_tip_pct_long_trip
1	2588060	44.878878967257556	34.6448725464399
2	3182850	44.95026347455895	15.70121635420941

```
In [0]: %sql
SELECT CASE
```

```

        WHEN trip_duration < 10 THEN 'Short'
        WHEN trip_duration BETWEEN 10 AND 30 THEN 'Medium'
        ELSE 'Long'
    END AS trip_type,
    COUNT(*) AS trip_count,
    AVG(fare_amount) AS avg_fare,
    AVG(try_divide(tip_amount, fare_amount) * 100) AS avg_tip_pct
FROM taxi_trips
GROUP BY trip_type
ORDER BY trip_type;

```

trip_type	trip_count	avg_fare	avg_tip_pct
Long	5977130	39.052190181575746	24.944778656206473
Medium	56280970	14.931591365605742	22.235290895705212
Short	62044130	6.470373798455994	28.21111438534907

```
In [0]: from pyspark.ml.feature import VectorAssembler
from pyspark.ml.regression import LinearRegression
from pyspark.ml.evaluation import RegressionEvaluator
```

```
In [0]: # Select important numeric columns for prediction
feature_cols = ["trip_distance", "trip_duration", "passenger_count", ""]

# Vector assembler for MLlib
assembler = VectorAssembler(inputCols=feature_cols, outputCol="features")

df_ml = assembler.transform(df_filtered).select("features", "fare_amount")
```

```
In [0]: train_df, test_df = df_ml.randomSplit([0.8, 0.2], seed=42)
```

```
In [0]: lr = LinearRegression(featuresCol="features", labelCol="fare_amount")
lr_model = lr.fit(train_df)
```

```
In [0]: predictions = lr_model.transform(test_df)

evaluator = RegressionEvaluator(
    labelCol="fare_amount", predictionCol="prediction", metricName="rmse")

rmse = evaluator.evaluate(predictions)
r2 = RegressionEvaluator(
    labelCol="fare_amount", predictionCol="prediction", metricName="r2")
    .evaluate(predictions)

print(f"Root Mean Squared Error (RMSE): {rmse}")
print(f"R-squared (R2): {r2}")
```

Root Mean Squared Error (RMSE): 8.457218703796803  
R-squared (R2): 0.2756538664193098

In [0]: `display(predictions.select("features", "fare_amount", "prediction"))`

features
{"type":"1","size":null,"indices":null,"values": ["0.01","0.0333333333333333","5.0","1.851125934112553","0.2416871612181792"]}
{"type":"1","size":null,"indices":null,"values": ["0.01","0.0666666666666667","1.0","0.9","0.2416871612181792"]}
{"type":"1","size":null,"indices":null,"values": ["0.01","0.0666666666666667","1.0","19.66","0.2416871612181792"]}
{"type":"1","size":null,"indices":null,"values": ["0.01","0.0833333333333333","1.0","1.851125934112553","0.2416871612181792"]}
{"type":"1","size":null,"indices":null,"values": ["0.01","0.1","6.0","1.851125934112553","0.2416871612181792"]}
{"type":"1","size":null,"indices":null,"values": ["0.01","0.1166666666666667","1.0","1.851125934112553","0.2416871612181792"]}

In [0]: `df_filtered.explain(True)`

```
-- Parsed Logical Plan --
Project [VendorID#14961, tpep_pickup_datetime#14962, tpep_dropoff_datetime#14963, passenger_count#14964, trip_distance#14965, pickup_longitude#14966, pickup_latitude#14967, RateCodeID#14968, store_and_fwd_flag#14969, dropoff_longitude#14970, dropoff_latitude#14971, payment_type#14972, fare_amount#14973, extra#15158, mta_tax#15160, tip_amount#15152, tolls_amount#15154, improvement_surcharge#15156, total_amount#14979, trip_duration#15901, round(CASE WHEN NOT (fare_amount#14973 = cast(0 as double)) THEN ((tip_amount#15152 / fare_amount#14973) * cast(100 as double)) ELSE cast(null as double) END, 2) AS tip_percentage#17354]
+- Project [VendorID#14961, tpep_pickup_datetime#14962, tpep_dropoff_datetime#14963, passenger_count#14964, trip_distance#14965, pickup_longitude#14966, pickup_latitude#14967, RateCodeID#14968, store_and_fwd_flag#14969, dropoff_longitude#14970, dropoff_latitude#14971, payment_type#14972, fare_amount#14973, extra#15158, mta_tax#15160, tip_amount#15152, tolls_amount#15154, improvement_surcharge#15156, total_amount#14979, trip_duration#15901, round(((tip_amount#15152 / fare_amount#14973) * cast(100 as double)), 2) AS tip_percentage#15903]
+- Project [VendorID#14961, tpep_pickup_datetime#14962, tpep_dropoff_datetime#14963, passenger_count#14964, trip_distance#14965, pickup_longitude#14966, pickup_latitude#14967, RateCodeID#14968, store_and_fwd_flag#14969, dropoff_longitude#14970, dropoff_latitude#14971, payment_type#14972, fare_amount#14973, extra#15158, mta_tax#15160, tip_amount#15152, tolls_amount#15154, improvement_surcharge#15156, total_amount#14979, (cast((unix_timestamp(tpep_dropoff_datetime#14963, yyyy-MM-dd HH:mm:ss, Some(Etc/UTC), true) - unix_timestamp(tpep_pickup_datetime#14962, yyyy-MM-dd HH:mm:ss, Some(Etc/UTC), true)) as double) / cast(60 as double))
```

```
AS trip_duration#15901]
    +- Filter (((pickup_latitude#14967 >= 40.5) AND (pickup_latitude#14967 <= cast(41 as double))) AND ((pickup_longitude#14966 >= -74.5) AND (pickup_longitude#14966 <= -73.5)))
        +- Filter (trip_duration#15162 > cast(0 as double))
            +- Project [VendorID#14961, tpep_pickup_datetime#14962, tpep_dropoff_datetime#14963, passenger_count#14964, trip_distance#14965, pickup_longitude#14966, pickup_latitude#14967, RateCodeID#14968, store_and_fwd_flag#14969, dropoff_longitude#14970, dropoff_latitude#14971, payment_type#14972, fare_amount#14973, extra#15158, mta_tax#15160, tip_amount#15152, tolls_amount#15154, improvement_surcharge#15156, total_amount#14979, (cast((unix_timestamp(tpep_dropoff_datetime#14963, yyyy-MM-dd HH:mm:ss, Some(Etc/UTC), true) - unix_timestamp(tpep_pickup_datetime#14962, yyyy-MM-dd HH:mm:ss, Some(Etc/UTC), true)) as double) / cast(60 as double)) AS trip_duration#15162]
                +- Filter (isnotnull(tpep_pickup_datetime#14962) AND isnotnull(tpep_dropoff_datetime#14963))
                    +- Project [VendorID#14961, tpep_pickup_datetime#14962, tpep_dropoff_datetime#14963, passenger_count#14964, trip_distance#14965, pickup_longitude#14966, pickup_latitude#14967, RateCodeID#14968, store_and_fwd_flag#14969, dropoff_longitude#14970, dropoff_latitude#14971, payment_type#14972, fare_amount#14973, extra#15158, CASE WHEN (isnull(mta_tax#14975) OR (mta_tax#14975 = cast(0 as double))) THEN 0.4987789140890973 ELSE mta_tax#14975 END AS mta_tax#15160, tip_amount#15152, tolls_amount#15154, improvement_surcharge#15156, total_amount#14979]
                        +- Project [VendorID#14961, tpep_pickup_datetime#14962, tpep_dropoff_datetime#14963, passenger_count#14964, trip_distance#14965, pickup_longitude#14966, pickup_latitude#14967, RateCodeID#14968, store_and_fwd_flag#14969, dropoff_longitude#14970, dropoff_latitude#14971, payment_type#14972, fare_amount#14973, CASE WHEN (isnull(extra#14974) OR (extra#14974 = cast(0 as double))) THEN 0.30903683757945827 ELSE extra#14974 END AS extra#15158, mta_tax#14975, tip_amount#15152, tolls_amount#15154, improvement_surcharge#15156, total_amount#14979]
                            +- Project [VendorID#14961, tpep_pickup_datetime#14962, tpep_dropoff_datetime#14963, passenger_count#14964, trip_distance#14965, pickup_longitude#14966, pickup_latitude#14967, RateCodeID#14968, store_and_fwd_flag#14969, dropoff_longitude#14970, dropoff_latitude#14971, payment_type#14972, fare_amount#14973, extra#14974, mta_tax#14975, tip_amount#15152, tolls_amount#15154, CASE WHEN (isnull(improvement_surcharge#14978) OR (improvement_surcharge#14978 = cast(0 as double))) THEN 0.2832208678472794 ELSE improvement_surcharge#14978 END AS improvement_surcharge#15156, total_amount#14979]
                                +- Project [VendorID#14961, tpep_pickup_datetime#14962, tpep_dropoff_datetime#14963, passenger_count#14964, trip_distance#14965, pickup_longitude#14966, pickup_latitude#14967, RateCodeID#14968, store_and_fwd_flag#14969, dropoff_longitude#14970, dropoff_latitude#14971, payment_type#14972, fare_amount#14973, extra#14974, mta_tax#14975, tip_amount#15152, CASE WHEN (isnull(tolls_amount#14977) OR (tolls_amount#14977 = cast(0 as double))) THEN 0.2416871612181792 ELSE tolls_amount#14977 END AS tolls_amount#15154, improvement_surcharge#14978, total_amount#14979]
                                    +- Project [VendorID#14961, tpep_pickup_datetime#14962, tpep_dropoff_datetime#14963, passenger_count#14964, trip_distance#14965, pickup_longitude#14966, pickup_latitude#14967, RateCodeID#14968, store_and_fwd_flag#14969, dropoff_longitude#14970, dropoff_latitude#14971, payment_type#14972, fare_amount#14973, extra#14974, mta_tax#14975, tip_amount#15152, tolls_amount#15154, CASE WHEN (isnull(improvement_surcharge#14978) OR (improvement_surcharge#14978 = cast(0 as double))) THEN 0.2832208678472794 ELSE improvement_surcharge#14978 END AS improvement_surcharge#15156, total_amount#14979]
```

```
atetime#14962, tpep_dropoff_datetime#14963, passenger_count#14964, trip
_distance#14965, pickup_longitude#14966, pickup_latitude#14967, RateCod
eID#14968, store_and_fwd_flag#14969, dropoff_longitude#14970, dropoff_l
atitude#14971, payment_type#14972, fare_amount#14973, extra#14974, mta_
tax#14975, CASE WHEN (isnull(tip_amount#14976) OR (tip_amount#14976 = c
ast(0 as double))) THEN 1.851125934112553 ELSE tip_amount#14976 END AS
tip_amount#15152, tolls_amount#14977, improvement_surcharge#14978, tota
l_amount#14979]
          +- Filter (total_amount#14979 > cast(0
as double))
          +- Filter (fare_amount#14973 > cas
t(0 as double))
          +- Filter (trip_distance#14965 >
cast(0 as double))
          +- Filter (passenger_count#14
964 > 0)
          +- Union false, false
            :- Union false, false
            :  :- Union false, fals
e
            :  :  :- Union false, f
alse
            :  :  :  :- Union fals
e, false
            :  :  :  :  :- Union fa
lse, false
            :  :  :  :  :  :- Union
false, false
            :  :  :  :  :  :  :- Un
ion false, false
            :  :  :  :  :  :  :  :- 
Union false, false
            :  :  :  :  :  :  :  :
:- SubqueryAlias workspace.default.yellow_tripdata_2015_01
            :  :  :  :  :  :  :
:  +- Relation workspace.default.yellow_tripdata_2015_01[VendorID#1496
1,tpep_pickup_datetime#14962,tpep_dropoff_datetime#14963,passenger_coun
t#14964,trip_distance#14965,pickup_longitude#14966,pickup_latitude#1496
7,RateCodeID#14968,store_and_fwd_flag#14969,dropoff_longitude#14970,dro
poff_latitude#14971,payment_type#14972,fare_amount#14973,extra#14974,mt
a_tax#14975,tip_amount#14976,tolls_amount#14977,improvement_surcharge#1
4978,total_amount#14979] parquet
            :  :  :  :  :  :  :
+- SubqueryAlias workspace.default.yellow_tripdata_2015_01
            :  :  :  :  :  :  :
+- Relation workspace.default.yellow_tripdata_2015_01[VendorID#14980,tp
ep_pickup_datetime#14981,tpep_dropoff_datetime#14982,passenger_count#14
983,trip_distance#14984,pickup_longitude#14985,pickup_latitude#14986,Ra
teCodeID#14987,store_and_fwd_flag#14988,dropoff_longitude#14989,dropoff_
latitude#14990,payment_type#14991,fare_amount#14992,extra#14993,mta_ta
x#14994,tip_amount#14995,tolls_amount#14996,improvement_surcharge#149
7,total_amount#14998] parquet
```

```
          : : : : : : : +-  
SubqueryAlias workspace.default.yellow_tripdata_2015_01  
          : : : : : : :  
+- Relation workspace.default.yellow_tripdata_2015_01[VendorID#14999, tp  
ep_pickup_datetime#15000, tpep_dropoff_datetime#15001, passenger_count#15  
002, trip_distance#15003, pickup_longitude#15004, pickup_latitude#15005, Ra  
teCodeID#15006, store_and_fwd_flag#15007, dropoff_longitude#15008, dropoff  
_latitude#15009, payment_type#15010, fare_amount#15011, extra#15012, mta_ta  
x#15013, tip_amount#15014, tolls_amount#15015, improvement_surcharge#1501  
6, total_amount#15017] parquet  
          : : : : : : : +- Su  
bqueryAlias workspace.default.yellow_tripdata_2015_01  
          : : : : : : :  
Relation workspace.default.yellow_tripdata_2015_01[VendorID#15018, tpep_  
pickup_datetime#15019, tpep_dropoff_datetime#15020, passenger_count#1502  
1, trip_distance#15022, pickup_longitude#15023, pickup_latitude#15024, Rate  
CodeID#15025, store_and_fwd_flag#15026, dropoff_longitude#15027, dropoff_l  
atitude#15028, payment_type#15029, fare_amount#15030, extra#15031, mta_tax#  
15032, tip_amount#15033, tolls_amount#15034, improvement_surcharge#15035, t  
otal_amount#15036] parquet  
          : : : : : : : +- Subqu  
eryAlias workspace.default.yellow_tripdata_2015_01  
          : : : : : : :  
+- Relation workspace.default.yellow_tripdata_2015_01[VendorID#15037, tpep_pi  
ckup_datetime#15038, tpep_dropoff_datetime#15039, passenger_count#15040, t  
rip_distance#15041, pickup_longitude#15042, pickup_latitude#15043, RateCod  
eID#15044, store_and_fwd_flag#15045, dropoff_longitude#15046, dropoff_lati  
tude#15047, payment_type#15048, fare_amount#15049, extra#15050, mta_tax#15  
051, tip_amount#15052, tolls_amount#15053, improvement_surcharge#15054, tota  
l_amount#15055] parquet  
          : : : : : : : +- Subquery  
Alias workspace.default.yellow_tripdata_2015_01  
          : : : : : : :  
+- Relation workspace.default.yellow_tripdata_2015_01[VendorID#15056, tpep_picku  
p_datetime#15057, tpep_dropoff_datetime#15058, passenger_count#15059, trip_  
_distance#15060, pickup_longitude#15061, pickup_latitude#15062, RateCodeI  
D#15063, store_and_fwd_flag#15064, dropoff_longitude#15065, dropoff_lati  
tude#15066, payment_type#15067, fare_amount#15068, extra#15069, mta_tax#150  
70, tip_amount#15071, tolls_amount#15072, improvement_surcharge#15073, total  
_amount#15074] parquet  
          : : : : : : : +- SubqueryAli  
as workspace.default.yellow_tripdata_2015_01  
          : : : : : : :  
+- Relation workspace.default.yellow_tripdata_2015_01[VendorID#15075, tpep_pickup_da  
tetime#15076, tpep_dropoff_datetime#15077, passenger_count#15078, trip_dis  
tance#15079, pickup_longitude#15080, pickup_latitude#15081, RateCodeID#150  
82, store_and_fwd_flag#15083, dropoff_longitude#15084, dropoff_latitude#15  
085, payment_type#15086, fare_amount#15087, extra#15088, mta_tax#15089, tip_  
amount#15090, tolls_amount#15091, improvement_surcharge#15092, total_amoun  
t#15093] parquet  
          : : : : : : : +- SubqueryAlias  
workspace.default.yellow_tripdata_2015_01
```

```

: :      +- Relation works
rkspace.default.yellow_tripdata_2015_01[VendorID#15094,tpep_pickup_date
time#15095,tpep_dropoff_datetime#15096,passenger_count#15097,trip_dista
nce#15098,pickup_longitude#15099,pickup_latitude#15100,RateCodeID#1510
1,store_and_fwd_flag#15102,dropoff_longitude#15103,dropoff_latitude#151
04,payment_type#15105,fare_amount#15106,extra#15107,mta_tax#15108,tip_a
mount#15109,tolls_amount#15110,improvement_surcharge#15111,total_amoun
t#15112] parquet
:      +- SubqueryAlias works
kspace.default.yellow_tripdata_2015_01
:      +- Relation works
pace.default.yellow_tripdata_2015_01[VendorID#15113,tpep_pickup_datetim
e#15114,tpep_dropoff_datetime#15115,passenger_count#15116,trip_distanc
e#15117,pickup_longitude#15118,pickup_latitude#15119,RateCodeID#15120,s
tore_and_fwd_flag#15121,dropoff_longitude#15122,dropoff_latitude#15123,
payment_type#15124,fare_amount#15125,extra#15126,mta_tax#15127,tip_amou
nt#15128,tolls_amount#15129,improvement_surcharge#15130,total_amount#15
131] parquet
:      +- SubqueryAlias works
ace.default.yellow_tripdata_2015_01
:      +- Relation workspace
e.default.yellow_tripdata_2015_01[VendorID#15132,tpep_pickup_datetime#1
5133,tpep_dropoff_datetime#15134,passenger_count#15135,trip_distance#15
136,pickup_longitude#15137,pickup_latitude#15138,RateCodeID#15139,store
_and_fwd_flag#15140,dropoff_longitude#15141,dropoff_latitude#15142,paym
ent_type#15143,fare_amount#15144,extra#15145,mta_tax#15146,tip_amount#1
5147,tolls_amount#15148,improvement_surcharge#15149,total_amount#15150]
parquet

== Analyzed Logical Plan ==
VendorID: int, tpep_pickup_datetime: timestamp, tpep_dropoff_datetime:
timestamp, passenger_count: int, trip_distance: double, pickup_longitud
e: double, pickup_latitude: double, RateCodeID: int, store_and_fwd fla
g: string, dropoff_longitude: double, dropoff_latitude: double, payment
_type: int, fare_amount: double, extra: double, mta_tax: double, tip_am
ount: double, tolls_amount: double, improvement_surcharge: double, tota
l_amount: double, trip_duration: double, tip_percentage: double
Project [VendorID#14961, tpep_pickup_datetime#14962, tpep_dropoff_datet
ime#14963, passenger_count#14964, trip_distance#14965, pickup_longitud
e#14966, pickup_latitude#14967, RateCodeID#14968, store_and_fwd_flag#14
969, dropoff_longitude#14970, dropoff_latitude#14971, payment_type#149
2, fare_amount#14973, extra#15158, mta_tax#15160, tip_amount#15152, tol
ls_amount#15154, improvement_surcharge#15156, total_amount#14979, trip_
duration#15901, round(CASE WHEN NOT (fare_amount#14973 = cast(0 as doubl
e)) THEN ((tip_amount#15152 / fare_amount#14973) * cast(100 as doubl
e)) ELSE cast(null as double) END, 2) AS tip_percentage#17354]
+- Project [VendorID#14961, tpep_pickup_datetime#14962, tpep_dropoff_da
tetime#14963, passenger_count#14964, trip_distance#14965, pickup_longit
ude#14966, pickup_latitude#14967, RateCodeID#14968, store_and_fwd_flag#
14969, dropoff_longitude#14970, dropoff_latitude#14971, payment_type#14
972, fare_amount#14973, extra#15158, mta_tax#15160, tip_amount#15152, t
olls_amount#15154, improvement_surcharge#15156, total_amount#14979, tri

```

```
p_duration#15901, round(((tip_amount#15152 / fare_amount#14973) * cast(100 as double)), 2) AS tip_percentage#15903]
    +- Project [VendorID#14961, tpep_pickup_datetime#14962, tpep_dropoff_datetime#14963, passenger_count#14964, trip_distance#14965, pickup_longitude#14966, pickup_latitude#14967, RateCodeID#14968, store_and_fwd_flag#14969, dropoff_longitude#14970, dropoff_latitude#14971, payment_type#14972, fare_amount#14973, extra#15158, mta_tax#15160, tip_amount#15152, tolls_amount#15154, improvement_surcharge#15156, total_amount#14979, (cast((unix_timestamp(tpep_dropoff_datetime#14963, yyyy-MM-dd HH:mm:ss, Some(Etc/UTC), true) - unix_timestamp(tpep_pickup_datetime#14962, yyyy-MM-dd HH:mm:ss, Some(Etc/UTC), true)) as double) / cast(60 as double)) AS trip_duration#15901]
        +- Filter (((pickup_latitude#14967 >= 40.5) AND (pickup_latitude#14967 <= cast(41 as double))) AND ((pickup_longitude#14966 >= -74.5) AND (pickup_longitude#14966 <= -73.5)))
            +- Filter (trip_duration#15162 > cast(0 as double))
                +- Project [VendorID#14961, tpep_pickup_datetime#14962, tpep_dropoff_datetime#14963, passenger_count#14964, trip_distance#14965, pickup_longitude#14966, pickup_latitude#14967, RateCodeID#14968, store_and_fwd_flag#14969, dropoff_longitude#14970, dropoff_latitude#14971, payment_type#14972, fare_amount#14973, extra#15158, mta_tax#15160, tip_amount#15152, tolls_amount#15154, improvement_surcharge#15156, total_amount#14979, (cast((unix_timestamp(tpep_dropoff_datetime#14963, yyyy-MM-dd HH:mm:ss, Some(Etc/UTC), true) - unix_timestamp(tpep_pickup_datetime#14962, yyyy-MM-dd HH:mm:ss, Some(Etc/UTC), true)) as double) / cast(60 as double)) AS trip_duration#15162]
                    +- Filter (isnotnull(tpep_pickup_datetime#14962) AND isnotnull(tpep_dropoff_datetime#14963))
                        +- Project [VendorID#14961, tpep_pickup_datetime#14962, tpep_dropoff_datetime#14963, passenger_count#14964, trip_distance#14965, pickup_longitude#14966, pickup_latitude#14967, RateCodeID#14968, store_and_fwd_flag#14969, dropoff_longitude#14970, dropoff_latitude#14971, payment_type#14972, fare_amount#14973, extra#15158, CASE WHEN (isnull(mta_tax#14975) OR (mta_tax#14975 = cast(0 as double))) THEN 0.4987789140890973 ELSE mta_tax#14975 END AS mta_tax#15160, tip_amount#15152, tolls_amount#15154, improvement_surcharge#15156, total_amount#14979]
                            +- Project [VendorID#14961, tpep_pickup_datetime#14962, tpep_dropoff_datetime#14963, passenger_count#14964, trip_distance#14965, pickup_longitude#14966, pickup_latitude#14967, RateCodeID#14968, store_and_fwd_flag#14969, dropoff_longitude#14970, dropoff_latitude#14971, payment_type#14972, fare_amount#14973, CASE WHEN (isnull(extra#14974) OR (extra#14974 = cast(0 as double))) THEN 0.30903683757945827 ELSE extra#14974 END AS extra#15158, mta_tax#14975, tip_amount#15152, tolls_amount#15154, improvement_surcharge#15156, total_amount#14979]
                                +- Project [VendorID#14961, tpep_pickup_datetime#14962, tpep_dropoff_datetime#14963, passenger_count#14964, trip_distance#14965, pickup_longitude#14966, pickup_latitude#14967, RateCodeID#14968, store_and_fwd_flag#14969, dropoff_longitude#14970, dropoff_latitude#14971, payment_type#14972, fare_amount#14973, extra#14974, mta_tax#14975, tip_amount#15152, tolls_amount#15154, CASE WHEN (isnull(improvement_surcharge#14978) OR (improvement_surcharge#14978 = cast(0 as double))) THEN 0.2832208678472794 ELSE improvement_surcharge#14978 END AS im
```



```
4978, total_amount#14979] parquet
: : : : : : : : :
+- SubqueryAlias workspace.default.yellow_tripdata_2015_01
: : : : : : : :
+- Relation workspace.default.yellow_tripdata_2015_01[VendorID#14980, tp
ep_pickup_datetime#14981, tpep_dropoff_datetime#14982, passenger_count#14
983, trip_distance#14984, pickup_longitude#14985, pickup_latitude#14986, Ra
teCodeID#14987, store_and_fwd_flag#14988, dropoff_longitude#14989, dropoff
_latitude#14990, payment_type#14991, fare_amount#14992, extra#14993, mta_ta
x#14994, tip_amount#14995, tolls_amount#14996, improvement_surcharge#1499
7, total_amount#14998] parquet
: : : : : : : :
+- SubqueryAlias workspace.default.yellow_tripdata_2015_01
: : : : : : :
+- Relation workspace.default.yellow_tripdata_2015_01[VendorID#14999, tp
ep_pickup_datetime#15000, tpep_dropoff_datetime#15001, passenger_count#15
002, trip_distance#15003, pickup_longitude#15004, pickup_latitude#15005, Ra
teCodeID#15006, store_and_fwd_flag#15007, dropoff_longitude#15008, dropoff
_latitude#15009, payment_type#15010, fare_amount#15011, extra#15012, mta_ta
x#15013, tip_amount#15014, tolls_amount#15015, improvement_surcharge#1501
6, total_amount#15017] parquet
: : : : : : : :
+- SubqueryAlias workspace.default.yellow_tripdata_2015_01
: : : : : : :
+- Relation workspace.default.yellow_tripdata_2015_01[VendorID#15018, tpep_
pickup_datetime#15019, tpep_dropoff_datetime#15020, passenger_count#1502
1, trip_distance#15022, pickup_longitude#15023, pickup_latitude#15024, Rate
CodeID#15025, store_and_fwd_flag#15026, dropoff_longitude#15027, dropoff_l
atitude#15028, payment_type#15029, fare_amount#15030, extra#15031, mta_tax#
15032, tip_amount#15033, tolls_amount#15034, improvement_surcharge#15035, t
otal_amount#15036] parquet
: : : : : : :
+- SubqueryAlias workspace.default.yellow_tripdata_2015_01
: : : : : : :
+- Relation workspace.default.yellow_tripdata_2015_01[VendorID#15037, tpep_pi
ckup_datetime#15038, tpep_dropoff_datetime#15039, passenger_count#15040, t
rip_distance#15041, pickup_longitude#15042, pickup_latitude#15043, RateCod
eID#15044, store_and_fwd_flag#15045, dropoff_longitude#15046, dropoff_lati
tude#15047, payment_type#15048, fare_amount#15049, extra#15050, mta_tax#15
051, tip_amount#15052, tolls_amount#15053, improvement_surcharge#15054, tota
l_amount#15055] parquet
: : : : :
+- Subquery
Alias workspace.default.yellow_tripdata_2015_01
: : : : :
+- Relation workspace.default.yellow_tripdata_2015_01[VendorID#15056, tpep_picku
p_datetime#15057, tpep_dropoff_datetime#15058, passenger_count#15059, trip
_distance#15060, pickup_longitude#15061, pickup_latitude#15062, RateCodeI
D#15063, store_and_fwd_flag#15064, dropoff_longitude#15065, dropoff_lati
tude#15066, payment_type#15067, fare_amount#15068, extra#15069, mta_tax#150
70, tip_amount#15071, tolls_amount#15072, improvement_surcharge#15073, total
_amount#15074] parquet
: : : :
+- SubqueryAli
```

```

as workspace.default.yellow_tripdata_2015_01
: : : +- Relation
workspace.default.yellow_tripdata_2015_01[VendorID#15075,tpep_pickup_da
tetime#15076,tpep_dropoff_datetime#15077,passenger_count#15078,trip_dis
tance#15079,pickup_longitude#15080,pickup_latitude#15081,RateCodeID#150
82,store_and_fwd_flag#15083,dropoff_longitude#15084,dropoff_latitude#15
085,payment_type#15086,fare_amount#15087,extra#15088,mta_tax#15089,tip_
amount#15090,tolls_amount#15091,improvement_surcharge#15092,total_amoun
t#15093] parquet
: : +- SubqueryAlias
workspace.default.yellow_tripdata_2015_01
: : +- Relation wo
rkspace.default.yellow_tripdata_2015_01[VendorID#15094,tpep_pickup_date
time#15095,tpep_dropoff_datetime#15096,passenger_count#15097,trip_dista
nce#15098,pickup_longitude#15099,pickup_latitude#15100,RateCodeID#1510
1,store_and_fwd_flag#15102,dropoff_longitude#15103,dropoff_latitude#151
04,payment_type#15105,fare_amount#15106,extra#15107,mta_tax#15108,tip_a
mount#15109,tolls_amount#15110,improvement_surcharge#15111,total_amoun
t#15112] parquet
: +- SubqueryAlias wor
kspace.default.yellow_tripdata_2015_01
: +- Relation works
pace.default.yellow_tripdata_2015_01[VendorID#15113,tpep_pickup_datetim
e#15114,tpep_dropoff_datetime#15115,passenger_count#15116,trip_distanc
e#15117,pickup_longitude#15118,pickup_latitude#15119,RateCodeID#15120,s
tore_and_fwd_flag#15121,dropoff_longitude#15122,dropoff_latitude#15123,
payment_type#15124,fare_amount#15125,extra#15126,mta_tax#15127,tip_amou
nt#15128,tolls_amount#15129,improvement_surcharge#15130,total_amount#15
131] parquet
+- SubqueryAlias worksp
ace.default.yellow_tripdata_2015_01
+- Relation workspac
e.default.yellow_tripdata_2015_01[VendorID#15132,tpep_pickup_datetime#1
5133,tpep_dropoff_datetime#15134,passenger_count#15135,trip_distance#15
136,pickup_longitude#15137,pickup_latitude#15138,RateCodeID#15139,store
_and_fwd_flag#15140,dropoff_longitude#15141,dropoff_latitude#15142,paym
ent_type#15143,fare_amount#15144,extra#15145,mta_tax#15146,tip_amount#1
5147,tolls_amount#15148,improvement_surcharge#15149,total_amount#15150]
parquet

== Optimized Logical Plan ==
Union false, false
:- Project [VendorID#14961, tpep_pickup_datetime#14962, tpep_dropoff_da
tetime#14963, passenger_count#14964, trip_distance#14965, pickup_longit
ude#14966, pickup_latitude#14967, RateCodeID#14968, store_and_fwd_flag#
14969, dropoff_longitude#14970, dropoff_latitude#14971, payment_type#14
972, fare_amount#14973, CASE WHEN (isnull(extra#14974) OR (extra#14974
= 0.0)) THEN 0.30903683757945827 ELSE extra#14974 END AS extra#15158, C
ASE WHEN (isnull(mta_tax#14975) OR (mta_tax#14975 = 0.0)) THEN 0.498778
9140890973 ELSE mta_tax#14975 END AS mta_tax#15160, tip_amount#15152, C
ASE WHEN (isnull(tolls_amount#14977) OR (tolls_amount#14977 = 0.0)) THE
N 0.2416871612181792 ELSE tolls_amount#14977 END AS tolls_amount#15154,
```

```

CASE WHEN (isnull(improvement_surcharge#14978) OR (improvement_surcharge#14978 = 0.0)) THEN 0.2832208678472794 ELSE improvement_surcharge#14978 END AS improvement_surcharge#15156, total_amount#14979, (cast((unix_timestamp(tpep_dropoff_datetime#14963, yyyy-MM-dd HH:mm:ss, Some(Etc/UTC), true) - unix_timestamp(tpep_pickup_datetime#14962, yyyy-MM-dd HH:mm:ss, Some(Etc/UTC), true)) as double) / 60.0) AS trip_duration#15901, CASE WHEN NOT (fare_amount#14973 = 0.0) THEN round(((tip_amount#15152 / fare_amount#14973) * 100.0), 2) END AS tip_percentage#17354]
: +- Project [VendorID#14961, tpep_pickup_datetime#14962, tpep_dropoff_datetime#14963, passenger_count#14964, trip_distance#14965, pickup_longitude#14966, pickup_latitude#14967, RateCodeID#14968, store_and_fwd_flag#14969, dropoff_longitude#14970, dropoff_latitude#14971, payment_type#14972, fare_amount#14973, extra#14974, mta_tax#14975, CASE WHEN (isnull(tip_amount#14976) OR (tip_amount#14976 = 0.0)) THEN 1.85112593411255 ELSE tip_amount#14976 END AS tip_amount#15152, tolls_amount#14977, improvement_surcharge#14978, total_amount#14979]
:     +- Filter (((((((((isnotnull(passenger_count#14964) AND isnotnull(trip_distance#14965)) AND isnotnull(fare_amount#14973)) AND isnotnull(total_amount#14979)) AND isnotnull(pickup_latitude#14967)) AND isnotnull(pickup_longitude#14966)) AND isnotnull(tpep_pickup_datetime#14962)) AND isnotnull(tpep_dropoff_datetime#14963)) AND (passenger_count#14964 > 0)) AND (trip_distance#14965 > 0.0)) AND (fare_amount#14973 > 0.0)) AND (total_amount#14979 > 0.0)) AND (pickup_latitude#14967 >= 40.5)) AND (pickup_latitude#14967 <= 41.0)) AND (pickup_longitude#14966 >= -74.5)) AND (pickup_longitude#14966 <= -73.5)) AND ((cast((unix_timestamp(tpep_dropoff_datetime#14963, yyyy-MM-dd HH:mm:ss, Some(Etc/UTC), true) - unix_timestamp(tpep_pickup_datetime#14962, yyyy-MM-dd HH:mm:ss, Some(Etc/UTC), true)) as double) / 60.0) > 0.0))
:         +- Relation workspace.default.yellow_tripdata_2015_01[VendorID#14961,tpep_pickup_datetime#14962,tpep_dropoff_datetime#14963,passenger_count#14964,trip_distance#14965,pickup_longitude#14966,pickup_latitude#14967,RateCodeID#14968,store_and_fwd_flag#14969,dropoff_longitude#14970,dropoff_latitude#14971,payment_type#14972,fare_amount#14973,extra#14974,mta_tax#14975,tip_amount#14976,tolls_amount#14977,improvement_surcharge#14978,total_amount#14979] parquet
:- Project [VendorID#14980, tpep_pickup_datetime#14981, tpep_dropoff_datetime#14982, passenger_count#14983, trip_distance#14984, pickup_longitude#14985, pickup_latitude#14986, RateCodeID#14987, store_and_fwd_flag#14988, dropoff_longitude#14989, dropoff_latitude#14990, payment_type#14991, fare_amount#14992, CASE WHEN (isnull(extra#14993) OR (extra#14993 = 0.0)) THEN 0.30903683757945827 ELSE extra#14993 END AS extra#24260, CASE WHEN (isnull(mta_tax#14994) OR (mta_tax#14994 = 0.0)) THEN 0.4987789140890973 ELSE mta_tax#14994 END AS mta_tax#24261, tip_amount#24251, CASE WHEN (isnull(tolls_amount#14996) OR (tolls_amount#14996 = 0.0)) THEN 0.2416871612181792 ELSE tolls_amount#14996 END AS tolls_amount#24262, CASE WHEN (isnull(improvement_surcharge#14997) OR (improvement_surcharge#14997 = 0.0)) THEN 0.2832208678472794 ELSE improvement_surcharge#14997 END AS improvement_surcharge#24263, total_amount#14998, (cast((unix_timestamp(tpep_dropoff_datetime#14982, yyyy-MM-dd HH:mm:ss, Some(Etc/UTC), true) - unix_timestamp(tpep_pickup_datetime#14981, yyyy-MM-dd HH:mm:ss, Some(Etc/UTC), true)) as double) / 60.0) AS trip_duration#24264, CASE WHEN NOT (fare_amount#14992 = 0.0) THEN round(((tip_amount#24251 /

```

```
fare_amount#14992) * 100.0), 2) END AS tip_percentage#24265]
: +- Project [VendorID#14980, tpep_pickup_datetime#14981, tpep_dropoff_datetime#14982, passenger_count#14983, trip_distance#14984, pickup_longitude#14985, pickup_latitude#14986, RateCodeID#14987, store_and_fwd_flag#14988, dropoff_longitude#14989, dropoff_latitude#14990, payment_type#14991, fare_amount#14992, extra#14993, mta_tax#14994, CASE WHEN (isnull(tip_amount#14995) OR (tip_amount#14995 = 0.0)) THEN 1.85112593411255
3 ELSE tip_amount#14995 END AS tip_amount#24251, tolls_amount#14996, improvement_surcharge#14997, total_amount#14998]
:     +- Filter (((((((((isnotnull(passenger_count#14983) AND isnotnull(trip_distance#14984)) AND isnotnull(fare_amount#14992)) AND isnotnull(total_amount#14998)) AND isnotnull(pickup_latitude#14986)) AND isnotnull(pickup_longitude#14985)) AND isnotnull(tpep_pickup_datetime#14981)) AND isnotnull(tpep_dropoff_datetime#14982)) AND (passenger_count#14983 > 0)) AND (trip_distance#14984 > 0.0)) AND (fare_amount#14992 > 0.0)) AND (total_amount#14998 > 0.0)) AND (pickup_latitude#14986 >= 40.5)) AND (pickup_latitude#14986 <= 41.0)) AND (pickup_longitude#14985 >= -74.5)) AND (pickup_longitude#14985 <= -73.5)) AND ((cast((unix_timestamp(tpep_dropoff_datetime#14982, yyyy-MM-dd HH:mm:ss, Some(Etc/UTC), true) - unix_timestamp(tpep_pickup_datetime#14981, yyyy-MM-dd HH:mm:ss, Some(Etc/UTC), true)) as double) / 60.0) > 0.0))
:         +- Relation workspace.default.yellow_tripdata_2015_01[VendorID#14980,tpep_pickup_datetime#14981,tpep_dropoff_datetime#14982,passenger_count#14983,trip_distance#14984,pickup_longitude#14985,pickup_latitude#14986,RateCodeID#14987,store_and_fwd_flag#14988,dropoff_longitude#14989,dropoff_latitude#14990,payment_type#14991,fare_amount#14992,extra#14993,mta_tax#14994,tip_amount#14995,tolls_amount#14996,improvement_surcharge#14997,total_amount#14998] parquet
:- Project [VendorID#14999, tpep_pickup_datetime#15000, tpep_dropoff_datetime#15001, passenger_count#15002, trip_distance#15003, pickup_longitude#15004, pickup_latitude#15005, RateCodeID#15006, store_and_fwd_flag#15007, dropoff_longitude#15008, dropoff_latitude#15009, payment_type#15010, fare_amount#15011, CASE WHEN (isnull(extra#15012) OR (extra#15012 = 0.0)) THEN 0.30903683757945827 ELSE extra#15012 END AS extra#24266, CASE WHEN (isnull(mta_tax#15013) OR (mta_tax#15013 = 0.0)) THEN 0.4987789140890973 ELSE mta_tax#15013 END AS mta_tax#24267, tip_amount#24252, CASE WHEN (isnull(tolls_amount#15015) OR (tolls_amount#15015 = 0.0)) THEN 0.2416871612181792 ELSE tolls_amount#15015 END AS tolls_amount#24268, CASE WHEN (isnull(improvement_surcharge#15016) OR (improvement_surcharge#15016 = 0.0)) THEN 0.2832208678472794 ELSE improvement_surcharge#15016 END AS improvement_surcharge#24269, total_amount#15017, (cast((unix_timestamp(tpep_dropoff_datetime#15001, yyyy-MM-dd HH:mm:ss, Some(Etc/UTC), true) - unix_timestamp(tpep_pickup_datetime#15000, yyyy-MM-dd HH:mm:ss, Some(Etc/UTC), true)) as double) / 60.0) AS trip_duration#24270, CASE WHEN NOT (fare_amount#15011 = 0.0) THEN round(((tip_amount#24252 / fare_amount#15011) * 100.0), 2) END AS tip_percentage#24271]
: +- Project [VendorID#14999, tpep_pickup_datetime#15000, tpep_dropoff_datetime#15001, passenger_count#15002, trip_distance#15003, pickup_longitude#15004, pickup_latitude#15005, RateCodeID#15006, store_and_fwd_flag#15007, dropoff_longitude#15008, dropoff_latitude#15009, payment_type#15010, fare_amount#15011, extra#15012, mta_tax#15013, CASE WHEN (isnull(tip_amount#15014) OR (tip_amount#15014 = 0.0)) THEN 1.85112593411255
```

```

3 ELSE tip_amount#15014 END AS tip_amount#24252, tolls_amount#15015, im
provement_surcharge#15016, total_amount#15017]
:      +- Filter (((((((((isnotnull(passenger_count#15002) AND is
notnull(trip_distance#15003)) AND isnotnull(fare_amount#15011)) AND is
notnull(total_amount#15017)) AND isnotnull(pickup_latitude#15005)) AND i
snotnull(pickup_longitude#15004)) AND isnotnull(tpep_pickup_datetime#15
000)) AND isnotnull(tpep_dropoff_datetime#15001)) AND (passenger_count#
15002 > 0)) AND (trip_distance#15003 > 0.0)) AND (fare_amount#15011 >
0.0)) AND (total_amount#15017 > 0.0)) AND (pickup_latitude#15005 >=
40.
5)) AND (pickup_latitude#15005 <= 41.0)) AND (pickup_longitude#15004 >=
-74.5)) AND (pickup_longitude#15004 <= -73.5)) AND ((cast((unix_timesta
mp(tpep_dropoff_datetime#15001, yyyy-MM-dd HH:mm:ss, Some(Etc/UTC), tru
e) - unix_timestamp(tpep_pickup_datetime#15000, yyyy-MM-dd HH:mm:ss, So
me(Etc/UTC), true)) as double) / 60.0) > 0.0))
:          +- Relation workspace.default.yellow_tripdata_2015_01[VendorI
D#14999, tpep_pickup_datetime#15000, tpep_dropoff_datetime#15001, passenge
r_count#15002, trip_distance#15003, pickup_longitude#15004, pickup_latitud
e#15005, RateCodeID#15006, store_and_fwd_flag#15007, dropoff_longitude#15
08, dropoff_latitude#15009, payment_type#15010, fare_amount#15011, extra#15
012, mta_tax#15013, tip_amount#15014, tolls_amount#15015, improvement_surch
arge#15016, total_amount#15017] parquet
:- Project [VendorID#15018, tpep_pickup_datetime#15019, tpep_dropoff_da
tetime#15020, passenger_count#15021, trip_distance#15022, pickup_longit
ude#15023, pickup_latitude#15024, RateCodeID#15025, store_and_fwd_flag#15
026, dropoff_longitude#15027, dropoff_latitude#15028, payment_type#15
029, fare_amount#15030, CASE WHEN (isnull(extra#15031) OR (extra#15031
= 0.0)) THEN 0.30903683757945827 ELSE extra#15031 END AS extra#24272, C
ASE WHEN (isnull(mta_tax#15032) OR (mta_tax#15032 = 0.0)) THEN 0.498778
9140890973 ELSE mta_tax#15032 END AS mta_tax#24273, tip_amount#24253, C
ASE WHEN (isnull(tolls_amount#15034) OR (tolls_amount#15034 = 0.0)) THE
N 0.2416871612181792 ELSE tolls_amount#15034 END AS tolls_amount#24274,
CASE WHEN (isnull(improvement_surcharge#15035) OR (improvement_surcharg
e#15035 = 0.0)) THEN 0.2832208678472794 ELSE improvement_surcharge#1503
5 END AS improvement_surcharge#24275, total_amount#15036, (cast((unix_t
imestamp(tpep_dropoff_datetime#15020, yyyy-MM-dd HH:mm:ss, Some(Etc/UT
C), true) - unix_timestamp(tpep_pickup_datetime#15019, yyyy-MM-dd HH:m
m:ss, Some(Etc/UTC), true)) as double) / 60.0) AS trip_duration#24276,
CASE WHEN NOT (fare_amount#15030 = 0.0) THEN round(((tip_amount#24253 /
fare_amount#15030) * 100.0), 2) END AS tip_percentage#24277]
:  +- Project [VendorID#15018, tpep_pickup_datetime#15019, tpep_dropoff
_datetime#15020, passenger_count#15021, trip_distance#15022, pickup_lon
gitude#15023, pickup_latitude#15024, RateCodeID#15025, store_and_fwd_fl
ag#15026, dropoff_longitude#15027, dropoff_latitude#15028, payment_typ
e#15029, fare_amount#15030, extra#15031, mta_tax#15032, CASE WHEN (isn
ull(tip_amount#15033) OR (tip_amount#15033 = 0.0)) THEN 1.85112593411255
3 ELSE tip_amount#15033 END AS tip_amount#24253, tolls_amount#15034, im
provement_surcharge#15035, total_amount#15036]
:      +- Filter (((((((((isnotnull(passenger_count#15021) AND is
notnull(trip_distance#15022)) AND isnotnull(fare_amount#15030)) AND is
notnull(total_amount#15036)) AND isnotnull(pickup_latitude#15024)) AND i
snotnull(pickup_longitude#15023)) AND isnotnull(tpep_pickup_datetime#15
019)) AND isnotnull(tpep_dropoff_datetime#15020)) AND (passenger_count#

```

```
15021 > 0)) AND (trip_distance#15022 > 0.0)) AND (fare_amount#15030 > 0.0)) AND (total_amount#15036 > 0.0)) AND (pickup_latitude#15024 >= 40.5)) AND (pickup_latitude#15024 <= 41.0)) AND (pickup_longitude#15023 >= -74.5)) AND (pickup_longitude#15023 <= -73.5)) AND ((cast((unix_timestamp(tpep_dropoff_datetime#15020, yyyy-MM-dd HH:mm:ss, Some(Etc/UTC), true) - unix_timestamp(tpep_pickup_datetime#15019, yyyy-MM-dd HH:mm:ss, Some(Etc/UTC), true)) as double) / 60.0) > 0.0))  
:      +- Relation workspace.default.yellow_tripdata_2015_01[VendorID#15018, tpep_pickup_datetime#15019, tpep_dropoff_datetime#15020, passenger_count#15021, trip_distance#15022, pickup_longitude#15023, pickup_latitude#15024, RateCodeID#15025, store_and_fwd_flag#15026, dropoff_longitude#15027, dropoff_latitude#15028, payment_type#15029, fare_amount#15030, extra#15031, mta_tax#15032, tip_amount#15033, tolls_amount#15034, improvement_surcharge#15035, total_amount#15036] parquet  
:- Project [VendorID#15037, tpep_pickup_datetime#15038, tpep_dropoff_datetime#15039, passenger_count#15040, trip_distance#15041, pickup_longitude#15042, pickup_latitude#15043, RateCodeID#15044, store_and_fwd_flag#15045, dropoff_longitude#15046, dropoff_latitude#15047, payment_type#15048, fare_amount#15049, CASE WHEN (isnull(extra#15050) OR (extra#15050 = 0.0)) THEN 0.30903683757945827 ELSE extra#15050 END AS extra#24278, CASE WHEN (isnull(mta_tax#15051) OR (mta_tax#15051 = 0.0)) THEN 0.4987789140890973 ELSE mta_tax#15051 END AS mta_tax#24279, tip_amount#24254, CASE WHEN (isnull(tolls_amount#15053) OR (tolls_amount#15053 = 0.0)) THEN 0.2416871612181792 ELSE tolls_amount#15053 END AS tolls_amount#24280, CASE WHEN (isnull(improvement_surcharge#15054) OR (improvement_surcharge#15054 = 0.0)) THEN 0.2832208678472794 ELSE improvement_surcharge#15054 END AS improvement_surcharge#24281, total_amount#15055, (cast((unix_timestamp(tpep_dropoff_datetime#15039, yyyy-MM-dd HH:mm:ss, Some(Etc/UTC), true) - unix_timestamp(tpep_pickup_datetime#15038, yyyy-MM-dd HH:mm:ss, Some(Etc/UTC), true)) as double) / 60.0) AS trip_duration#24282, CASE WHEN NOT (fare_amount#15049 = 0.0) THEN round(((tip_amount#24254 / fare_amount#15049) * 100.0), 2) END AS tip_percentage#24283]  
:      +- Project [VendorID#15037, tpep_pickup_datetime#15038, tpep_dropoff_datetime#15039, passenger_count#15040, trip_distance#15041, pickup_longitude#15042, pickup_latitude#15043, RateCodeID#15044, store_and_fwd_flag#15045, dropoff_longitude#15046, dropoff_latitude#15047, payment_type#15048, fare_amount#15049, extra#15050, mta_tax#15051, CASE WHEN (isnull(tip_amount#15052) OR (tip_amount#15052 = 0.0)) THEN 1.851125934112553 ELSE tip_amount#15052 END AS tip_amount#24254, tolls_amount#15053, improvement_surcharge#15054, total_amount#15055]  
:      +- Filter (((((((((isnotnull(passenger_count#15040) AND isnotnull(trip_distance#15041)) AND isnotnull(fare_amount#15049)) AND isnotnull(total_amount#15055)) AND isnotnull(pickup_latitude#15043)) AND isnotnull(pickup_longitude#15042)) AND isnotnull(tpep_pickup_datetime#15038)) AND isnotnull(tpep_dropoff_datetime#15039)) AND (passenger_count#15040 > 0)) AND (trip_distance#15041 > 0.0)) AND (fare_amount#15049 > 0.0)) AND (total_amount#15055 > 0.0)) AND (pickup_latitude#15043 >= 40.5)) AND (pickup_latitude#15043 <= 41.0)) AND (pickup_longitude#15042 >= -74.5)) AND (pickup_longitude#15042 <= -73.5)) AND ((cast((unix_timestamp(tpep_dropoff_datetime#15039, yyyy-MM-dd HH:mm:ss, Some(Etc/UTC), true) - unix_timestamp(tpep_pickup_datetime#15038, yyyy-MM-dd HH:mm:ss, Some(Etc/UTC), true)) as double) / 60.0) > 0.0))
```

```
:      +- Relation workspace.default.yellow_tripdata_2015_01[VendorI
D#15037,tpep_pickup_datetime#15038,tpep_dropoff_datetime#15039,passenge
r_count#15040,trip_distance#15041,pickup_longitude#15042,pickup_latitud
e#15043,RateCodeID#15044,store_and_fwd_flag#15045,dropoff_longitude#150
46,dropoff_latitude#15047,payment_type#15048,fare_amount#15049,extra#15
050,mta_tax#15051,tip_amount#15052,tolls_amount#15053,improvement_surch
arge#15054,total_amount#15055] parquet
:- Project [VendorID#15056, tpep_pickup_datetime#15057, tpep_dropoff_da
tetime#15058, passenger_count#15059, trip_distance#15060, pickup_longit
ude#15061, pickup_latitude#15062, RateCodeID#15063, store_and_fwd_flag#
15064, dropoff_longitude#15065, dropoff_latitude#15066, payment_type#15
067, fare_amount#15068, CASE WHEN (isnull(extra#15069) OR (extra#15069
= 0.0)) THEN 0.30903683757945827 ELSE extra#15069 END AS extra#24284, C
ASE WHEN (isnull(mta_tax#15070) OR (mta_tax#15070 = 0.0)) THEN 0.498778
9140890973 ELSE mta_tax#15070 END AS mta_tax#24285, tip_amount#24255, C
ASE WHEN (isnull(tolls_amount#15072) OR (tolls_amount#15072 = 0.0)) THE
N 0.2416871612181792 ELSE tolls_amount#15072 END AS tolls_amount#24286,
CASE WHEN (isnull(improvement_surcharge#15073) OR (improvement_surcharg
e#15073 = 0.0)) THEN 0.2832208678472794 ELSE improvement_surcharge#1507
3 END AS improvement_surcharge#24287, total_amount#15074, (cast((unix_t
imestamp(tpep_dropoff_datetime#15058, yyyy-MM-dd HH:mm:ss, Some(Etc/UT
C), true) - unix_timestamp(tpep_pickup_datetime#15057, yyyy-MM-dd HH:m
m:ss, Some(Etc/UTC), true)) as double) / 60.0) AS trip_duration#24288,
CASE WHEN NOT (fare_amount#15068 = 0.0) THEN round(((tip_amount#24255 /
fare_amount#15068) * 100.0), 2) END AS tip_percentage#24289]
:  +- Project [VendorID#15056, tpep_pickup_datetime#15057, tpep_dropoff
_datetime#15058, passenger_count#15059, trip_distance#15060, pickup_lon
gitude#15061, pickup_latitude#15062, RateCodeID#15063, store_and_fwd_fl
ag#15064, dropoff_longitude#15065, dropoff_latitude#15066, payment_typ
e#15067, fare_amount#15068, extra#15069, mta_tax#15070, CASE WHEN (isnu
ll(tip_amount#15071) OR (tip_amount#15071 = 0.0)) THEN 1.85112593411255
3 ELSE tip_amount#15071 END AS tip_amount#24255, tolls_amount#15072, im
provement_surcharge#15073, total_amount#15074]
:      +- Filter (((((((((isnotnull(passenger_count#15059) AND is
notnull(trip_distance#15060)) AND isnotnull(fare_amount#15068)) AND isn
otnull(total_amount#15074)) AND isnotnull(pickup_latitude#15062)) AND i
snotnull(pickup_longitude#15061)) AND isnotnull(tpep_pickup_datetime#15
057)) AND isnotnull(tpep_dropoff_datetime#15058)) AND (passenger_count#
15059 > 0)) AND (trip_distance#15060 > 0.0)) AND (fare_amount#15068 >
0.0)) AND (total_amount#15074 > 0.0)) AND (pickup_latitude#15062 >=
40.5)) AND (pickup_latitude#15062 <= 41.0)) AND (pickup_longitude#15061 >=
-74.5)) AND (pickup_longitude#15061 <= -73.5)) AND ((cast((unix_timesta
mp(tpep_dropoff_datetime#15058, yyyy-MM-dd HH:mm:ss, Some(Etc/UTC), tru
e) - unix_timestamp(tpep_pickup_datetime#15057, yyyy-MM-dd HH:mm:ss, So
me(Etc/UTC), true)) as double) / 60.0) > 0.0))
:      +- Relation workspace.default.yellow_tripdata_2015_01[VendorI
D#15056,tpep_pickup_datetime#15057,tpep_dropoff_datetime#15058,passenge
r_count#15059,trip_distance#15060,pickup_longitude#15061,pickup_latitud
e#15062,RateCodeID#15063,store_and_fwd_flag#15064,dropoff_longitude#150
65,dropoff_latitude#15066,payment_type#15067,fare_amount#15068,extra#15
069,mta_tax#15070,tip_amount#15071,tolls_amount#15072,improvement_surch
arge#15073,total_amount#15074] parquet
```

```

:- Project [VendorID#15075, tpep_pickup_datetime#15076, tpep_dropoff_datetime#15077, passenger_count#15078, trip_distance#15079, pickup_longitude#15080, pickup_latitude#15081, RateCodeID#15082, store_and_fwd_flag#15083, dropoff_longitude#15084, dropoff_latitude#15085, payment_type#15086, fare_amount#15087, CASE WHEN (isnull(extra#15088) OR (extra#15088 = 0.0)) THEN 0.30903683757945827 ELSE extra#15088 END AS extra#24290, CASE WHEN (isnull(mta_tax#15089) OR (mta_tax#15089 = 0.0)) THEN 0.4987789140890973 ELSE mta_tax#15089 END AS mta_tax#24291, tip_amount#24256, CASE WHEN (isnull(tolls_amount#15091) OR (tolls_amount#15091 = 0.0)) THEN 0.2416871612181792 ELSE tolls_amount#15091 END AS tolls_amount#24292, CASE WHEN (isnull(improvement_surcharge#15092) OR (improvement_surcharge#15092 = 0.0)) THEN 0.2832208678472794 ELSE improvement_surcharge#15092 END AS improvement_surcharge#24293, total_amount#15093, (cast((unix_timestamp(tpep_dropoff_datetime#15077, yyyy-MM-dd HH:mm:ss, Some(Etc/UTC), true) - unix_timestamp(tpep_pickup_datetime#15076, yyyy-MM-dd HH:mm:ss, Some(Etc/UTC), true)) as double) / 60.0) AS trip_duration#24294, CASE WHEN NOT (fare_amount#15087 = 0.0) THEN round(((tip_amount#24256 / fare_amount#15087) * 100.0), 2) END AS tip_percentage#24295]
: +- Project [VendorID#15075, tpep_pickup_datetime#15076, tpep_dropoff_datetime#15077, passenger_count#15078, trip_distance#15079, pickup_longitude#15080, pickup_latitude#15081, RateCodeID#15082, store_and_fwd_flag#15083, dropoff_longitude#15084, dropoff_latitude#15085, payment_type#15086, fare_amount#15087, extra#15088, mta_tax#15089, CASE WHEN (isnull(tip_amount#15090) OR (tip_amount#15090 = 0.0)) THEN 1.851125934112553 ELSE tip_amount#15090 END AS tip_amount#24256, tolls_amount#15091, improvement_surcharge#15092, total_amount#15093]
:     +- Filter (((((((((isnotnull(passenger_count#15078) AND isnotnull(trip_distance#15079)) AND isnotnull(fare_amount#15087)) AND isnotnull(total_amount#15093)) AND isnotnull(pickup_latitude#15081)) AND isnotnull(pickup_longitude#15080)) AND isnotnull(tpep_pickup_datetime#15076)) AND isnotnull(tpep_dropoff_datetime#15077)) AND (passenger_count#15078 > 0)) AND (trip_distance#15079 > 0.0)) AND (fare_amount#15087 > 0.0)) AND (total_amount#15093 > 0.0)) AND (pickup_latitude#15081 >= 40.5)) AND (pickup_latitude#15081 <= 41.0)) AND (pickup_longitude#15080 >= -74.5)) AND (pickup_longitude#15080 <= -73.5)) AND ((cast((unix_timestamp(tpep_dropoff_datetime#15077, yyyy-MM-dd HH:mm:ss, Some(Etc/UTC), true) - unix_timestamp(tpep_pickup_datetime#15076, yyyy-MM-dd HH:mm:ss, Some(Etc/UTC), true)) as double) / 60.0) > 0.0))
:         +- Relation workspace.default.yellow_tripdata_2015_01[VendorID#15075,tpep_pickup_datetime#15076,tpep_dropoff_datetime#15077,passenger_count#15078,trip_distance#15079,pickup_longitude#15080,pickup_latitude#15081,RateCodeID#15082,store_and_fwd_flag#15083,dropoff_longitude#15084,dropoff_latitude#15085,payment_type#15086,fare_amount#15087,extra#15088,mta_tax#15089,tip_amount#15090,tolls_amount#15091,improvement_surcharge#15092,total_amount#15093] parquet
:- Project [VendorID#15094, tpep_pickup_datetime#15095, tpep_dropoff_datetime#15096, passenger_count#15097, trip_distance#15098, pickup_longitude#15099, pickup_latitude#15100, RateCodeID#15101, store_and_fwd_flag#15102, dropoff_longitude#15103, dropoff_latitude#15104, payment_type#15105, fare_amount#15106, CASE WHEN (isnull(extra#15107) OR (extra#15107 = 0.0)) THEN 0.30903683757945827 ELSE extra#15107 END AS extra#24296, CASE WHEN (isnull(mta_tax#15108) OR (mta_tax#15108 = 0.0)) THEN 0.498778

```



```

C), true) - unix_timestamp(tpep_pickup_datetime#15114, yyyy-MM-dd HH:mm:ss, Some(Etc/UTC), true)) as double) / 60.0) AS trip_duration#24306,
CASE WHEN NOT (fare_amount#15125 = 0.0) THEN round(((tip_amount#24258 / fare_amount#15125) * 100.0), 2) END AS tip_percentage#24307]
: +- Project [VendorID#15113, tpep_pickup_datetime#15114, tpep_dropoff_datetime#15115, passenger_count#15116, trip_distance#15117, pickup_longitude#15118, pickup_latitude#15119, RateCodeID#15120, store_and_fwd_flag#15121, dropoff_longitude#15122, dropoff_latitude#15123, payment_type#15124, fare_amount#15125, extra#15126, mta_tax#15127, CASE WHEN (isnull(tip_amount#15128) OR (tip_amount#15128 = 0.0)) THEN 1.85112593411255 ELSE tip_amount#15128 END AS tip_amount#24258, tolls_amount#15129, improvement_surcharge#15130, total_amount#15131]
:     +- Filter (((((((((isnotnull(passenger_count#15116) AND isnotnull(trip_distance#15117)) AND isnotnull(fare_amount#15125)) AND isnotnull(total_amount#15131)) AND isnotnull(pickup_latitude#15119)) AND isnotnull(pickup_longitude#15118)) AND isnotnull(tpep_pickup_datetime#15114)) AND isnotnull(tpep_dropoff_datetime#15115)) AND (passenger_count#15116 > 0)) AND (trip_distance#15117 > 0.0)) AND (fare_amount#15125 > 0.0)) AND (total_amount#15131 > 0.0)) AND (pickup_latitude#15119 >= 40.5)) AND (pickup_latitude#15119 <= 41.0)) AND (pickup_longitude#15118 >= -74.5)) AND (pickup_longitude#15118 <= -73.5)) AND ((cast((unix_timestamp(tpep_dropoff_datetime#15115, yyyy-MM-dd HH:mm:ss, Some(Etc/UTC), true) - unix_timestamp(tpep_pickup_datetime#15114, yyyy-MM-dd HH:mm:ss, Some(Etc/UTC), true)) as double) / 60.0) > 0.0))
:         +- Relation workspace.default.yellow_tripdata_2015_01[VendorID#15113, tpep_pickup_datetime#15114, tpep_dropoff_datetime#15115, passenger_count#15116, trip_distance#15117, pickup_longitude#15118, pickup_latitude#15119, RateCodeID#15120, store_and_fwd_flag#15121, dropoff_longitude#15122, dropoff_latitude#15123, payment_type#15124, fare_amount#15125, extra#15126, mta_tax#15127, tip_amount#15128, tolls_amount#15129, improvement_surcharge#15130, total_amount#15131] parquet
+- Project [VendorID#15132, tpep_pickup_datetime#15133, tpep_dropoff_datetime#15134, passenger_count#15135, trip_distance#15136, pickup_longitude#15137, pickup_latitude#15138, RateCodeID#15139, store_and_fwd_flag#15140, dropoff_longitude#15141, dropoff_latitude#15142, payment_type#15143, fare_amount#15144, CASE WHEN (isnull(extra#15145) OR (extra#15145 = 0.0)) THEN 0.30903683757945827 ELSE extra#15145 END AS extra#24308, CASE WHEN (isnull(mta_tax#15146) OR (mta_tax#15146 = 0.0)) THEN 0.4987789140890973 ELSE mta_tax#15146 END AS mta_tax#24309, tip_amount#24259, CASE WHEN (isnull(tolls_amount#15148) OR (tolls_amount#15148 = 0.0)) THEN 0.2416871612181792 ELSE tolls_amount#15148 END AS tolls_amount#24310, CASE WHEN (isnull(improvement_surcharge#15149) OR (improvement_surcharge#15149 = 0.0)) THEN 0.2832208678472794 ELSE improvement_surcharge#15149 END AS improvement_surcharge#24311, total_amount#15150, (cast((unix_timestamp(tpep_dropoff_datetime#15134, yyyy-MM-dd HH:mm:ss, Some(Etc/UTC), true) - unix_timestamp(tpep_pickup_datetime#15133, yyyy-MM-dd HH:mm:ss, Some(Etc/UTC), true)) as double) / 60.0) AS trip_duration#24312, CASE WHEN NOT (fare_amount#15144 = 0.0) THEN round(((tip_amount#24259 / fare_amount#15144) * 100.0), 2) END AS tip_percentage#24313]
:     +- Project [VendorID#15132, tpep_pickup_datetime#15133, tpep_dropoff_datetime#15134, passenger_count#15135, trip_distance#15136, pickup_longitude#15137, pickup_latitude#15138, RateCodeID#15139, store_and_fwd_flag#15140, dropoff_longitude#15141, dropoff_latitude#15142, payment_type#15143, fare_amount#15144, extra#15145, mta_tax#15146, tolls_amount#15148, improvement_surcharge#15149, total_amount#15150, trip_duration#24312, tip_percentage#24313]

```

```

ag#15140, dropoff_longitude#15141, dropoff_latitude#15142, payment_type#15143, fare_amount#15144, extra#15145, mta_tax#15146, CASE WHEN (isnull(tip_amount#15147) OR (tip_amount#15147 = 0.0)) THEN 1.85112593411255 ELSE tip_amount#15147 END AS tip_amount#24259, tolls_amount#15148, improvement_surcharge#15149, total_amount#15150]
    +- Filter (((((((((isnotnull(passenger_count#15135) AND isnotnull(trip_distance#15136)) AND isnotnull(fare_amount#15144)) AND isnotnull(total_amount#15150)) AND isnotnull(pickup_latitude#15138)) AND isnotnull(pickup_longitude#15137)) AND isnotnull(tpep_pickup_datetime#15133)) AND isnotnull(tpep_dropoff_datetime#15134)) AND (passenger_count#15135 > 0)) AND (trip_distance#15136 > 0.0)) AND (fare_amount#15144 > 0.0)) AND (total_amount#15150 > 0.0)) AND (pickup_latitude#15138 >= 40.5)) AND (pickup_latitude#15138 <= 41.0)) AND (pickup_longitude#15137 >= -74.5)) AND (pickup_longitude#15137 <= -73.5)) AND ((cast((unix_timestamp(tpep_dropoff_datetime#15134, yyyy-MM-dd HH:mm:ss, Some(Etc/UTC), true) - unix_timestamp(tpep_pickup_datetime#15133, yyyy-MM-dd HH:mm:ss, Some(Etc/UTC), true)) as double) / 60.0) > 0.0))
    +- Relation workspace.default.yellow_tripdata_2015_01[VendorID#15132, tpep_pickup_datetime#15133, tpep_dropoff_datetime#15134, passenger_count#15135, trip_distance#15136, pickup_longitude#15137, pickup_latitude#15138, RateCodeID#15139, store_and_fwd_flag#15140, dropoff_longitude#15141, dropoff_latitude#15142, payment_type#15143, fare_amount#15144, extra#15145, mta_tax#15146, tip_amount#15147, tolls_amount#15148, improvement_surcharge#15149, total_amount#15150] parquet

== Physical Plan ==
*(1) ColumnarToRow
+- PhotonResultStage
    +- PhotonUnion Generic
        +- PhotonProject [VendorID#14961, tpep_pickup_datetime#14962, tpep_dropoff_datetime#14963, passenger_count#14964, trip_distance#14965, pickup_longitude#14966, pickup_latitude#14967, RateCodeID#14968, store_and_fwd_flag#14969, dropoff_longitude#14970, dropoff_latitude#14971, payment_type#14972, fare_amount#14973, CASE WHEN (isnull(extra#14974) OR (extra#14974 = 0.0)) THEN 0.30903683757945827 ELSE extra#14974 END AS extra#15158, CASE WHEN (isnull(mta_tax#14975) OR (mta_tax#14975 = 0.0)) THEN 0.4987789140890973 ELSE mta_tax#14975 END AS mta_tax#15160, tip_amount#15152, CASE WHEN (isnull(tolls_amount#14977) OR (tolls_amount#14977 = 0.0)) THEN 0.2416871612181792 ELSE tolls_amount#14977 END AS tolls_amount#15154, CASE WHEN (isnull(improvement_surcharge#14978) OR (improvement_surcharge#14978 = 0.0)) THEN 0.2832208678472794 ELSE improvement_surcharge#14978 END AS improvement_surcharge#15156, total_amount#14979, (cast((unix_timestamp(tpep_dropoff_datetime#14963, yyyy-MM-dd HH:mm:ss, Some(Etc/UTC), true) - unix_timestamp(tpep_pickup_datetime#14962, yyyy-MM-dd HH:mm:ss, Some(Etc/UTC), true)) as double) / 60.0) AS trip_duration#15901, CASE WHEN NOT (fare_amount#14973 = 0.0) THEN round(((tip_amount#15152 / fare_amount#14973) * 100.0), 2) END AS tip_percentage#17354]
        :   +- PhotonProject [VendorID#14961, tpep_pickup_datetime#14962, tpep_dropoff_datetime#14963, passenger_count#14964, trip_distance#14965, pickup_longitude#14966, pickup_latitude#14967, RateCodeID#14968, store_and_fwd_flag#14969, dropoff_longitude#14970, dropoff_latitude#14971, payment_type#14972, fare_amount#14973, extra#14974, mta_tax#14975, CASE

```

```

WHEN (isnull(tip_amount#14976) OR (tip_amount#14976 = 0.0)) THEN 1.8511
25934112553 ELSE tip_amount#14976 END AS tip_amount#15152, tolls_amoun
t#14977, improvement_surcharge#14978, total_amount#14979]
    :  +- PhotonScan parquet workspace.default.yellow_tripdata_201
5_01[VendorID#14961,tpep_pickup_datetime#14962,tpep_dropoff_datetime#14
963,passenger_count#14964,trip_distance#14965,pickup_longitude#14966,pi
ckup_latitude#14967,RateCodeID#14968,store_and_fwd_flag#14969,dropoff_l
ongitude#14970,dropoff_latitude#14971,payment_type#14972,fare_amount#14
973,extra#14974,mta_tax#14975,tip_amount#14976,tolls_amount#14977,impro
vement_surcharge#14978,total_amount#14979] DataFilters: [isnotnull(pass
enger_count#14964), isnotnull(trip_distance#14965), isnotnull(fare_amou
nt#14973), ..., DictionaryFilters: [(passenger_count#14964 > 0), (trip_
distance#14965 > 0.0), (total_amount#14979 > 0.0), ((pickup_l..., Forma
t: parquet, Location: PreparedDeltaFileIndex(1 paths)[s3://dbstorage-pr
od-c51pt/uc/1fcfbb48-e718-4ab4-85bc-b2315e6ce120..., OptionalDataFilter
s: [], PartitionFilters: [], ReadSchema: struct<VendorID:int,tpep_picku
p_datetime:timestamp,tpep_dropoff_datetime:timestamp,passenger_coun..., RequiredDataFilters: [isnotnull(passenger_count#14964), isnotnull(trip_
distance#14965), isnotnull(fare_amount#14973), ...
    :- PhotonProject [VendorID#14980, tpep_pickup_datetime#14981, tpe
p_dropoff_datetime#14982, passenger_count#14983, trip_distance#14984, p
ickup_longitude#14985, pickup_latitude#14986, RateCodeID#14987, store_a
nd_fwd_flag#14988, dropoff_longitude#14989, dropoff_latitude#14990, pay
ment_type#14991, fare_amount#14992, CASE WHEN (isnull(extra#14993) OR (
extra#14993 = 0.0)) THEN 0.30903683757945827 ELSE extra#14993 END AS ex
tra#24260, CASE WHEN (isnull(mta_tax#14994) OR (mta_tax#14994 = 0.0)) T
HEN 0.4987789140890973 ELSE mta_tax#14994 END AS mta_tax#24261, tip_amo
unt#24251, CASE WHEN (isnull(tolls_amount#14996) OR (tolls_amount#14996
= 0.0)) THEN 0.2416871612181792 ELSE tolls_amount#14996 END AS tolls_am
ount#24262, CASE WHEN (isnull(improvement_surcharge#14997) OR (improvem
ent_surcharge#14997 = 0.0)) THEN 0.2832208678472794 ELSE improvement_su
rcharge#14997 END AS improvement_surcharge#24263, total_amount#14998, ((
cast((unix_timestamp(tpep_dropoff_datetime#14982, yyyy-MM-dd HH:mm:ss,
Some(Etc/UTC), true) - unix_timestamp(tpep_pickup_datetime#14981, yyyy-
MM-dd HH:mm:ss, Some(Etc/UTC), true)) as double) / 60.0) AS trip_duri
on#24264, CASE WHEN NOT (fare_amount#14992 = 0.0) THEN round(((tip_amou
nt#24251 / fare_amount#14992) * 100.0), 2) END AS tip_percentage#24265]
    :  +- PhotonProject [VendorID#14980, tpep_pickup_datetime#14981,
tpep_dropoff_datetime#14982, passenger_count#14983, trip_distance#1498
4, pickup_longitude#14985, pickup_latitude#14986, RateCodeID#14987, sto
re_and_fwd_flag#14988, dropoff_longitude#14989, dropoff_latitude#14990,
payment_type#14991, fare_amount#14992, extra#14993, mta_tax#14994, CASE
WHEN (isnull(tip_amount#14995) OR (tip_amount#14995 = 0.0)) THEN 1.8511
25934112553 ELSE tip_amount#14995 END AS tip_amount#24251, tolls_amoun
t#14996, improvement_surcharge#14997, total_amount#14998]
    :  +- PhotonScan parquet workspace.default.yellow_tripdata_201
5_01[VendorID#14980,tpep_pickup_datetime#14981,tpep_dropoff_datetime#14
982,passenger_count#14983,trip_distance#14984,pickup_longitude#14985,pi
ckup_latitude#14986,RateCodeID#14987,store_and_fwd_flag#14988,dropoff_l
ongitude#14989,dropoff_latitude#14990,payment_type#14991,fare_amount#14
992,extra#14993,mta_tax#14994,tip_amount#14995,tolls_amount#14996,impro
vement_surcharge#14997,total_amount#14998] DataFilters: [isnotnull(pass
enger_count#14983), ...

```

```

enger_count#14983), isnotnull(trip_distance#14984), isnotnull(fare_amou
nt#14992), ..., DictionaryFilters: [((pickup_latitude#14986 >= 40.5) AN
D (pickup_latitude#14986 <= 41.0)), ((pickup_longitude#14985 ... , Forma
t: parquet, Location: PreparedDeltaFileIndex(1 paths)[s3://dbstorage-pr
od-c51pt/uc/1fcfbb48-e718-4ab4-85bc-b2315e6ce120... , OptionalDataFilter
s: [], PartitionFilters: [], ReadSchema: struct<VendorID:int,tpep_picku
p_datetime:timestamp,tpep_dropoff_datetime:timestamp,passenger_coun..., Re
quiredDataFilters: [isnotnull(passenger_count#14983), isnotnull(trip_
distance#14984), isnotnull(fare_amount#14992), ...
:- PhotonProject [VendorID#14999, tpep_pickup_datetime#15000, tpe
p_dropoff_datetime#15001, passenger_count#15002, trip_distance#15003, p
ickup_longitude#15004, pickup_latitude#15005, RateCodeID#15006, store_a
nd_fwd_flag#15007, dropoff_longitude#15008, dropoff_latitude#15009, pay
ment_type#15010, fare_amount#15011, CASE WHEN (isnull(extra#15012) OR (
extra#15012 = 0.0)) THEN 0.30903683757945827 ELSE extra#15012 END AS ex
tra#24266, CASE WHEN (isnull(mta_tax#15013) OR (mta_tax#15013 = 0.0)) T
HEN 0.4987789140890973 ELSE mta_tax#15013 END AS mta_tax#24267, tip_amo
unt#24252, CASE WHEN (isnull(tolls_amount#15015) OR (tolls_amount#15015
= 0.0)) THEN 0.2416871612181792 ELSE tolls_amount#15015 END AS tolls_am
ount#24268, CASE WHEN (isnull(improvement_surcharge#15016) OR (improvem
ent_surcharge#15016 = 0.0)) THEN 0.2832208678472794 ELSE improvement_su
rcharge#15016 END AS improvement_surcharge#24269, total_amount#15017, (c
ast((unix_timestamp(tpep_dropoff_datetime#15001, yyyy-MM-dd HH:mm:ss,
Some(Etc/UTC), true) - unix_timestamp(tpep_pickup_datetime#15000, yyyy-
MM-dd HH:mm:ss, Some(Etc/UTC), true)) as double) / 60.0) AS trip_durati
on#24270, CASE WHEN NOT (fare_amount#15011 = 0.0) THEN round(((tip_amou
nt#24252 / fare_amount#15011) * 100.0), 2) END AS tip_percentage#24271]
: +- PhotonProject [VendorID#14999, tpep_pickup_datetime#15000,
tpep_dropoff_datetime#15001, passenger_count#15002, trip_distance#150
3, pickup_longitude#15004, pickup_latitude#15005, RateCodeID#15006, sto
re_and_fwd_flag#15007, dropoff_longitude#15008, dropoff_latitude#15009,
payment_type#15010, fare_amount#15011, extra#15012, mta_tax#15013, CASE
WHEN (isnull(tip_amount#15014) OR (tip_amount#15014 = 0.0)) THEN 1.8511
25934112553 ELSE tip_amount#15014 END AS tip_amount#24252, tolls_amoun
t#15015, improvement_surcharge#15016, total_amount#15017]
:     +- PhotonScan parquet workspace.default.yellow_tripdata_201
5_01[VendorID#14999,tpep_pickup_datetime#15000,tpep_dropoff_datetime#15
001,passenger_count#15002,trip_distance#15003,pickup_longitude#15004,pi
ckup_latitude#15005,RateCodeID#15006,store_and_fwd_flag#15007,dropoff_l
ongitude#15008,dropoff_latitude#15009,payment_type#15010,fare_amount#15
011,extra#15012,mta_tax#15013,tip_amount#15014,tolls_amount#15015,impr
ovement_surcharge#15016,total_amount#15017] DataFilters: [isnotnull(pass
enger_count#15002), isnotnull(trip_distance#15003), isnotnull(fare_amou
nt#15011), ..., DictionaryFilters: [(passenger_count#15002 > 0), ((pick
up_latitude#15005 >= 40.5) AND (pickup_latitude#15005 <= 41.0... , Forma
t: parquet, Location: PreparedDeltaFileIndex(1 paths)[s3://dbstorage-pr
od-c51pt/uc/1fcfbb48-e718-4ab4-85bc-b2315e6ce120... , OptionalDataFilter
s: [], PartitionFilters: [], ReadSchema: struct<VendorID:int,tpep_picku
p_datetime:timestamp,tpep_dropoff_datetime:timestamp,passenger_coun..., Re
quiredDataFilters: [isnotnull(passenger_count#15002), isnotnull(trip_
distance#15003), isnotnull(fare_amount#15011), ...
:- PhotonProject [VendorID#15018, tpep_pickup_datetime#15019, tpe

```

```

p_dropoff_datetime#15020, passenger_count#15021, trip_distance#15022, p
ickup_longitude#15023, pickup_latitude#15024, RateCodeID#15025, store_a
nd_fwd_flag#15026, dropoff_longitude#15027, dropoff_latitude#15028, pay
ment_type#15029, fare_amount#15030, CASE WHEN (isnull(extra#15031) OR (
extra#15031 = 0.0)) THEN 0.30903683757945827 ELSE extra#15031 END AS ex
tra#24272, CASE WHEN (isnull(mta_tax#15032) OR (mta_tax#15032 = 0.0)) T
HEN 0.4987789140890973 ELSE mta_tax#15032 END AS mta_tax#24273, tip_amo
unt#24253, CASE WHEN (isnull(tolls_amount#15034) OR (tolls_amount#15034
= 0.0)) THEN 0.2416871612181792 ELSE tolls_amount#15034 END AS tolls_am
ount#24274, CASE WHEN (isnull(improvement_surcharge#15035) OR (improvem
ent_surcharge#15035 = 0.0)) THEN 0.2832208678472794 ELSE improvement_su
rcharge#15035 END AS improvement_surcharge#24275, total_amount#15036, (
cast((unix_timestamp(tpep_dropoff_datetime#15020, yyyy-MM-dd HH:mm:ss,
Some(Etc/UTC), true) - unix_timestamp(tpep_pickup_datetime#15019, yyyy-
MM-dd HH:mm:ss, Some(Etc/UTC), true)) as double) / 60.0) AS trip_durati
on#24276, CASE WHEN NOT (fare_amount#15030 = 0.0) THEN round(((tip_amou
nt#24253 / fare_amount#15030) * 100.0), 2) END AS tip_percentage#24277]
    : +- PhotonProject [VendorID#15018, tpep_pickup_datetime#15019,
tpep_dropoff_datetime#15020, passenger_count#15021, trip_distance#1502
2, pickup_longitude#15023, pickup_latitude#15024, RateCodeID#15025, sto
re_and_fwd_flag#15026, dropoff_longitude#15027, dropoff_latitude#15028,
payment_type#15029, fare_amount#15030, extra#15031, mta_tax#15032, CASE
WHEN (isnull(tip_amount#15033) OR (tip_amount#15033 = 0.0)) THEN 1.8511
25934112553 ELSE tip_amount#15033 END AS tip_amount#24253, tolls_amoun
t#15034, improvement_surcharge#15035, total_amount#15036]
    :     +- PhotonScan parquet workspace.default.yellow_tripdata_201
5_01[VendorID#15018,tpep_pickup_datetime#15019,tpep_dropoff_datetime#15
020,passenger_count#15021,trip_distance#15022,pickup_longitude#15023,pi
ckup_latitude#15024,RateCodeID#15025,store_and_fwd_flag#15026,dropoff_l
ongitude#15027,dropoff_latitude#15028,payment_type#15029,fare_amount#15
030,extra#15031,mta_tax#15032,tip_amount#15033,tolls_amount#15034,impro
vement_surcharge#15035,total_amount#15036] DataFilters: [isnotnull(pass
enger_count#15021), isnotnull(trip_distance#15022), isnotnull(fare_amou
nt#15030), ...], DictionaryFilters: [((pickup_longitude#15023 >= -74.5)
AND (pickup_longitude#15023 <= -73.5)), (trip_distance#15022 ...], Forma
t: parquet, Location: PreparedDeltaFileIndex(1 paths)[s3://dbstorage-pr
od-c51pt/uc/1fcfbb48-e718-4ab4-85bc-b2315e6ce120...], OptionalDataFilter
s: [], PartitionFilters: [], ReadSchema: struct<VendorID:int,tpep_picku
p_datetime:timestamp,tpep_dropoff_datetime:timestamp,passenger_coun..., Re
quiredDataFilters: [isnotnull(passenger_count#15021), isnotnull(trip_
distance#15022), isnotnull(fare_amount#15030), ...]
    :- PhotonProject [VendorID#15037, tpep_pickup_datetime#15038, tpe
p_dropoff_datetime#15039, passenger_count#15040, trip_distance#15041, p
ickup_longitude#15042, pickup_latitude#15043, RateCodeID#15044, store_a
nd_fwd_flag#15045, dropoff_longitude#15046, dropoff_latitude#15047, pay
ment_type#15048, fare_amount#15049, CASE WHEN (isnull(extra#15050) OR (
extra#15050 = 0.0)) THEN 0.30903683757945827 ELSE extra#15050 END AS ex
tra#24278, CASE WHEN (isnull(mta_tax#15051) OR (mta_tax#15051 = 0.0)) T
HEN 0.4987789140890973 ELSE mta_tax#15051 END AS mta_tax#24279, tip_amo
unt#24254, CASE WHEN (isnull(tolls_amount#15053) OR (tolls_amount#15053
= 0.0)) THEN 0.2416871612181792 ELSE tolls_amount#15053 END AS tolls_am
ount#24280, CASE WHEN (isnull(improvement_surcharge#15054) OR (improvem

```

```
ent_surcharge#15054 = 0.0)) THEN 0.2832208678472794 ELSE improvement_su
rcharge#15054 END AS improvement_surcharge#24281, total_amount#15055, (
cast((unix_timestamp(tpep_dropoff_datetime#15039, yyyy-MM-dd HH:mm:ss,
Some(Etc/UTC), true) - unix_timestamp(tpep_pickup_datetime#15038, yyyy-
MM-dd HH:mm:ss, Some(Etc/UTC), true)) as double) / 60.0) AS trip_durati
on#24282, CASE WHEN NOT (fare_amount#15049 = 0.0) THEN round(((tip_amou
nt#24254 / fare_amount#15049) * 100.0), 2) END AS tip_percentage#24283]
    : +- PhotonProject [VendorID#15037, tpep_pickup_datetime#15038,
tpep_dropoff_datetime#15039, passenger_count#15040, trip_distance#1504
1, pickup_longitude#15042, pickup_latitude#15043, RateCodeID#15044, sto
re_and_fwd_flag#15045, dropoff_longitude#15046, dropoff_latitude#15047,
payment_type#15048, fare_amount#15049, extra#15050, mta_tax#15051, CASE
WHEN (isnull(tip_amount#15052) OR (tip_amount#15052 = 0.0)) THEN 1.8511
25934112553 ELSE tip_amount#15052 END AS tip_amount#24254, tolls_amoun
t#15053, improvement_surcharge#15054, total_amount#15055]
    :     +- PhotonScan parquet workspace.default.yellow_tripdata_201
5_01[VendorID#15037,tpep_pickup_datetime#15038,tpep_dropoff_datetime#15
039,passenger_count#15040,trip_distance#15041,pickup_longitude#15042,pi
ckup_latitude#15043,RateCodeID#15044,store_and_fwd_flag#15045,dropoff_l
ongitude#15046,dropoff_latitude#15047,payment_type#15048,fare_amount#15
049,extra#15050,mta_tax#15051,tip_amount#15052,tolls_amount#15053,impro
vement_surcharge#15054,total_amount#15055] DataFilters: [isnotnull(pass
enger_count#15040), isnotnull(trip_distance#15041), isnotnull(fare_amou
nt#15049), ...], DictionaryFilters: [(total_amount#15055 > 0.0), (passen
ger_count#15040 > 0), (trip_distance#15041 > 0.0), (fare_amou..., Forma
t: parquet, Location: PreparedDeltaFileIndex(1 paths)[s3://dbstorage-pr
od-c51pt/uc/1fcfbb48-e718-4ab4-85bc-b2315e6ce120..., OptionalDataFilter
s: [], PartitionFilters: [], ReadSchema: struct<VendorID:int,tpep_picku
p_datetime:timestamp,tpep_dropoff_datetime:timestamp,passenger_coun..., Re
quiredDataFilters: [isnotnull(passenger_count#15040), isnotnull(trip_
distance#15041), isnotnull(fare_amount#15049), ...]
    :- PhotonProject [VendorID#15056, tpep_pickup_datetime#15057, tpe
p_dropoff_datetime#15058, passenger_count#15059, trip_distance#15060, p
ickup_longitude#15061, pickup_latitude#15062, RateCodeID#15063, store_a
nd_fwd_flag#15064, dropoff_longitude#15065, dropoff_latitude#15066, pay
ment_type#15067, fare_amount#15068, CASE WHEN (isnull(extra#15069) OR (
extra#15069 = 0.0)) THEN 0.30903683757945827 ELSE extra#15069 END AS ex
tra#24284, CASE WHEN (isnull(mta_tax#15070) OR (mta_tax#15070 = 0.0)) T
HEN 0.4987789140890973 ELSE mta_tax#15070 END AS mta_tax#24285, tip_amo
unt#24255, CASE WHEN (isnull(tolls_amount#15072) OR (tolls_amount#15072
= 0.0)) THEN 0.2416871612181792 ELSE tolls_amount#15072 END AS tolls_am
ount#24286, CASE WHEN (isnull(improvement_surcharge#15073) OR (improvem
ent_surcharge#15073 = 0.0)) THEN 0.2832208678472794 ELSE improvement_su
rcharge#15073 END AS improvement_surcharge#24287, total_amount#15074, (
cast((unix_timestamp(tpep_dropoff_datetime#15058, yyyy-MM-dd HH:mm:ss,
Some(Etc/UTC), true) - unix_timestamp(tpep_pickup_datetime#15057, yyyy-
MM-dd HH:mm:ss, Some(Etc/UTC), true)) as double) / 60.0) AS trip_durati
on#24288, CASE WHEN NOT (fare_amount#15068 = 0.0) THEN round(((tip_amou
nt#24255 / fare_amount#15068) * 100.0), 2) END AS tip_percentage#24289]
    : +- PhotonProject [VendorID#15056, tpep_pickup_datetime#15057,
tpep_dropoff_datetime#15058, passenger_count#15059, trip_distance#1506
0, pickup_longitude#15061, pickup_latitude#15062, RateCodeID#15063, sto
```

```

re_and_fwd_flag#15064, dropoff_longitude#15065, dropoff_latitude#15066,
payment_type#15067, fare_amount#15068, extra#15069, mta_tax#15070, CASE
WHEN (isnull(tip_amount#15071) OR (tip_amount#15071 = 0.0)) THEN 1.8511
25934112553 ELSE tip_amount#15071 END AS tip_amount#24255, tolls_amoun
t#15072, improvement_surcharge#15073, total_amount#15074]
      :      +- PhotonScan parquet workspace.default.yellow_tripdata_201
5_01[VendorID#15056,tpep_pickup_datetime#15057,tpep_dropoff_datetime#15
058,passenger_count#15059,trip_distance#15060,pickup_longitude#15061,pi
ckup_latitude#15062,RateCodeID#15063,store_and_fwd_flag#15064,dropoff_l
ongitude#15065,dropoff_latitude#15066,payment_type#15067,fare_amount#15
068,extra#15069,mta_tax#15070,tip_amount#15071,tolls_amount#15072,impro
vement_surcharge#15073,total_amount#15074] DataFilters: [isnotnull(pass
enger_count#15059), isnotnull(trip_distance#15060), isnotnull(fare_amou
nt#15068), ...], DictionaryFilters: [(trip_distance#15060 > 0.0), ((pick
up_longitude#15061 >= -74.5) AND (pickup_longitude#15061 <= -..., Forma
t: parquet, Location: PreparedDeltaFileIndex(1 paths)[s3://dbstorage-pr
od-c51pt/uc/1fcfb848-e718-4ab4-85bc-b2315e6ce120..., OptionalDataFilter
s: [], PartitionFilters: [], ReadSchema: struct<VendorID:int,tpep_picku
p_datetime:timestamp,tpep_dropoff_datetime:timestamp,passenger_coun..., Re
quiredDataFilters: [isnotnull(passenger_count#15059), isnotnull(trip_
distance#15060), isnotnull(fare_amount#15068), ...]
      :- PhotonProject [VendorID#15075, tpep_pickup_datetime#15076, tpe
p_dropoff_datetime#15077, passenger_count#15078, trip_distance#15079, p
ickup_longitude#15080, pickup_latitude#15081, RateCodeID#15082, store_a
nd_fwd_flag#15083, dropoff_longitude#15084, dropoff_latitude#15085, pay
ment_type#15086, fare_amount#15087, CASE WHEN (isnull(extra#15088) OR
(extra#15088 = 0.0)) THEN 0.30903683757945827 ELSE extra#15088 END AS ex
tra#24290, CASE WHEN (isnull(mta_tax#15089) OR (mta_tax#15089 = 0.0)) T
HEN 0.4987789140890973 ELSE mta_tax#15089 END AS mta_tax#24291, tip_amo
unt#24256, CASE WHEN (isnull(tolls_amount#15091) OR (tolls_amount#15091
= 0.0)) THEN 0.2416871612181792 ELSE tolls_amount#15091 END AS tolls_am
ount#24292, CASE WHEN (isnull(improvement_surcharge#15092) OR (improvem
ent_surcharge#15092 = 0.0)) THEN 0.2832208678472794 ELSE improvement_su
rcharge#15092 END AS improvement_surcharge#24293, total_amount#15093, (c
ast((unix_timestamp(tpep_dropoff_datetime#15077, yyyy-MM-dd HH:mm:ss,
Some(Etc/UTC), true) - unix_timestamp(tpep_pickup_datetime#15076, yyyy-
MM-dd HH:mm:ss, Some(Etc/UTC), true)) as double) / 60.0) AS trip_duri
ation#24294, CASE WHEN NOT (fare_amount#15087 = 0.0) THEN round(((tip_amou
nt#24256 / fare_amount#15087) * 100.0), 2) END AS tip_percentage#24295]
      :      +- PhotonProject [VendorID#15075, tpep_pickup_datetime#15076,
tpep_dropoff_datetime#15077, passenger_count#15078, trip_distance#1507
9, pickup_longitude#15080, pickup_latitude#15081, RateCodeID#15082, sto
re_and_fwd_flag#15083, dropoff_longitude#15084, dropoff_latitude#15085,
payment_type#15086, fare_amount#15087, extra#15088, mta_tax#15089, CASE
WHEN (isnull(tip_amount#15090) OR (tip_amount#15090 = 0.0)) THEN 1.8511
25934112553 ELSE tip_amount#15090 END AS tip_amount#24256, tolls_amoun
t#15091, improvement_surcharge#15092, total_amount#15093]
      :      +- PhotonScan parquet workspace.default.yellow_tripdata_201
5_01[VendorID#15075,tpep_pickup_datetime#15076,tpep_dropoff_datetime#15
077,passenger_count#15078,trip_distance#15079,pickup_longitude#15080,pi
ckup_latitude#15081,RateCodeID#15082,store_and_fwd_flag#15083,dropoff_l
ongitude#15084,dropoff_latitude#15085,payment_type#15086,fare_amount#15
087]
  
```

```
087,extra#15088,mta_tax#15089,tip_amount#15090,tolls_amount#15091,improvement_surcharge#15092,total_amount#15093] DataFilters: [isnotnull(passenger_count#15078), isnotnull(trip_distance#15079), isnotnull(fare_amount#15087), ..., DictionaryFilters: [(total_amount#15093 > 0.0), (passenger_count#15078 > 0), (trip_distance#15079 > 0.0), ((pickup_longitude#15099 > 0) AND (pickup_latitude#15100 > 0))], Format: parquet, Location: PreparedDeltaFileIndex(1 paths)[s3://dbstorage-pr0d-c51pt/uc/1fcfbb48-e718-4ab4-85bc-b2315e6ce120...], OptionalDataFilters: [], PartitionFilters: [], ReadSchema: struct<VendorID:int,tpep_pickup_datetime:timestamp,tpep_dropoff_datetime:timestamp,passenger_count:int,toll_amount:double,tip_amount:double,trip_distance:double,trip_duration:double,trip_fare:double,trip_id:double,trip_mta_tax:double,trip_payment_type:double,trip_tip:double,trip_tolls:double,trip_total:double>, RequiredDataFilters: [isnotnull(passenger_count#15078), isnotnull(trip_distance#15079), isnotnull(fare_amount#15087), ...]

:- PhotonProject [VendorID#15094, tpep_pickup_datetime#15095, tpep_dropoff_datetime#15096, passenger_count#15097, trip_distance#15098, pickup_longitude#15099, pickup_latitude#15100, RateCodeID#15101, store_and_fwd_flag#15102, dropoff_longitude#15103, dropoff_latitude#15104, payment_type#15105, fare_amount#15106, CASE WHEN (isnull(extra#15107) OR (extra#15107 = 0.0)) THEN 0.30903683757945827 ELSE extra#15107 END AS extra#24296, CASE WHEN (isnull(mta_tax#15108) OR (mta_tax#15108 = 0.0)) THEN 0.4987789140890973 ELSE mta_tax#15108 END AS mta_tax#24297, tip_amount#24257, CASE WHEN (isnull(tolls_amount#15110) OR (tolls_amount#15110 = 0.0)) THEN 0.2416871612181792 ELSE tolls_amount#15110 END AS tolls_amount#24298, CASE WHEN (isnull(improvement_surcharge#15111) OR (improvement_surcharge#15111 = 0.0)) THEN 0.2832208678472794 ELSE improvement_surcharge#15111 END AS improvement_surcharge#24299, total_amount#15112, (cast((unix_timestamp(tpep_dropoff_datetime#15096, yyyy-MM-dd HH:mm:ss, Some(Etc/UTC), true) - unix_timestamp(tpep_pickup_datetime#15095, yyyy-MM-dd HH:mm:ss, Some(Etc/UTC), true)) as double) / 60.0) AS trip_duration#24300, CASE WHEN NOT (fare_amount#15106 = 0.0) THEN round(((tip_amount#24257 / fare_amount#15106) * 100.0), 2) END AS tip_percentage#24301]

: +- PhotonProject [VendorID#15094, tpep_pickup_datetime#15095, tpep_dropoff_datetime#15096, passenger_count#15097, trip_distance#15098, pickup_longitude#15099, pickup_latitude#15100, RateCodeID#15101, store_and_fwd_flag#15102, dropoff_longitude#15103, dropoff_latitude#15104, payment_type#15105, fare_amount#15106, extra#15107, mta_tax#15108, CASE WHEN (isnull(tip_amount#15109) OR (tip_amount#15109 = 0.0)) THEN 1.851125934112553 ELSE tip_amount#15109 END AS tip_amount#24257, tolls_amount#15110, improvement_surcharge#15111, total_amount#15112]

:     +- PhotonScan parquet workspace.default.yellow_tripdata_2015_01[VendorID#15094,tpep_pickup_datetime#15095,tpep_dropoff_datetime#15096,passenger_count#15097,trip_distance#15098,pickup_longitude#15099,pickup_latitude#15100,RateCodeID#15101,store_and_fwd_flag#15102,dropoff_longitude#15103,dropoff_latitude#15104,payment_type#15105,fare_amount#15106,extra#15107,mta_tax#15108,tip_amount#15109,tolls_amount#15110,improvement_surcharge#15111,total_amount#15112] DataFilters: [isnotnull(passenger_count#15097), isnotnull(trip_distance#15098), isnotnull(fare_amount#15106), ..., DictionaryFilters: [(trip_distance#15098 > 0.0), (passenger_count#15097 > 0), ((pickup_latitude#15100 >= 40.5) AND (...)), Format: parquet, Location: PreparedDeltaFileIndex(1 paths)[s3://dbstorage-pr0d-c51pt/uc/1fcfbb48-e718-4ab4-85bc-b2315e6ce120...], OptionalDataFilters: [], PartitionFilters: [], ReadSchema: struct<VendorID:int,tpep_pickup_datetime:timestamp,tpep_dropoff_datetime:timestamp,passenger_count:int,toll_amount:double,tip_amount:double,trip_distance:double,trip_fare:double,trip_id:double,trip_mta_tax:double,trip_payment_type:double,trip_tip:double,trip_tolls:double,trip_total:double>, RequiredDataFilters: [isnotnull(passenger_count#15097), isnotnull(trip_distance#15098), isnotnull(fare_amount#15106), ...]
```



```

= 0.0)) THEN 0.2416871612181792 ELSE tolls_amount#15148 END AS tolls_amount#24310, CASE WHEN (isnull(improvement_surcharge#15149) OR (improvement_surcharge#15149 = 0.0)) THEN 0.2832208678472794 ELSE improvement_surcharge#15149 END AS improvement_surcharge#24311, total_amount#15150, (cast((unix_timestamp(tpep_dropoff_datetime#15134, yyyy-MM-dd HH:mm:ss, Some(Etc/UTC), true) - unix_timestamp(tpep_pickup_datetime#15133, yyyy-MM-dd HH:mm:ss, Some(Etc/UTC), true)) as double) / 60.0) AS trip_duration#24312, CASE WHEN NOT (fare_amount#15144 = 0.0) THEN round(((tip_amount#24259 / fare_amount#15144) * 100.0), 2) END AS tip_percentage#24313]
    +- PhotonProject [VendorID#15132, tpep_pickup_datetime#15133, tpep_dropoff_datetime#15134, passenger_count#15135, trip_distance#15136, pickup_longitude#15137, pickup_latitude#15138, RateCodeID#15139, store_and_fwd_flag#15140, dropoff_longitude#15141, dropoff_latitude#15142, payment_type#15143, fare_amount#15144, extra#15145, mta_tax#15146, CASE WHEN (isnull(tip_amount#15147) OR (tip_amount#15147 = 0.0)) THEN 1.851125934112553 ELSE tip_amount#15147 END AS tip_amount#24259, tolls_amount#15148, improvement_surcharge#15149, total_amount#15150]
    +- PhotonScan parquet workspace.default.yellow_tripdata_2015_01[VendorID#15132,tpep_pickup_datetime#15133,tpep_dropoff_datetime#15134,passenger_count#15135,trip_distance#15136,pickup_longitude#15137,pickup_latitude#15138,RateCodeID#15139,store_and_fwd_flag#15140,dropoff_longitude#15141,dropoff_latitude#15142,payment_type#15143,fare_amount#15144,extra#15145,mta_tax#15146,tip_amount#15147,tolls_amount#15148,improvement_surcharge#15149,total_amount#15150] DataFilters: [isnotnull(passenger_count#15135), isnotnull(trip_distance#15136), isnotnull(fare_amount#15144), ...], DictionaryFilters: [((pickup_latitude#15138 >= 40.5) AND (pickup_latitude#15138 <= 41.0)), (total_amount#15150 > 0.0...], Format: parquet, Location: PreparedDeltaFileIndex(1 paths)[s3://dbstorage-prd-c51pt/uc/1fcfbb48-e718-4ab4-85bc-b2315e6ce120...], OptionalDataFilters: [], PartitionFilters: [], ReadSchema: struct<VendorID:int,tpep_pickup_datetime:timestamp,tpep_dropoff_datetime:timestamp,passenger_count:int,trip_distance:double,pickup_longitude:double,pickup_latitude:double,RateCodeID:int,store_and_fwd_flag:int,dropoff_longitude:double,dropoff_latitude:double,payment_type:int,fare_amount:double,extra:double,mta_tax:double,tip_amount:double,tolls_amount:double,improvement_surcharge:double,total_amount:double>, RequiredDataFilters: [isnotnull(passenger_count#15135), isnotnull(trip_distance#15136), isnotnull(fare_amount#15144), ...]

== Photon Explanation ==
The query is fully supported by Photon.
== Optimizer Statistics (table names per statistics state) ==
    missing =
    partial =
    full    = yellow_tripdata_2015_01, yellow_tripdata_2015_01

```

In [0]: `from pyspark.sql.functions import col`

```

# Transformation (lazy)
df_transformed = df_filtered.withColumn("fare_per_mile", col("fare_amo..."))

# No computation happens yet
print("Transformation defined but not executed.")

```

```
# Action (eager)
row_count = df_transformed.count() # triggers execution
print(f"Action executed. Total rows: {row_count}")
```

Transformation defined but not executed.

Action executed. Total rows: 124302230

```
In [0]: # Write as Delta table (default format)
df_filtered.write \
    .mode("overwrite") \
    .partitionBy("VendorID") \
    .saveAsTable("nyc_taxi_processed")
```

```
In [0]: # Read it back
df = spark.table("nyc_taxi_processed")
df.show()
```

VendorID	tpep_pickup_datetime	tpep_dropoff_datetime	passenger_count	trip_distance	pickup_longitude	pickup_latitude	RateCodeID	store_and_fwd_flag	dropoff_longitude	dropoff_latitude	payment_type	fare_amount	extra	mta_tax	tip_amount	tolls_amount	improvement_surcharge	total_amount	trip_duration	tip_percentage	
2	2015-01-04 19:08:08	2015-01-04 19:09:32	1	0.44	-73.95577239990234	40.77932357788086	1	N	-73.96114349365234	40.77193069458008	2	3.5	0.3090	3683757945827	0.5	1.851125934112553	0.2416871612181792	0.3	4.3	1.4	52.89
2	2015-01-04 19:08:09	2015-01-04 19:18:27	1	1.6	-73.98239135742188	40.757545471191406	1	N	-73.98091125488281	40.772483825683594	1	9.0	0.30903	683757945827	0.5	1.8	0.2416871612181792	0.3	11.6	10.3	20.0
2	2015-01-04 19:08:09	2015-01-04 19:14:19	1	1.69	-73.9889144897461	40.7738037109375	1	N	-73.97200012207031	40.75933074951172	1	7.0	0.3090	3683757945827	0.5	1.4	0.2416871612181792	0.3	9.2	6.166666666666667	20.0
2	2015-01-04 19:08:10	2015-01-04 19:24:59	4	9.87	-73.99996948242188	40.71793746948242	1	N	-73.9144058227539	40.75789260864258	1	27.5	0.3090	3683757945827	0.5	2.5	0.2416871612181792	0.3	30.8	16.81666666666666	9.09
2	2015-01-04 19:08:11	2015-01-04 19:13:21	1																		

0.57 | -73.9782485961914 | 40.745670318603516 | 1 |  
 N | -73.98542785644531 | 40.74728012084961 | 1 | 5.0 | 0.3090  
 3683757945827 | 0.5 | 1.0 | 0.2416871612181792 |  
 0.3 | 6.8 | 5.166666666666667 | 20.0 |  
 | 2 | 2015-01-04 19:08:11 | 2015-01-04 19:17:29 | 3 |  
 1.82 | -73.95823669433594 | 40.77878189086914 | 1 |  
 N | -73.98169708251953 | 40.773616790771484 | 1 | 9.0 | 0.3090  
 3683757945827 | 0.5 | 1.2 | 0.2416871612181792 |  
 0.3 | 11.0 | 9.3 | 13.33 |  
 | 2 | 2015-01-04 19:08:11 | 2015-01-04 19:36:27 | 2 |  
 16.3 | -73.78997802734375 | 40.647098541259766 | 1 |  
 N | -73.84770202636719 | 40.83883285522461 | 1 | 44.5 | 0.3090  
 3683757945827 | 0.5 | 1.851125934112553 | 5.33 |  
 0.3 | 50.63 | 28.266666666666666 | 4.16 |  
 | 2 | 2015-01-04 19:08:12 | 2015-01-04 19:16:48 | 1 |  
 1.78 | -73.98756408691406 | 40.77362060546875 | 1 |  
 N | -73.97200775146484 | 40.7922477722168 | 2 | 8.5 | 0.3090  
 3683757945827 | 0.5 | 1.851125934112553 | 0.2416871612181792 |  
 0.3 | 9.3 | 8.6 | 21.78 |  
 | 2 | 2015-01-04 19:08:12 | 2015-01-04 19:14:20 | 1 |  
 1.63 | -73.9703598022461 | 40.752044677734375 | 1 |  
 N | -73.98538970947266 | 40.731693267822266 | 1 | 7.0 | 0.3090  
 3683757945827 | 0.5 | 1.0 | 0.2416871612181792 |  
 0.3 | 8.8 | 6.13333333333334 | 14.29 |  
 | 2 | 2015-01-04 19:08:12 | 2015-01-04 19:26:40 | 1 |  
 5.8 | -73.95423126220703 | 40.774593353271484 | 1 | N  
 | -73.99592590332031 | 40.72350311279297 | 1 | 20.0 | 0.30903  
 683757945827 | 0.5 | 4.0 | 0.2416871612181792 |  
 0.3 | 24.8 | 18.466666666666665 | 20.0 |  
 | 2 | 2015-01-04 19:08:12 | 2015-01-04 19:11:58 | 1 |  
 0.95 | -73.96778106689453 | 40.76289749145508 | 1 |  
 N | -73.96028900146484 | 40.7713508605957 | 1 | 5.0 | 0.3090  
 3683757945827 | 0.5 | 1.0 | 0.2416871612181792 |  
 0.3 | 6.8 | 3.766666666666666 | 20.0 |  
 | 2 | 2015-01-11 19:46:21 | 2015-01-11 19:48:41 | 1 |  
 0.63 | -73.97464752197266 | 40.752418518066406 | 1 |  
 N | -73.97885131835938 | 40.746192932128906 | 2 | 4.0 | 0.3090  
 3683757945827 | 0.5 | 1.851125934112553 | 0.2416871612181792 |  
 0.3 | 4.8 | 2.333333333333335 | 46.28 |  
 | 2 | 2015-01-11 19:46:22 | 2015-01-11 19:51:52 | 6 |  
 0.86 | -73.98411560058594 | 40.76070022583008 | 1 |  
 N | -73.98731994628906 | 40.751163482666016 | 2 | 5.5 | 0.3090  
 3683757945827 | 0.5 | 1.851125934112553 | 0.2416871612181792 |  
 0.3 | 6.3 | 5.5 | 33.66 |  
 | 2 | 2015-01-11 19:46:23 | 2015-01-11 19:52:21 | 1 |  
 1.06 | -74.00928497314453 | 40.71708297729492 | 1 |  
 N | -74.00005340576172 | 40.72431182861328 | 1 | 6.0 | 0.3090  
 3683757945827 | 0.5 | 1.2 | 0.2416871612181792 |  
 0.3 | 8.0 | 5.966666666666667 | 20.0 |  
 | 2 | 2015-01-11 19:46:24 | 2015-01-11 19:50:03 | 5 |  
 0.67 | -74.00414276123047 | 40.707210540771484 | 1 |  
 N | -73.9966812133789 | 40.714908599853516 | 1 | 4.5 | 0.3090

3683757945827	0.5	1.0	0.2416871612181792
0.3	6.3	3.65	22.22
2	2015-01-11 19:46:24	2015-01-11 19:51:23	5
1.37	-74.0005874633789	40.73724365234375	1
N -74.00826263427734	40.71992111206055	1	6.0 0.3090
3683757945827	0.5	2.0	0.2416871612181792
0.3	8.8	4.983333333333333	33.33
2	2015-01-11 19:46:24	2015-01-11 19:52:23	1
0.93	-74.00054931640625	40.729888916015625	1
N -73.9961929321289	40.73809814453125	2	5.5 0.3090
3683757945827	0.5	1.851125934112553	0.2416871612181792
0.3	6.3	5.983333333333333	33.66
2	2015-01-11 19:46:24	2015-01-11 20:01:00	1
1.98	-73.99056243896484	40.686641693115234	1
N -73.95591735839844	40.67766571044922	1	11.0 0.3090
3683757945827	0.5	2.75	0.2416871612181792
0.3	14.55	14.6	25.0
2	2015-01-11 19:46:25	2015-01-11 19:51:03	2
1.36	-73.99972534179688	40.73329162597656	1
N -73.98614501953125	40.7436637878418	1	6.0 0.3090
3683757945827	0.5	1.2	0.2416871612181792
0.3	8.0	4.633333333333334	20.0
2	2015-01-11 19:46:26	2015-01-11 19:56:30	3
2.17	-73.99606323242188	40.74431610107422	1
N -73.99635314941406	40.721290588378906	1	9.0 0.3090
3683757945827	0.5	1.0	0.2416871612181792
0.3	10.8	10.066666666666666	11.11

only showing top 20 rows