

Pickup drop off Script

```
T = Load 'hpc_join_12' using PigStorage('\t') as
(medallion:chararray,hack_license:chararray,vendor_id:chararray,rate_code:int,store_and_fwd_flag:chararray,pickup_datetime:chararray,dropoff_datetime:chararray,passenger_count:chararray,trip_time_in_secs:chararray,trip_distance:float,pickup_longitude:chararray,pickup_latitude:chararray,dropoff_longitude:chararray,dropoff_latitude:chararray,medallion_2:chararray,hack_licence_2:chararray,vendor_id_2:chararray,time:chararray,payment_type:chararray,fare_amount:float,surcharge:float,mta_tax:float,tip_amount:float,tolls_amount:chararray,
total_amount:float,pickup_longitude_2:chararray,pickup_latitude_2:chararray,pickup_region:chararray,dropoff_longitude_2:chararray,dropoff_latitude_2:chararray,dropoff_region:chararray);

O = FOREACH T GENERATE SUBSTRING(pickup_datetime,11,13) as
hour_p,medallion,hack_license,vendor_id,rate_code,store_and_fwd_flag,pickup_datetime,dropoff_datetime,passenger_count,trip_time_in_secs,trip_distance,pickup_longitude,pickup_latitude,dropoff_longitude,dropoff_latitude,medallion_2,hack_licence_2,vendor_id_2,time,payment_type,fare_amount,surcharge,mta_tax,tip_amount,tolls_amount,
total_amount,pickup_longitude_2,pickup_latitude_2,pickup_region,dropoff_longitude_2,dropoff_latitude_2,dropoff_region;

LOGS_GROUP= GROUP O by (pickup_region,dropoff_region);

C = FOREACH LOGS_GROUP GENERATE FLATTEN(group) as (pickup_region,dropoff_region),
SUM(O.total_amount) as Total_amount;

countof_amount1 = order C by Total_amount ASC;

Store countof_amount1 into 'Pickup_dropoff_12';
```

Best Driver Script

```
T = Load 'join_1' using PigStorage('\t') as
(medallion:chararray,hack_license:chararray,vendor_id:chararray,rate_code:int,store_and_fwd_flag:chararray,pickup_datetime:chararray,dropoff_datetime:chararray,passenger_count:chararray,trip_time_in_secs:chararray,trip_distance:float,pickup_longitude:chararray,pickup_latitude:chararray,dropoff_longitude:chararray,dropoff_latitude:chararray,medallion_2:chararray,hack_licence_2:chararray,vendor_id_2:chararray,time:chararray,payment_type:chararray,fare_amount:float,surcharge:float,mta_tax:float,tip_amount:float,tolls_amount:chararray,
total_amount:float,pickup_longitude_2:chararray,pickup_latitude_2:chararray,pickup_region:chararray,dropoff_longitude_2:chararray,dropoff_latitude_2:chararray,dropoff_region:chararray);
```

```
O = FOREACH T GENERATE SUBSTRING(pickup_datetime,8,10) as
hour_p,medallion,hack_license,vendor_id,rate_code,store_and_fwd_flag,pickup_datetime,dropoff_datetime,passenger_count,trip_time_in_secs,trip_distance,pickup_longitude,pickup_latitude,dropoff_longitude,dropoff_latitude,medallion_2,hack_licence_2,vendor_id_2,time,payment_type,fare_amount,surcharge,mta_tax,tip_amount,tolls_amount,
total_amount,pickup_longitude_2,pickup_latitude_2,pickup_region,dropoff_longitude_2,dropoff_latitude_2,dropoff_region;
```

```
grp = group O by hack_license;
```

```
mean = foreach grp { sum = SUM (O.fare_amount); sum2 = SUM (O.tip_amount);count1= COUNT(O);
GENERATE group as hack_license, (sum2/sum)*100 as tip_by_fare, sum as fare_total,sum2 as
tip_total,count1 as count;};
```

```
X = ORDER mean BY tip_total DESC;
```

```
Store X into 'Diver1';
```

Part II to find the region to get maximum tip

```
grunt> T = load 'join_8' using PigStorage('\t') as
(medallion:chararray,hack_license:chararray,vendor_id:chararray,rate_code:int,store_and_fwd_flag:chararray,pickup_datetime:chararray,dropoff_datetime:chararray,passenger_count:chararray,trip_time_in_secs:chararray,trip_distance:float,pickup_longitude:chararray,pickup_latitude:chararray,dropoff_longitude:chararray,dropoff_latitude:chararray,medallion_2:chararray,hack_licence_2:chararray,vendor_id_2:chararray,time:chararray,payment_type:chararray,fare_amount:float,surcharge:float,mta_tax:float,tip_amount:float,tolls_amount:chararray,
total_amount:float,pickup_longitude_2:chararray,pickup_latitude_2:chararray,pickup_region:chararray,dropoff_longitude_2:chararray,dropoff_latitude_2:chararray,dropoff_region:chararray);
```

```
grunt> W = Load 'driver_analysis' using PigStorage(',') as
(hack_license:chararray,percent_tips:chararray,total_fare:chararray,tip_total:chararray,count_rides:float);
grunt> H = Join W by hack_license,T by hack_license;
grunt> R = Group H by T::pickup_region;
```

```
grunt> mean = foreach R {count1= COUNT(H); GENERATE group as pickup_region,count1 as count;};
grunt> dump mean;
```

Passenger count Script

```
T = Load 'join_12' using PigStorage('\t') as
(medallion:chararray,hack_license:chararray,vendor_id:chararray,rate_code:int,store_and_fwd_flag:chararray,pickup_datetime:chararray,dropoff_datetime:chararray,passenger_count:chararray,trip_time_in_secs:chararray,trip_distance:float,pickup_longitude:chararray,pickup_latitude:chararray,dropoff_longitude:chararray,dropoff_latitude:chararray,medallion_2:chararray,hack_licence_2:chararray,vendor_id_2:chararray,time:chararray,payment_type:chararray,fare_amount:float,surcharge:float,mta_tax:float,tip_amount:float,tolls_amount:chararray,
total_amount:float,pickup_longitude_2:chararray,pickup_latitude_2:chararray,pickup_region:chararray,dropoff_longitude_2:chararray,dropoff_latitude_2:chararray,dropoff_region:chararray);

O = FOREACH T GENERATE SUBSTRING(pickup_datetime,11,13) as
hour_p,medallion,hack_license,vendor_id,rate_code,store_and_fwd_flag,pickup_datetime,dropoff_datetime,passenger_count,trip_time_in_secs,trip_distance,pickup_longitude,pickup_latitude,dropoff_longitude,dropoff_latitude,medallion_2,hack_licence_2,vendor_id_2,time,payment_type,fare_amount,surcharge,mta_tax,tip_amount,tolls_amount,
total_amount,pickup_longitude_2,pickup_latitude_2,pickup_region,dropoff_longitude_2,dropoff_latitude_2,dropoff_region;

grp = group O by passenger_count;

mean = foreach grp { sum = SUM (O.fare_amount); sum2 = SUM (O.tip_amount);count1= COUNT(O);
GENERATE group as passenger_count, (sum2/sum)*100 as tip_by_fare, sum as fare_total,sum2 as tip_total,count1 as count;};
```

Less than 1 mile trips script.

```
O = FOREACH T GENERATE SUBSTRING(pickup_datetime,8,10) as
hour_p,medallion,hack_license,vendor_id,rate_code,store_and_fwd_flag,pickup_datetime,dropoff_datetime,passenger_count,trip_time_in_secs,trip_distance,pickup_longitude,pickup_latitude,dropoff_longitude,dropoff_latitude,medallion_2,hack_licence_2,vendor_id_2,time,payment_type,fare_amount,surcharge,mta_tax,tip_amount,tolls_amount,
total_amount,pickup_longitude_2,pickup_latitude_2,pickup_region,dropoff_longitude_2,dropoff_latitude_2,dropoff_region;
```

```
D = FILTER O by trip_distance < 1;
```

```
LOGS_GROUP= GROUP D ALL;
```

```
LOG_COUNT = FOREACH LOGS_GROUP GENERATE COUNT(D);
```

```
dump LOG_COUNT;
```