Data analyzing about Citi Bike

#Find the most popular route with different user types – concat function#

Select usertype, concat(start\_station\_name, ‘to’, end\_station\_name) as route,

Count(\*) as num\_trips,

Round(AVG(cast(tripduration as int64)/60),2) as duration

From New\_york.citibike\_trips

Group by start\_station\_name, end\_station\_name, usertype

Order by num\_trips desc

Limit 10

#compare the number of bikes available at a station to the average number of bikes available#

select

Station\_id,

Num\_bikes\_available,

(select avg(num\_bikes\_available) from new\_york.citibike\_stations) as avg\_num\_bikes\_available

From

New\_york.citibike\_stations

#Calculating the numbers of rides have started at each station over time in descending order#

Select Station\_id, Name, Number\_of\_rides as number\_of\_rides\_starting\_at\_station

From

(select Start\_station\_id, Count(\*) number\_of\_rides

From New\_york.citibike\_stations

Group by  Start\_station\_id)

As station\_num\_trips

Inner join  New\_york.citibike\_stations

On station\_id = start\_station\_id

Order by number\_of\_rides desc

#Want a list of station subscribers used#

Select  station\_id,name

From New\_york.citibike\_stations

Where station\_id in

(select start\_station\_id

From New\_york.citibike\_trips

Where usertype = ‘subscriber’)

#Data calculation about New York citibike data set:  Find out how many rides people took on these bikes per year#

Select extract(Year from starttime) as year,

Count(\*) as number\_of\_rides

From new\_york.citibike\_trips

Group by year

Order by year desc

#Temporary table: analyze data for bike trips that  were over 60 mins#

With trips\_over\_1h as (

Select \* from new\_york.citibike\_trips

Where tripduration >= 60)

#count how manytrips  are 60+ minutes long#

Select count(\*) as cnt

From trips\_over\_1h