```
CREATE TABLE station_totals (
 STATION TEXT, DAILY_ACTIVITY INTEGER,
PRIMARY KEY (STATION,DAILY_ACTIVITY) );

CREATE TABLE mta_zc (
STATION TEXT, DAILY_ACTIVITY INTEGER, ZIPCODE INTEGER,
PRIMARY KEY (STATION,DAILY_ACTIVITY,ZIPCODE) );
CREATE TABLE dropoff_boxes (
Address TEXT, BBL INTEGER, BIN INTEGER, CouncilDist INTEGER, DropOff_SiteName TEXT,
Latitude FLOAT, Longitude FLOAT, NTAName TEXT, Number TEXT, ObjectId INTEGER, Street
TEXT, Zipcode INTEGER, boroCD INTEGER, ct2010 INTEGER, point TEXT,
PRIMARY KEY
(Address,BBL,BIN,CouncilDist,DropOff_SiteName,Latitude,Longitude,NTAName,Number,ObjectI
d,Street,Zipcode,boroCD,ct2010,point) );

CREATE TABLE mta_all_data (
STATION TEXT, DAILY_ACTIVITY INTEGER, BIN INTEGER, ZIPCODE INTEGER, Address TEXT,
PRIMARY KEY (STATION,DAILY_ACTIVITY,BIN,ZIPCODE,Address) );
#created table shells, data was pulled in directly through DB Browser and not any queries

SELECT Address, BIN, Zipcode, count()
from dropoff_boxes
GROUP BY Address, BIN, Zipcode
HAVING count() > 1
#determined what my duplicates are

INSERT INTO dropoff_boxes
SELECT Address, BIN, Zipcode, count()
OVER (PARTITION BY Address) as bin_count from dropoff_boxes
GROUP BY Address, BIN, Zipcode HAVING count() = 1
#created bin_count column

DELETE FROM dropoff_boxes
WHERE BIN_COUNT IS NULL;
#deleted duplicate bins from dataset

INSERT INTO mta_all_data
SELECT z.STATION, z.ZIPCODE, b.Address, b.BIN_COUNT, s.DAILY_ACTIVITY
FROM mta_zc z INNER JOIN dropoff_boxes b ON z.ZIPCODE = b.ZIPCODE INNER JOIN
station_totals s ON z.STATION = s.STATION
#joined all three datasets together

SELECT ZIPCODE, SUM(BIN_COUNT), SUM(DAILY_ACTIVITY)
round(SUM(DAILY_ACTIVITY)/SUM(BIN_COUNT)) AS RATIO
```

```
 From mta_all_data
 GROUP BY zipcode;
#aggregated data to find total activity and bin count within a zipcode and used that to create
ratio column
```