CREATE TABLE fare (

medallion VARCHAR(255) NOT NULL ,

hack\_license VARCHAR(255) NOT NULL,

vendor\_id char(3) NOT NULL,

pickup\_datetime datetime NOT NULL,

payment\_type char(3) NOT NULL,

fare\_amount decimal(10,2) NOT NULL,

surcharge decimal(10,2) NOT NULL,

mta\_tax decimal(10,2) NOT NULL,

tip\_amount decimal(10,2) NOT NULL,

tolls\_amount decimal(10,2) NOT NULL,

total\_amount decimal(10,2) NOT NULL,

PRIMARY KEY (medallion, hack\_license, vendor\_id, pickup\_datetime)

);

LOAD DATA LOCAL INFILE '/Users/maloney/Documents/trip\_fare\_3.csv'

INTO TABLE fare

FIELDS TERMINATED BY ','

-- ENCLOSED BY '"'

LINES TERMINATED BY '\n'

IGNORE 1 ROWS;

CREATE TABLE trip\_data (

medallion VARCHAR(255) NOT NULL ,

hack\_license VARCHAR(255) NOT NULL,

vendor\_id varchar(3) NOT NULL,

rate\_code varchar(3) NOT NULL,

store\_and\_fwd\_flag varchar(3) NOT NULL,

pickup\_datetime datetime NOT NULL,

dropoff\_datetime datetime NOT NULL,

passenger\_count int NOT NULL,

trip\_time\_in\_secs float NOT NULL,

trip\_distance float NOT NULL,

pickup\_longitude float NOT NULL,

pickup\_latitude float NOT NULL,

dropoff\_longitude float NOT NULL,

dropoff\_latitude float NOT NULL,

PRIMARY KEY (medallion, hack\_license, vendor\_id, pickup\_datetime)

);

LOAD DATA LOCAL INFILE '/Users/maloney/Documents/trip\_data\_3.csv'

INTO TABLE trip\_data

FIELDS TERMINATED BY ','

-- ENCLOSED BY '"'

LINES TERMINATED BY '\n'

IGNORE 1 ROWS;

-- ###What fraction of payments under $5 use a credit card \*

-- TABLE DECLARATION ##################################################

CREATE TEMPORARY TABLE TABLE1 (NUMERATOR INT, DENOMINATOR INT, RATIO float) ENGINE=MEMORY;

-- #####################################################################

-- WHAT GETS INSERTED INTO TABLE 1

INSERT INTO TABLE1

SELECT

A.NUM,

A.DENOM,

(convert(A.NUM,DECIMAL(20,10))/convert(A.DENOM,DECIMAL(20,10))) as ratio

FROM

(

-- COLUMN SELECTION. TWO NUMBERS WILL REPRESENT A NUM AND A DENOM

SELECT

(select count(\*) from fare

where

payment\_type = "CRD"

and

total\_amount < 5)

AS NUM,

(select count(\*) from fare

where

total\_amount < 5)

AS DENOM

)A;

-- select cast(NUMERATOR as FLOAT) from TABLE1;

SELECT NUMERATOR, DENOMINATOR, RATIO

FROM TABLE1;

-- 0.0887821

DROP TABLE TABLE1;

-- ###What fraction of payments over $50 use a credit card\*

-- 0.681673

-- TABLE DECLARATION ##################################################

CREATE TEMPORARY TABLE TABLE1 (NUMERATOR INT, DENOMINATOR INT, RATIO float) ENGINE=MEMORY;

-- #####################################################################

-- WHAT GETS INSERTED INTO TABLE 1

INSERT INTO TABLE1

SELECT

A.NUM,

A.DENOM,

(convert(A.NUM,DECIMAL(20,10))/convert(A.DENOM,DECIMAL(20,10))) as ratio

FROM

(

-- COLUMN SELECTION. TWO NUMBERS WILL REPRESENT A NUM AND A DENOM

SELECT

(select count(\*) from fare

where

payment\_type = "CRD"

and

total\_amount > 50)

AS NUM,

(select count(\*) from fare

where

total\_amount > 50)

AS DENOM

)A;

-- select cast(NUMERATOR as FLOAT) from TABLE1;

SELECT NUMERATOR, DENOMINATOR, RATIO

FROM TABLE1;

DROP TABLE TABLE1;

-- CREATE A MERGED TABLE TO ANSWER THE OTHER QUESTIONS

-- TABLE DECLARATION ##################################################

CREATE TABLE TABLE1

( medallion VARCHAR(255) NOT NULL ,

hack\_license VARCHAR(255) NOT NULL,

vendor\_id varchar(3) NOT NULL,

rate\_code varchar(3) NOT NULL,

store\_and\_fwd\_flag varchar(3) NOT NULL,

pickup\_datetime datetime NOT NULL,

dropoff\_datetime datetime NOT NULL,

passenger\_count int NOT NULL,

trip\_time\_in\_secs float NOT NULL,

trip\_distance float NOT NULL,

pickup\_longitude float NOT NULL,

pickup\_latitude float NOT NULL,

dropoff\_longitude float NOT NULL,

dropoff\_latitude float NOT NULL,

payment\_type char(3) NOT NULL,

fare\_amount decimal(10,2) NOT NULL,

surcharge decimal(10,2) NOT NULL,

mta\_tax decimal(10,2) NOT NULL,

tip\_amount decimal(10,2) NOT NULL,

tolls\_amount decimal(10,2) NOT NULL,

total\_amount decimal(10,2) NOT NULL,

PRIMARY KEY (medallion, hack\_license, vendor\_id, pickup\_datetime)

);

-- #####################################################################

-- WHAT GETS INSERTED INTO TABLE 1

INSERT INTO TABLE1

SELECT

trip\_data.medallion,

trip\_data.hack\_license,

trip\_data.vendor\_id,

trip\_data.rate\_code,

trip\_data.store\_and\_fwd\_flag,

trip\_data.pickup\_datetime,

trip\_data.dropoff\_datetime,

trip\_data.passenger\_count,

trip\_data.trip\_time\_in\_secs,

trip\_data.trip\_distance,

trip\_data.pickup\_longitude,

trip\_data.pickup\_latitude,

trip\_data.dropoff\_longitude,

trip\_data.dropoff\_latitude,

fare.payment\_type,

fare.fare\_amount,

fare.surcharge,

fare.mta\_tax,

fare.tip\_amount,

fare.tolls\_amount,

fare.total\_amount

FROM

trip\_data trip\_data,

fare fare

where

trip\_data.medallion = fare.medallion

and

trip\_data.hack\_license = fare.hack\_license

and

trip\_data.vendor\_id = fare.vendor\_id

and

trip\_data.pickup\_datetime = fare.pickup\_datetime;

library("RODBC")

odbcDataSources()

ch <- odbcConnect("database1")

#### What is the mean fare per minute driven?\*

df3 <- sqlQuery(ch, paste('select avg((fare\_amount) / (trip\_time\_in\_secs/60)) as mean\_fare\_per\_minute from TABLE1'))

transform(df3, mean\_fare\_per\_minute = as.numeric(mean\_fare\_per\_minute))

mean(df3$mean\_fare\_per\_minute, na.rm=TRUE)

####1.412929

### What is the median of the taxi's fare per mile driven?\*\*

df4 <- sqlQuery(ch, paste('select cast(fare\_amount /trip\_distance as DECIMAL) as fare\_per\_mile from TABLE1'))

transform(df5, fare\_per\_mile = as.numeric(fare\_per\_mile))

median(df4$fare\_per\_mile, na.rm=TRUE)

###$6.27

###What is the 95 percentile of the taxi's average driving speed in miles per hour?\*

df5 <- sqlQuery(ch, paste('select

(trip\_distance / (trip\_time\_in\_secs / 3600)) as ave\_drive\_speed

from TABLE1'))

transform(df5, ave\_drive\_speed = as.numeric(ave\_drive\_speed))

quantile(df5$ave\_drive\_speed, c(.25, .50, .95), na.rm=TRUE)

###26.58261

### What is the average ratio of the distance between the pickup and dropoff divided by the distance driven?\*

df6 <- sqlQuery(ch, paste('SELECT trip\_distance,

(69 \* (DEGREES(ACOS(COS(RADIANS(pickup\_latitude)) \* COS(RADIANS(dropoff\_latitude)) \*

COS(RADIANS(pickup\_longitude) - RADIANS(dropoff\_longitude)) +

SIN(RADIANS(pickup\_latitude)) \* SIN(RADIANS(dropoff\_latitude)))))) AS actual\_ml

FROM TABLE1

where

trip\_distance <> 0

and

(pickup\_latitude <> 0 and dropoff\_latitude <> 0 and pickup\_longitude <> 0 and dropoff\_longitude <> 0 )'))

mean(df6$actual\_ml/df6$trip\_distance, na.rm=TRUE)

### 0.9713058

### What is the average tip for rides from JFK?\*

###Perimeter Road, Jamaica, NY 11430, USA"

### "lat" : 40.6409891,

### "lng" : -73.77432019999999

### using within a 30 mile radius of googles accepted location as it is 880 square miles

df7 <- sqlQuery(ch, paste('select tip\_amount,

(69 \* (DEGREES(ACOS(COS(RADIANS(pickup\_latitude)) \* COS(RADIANS(40.6409891)) \*

COS(RADIANS(pickup\_longitude) - RADIANS(-73.77432019999999)) +

SIN(RADIANS(pickup\_latitude)) \* SIN(RADIANS(40.6409891)))))) AS distance\_from\_jfk

from TABLE1

HAVING distance\_from\_jfk < 30'))

mean(df7$tip\_amount, na.rm=TRUE)

### 1.303936

### What is the median March revenue of a taxi driver?\*

### DEFINE REVENUE AS TOTAL AMMOUNT

### DEFINE A DRIVER AS DISTINCT HACK LICENSE AND NOT A COMBO OF MEDIALION AS THEY COULD CHANGE CABS (32,991)

### ASSUME DRIVER KEEP ANY FARE STARTED IN MARCH FOR MARCH (THERE ARE SOME THAT END IN APRIL, THIS COULD BE HANDLE WITH EXCLUDING END TIMES)

df8 <- sqlQuery(ch, paste('select hack\_license, SUM(total\_amount) as revenue

from TABLE1

group by 1'))

str(df8)

median(df8$revenue, na.rm=TRUE)

### 7221.2