Name: XinYue Liu

Student ID: 1332-0443-43

Email: liuxinyu@usc.edu

Part 1:

Query #1

```
#standardSQL
SELECT
 name, count
FROM
 `babynames.names_2014`
WHERE
 gender = 'F'
 AND
 name LIKE "_i%"
ORDER BY count DESC;
```

Query #2

```
#standardSQL
SELECT
 SUM(count)
FROM
 `babynames.names_2014`
WHERE
 name LIKE "Xin%";
```

result: 32

Part 2:

Query 2nd cell:

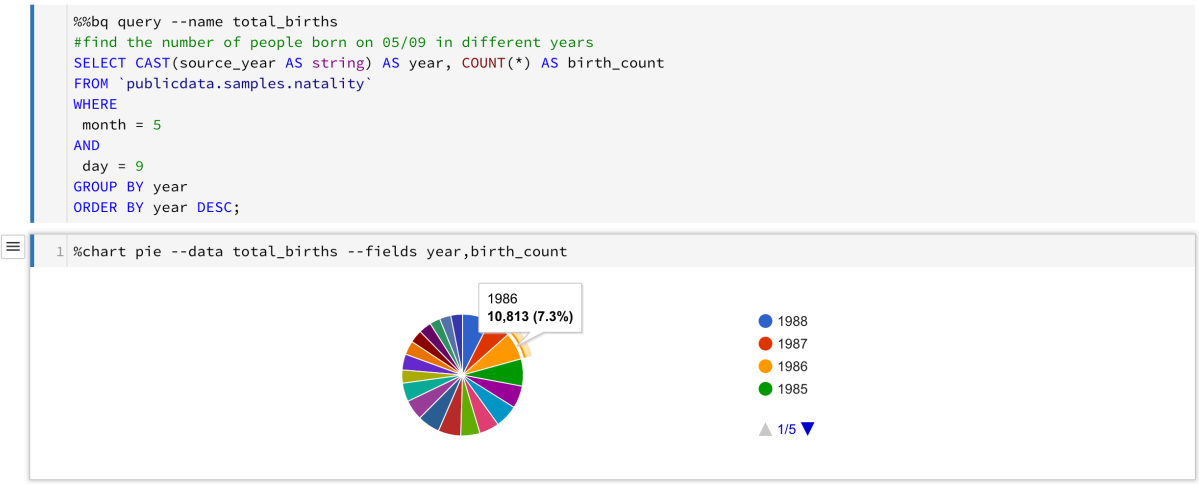
```
%%bq query
#find the weekday of 1994/05/09
SELECT wday
FROM
 `publicdata.samples.natality`
WHERE
 year = 1994
 AND
 month = 5
 AND
 day = 9;
```

Query 3rd cell:

```
%%bq query --name total_births
#find the number of people born on 05/09 in different years
SELECT CAST(source_year AS string) AS year, COUNT(*) AS birth_count
FROM `publicdata.samples.natality`
WHERE
 month = 5
AND
 day = 9
GROUP BY year
ORDER BY year DESC;

%chart pie --data total_births --fields year,birth_count
```

Snapshot of visualization:



Part 3:

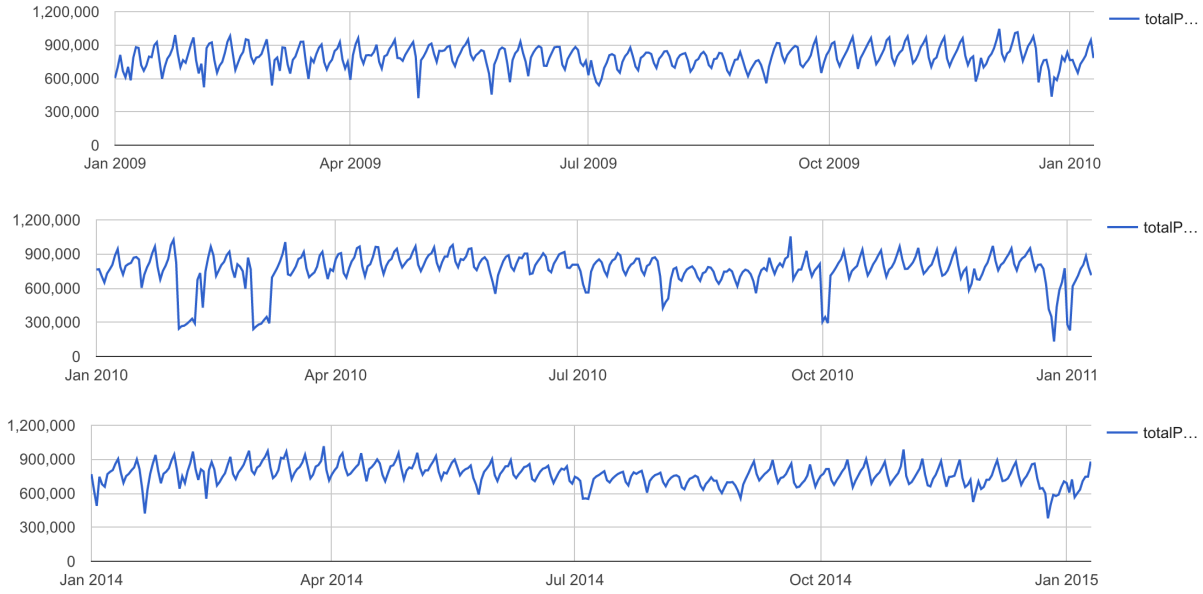
1. BigQuery:

```
SELECT DATE(pickup_datetime) AS dateTime, SUM(passenger_count) AS totalPassenger
FROM `nyc-tlc.yellow.trips`
WHERE DATE(pickup_datetime) < '2015-01-01'
AND DATE(dropoff_datetime) < '2015-01-01'
GROUP BY dateTime
ORDER BY dateTime;
```

2.Datalab:

```
%%bq query
SELECT DATE(pickup_datetime) AS dateTime, SUM(passenger_count) AS totalPassenger
FROM `nyc-tlc.yellow.trips`
WHERE DATE(pickup_datetime) < '2015-01-01'
AND DATE(dropoff_datetime) < '2015-01-01'
GROUP BY dateTime
ORDER BY dateTime;
```

Visualization:



Semi-periodical pattern:

Based on my observation, I think a peak of a week is Sunday.

Query to prove:

```
%%bq query
SELECT dateTime,EXTRACT(DAYOFWEEK FROM dateTime) dayOfweek
FROM
(SELECT DATE(pickup_datetime) AS dateTime, SUM(passenger_count) AS totalPassenger
FROM `nyc-tlc.yellow.trips`
WHERE DATE(pickup_datetime) BETWEEN '2010-01-01' AND '2011-01-10'
AND DATE(dropoff_datetime) BETWEEN '2010-01-01' AND '2011-01-10'
GROUP BY dateTime
HAVING totalPassenger > 950000
ORDER BY dateTime);
```

```
%%bq query
SELECT dateTime,EXTRACT(DAYOFWEEK FROM dateTime) dayOfweek
FROM
(SELECT DATE(pickup_datetime) AS dateTime, SUM(passenger_count) AS totalPassenger
FROM `nyc-tlc.yellow.trips`
WHERE DATE(pickup_datetime) BETWEEN '2014-01-01' AND '2015-01-10'
AND DATE(dropoff_datetime) BETWEEN '2014-01-01' AND '2015-01-10'
GROUP BY dateTime
HAVING totalPassenger > 950000
ORDER BY dateTime);
```

dateTime	dayOfweek
2014-03-01	7
2014-04-05	7
2014-03-15	7
2014-11-01	7
2014-04-26	7
2014-03-29	7
2014-05-03	7
2014-02-08	7
2014-03-08	7
2014-04-12	7

(rows: 10, time: 0.2s, cached, job: job_AEBjkuaOxF5HWD_G_xGyH-wyfC0)

Two unusual patterns (two holidays):

1. **Martin Luther King Jr. Day**
2. **Christmas Day**

Explanation: Due to those two holidays, people are more willing to stay home with their families. Therefore, there are big decreases in numbers around these two holidays.

Minimum Point:

2011-08-28 Hurricane Irene hits New York
2012-10-29 Time Lapse of Hurricane Sandy hitting New York City October 29th ...

2013-08-04 In August 2013 the City of New York introduced a restricted class of “boro” (or “green”) taxis. Boro taxis provide an equivalent service to that of “yellow” medallion taxis, but are restricted from pickups in Manhattan south of a boundary along East 96th St and West 110th St.

Bonus:

Query:

```
#legacySQL
SELECT pickup_longitude,pickup_latitude
FROM [nyc-tlc:yellow.trips]
WHERE DATE(pickup_datetime) < '2013-01-01'
AND DATE(dropoff_datetime) < '2013-12-31'
AND HOUR(pickup_datetime) > 18
AND pickup_longitude != 0
AND pickup_latitude != 0
AND total_amount >= 300.0
AND total_amount <= 400.0;
```

[Download as pickup.csv]

```
#legacySQL
SELECT dropoff_longitude,dropoff_latitude
FROM [nyc-tlc:yellow.trips]
WHERE DATE(pickup_datetime) < '2013-01-01'
AND DATE(dropoff_datetime) < '2013-12-31'
AND HOUR(pickup_datetime) > 18
AND dropoff_longitude != 0
AND dropoff_latitude != 0
AND total_amount >= 300.0
AND total_amount <= 400.0;
```

[Download as dropoff.csv]

Load pickup.csv and dropoff.csv into google my maps like below:
https://drive.google.com/open?id=1_aUWKbNzHknMP9w7jkke1TYdYT4&usp=sharing