Bigquery public data: https://cloud.google.com/bigquery/public-data/bay-bike-share

**1. Please look further into the source data and see if bike shares were downward trending in 2016. (Hints: monthly data)**

SELECT

EXTRACT(MONTH

FROM

start\_date) AS MONTH,

COUNT(\*) AS Monthly\_trips

FROM

`bigquery-public-data.san\_francisco.bikeshare\_trips`

WHERE

EXTRACT(year

FROM

start\_date) = 2016

GROUP BY

MONTH

ORDER BY

MONTH

**2. How is subscriber monthly usage growth comparing to customer ?**

SELECT

EXTRACT(YEAR

FROM

start\_date) AS YEAR,

EXTRACT(MONTH

FROM

start\_date) AS MONTH,

subscriber\_type,

COUNT(\*) AS Monthly\_trips

FROM

`bigquery-public-data.san\_francisco.bikeshare\_trips`

GROUP BY

YEAR,

MONTH,

subscriber\_type

ORDER BY

YEAR,

MONTH

**3. How does seasonality affect bike shares? We can define Dec-Feb as Winter season, etc. use case statement.**

**For this question, I thought about two methods. The first one is to consider the seasonality without year; the second one is to consider it with year. I finally chose the latter to make the answer more precise. But to conduct this method, I have to use other tools to get the final output.**

**Step1: standard SQL**

select

table\_1.month, table\_1.year, table\_1.monthly\_trips,

CASE

WHEN table\_1.month IN (12, 1, 2) THEN "Winter"

WHEN table\_1.month IN (3,4,5) THEN "Spring"

WHEN table\_1.month in (6,7,8) THEN "Summer"

ELSE "Autumn"

END as season

from

(SELECT

EXTRACT(month

FROM

start\_date) AS month,

EXTRACT(YEAR

FROM

start\_date) AS year,

COUNT(\*) AS monthly\_trips

FROM

`bigquery-public-data.san\_francisco.bikeshare\_trips`

GROUP BY

year, month) as table\_1

**Step2:** use “if” function in excel to change the year when the month is 1 or 2. Here, I assume that January or February belongs to last year’s winter. For example, 2014.1 should be 2013 winter.

**Step3**: use excel to make pivotchart

From this picture, we can see that in winter bike shares will decrease significantly. And there is no significant evidence showing that other three seasons have stronger influence in bike shares than each other.

**4. What is avg lifespan of the bikes in this dataset?**

select round(avg(lifespan),2) as average\_lifespan from

(select Date\_diff(max(date(end\_date)),min(date(start\_date)),day) as lifespan, bike\_number

from `bigquery-public-data.san\_francisco.bikeshare\_trips`

group by bike\_number)

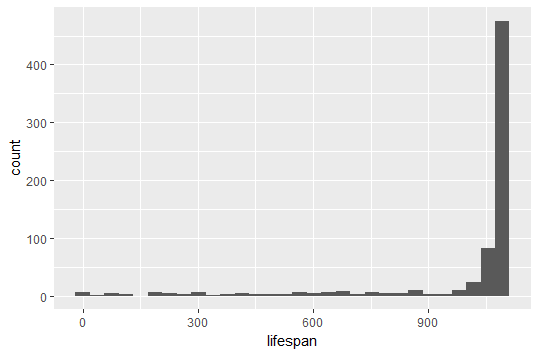
997.4986

**histogram and show distribution of your results**

select Date\_diff(max(date(end\_date)),min(date(start\_date)),day) as lifespan, bike\_number

from `bigquery-public-data.san\_francisco.bikeshare\_trips`

group by bike\_number



**5. What are the top 3 start-end station pairings each year?**

select \* from

(select t1.year, t1.start\_station\_name, t1.end\_station\_name, t1.trips,

rank() over (partition by t1.year order by t1.trips desc) as ranking

from

(select extract(year from start\_date) as year,

start\_station\_name, end\_station\_name,

count(\*) as trips

from `bigquery-public-data.san\_francisco.bikeshare\_trips`

group by year, start\_station\_name, end\_station\_name

order by year

)t1

)t2

where ranking <=3

order by year

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| year | start\_station\_name | end\_station\_name | trips | the\_rank |
| 2013 | Harry Bridges Plaza (Ferry Building) | Embarcadero at Sansome | 945 | 1 |
| 2013 | Townsend at 7th | San Francisco Caltrain (Townsend at 4th) | 807 | 2 |
| 2013 | San Francisco Caltrain 2 (330 Townsend) | Townsend at 7th | 716 | 3 |
| 2014 | Townsend at 7th | San Francisco Caltrain (Townsend at 4th) | 3158 | 1 |
| 2014 | San Francisco Caltrain 2 (330 Townsend) | Townsend at 7th | 2937 | 2 |
| 2014 | Harry Bridges Plaza (Ferry Building) | Embarcadero at Sansome | 2826 | 3 |
| 2015 | Townsend at 7th | San Francisco Caltrain 2 (330 Townsend) | 3603 | 1 |
| 2015 | San Francisco Caltrain 2 (330 Townsend) | Townsend at 7th | 3435 | 2 |
| 2015 | Harry Bridges Plaza (Ferry Building) | Embarcadero at Sansome | 3253 | 3 |
| 2016 | Market at 10th | San Francisco Caltrain 2 (330 Townsend) | 2196 | 1 |
| 2016 | Harry Bridges Plaza (Ferry Building) | Embarcadero at Sansome | 2126 | 2 |
| 2016 | Townsend at 7th | San Francisco Caltrain 2 (330 Townsend) | 2068 | 3 |

**6. Prepare dataset to visualize # of bikes available on Sept 1, 2015 on a map view (join)**

select distinct(t1.station\_id), t1.bikes\_available, landmark

from

(select distinct (`bigquery-public-data.san\_francisco.bikeshare\_status`.station\_id) as station\_id, max(bikes\_available) as bikes\_available

from `bigquery-public-data.san\_francisco.bikeshare\_status`

where bikes\_available > 0 and date( time) = date(2015,9,1)

group by `bigquery-public-data.san\_francisco.bikeshare\_status`.station\_id) as t1

inner join `bigquery-public-data.san\_francisco.bikeshare\_stations`

on t1.station\_id = `bigquery-public-data.san\_francisco.bikeshare\_stations`.station\_id

**7. Do we have any station without any activities on a particular day of 2015? (left join)**

For this question, I assume that in table *`bigquery-public-data.san\_francisco.bikeshare\_status`,* there are data for the existing bike share stations every single day.

select t1.date,t1.station

from

(select distinct(station\_id) as station, extract(date from time) as date

from `bigquery-public-data.san\_francisco.bikeshare\_status`

where extract(year from time) = 2015

)t1

left join

(select distinct(start\_station\_id) as start\_station,

extract(date from start\_date) as date

from `bigquery-public-data.san\_francisco.bikeshare\_trips`

where extract(year from start\_date) = 2015

)t2

on t1.station = t2.start\_station and t1.date = t2.date

left join

(select distinct(end\_station\_id) as end\_station,

extract(date from end\_date) as date

from `bigquery-public-data.san\_francisco.bikeshare\_trips`

where extract(year from end\_date) = 2015

)t3

on t1.station = t3.end\_station and t1.date = t3.date

where t2.start\_station is null and t3.end\_station is null

order by t1.date,t1.station