Assignment No. 3 – Hubway Bikes

Syed Umair Hassan Kazmi

Q1. What was the duration of the longest trip? What was the average trip duration? What was the smallest trip duration?

Ans) 1 - The duration for the **longest** trip was **9999 seconds**
2 - **Average** trip duration was **912.4097 seconds**
3 - The **smallest** trip duration was **0 seconds**

SELECT max(duration) as max_trip_duration , avg(duration) as avg_trip_duration , min(duration) as min_trip_duration FROM trips;

	max_trip_duration	avg_trip_duration	min_trip_duration
1	9999	912.409681904661	0

Q2. How many trips were taken by 'Registered' users?

Ans) The number of trips taken by registered user are **1105192**

1 - SELECT sub_type, count(id) FROM trips GROUP BY sub_type;

2 - SELECT sub_type, count(id) FROM trips WHERE sub_type= 'Registered';

3 - SELECT sub_type, count(id) FROM trips WHERE lower(sub_type) like '%register%'

All 3 queries work but the group by query also shows other groups and their respected user but in terms of speed , it ranks 2nd. 'Where =' ranks 1st while 'Where like' ranks 3rd

	sub_type	count(id)
1	Casual	464809
2	Registered	1105192

Q3. How many trips were taken by male users in comparison to Female users? for Registered users only

Ans) Number of registered **male** users are **833858** while Numbers of **female** registered users are (**271333 + 1) = 271334**

SELECT gender, sub_type, count(*) FROM trips WHERE sub_type = 'Registered' GROUP BY gender;

	gender	sub_type	count(*)
1	Female	Registered	271333
2	Female	Registered	1
3	Male	Registered	833858

Q4. Do registered or casual users take longer trips?

Ans) On Average casual user take longer time to travel then registered users

SELECT sub_type, avg(duration) FROM trips GROUP BY sub_type;

	sub_type	avg(duration)
1	Casual	1519.64389674038
2	Registered	657.02606696393

Q5. Which bike was used for the most trips?

Ans) Bike **B00490** was used the most for a number of **2120** trips

SELECT bike_number, count(bike_number) as no_of_bikes FROM trips
GROUP BY bike_number ORDER BY count(bike_number) DESC LIMIT 1;

	bike_number	no_of_bikes
1	B00490	2120

Q6. What is the average duration of trips by users over the age of 30?

Ans) Average duration of trips for age over 30 is **913.929177395571** seconds

SELECT avg(duration) AS avg_duration_over_30 FROM trips WHERE (DATE('now') - birth_date) > 30;

	avg_duration_over_30
1	913.929177395571

Q7. Which stations are most frequently used for round trips?

Ans) **Station 58, 36, 42, 53, 52** are the most frequent station for round trips. Respective names can be seen in the query

```
SELECT B.station, A.start_station as station_id, count(A.start_station) as station_frequency
FROM trips  as A
JOIN stations as B on A.start_station = B.id
WHERE start_station = end_station GROUP BY start_station ORDER BY count(*) DESC
LIMIT 5;
```

	station	station_id	station_frequency
1	The Esplanade - Beacon St.at Arlington St.	58	3064
2	Boston Public Library - 700 Boylston St.	36	2548
3	Boylston St.at Arlington St.	42	2163
4	Beacon St / Mass Ave	53	2144
5	Newbury St / Hereford St	52	1636

Q8. How many trips start and end in different municipalities? Station Table has the Municipality Attribute

Ans) **309748** trips end in different municipalities

```
SELECT  count(*) as no_trips_with_diff_municipality FROM trips  as  A
LEFT JOIN stations as B on A.start_station = B.id
LEFT JOIN stations as C on A.end_station = C.id
WHERE B.municipality <> C.municipality;
```

	no_trips_with_diff_municipality
1	309748

Q9. How many trips incurred additional fees (lasted longer than 30 minutes)?

Ans) **119917** trips incurred additional fees

SELECT count(*) as additional_fees FROM trips WHERE (duration/60) > 30

	additional_fees
1	119917

Q10. Which bike was used for the longest total time? Provide the answer in Hours

Ans) Bike No. **B00490** has the most duration in hrs – **571 hrs**

SELECT bike_number, duration_in_hrs/3600 as duration_in_hrs FROM
(SELECT bike_number, sum(duration) as duration_in_hrs FROM trips GROUP BY
bike_number ORDER BY sum(duration) DESC LIMIT 1);

	bike_number	duration_in_hrs
1	B00490	571

Q11. Did registered or casual users take more round trips? A Round Trip is considered as Same Start and End Destination

Ans) **Casual** users took more round trips

SELECT sub_type, count(*) FROM trips
WHERE start_station = end_station GROUP BY sub_type;

	sub_type	count(*)
1	Casual	41427
2	Registered	31641

Q12. Which municipality had the most frequent Station End?

Ans) **Boston** had the most frequent **station end**

SELECT municipality, count(*) FROM trips as A
LEFT JOIN stations as B on A.end_station = B.id

GROUP BY B.municipality ORDER BY count(*) DESC LIMIT 1;

	municipality	count(*)
1	Boston	1212364

Q13. Which From and To Routes are Frequent with Regards to Municipalities?

Ans) **Boston to Boston** is the most frequent from and to route

```
SELECT  B.municipality as start_municipality, C.municipality as end_municipality, count(*)
FROM trips  as  A
LEFT JOIN stations as B on A.start_station = B.id
LEFT JOIN stations as C on A.end_station = C.id
GROUP BY B.municipality, C.municipality
ORDER BY count(*) DESC
LIMIT 5;
```

	start_municipality	end_municipality	count(*)
1	Boston	Boston	1081805
2	Cambridge	Cambridge	162538
3	Boston	Cambridge	110968
4	Cambridge	Boston	110078
5	Cambridge	Somerville	20998

Q14. Which Age Band has the most Rides? Calculate Age Bands based on interval of 10 and Cut Off by 70 or Above

Ans) **30 – 39 age gap** have the most rides **after** a group which **didn't mention their ages.**

```
SELECT
(CASE WHEN (DATE('now') - birth_date) < 20 THEN '0 - 19'
   WHEN DATE('now') - birth_date  >= 20 And DATE('now') - birth_date < 30 THEN '20 -
29'
   WHEN DATE('now') - birth_date  >= 30 And DATE('now') - birth_date < 40 THEN '30 -
39'
   WHEN DATE('now') - birth_date  >= 40 And DATE('now') - birth_date < 50 THEN '40 -
49'
   WHEN DATE('now') - birth_date  >= 50 And DATE('now') - birth_date < 60 THEN '50 -
59'
   WHEN DATE('now') - birth_date  >= 60 And DATE('now') - birth_date < 70 THEN '60 -
69'
```

```
   WHEN DATE('now') - birth_date  >= 70 And DATE('now') - birth_date < 130 THEN '70+'
   ELSE 'age not mentioned'
   END) AS age_group, count(*)
FROM trips GROUP BY age_group ORDER BY count(*) DESC;
```

	age_group	count(*)
1	age not mentioned	1219985
2	30 - 39	127764
3	40 - 49	107600
4	50 - 59	60527
5	60 - 69	40883
6	70+	10338
7	20 - 29	2904

Q15. What Day of the Week is most Popular for Rides? What Time (24 Hrs Format) is Popular on that Particular Date?

Ans) **Wednesday** is most popular and **17th hour** is most popular hour of that day.

```
SELECT  (CASE CAST(A.Day_n as INTEGER)
  when 0 then 'Sunday'
  when 1 then 'Monday'
  when 2 then 'Tuesday'
  when 3 then 'Wednesday'
  when 4 then 'Thursday'
  when 5 then 'Friday'
  else 'Saturday'
  end) as Day,  A.*
FROM ( Select strftime('%w',start_date)  as Day_n, strftime('%H',start_date) as hour,
count(*) as no_trips_by_hour
FROM trips GROUP BY Day_n, hour ORDER BY count(*)) as A
JOIN (SELECT strftime('%w',start_date) as Day_n
FROM trips  GROUP BY Day_n ORDER BY count(*) DESC LIMIT 1) AS B
ON A.Day_n = B.Day_n ORDER BY A.no_trips_by_hour DESC limit 1;
```

	Day	Day_n	hour	no_trips_by_hour
1	Wednesday	3	17	29582

Q.16) What hour along with its day is most popular?

Ans) Also, most popular hour along with its day is **Monday's 17th hour**

```
SELECT  (CASE CAST( strftime('%w',start_date) as INTEGER)
  when 0 then 'Sunday'
  when 1 then 'Monday'
  when 2 then 'Tuesday'
  when 3 then 'Wednesday'
  when 4 then 'Thursday'
  when 5 then 'Friday'
  else 'Saturday'
  end) as Day, strftime('%H',start_date) as hour, count(*)
  FROM trips GROUP BY day, hour ORDER BY count(*) DESC LIMIT 5;
```

	Day	hour	count(*)
1	Monday	17	29733
2	Tuesday	17	29693
3	Wednesday	17	29582
4	Thursday	17	28798
5	Wednesday	08	25397

Do give the review sir, thanks.