# CYCLISTIC ANNUAL TREND - LOGBOOK

1. Download 12 csv files :
    a. from 202008-divvy-tripdata.csv to 202107-divvy-tripdata.csv
    b. unzip the files in: Course 8 Google Data Analytics Capstone\Project 1 – Data\Cyclistic Zip files\tripdata_unzip_files

2. Open the csv file in excel: divvy_tripdata
3. Create column : ride_lenght, days_of_week, d_o_w_no
4. edit the 12 csv
5. save files in Course 8 Google Data Analytics Capstone\Project 1 – Data\Cyclistic Zip files\tripdata_unzip_files\v2_corrected

6. Import all 12 .csv into SQLite Database: tripdata  TABLE: Details

12 months from August 2020 to July 2021 totaling 4,731,081 rows

For each files: add 3 columns:

ride_length = ended_at – started_at format: dd hh:mm:ss

days =DAY(ride_length) format: number (no decimal)

hours =HOUR(ride_length) format: number (no decimal)

minutes =MINUTE(ride_length) format: number(no decimal)

seconds =SECOND(ride_length) format: number(no decimal)

total_seconds = (days*86400)+(hours*3600)+(minutes*60)+seconds

doy_of_week =WEEKDAY(<cells>,1) formatted as dddd (Sunday…Saturday)

dow_no = WEEKDAY(<cells>,1) formatted as d (1…7)

After apply a filter on ride_length to clean the data, then save as v2 file

a. 202107: 822397 rows +
    i. 13 rows deleted because ended_at was before started_at
    ii. 12695 rows has ride_length < 1 minute (will be deleted in SQL)
b. 202106: 729590 rows +
    i. 5 rows deleted because ended_at was before started_at
    ii. 11891 rows has ride_length < 1 minute (will be deleted in SQL)
c. 202105: 531631 rows +
    i. 2 rows deleted because ended_at was before started_at
    ii. 8388 rows has ride_length < 1 minute (will be deleted in SQL)
d. 202104: 337225 rows +
    i. 5 rows deleted because ended_at was before started_at
    ii. 4803 rows has ride_length < 1 minute (will be deleted in SQL)
e. 202103: 228494 rows +
    i. 2 rows deleted because ended_at was before started_at
    ii. 2971 rows has ride_length < 1 minute (will be deleted in SQL)
f. 202102: 49622 rows +
    i. 0 rows deleted because ended_at was before started_at
    ii. 973 rows has ride_length < 1 minute (will be deleted in SQL)
g. 202101: 96832 rows +
    i. 2 rows deleted because ended_at was before started_at
    ii. 1417 rows has ride_length < 1 minute (will be deleted in SQL)
h. 202012: 131139 rows +
    i. 434 rows deleted because ended_at was before started_at
    ii. 1642 rows has ride_length < 1 minute (will be deleted in SQL)
i. 202011: 258851 rows +
    i. 865 rows deleted because ended_at was before started_at
    ii. 3689 rows has ride_length < 1 minute (will be deleted in SQL)
j. 202010: 386742 rows +
    i. 1911 rows deleted because ended_at was before started_at
    ii. 6838 rows has ride_length < 1 minute (will be deleted in SQL)
k. 202009: 530826 rows +
    i. 2132 rows deleted because ended_at was before than started_at
    ii. 8525 rows has ride_length < 1 minute (will be deleted in SQL)
l. 202008: 622361 rows +
    i. 2769 rows deleted because ended_at was before started_at
    ii. 10152 rows has ride_length < 1 minute (will be deleted in SQL)

TABLE FIELDS

ride_id

rideable_type

started_at

ended_at

ride_length

days

hours

minutes

seconds

total_seconds

day_of_week

dow_no

start_station_name

start_station_id

end_station_name

end_station_id

start_lat

start_lng

end_lat

end_lng

member_casual

7. Create an Excel files Summary_stat.xlsx to keep some information and stats about the dataset
8. SQLite Query

```
SELECT member_casual, AVG(total_seconds)/60 as total_minutes
FROM tripdata.details
GROUP BY member_casual
```

RESULT

| | |
|---|---|
| casual | 36.53106698620879 |
| member | 14.73532070024562 |

```
SELECT rideable_type,COUNT(rideable_type)
FROM tripdata.details
GROUP BY rideable_type
```

RESULT

| | |
|---|---|
| classic_bike | 1785493 |
| docked_bike | 1550445 |
| electric_bike | 1387003 |

```
SELECT member_casual,COUNT(member_casual)
FROM tripdata.details
GROUP BY member_casual
```

RESULT

| | |
|---|---|
| casual | 2099671 |
| member | 2623270 |

```
SELECT COUNT(DISTINCT(start_station_id))
FROM tripdata.details
```

RESULT

1280

```
SELECT COUNT(DISTINCT(start_station_name))
FROM tripdata.details
```

RESULT

739

```
SELECT COUNT(DISTINCT(end_station_id))
FROM
tripdata.details
```

RESULT

1279

```
SELECT COUNT(DISTINCT(end_station_name))
FROM
tripdata.details
```

RESULT

736

```
SELECT day_of_week, AVG(ride_length) AS avg_ride
FROM tripdata.details
GROUP BY (day_of_week)
ORDER BY (avg_ride) DESC
```

RESULT

| | |
|---|---|
| Sunday | 0.00280914424887 |
| Saturday | 0.00222737777095 |
| Friday | 0.00189900146945 |
| Monday | 0.00164934622613 |
| Thursday | 0.00143996835665 |
| Wednesday | 0.00142693429753 |
| Tuesday | 0.00119346181834 |

```
SELECT day_of_week, AVG(total_seconds)/60 AS avg_ride_minutes
FROM tripdata.details
GROUP BY (day_of_week)
ORDER BY (avg_ride_minutes) DESC
```

RESULT

| | |
|---|---|
| Sunday | 30.63388291542123 |
| Saturday | 28.91770012326631 |
| Friday | 23.26415876982887 |
| Monday | 22.9689842804972 |
| Tuesday | 20.82917069321898 |
| Wednesday | 20.73689757394715 |
| Thursday | 20.62229687211414 |

```
SELECT COUNT(*)
FROM tripdata.details
WHERE CAST(total_seconds as int) < 60
```

RESULT

73887

```
DELETE
FROM tripdata.details
WHERE CAST(total_seconds as int) < 60
```

RESULT

Query finished in 4.465 second(s). Rows affected: 73887

```
SELECT day_of_week, AVG(total_seconds)/60 AS avg_ride_minutes
FROM tripdata.details
GROUP BY (day_of_week)
ORDER BY (avg_ride_minutes) DESC
```

RESULT

| | |
|---|---|
| Sunday | 31.11970452827306 |
| Saturday | 29.36623376474381 |
| Friday | 23.62749380216491 |
| Monday | 23.33688225943991 |
| Tuesday | 21.15533760734856 |
| Wednesday | 21.05282258757791 |
| Thursday | 20.94503242143717 |

```
SELECT member_casual, AVG(total_seconds)/60 AS avg_ride_minutes
FROM tripdata.details
    GROUP BY member_casual
```

RESULT

| | |
|---|---|
| casual | 37.05418266317892 |
| member | 14.97988776422216 |

```
SELECT member_casual, started_at, ended_at, total_seconds, MAX(total_seconds)/86400 AS
max_ride_days
FROM tripdata.details
GROUP BY member_casual
```

| member_casual | started_at | ended_at | total_seconds | max_ride_days |
|---|---|---|---|---|
| casual | 2020-09-06 23:04 | 2020-10-07 22:08 | 2675079 | 30 |
| member | 2020-08-03 15:50:18 | 2020-08-26 20:51:40 | 2005282 | 23 |

```
SELECT member_casual, AVG(total_seconds)/60 AS avg_ride_minutes
FROM tripdata.details
WHERE started_at BETWEEN "2020-08-01" and "2021-10-31"
GROUP BY member_casual
```

RESULT

| | |
|---|---|
| casual | 37.05418266317892 |
| member | 14.97988776422216 |

```
SELECT member_casual, AVG(total_seconds)/60 AS avg_ride_minutes
FROM tripdata.details
WHERE started_at BETWEEN "2020-11-01" and "2021-03-31"
GROUP BY member_casual
```

RESULT

| | |
|---|---|
| casual | 34.24992936851313 |
| member | 13.96762237869367 |

```
SELECT member_casual, AVG(total_seconds)/60 AS avg_ride_minutes,
MAX(total_seconds)/60 AS max_ride_minutes, MIN(total_seconds)/60 as min_ride_minutes
FROM tripdata.details
WHERE started_at BETWEEN "2020-08-01" and "2020-09-30"
GROUP BY member_casual
```

RESULT

| | avg | max | min |
|---|---|---|---|
| casual | 42.45570601164682 | 44584 | 1 |
| member | 16.55668696656189 | 33421 | 1 |

```
SELECT member_casual, AVG(total_seconds)/60 AS avg_ride_minutes,
MAX(total_seconds)/60 AS max_ride_minutes, MIN(total_seconds)/60 as min_ride_minutes
FROM tripdata.details
WHERE started_at BETWEEN "2020-10-01" and "2020-11-30"
GROUP BY member_casual
```

RESULT

| | avg | max | min |
|---|---|---|---|
| casual | 31.39467270754935 | 35934 | 1 |
| member | 14.10887283119915 | 9557 | 1 |

```
SELECT member_casual, AVG(total_seconds)/60 AS avg_ride_minutes,
MAX(total_seconds)/60 AS max_ride_minutes, MIN(total_seconds)/60 as min_ride_minutes
FROM tripdata.details
WHERE started_at BETWEEN "2020-12-01" and "2021-01-31"
GROUP BY member_casual
```

RESULT

| | avg | max | min |
|---|---|---|---|
| casual | 26.69417429644319 | 19825 | 1 |
| member | 12.95888233023602 | 1499 | 1 |

```
SELECT member_casual, AVG(total_seconds)/60 AS avg_ride_minutes,
MAX(total_seconds)/60 AS max_ride_minutes, MIN(total_seconds)/60 as min_ride_minutes
FROM tripdata.details
WHERE started_at BETWEEN "2021-02-01" and "2021-03-31"
GROUP BY member_casual
```

RESULT

| | avg | max | min |
|---|---|---|---|
| casual | 40.05467867946767 | 31681 | 1 |
| member | 15.14043094031244 | 1559 | 1 |

```
SELECT member_casual, AVG(total_seconds)/60 AS avg_ride_minutes,
MAX(total_seconds)/60 AS max_ride_minutes, MIN(total_seconds)/60 as min_ride_minutes
FROM tripdata.details
WHERE started_at BETWEEN "2021-04-01" and "2021-05-31"
GROUP BY member_casual
```

RESULT

| | avg | max | min |
|---|---|---|---|
| casual | 38.28243161183273 | 38963 | 1 |
| member | 14.85336046194912 | 1499 | 1 |

```
SELECT member_casual, AVG(total_seconds)/60 AS avg_ride_minutes,
MAX(total_seconds)/60 AS max_ride_minutes, MIN(total_seconds)/60 as min_ride_minutes
FROM tripdata.details
WHERE started_at BETWEEN "2021-06-01" and "2021-07-31"
GROUP BY member_casual
```

RESULT

| | avg | max | min |
|---|---|---|---|
| casual | 35.01507722488874 | 41645 | 1 |
| member | 14.66511180901766 | 1499 | 1 |

Less than 1 hr

```
SELECT member_casual, COUNT(*)
FROM tripdata.details
WHERE total_seconds < 3600
GROUP BY member_casual
```

RESULT

| | |
|---|---|
| casual | 1831777 |
| member | 2555784 |

More than 1hr

```
SELECT member_casual, COUNT(*)
FROM tripdata.details
WHERE total_seconds > 3600
GROUP BY member_casual
```

RESULT

| | |
|---|---|
| casual | 237856 |
| member | 23531 |

More than 2hrs

```
SELECT member_casual, COUNT(*)
FROM tripdata.details
WHERE total_seconds > 7200
GROUP BY member_casual
```

RESULT

| | |
|---|---|
| casual | 63089 |
| member | 6080 |

More than 6 hours

```
SELECT member_casual, COUNT(*)
FROM tripdata.details
WHERE total_seconds > 21600
GROUP BY member_casual
```

RESULT

| | |
|---|---|
| casual | 9033 |

| | |
|---|---|
| member | 1571 |

More than 12hrs

```
SELECT member_casual, COUNT(*)
FROM tripdata.details
WHERE total_seconds > 43200
GROUP BY member_casual
```

RESULT

| | |
|---|---|
| casual | 5860 |
| member | 906 |

More than 1 day

```
SELECT member_casual, COUNT(*)
FROM tripdata.details
WHERE total_seconds > 86400
GROUP BY member_casual
```

RESULT

| | |
|---|---|
| casual | 3051 |
| member | 396 |

More than 2 days

```
SELECT member_casual, COUNT(*)
FROM tripdata.details
WHERE total_seconds > 172800
GROUP BY member_casual
```

RESULT

| | |
|---|---|
| casual | 818 |
| member | 22 |

More than 7 days

```
SELECT member_casual, COUNT(*)
FROM tripdata.details
WHERE total_seconds > 604800
GROUP BY member_casual
```

RESULT

| | |
|---|---|
| casual | 315 |
| member | 11 |

More than 14 days

```
SELECT member_casual, COUNT(*)
FROM tripdata.details
WHERE total_seconds > 1209600
GROUP BY member_casual
```

RESULT

| | |
|---|---|
| casual | 132 |
| member | 4 |

More than 30 days

```
SELECT member_casual, COUNT(*)
FROM tripdata.details
WHERE total_seconds > 2592000
GROUP BY member_casual
```

RESULT

| | |
|---|---|
| casual | 1 |
| member | 0 |

CALCULATE THE MODE

```
SELECT COUNT(*), total_seconds
FROM tripdata.details
GROUP BY total_seconds
ORDER BY COUNT(*) DESC
```

RESULT

| COUNT | Total_seconds |
|---|---|
| 4134 | 410 |
| 4115 | 364 |
| 4115 | 358 |
| 4098 | 385 |
| 4092 | 432 |
| 4084 | 443 |
| 4083 | 406 |
| 4066 | 407 |
| 4062 | 404 |
| 4057 | 356 |

SO THE MODE IS 410 SECONDS: 6 minutes 50 seconds FOR ALL

MODE FOR MEMBER

```
SELECT COUNT(*), total_seconds
FROM tripdata.details
WHERE member_casual = "member"
GROUP BY total_seconds
ORDER BY COUNT(*) DESC
```

RESULT

| COUNT | Total_seconds |
|---|---|
| 2995 | 338 |
| 2975 | 364 |
| 2964 | 358 |
| 2958 | 331 |
| 2925 | 319 |
| 2922 | 316 |
| 2914 | 347 |
| 2911 | 307 |
| 2904 | 322 |

SO THE MODE IS 338 SECONDS: 5 minutes 38 seconds FOR MEMBER

MODE FOR CASUAL

```
SELECT COUNT(*), total_seconds
FROM tripdata.details
WHERE member_casual = "casual"
GROUP BY total_seconds
ORDER BY COUNT(*) DESC
```

RESULT

| COUNT | Total_seconds |
|---|---|
| 1415 | 550 |
| 1406 | 555 |
| 1401 | 642 |
| 1399 | 496 |
| 1398 | 473 |
| 1397 | 563 |
| 1396 | 512 |
| 1395 | 552 |

SO THE MODE IS 550 SECONDS: 9 minutes 10 seconds FOR CASUAL

```
SELECT strftime('%m', started_at) as month, member_casual, (AVG(total_seconds)/60) as
avg_minutes
FROM tripdata.details
WHERE member_casual = "casual"
GROUP BY month
```

RESULT

| | | |
|---|---|---|
| 01 | casual | 26.07430968127639 |
| 02 | casual | 50.19325070252911 |
| 03 | casual | 39.27182329177746 |
| 04 | casual | 38.95807915804175 |
| 05 | casual | 39.17886036745406 |
| 06 | casual | 36.83054090616895 |
| 07 | casual | 33.45426701890985 |
| 08 | casual | 45.60053350915751 |
| 09 | casual | 38.49149794046244 |
| 10 | casual | 31.29236885759377 |
| 11 | casual | 32.24789884315308 |
| 12 | casual | 27.14590370245519 |
| AVG | casual | 33.8462198691177 |

```
SELECT strftime('%m', started_at) as month, member_casual, (AVG(total_seconds)/60) as
avg_minutes
FROM tripdata.details
WHERE member_casual = "member"
GROUP BY month
```

RESULT

| | | |
|---|---|---|
| AVG | member | 13.45267623710818 |
| 01 | member | 13.05595460619889 |
| 02 | member | 18.38942354917927 |
| 03 | member | 14.57718966161899 |
| 04 | member | 15.33038356112947 |
| 05 | member | 15.21823868957857 |
| 06 | member | 15.30141534075156 |
| 07 | member | 14.80258639918349 |
| 08 | member | 17.13155522741606 |
| 09 | member | 16.07171336873708 |
| 10 | member | 14.62313856424451 |
| 11 | member | 13.78526873357398 |
| 12 | member | 12.91166893714774 |

```
SELECT member_casual, COUNT(*)
FROM tripdata.details
--WHERE member_casual = "casual"
GROUP BY (member_casual)
```

RESULT

| | |
|---|---|
| casual | 2069722 |
| member | 2579332 |

```
SELECT member_casual, COUNT(*)
FROM tripdata.details
WHERE total_seconds > 600
GROUP BY (member_casual)
```

RESULT

| | |
|---|---|
| casual | 1578274 |
| member | 1372466 |

```
SELECT member_casual, COUNT(*)
FROM tripdata.details
WHERE total_seconds > 3600
GROUP BY (member_casual)
```

RESULT

| | |
|---|---|
| casual | 237856 |
| member | 23531 |

```
SELECT member_casual, COUNT(*)
FROM tripdata.details
WHERE total_seconds > 7200
GROUP BY (member_casual)
```

RESULT

| | |
|---|---|
| casual | 63089 |
| member | 6080 |

```
SELECT member_casual, COUNT(*)
FROM tripdata.details
WHERE total_seconds > 86400
GROUP BY (member_casual)
```

RESULT

| | |
|---|---|
| casual | 3051 |
| member | 396 |