Miguel_Diaz_ClassWork_Sesson_7

SQL locally with PostgresSQL

Activity 1

Creating tables:

```
3 ➤ CREATE TABLE gps_data (
       datapointid VARCHAR(40) PRIMARY KEY,
5
       journeyid VARCHAR(40),
6
      latitude NUMERIC(10, 7),
7
       longitude NUMERIC(10, 7),
8
      month INTEGER,
9 day INTEGER,10 hour INTEGER
11 );
12
13 v COPY gps_data (datapointid, journeyid, latitude, longitude, month, day, hour)
14 FROM 'C:/adi10136/Documents/CE-5900-16/gps_data.csv'
15 DELIMITER ','
16 CSV HEADER;
17
```

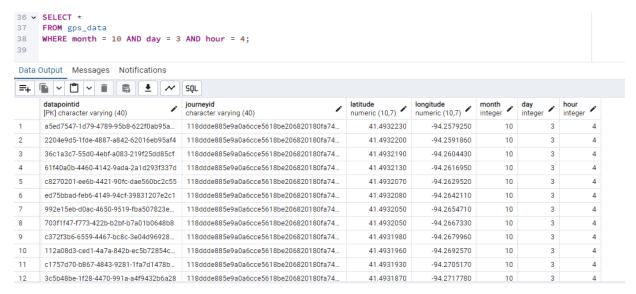
	datapointid [PK] character varying (40)	journeyid character varying (40)	latitude numeric (10,7)	longitude numeric (10,7)	month integer	day integer	hour integer /
1	000074d0-2968-49ff-aa61-8f0ff648a3bd	80b68e1a324d2c47f413faaef00cbf55033b6e73	41.5033700	-94.6438390	10	2	7
2	0000b449-4250-4b47-a4b9-f9c955db6cdd	9cc28b887c54bfd910e8f0c49cfe819122432d79	41.3631620	-94.6993880	10	1	6
3	000133f0-e2d7-446e-9036-8dcc1c11a74a	8cf32f521244f66a0795041ed458149ad245e996	41.4951030	-94.5879900	10	3	7
4	00017452-b680-4ef8-b84a-02d9b2d55e00	fdc51c3bf1171ad738d4d4180ed003b4372e02b9	41.3130080	-94.4531930	10	1	6
5	0002283c-6b03-4cce-9521-c2c0d5f9cc14	21dfe4b9d65fe92c3833eca292b14618d33ac44d	41.4797720	-94.6829570	10	2	7
6	00026fca-6f10-48bb-bfa8-b65d85843c17	3dadf6e8e74bfbecbcb6551c27ac47c2e4e8d79f	41.4930800	-94.6671800	10	1	7
7	00029e52-a2a3-42c0-aecc-34f8bcd8439d	9ab0a042f7a696bf326cb68c4cce050544e63127	41.3019040	-94.5076160	10	2	2
8	000362e8-6a1e-46c3-9c18-a88d8110d1	7447f426dba0f50f4491ac9b27d05b473c6bc64f	41.3019130	-94.3448860	10	3	6
9	00039125-10bf-4848-98bd-89835a030b90	d35bc46b290bce6d2f0dda65d3ac733d9b75b91c	41.4960780	-94.5104200	10	1	7
10	0003c83d-26ca-4f23-8de0-ccdbb5a441c1	aad413f86894c26134939e90bc528c3a4a49eb21	41.4931360	-94.2960720	10	2	7
11	00043376-1273-40eb-81d8-44e1766cec	933189643d69d8a76afdd785cbe256aab4597a3a	41.2907940	-94.5626500	10	2	1
12	00044a58-335d-41d0-aeff-796288aa4e9a	1a1cc638f1d62d7428ef0d8117e04d6e0417cc3a	41.1862740	-94.4717440	10	1	7
13	0004b934-396d-4127-a531-daea3f60cef4	b05102feef036405342947580880007d5d205c40	41.4968610	-94.5751210	10	1	7
14	0004dc3c-3a3d-48a0-b262-ec41f948c66b	bed5861d4e06cef32464704d743cd56886831703	41.4979560	-94.3176780	10	3	3
15	0005113f-da20-4887-b4b7-866a31d94297	47807fab870bd1775bf5218280a2d00922c16479	41.4929490	-94.2894780	10	1	7
16	00051863-9430-404c-931d-5e98e2d1c6	79f272e334eccc7adce90ab541340b6db518156f	41.4930050	-94.3587250	10	3	5
17	00054749-a3ad-4d9f-bf3d-493aab8ead23	167458bea3f34cfe5969db1e4d9dd41d68dd1d45	41.4929800	-94.2630920	10	2	3
18	0006dd14-a24b-4c16-9150-c86460c962	16ef58af500c1498c30ef52a0753cabb6c17d142	41.4968060	-94.5523840	10	2	2
19	0006f1de-e423-451a-aabd-4bc00981b219	52ec83558b66c61028a5ecb1f1628ce43d4b965f	41.4956930	-94.6429710	10	2	7
20	00072faf-d4b6-4284-a9fd-0710264d945a	7b6ec23ed88d5781fefb8632e523d87e8696d7ff	41.3742780	-94.5192590	10	1	7
21	0007c88f-4b42-41da-ba40-b6322aef406e	b7884bd1ad212cb20868c4512540d4d1476573	41.4930620	-94.2500950	10	3	6
22	000812f5-d195-4e13-b9a5-12e66ddea10d	a2f9b7e63ba014946f2cd8e2a426102080c4a193	41.4969530	-94.5624010	10	2	7
23	00089760-60fe-40c2-91db-e329595f0516	78a049d6d9c86667d971d946373549aa3a1625	41.4965330	-94.3191460	10	1	7
24	0008bad9-4429-4773-8b00-3effd6a5306a	f8e18d6a9d5e8a6e3ba3c5aaab7aea7588458658	41.2873810	-94.4796770	10	3	7

```
20 v CREATE TABLE vehicle_data (
         datapointid VARCHAR(40) PRIMARY KEY,
21
22
         geohash VARCHAR(12),
23
         speed NUMERIC(5, 2),
24
         make VARCHAR(50),
25
         model VARCHAR(50),
26
         route_id VARCHAR(20),
27
         segment_start_measure NUMERIC(10, 3)
28
    );
29
30 v COPY vehicle_data (datapointid, geohash, speed, make, model, route_id, segment_start_measure)
     FROM 'C:/adi10136/Documents/CE-5900-16/vehicle_data.csv'
32
     DELIMITER ','
33 CSV HEADER;
```

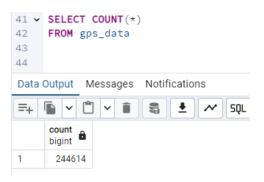
	datapointid [PK] character varying (40)	geohash character varying (12)	speed numeric (5,2)	make character varying (50)	model character varying (50)	route_id character varying (20)	segment_start_measure numeric (10,3)
1	000074d0-2968-49ff-aa61-8f0ff648a3bd	9zksn7	27.64	CHEVROLET	SILVERADO_HD	M002246110E	0.689
2	0000b449-4250-4b47-a4b9-f9c955db6cdd	9zkem1	0.00	GMC	SIERRA_HD	C000141000N	9.110
3	000133f0-e2d7-446e-9036-8dcc1c11a74a	9zksp9	122.11	GMC	SIERRA	S001910080E	78.961
4	00017452-b680-4ef8-b84a-02d9b2d55e00	9zkg48	61.05	GMC	ACADIA	S001930025N	48.283
5	0002283c-6b03-4cce-9521-c2c0d5f9cc14	9zkevr	97.91	CHEVROLET	TRAVERSE	C000141010N	1.254
6	00026fca-6f10-48bb-bfa8-b65d85843c17	9zksjc	124.41	GMC	SIERRA	S001910080E	74.798
7	00029e52-a2a3-42c0-aecc-34f8bcd8439d	9zkfcq	86.39	GMC	TERRAIN	S001930092E	77.287
8	000362e8-6a1e-46c3-9c18-a88d8110d1	9zkfvn	94.46	CHEVROLET	SILVERADO_HD	S001930092E	85.787
9	00039125-10bf-4848-98bd-89835a030b90	9zku16	129.02	CHEVROLET	TRAX_TRACKER_TRAILBLAZER_B	S001910080E	82.973
10	0003c83d-26ca-4f23-8de0-ccdbb5a441c1	9zkun1	118.65	CHEVROLET	SILVERADO_HD	S001910080W	211.620
11	00043376-1273-40eb-81d8-44e1766cec	9zkfbh	9.21	CHEVROLET	EQUINOX	M264741070N	0.203
12	00044a58-335d-41d0-aeff-796288aa4e9a	9zkf61	0.00	CHEVROLET	SILVERADO	C000146050E	10.115
13	0004b934-396d-4127-a531-daea3f60cef4	9zkspf	116.35	CHEVROLET	EQUINOX	S001910080W	226.135
14	0004dc3c-3a3d-48a0-b262-ec41f948c66b	9zkujd	2.30	CHEVROLET	TRAVERSE	M745746070E	0.435
15	0005113f-da20-4887-b4b7-866a31d94297	9zkun3	119.80	CHEVROLET	IMPALA	S001910080E	94.501
16	00051863-9430-404c-931d-5e98e2d1c6	9zkuhc	120.96	CHEVROLET	SILVERADO	S001910080E	90.930
17	00054749-a3ad-4d9f-bf3d-493aab8ead23	9zkunc	116.35	CHEVROLET	SILVERADO	S001910080E	95.896
18	0006dd14-a24b-4c16-9150-c86460c962	9zku06	118.65	GMC	TERRAIN	S001910080W	225.000

Activity 2

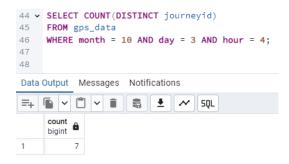
Select datapoints from October 3 at 4 AM



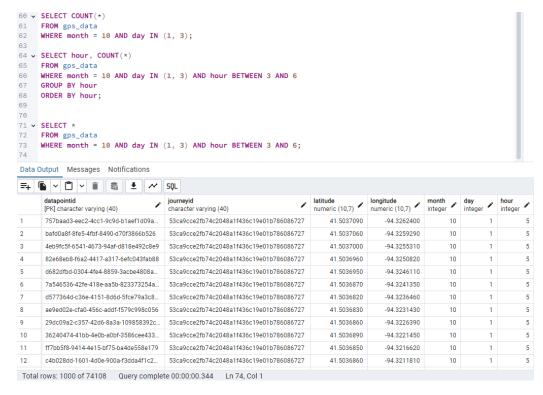
2. Count how many rows in gps_data table (Expected Answer: 244614)



3. Find how many unique journey_ids in the data filtered as per (1) (Ans: 7)

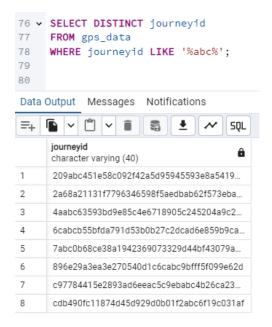


4. Filter data on October 1st and 3rd, and hour between 3 AM and 6 AM (Expected Answer: 101116 rows)



```
60 V SELECT COUNT(*)
FROM gps_data
WHERE month = 10 AND day IN (1, 3);
64 • SELECT hour, COUNT(*)
65 FROM gps data
    WHERE month = 10 AND day IN (1, 3) AND hour BETWEEN 3 AND 6
    GROUP BY hour
68 ORDER BY hour;
   -- SELECT *
72
   -- FROM gps_data
    -- WHERE month = 10 AND day IN (1, 3) AND hour BETWEEN 3 AND 6;
Data Output Messages Notifications
=+ 6 ∨ 1 ∨ 1 3 4 7 SQL
               5715
2
              10578
3
               13249
               44566
```

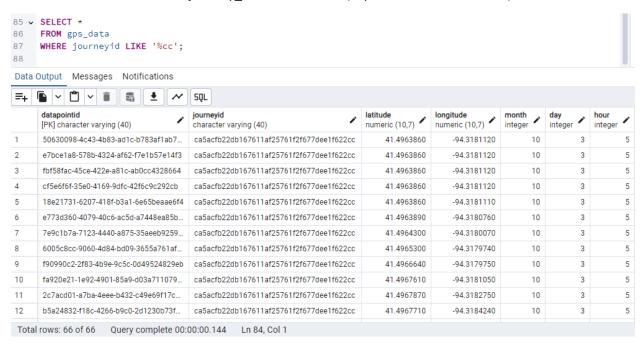
5. Select all journey_ids containing "abc" from the whole gps_data table (Expected Answer: 8)



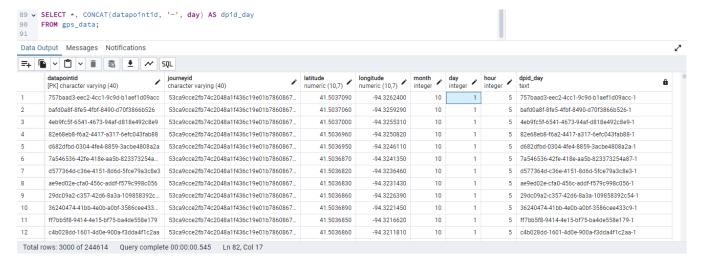
6. Count how many datapoint_ids start with "9" (Expected Answer: 15256)



7. Select all rows where journey_id ends with "cc" (Expected Answer: 66 rows)

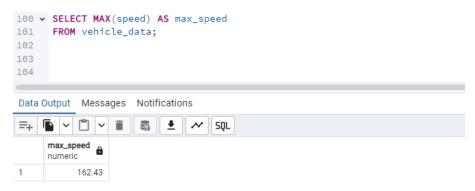


Select all columns and create a new column that combines datapoint_id and day with a "-" symbol



Activity - 3: Aggregations

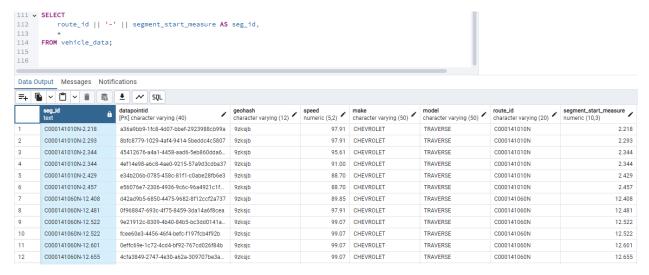
1. Find maximum speed from table vehicle_data (Ans: 162.43)



2. Find average, minimum, maximum and standard deviations of speeds for "CHEVROLET" make



3. Create seg_id by combining "route_id" and "segment_start_measure"



4. Add this seg_id to existing vehicle_data table and verify it

• Alter_table and update

```
111 v SELECT route_id || '-' || segment_start_measure AS seg_id,

113 *

FROM vehicle_data;

116 v ALTER TABLE vehicle_data

117 ADD COLUMN seg_id VARCHAR(60);

118

119 v UPDATE vehicle_data

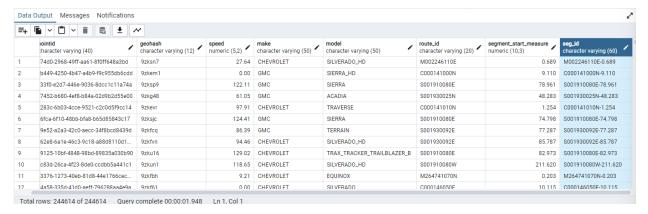
120 SET seg_id = route_id || '-' || segment_start_measure;

121

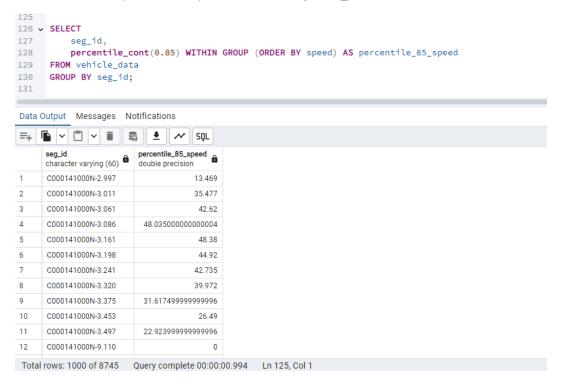
122 v SELECT seg_id, *

FROM vehicle_data

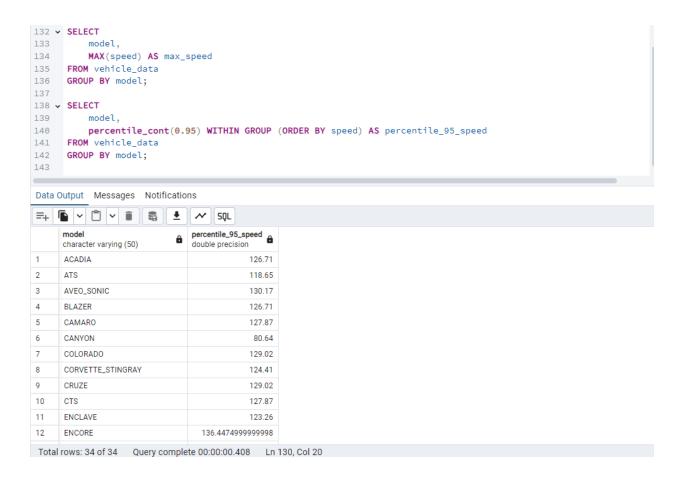
LIMIT 10;
```



5. Find 85th percentiles speeds for each segment_id

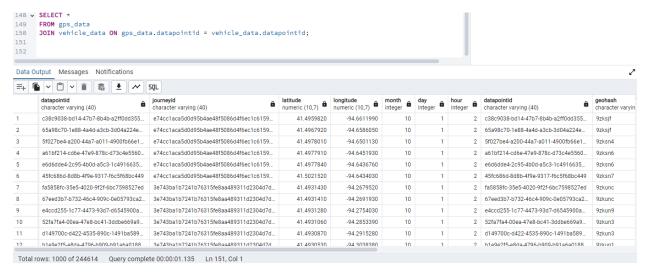


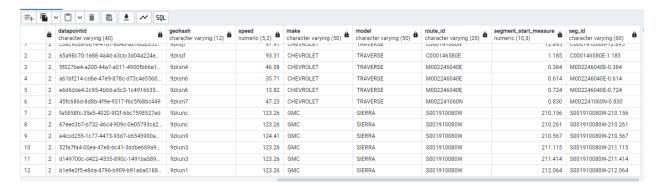
6. Find the highest speeds for each model? (Compare 95 th percentile speeds for each model)



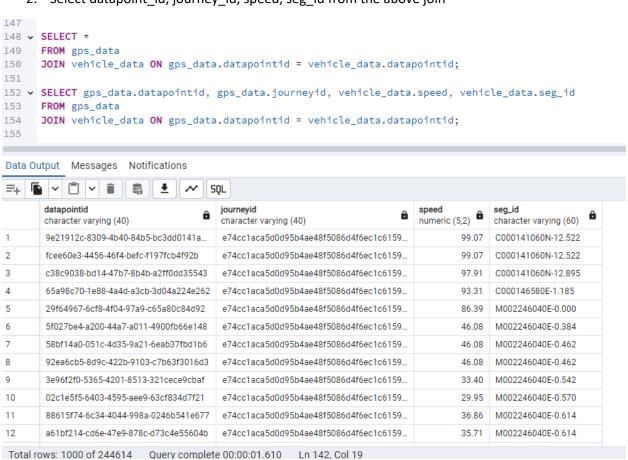
ACTIVITY - 4: Joins

1. Join the "gps_data" and "vehicle_data" tables using "JOIN" on datapoint_id





2. Select datapoint_id, journey_id, speed, seg_id from the above join



3. Count the unique journey_ids that came to stop at least once (Speed = 0)



ANEXUS (IMPLEMENTED CODE):
ACTIVITY 1 - CREATE TABLES
DROP TABLE IF EXISTS gps_data;
CREATE TABLE gps_data (
datapointid VARCHAR(40) PRIMARY KEY,
- journeyid VARCHAR(40),
- latitude NUMERIC(10, 7),
- longitude NUMERIC(10, 7),
- month INTEGER,
- day INTEGER,
hour INTEGER
-);
COPY gps_data (datapointid, journeyid, latitude, longitude, month, day, hour)
FROM 'C:/adi10136/Documents/CE-5900-16/gps_data.csv'
- DELIMITER ','
CSV HEADER;
DROP TABLE IF EXISTS vehicle_data;
CREATE TABLE vehicle_data (
datapointid VARCHAR(40) PRIMARY KEY,
geohash VARCHAR(12),
speed NUMERIC(5, 2),
make VARCHAR(50),
model VARCHAR(50),
- route_id VARCHAR(20),
segment_start_measure NUMERIC(10, 3)
-);
COPY vehicle_data (datapointid, geohash, speed, make, model, route_id, segment_start_measure)
FROM 'C:/adi10136/Documents/CE-5900-16/vehicle_data.csv'
- DELIMITER ','
- CSV HEADER;

SELECT *
FROM gps_data
WHERE month = 10 AND day = 3 AND hour = 4;
SELECT COUNT(*)
FROM gps_data
SELECT COUNT(DISTINCT journeyid)
FROM gps_data
WHERE month = 10 AND day = 3 AND hour = 4;
Query to select the rows
- SELECT *
- FROM gps_data
WHERE month = 10 AND day IN (1, 3) AND hour BETWEEN 3 AND 6;
- WHERE HOURT - 10 AND DAY IN (1, 5) AND HOUR BETWEEN 3 AND 0,
Query to count the rows that match the condition
SELECT COUNT(*)
FROM gps_data
WHERE month = 10 AND day IN (1, 3) AND hour BETWEEN 3 AND 6;
SELECT COUNT(*)
- SELECT COUNT(*) - FROM gps_data
FROM gps_data
FROM gps_data
FROM gps_data WHERE month = 10 AND day IN (1, 3);
- FROM gps_data - WHERE month = 10 AND day IN (1, 3); - SELECT hour, COUNT(*)
FROM gps_data WHERE month = 10 AND day IN (1, 3); SELECT hour, COUNT(*) FROM gps_data
- FROM gps_data - WHERE month = 10 AND day IN (1, 3); - SELECT hour, COUNT(*) - FROM gps_data - WHERE month = 10 AND day IN (1, 3) AND hour BETWEEN 3 AND 6
FROM gps_data WHERE month = 10 AND day IN (1, 3); SELECT hour, COUNT(*) FROM gps_data WHERE month = 10 AND day IN (1, 3) AND hour BETWEEN 3 AND 6 GROUP BY hour
FROM gps_data WHERE month = 10 AND day IN (1, 3); SELECT hour, COUNT(*) FROM gps_data WHERE month = 10 AND day IN (1, 3) AND hour BETWEEN 3 AND 6 GROUP BY hour
FROM gps_data WHERE month = 10 AND day IN (1, 3); SELECT hour, COUNT(*) FROM gps_data WHERE month = 10 AND day IN (1, 3) AND hour BETWEEN 3 AND 6 GROUP BY hour
- FROM gps_data - WHERE month = 10 AND day IN (1, 3); - SELECT hour, COUNT(*) - FROM gps_data - WHERE month = 10 AND day IN (1, 3) AND hour BETWEEN 3 AND 6 - GROUP BY hour;
FROM gps_data WHERE month = 10 AND day IN (1, 3); SELECT hour, COUNT(*) FROM gps_data WHERE month = 10 AND day IN (1, 3) AND hour BETWEEN 3 AND 6 GROUP BY hour ORDER BY hour;
- FROM gps_data - WHERE month = 10 AND day IN (1, 3); - SELECT hour, COUNT(*) - FROM gps_data - WHERE month = 10 AND day IN (1, 3) AND hour BETWEEN 3 AND 6 - GROUP BY hour - ORDER BY hour;
- FROM gps_data - WHERE month = 10 AND day IN (1, 3); - SELECT hour, COUNT(*) - FROM gps_data - WHERE month = 10 AND day IN (1, 3) AND hour BETWEEN 3 AND 6 - GROUP BY hour - ORDER BY hour;
FROM gps_data WHERE month = 10 AND day IN (1, 3); SELECT hour, COUNT(*) FROM gps_data WHERE month = 10 AND day IN (1, 3) AND hour BETWEEN 3 AND 6 GROUP BY hour; ORDER BY hour; SELECT * FROM gps_data WHERE month = 10 AND day IN (1, 3) AND hour BETWEEN 3 AND 6;
- FROM gps_data - WHERE month = 10 AND day IN (1, 3); - SELECT hour, COUNT(*) - FROM gps_data - WHERE month = 10 AND day IN (1, 3) AND hour BETWEEN 3 AND 6 - GROUP BY hour - ORDER BY hour; - SELECT * - FROM gps_data - WHERE month = 10 AND day IN (1, 3) AND hour BETWEEN 3 AND 6;
FROM gps_data WHERE month = 10 AND day IN (1, 3); SELECT hour, COUNT(*) FROM gps_data WHERE month = 10 AND day IN (1, 3) AND hour BETWEEN 3 AND 6 GROUP BY hour; ORDER BY hour; SELECT * FROM gps_data WHERE month = 10 AND day IN (1, 3) AND hour BETWEEN 3 AND 6;

SELECT COUNT(*)
FROM gps_data
WHERE datapointid LIKE '9%';
- SELECT *
FROM gps_data
WHERE journeyid LIKE '%cc';
SELECT *, CONCAT(datapointid, '-', day) AS dpid_day
FROM gps_data;
ACTIVITY - 3 - Aggregations
- SELECT MAX(speed) AS max_speed
FROM vehicle_data;
SELECT
AVG(speed) AS avg_speed,
MIN(speed) AS min_speed,
MAX(speed) AS max_speed,
STDDEV(speed) AS stddev_speed
FROM vehicle_data
WHERE make = 'CHEVROLET';
SELECT
- route_id '-' segment_start_measure AS seg_id,
- *
FROM vehicle_data;
······································
ALTER TABLE vehicle_data
- ADD COLUMN seg_id VARCHAR(60);
UPDATE vehicle_data
SET seg_id = route_id '-' segment_start_measure;
- SELECT seg_id, *
FROM vehicle_data
LIMIT 10;
- SELECT
seg_id,
- percentile_cont(0.85) WITHIN GROUP (ORDER BY speed) AS percentile_85_speed
FROM vehicle_data

GROUP BY seg_id;
GROUP BT SER_IU;
SELECT
model,
MAX(speed) AS max_speed
FROM vehicle_data
GROUP BY model;
SELECT
model,
percentile_cont(0.95) WITHIN GROUP (ORDER BY speed) AS percentile_95_speed
FROM vehicle_data
GROUP BY model;
ACTIVITY - 4 : Joins
SELECT *
FROM gps_data
JOIN vehicle_data ON gps_data.datapointid = vehicle_data.datapointid;
SELECT gps_data.datapointid, gps_data.journeyid, vehicle_data.speed, vehicle_data.seg_id
FROM gps_data
JOIN vehicle_data ON gps_data.datapointid = vehicle_data.datapointid;
SELECT COUNT(DISTINCT gps_data.journeyid) AS stopped_journey_count
FROM gps_data
JOIN vehicle_data ON gps_data.datapointid = vehicle_data.datapointid
WHERE vehicle_data.speed = 0;