Question 1

Create a materialized view to compute the average, min and max trip time between each taxi zone.

Note that we consider the do not consider `a->b` and `b->a` as the same trip pair. So as an example, you would consider the following trip pairs as different pairs:

```
Yorkville East -> Steinway
Steinway -> Yorkville East
```

From this MV, find the pair of taxi zones with the highest average trip time. You may need to use the dynamic filter pattern for this.

Bonus (no marks): Create an MV which can identify anomalies in the data. For example, if the average trip time between two zones is 1 minute, but the max trip time is 10 minutes and 20 minutes respectively.

Options:

1. **Yorkville East, Steinway**
2. Murray Hill, Midwood
3. East Flatbush/Farragut, East Harlem North
4. Midtown Center, University Heights/Morris Heights

p.s. The trip time between taxi zones does not take symmetricity into account, i.e. `A -> B` and `B -> A` are considered different trips. This applies to subsequent questions as well.

Question 2

Recreate the MV(s) in question 1, to also find the number of trips for the pair of taxi zones with the highest average trip time.

Options:

1. **5**
2. 3
3. 10

4. 1

Answer:
Yorkville East, Steinway is the only one options matched in my listed result, with 7 trips count which is closed to 5 in options provided.

```
CREATE MATERIALIZED VIEW trip_time_view AS
SELECT
        tpep_pickup_datetime as pickup_time,
        tpep_dropoff_datetime as dropoff_time,
   concat(taxi_zone_pu.Zone, ' to ', taxi_zone_do.Zone) as trip_route,
        tpep_dropoff_datetime- tpep_pickup_datetime as trip_time
FROM
   trip_data
   JOIN taxi_zone as taxi_zone_pu
      ON trip_data.PULocationID = taxi_zone_pu.location_id
   JOIN taxi_zone as taxi_zone_do
      ON trip_data.DOLocationID = taxi_zone_do.location_id;
```

```
select trip_route, count(trip_route), avg(trip_time) as trip_time_avg from trip_time_view group by trip_route order by trip_time_avg desc limit 50;
```

```
wongs=# select trip_route, count(trip_route), avg(trip_time) as trip_time_avg from trip_time_view group by trip_route order by trip_time_avg des
c limit 50;
                          trip_route                           | count |  trip_time_avg
---------------------------------------------------------------+-------+-----------------
 Saint Michaels Cemetery/Woodside to Long Island City/Hunters Point |     1 | 1 day 23:50:38
 Stuy Town/Peter Cooper Village to Murray Hill-Queens          |     1 | 23:58:44
 Washington Heights North to Highbridge Park                   |     1 | 23:58:40
 Clinton Hill to JFK Airport                                   |     1 | 23:49:41
 Upper East Side South to Soundview/Castle Hill                |     1 | 23:45:28
 Times Sq/Theatre District to Erasmus                          |     1 | 23:38:46
 Clinton Hill to South Williamsburg                            |     2 | 12:02:23
 Baisley Park to Chinatown                                     |     2 | 12:01:06.5
 Queensbridge/Ravenswood to Astoria                            |    17 | 09:49:23.705882
 Upper East Side South to Kew Gardens Hills                    |     3 | 08:22:39
 Long Island City/Queens Plaza to Bedford                      |     3 | 08:16:05
 UN/Turtle Bay South to Stuyvesant Heights                     |     3 | 08:15:48.666667
 Midtown Center to Highbridge                                  |     3 | 08:13:23
 Astoria to Midtown Center                                     |     3 | 08:07:52.333333
 NV to Dyker Heights                                           |     2 | 07:46:41.5
 Upper West Side South to Crown Heights North                  |     4 | 06:36:12.25
 Clinton East to Prospect-Lefferts Gardens                     |     4 | 06:33:01.75
 West Village to Borough Park                                  |     4 | 06:18:58.75
 World Trade Center to Crown Heights South                     |     4 | 06:17:21.5
 SoHo to Red Hook                                              |     4 | 06:12:29.25
 Union Sq to Woodside                                          |     5 | 05:06:31.2
 Greenwich Village South to Elmhurst                           |     5 | 05:05:01.4
 Manhattan Valley to Chinatown                                 |     5 | 05:03:17.6
 Two Bridges/Seward Park to Bushwick South                     |     5 | 05:02:20.4
 Marine Park/Floyd Bennett Field to NV                         |     1 | 04:31:17
 Two Bridges/Seward Park to Upper West Side North              |     6 | 04:20:45.333333
 Rosedale to Arden Heights                                     |     1 | 04:19:13
 Upper East Side South to NA                                   |     7 | 03:57:42.571429
 Clinton West to Crown Heights North                           |     7 | 03:57:18
 West Village to Flatbush/Ditmas Park                          |    13 | 03:55:44.923077
 Chinatown to Lenox Hill West                                  |    13 | 03:54:55.615385
 West Village to Elmhurst/Maspeth                              |     2 | 03:41:56
 Yorkville East to Steinway                                    |     7 | 03:41:42.428571
```

Question 3

From the latest pickup time to 17 hours before, what are the top 3 busiest zones in terms of number of pickups? For example if the latest pickup time is 2020-01-01 17:00:00, then the query should return the top 3 busiest zones from 2020-01-01 00:00:00 to 2020-01-01 17:00:00.

HINT: You can use dynamic filter pattern to create a filter condition based on the latest pickup time.

NOTE: For this question `17 hours` was picked to ensure we have enough data to work with.

Options:

1. Clinton East, Upper East Side North, Penn Station
2. LaGuardia Airport, Lincoln Square East, JFK Airport
3. **Midtown Center, Upper East Side South, Upper East Side North**
4. LaGuardia Airport, Midtown Center, Upper East Side North

Answer:

```
CREATE MATERIALIZED VIEW busiest_zones_1_min AS SELECT
   taxi_zone.Zone AS pickup_zone,
   count(*) AS lastest_17_hr_pickup_cnt
FROM
   trip_data
      JOIN taxi_zone
         ON trip_data.PULocationID = taxi_zone.location_id
WHERE
   trip_data.tpep_pickup_datetime > (trip_data.tpep_pickup_datetime  - INTERVAL '17' HOUR)
GROUP BY
   taxi_zone.Zone
ORDER BY lastest_17_hr_pickup_cnt desc
   LIMIT 10;
```

pickup_zone	lastest_17_hr_pickup_cnt
JFK Airport	50417
Upper East Side North	49984
Upper East Side South	49846
Midtown Center	36969
Penn Station/Madison Sq West	33042
Lincoln Square East	32185
Clinton East	31163
Lenox Hill West	30938
Upper West Side South	30793
Murray Hill	30610

(10 rows)