Question 1: What is count for fhv vehicles data for year 2019?

```
select count(*) from `datasciene-338718.trips_data_all.fhv_data_non_partitioned`;
```

Row	f0_
1	42084899

Question 2: How many distinct dispatching_base_num we have in fhv for 2019?

```
select count(distinct(dispatching_base_num)) from `datasciene-338718.trips_data_all.fhv_data_non_partitioned`
```

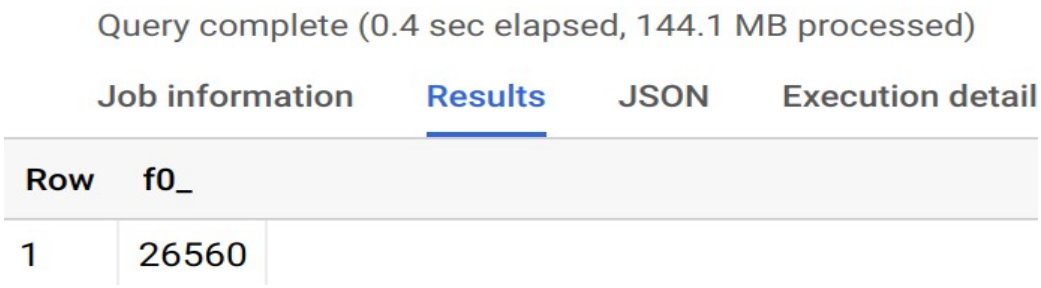
Row	f0_
1	792

Question 3: Best strategy to optimise if query always filter by dropoff_datetime and order by dispatching_base_num?

Ans: Partition by dropoff_datetime and cluster by dispatching_base_num

Question 4: What is the count, estimated and actual data processed for query which counts trip between 2019/01/01 and 2019/03/31 for dispatching_base_num B00987, B02060, B02279?

```
select count(*) from `datasciene-338718.trips_data_all.fhv_data_partitioned_clustered` where pickup_datetime between '2019-01-01' and '2019-03-31' and dispatching_base_num in ('B02279','B02060','B00987');
```



Question 5: What will be the best partitioning or clustering strategy when filtering on dispatching_base_num and SR_Flag?

Ans: Partition by dispatching_base_num and cluster by SR_Flag

Question 6: What improvements can be seen by partitioning and clustering for data size less than 1 GB?

Ans: Can be worse due to metadata

Question 7: In which format does BigQuery save data?

Ans: Columnar