Problem 3: Ingest Data

```
INSERT INTO customer (SELECT * FROM
snowflake_sample_data.tpch_sf10.customer);
```

Run Time: 7.13s

```
INSERT INTO lineitem (SELECT * FROM
snowflake_sample_data.tpch_sf10.lineitem);
```

Run Time: 37.26s

```
INSERT INTO nation (SELECT * FROM
snowflake_sample_data.tpch_sf10.nation);
```

Run Time: 1.22s

```
INSERT INTO orders (SELECT * FROM
snowflake_sample_data.tpch_sf10.orders);
```

Run Time: 12.71s

```
INSERT INTO part (SELECT * FROM snowflake_sample_data.tpch_sf10.part);
```

Run Time: 4.59s

```
INSERT INTO partsupp (SELECT * FROM
snowflake_sample_data.tpch_sf10.partsupp);
```

Run Time: 9.29s

```
INSERT INTO region (SELECT * FROM
snowflake_sample_data.tpch_sf10.region);
```

Run Time: 0.55s

```
INSERT INTO supplier (SELECT * FROM
snowflake_sample_data.tpch_sf10.supplier);
```

Run Time: 1.75s

Problem 4: Run Queries (All queries are run 5 times)

1. What is the total number of parts offered by each supplier?

Run Times:

78 ms, 69ms, 64ms, 64ms, 50ms,

Max Run Time: 78ms

Min RUn Time: 50ms

Avg Run Time: 65ms

# of Rows: 100,000

First two rows:

Row	supplier	count
1	Supplier#000030027	80
2	Supplier#000080030	80

2. What is the cost of the most expensive part by any supplier?

Run Times:

275ms, 90ms, 46ms, 51ms, 57ms,

Max Run Time: 275ms

Min RUn Time: 46ms

Avg Run Time: 103.8ms

# of Rows: 1

First two rows:

Row	SUPPLIER_MAX_COST.MAX_COST
1	1000.00

3. What is the cost of the most expensive part for each supplier?

Run Times:

50ms, 52ms, 49ms, 49ms, 60ms,

Max Run Time: 60ms

Min RUn Time: 49ms

Avg Run Time: 52ms

# of Rows: 100,000

First two rows:

Row	SUPPLIER	cost
1	Supplier#000036437	986.21
2	Supplier#000061409	978.94

4. What is the total number of customers per nation?

Run Times:

83ms, 60ms, 92ms, 78ms, 49ms,

Max Run Time: 92ms

Min RUn Time: 49ms

Avg Run Time: 74.4ms

# of Rows: 25

First two rows:

Row	nation	customer_count
1	ALGERIA	59916
2	KENYA	59476

5. What is number of parts shipped between 10 October, 1996 and 10 November, 1996 for each supplier?

Run Times:

73ms, 56ms, 73ms, 49ms, 51ms,
Max Run Time: 73ms
Min RUn Time: 49ms
Avg Run Time: 60.4ms

# of Rows: 99,964

First two rows:

Row	supplier	count
1	Supplier#000003908	4
2	Supplier#000062428	12

Brief Discussion:

First of all, I observed that queries 1, 3, 4, 5 have similary run time. I think the reason for this is that all these queries are performing very similar logic, which is joining two tables on a specific key.

Query 2 on the other hand, only deals with one table, Partsupp, but the way I structed the query makes it have a longer run time. This is because I first get a temporary table for the most expensive part of each supplier, and then find the max cost of this temporary table. I chose to structure the query this way because the logic here can be reused in query 2.

Another thing I observed that is not shown in the information required by the assignment is that. After the first query run, which typically have a significant longer run time, the second query run often also have a longer run time compared to the average run time listed above. This probably has something to do with how Snowflake handles repeated queries.