11-1
SELECT * FROM
peak-theorem-466601-f1.mudulabs_project.data
LIMIT 10

작업 정보 결과 차트 JSON 실행 세부정보 실행 그래프

행	InvoiceNo	StockCode	Quantity	InvoiceDate	UnitPrice	CustomerID	Country	Description
1	574469	22385	12	2011-11-04 11:55:00 UTC	0.0	12431	Australia	JUMBO BAG SPACEBOY DESIGN
2	554037	22619	80	2011-05-20 14:13:00 UTC	0.0	12415	Australia	SET OF 6 SOLDIER SKITTLES
3	574138	23234	216	2011-11-03 11:26:00 UTC	0.0	12415	Australia	BISCUIT TIN VINTAGE CHRISTM...
4	539722	22423	10	2010-12-21 13:45:00 UTC	0.0	14911	EIRE	REGENCY CAKESTAND 3 TIER
5	562973	23157	240	2011-08-11 11:42:00 UTC	0.0	14911	EIRE	SET OF 6 NATIVITY MAGNETS
6	574252	M	1	2011-11-03 13:24:00 UTC	0.0	12437	France	MANUAL
7	577168	M	1	2011-11-18 10:42:00 UTC	0.0	12603	Germany	MANUAL
8	537197	22841	1	2010-12-05 14:02:00 UTC	0.0	12647	Germany	ROUND CAKE TIN VINTAGE GRE...
9	564651	23270	96	2011-08-26 14:19:00 UTC	0.0	14646	Netherlands	SET OF 2 CERAMIC PAINTED H...
10	564651	22955	144	2011-08-26 14:19:00 UTC	0.0	14646	Netherlands	36 FOIL STAR CAKE CASES

11.3 데이터 전처리

```sql
SELECT
    COUNT(InvoiceNo) AS COUNT_InvoiceNo,
    COUNT(StockCode) AS COUNT_StockCode,
    COUNT(Description) AS COUNT_Description,
    COUNT(Quantity) AS COUNT_Quantity,
    COUNT(InvoiceDate) AS COUNT_InvoiceDate,
    COUNT(UnitPrice) AS COUNT_UnitPrice,
    COUNT(CustomerID) AS COUNT_CustomerID,
    COUNT(Country) AS COUNT_Country,

    -- COUNTIF(InvoiceNo IS NULL) AS COUNT_InvoiceNo_NULL,
    -- COUNTIF(StockCode IS NULL) AS COUNT_StockCode_NULL,
    -- COUNTIF(Description IS NULL) AS COUNT_Description_NULL,
    -- COUNTIF(Quantity IS NULL) AS COUNT_Quantity_NULL,
    -- COUNTIF(InvoiceDate IS NULL) AS COUNT_InvoiceDate_NULL,
    -- COUNTIF(UnitPrice IS NULL) AS COUNT_UnitPrice_NULL,
    -- COUNTIF(CustomerID IS NULL) AS COUNT_CustomerID_NULL,
    -- COUNTIF(Country IS NULL) AS COUNT_Country_NULL,
    -- COUNT(*) AS Total_rows
FROM `peak-theorem-466601-f1.mudulabs_project.data`
WHERE
    InvoiceNo LIKE '%C%'
LIMIT 100;
```

작업 정보 결과 차트 JSON 실행 세부정보 실행 그래프

행	COUNT_InvoiceNo	COUNT_StockCode	COUNT_Descripti...	COUNT_Quantity	COUNT_InvoiceD...	COUNT_UnitPrice	COUNT_Custome...	COUNT_Country
1	0	0	0	0	0	0	0	0

11-4 데이터 전처리(1)

```sql
SELECT
    'InvoiceNo' AS InvoiceNo_count,
    ROUND(SUM(CASE WHEN InvoiceNo IS NULL THEN 1 ELSE 0 END) / COUNT(*) * 100, 2)
AS INVO_ms_value

FROM `peak-theorem-466601-f1.mudulabs_project.data`

UNION ALL

SELECT
    'StockCode' AS StockCode_count,
    ROUND(SUM(CASE WHEN StockCode IS NULL THEN 1 ELSE 0 END) / COUNT(*) * 100, 2)
AS Sto_ms_value

FROM `peak-theorem-466601-f1.mudulabs_project.data`

UNION ALL

SELECT
    'UnitPrice' AS UnitPrice_count,
    ROUND(SUM(CASE WHEN UnitPrice IS NULL THEN 1 ELSE 0 END) / COUNT(*) * 100, 2)
AS Uni_ms_value

FROM `peak-theorem-466601-f1.mudulabs_project.data`
UNION ALL

SELECT
    'Description' AS Description_count,
    ROUND(SUM(CASE WHEN Description IS NULL THEN 1 ELSE 0 END) / COUNT(*) * 100, 2)
AS Des_ms_value

FROM `peak-theorem-466601-f1.mudulabs_project.data`
UNION ALL

SELECT
    'Quantity' AS Quantity_count,
    ROUND(SUM(CASE WHEN Quantity IS NULL THEN 1 ELSE 0 END) / COUNT(*) * 100, 2) AS
Qua_ms_value

FROM `peak-theorem-466601-f1.mudulabs_project.data`

UNION ALL

SELECT
    'InvoiceDate' AS InvoiceDate_count,
    ROUND(SUM(CASE WHEN InvoiceDate IS NULL THEN 1 ELSE 0 END) / COUNT(*) * 100, 2)
AS IND_ms_value

FROM `peak-theorem-466601-f1.mudulabs_project.data`
```

```sql
UNION ALL

SELECT
    'CustomerID' AS CustomerID_count,
    ROUND(SUM(CASE WHEN CustomerID IS NULL THEN 1 ELSE 0 END) / COUNT(*) * 100, 2)
AS Cus_ms_value

FROM `peak-theorem-466601-f1.mudulabs_project.data`

UNION ALL

SELECT
    'Country' AS Country_count,
    ROUND(SUM(CASE WHEN Country IS NULL THEN 1 ELSE 0 END) / COUNT(*) * 100, 2) AS
Con_ms_value

FROM `peak-theorem-466601-f1.mudulabs_project.data`

ORDER BY INVO_ms_value DESC;
```

행	InvoiceNo_count	INVO_ms_value
1	Country	0.0
2	InvoiceNo	0.0
3	StockCode	0.0
4	Description	0.0
5	InvoiceDate	0.0
6	Quantity	0.0
7	UnitPrice	0.0
8	CustomerID	0.0

```sql
SELECT
    COUNT(Description),

FROM `peak-theorem-466601-f1.mudulabs_project.data`
WHERE
    Description LIKE '%?%'
    AND Description LIKE '%w___%'
    AND Description LIKE '%EA%'
    AND StockCode = '85123A'
ORDER BY
    MAX(CASE WHEN StockCode = '85123A' THEN 1 ELSE 0 END) DESC;
```

행	f0_
1	0

```sql
SELECT
    StockCode,
    COUNT(DISTINCT Description) AS distinct_description_count,
    ARRAY_AGG(DISTINCT Description ORDER BY Description LIMIT 5) AS
sample_descriptions
FROM
    `peak-theorem-466601-f1.mudulabs_project.data`
WHERE
    Description IS NOT NULL

GROUP BY
    StockCode
HAVING
    COUNT(DISTINCT Description) > 1
ORDER BY
    distinct_description_count DESC;
```

결측치

```sql
DELETE FROM peak-theorem-466601-f1.mudulabs_project.data
WHERE Description IS NULL AND CustomerID IS NULL;
```

```sql
SELECT
    CREATE OR REPLACE TABLE `peak-theorem-466601-f1.mudulabs_project.data` (
    COUNT(DISTINCT *)) ;

FROM
    `peak-theorem-466601-f1.mudulabs_project.data`

CREATE OR REPLACE TABLE `peak-theorem-466601-f1.mudulabs_project.data` AS
SELECT DISTINCT *
FROM `peak-theorem-466601-f1.mudulabs_project.data`;
```

```sql
SELECT
    COUNT(DISTINCT InvoiceNo)
FROM `peak-theorem-466601-f1.mudulabs_project.data`;

WHERE
    InvoiceNo LIKE 'C%'

ORDER BY
    InvoiceNo DESC
LIMIT 100;
```

```sql
SELECT *
    -- DISTINCT InvoiceNo

FROM
   `peak-theorem-466601-f1.mudulabs_project.data`

WHERE
    InvoiceNo LIKE 'C%'

ORDER BY
    InvoiceNo DESC
LIMIT 100;
```

11-5 전처리 2

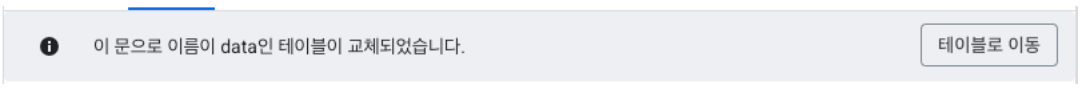
```sql
SELECT InvoiceNo, StockCode, Description, Quantity, InvoiceDate, UnitPrice,
CustomerID, Country,
    COUNT(*) AS count_rec

FROM
    `peak-theorem-466601-f1.mudulabs_project.data`

GROUP BY
    InvoiceNo, StockCode, Description, Quantity, InvoiceDate, UnitPrice,
CustomerID, Country
HAVING
    COUNT(*) > 1;
```

50 ▾ 1 – 50 (전체 4879행)

```sql
CREATE OR REPLACE TABLE `peak-theorem-466601-f1.mudulabs_project.data` AS
SELECT DISTINCT *
FROM `peak-theorem-466601-f1.mudulabs_project.data`;
```



11-6 데이터 전처리 3

```sql
SELECT
    DISTINCT InvoiceNo LIKE ‘c%’
```

```
FROM
    `peak-theorem-466601-f1.mudulabs_project.data`

ORDER BY
    InvoiceNo DESC
-- LIMIT 100;
```

행	InvoiceNo
1	536544
2	536555
3	536558
4	536565
5	536592
6	536596
7	536640
8	536755
9	536756

페이지당 결과 수: 50 1 – 50 (전체 100행)

```
SELECT
    StockCode, COUNT(*) AS sell_cnt

FROM
   `peak-theorem-466601-f1.mudulabs_project.data`

GROUP BY
    StockCode
ORDER BY
   sell_cnt DESC

LIMIT 10;
```

행	StockCode	sell_cnt
1	85123A	2301
2	22423	2192
3	85099B	2156
4	47566	1720
5	20725	1626
6	84879	1489
7	22197	1468
8	22720	1465
9	21212	1367
10	22383	1328

P.S 정리하면서 진행을 했어야 하나 망각하고 다 끝난뒤에 다시 정리하려니 문제가 생기고 있습니다.

일단 올리고 과제는 다시 복습해야 될것 같아 해보고 완성되면 추후 다시 올리겠습니다.