**OBJECTIVE**: To perform ad-hoc analysis using SQL queries on media datasets and generate insights for circulation, revenue, efficiency, and digital readiness.

## DATASET DESCRIPTION:

- fact_print_sales → Contains data on copies printed and net circulation across cities and months.
- fact_ad_revenue → Stores advertisement revenue details by year and ad category.
- fact_city_readiness → Tracks city-level literacy rate, smartphone penetration, and internet penetration over time.
- fact_digital_pilot → Includes digital engagement metrics such as active users, sessions, and engagement rate.
- dim_city → Dimension table with city attributes such as ID, name, state, and tier classification.
- dim_ad_category → Dimension table for advertisement categories with category IDs and names.

## Business Request – 1:

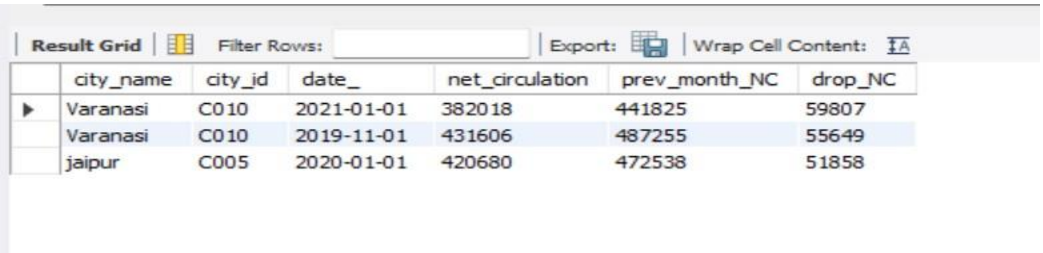
*Monthly Circulation Drop Check Generate a report showing the top 3 months (2019–2024) where any city recorded the sharpest month-over-month decline in net_circulation.*

- *Fields: • city_name • month (YYYY-MM) • net_circulation*

## Solution:

```
with cte as

(

select city_id, date_, net_circulation,

lag(net_circulation,1) over(partition by city_id order by city_id, date_) as prev_month_NC,
((lag(net_circulation,1) over(partition by city_id order by city_id, date_)) - net_circulation) as
drop_NC

from fact_print

order by city_id

)
```

```
select city_name, cte.* from cte
join dim_city using(city_id)
where prev_month_NC is not null order by drop_nc desc limit 3;
```



Result Grid | Filter Rows: | Export: | Wrap Cell Content:

| city_name | city_id | date_ | net_circulation | prev_month_NC | drop_NC |
|---|---|---|---|---|---|
| Varanasi | C010 | 2021-01-01 | 382018 | 441825 | 59807 |
| Varanasi | C010 | 2019-11-01 | 431606 | 487255 | 55649 |
| jaipur | C005 | 2020-01-01 | 420680 | 472538 | 51858 |

**Business Request – 2:**

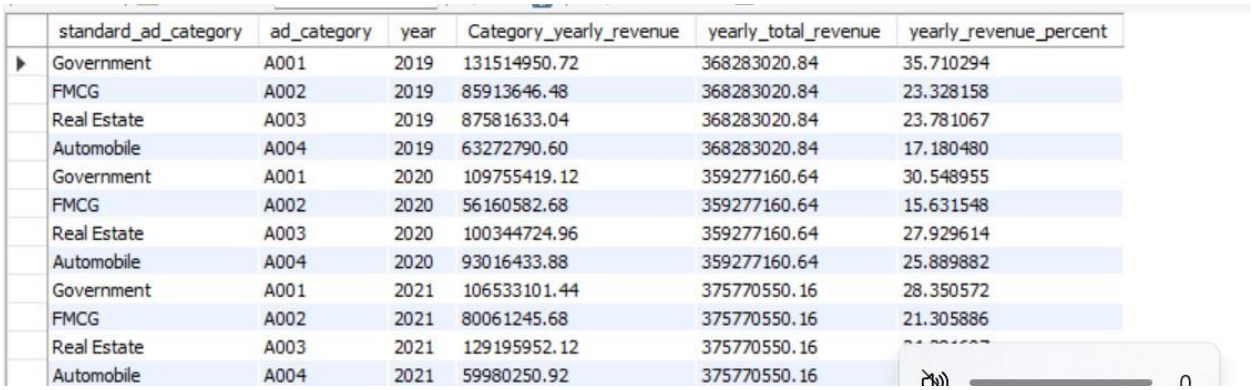
*Yearly Revenue Concentration by Category Identify ad categories that contributed > 50% of total yearly ad revenue.*

- *Fields: • year • category_name • category_revenue • total_revenue_year • pct_of_year_total*

**Solution:**

```
with cte as(
SELECT ad_category, quarter_,
case
when SUBSTRING_INDEX(quarter_, '-', 1) like "%Q%" then SUBSTRING_INDEX(quarter_, '-', 1)
else SUBSTRING_INDEX(quarter_, '-', -1) end as qtr,
case
when SUBSTRING_INDEX(quarter_, '-', 1) not like "%Q%" then SUBSTRING_INDEX(quarter_, '-', 1)
else SUBSTRING_INDEX(quarter_, '-', -1) end as year,
ad_revenue, currency,
case
when currency = "EUR" then ad_revenue*104
```

```
when currency = "USD" then ad_revenue*88
else ad_revenue end as revenue_INR
FROM fact_revenue),

cte2 as(
 select ad_category, year, sum(revenue_INR) as Category_yearly_revenue
 from cte
 group by ad_category, year)

 select standard_ad_category, cte2.*, sum(category_yearly_revenue) over(partition by year) as yearly_total_revenue,
 category_yearly_revenue / sum(category_yearly_revenue) over(partition by year)*100 as yearly_revenue_percent
 from cte2
 join dim_ad_category c on c.ad_category_id=cte2.ad_category;
```



| standard_ad_category | ad_category | year | Category_yearly_revenue | yearly_total_revenue | yearly_revenue_percent |
|---|---|---|---|---|---|
| Government | A001 | 2019 | 131514950.72 | 368283020.84 | 35.710294 |
| FMCG | A002 | 2019 | 85913646.48 | 368283020.84 | 23.328158 |
| Real Estate | A003 | 2019 | 87581633.04 | 368283020.84 | 23.781067 |
| Automobile | A004 | 2019 | 63272790.60 | 368283020.84 | 17.180480 |
| Government | A001 | 2020 | 109755419.12 | 359277160.64 | 30.548955 |
| FMCG | A002 | 2020 | 56160582.68 | 359277160.64 | 15.631548 |
| Real Estate | A003 | 2020 | 100344724.96 | 359277160.64 | 27.929614 |
| Automobile | A004 | 2020 | 93016433.88 | 359277160.64 | 25.889882 |
| Government | A001 | 2021 | 106533101.44 | 375770550.16 | 28.350572 |
| FMCG | A002 | 2021 | 80061245.68 | 375770550.16 | 21.305886 |
| Real Estate | A003 | 2021 | 129195952.12 | 375770550.16 | |
| Automobile | A004 | 2021 | 59980250.92 | 375770550.16 | |

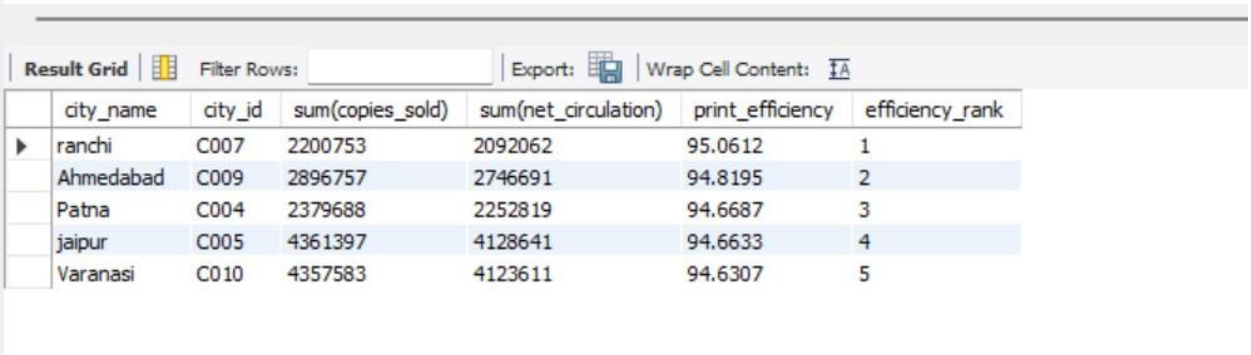
**Business Request – 3:**

*2024 Print Efficiency Leaderboard For 2024, rank cities by print efficiency = net_circulation / copies_printed. Return top 5.*

- *Fields: • city_name • copies_printed_2024 • net_circulation_2024 • efficiency_ratio = net_circulation_2024 / copies_printed_2024 • efficiency_rank_2024*

**Solution:**

```
with cte as
 (
 select city_id, sum(copies_sold), sum(net_circulation), sum(net_circulation)/sum(copies_sold)*100
as print_efficiency,
 rank() over(order by sum(net_circulation)/sum(copies_sold)*100 desc) as efficiency_rank
 from fact_print
 where year(date_) = 2024
 group by city_id
 )
select c.city_name, cte.*
from cte
join dim_city c using(city_id)
 order by efficiency_rank limit 5;
```



Result Grid | Filter Rows: | Export: | Wrap Cell Content:

| city_name | city_id | sum(copies_sold) | sum(net_circulation) | print_efficiency | efficiency_rank |
|---|---|---|---|---|---|
| ranchi | C007 | 2200753 | 2092062 | 95.0612 | 1 |
| Ahmedabad | C009 | 2896757 | 2746691 | 94.8195 | 2 |
| Patna | C004 | 2379688 | 2252819 | 94.6687 | 3 |
| jaipur | C005 | 4361397 | 4128641 | 94.6633 | 4 |
| Varanasi | C010 | 4357583 | 4123611 | 94.6307 | 5 |

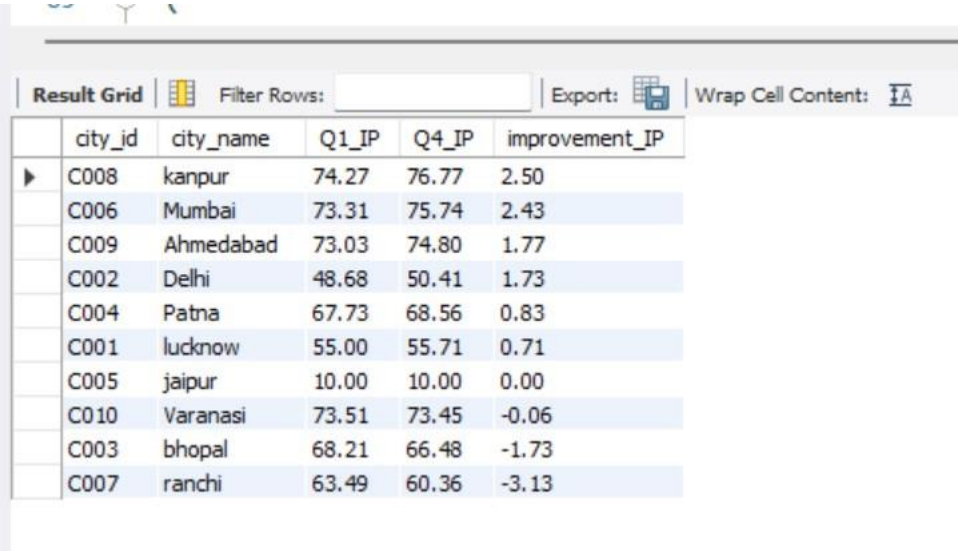
**Business Request – 4 :**

*Internet Readiness Growth (2021) For each city, compute the change in internet penetration from Q1-2021 to Q4-2021 and identify the city with the highest improvement.*

- *Fields: • city_name • internet_rate_q1_2021 • internet_rate_q4_2021 • delta_internet_rate = internet_rate_q4_2021 − internet_rate_q1_2021*

**Solution:**

```
with cte1 as
 (
 select city_id, quarter_,
 internet_penetration as Q1_IP
 from city_readiness
 where quarter_ like "2021%Q1"
 ),
 cte2 as
 (
 select city_id, quarter_,
 internet_penetration as Q4_IP
 from city_readiness
 where quarter_ like "2021%Q4"
 )

 select cte1.city_id, city_name, Q1_IP, Q4_IP, Q4_IP-Q1_IP as improvement_IP
 from cte1
 join cte2
 on cte1.city_id=cte2.city_id
 join dim_city c on cte1.city_id = c.city_id
 order by improvement_IP desc;
```



Result Grid | Filter Rows: | Export: | Wrap Cell Content:

| city_id | city_name | Q1_IP | Q4_IP | improvement_IP |
|---|---|---|---|---|
| C008 | kanpur | 74.27 | 76.77 | 2.50 |
| C006 | Mumbai | 73.31 | 75.74 | 2.43 |
| C009 | Ahmedabad | 73.03 | 74.80 | 1.77 |
| C002 | Delhi | 48.68 | 50.41 | 1.73 |
| C004 | Patna | 67.73 | 68.56 | 0.83 |
| C001 | lucknow | 55.00 | 55.71 | 0.71 |
| C005 | jaipur | 10.00 | 10.00 | 0.00 |
| C010 | Varanasi | 73.51 | 73.45 | -0.06 |
| C003 | bhopal | 68.21 | 66.48 | -1.73 |
| C007 | ranchi | 63.49 | 60.36 | -3.13 |

**Business Request – 5:**

*Consistent Multi-Year Decline (2019→2024) Find cities where both net_circulation and ad_revenue decreased every year from 2019 through 2024 (strictly decreasing sequences).*

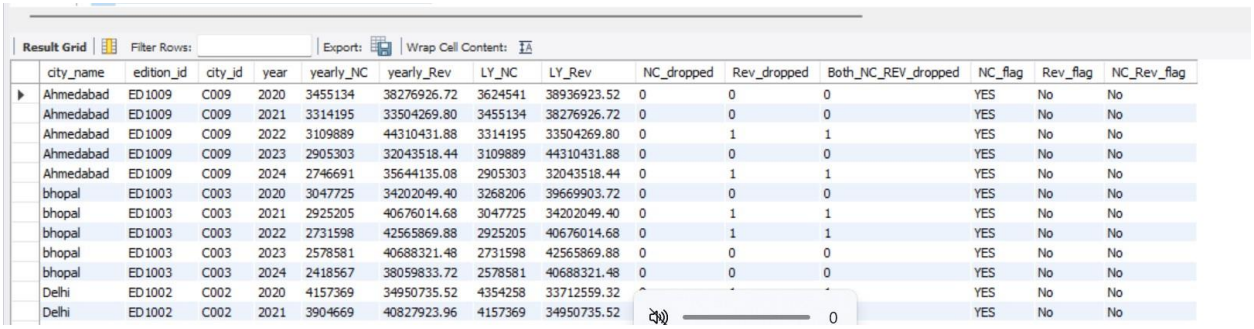
- *Fields: • city_name • year • yearly_net_circulation • yearly_ad_revenue • is_declining_print (Yes/No per city over 2019–2024) • is_declining_ad_revenue (Yes/No) • is_declining_both (Yes/No)*

**Solution:**

```
With Print_Rev_data as(
 with Print_data as
 (
select edition_ID, city_id, year(date_) as year, sum(net_circulation) as yearly_NC
from fact_print
group by edition_ID, city_id, year(date_)
),
revenue_data as
(with sub_revenue as
        (
SELECT edition_ID, quarter_,
```

```
case
when SUBSTRING_INDEX(quarter_, '-', 1) like "%Q%" then SUBSTRING_INDEX(quarter_, '-', 1)
else SUBSTRING_INDEX(quarter_, '-', -1) end as qtr,
case
when SUBSTRING_INDEX(quarter_, '-', 1) not like "%Q%" then SUBSTRING_INDEX(quarter_, '-', 1)
else SUBSTRING_INDEX(quarter_, '-', -1) end as year,
ad_revenue, currency,
case
when currency = "EUR" then ad_revenue*104
when currency = "USD" then ad_revenue*88
else ad_revenue end as revenue_INR
FROM fact_revenue
)
select edition_ID, year, sum(revenue_INR) as yearly_Rev
from sub_revenue
group by edition_ID, year
)

select p.edition_id , p.city_id, p.year, yearly_NC, yearly_Rev,
lag(yearly_NC,1) over(partition by p.edition_id order by p.edition_id, year ) as LY_NC,
lag(yearly_Rev,1) over(partition by p.edition_id order by p.edition_id, year ) as LY_Rev,
if (yearly_NC < lag(yearly_NC,1) over(partition by p.edition_id order by p.edition_id, year ), 0, 1)
as NC_dropped,
if (yearly_Rev < lag(yearly_Rev,1) over(partition by p.edition_id order by p.edition_id, year ), 0, 1)
as Rev_dropped,
if (yearly_NC < lag(yearly_NC,1) over(partition by p.edition_id order by p.edition_id, year ) and
yearly_Rev < lag(yearly_Rev,1) over(partition by p.edition_id order by p.edition_id, year ) , 0, 1) as
Both_NC_REV_dropped
```

```
from print_data p
join revenue_data r using (edition_id, year)
order by edition_id, year
)

select c.city_name , pr.*, if(pr2.a > 0, "No", "YES") as NC_flag, if(pr2.b > 0, "No", "YES") as Rev_flag, if(pr2.c > 0, "No", "YES") as NC_Rev_flag
from print_rev_data pr
join dim_city c using(city_id)
join (select city_id, sum(NC_dropped) as a,sum(rev_dropped) as b, sum(both_NC_rev_dropped) as c
from print_rev_data where LY_NC is not null
group by city_id ) as pr2 on pr2.city_id=pr.city_id
where LY_NC is not null
order by city_name, year;
```



Result Grid | Filter Rows: | Export: | Wrap Cell Content:

| city_name | edition_id | city_id | year | yearly_NC | yearly_Rev | LY_NC | LY_Rev | NC_dropped | Rev_dropped | Both_NC_REV_dropped | NC_flag | Rev_flag | NC_Rev_flag |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Ahmedabad | ED1009 | C009 | 2020 | 3455134 | 38276926.72 | 3624541 | 38936923.52 | 0 | 0 | 0 | YES | No | No |
| Ahmedabad | ED1009 | C009 | 2021 | 3314195 | 33504269.80 | 3455134 | 38276926.72 | 0 | 0 | 0 | YES | No | No |
| Ahmedabad | ED1009 | C009 | 2022 | 3109889 | 44310431.88 | 3314195 | 33504269.80 | 0 | 1 | 1 | YES | No | No |
| Ahmedabad | ED1009 | C009 | 2023 | 2905303 | 32043518.44 | 3109889 | 44310431.88 | 0 | 0 | 0 | YES | No | No |
| Ahmedabad | ED1009 | C009 | 2024 | 2746691 | 35644135.08 | 2905303 | 32043518.44 | 0 | 1 | 1 | YES | No | No |
| bhopal | ED1003 | C003 | 2020 | 3047725 | 34202049.40 | 3268206 | 39669903.72 | 0 | 0 | 0 | YES | No | No |
| bhopal | ED1003 | C003 | 2021 | 2925205 | 40676014.68 | 3047725 | 34202049.40 | 0 | 1 | 1 | YES | No | No |
| bhopal | ED1003 | C003 | 2022 | 2731598 | 42565869.88 | 2925205 | 40676014.68 | 0 | 1 | 1 | YES | No | No |
| bhopal | ED1003 | C003 | 2023 | 2578581 | 40688321.48 | 2731598 | 42565869.88 | 0 | 0 | 0 | YES | No | No |
| bhopal | ED1003 | C003 | 2024 | 2418567 | 38059833.72 | 2578581 | 40688321.48 | 0 | 0 | 0 | YES | No | No |
| Delhi | ED1002 | C002 | 2020 | 4157369 | 34950735.52 | 4354258 | 33712559.32 | | | | YES | No | No |
| Delhi | ED1002 | C002 | 2021 | 3904669 | 40827923.96 | 4157369 | 34950735.52 | | | | YES | No | No |

0

## Business Request – 6 :

*2021 Readiness vs Pilot Engagement Outlier In 2021, identify the city with the highest digital readiness score but among the bottom 3 in digital pilot engagement. readiness_score = AVG(smartphone_rate, internet_rate, literacy_rate) "Bottom 3 engagement" uses the chosen engagement metric provided (e.g., engagement_rate, active_users, or sessions).*

- *Fields: • city_name • readiness_score_2021 • engagement_metric_2021 • readiness_rank_desc • engagement_rank_asc • is_outlier (Yes/No)*

**Solution:**

```
with cte1 as
(
select city_id, round(avg(city_Readiness_score),2) as readiness_score
from
(
select *, Round((literacy_rate+smartphone_penetration+internet_penetration)/3,2) as City_readiness_score
from city_readiness
where quarter_ like "%2021%" ) as abc
group by city_id
),
cte2 as
(
select city_id, sum(users_reached) as reached_count, sum(downloads_or_accesses) as download_count, round(avg(avg_bounce_rate),2) as bounce_rate
from fact_pilot
group by city_id
),
cte3 as
(
 select c.city_name, cte2.*, cte1.readiness_score,
 dense_rank() over(order by cte1.readiness_score desc) as readiness_rank,
 dense_rank() over(order by cte2.reached_count asc) as engagement_rank
 from cte1
 join cte2 on cte1.city_id = cte2.city_id
```

```
join dim_city c on cte1.city_id = c.city_id
order by engagement_rank desc
)
select *,
case
when readiness_rank <= 3 and engagement_rank <= 3 then "Outlier"
else "Not an Outlier" end as Outlier_Flag
from cte3
where engagement_rank < 4 order by readiness_rank limit 1;
```

```
138  (
139      select *, Round((literacy_rate+smartphone_penetration+internet_penetration)/3,2) as City_readiness_score
```

Result Grid | Filter Rows: | Export: | Wrap Cell Content:

| city_name | city_id | reached_count | download_count | bounce_rate | readiness_score | readiness_rank | engagement_rank | Outlier_Flag |
|---|---|---|---|---|---|---|---|---|
| kanpur | C008 | 88749 | 36289 | 72.80 | 75.23 | 1 | 1 | Outlier |