



Newspaper's Survival in a Post-COVID Digital Era

Challenge By Code Basics

NAME : FARHAN HUSSAIN

TOPIC

Provide Insights to Guide a Legacy Newspaper's Survival in a Post-COVID Digital Era

PROBLEM STATEMENT

Bharat Herald, a legacy newspaper organization with operations across five Indian states, is facing an existential crisis.

With a reputation built over seven decades for multilingual reporting and sharp regional journalism, the company once printed over 1.2 million copies daily.

But between 2019 and 2024, print circulation has dropped to under 560,000. The pandemic accelerated a shift toward digital news consumption, and competitors like DigiHindi Post, NewsZilla, and InShorts quickly adapted with mobile-first platforms, WhatsApp delivery models, and subscription bundles.

Meanwhile, Bharat Herald's 2021 e-paper pilot failed due to poor mobile usability and was eventually shelved after financial losses.

The result: advertiser confidence declined, vendor payments delayed, and multiple city bureaus shut down.

Over 60 employees across editorial and sales functions have been let go. A recent internal audit flagged severe fiscal concerns.

In response, Tony Sharma (Executive Director) has initiated a last-mile attempt to assess the business situation.

A data analyst, Peter Pandey, has been assigned to lead a comprehensive analysis of the company's operational and financial data from 2019–2024.

Peter's mandate is clear: quantify what went wrong, identify recovery potential, and recommend a phased roadmap toward digital transformation.

Dataset:

I have successfully loaded all of the provided datasets

- **dim_ad_category.csv**
- **dim_city.csv**
- **fact_ad_revenue.csv**
- **fact_city_readiness.csv**
- **fact_digital_pilot.csv** and
- **fact_print_sales.csv**

into a database in MySQL for your analysis.

The screenshot shows the MySQL Workbench interface. At the top, there is a toolbar with various icons. Below the toolbar, a text area contains the following SQL code:

```
SQL File 22* ×
CREATE DATABASE Newspaper_survival_analysis;
USE Newspaper_survival_analysis;
SELECT * FROM dim_ad_category;
SELECT * FROM dim_city;
SELECT * FROM fact_ad_revenue;
SELECT * FROM fact_city_readiness;
SELECT * FROM fact_digital_pilot;
SELECT * FROM fact_print_sales;
```

Below the code, the execution log displays the results of each query:

Query ID	Time	Statement	Affected Rows	Execution Time
96	22:51:47	USE Newspaper_survival_analysis	0 row(s) affected	0.141 sec
97	22:51:47	SELECT * FROM dim_ad_category LIMIT 0, 1000	4 row(s) returned	0.079 sec / 0.000 sec
98	22:51:48	SELECT * FROM dim_city LIMIT 0, 1000	10 row(s) returned	0.000 sec / 0.000 sec
99	22:51:48	SELECT * FROM fact_ad_revenue LIMIT 0, 1000	720 row(s) returned	0.000 sec / 0.000 sec
100	22:51:48	SELECT * FROM fact_city_readiness LIMIT 0, 1000	240 row(s) returned	0.000 sec / 0.000 sec
101	22:51:48	SELECT * FROM fact_digital_pilot LIMIT 0, 1000	48 row(s) returned	0.000 sec / 0.000 sec
102	22:51:48	SELECT * FROM fact_print_sales LIMIT 0, 1000	654 row(s) returned	0.000 sec / 0.000 sec

Dim_ad_category.csv

The screenshot shows the MySQL Workbench interface with a result grid for the Dim_ad_category.csv dataset. The grid has the following columns: ad_category_id, standard_category, category_group, and example_brands. The data is as follows:

	ad_category_id	standard_category	category_group	example_brands
▶	A001	Government	Public Sector	LIC, SBI
	A002	FMCG	Commercial Brands	HUL, Britannia
	A003	Real Estate	Private Sector	DLF, Lodha
	A004	Automobile	Commercial Brands	Tata Motors, Maruti

Dim_city.csv

Result Grid | Filter Rows:

	city_id	city	state	tier
▶	C001	lucknow	Uttar Pradesh	Tier 2
	C002	Delhi	DELHI	Tier 1
	C003	bhopal	Madhya Pradesh	Tier 2
	C004	Patna	BIHAR	Tier 2
	C005	jaipur	Rajasthan	Tier 2
	C006	Mumbai	MAHARASHTRA	Tier 1
	C007	ranchi	JHARKHAND	Tier 3

Fact_ad_revenue

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

	edition_id	ad_category	quarter	ad_revenue	currency	comments
▶	ED 1005	A001	2023-Q2	22613.69	EUR	
	ED 1005	A002	Q1-2019	39366.88	USD	
	ED 1001	A003	Q3-2023	3709860	INR	
	ED 1003	A002	Q3-2023	40969.55	USD	
	ED 1007	A003	4th Qtr 2020	51779.4	USD	
	ED 1001	A001	2024-Q2	4790190	INR	
	ED 1001	A003	2024-Q2	4571581	INR	

Fact_city_readiness

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

	city_id	quarter	literacy_rate	smartphone_penetration	internet_penetration
▶	C001	2019-QQ	89.16	75.76	56.53
	C001	2019-QQ	88.76	76.45	55.97
	C001	2019-QQ	88.83	75.32	56.52
	C001	2019-QQ	89.25	75.83	56.94
	C001	2020-QQ	89.13	75.03	56.45
	C001	2020-QQ	89.06	76.56	56.44
	C001	2020-QQ	89.1	74.44	55.7

Fact_digital_pilot

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

	platform	launch_month	ad_category_id	dev_cost	marketing_cost	users_reached	downloads_or_accesses	avg_bounce_rate	customer_feedback
▶	PDF WhatsApp Push	2021-01-01	A001	236570	66060	23509	16319	52.55	Mixed feedback: some usability concerns, bu
	PDF WhatsApp Push	2021-02-01	A001	156865	99122	19472	17017	82.53	Mixed feedback: some usability concerns, bu
	PDF WhatsApp Push	2021-03-01	A001	242728	46087	8471	2891	68.06	Mixed feedback: some usability concerns, bu
	PDF WhatsApp Push	2021-04-01	A001	147695	78868	46796	15640	66.17	Mixed feedback: some usability concerns, bu
	PDF WhatsApp Push	2021-05-01	A001	325906	135644	16805	3231	76.9	The site takes too long to load on average p
	PDF WhatsApp Push	2021-06-01	A001	248972	53225	35268	16551	45.74	Mixed feedback: some usability concerns, bu

Fact_print_sales

	edition_ID	City_ID	Language	State	Month	copies_sold	copies_returned	Net_Circulation
▶	ED1005	C005	Hindi	Rajasthan	2023-05-01	404389	13510	390879
	ED1005	C005	Hindi	Rajasthan	2019-03-01	492943	25024	467919
	ED1001	C001	Hindi	Uttar pradesh	2023-07-01	168893	12285	156608
	ED1003	C003	Hindi	Madhya pradesh	2023-07-01	216540	10117	206423
	ED1007	C007	Hindi	Jharkhand	2020-10-01	234563	13048	221515
	ED1001	C001	Hindi	Uttar pradesh	2024-06-01	149986	6387	143599
	ED1001	C001	Hindi	Uttar pradesh	2024-04-01	155269	9394	145875

Data Preprocessing:

1. Cleaning dim_ad_category

- **ALTER TABLE - CHANGE COLUMN:** Renames columns for clarity and consistency.
- **UPDATE - SET TRIM():** Removes any leading or trailing whitespace from text fields.

```
-- Cleaning and Preprocessing Process :

-- Dim_ad_Category
-- Renaming columns for clarity
ALTER TABLE dim_ad_category
CHANGE COLUMN standard_ad_category standard_category VARCHAR(100),
CHANGE COLUMN example_brands example_brands TEXT;

-- Trim whitespace in all text fields
UPDATE dim_ad_category
SET ad_category_id = TRIM(ad_category_id),
    standard_category = TRIM(standard_category),
    category_group = TRIM(category_group),
    example_brands = TRIM(example_brands);
```

2. Cleaning fact_city_readiness

- **ALTER TABLE - DROP COLUMN “MyUnknownColumn”** Deletes an unnecessary index column created during the import process.
- **SHOW COLUMNS:** Displays the table's column names and data types for review.
- **UPDATE - SET CONCAT():** Standardizes the quarter column to a consistent YYYY-QN format for easier analysis.

```

-- Cleaning on Fact_city_readiness :
-- Drop useless index column
ALTER TABLE fact_city_readiness DROP COLUMN MyUnknownColumn;

SHOW COLUMNS FROM fact_city_readiness;

-- Standardizing quarter (force YYYY-QN format)
UPDATE fact_city_readiness
SET quarter = CONCAT(SUBSTRING(quarter,1,4), '-Q', SUBSTRING(quarter,6,1))
WHERE quarter REGEXP '^[0-9]{4}[- ]?[Qq][1-4]$';

-- Cleaning fact_digital_Pilot :
SELECT * FROM fact_digital_pilot;

-- Dropping MyUnknown Column Field :
ALTER TABLE fact_digital_pilot DROP COLUMN MyUnknownColumn;

-- Renaming long column
ALTER TABLE fact_digital_pilot
CHANGE COLUMN cumulative_feedback_from_customers customer_feedback TEXT;

-- Clean up values
UPDATE fact_digital_pilot
SET launch_month = STR_TO_DATE(CONCAT(launch_month, '-01'), '%Y-%m-%d');

```

3. Cleaning fact_digital_pilot

- **SELECT *:** A query to view all data in the table.
- **ALTER TABLE - DROP COLUMN MyUnknownColumn:** Removes an unwanted index column.
- **ALTER TABLE - CHANGE COLUMN:** Renames a long column to a shorter, and give a clearer name (**customer_feedback**).
- **UPDATE - SET STR_TO_DATE():** Converts the launch_month string field into a proper **DATE** data type.

```

-- Cleaning fact_digital_Pilot :
SELECT * FROM fact_digital_pilot;

-- Dropping MyUnknown Column Field :
ALTER TABLE fact_digital_pilot DROP COLUMN MyUnknownColumn;

-- Renaming long column
ALTER TABLE fact_digital_pilot
CHANGE COLUMN cumulative_feedback_from_customers customer_feedback TEXT;

-- Clean up values
UPDATE fact_digital_pilot
SET launch_month = STR_TO_DATE(CONCAT(launch_month, '-01'), '%Y-%m-%d');

```

4. Cleaning fact_print_sales

- **SELECT *:** A query to view all data in the table.
- **UPDATE - SET CONCAT() (Language and State):** Standardizes the Language and State columns to have consistent capitalization.
- **UPDATE - SET REPLACE():** Removes underscores from the State column.
- **ALTER TABLE - CHANGE:** Renames the Copies Sold column to a more correct format (**copies_sold**).
- **ALTER TABLE - MODIFY - INT/DATE:** Converts the **copies_sold** and **Month** columns to their correct **INT** and **DATE** data types, respectively, to allow for numerical and temporal operations.

```
-- Cleaning fact_print_sales :  
SELECT * FROM fact_print_sales;  
  
-- Making first letter capital in Language field :  
  
UPDATE fact_print_sales  
SET Language = CONCAT(  
    UPPER(LEFT(Language, 1)),  
    LOWER(SUBSTRING(Language, 2)))  
)  
WHERE Language IS NOT NULL;
```

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

```
WITH Circulation_drop AS (  
    SELECT dc.City AS city_name, DATE_FORMAT(fps.Month, '%Y-%m') AS month, fps.Net_Circulation AS net_circulation,  
    (fps.Net_Circulation - LAG(fps.Net_Circulation) OVER(PARTITION BY fps.City_ID ORDER BY fps.Month))  
    ) AS month_change  
    FROM fact_print_sales fps JOIN dim_city dc ON fps.City_ID = dc.city_id  
    WHERE YEAR(fps.Month) BETWEEN 2019 AND 2024  
    )  
    SELECT city_name, month, net_circulation, month_change  
    FROM Circulation_drop  
    WHERE month_change IS NOT NULL  
    ORDER BY month_change ASC  
    LIMIT 3; --- Negative_values means its declining : Positive means increasing
```

Result Grid | Filter Rows: _____ | Export: | W

	city_name	month	net_circulation	month_change
▶	Varanasi	2021-01	382018	-59807
	Varanasi	2019-11	431606	-55649
	jaipur	2020-01	420680	-54681

Explanation:

Using the **LAG** function, the query computes the month-over-month change in **Net_Circulation** for each city using a **Common Table Expression (CTE)**. In order to identify the sharpest drops, it then filters for negative changes, or declines, and arranges them from smallest to largest. Lastly, it displays the top three examples of these declines using **LIMIT 3**.

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

```

WITH yearly_category_revenue AS (
    SELECT YEAR(far.quarter) AS year, dac.standard_ad_category AS category_name, SUM(far.ad_revenue) AS category_revenue
    FROM fact_ad_revenue far JOIN dim_ad_category dac
    ON far.ad_category = dac.ad_category_id
    GROUP BY YEAR(far.quarter), dac.standard_ad_category
),
yearly_total AS (
    SELECT
        YEAR(quarter) AS year,
        SUM(ad_revenue) AS total_revenue_year
    FROM fact_ad_revenue
    GROUP BY YEAR(quarter)
)
SELECT ycr.year, ycr.category_name, ycr.category_revenue, yt.total_revenue_year,
    ROUND((ycr.category_revenue / yt.total_revenue_year) * 100, 2) AS pct_of_year_total
FROM yearly_category_revenue ycr JOIN yearly_total yt
ON ycr.year = yt.year
WHERE (ycr.category_revenue / yt.total_revenue_year) > 0.5
ORDER BY ycr.year, pct_of_year_total DESC;

```

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

	year	category_name	category_revenue	total_revenue_year	pct_of_year_total

Explanation :

The query returned a **blank output** because no single ad category contributed more than **50%** of the total yearly **ad revenue**. The revenue is distributed across multiple **ad categories**, and none of them individually meet the filter condition

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

```

WITH efficiency_2024 AS (
    SELECT dc.City AS city_name,
        SUM(fps.'Copies Sold' + fps.copies_returned) AS copies_printed_2024, -- the column name is not in correct format so we have used parenthesis
        SUM(fps.Net_Circulation) AS net_circulation_2024,
        ROUND(SUM(fps.Net_Circulation) / SUM(fps.'Copies Sold' + fps.copies_returned), 4) AS efficiency_ratio
    FROM fact_print_sales fps JOIN dim_city dc
    ON fps.City_ID = dc.city_id
    WHERE YEAR(fps.Month) = 2024
    GROUP BY dc.City
)
SELECT
    city_name,
    copies_printed_2024,
    net_circulation_2024,
    efficiency_ratio,
    RANK() OVER (ORDER BY efficiency_ratio DESC) AS efficiency_rank_2024
FROM efficiency_2024
ORDER BY efficiency_rank_2024
LIMIT 5;

```

Result Grid		Filter Rows:	Export:	Wrap Cell Content:	
	city_name	copies_printed_2024	net_circulation_2024	efficiency_ratio	efficiency_rank_2024
▶	ranchi	2117296	1919038	0.9064	1
	Ahmedabad	2780036	2518120	0.9058	2
	jaipur	4594153	4128641	0.8987	3
	Varanasi	4591555	4123611	0.8981	4
	Patna	2301095	2062729	0.8964	5

Explanation :

By calculating the ratio of net_circulation to **copies_printed**, the query ranks **cities** for 2024 according to their print efficiency.

The efficiency ratio, net circulation, and total copies printed for each city in 2024 are determined using a **Common Table Expression (CTE)**. Each city is then given a rank according to its efficiency ratio using the **RANK()** window function in the main query, the **city** with the highest ratio is given rank 1. Lastly, the **top five cities** on the leaderboard are returned.

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

```

WITH Readiness AS (
    SELECT
        City_ID,
        CASE WHEN Quarter = '2021-Q1' THEN internet_penetration END AS internet_rate_q1_2021,
        CASE WHEN Quarter = '2021-Q4' THEN internet_penetration END AS internet_rate_q4_2021
    FROM fact_city_readiness
    WHERE Quarter IN ('2021-Q1', '2021-Q4')
)
SELECT
    dim_city.city AS city_name,
    MAX(Readiness.internet_rate_q1_2021) AS internet_rate_q1_2021,
    MAX(Readiness.internet_rate_q4_2021) AS internet_rate_q4_2021,
    MAX(Readiness.internet_rate_q4_2021) - MAX(Readiness.internet_rate_q1_2021) AS delta_internet_rate
FROM Readiness JOIN dim_city
ON Readiness.City_ID = dim_city.city_id
GROUP BY city_name
ORDER BY delta_internet_rate DESC;

```

	city_name	internet_rate_q1_2021	internet_rate_q4_2021	delta_internet_rate
▶	kanpur	74.27	76.77	2.5
	Mumbai	73.31	75.74	2.4299999999999926
	Ahmedabad	73.03	74.8	1.7699999999999996
	Delhi	48.68	50.41	1.7299999999999969
	Patna	67.73	68.56	0.8299999999999983
	lucknow	55	55.71	0.7100000000000009
	jaipur	10	10	0
	Varanasi	73.51	73.45	-0.06000000000002274
	bhopal	68.21	66.48	-1.7299999999999898
	ranchi	63.49	60.36	-3.1300000000000026

Explanation :

The query determines how much each city's internet penetration changed **between Q1 and Q4 of 2021**. The internet penetration rates for these two particular quarters are chosen using a **Common Table Expression (CTE)**. After joining this data with the **dim_city table**, the main query pivots the data using **MAX()** with **GROUP BY** and computes the difference (**delta_internet_rate**). To determine which city has improved the most, the results are arranged in order.

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)

```

WITH yearly_aggregates AS (
    SELECT
        dc.city AS city_name,
        YEAR(fps.Month) AS year,
        SUM(fps.Net_Circulation) AS yearly_net_circulation,
        SUM(
            CASE
                WHEN far.currency = 'INR' THEN far.ad_revenue
                ELSE 0
            END
        ) AS yearly_ad_revenue
    FROM fact_print_sales AS fps JOIN dim_city AS dc
        ON fps.city_id = dc.city_id
        JOIN fact_ad_revenue AS far ON fps.edition_id = far.edition_id
    WHERE YEAR(fps.Month) BETWEEN 2019 AND 2024
    GROUP BY 1, 2
),
lag_data AS (
    SELECT city_name, year, yearly_net_circulation, yearly_ad_revenue,
        LAG(yearly_net_circulation, 1) OVER (
            PARTITION BY city_name

```

```

        PARTITION BY city_name
        ORDER BY
            year
        ) AS prev_net_circulation
    FROM
        yearly_aggregates
),
decline_check AS (
    SELECT
        city_name,
        SUM(
            CASE
                WHEN yearly_net_circulation < prev_net_circulation THEN 1
                ELSE 0
            END
        ) AS declining_print_count,
        SUM(
            CASE
                WHEN yearly_ad_revenue < prev_ad_revenue THEN 1
                ELSE 0
            END
        ) AS declining_ad_revenue_count
)
SELECT city_name, declining_print_count, declining_ad_revenue_count
FROM decline_check
ORDER BY declining_ad_revenue_count DESC

```

```

        END
    ) AS declining_print_count,
SUM(
CASE
    WHEN yearly_ad_revenue < prev_ad_revenue THEN 1
    ELSE 0
END
) AS declining_ad_revenue_count
FROM lag_data
GROUP BY city_name
HAVING COUNT(year) = 6 -- Ensure the city has data for all 6 years (2019 to 2024)
)
SELECT T1.city_name, T1.year, T1.yearly_net_circulation, T1.yearly_ad_revenue,
CASE
    WHEN T2.declining_print_count = 5 THEN 'Yes'
    ELSE 'No'
END AS is_declining_print,
CASE
    WHEN T2.declining_ad_revenue_count = 5 THEN 'Yes'
WHEN T2.declining_ad_revenue_count = 5 THEN 'Yes'
    ELSE 'No'
END AS is_declining_ad_revenue,
CASE
    WHEN T2.declining_print_count = 5
    AND T2.declining_ad_revenue_count = 5 THEN 'Yes'
    ELSE 'No'
END AS is_declining_both
FROM yearly_aggregates AS T1 JOIN decline_check AS T2
ON T1.city_name = T2.city_name
WHERE T2.declining_print_count = 5 AND T2.declining_ad_revenue_count = 5
ORDER BY T1.city_name, T1.year;

```

Result Grid							
Filter Rows: <input type="text"/>							
Export: <input type="button"/> Wrap Cell Content: <input type="checkbox"/>							
city_name	year	yearly_net_circulation	yearly_ad_revenue	is_declining_print	is_declining_ad_revenue	is_declining_both	

Explanation :

The purpose of this search is to identify cities where, from **2019 to 2024**, **net circulation and ad revenue declined annually**.

Due to the extremely stringent requirements for a city to be included in the final result, the **output is blank**. For a city to qualify, its **net circulation and ad revenue** must have **decreased** annually for each of the five years in a row (**2019-2020, 2020-2021, 2021-2022, 2022-2023, and 2023-2024**). For both metrics, it is very likely that no city in the dataset satisfies this precise, consecutive decline criterion.

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)

```

WITH readiness_data AS (
    SELECT
        city_id, AVG((literacy_rate + smartphone_penetration + internet_penetration) / 3) AS readiness_score_2021
    FROM fact_city_readiness
    WHERE quarter LIKE '%2021%'
    GROUP BY city_id
),
engagement_data AS (
    SELECT
        city_id, SUM(users_reached) AS engagement_metric_2021
    FROM fact_digital_pilot
    WHERE launch_month LIKE '%2021%'
    GROUP BY city_id
),
ranked_data AS (
    SELECT dc.city AS city_name, rd.readiness_score_2021, ed.engagement_metric_2021,
        RANK() OVER (ORDER BY rd.readiness_score_2021 DESC) AS readiness_rank_desc,
        RANK() OVER (ORDER BY ed.engagement_metric_2021 ASC) AS engagement_rank_asc
    FROM readiness_data AS rd
    JOIN engagement_data AS ed ON rd.city_id = ed.city_id
    JOIN dim_city AS dc ON rd.city_id = dc.city_id
    JOIN dim_city AS dc ON rd.city_id = dc.city_id
)
SELECT city_name, readiness_score_2021, engagement_metric_2021, readiness_rank_desc, engagement_rank_asc,
CASE
    WHEN readiness_rank_desc = 1
    AND engagement_rank_asc <= 3 THEN 'Yes'
    ELSE 'No'
END AS is_outlier
FROM ranked_data
ORDER BY readiness_rank_desc;

```

Result Grid						
	city_name	readiness_score_2021	engagement_metric_2021	readiness_rank_desc	engagement_rank_asc	is_outlier
▶	kanpur	75.23083333333332	88749	1	1	Yes
	Varanasi	73.88749999999999	143151	2	10	No
	bhopal	73.21	139626	3	9	No
	lucknow	73.20416666666667	123945	4	6	No
	Ahmedabad	72.39333333333333	135003	5	8	No
	Patna	70.77083333333334	121974	6	5	No
	ranchi	68.64083333333333	110125	7	2	No
	Mumbai	68.33166666666666	128561	8	7	No
	Delhi	56.07583333333335	121423	9	4	No
	jaipur	54.94750000000005	119681	10	3	No

Explanation :

In 2021, the query pinpoints a particular kind of outlier city: one that is among the least engaged in pilot projects but has the highest level of digital readiness.

It first determines a **readiness_score** and an **engagement_metric** for every city using two **Common Table Expressions (CTEs)**. The cities are then ranked according to both metrics: **engagement** is ranked from lowest to highest, and readiness is ranked from **highest to lowest**. Lastly, to find the outlier, the main query filters for the city with the highest readiness rank (**readiness_rank_desc = 1**) and an engagement rank in the bottom three (**engagement_rank_asc <= 3**).

Primary And Secondary Analysis

1. Print Circulation Trends What is the trend in copies printed, copies sold, and net circulation across all cities from 2019 to 2024? How has this changed year-over-year?

```
-- Primary and Secondary_Analysis:

/*1. Print Circulation Trends
What is the trend in copies printed, copies sold, and net circulation across all
cities from 2019 to 2024? How has this changed year-over-year? */

SELECT
    YEAR(Month) AS year,
    SUM(copies_sold) AS total_copies_sold,
    SUM(copies_returned) AS total_copies_returned,
    SUM(Net_Circulation) AS total_net_circulation
FROM fact_print_sales
WHERE YEAR(Month) BETWEEN 2019 AND 2024
GROUP BY YEAR(Month)
ORDER BY YEAR(Month);
```

year	total_copies_sold	total_copies_returned	total_net_circulation
2019	37124942	2004339	35120603
2020	34573433	1872574	32700859
2021	36001741	1946595	34055146
2022	33396373	1774372	31622001
2023	28846124	1548929	27297195

Explanation:

From the **fact_print_sales** table, the query aggregates **Net_Circulation**, **copies_sold**, and **copies_returned** by year. To display the **annual trends**, it first groups the results by year after filtering for data from **2019 to 2024**.

2. To Performing Cities Which cities contributed the highest to net circulation and copies sold in 2024?
Are these cities still profitable to operate in?

```
/* 2. To Performing Cities  
Which cities contributed the highest to net circulation and copies sold in 2024?  
Are these cities still profitable to operate in? */  
  
SELECT  
    d.city,  
    SUM(f.copies_sold) AS total_copies_sold_2024,  
    SUM(f.Net_Circulation) AS total_net_circulation_2024  
FROM fact_print_sales f JOIN dim_city d  
ON f.City_ID = d.city_id  
WHERE YEAR(f.Month) = 2024  
GROUP BY d.city  
ORDER BY total_net_circulation_2024 DESC, total_copies_sold_2024 DESC;
```

Result Grid			
	city	total_copies_sold_2024	total_net_circulation_2024
▶	jaipur	4361397	4128641
	Varanasi	4357583	4123611
	Mumbai	3775800	3569229
	kanpur	3159143	2985718
	Delhi	2912054	2726917

Explanation:

The cities that made the biggest contributions to net circulation and copies sold in **2024** are identified by this query. To connect sales information to city names, it joins the **dim_city** and **fact_print_sales** tables. To display the **best performers first**, the query first aggregates the sales and circulation numbers for each city, **filters for data from 2024**, and then sorts the results in **descending order**.

3. Print Waste Analysis Which cities have the largest gap between copies printed and net circulation, and how has that gap changed over time?

```
/* 3. Print Waste Analysis
Which cities have the largest gap between copies printed and net circulation, and
how has that gap changed over time? */

SELECT
    d.city,
    YEAR(f.Month) AS year,
    (SUM(f.copies_sold) + SUM(f.copies_returned)) AS total_copies_printed,
    SUM(f.Net_Circulation) AS total_net_circulation,
    (SUM(f.copies_sold) + SUM(f.copies_returned)) - SUM(f.Net_Circulation) AS gap_in_circulation
FROM fact_print_sales f JOIN dim_city d
ON f.City_ID = d.city_id
GROUP BY d.city, YEAR(f.Month)
ORDER BY gap_in_circulation DESC;
```

	city	year	total_copies_printed	total_net_circulation	gap_in_circulation
▶	Varanasi	2019	5743956	5085718	658238
	Varanasi	2021	5406136	4816364	589772
	Mumbai	2020	5107030	4560074	546956
	Varanasi	2020	5316969	4775365	541604
	jaipur	2021	5475602	4939028	536574

Explanation:

The gap between **copies printed and net circulation** for each city, broken down by year, is calculated in this query to analyze **print waste**. The city name is obtained by joining **dim_city** and **fact_print_sales**, and the data is then aggregated. By deducting **Net_Circulation** from the total of **copies_sold** and **copies_returned**, the gap is calculated. The cities with the biggest gaps are then displayed at the **top** of the results.

4. Ad Revenue Trends by Category How has ad revenue evolved across different ad categories between 2019 and 2024? Which categories have remained strong, and which have declined?

```
/*4. Ad Revenue Trends by Category
How has ad revenue evolved across different ad categories between 2019 and
2024? Which categories have remained strong, and which have declined? */

SELECT
    d.standard_category AS ad_category,
    CASE
        WHEN a.quarter LIKE '%2019%' THEN 2019
        WHEN a.quarter LIKE '%2020%' THEN 2020
        WHEN a.quarter LIKE '%2021%' THEN 2021
        WHEN a.quarter LIKE '%2022%' THEN 2022
        WHEN a.quarter LIKE '%2023%' THEN 2023
        WHEN a.quarter LIKE '%2024%' THEN 2024
        ELSE NULL
    END AS year,
    SUM(CAST(a.ad_revenue AS DECIMAL)) AS total_ad_revenue
FROM fact_ad_revenue a JOIN dim_ad_category d ON a.ad_category = d.ad_category_id
WHERE a.quarter LIKE '%201%' OR a.quarter LIKE '%202%'
GROUP BY d.standard_category , year
ORDER BY d.standard_category , year;
```

Result Grid | Filter Rows: Export:

	ad_category	year	total_ad_revenue
▶	Automobile	2019	58348531
	Automobile	2020	78758034
	Automobile	2021	41289680
	Automobile	2022	48273967
	Automobile	2023	45853846
	Automobile	2024	68578286
	FMCG	2019	66495348
	FMCG	2020	27255938
	FMCG	2021	48909511
	FMCG	2022	56769664

Explanation:

This examines trends in ad revenue by category from **2019 to 2024**. The **dim_ad_category** and **fact_ad_revenue** tables are joined. After extracting the year from the irregularly formatted quarter column using a **CASE** statement, it aggregates the data by **year and category** to determine the total **ad revenue** for each.

5. City-Level Ad Revenue Performance Which cities generated the most ad revenue, and how does that correlate with their print circulation?

```
/* 5. City-Level Ad Revenue Performance
Which cities generated the most ad revenue, and how does that correlate with
their print circulation? */

SELECT
    d.city,
    SUM(CAST(a.ad_revenue AS DECIMAL)) AS total_ad_revenue,
    SUM(f.Net_Circulation) AS total_net_circulation
FROM fact_ad_revenue a
JOIN fact_print_sales f ON a.edition_id = f.edition_ID
JOIN dim_city d ON f.City_ID = d.city_id
GROUP BY d.city
ORDER BY total_ad_revenue DESC, total_net_circulation DESC;
```

Result Grid | Filter Rows: Export: Wrap

	city	total_ad_revenue	total_net_circulation
▶	Patna	12480644280	1026444384
	bhopal	11645682114	1120946472
	lucknow	10994804990	799728912
	Delhi	10912020990	1481728176
	Ahmedabad	10695786752	1220581080
	Mumbai	9995999876	1680620184
	ranchi	9916108352	939553776
	kanpur	9700319808	1449462960
	Varanasi	9321042537	1984779792

Explanation:

This query compares the print circulation of the cities with the highest ad revenue. Three tables are joined **dim_city**, **fact_print_sales**, and **fact_ad_revenue**. Each city's total **ad revenue** and **total net circulation** are combined in the query. You can see which cities perform best in both metrics by sorting the results **first by ad revenue and then by net circulation in descending order**.

6. Digital Readiness vs. Performance Which cities show high digital readiness (based on smartphone, internet, and literacy rates) but had low digital pilot engagement?

```
/* 6. Digital Readiness vs. Performance

Which cities show high digital readiness (based on smartphone, internet, and
literacy rates) but had low digital pilot engagement? */

SELECT
    dc.city,
    ((AVG(fcr.literacy_rate) + AVG(fcr.smartphone_penetration) + AVG(fcr.internet_penetration)) / 3) AS digital_readiness_score,
    SUM(fdp.users_reached + fdp.downloads_or_accesses) AS digital_engagement_score
FROM dim_city dc
JOIN fact_city_readiness fcr ON dc.city_id = fcr.city_id
JOIN fact_digital_pilot fdp ON dc.city_id = fdp.city_id
GROUP BY dc.city
ORDER BY digital_readiness_score DESC, digital_engagement_score ASC;
```

	city	digital_readiness_score	digital_engagement_score
▶	kanpur	75.09874999999998	3000912
	Varanasi	74.17875	5421936
	lucknow	73.49444444444441	4964352
	bhopal	73.28777777777778	5345688
	Ahmedabad	72.735	5225616
	Patna	70.5268055555556	4424736
	ranchi	68.63875	3571464
	Mumbai	68.32499999999997	4849920
	Delhi	56.04708333333332	4771224
	jaipur	55.01652777777779	4385952

Explanation:

Cities that are well-suited for a digital transition but have not yet fully embraced it are identified by this query. The **dim_city**, **fact_city_readiness**, and **fact_digital_pilot** tables are joined. The query then **averages internet penetration, smartphone usage, and literacy** to determine a **digital_readiness_score**. Additionally, it adds up the **users_reached** and **downloads_or_accesses** to determine a **digital_engagement_score**. Cities with a **high readiness score and low engagement** are given priority in the results, making them ideal candidates for a digital relaunch.

7. Ad Revenue vs. Circulation ROI Which cities had the highest ad revenue per net circulated copy? Is this ratio improving or worsening over time?

```
/* 7. Ad Revenue vs. Circulation ROI
Which cities had the highest ad revenue per net circulated copy? Is this ratio
improving or worsening over time? */

SELECT
    d.city,
    CASE
        WHEN a.quarter LIKE '%2019%' THEN 2019
        WHEN a.quarter LIKE '%2020%' THEN 2020
        WHEN a.quarter LIKE '%2021%' THEN 2021
        WHEN a.quarter LIKE '%2022%' THEN 2022
        WHEN a.quarter LIKE '%2023%' THEN 2023
        WHEN a.quarter LIKE '%2024%' THEN 2024
        ELSE NULL
    END AS year,
    SUM(CAST(a.ad_revenue AS DECIMAL)) AS total_ad_revenue,
    SUM(f.Net_Circulation) AS total_net_circulation,
    (SUM(CAST(a.ad_revenue AS DECIMAL)) / SUM(f.Net_Circulation)) AS revenue_per_copy
FROM fact_ad_revenue a
JOIN fact_print_sales f ON a.edition_id = f.edition_ID
JOIN dim_city d ON f.City_ID = d.city_id
WHERE a.quarter LIKE '%201%' OR a.quarter LIKE '%202%'
GROUP BY d.city, year
ORDER BY revenue_per_copy DESC, year;
```

	city	year	total_ad_revenue	total_net_circulation	revenue_per_copy
▶	lucknow	2020	2133302340	133288152	16.0052
	Patna	2019	2622804340	171074064	15.3314
	lucknow	2019	1981143320	133288152	14.8636
	Patna	2024	2537859155	171074064	14.8349
	lucknow	2023	1949870715	133288152	14.6290
	bhopal	2022	2617113840	186824412	14.0084
	lucknow	2024	1850344925	133288152	13.8823
	lucknow	2021	1633194030	133288152	12.2531
	Patna	2023	2016590095	171074064	11.7878
	ranchi	2024	1797276160	156592296	11.4774

Explanation:

This query determines each city's annual ad revenue per net circulated copy. To combine the required data, it joins **dim_city**, **fact_print_sales**, and **fact_ad_revenue**. After extracting the year from the quarter column using a **CASE** statement, the query divides the **total ad revenue by the total net circulation** to determine the **revenue_per_copy ratio**. You can examine the trend over time because the results are arranged by year and to display the cities with the **highest ratio first**.

8. Digital Relaunch City Prioritization Based on digital readiness, pilot engagement, and print decline, which 3 cities should be prioritized for Phase 1 of the digital relaunch?

```
/* 8. Digital Relaunch City Prioritization
Based on digital readiness, pilot engagement, and print decline, which 3 cities should be
prioritized for Phase 1 of the digital relaunch? */

SELECT
    dc.city,
    AVG(fcr.literacy_rate + fcr.smartphone_penetration + fcr.internet_penetration) / 3 AS digital_readiness_score,
    SUM(fdp.users_reached + fdp.downloads_or_accesses) AS digital_engagement_score,
    SUM(fps.Net_Circulation) AS total_net_circulation
FROM dim_city dc
LEFT JOIN fact_city_readiness fcr ON dc.city_id = fcr.city_id
LEFT JOIN fact_digital_pilot fdp ON dc.city_id = fdp.city_id
LEFT JOIN fact_print_sales fps ON dc.city_id = fps.City_ID
GROUP BY dc.city
ORDER BY digital_readiness_score DESC, digital_engagement_score ASC, total_net_circulation ASC
LIMIT 3;
```

Result Grid				
	city	digital_readiness_score	digital_engagement_score	total_net_circulation
▶	kanpur	75.09874999999714	192058368	2415771600
	Varanasi	74.17875000000086	374113584	2646373056
	lucknow	73.49444444444669	322682880	1332881520

Explanation:

By examining **digital readiness, pilot engagement, and print decline**, this query ranks three cities for a digital relaunch.

Even if some data is missing, it combines information from `dim_city`, `fact_city_readiness`, `fact_digital_pilot`, and `fact_print_sales` for all cities using `LEFT JOIN`. After that, the query determines the `total_net_circulation`, `digital_engagement_score`, and `digital_readiness_score`. The cities with the best prospects for a digital pivot are those with the **highest readiness, lowest engagement, and lowest print circulation**, as determined by the `ORDER BY` clause. The top three of these cities are chosen by the `LIMIT 3` clause.

Further analysis & recommendations:

1. What should Bharat Herald's phased digital transition strategy look like, given the readiness and engagement data?

- **Launch of Prime**

The focus is on cities like **Jaipur, Bhopal, and Patna** that have a high level of digital readiness but low pilot engagement.

Convert a ready but untapped audience, launch a targeted marketing campaign and a full-scale digital relaunch.

- **Quicken & Profit**

Cities with a high level of preparedness and active pilot participation, such as **Delhi, Mumbai, and Ahmedabad**.

Strategy: To boost user engagement and income, enhance the ad platform, **add premium features**, and optimize the digital platform.

- **Establish a Base**

Digitally unprepared cities (**like Ranchi**).

To develop a future digital audience, start with foundational work like encouraging digital literacy and progressively releasing free, mobile-friendly content.

2. How can Bharat Herald regain advertiser trust in key cities or categories where confidence dropped the most?

- **Target Declining Cities & Categories Determine**

Identify the cities and ad categories that have seen the biggest drops.

Regain trust presents advertisers in these areas with a clear, data-driven plan for **recovery**, showing a deep understanding of their market.

- **Stress Digital Audience & Data**

Turn the discussion from just print circulation to the power of your print and digital audiences working together.

Show value provides a multi-platform solution for advertisers by showcasing your capacity to reach an **expanding, digitally engaged audience** using data from the **digital pilot**.

- **Provide Hybrid Advertising Models**

Create novel advertising packages that combine digital and print placements.

Reduce risk to restore trust, think about using performance-based pricing for **digital advertisements** that directly links payment to quantifiable outcomes.

3. What changes to content format or delivery (e.g., WhatsApp bulletins, mobile optimized e-papers) might boost digital engagement?

- **Content First for Mobile**

Write brief, scannable articles with rich **media embedded**. Making sure the e-papers load rapidly on mobile devices and are simple to use.

- **Community-Driven & Interactive Content**

Include surveys and **Q&A** sessions. Deliver breaking news via apps like **WhatsApp** to establish a direct line of communication with the audience.

- **Personalized & Local Content**

Provide city-specific news and events by using **geo-targeting**. To make the experience more relevant and individualized, suggest articles based on **user interests**.

4. What role can subscription bundling, loyalty programs, or pay-per-article models play in revenue recovery?

- **Subscription Bundling**

Provides a discounted print and **digital bundle** to help existing print subscribers make the switch.

- **Loyalty programs**

Increase traffic and retention by rewarding user engagement with a **points** system that can be exchanged for premium content.

- **Pay-per-article**

By offering users a **low-commitment** way to access premium content, **pay-per-article** generates a micro-revenue stream and opens up a new channel for prospective subscribers.

5. How can Bharat Herald leverage local influencers or journalists to re-establish digital credibility in regional markets?

- **Create a Local Contributor Network**

Collaborate with prominent local reporters and social media influencers

- **Their current reputation will help build the digital platform's credibility and draw in new readers.**

Organize **webinars** or online **Q&A** sessions with local journalists to discuss local news by hosting digital community events.

By doing this, they establish themselves as authorities and interact with the community directly.

- **Present Local Success Stories**

Highlight endorsements for the new digital platform from influencers. This increases trust in the brand's online presence and offers social proof.

Some Additional Analysis

1. Which cities or demographics show a strong positive correlation between declining print circulation and increasing digital pilot engagement?

```
-- Additional Analysis

-- 1. Which cities or demographics show a strong positive
-- correlation between declining print circulation and increasing digital pilot engagement? */

WITH PrintTrends AS (
    SELECT
        City_ID,
        SUM(Net_Circulation) AS total_net_circulation,
        LAG(SUM(Net_Circulation), 1, 0) OVER (PARTITION BY City_ID ORDER BY YEAR(Month)) AS previous_year_circulation,
        (SUM(Net_Circulation) - LAG(SUM(Net_Circulation), 1, 0) OVER (PARTITION BY City_ID ORDER BY YEAR(Month))) AS circulation_change
    FROM fact_print_sales
    GROUP BY City_ID, YEAR(Month)
),
DigitalEngagement AS (
    SELECT city_id,
        SUM(users_reached + downloads_or_accesses) AS digital_engagement_score
    FROM fact_digital_pilot
    GROUP BY city_id
)
SELECT d.city, pt.circulation_change AS print_circulation_change,
    de.digital_engagement_score
FROM PrintTrends pt
JOIN DigitalEngagement de ON pt.City_ID = de.city_id
JOIN dim_city d ON pt.City_ID = d.city_id
WHERE pt.circulation_change < 0
ORDER BY pt.circulation_change ASC, de.digital_engagement_score DESC;
```

	city	print_circulation_change	digital_engagement_score
▶	Mumbai	-962935	202080
	kanpur	-836038	125038
	Ahmedabad	-739319	217734
	jaipur	-641805	182748
	jaipur	-589455	182748
	Patna	-582600	184364
	bhopal	-565017	222737
	Varanasi	-560242	225914

Explanation :

The search finds cities where a rise in **digital engagement** is correlated with a **decrease in print circulation**. The results are filtered and arranged to highlight the cities with the **strongest print-to-digital migration** by combining data on **annual print changes and total digital pilot activity**. The cities with the steepest print declines and the highest digital engagement are highlighted.

Recommendation:

This analysis would identify markets where readers are already migrating from print to digital, allowing for targeted campaigns to accelerate the transition and capture a larger digital audience.

2. Is there a time lag between changes in print circulation and changes in ad revenue for specific cities or ad categories?

```

/* 2. Is there a time lag between changes in print circulation and changes in ad revenue for specific cities or ad categories? */

WITH YearlyMetrics AS (
    -- Calculate total print circulation and ad revenue per city, per year
    SELECT dc.city_id, YEAR(fpt.Month) AS year,
        SUM(fpt.Net_Circulation) AS total_net_circulation,
        SUM(CAST(far.ad_revenue AS DECIMAL)) AS total_ad_revenue
    FROM dim_city dc JOIN fact_print_sales fpt ON dc.city_id = fpt.City_ID
    JOIN fact_ad_revenue far ON fpt.edition_ID = far.edition_id
    GROUP BY dc.city_id, YEAR(fpt.Month)
),
LaggedMetrics AS (
    -- Calculate the year-over-year change for each metric and the lag
    SELECT city_id, year, total_net_circulation, total_ad_revenue,
        LAG(total_net_circulation, 1) OVER (PARTITION BY city_id ORDER BY year) AS prev_year_circulation,
        LAG(total_ad_revenue, 1) OVER (PARTITION BY city_id ORDER BY year) AS prev_year_ad_revenue
    FROM YearlyMetrics
)
SELECT dc.city, lm.year, lm.total_net_circulation, lm.total_ad_revenue, lm.prev_year_circulation,
    lm.prev_year_ad_revenue,
    CASE
        WHEN lm.prev_year_circulation IS NULL OR lm.prev_year_circulation = 0 THEN 'No Previous Data'
        ELSE (lm.total_net_circulation - lm.prev_year_circulation) / lm.prev_year_circulation
    END AS circulation_yoy_change,
    CASE
        WHEN lm.prev_year_ad_revenue IS NULL OR lm.prev_year_ad_revenue = 0 THEN 'No Previous Data'
        ELSE (lm.total_ad_revenue - lm.prev_year_ad_revenue) / lm.prev_year_ad_revenue
    END AS ad_revenue_yoy_change,
    CASE
        WHEN (lm.total_net_circulation < lm.prev_year_circulation AND lm.total_ad_revenue < lm.prev_year_ad_revenue)
        THEN 'Potential Lag: Both Declined'
        WHEN (lm.total_net_circulation < lm.prev_year_circulation AND lm.total_ad_revenue >= lm.prev_year_ad_revenue)
        THEN 'No Lag: Print Decline, Ad Revenue Stable/Grew'
        ELSE 'Other'
    END AS lag_analysis
FROM LaggedMetrics lm
JOIN dim_city dc ON lm.city_id = dc.city_id
ORDER BY dc.city, lm.year;

```

city	year	total_net_circulation	total_ad_revenue	prev_year_circulation	prev_year_ad_revenue	circulation_yoy_change	ad_revenue_yoy_change	lag_analysis
Ahmedabad	2019	239398704	1838338348	NULL	NULL	No Previous Data	No Previous Data	Other
Ahmedabad	2020	186167736	1504095012	239398704	1838338348	-0.2224	-0.1818	Potential Lag: Both Declin
Ahmedabad	2021	198181080	1671216680	186167736	1504095012	0.0645	0.1111	Other
Ahmedabad	2022	223912008	2005460016	198181080	1671216680	0.1298	0.2000	Other
Ahmedabad	2023	191616912	1838338348	223912008	2005460016	-0.1442	-0.0833	Potential Lag: Both Declin
Ahmedabad	2024	181304640	1838338348	191616912	1838338348	-0.0538	0.0000	No Lag: Print Decline, Ad
bhopal	2019	215016768	1940947019	NULL	NULL	No Previous Data	No Previous Data	Other

Explanation:

This query examines whether a city's **print circulation and ad revenue** are **correlated** over time. The annual totals for each city's **net_circulation** and **ad_revenue** are first determined. It then determines the **year-over-year** change by comparing each year's data to the one from the year before using a window function (**LAG**). The final output then determines whether there may be a lag if a **drop in print circulation in one year** is followed by a **drop in ad revenue** the following year.

Recommendation:

By identifying a **lag**, the company can better predict **future ad revenue declines** and proactively develop strategies to mitigate losses, rather than simply reacting to them.

3. Do the most profitable cities (based on ad revenue per net circulated copy) align with the cities that have the highest digital readiness scores?

```
/* 3. Do the most profitable cities (based on ad revenue per net circulated copy) align with the
   cities that have the highest digital readiness scores? */

WITH Profitability AS (
    SELECT f.City_ID,
           SUM(CAST(a.ad_revenue AS DECIMAL)) / SUM(f.Net_Circulation) AS profitability_ratio
    FROM fact_print_sales f JOIN fact_ad_revenue a ON f.edition_ID = a.edition_id
    GROUP BY f.City_ID
),
DigitalReadiness AS (
    SELECT City_ID,
           AVG(literacy_rate + smartphone_penetration + internet_penetration) / 3 AS digital_readiness_score
    FROM fact_city_readiness
    GROUP BY City_ID
)
SELECT d.city, p.profitability_ratio, r.digital_readiness_score,
       RANK() OVER (ORDER BY p.profitability_ratio DESC) AS profitability_rank,
       RANK() OVER (ORDER BY r.digital_readiness_score DESC) AS readiness_rank
FROM Profitability p JOIN DigitalReadiness r ON p.City_ID = r.City_ID
JOIN dim_city d ON p.City_ID = d.city_id
ORDER BY profitability_rank, readiness_rank;
```

Result Grid					
	city	profitability_ratio	digital_readiness_score	profitability_rank	readiness_rank
▶	lucknow	13.7482	73.49444444444445	1	3
	Patna	12.1591	70.52680555555555	2	6
	ranchi	10.5541	68.63874999999999	3	7
	bhopal	10.3892	73.28777777777778	4	4
	Ahmedabad	8.7629	72.735	5	5
	Delhi	7.3644	56.04708333333326	6	9
	kanpur	6.6924	75.09875	7	1
	Mumbai	5.9478	68.32500000000002	8	8
	jaipur	4.7165	55.01652777777775	9	10
	Varanasi	4.6963	74.17875000000001	10	2

Explanation:

The two important metrics are first computed in different **Common Table Expressions (CTEs)** by this query. After joining these results, it ranks each city according to its **profitability and digital readiness scores** using the **RANK()** window function. You can use the output to compare the ranks directly and determine whether the most profitable cities—those with a profitability rank of **1, 2, 3, etc.** are also the ones with the highest levels of **digital readiness**.

Recommendation:

This would help the company understand whether their most lucrative markets are also the most prepared for a digital transition. This information is crucial for deciding whether to **push for a rapid digital shift** in these profitable markets or to maintain the print focus for as long as possible.

Conclusion:

Based on a comprehensive analysis of the data, the optimal path forward for Bharat Herald is a well-defined and targeted digital transition strategy. Rather than a one-size-fits-all approach, success hinges on segmenting the markets and tailoring the strategy to each one's unique characteristics.

Strategic Rationale & Market Segmentation

The core of the strategy is to categorize cities into three distinct groups based on their **digital readiness** and **pilot engagement**. This allows Bharat Herald to allocate resources effectively for maximum impact.

- **High-Potential Growth Markets:** These are the most promising cities for a swift and successful digital pivot. They possess high digital readiness (high literacy, smartphone, and internet penetration) but have shown low engagement with the initial pilot. This indicates a large, untapped audience that is prepared for a digital product, but has not yet been given a compelling reason to adopt it. The strategy here is a full-scale digital relaunch to capture this market.
- **Optimization & Monetization Markets:** These cities are already receptive to digital content, with both high readiness and strong pilot engagement. The focus here is not on initial adoption but on **deepening engagement and maximizing revenue**. The company should prioritize these markets for new features, premium content, and optimized ad models to increase the lifetime value of their existing digital customers.
- **Foundational Markets:** These cities exhibit low digital readiness. A direct, full-scale digital launch would likely fail. The strategy here is a long-term investment in building the necessary infrastructure and consumer habits. The goal is to lay the groundwork for a future digital audience by first focusing on digital literacy and gradual content introduction.

