

MAVENFLIX STREAMING VIDEO SUBSCRIPTION ANALYSIS USING SQL

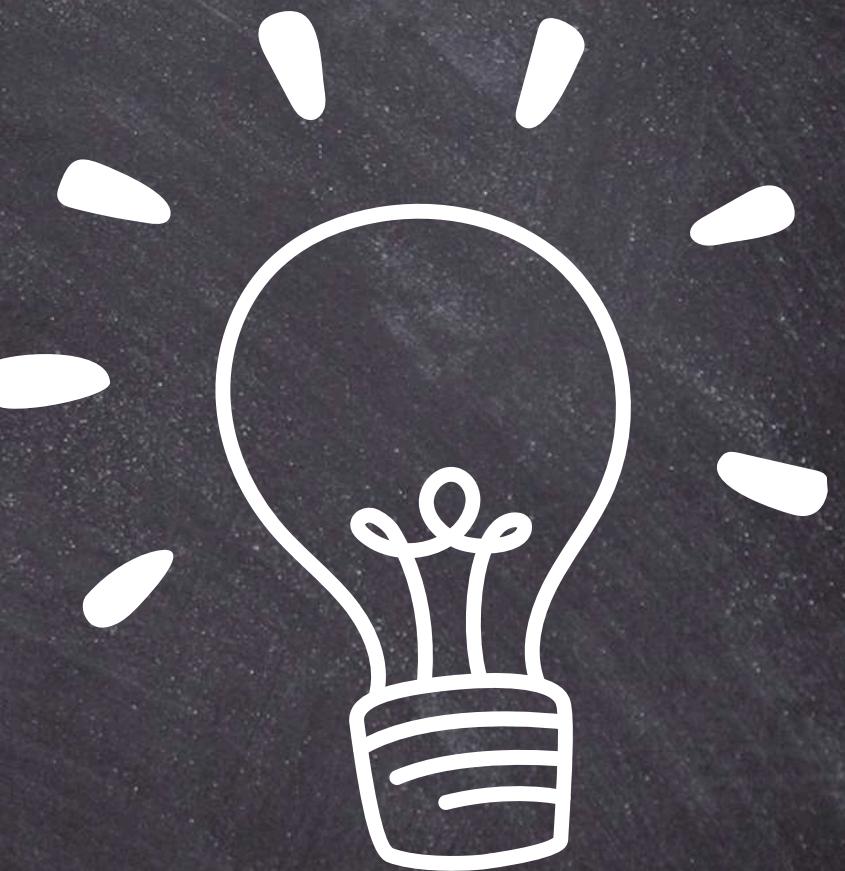


BY - SHUBHAM KUMAR

ABOUT DATA



SUBSCRIPTION RECORDS FOR MAVENFLIX, A FICTITIOUS VIDEO STREAMING PLATFORM. DATASET INCLUDES INFORMATION ABOUT ~2,800 SUBSCRIBERS FROM SEPTEMBER 2022 THROUGH SEPTEMBER 2023. EACH RECORD REPRESENTS AN INDIVIDUAL CUSTOMER'S SUBSCRIPTION, INCLUDING THE SUBSCRIPTION COST, CREATED/CANCELED DATE, INTERVAL, AND PAYMENT STATUS.



DATA PREPARATION & CLEANING

STEPS INVOLVED:-

- IMPORT CSV FILE INTO MYSQL WORKBENCH.
- CREATE SCHEMA AND TABLE FOR THE DATASET.
- AFTER IMPORTING DATASET CHECK THE DATA TYPES OF ALL COLUMNS AND CONVERTED TO RIGHT DATA TYPE.
- USING ALTER COMMAND TO ADD NEW COLUMNS IN TO THE TABLE.

=> CREATE NEW_CREATED_DATE NAME COLUMN USING ALTER

ALTER TABLE SUBSCRIPTION ADD COLUMN NEW_CREATED_DATE DATE;

DATA PREPARATION & CLEANING

=> CREATE NEW_CANCELED_DATE NAME COLUMN USING ALTER

ALTER TABLE SUBSCRIPTION ADD COLUMN NEW_CANCELED_DATE DATE;

=> UPDATE THE NEW COLUMN WITH THE CORRECT DATA TYPE

-- FOR NEW_CREATED_DATE:

UPDATE SUBSCRIPTION

SET

NEW_CREATED_DATE = STR_TO_DATE(CREATED_DATE, '%D-%M-%Y');

DATA PREPARATION & CLEANING

=> UPDATE THE NEW COLUMN WITH THE CORRECT DATA TYPE

-- FOR NEW_CANCELED_DATE:

```
UPDATE SUBSCRIPTION
SET
    NEW_CANCELED_DATE = CASE
        WHEN
            CANCELED_DATE IS NOT NULL
            AND CANCELED_DATE != ""
        THEN
            STR_TO_DATE(CANCELED_DATE, '%D-%M-%Y')
        ELSE NULL
    END;
```

DATA PREPARATION & CLEANING

=> DROP ORIGINAL CREATED_DATE COLUMN FROM TABLE

ALTER TABLE SUBSCRIPTION DROP COLUMN CREATED_DATE

=> DROP ORIGINAL CANCELED_DATE COLUMN FROM TABLE

ALTER TABLE SUBSCRIPTION DROP COLUMN CANCELED_DATE

PROBLEM STATEMENTS

IDENTIFY THE MAVENFLIX CUSTOMERS SUBSCRIPTION TRENDS AND PATTERNS.

- TOTAL PAID SUBSCRIPTION
- TOTAL UNPAID SUBSCRIPTION
- TOTAL CUSTOMER PAID SUBSCRIPTION
- SUBSCRIPTIONS TRENDS OVER TIME
- PERCENTAGE OF CUSTOMERS WITH 5 MONTH OR MORE SUBSCRIPTIONS
- WHICH MONTH HAD THE HIGHEST AND LOWEST RETENTION SUBSCRIBER RETENTION
- NUMBERS OF ACTIVE PAYING SUBSCRIBERS
- TOTAL AMOUNT GENERATED FROM PAID SUBSCRIBERS

TOTAL PAID SUBSCRIPTION

```
47
48      -- • Total Paid Subscription
49
50 • SELECT
51     COUNT(was_subscription_paid)
52     FROM
53     subscription
54     WHERE
55     was_subscription_paid = 'Yes'
56
```

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

count(was_subscription_paid)
▶ 2936

TOTAL PAID SUBSCRIPTION WAS 2936

TOTAL UNPAID SUBSCRIPTION

```
63
64  -- • Total Unpaid Subscription
65
66 • SELECT
67      COUNT(was_subscription_paid)
68  FROM
69      subscription
70  WHERE
71      was_subscription_paid = 'No'
72
73
```

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

count(was_subscription_paid)
133

TOTAL UNPAID SUBSCRIPTION WAS 133

TOTAL CUSTOMER PAID SUBSCRIPTION

```
75
76 -- • Total Customer Paid Subscription
77
78 SELECT COUNT(DISTINCT customer_id) as paid_subscription_customer
79 FROM subscription
80 WHERE was_subscription_paid = 'Yes'
81
82
83
84
85
86
```

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

paid_subscription_customer
2744

THE TOTAL DISTINCT CUSTOMER PAID SUBSCRIPTION IS 2744.

SUBSCRIPTIONS TREND OVER TIME

```
85
86  -- • Subscriptions trends over time
87
88  SELECT
89      DATE_FORMAT(new_created_date, '%Y-%m') AS monthly_trend,
90      COUNT(DISTINCT customer_id) AS subscribers
91  FROM
92      subscription
93  WHERE
94      was_subscription_paid = 'Yes'
95  GROUP BY 1
96
97
98
99
100
```

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

monthly_trend	subscribers
2022-09	206
2022-10	189
2022-11	176
2022-12	217
2023-01	254
2023-02	265
2023-03	248
2023-04	207

Result 14

PERCENTAGE OF CUSTOMERS WITH 5 MONTH OR MORE SUBSCRIPTIONS

```
98
99  -- • Percentage of customers with 5 month or more subscriptions
100
101 with cte as (
102     select customer_id,
103         IFNULL(DATEDIFF(new_canceled_date, new_created_date), 0)/30 as total_months_subscription
104     from subscription )
105
106 select
107     round((count(case when total_months_subscription >=5 then customer_id end) * 100.0 / count(distinct customer_id)),2) as 5_months_subscriptions_percent
108 from cte
109
110
```

Result Grid	Filter Rows:	Export:	Wrap Cell Content:
5_months_subscriptions_percent			5.60

THE PERCENTAGE OF CUSTOMER WITH 5 MONTHS OR MORE SUBSCRIPTIONS IS 5.60%

WHICH MONTH HAD THE HIGHEST AND LOWEST RETENTION SUBSCRIBER RETENTION

```
113
114  -- • Which month had the highest and lowest retention subscriber retention
115
116  with monthly_subs as (
117      SELECT count(customer_id) as new_customer,
118          date_format(new_created_date, '%Y-%m') as subs_month
119      from subscription
120      group by 2
121  ),
122  monthly_cancl as (
123      SELECT count(customer_id) as cancl_customer,
124          date_format(new_canceled_date, '%Y-%m') as cancl_month
125      from subscription
126      where new_canceled_date is not null
127      group by 2
128  ),
```

```
131
132  monthly_retention as (
133      select subs_month,
134          new_customer,
135          coalesce(cancl_customer, 0) as cancl_customer,
136          (new_customer - coalesce(cancl_customer, 0)) / new_customer * 100 as retention_rate
137      from monthly_subs
138      left join monthly_cancl
139      on subs_month = cancl_month
140  )
141
142  select subs_month, new_customer, cancl_customer, round(retention_rate,2)
143  from monthly_retention
144  order by retention_rate desc
145
146 ---
```

	subs_month	new_customer	cancl_customer	round(retention_rate,2)
▶	2023-07	306	207	3.05
	2023-05	285	215	2.84
	2023-02	275	171	2.74
	2023-06	275	225	2.74
	2023-01	273	149	2.72
	2023-08	268	219	2.67
	2023-03	266	224	2.65
	2022-12	230	123	2.29
	2023-04	223	194	2.22
	2022-09	217	33	2.17
	2022-10	196	81	1.96
	2022-11	100	100	1.00

NUMBER OF ACTIVE PAYING SUBSCRIBER

```
120
121
122
123     • Numbers of Active Paying Subscribers
124
125     SELECT COUNT(customer_id) AS active_subscriber_count
126     FROM subscription
127     WHERE new_canceled_date IS NULL AND was_subscription_paid = 'Yes'
128
129
130
131
```

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

active_subscriber_count
1055

THERE ARE TOTAL OF 1055 ACTIVE PAYING SUBSCRIBERS.

TOTAL AMOUNT GENERATED FROM PAID SUBSCRIBERS

```
130
131      -- • Total Amount Generated from Paid Subscribers
132
133      SELECT SUM(subscription_cost) AS paid_subscriber_generated_amount
134      FROM subscription
135      WHERE was_subscription_paid = 'Yes'
136
137
138
139
140
141
```

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

paid_subscriber_generated_amount
114504

THE TOTAL AMOUNT GENERATED FROM PAID SUBSCRIBERS IS 114,504.

CONCLUSION:

- THE TOTAL NUMBER OF PAID SUBSCRIPTIONS ARE 2936.
- THE TOTAL UNPAID SUBSCRIPTIONS ARE 133.
- THE TOTAL DISTINCT CUSTOMER PAID SUBSCRIPTIONS ARE 2744.
- THE PERCENTAGE OF CUSTOMER WITH 5 MONTHS OR MORE SUBSCRIPTIONS IS 5.60%.
- THE HIGHEST RETENTION RATE IS IN THE MONTH OF SEPTEMBER IN 2022 "2022-09" WHILE "2023-04" HAD THE LOWEST RETENTION RATE.
- THERE ARE TOTAL OF 1055 ACTIVE PAYING SUBSCRIBERS.
- THE TOTAL AMOUNT GENERATED FROM PAID SUBSCRIBERS ARE 114,504.