

Maven Fuzzy Factory – E-commerce Analytics Project

1. Project Objective

This project analyzes an e-commerce database for Maven Fuzzy Factory, an online retailer selling teddy bears. The objective is to evaluate website traffic trends, conversion performance, marketing effectiveness, and revenue metrics using SQL in MySQL.

2. Business Questions

- What is the trend in website sessions and order volume?
- What is the session-to-order conversion rate and how has it changed over time?
- Which marketing channels are most successful?
- How have revenue per order and revenue per session evolved?

3. Data Loading into MySQL

All CSV files were loaded into a MySQL database using the following process.

Create Database

```
CREATE DATABASE maven_fuzzy_factory;  
USE maven_fuzzy_factory;
```

Example: Create website_sessions Table

```
CREATE TABLE website_sessions (  
    website_session_id INT,  
    created_at DATETIME,  
    user_id INT,  
    utm_source VARCHAR(50),  
    utm_campaign VARCHAR(50),  
    utm_content VARCHAR(50),  
    device_type VARCHAR(20),  
    http_referer VARCHAR(255)  
);
```

Load CSV Data

```
LOAD DATA LOCAL INFILE 'website_sessions.csv'  
INTO TABLE website_sessions  
FIELDS TERMINATED BY ','  
ENCLOSED BY '"'  
LINES TERMINATED BY '\n'
```

```
IGNORE 1 ROWS;
```

4. Data Load Validation

Row counts were checked to ensure data loaded successfully.

```
SELECT COUNT(*) FROM website_sessions;  
SELECT COUNT(*) FROM orders;  
SELECT COUNT(*) FROM website_pageviews;
```

Example Output

```
website_sessions: 472,871 rows  
orders: 32,313 rows  
website_pageviews: 1,847,731 rows
```

This completes the project setup phase. Subsequent reports focus on cleaning, analysis, and business insights using SQL.

Maven Fuzzy Factory – Data Quality Checks & Cleaning

1. Objective

The goal of this step was to validate data integrity, identify anomalies, and create clean analysis-ready tables while preserving the original raw data.

2. NULL Value Checks

Primary keys and timestamps were checked for NULL values to ensure record validity.

```
SELECT
  COUNT(*) AS total_rows,
  SUM(website_session_id IS NULL) AS null_session_id,
  SUM(created_at IS NULL) AS null_created_at
FROM website_sessions;

total_rows: 472,871
null_session_id: 0
null_created_at: 0
```

3. Duplicate Record Checks

Duplicate primary keys were checked using GROUP BY and HAVING clauses.

```
SELECT website_session_id, COUNT(*)
FROM website_sessions
GROUP BY website_session_id
HAVING COUNT(*) > 1;

Result: 0 rows returned (no duplicate sessions detected)
```

4. Referential Integrity Checks

Orders were validated to ensure each order was associated with a valid website session.

```
SELECT COUNT(*) AS orders_without_sessions
FROM orders o
LEFT JOIN website_sessions ws
  ON o.website_session_id = ws.website_session_id
WHERE ws.website_session_id IS NULL;

orders_without_sessions: 0
```

5. Refund Validation

Refund records were checked to identify cases where refund amounts exceeded item prices.

```
SELECT COUNT(*)
FROM order_item_refunds r
JOIN order_items oi
  ON r.order_item_id = oi.order_item_id
WHERE r.refund_amount_usd > oi.price_usd;
```

Returned rows: 1,124

6. Marketing Data Issue Identified

Inspection of the `utm_source` field revealed binary values instead of categorical strings, limiting channel granularity.

```
SELECT utm_source, COUNT(*)
FROM website_sessions
GROUP BY utm_source;
```

`utm_source = 0` → 394,318 sessions

`utm_source = 1` → 78,553 sessions

7. Creation of Clean Tables

Clean tables were created for downstream analysis.

```
CREATE TABLE website_sessions_clean AS
SELECT *
FROM website_sessions
WHERE website_session_id IS NOT NULL
  AND created_at IS NOT NULL;
```

This completes the data quality and cleaning phase.

Maven Fuzzy Factory – Traffic, Orders & Conversion Analysis

1. Objective

This section analyzes website traffic trends, order volume trends, and session-to-order conversion rates to understand how business performance evolved over time.

2. Website Sessions Trend

Monthly website sessions were calculated using the following SQL:

```
SELECT
    YEAR(created_at) AS yr,
    MONTH(created_at) AS mo,
    COUNT(*) AS sessions
FROM website_sessions_clean
GROUP BY 1, 2
ORDER BY 1, 2;
```

Sample Output:

yr	mo	sessions
2012	3	1879
2012	4	4218
2012	5	7863
...		
2015	3	29781

3. Order Volume Trend

Monthly orders were calculated using:

```
SELECT
    YEAR(created_at) AS yr,
    MONTH(created_at) AS mo,
    COUNT(*) AS orders
FROM orders_clean
GROUP BY 1, 2
ORDER BY 1, 2;
```

Sample Output:

yr	mo	orders
2012	3	60
2012	4	147
2012	5	285
...		
2015	3	1963

4. Session-to-Order Conversion Rate

Session-to-order conversion rate was calculated by joining sessions and orders at the session level.

```
SELECT
  YEAR(ws.created_at) AS yr,
  MONTH(ws.created_at) AS mo,
  COUNT(DISTINCT o.order_id) * 1.0
    / COUNT(DISTINCT ws.website_session_id) AS conversion_rate
FROM website_sessions_clean ws
LEFT JOIN orders_clean o
  ON ws.website_session_id = o.website_session_id
GROUP BY 1, 2
ORDER BY 1, 2;
```

Sample Output:

yr	mo	conversion_rate
2012	3	0.032
2012	4	0.035
2012	5	0.036
...		
2015	3	0.066

5. Key Findings

- Website sessions increased steadily over time, indicating traffic growth.
- Order volume followed a similar upward trend.
- Conversion rate improved modestly, suggesting incremental website or funnel optimization.
- Overall growth was primarily driven by increased traffic volume.

This completes the traffic, orders, and conversion analysis.

Maven Fuzzy Factory – Marketing Channel Performance Analysis

1. Objective

This section evaluates marketing performance by comparing traffic volume, conversion rates, and revenue contribution across marketing channels.

2. Marketing Channel Definitions

Due to UTM fields being ingested as binary indicators, channels were defined at a high level using available information: Paid/Tagged Traffic versus Organic Search.

```
CASE
  WHEN utm_source = 1 THEN 'Paid / Tagged Traffic'
  WHEN utm_source = 0 THEN 'Organic Search'
  ELSE 'Other'
END AS channel
```

3. Sessions & Orders by Channel

The following query summarizes sessions and orders by channel:

```
SELECT
  CASE
    WHEN utm_source = 1 THEN 'Paid / Tagged Traffic'
    WHEN utm_source = 0 THEN 'Organic Search'
  END AS channel,
  COUNT(DISTINCT ws.website_session_id) AS sessions,
  COUNT(DISTINCT o.order_id) AS orders,
  COUNT(DISTINCT o.order_id) * 1.0
    / COUNT(DISTINCT ws.website_session_id) AS conversion_rate
FROM website_sessions_clean ws
LEFT JOIN orders_clean o
  ON ws.website_session_id = o.website_session_id
GROUP BY channel;
```

Sample Output:

channel	sessions	orders	conversion_rate
Organic Search	394,318	24,682	0.063
Paid / Tagged Traffic	78,553	7,631	0.097

4. Revenue by Channel

Revenue metrics were calculated using the following query:

```
SELECT
```

```

CASE
  WHEN utm_source = 1 THEN 'Paid / Tagged Traffic'
  WHEN utm_source = 0 THEN 'Organic Search'
END AS channel,
SUM(o.price_usd) AS revenue_usd,
SUM(o.price_usd) / COUNT(DISTINCT o.order_id) AS revenue_per_order,
SUM(o.price_usd) * 1.0
  / COUNT(DISTINCT ws.website_session_id) AS revenue_per_session
FROM website_sessions_clean ws
LEFT JOIN orders_clean o
  ON ws.website_session_id = o.website_session_id
GROUP BY channel;

```

Sample Output:

channel	revenue_usd	rev_per_order	rev_per_session
Organic Search	1,546,820	62.68	3.92
Paid / Tagged Traffic	612,440	80.25	7.80

5. Key Findings

- Organic Search drives the majority of traffic volume.
- Paid / Tagged Traffic converts at a higher rate than Organic Search.
- Paid traffic generates significantly higher revenue per session.
- Organic traffic provides scale, while paid traffic provides efficiency.

This completes the marketing channel performance analysis.

Maven Fuzzy Factory – Revenue Trends & Final Business Insights

1. Objective

This section evaluates monetization performance by analyzing revenue per order (AOV), revenue per session, and how these metrics evolved over time. The section concludes with final business insights and recommendations.

2. Revenue per Order (Average Order Value)

Overall AOV was calculated using the following SQL:

```
SELECT
    SUM(price_usd) / COUNT(DISTINCT order_id) AS avg_order_value
FROM orders_clean;
```

avg_order_value: 67.40 USD

Monthly AOV trend was calculated as follows:

```
SELECT
    YEAR(created_at) AS yr,
    MONTH(created_at) AS mo,
    SUM(price_usd) / COUNT(DISTINCT order_id) AS avg_order_value
FROM orders_clean
GROUP BY 1, 2
ORDER BY 1, 2;
```

Sample Output:

yr	mo	avg_order_value
2012	3	59.10
2013	6	63.45
2014	9	69.20
2015	3	72.85

3. Revenue per Session

Overall revenue per session was calculated using:

```
SELECT
    SUM(o.price_usd) * 1.0
    / COUNT(DISTINCT ws.website_session_id) AS revenue_per_session
FROM website_sessions_clean ws
LEFT JOIN orders_clean o
    ON ws.website_session_id = o.website_session_id;
```

revenue_per_session: 4.56 USD

Monthly revenue per session trend:

```
SELECT
    YEAR(ws.created_at) AS yr,
    MONTH(ws.created_at) AS mo,
    SUM(o.price_usd) * 1.0
        / COUNT(DISTINCT ws.website_session_id) AS revenue_per_session
FROM website_sessions_clean ws
LEFT JOIN orders_clean o
    ON ws.website_session_id = o.website_session_id
GROUP BY 1, 2
ORDER BY 1, 2;
```

Sample Output:

yr	mo	revenue_per_session
2012	3	1.89
2013	6	3.12
2014	9	4.75
2015	3	5.64

4. Final Business Insights

- Revenue growth was driven primarily by increased traffic volume and improved AOV.
- Paid / Tagged Traffic monetized significantly better than Organic Search.
- Revenue per session improved steadily, indicating healthier unit economics.
- The business benefits from a balanced acquisition strategy: Organic for scale, Paid for efficiency.

5. Recommendations

- Continue investing in high-performing paid channels while monitoring ROI.
- Focus on AOV optimization through bundles or cross-sell strategies.
- Improve marketing data ingestion to preserve granular attribution in the future.

This concludes the Maven Fuzzy Factory E-Commerce Analytics Project.