Data Analysis

By Prayas Botake

BUSINESS PROBLEM STATEMENT

The company wants to analyze sales performance, customer purchasing behavior, refund impact, and website funnel efficiency to increase revenue, reduce refunds, and improve conversion rate.

BUSINESS OBJECTIVES

Track overall revenue and growth trends

Understand refund impact on profitability

Identify high-performing and high-risk products

Analyze website conversion funnel (Sessions → Orders)

Evaluate traffic quality across marketing channels

Tools & Technologies Used

MySQL – Data cleaning, transformation, analytics queries

Power BI – Data modeling, DAX measures, dashboards

SQL Views – Reusable analytics logic

CSV Files – Source data format

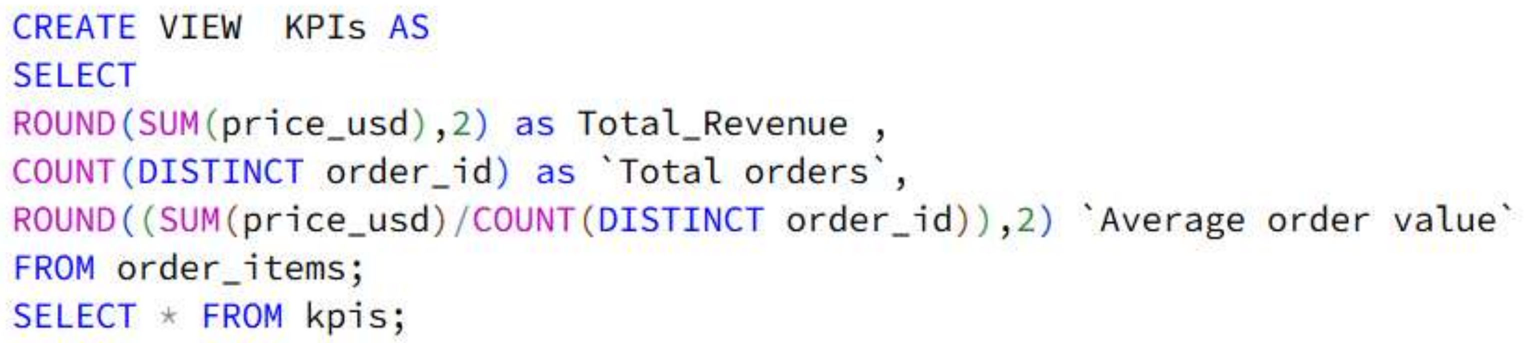
📁 DATASET DESCRIPTION

The dataset consists of **6 relational tables**:

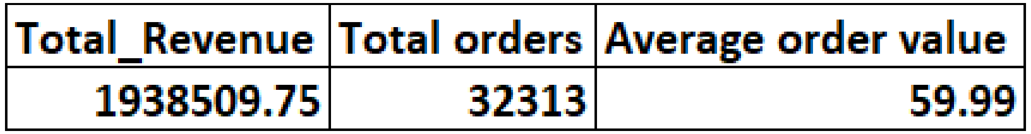
Table Name	**Description**
Orders	- Order-level transaction data
Order_items	- Product-level order details and revenue
order_item_refunds	- Refunded order itemsproducts Product master data
products	- Product master data
website_sessions	- Website visit/session data
website_pageviews	- Page-level user behavior

SQL Queries

1. What is total revenue, total orders, and AOV

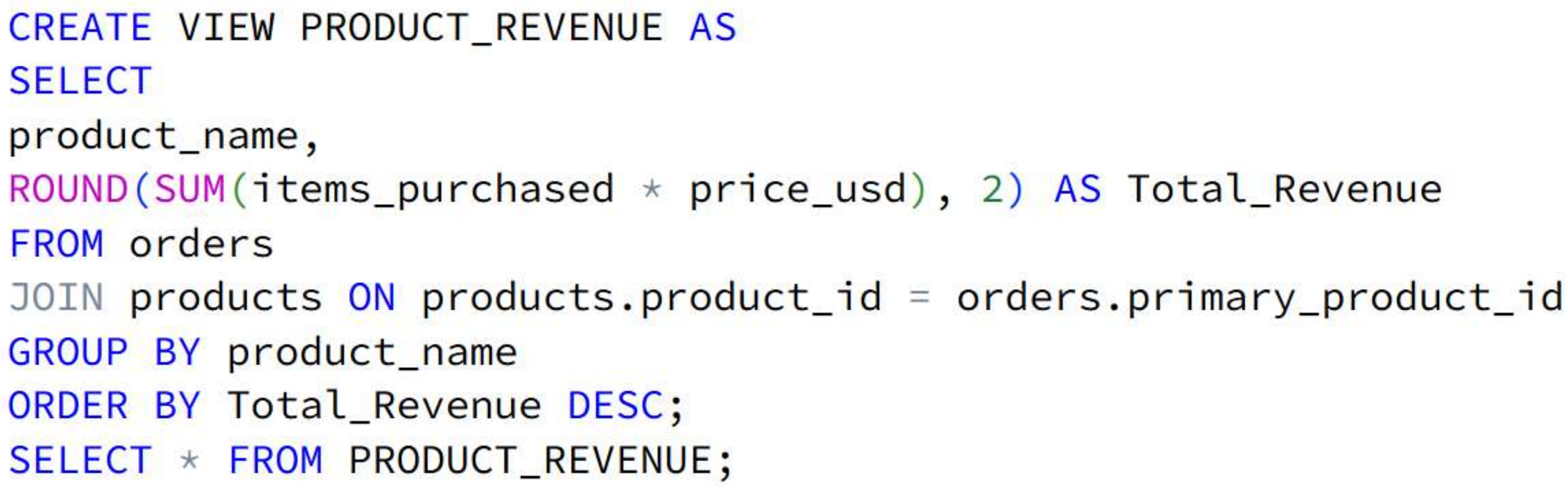


```
CREATE VIEW  KPIs AS
SELECT
ROUND(SUM(price_usd),2) as Total_Revenue ,
COUNT(DISTINCT order_id) as `Total orders`,
ROUND((SUM(price_usd)/COUNT(DISTINCT order_id)),2) `Average order value`
FROM order_items;
SELECT * FROM kpis;
```

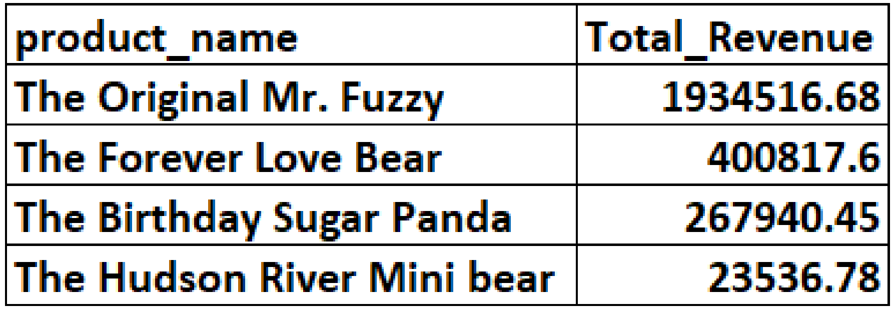


Total_Revenue	Total orders	Average order value
1938509.75	32313	59.99

1. Which products generate highest revenue

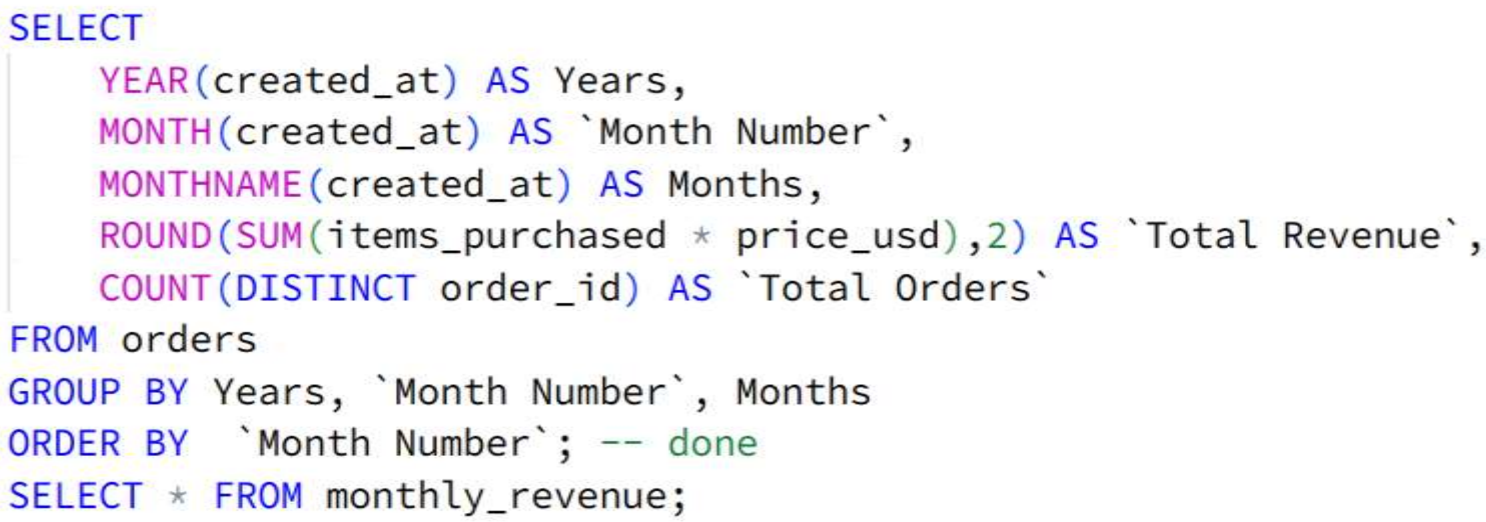


```sql
CREATE VIEW PRODUCT_REVENUE AS
SELECT
product_name,
ROUND(SUM(items_purchased * price_usd), 2) AS Total_Revenue
FROM orders
JOIN products ON products.product_id = orders.primary_product_id
GROUP BY product_name
ORDER BY Total_Revenue DESC;
SELECT * FROM PRODUCT_REVENUE;
```

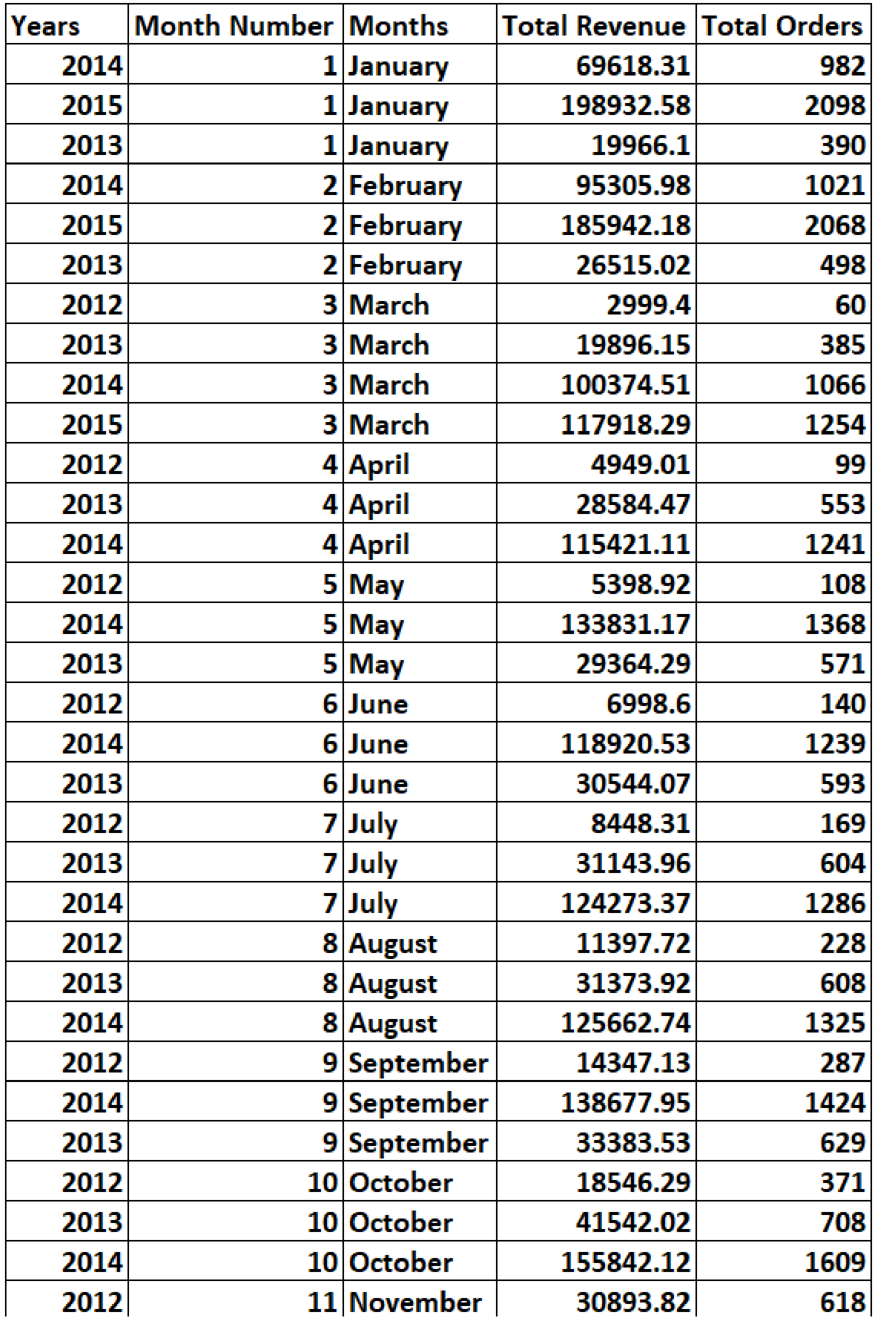


product_name	Total_Revenue
The Original Mr. Fuzzy	1934516.68
The Forever Love Bear	400817.6
The Birthday Sugar Panda	267940.45
The Hudson River Mini bear	23536.78

3. Monthly revenue & order trend.

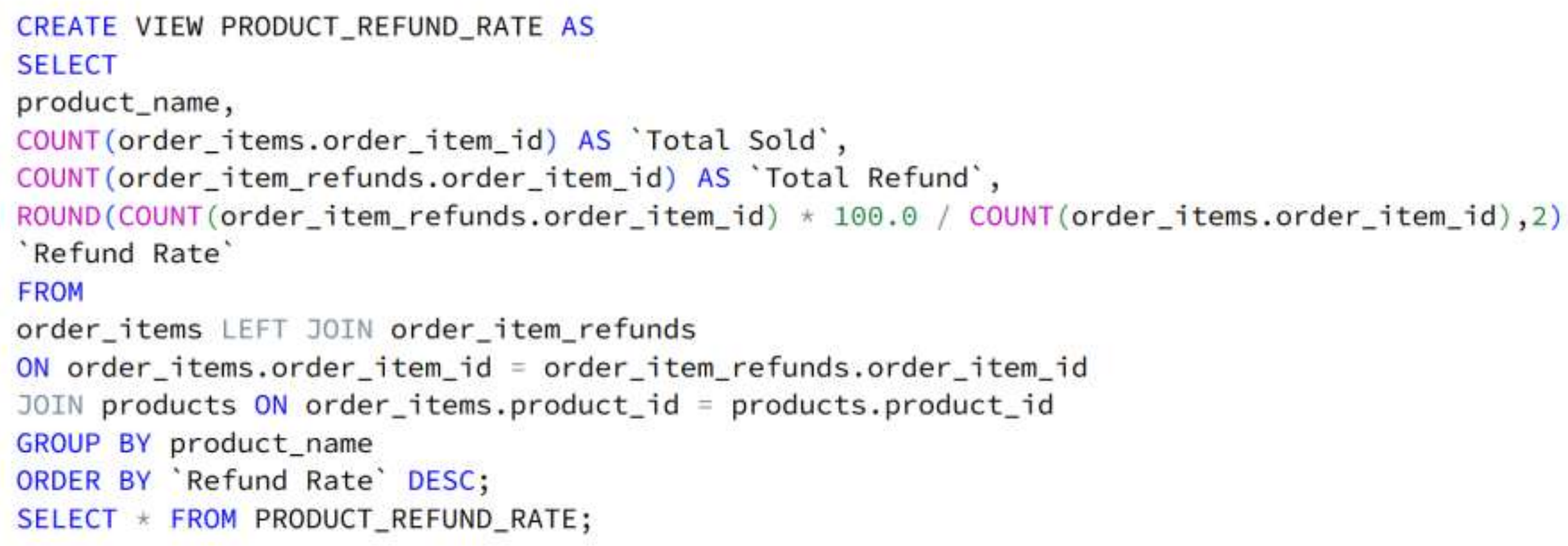


```
SELECT
    YEAR(created_at) AS Years,
    MONTH(created_at) AS `Month Number`,
    MONTHNAME(created_at) AS Months,
    ROUND(SUM(items_purchased * price_usd),2) AS `Total Revenue`,
    COUNT(DISTINCT order_id) AS `Total Orders`
FROM orders
GROUP BY Years, `Month Number`, Months
ORDER BY  `Month Number`; -- done
SELECT * FROM monthly_revenue;
```

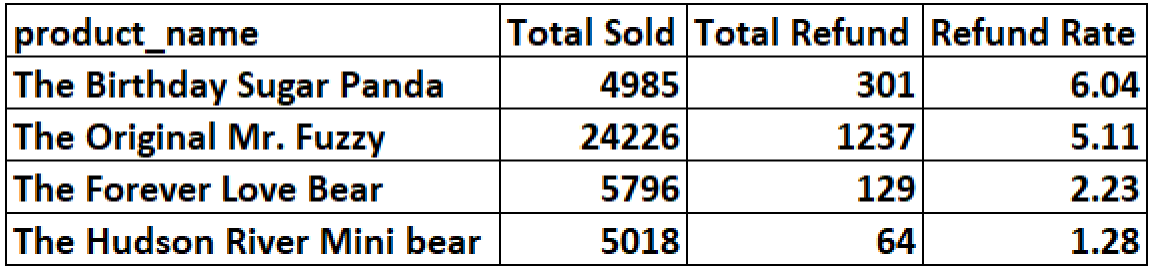


Years	Month Number	Months	Total Revenue	Total Orders
2014	1	January	69618.31	982
2015	1	January	198932.58	2098
2013	1	January	19966.1	390
2014	2	February	95305.98	1021
2015	2	February	185942.18	2068
2013	2	February	26515.02	498
2012	3	March	2999.4	60
2013	3	March	19896.15	385
2014	3	March	100374.51	1066
2015	3	March	117918.29	1254
2012	4	April	4949.01	99
2013	4	April	28584.47	553
2014	4	April	115421.11	1241
2012	5	May	5398.92	108
2014	5	May	133831.17	1368
2013	5	May	29364.29	571
2012	6	June	6998.6	140
2014	6	June	118920.53	1239
2013	6	June	30544.07	593
2012	7	July	8448.31	169
2013	7	July	31143.96	604
2014	7	July	124273.37	1286
2012	8	August	11397.72	228
2013	8	August	31373.92	608
2014	8	August	125662.74	1325
2012	9	September	14347.13	287
2014	9	September	138677.95	1424
2013	9	September	33383.53	629
2012	10	October	18546.29	371
2013	10	October	41542.02	708
2014	10	October	155842.12	1609
2012	11	November	30893.82	618

4. Which products have highest refund rate?

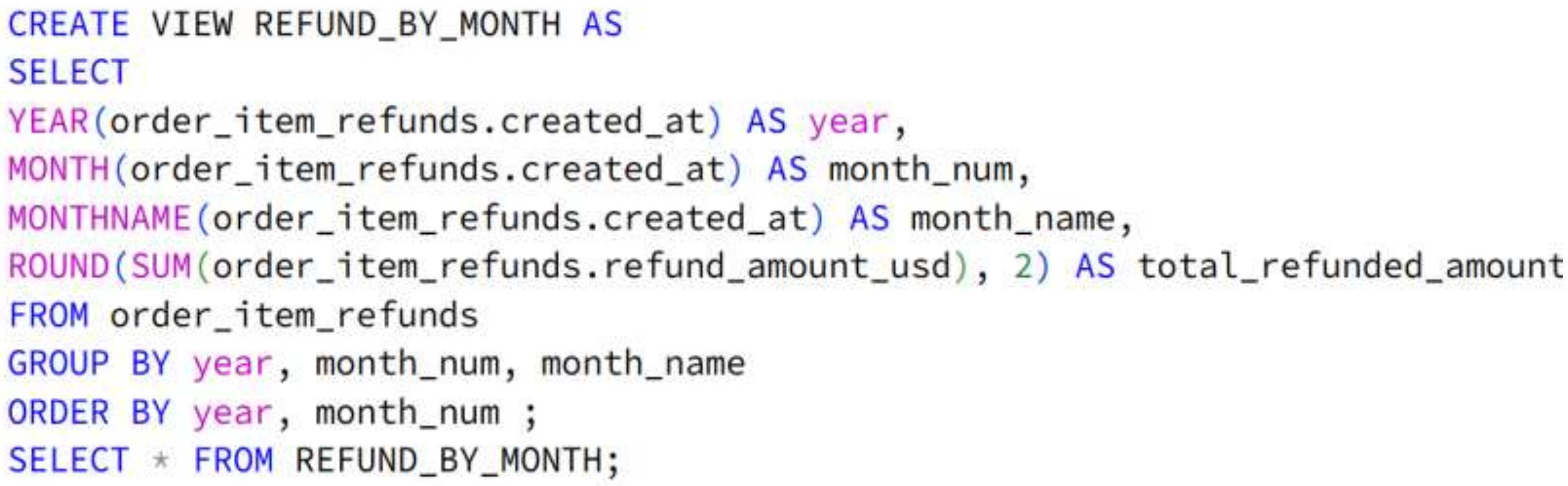


```
CREATE VIEW PRODUCT_REFUND_RATE AS
SELECT
product_name,
COUNT(order_items.order_item_id) AS `Total Sold`,
COUNT(order_item_refunds.order_item_id) AS `Total Refund`,
ROUND(COUNT(order_item_refunds.order_item_id) * 100.0 / COUNT(order_items.order_item_id),2)
`Refund Rate`
FROM
order_items LEFT JOIN order_item_refunds
ON order_items.order_item_id = order_item_refunds.order_item_id
JOIN products ON order_items.product_id = products.product_id
GROUP BY product_name
ORDER BY `Refund Rate` DESC;
SELECT * FROM PRODUCT_REFUND_RATE;
```

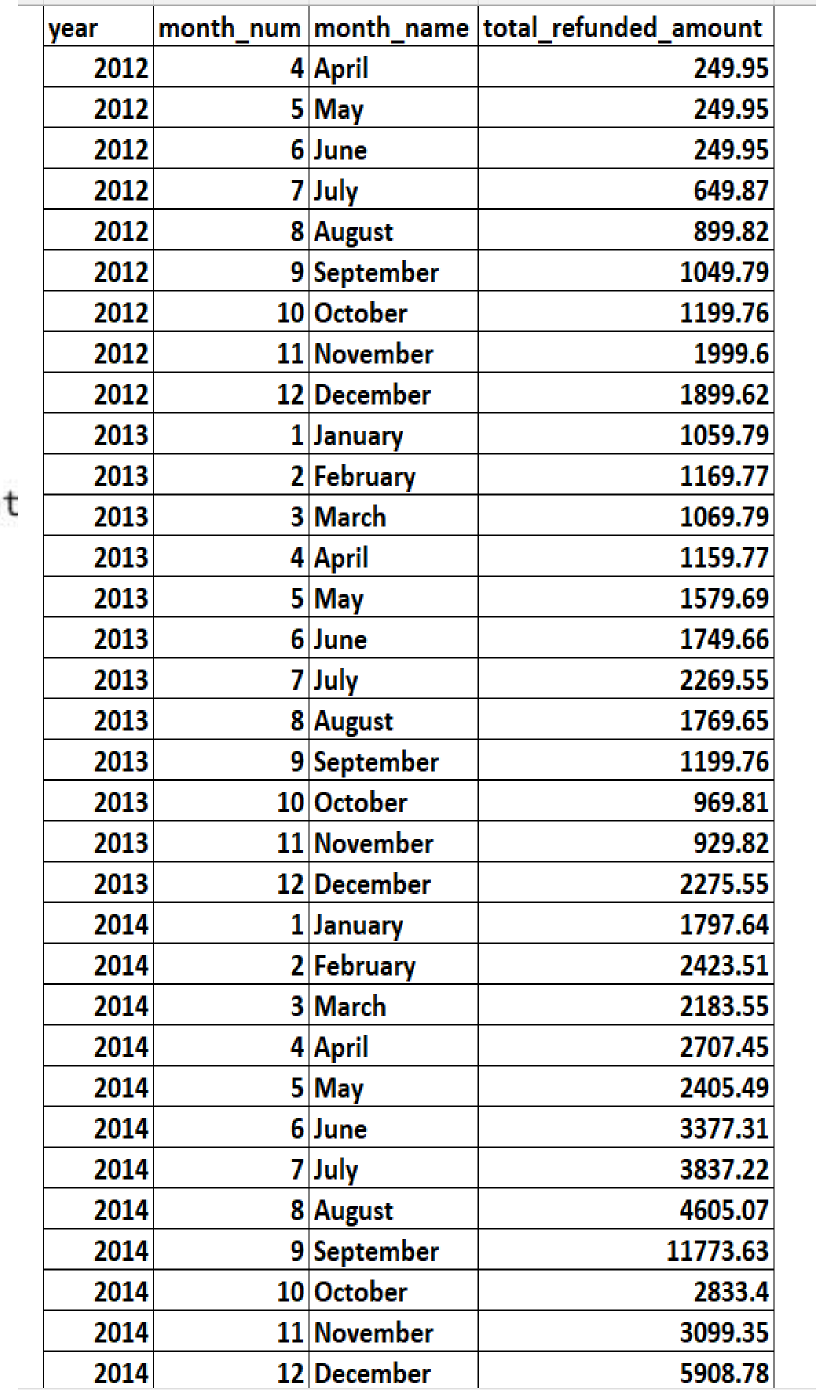


product_name	Total Sold	Total Refund	Refund Rate
The Birthday Sugar Panda	4985	301	6.04
The Original Mr. Fuzzy	24226	1237	5.11
The Forever Love Bear	5796	129	2.23
The Hudson River Mini bear	5018	64	1.28

5. Find the average customer rating per vehicle type

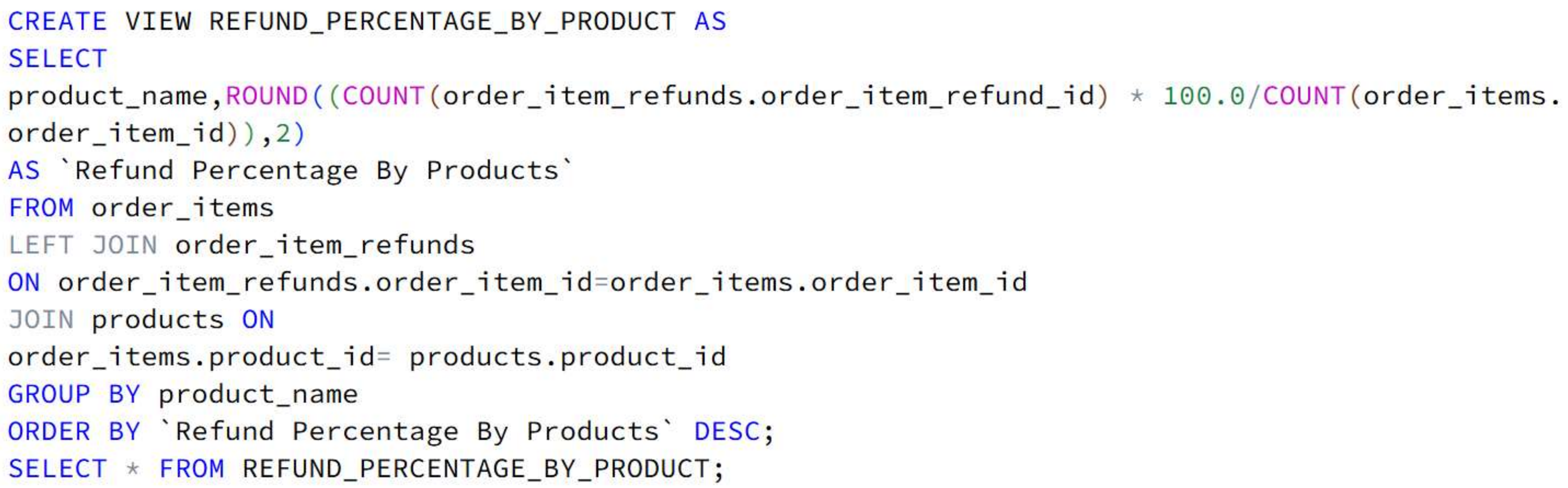


```
CREATE VIEW REFUND_BY_MONTH AS
SELECT
YEAR(order_item_refunds.created_at) AS year,
MONTH(order_item_refunds.created_at) AS month_num,
MONTHNAME(order_item_refunds.created_at) AS month_name,
ROUND(SUM(order_item_refunds.refund_amount_usd), 2) AS total_refunded_amount
FROM order_item_refunds
GROUP BY year, month_num, month_name
ORDER BY year, month_num ;
SELECT * FROM REFUND_BY_MONTH;
```

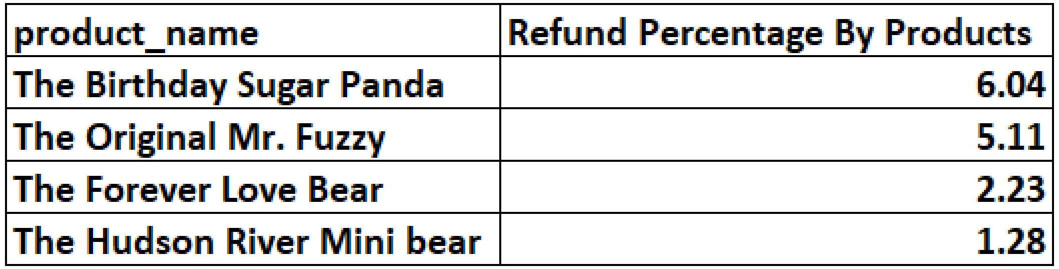


year	month_num	month_name	total_refunded_amount
2012	4	April	249.95
2012	5	May	249.95
2012	6	June	249.95
2012	7	July	649.87
2012	8	August	899.82
2012	9	September	1049.79
2012	10	October	1199.76
2012	11	November	1999.6
2012	12	December	1899.62
2013	1	January	1059.79
2013	2	February	1169.77
2013	3	March	1069.79
2013	4	April	1159.77
2013	5	May	1579.69
2013	6	June	1749.66
2013	7	July	2269.55
2013	8	August	1769.65
2013	9	September	1199.76
2013	10	October	969.81
2013	11	November	929.82
2013	12	December	2275.55
2014	1	January	1797.64
2014	2	February	2423.51
2014	3	March	2183.55
2014	4	April	2707.45
2014	5	May	2405.49
2014	6	June	3377.31
2014	7	July	3837.22
2014	8	August	4605.07
2014	9	September	11773.63
2014	10	October	2833.4
2014	11	November	3099.35
2014	12	December	5908.78

6. Refund percentage by product.

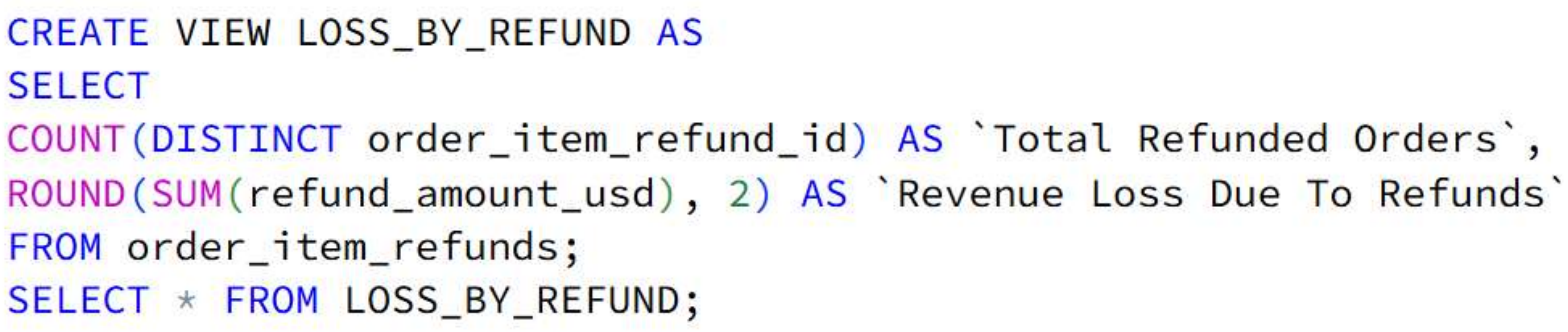


```sql
CREATE VIEW REFUND_PERCENTAGE_BY_PRODUCT AS
SELECT
product_name,ROUND((COUNT(order_item_refunds.order_item_refund_id) * 100.0/COUNT(order_items.
order_item_id)),2)
AS `Refund Percentage By Products`
FROM order_items
LEFT JOIN order_item_refunds
ON order_item_refunds.order_item_id=order_items.order_item_id
JOIN products ON
order_items.product_id= products.product_id
GROUP BY product_name
ORDER BY `Refund Percentage By Products` DESC;
SELECT * FROM REFUND_PERCENTAGE_BY_PRODUCT;
```

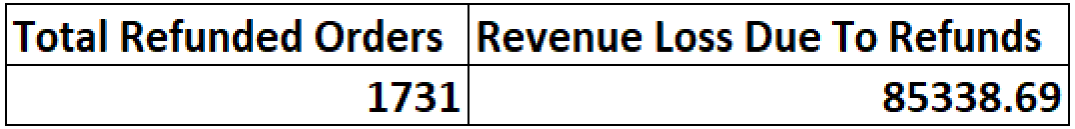


product_name	Refund Percentage By Products
The Birthday Sugar Panda	6.04
The Original Mr. Fuzzy	5.11
The Forever Love Bear	2.23
The Hudson River Mini bear	1.28

7. Revenue loss due to refunds.

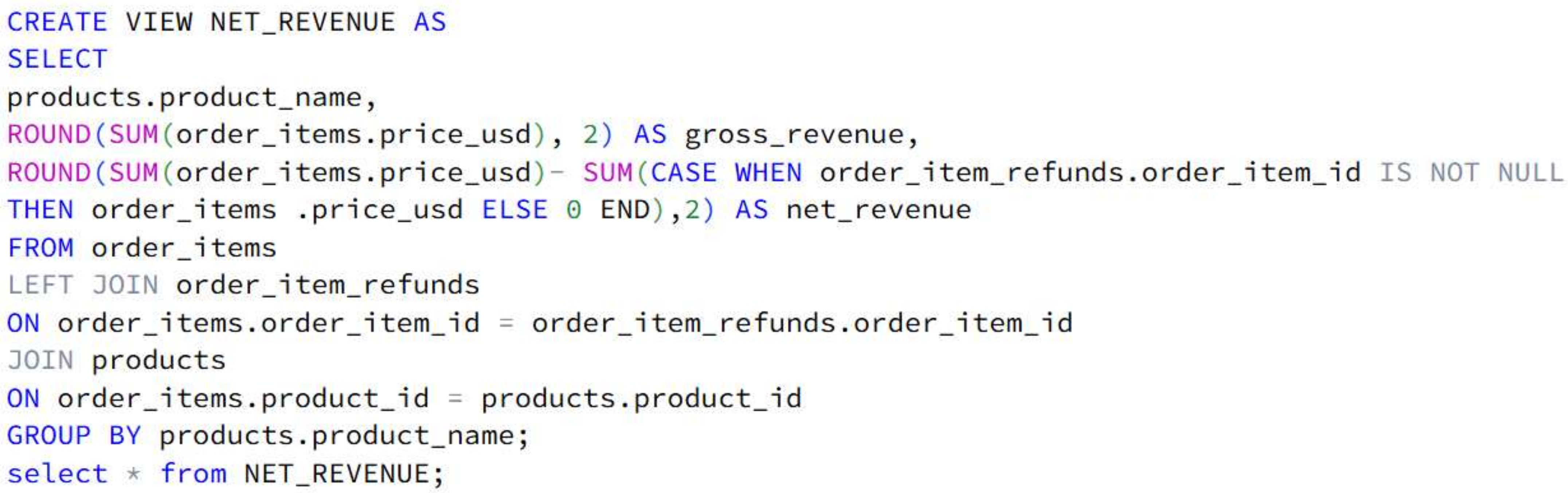


```
CREATE VIEW LOSS_BY_REFUND AS
SELECT
COUNT(DISTINCT order_item_refund_id) AS `Total Refunded Orders`,
ROUND(SUM(refund_amount_usd), 2) AS `Revenue Loss Due To Refunds`
FROM order_item_refunds;
SELECT * FROM LOSS_BY_REFUND;
```

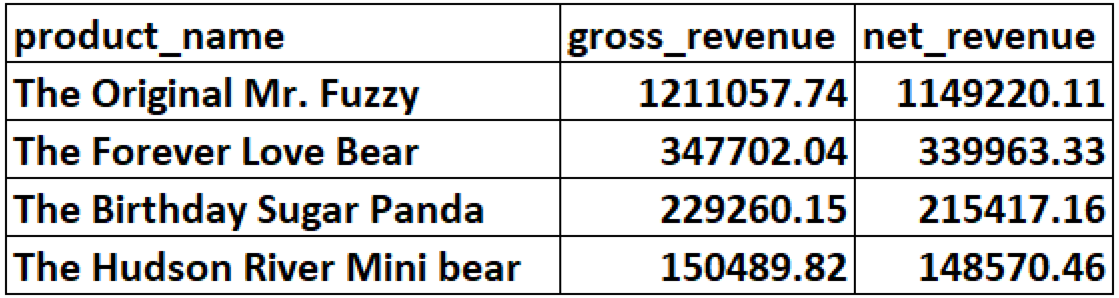


Total Refunded Orders	Revenue Loss Due To Refunds
1731	85338.69

8. Net revenue after refunds.

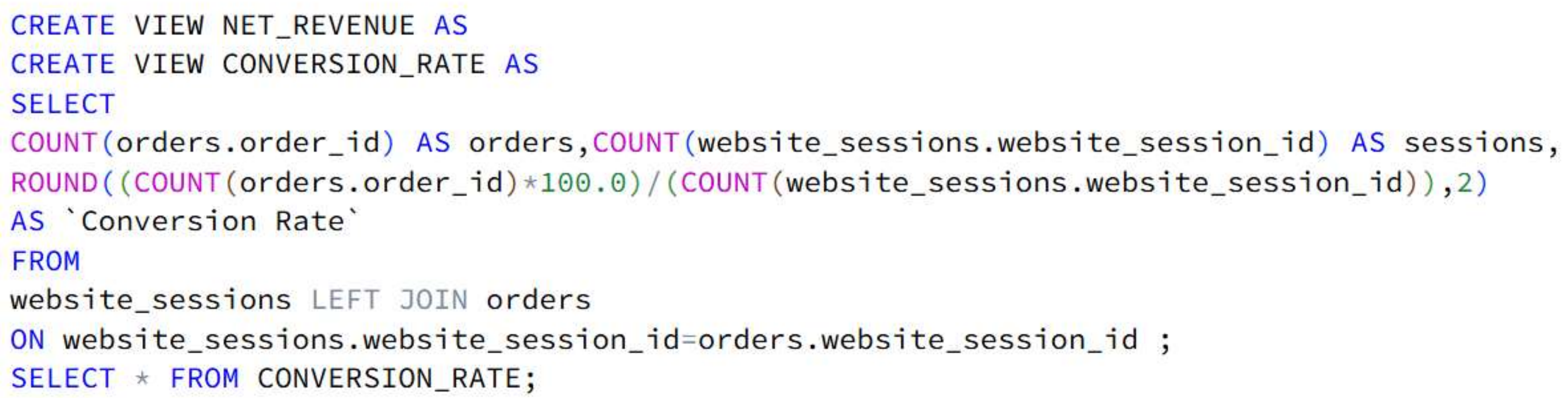


```
CREATE VIEW NET_REVENUE AS
SELECT
products.product_name,
ROUND(SUM(order_items.price_usd), 2) AS gross_revenue,
ROUND(SUM(order_items.price_usd)- SUM(CASE WHEN order_item_refunds.order_item_id IS NOT NULL
THEN order_items .price_usd ELSE 0 END),2) AS net_revenue
FROM order_items
LEFT JOIN order_item_refunds
ON order_items.order_item_id = order_item_refunds.order_item_id
JOIN products
ON order_items.product_id = products.product_id
GROUP BY products.product_name;
select * from NET_REVENUE;
```

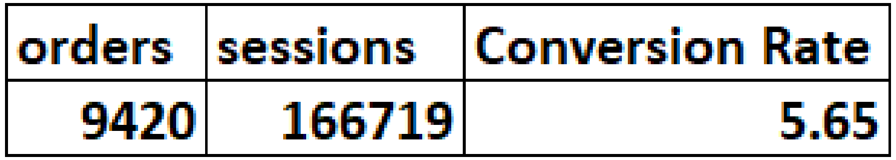


product_name	gross_revenue	net_revenue
The Original Mr. Fuzzy	1211057.74	1149220.11
The Forever Love Bear	347702.04	339963.33
The Birthday Sugar Panda	229260.15	215417.16
The Hudson River Mini bear	150489.82	148570.46

9. Sessions → Orders conversion rate.

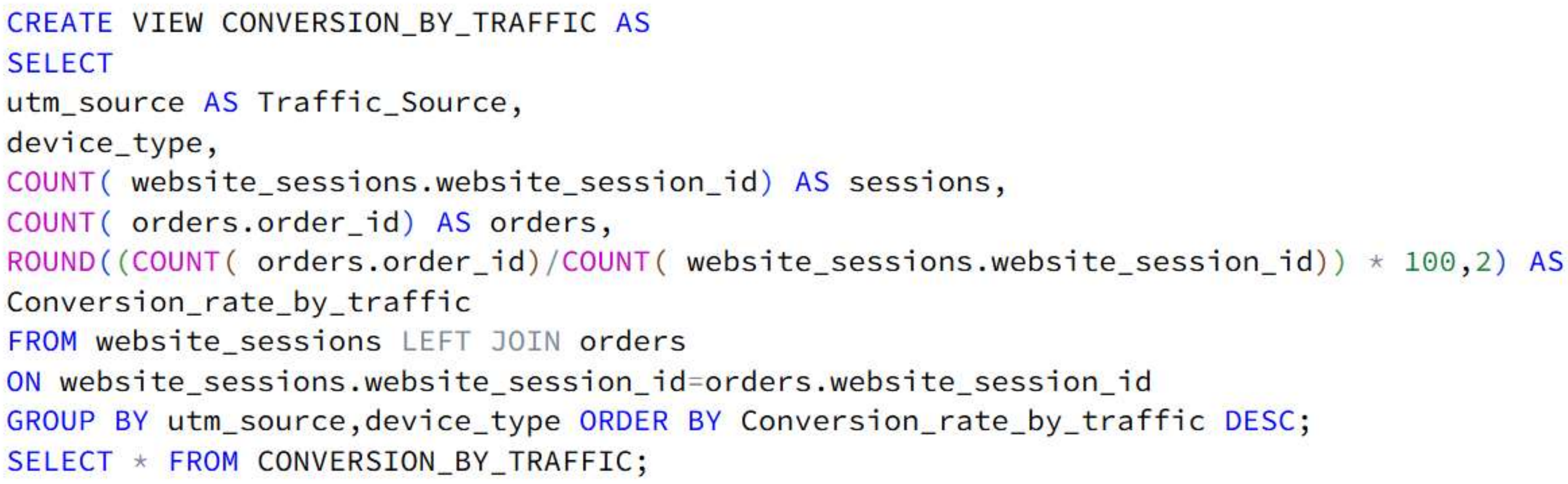


```
CREATE VIEW NET_REVENUE AS
CREATE VIEW CONVERSION_RATE AS
SELECT
COUNT(orders.order_id) AS orders,COUNT(website_sessions.website_session_id) AS sessions,
ROUND((COUNT(orders.order_id)*100.0)/(COUNT(website_sessions.website_session_id)),2)
AS `Conversion Rate`
FROM
website_sessions LEFT JOIN orders
ON website_sessions.website_session_id=orders.website_session_id ;
SELECT * FROM CONVERSION_RATE;
```

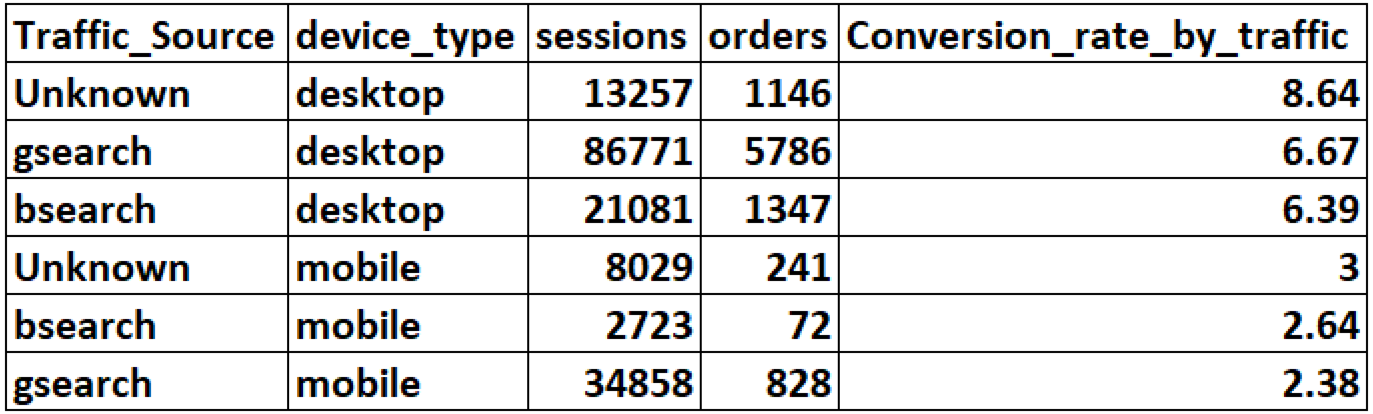


orders	sessions	Conversion Rate
9420	166719	5.65

10. Conversion rate by traffic source.

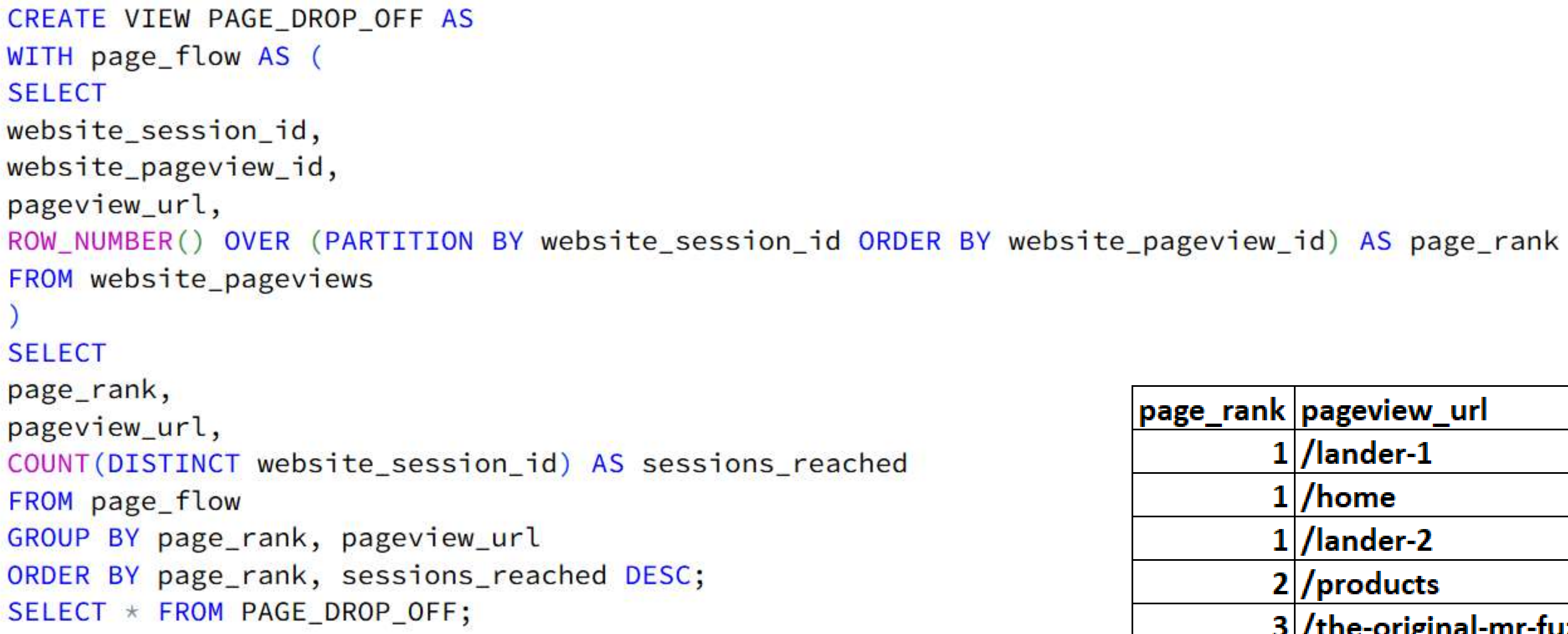


```
CREATE VIEW CONVERSION_BY_TRAFFIC AS
SELECT
utm_source AS Traffic_Source,
device_type,
COUNT( website_sessions.website_session_id) AS sessions,
COUNT( orders.order_id) AS orders,
ROUND((COUNT( orders.order_id)/COUNT( website_sessions.website_session_id)) * 100,2) AS
Conversion_rate_by_traffic
FROM website_sessions LEFT JOIN orders
ON website_sessions.website_session_id=orders.website_session_id
GROUP BY utm_source,device_type ORDER BY Conversion_rate_by_traffic DESC;
SELECT * FROM CONVERSION_BY_TRAFFIC;
```

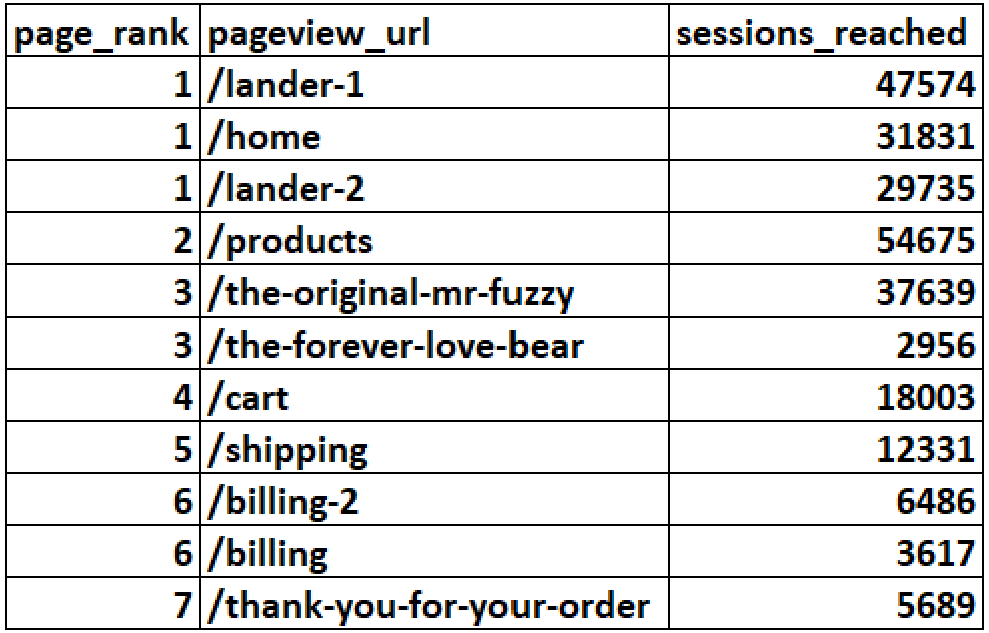


Traffic_Source	device_type	sessions	orders	Conversion_rate_by_traffic
Unknown	desktop	13257	1146	8.64
gsearch	desktop	86771	5786	6.67
bsearch	desktop	21081	1347	6.39
Unknown	mobile	8029	241	3
bsearch	mobile	2723	72	2.64
gsearch	mobile	34858	828	2.38

11. Page-wise drop-off analysis.

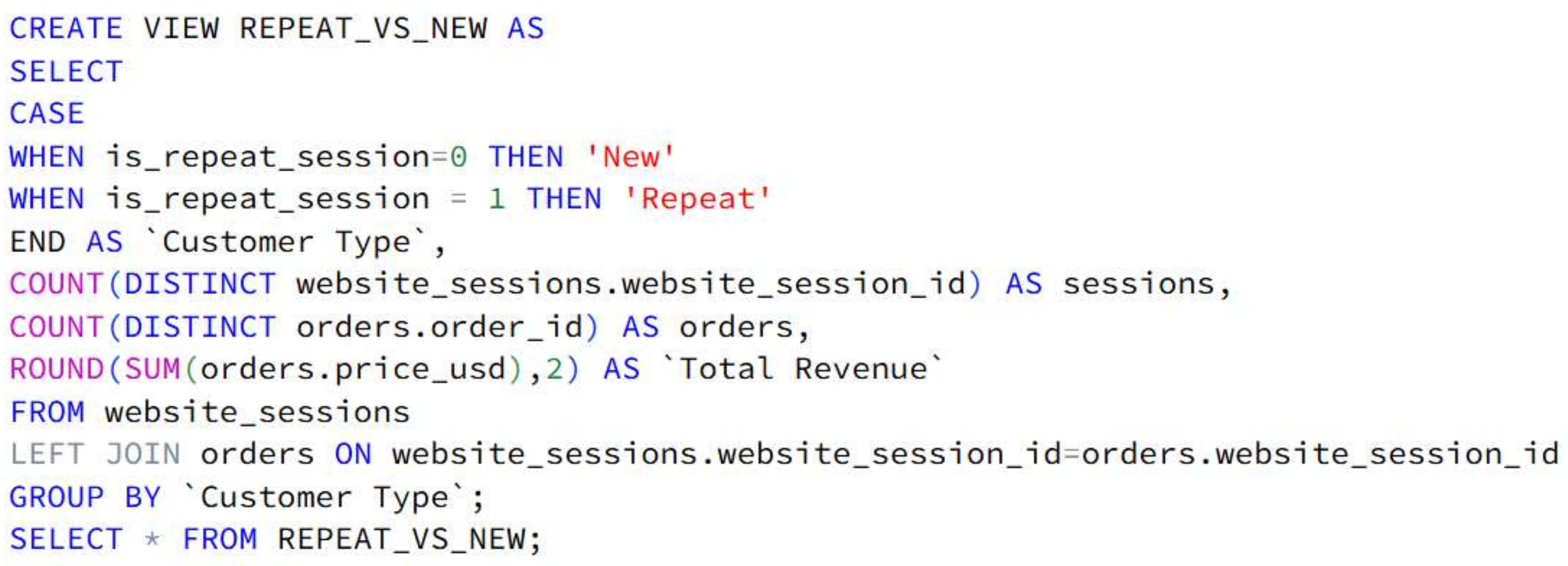


```
CREATE VIEW PAGE_DROP_OFF AS
WITH page_flow AS (
SELECT
website_session_id,
website_pageview_id,
pageview_url,
ROW_NUMBER() OVER (PARTITION BY website_session_id ORDER BY website_pageview_id) AS page_rank
FROM website_pageviews
)
SELECT
page_rank,
pageview_url,
COUNT(DISTINCT website_session_id) AS sessions_reached
FROM page_flow
GROUP BY page_rank, pageview_url
ORDER BY page_rank, sessions_reached DESC;
SELECT * FROM PAGE_DROP_OFF;
```

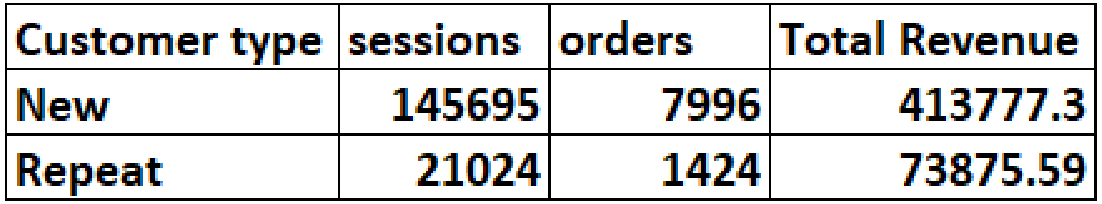


page_rank	pageview_url	sessions_reached
1	/lander-1	47574
1	/home	31831
1	/lander-2	29735
2	/products	54675
3	/the-original-mr-fuzzy	37639
3	/the-forever-love-bear	2956
4	/cart	18003
5	/shipping	12331
6	/billing-2	6486
6	/billing	3617
7	/thank-you-for-your-order	5689

12. Repeat vs New customer analysis.

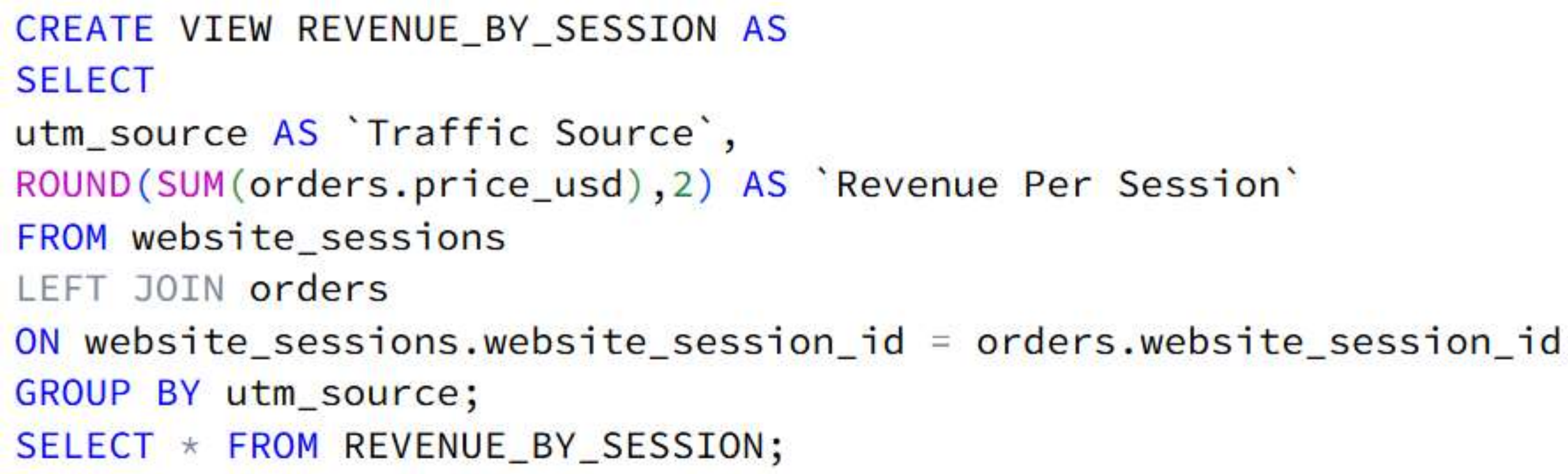


```
CREATE VIEW REPEAT_VS_NEW AS
SELECT
CASE
WHEN is_repeat_session=0 THEN 'New'
WHEN is_repeat_session = 1 THEN 'Repeat'
END AS `Customer Type`,
COUNT(DISTINCT website_sessions.website_session_id) AS sessions,
COUNT(DISTINCT orders.order_id) AS orders,
ROUND(SUM(orders.price_usd),2) AS `Total Revenue`
FROM website_sessions
LEFT JOIN orders ON website_sessions.website_session_id=orders.website_session_id
GROUP BY `Customer Type`;
SELECT * FROM REPEAT_VS_NEW;
```

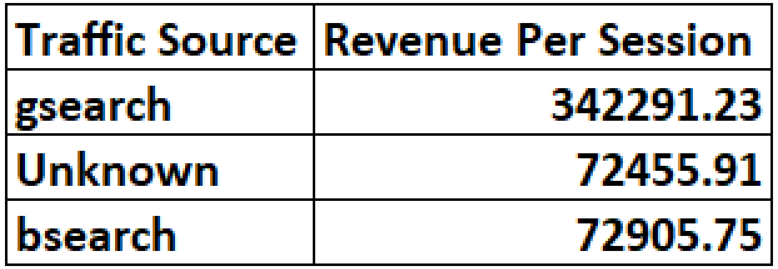


Customer type	sessions	orders	Total Revenue
New	145695	7996	413777.3
Repeat	21024	1424	73875.59

13. Revenue per session.



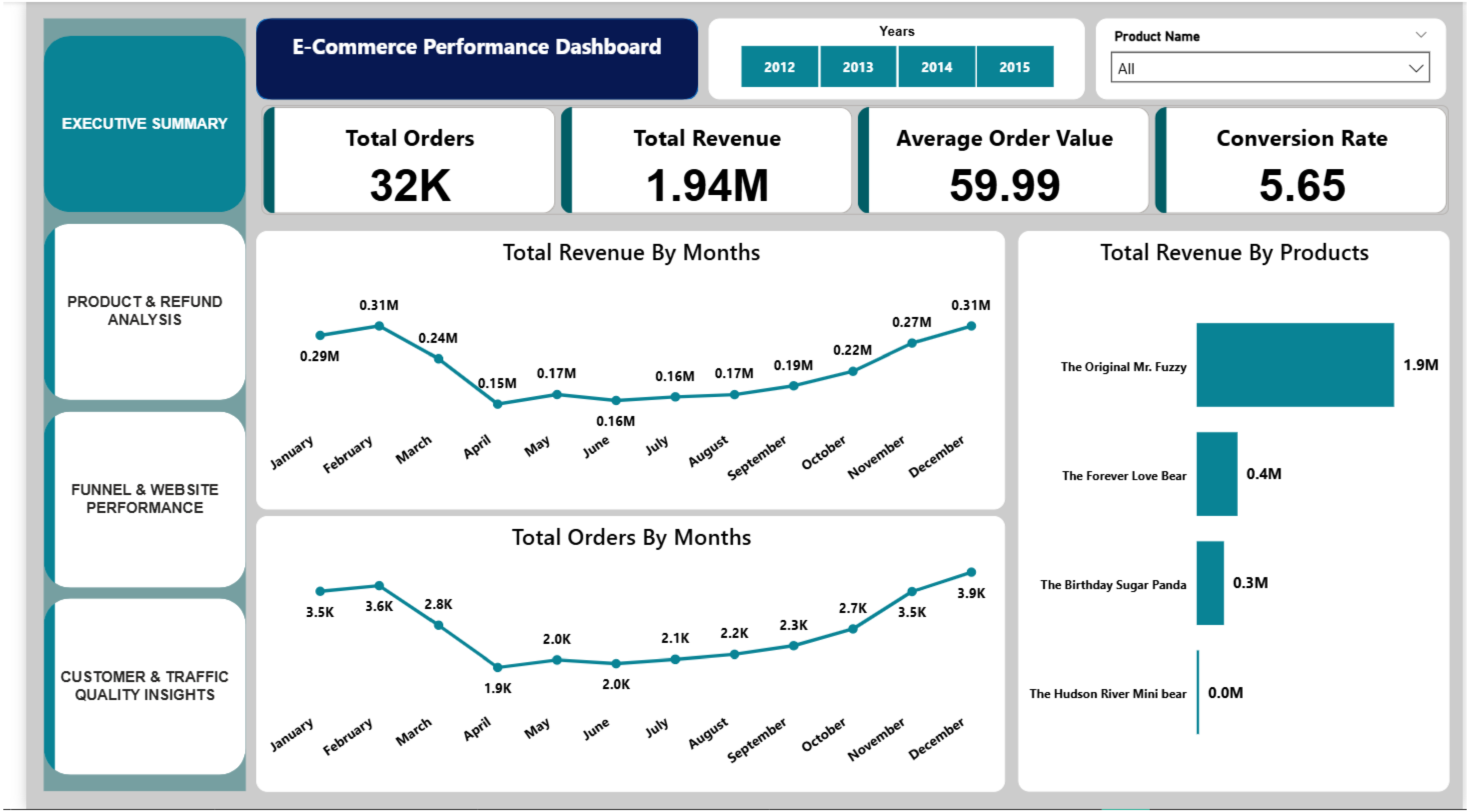
```
CREATE VIEW REVENUE_BY_SESSION AS
SELECT
utm_source AS `Traffic Source`,
ROUND(SUM(orders.price_usd),2) AS `Revenue Per Session`
FROM website_sessions
LEFT JOIN orders
ON website_sessions.website_session_id = orders.website_session_id
GROUP BY utm_source;
SELECT * FROM REVENUE_BY_SESSION;
```



Traffic Source	Revenue Per Session
gsearch	342291.23
Unknown	72455.91
bsearch	72905.75

Data Visualization

Executive Summary KPIs



E-Commerce Performance Dashboard

Years: 2012 2013 2014 2015

Product Name: All

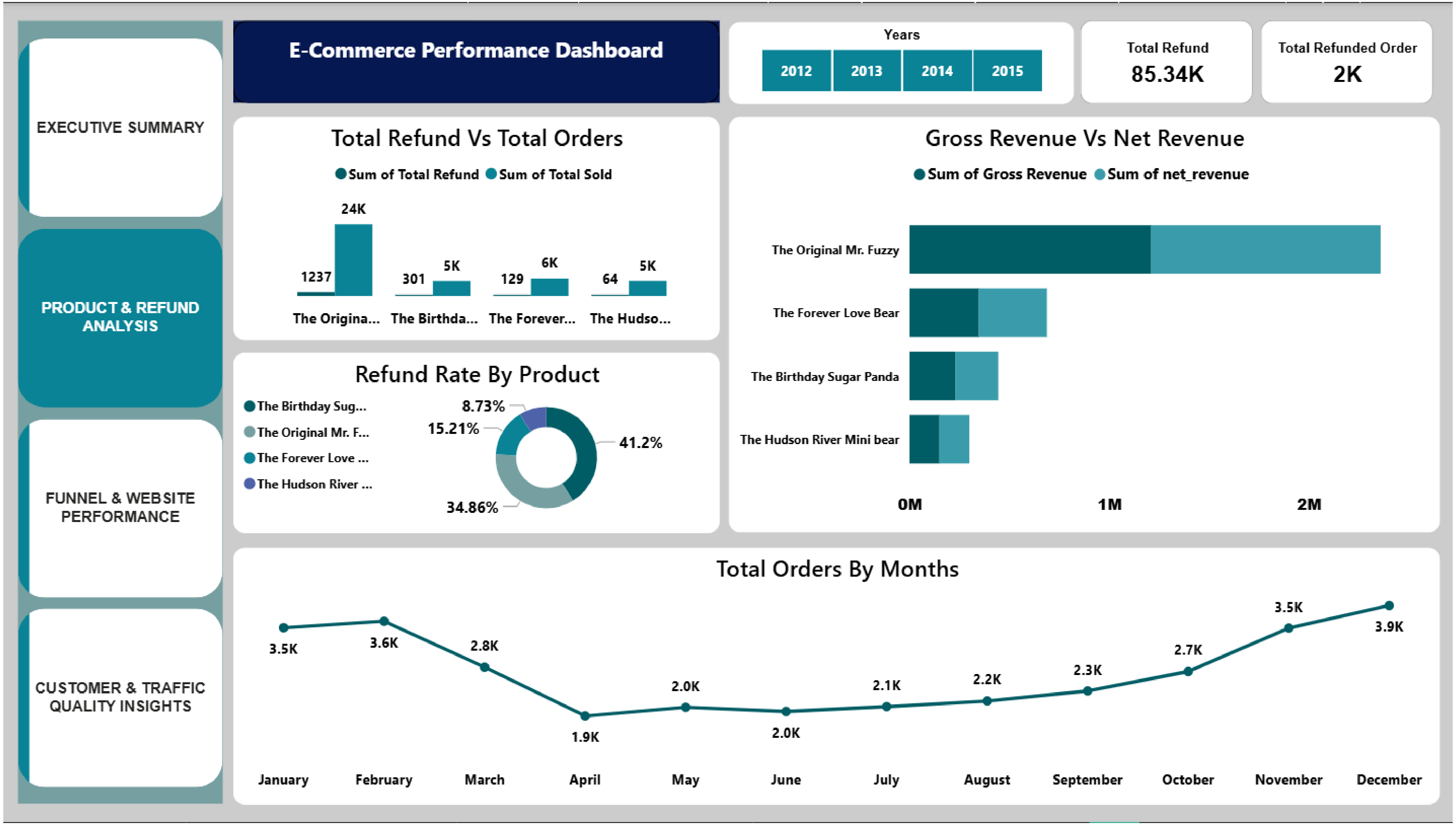
Total Orders 32K

Total Revenue 1.94M

Average Order Value 59.99

Conversion Rate 5.65

Product & Refund Analysis



EXECUTIVE SUMMARY

PRODUCT & REFUND ANALYSIS

FUNNEL & WEBSITE PERFORMANCE

CUSTOMER & TRAFFIC QUALITY INSIGHTS

E-Commerce Performance Dashboard

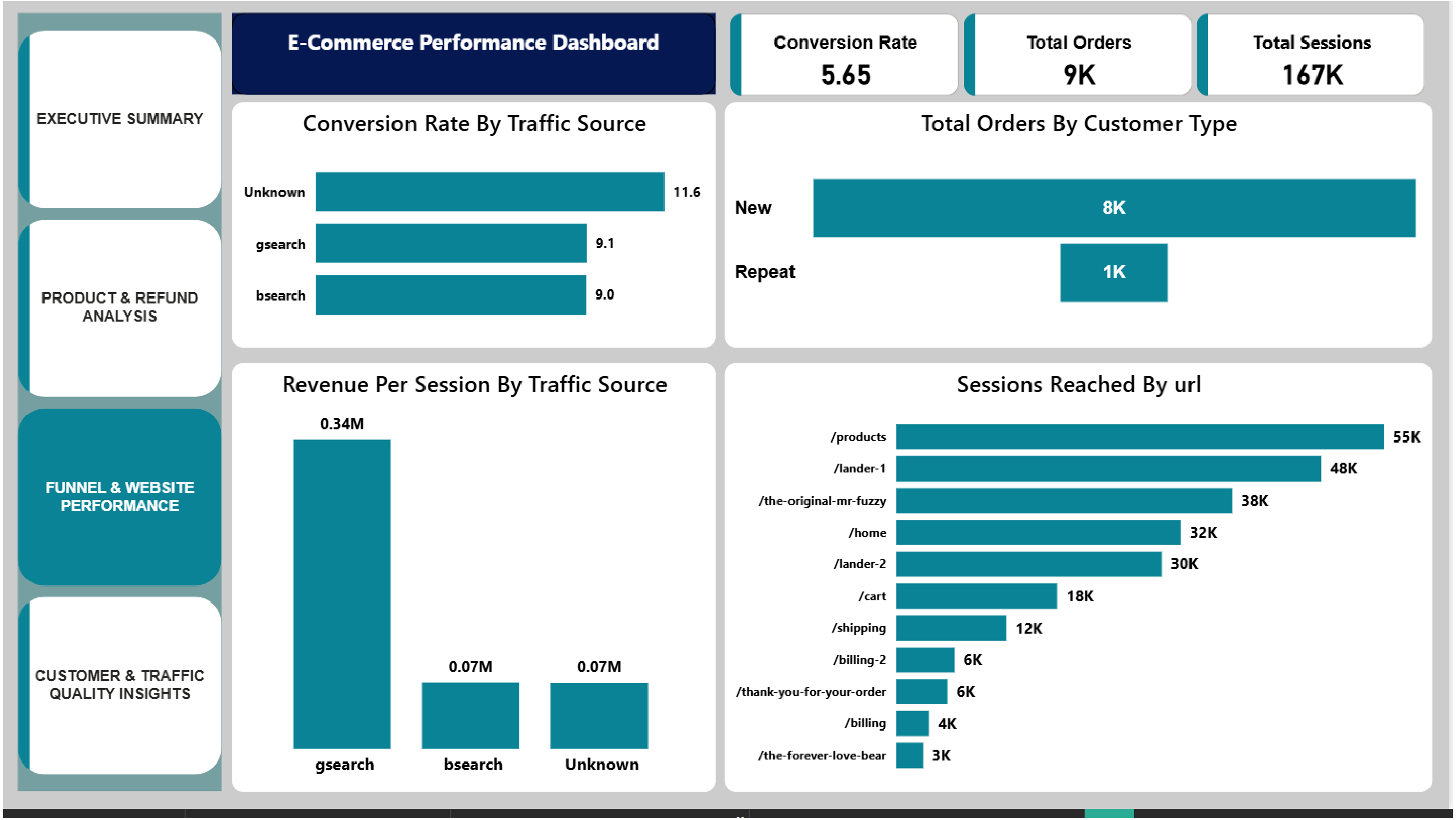
Years

2012 | 2013 | 2014 | 2015

Total Refund
85.34K

Total Refunded Order
2K

Funnel & Website Performance



EXECUTIVE SUMMARY

PRODUCT & REFUND ANALYSIS

FUNNEL & WEBSITE PERFORMANCE

CUSTOMER & TRAFFIC QUALITY INSIGHTS

E-Commerce Performance Dashboard

Conversion Rate 5.65

Total Orders 9K

Total Sessions 167K

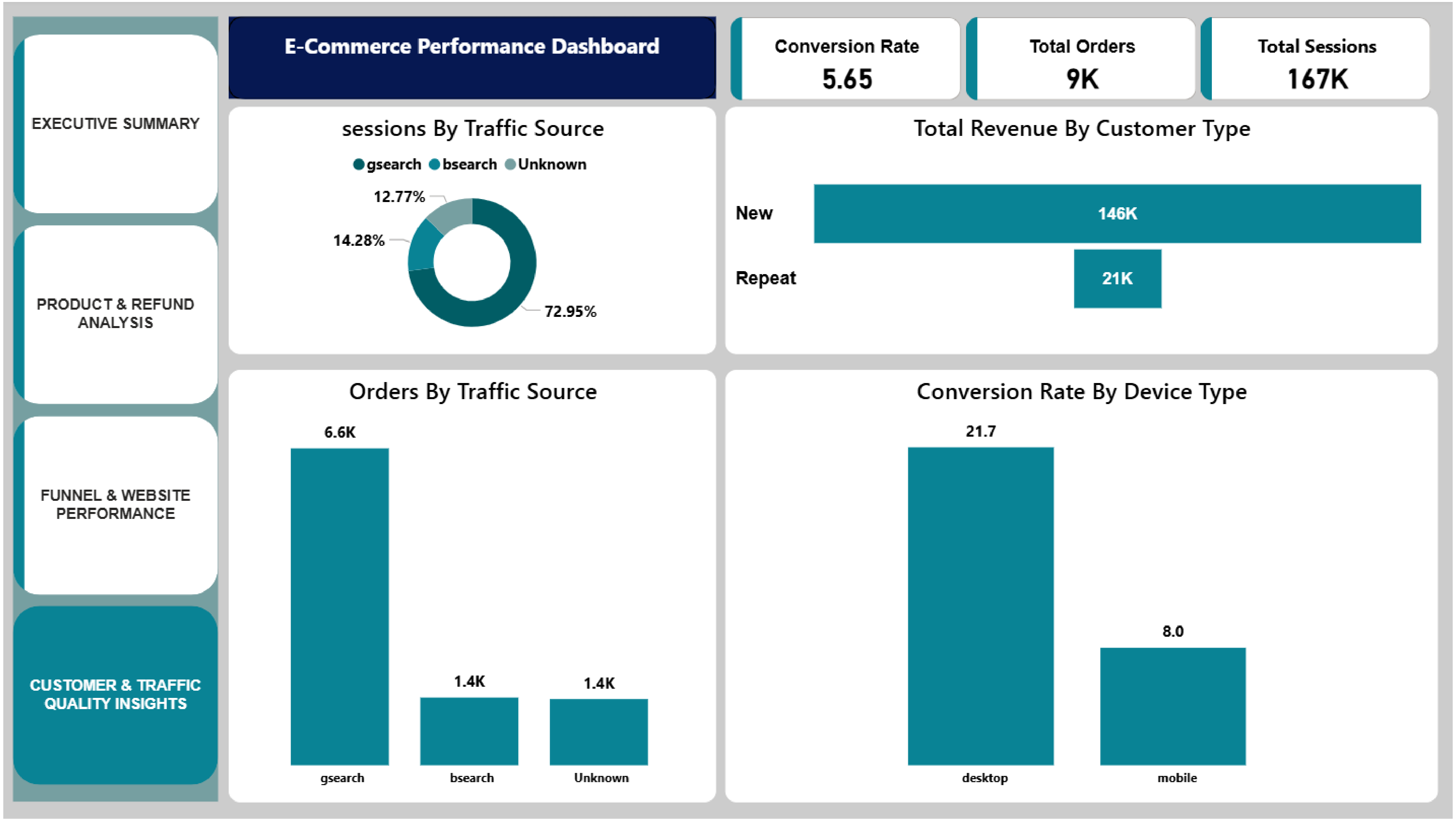
Conversion Rate By Traffic Source

Total Orders By Customer Type

Revenue Per Session By Traffic Source

Sessions Reached By url

Customer & Traffic Quality Insights



E-Commerce Performance Dashboard

Conversion Rate 5.65

Total Orders 9K

Total Sessions 167K

EXECUTIVE SUMMARY
PRODUCT & REFUND ANALYSIS
FUNNEL & WEBSITE PERFORMANCE
CUSTOMER & TRAFFIC QUALITY INSIGHTS
sessions By Traffic Source
● gsearch ● bsearch ● Unknown
12.77%
14.28%
72.95%
Total Revenue By Customer Type
New
146K
Repeat
21K
Orders By Traffic Source
6.6K
1.4K
1.4K
gsearch
bsearch
Unknown
Conversion Rate By Device Type
21.7
8.0
desktop
mobile

Business Insights & Recommendations

Business Insights

- Refunds significantly reduce net revenue despite strong gross sales
- A small number of products contribute disproportionately to refunds
- Conversion rates around 5–6% indicate strong e-commerce performance
- Organic and direct traffic channels deliver higher revenue per session
- Funnel analysis highlights optimization opportunities before checkout

Recommendations

- Improve product quality and descriptions for high-refund items
- Optimize checkout and landing pages to reduce funnel drop-offs
- Invest more in high-converting traffic sources
- Monitor refund trends as a key profitability metric

Thank You!!