E-COMMERCE SALES ANALYSIS & CUSTOMER SEGMENTATION (RFM)

AN END-TO-END DATA ANALYTICS PROJECT USING PYTHON & POWER BI

Presented by – Mayur Kumar

PROJECT OVERVIEW

- This project analyzes dummy dataset from a UK-based E-commerce retailer.
- Built complete end-to-end using Python (Pandas) and Power BI.
- Performed data cleaning, transformation, and feature engineering.
- Developed key dashboards to understand revenue trends, customer behavior, and product performance.
- Implemented RFM (Recency, Frequency, Monetary) segmentation to classify customer groups.
- Delivered actionable business insights for marketing and sales teams.

DATASET DESCRIPTION

- Total Records: 532,621 (2010–2011)
- Contains transactional-level sales data
- Used for analysing trends, customer behaviour, and product performance

BUSINESS PROBLEM

A UK-based online retail company experienced inconsistent sales, high customer drop-off, and unclear purchasing patterns across different markets.

Although the company collected large amounts of transactional data, it lacked the ability to convert this data into meaningful business insights.

MAIN PROBLEMS FACED BY THE BUSINESS

- Customers were not buying repeatedly, and the company didn't know why.
- The business did not know which customers were valuable or which ones were slipping away.
- Sales changed a lot across months, but there was no clear understanding of trends or peak seasons.
- The company did not know which products were top performers and which ones were not selling well.
- Lack of insights made marketing, customer retention, and inventory planning difficult.

What the Business Will Gain

- Identify the most valuable customers using RFM segmentation (Champions, Loyal, Potential Loyalists).
- Detect customers at risk of churn and long-inactive customers (At-Risk, Hibernating/Lost).
- Understand monthly sales trends and identify peak revenue periods.
- Know which products generate the highest revenue across all transactions.
- Understand which countries contribute most to sales and where growth potential exists.
- Use an interactive dashboard to monitor KPIs, customer behavior, and product performance in real time.

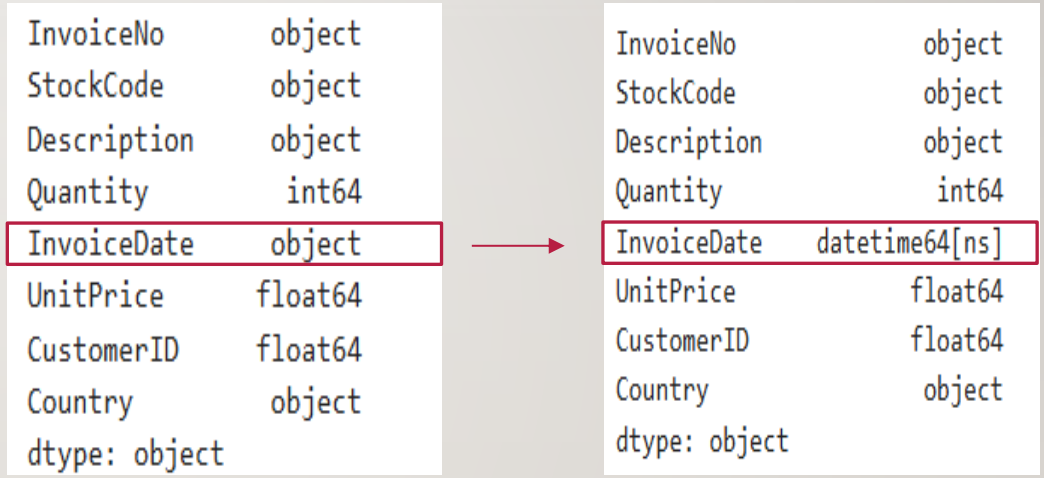
Process Involved

- Clean and prepare the data for analysis
- Calculate sales KPIs and monthly revenue trends
- Identify top products and top countries
- Perform RFM analysis to study customer behavior
- Segment customers using RFM scores
- Build an interactive Power BI dashboard to visualize insights

Data Cleaning

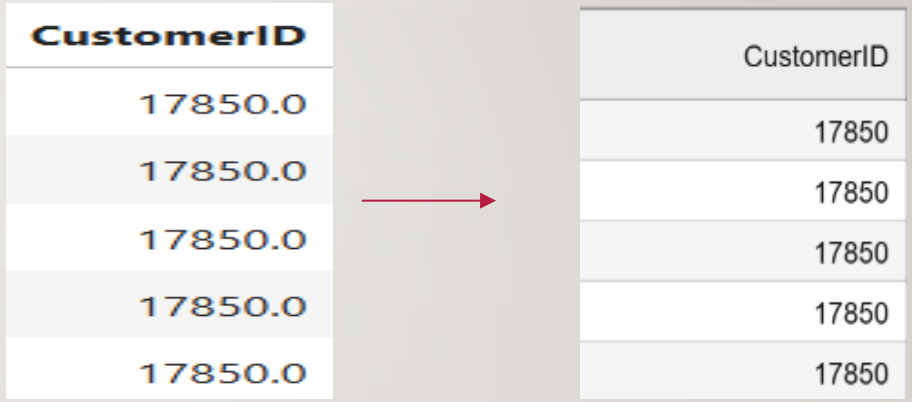
CONVERTING INVOICE DATE TO DATETIME

The Invoice Date column was stored as text, so it was converted it into proper datetime format to enable accurate time-based analysis.



STANDARDIZING CUSTOMER-ID DATA TYPE

Customer-ID was originally stored as a float with decimals. So it was converted it into an integer first and then into a string for accurate grouping.



CustomerID
17850.0
17850.0
17850.0
17850.0
17850.0

CustomerID
17850
17850
17850
17850
17850

REMOVING CANCELLED INVOICES

Invoice numbers starting with ‘C’ represent cancelled transactions. These records were removed to ensure accurate revenue calculations and prevent distortion in analysis.

InvoiceNo	StockCode
C553748	21658
C553748	22191
C553748	22194
C553763	23112
C553763	22487

REMOVING NEGATIVE AND ZERO QUANTITIES

Rows with negative or zero quantities represent product returns or invalid transactions. These records were removed to maintain accurate revenue and quantity analysis.

Quantity
-1
-1
-12
-24
-24

Quantity
6
6
8
6
6

CREATING THE REVENUE COLUMN

To analyze sales performance, A Revenue column was created by multiplying Quantity and UnitPrice. This value powers all KPIs and Power BI visualizations.

	Quantity		UnitPrice		Revenue
0	6	x	2.55	=	15.30
1	6	x	3.39	=	20.34
2	8	x	2.75	=	22.00
3	6	x	3.39	=	20.34
4	6	x	3.39	=	20.34

CLEANING PRODUCT DESCRIPTIONS

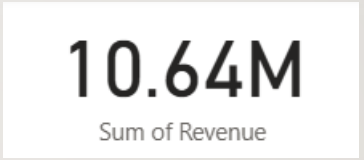
Product descriptions contained inconsistent text formatting, so they were standardized by converting all values to lowercase and removing leading or trailing spaces.

" WHITE Hanging Heart T-Light Holder "	"white hanging heart t-light holder"
" Red POLKA DOT BOWL "	"red polka dot bowl"
"Party Bunting "	"party bunting"

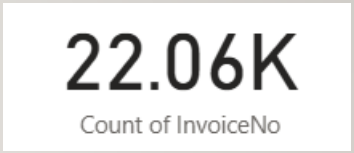
Calculating Key Metrics (KPIs)

CALCULATING KEY METRICS (KPIS)

After cleaning the dataset, key performance metrics such as total revenue, total orders, customer count, quantity sold, and AOV were calculated to understand overall business performance.



Total Revenue



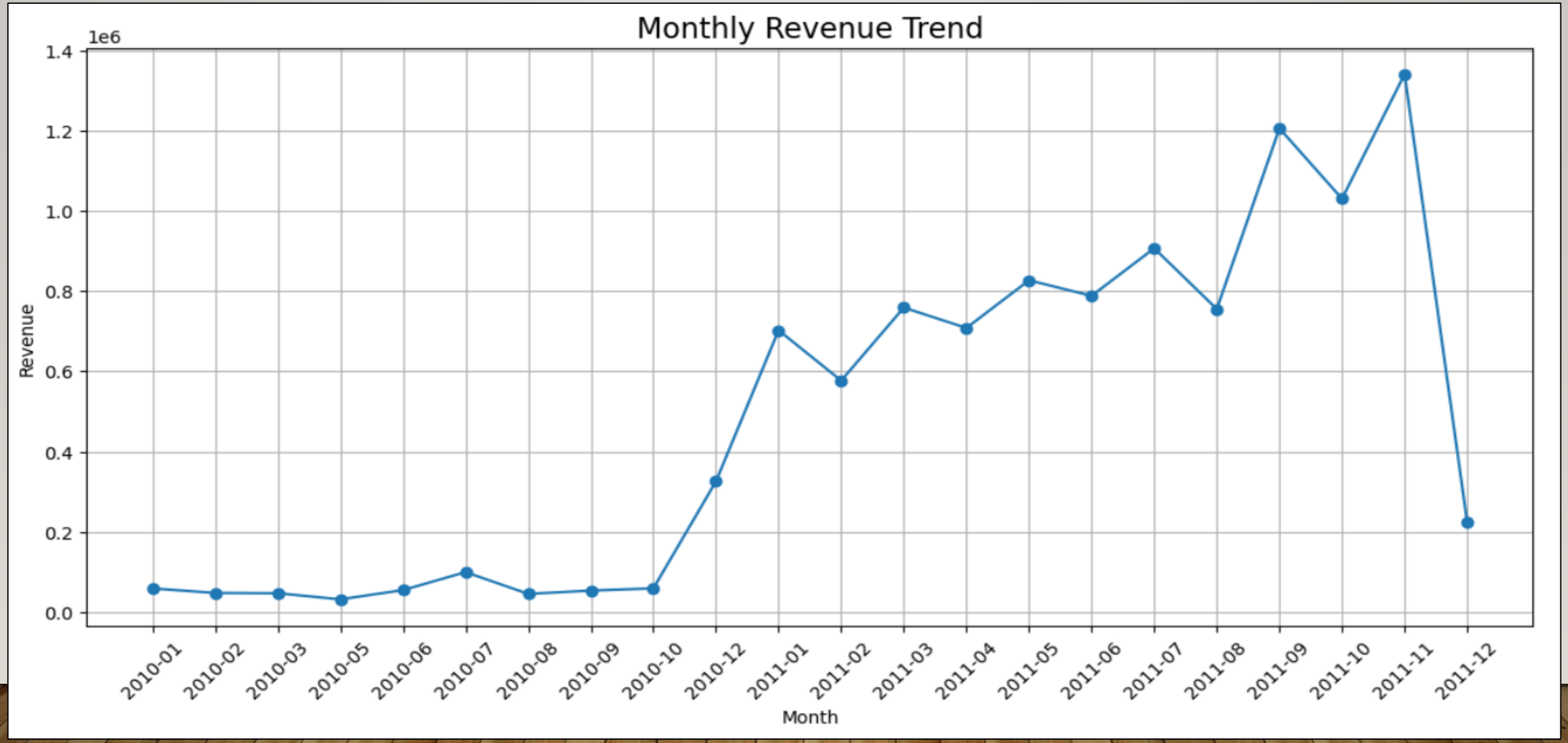
Total Orders

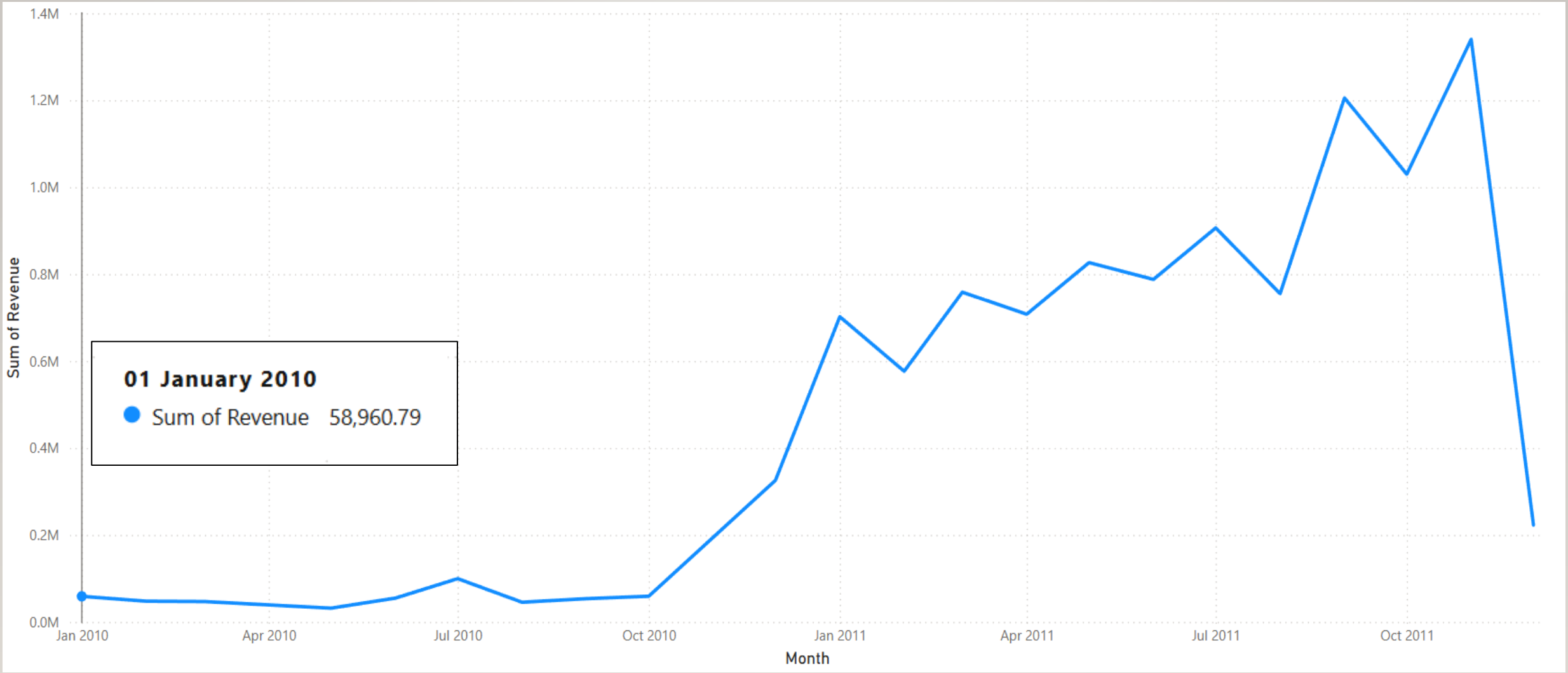
Total Customers

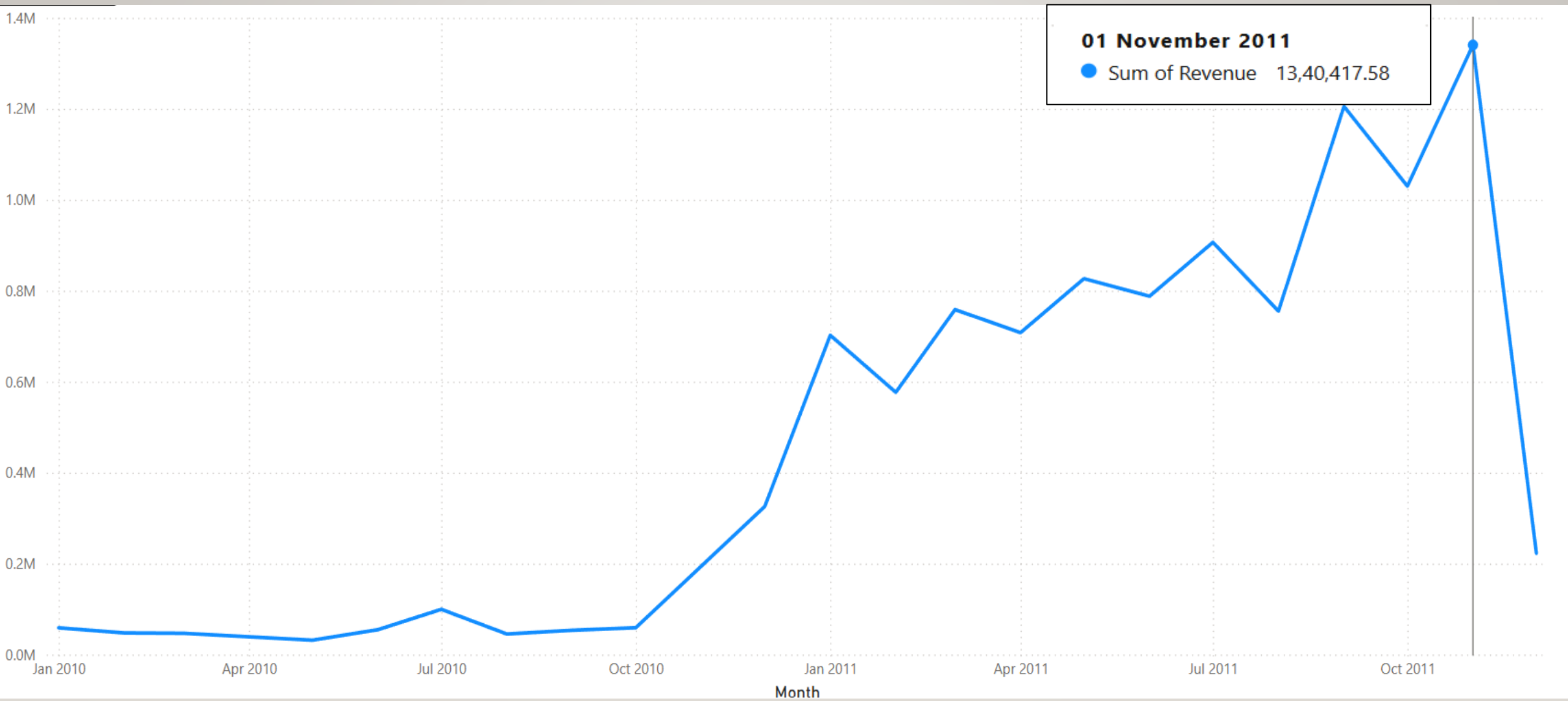
Total Quantity of Products

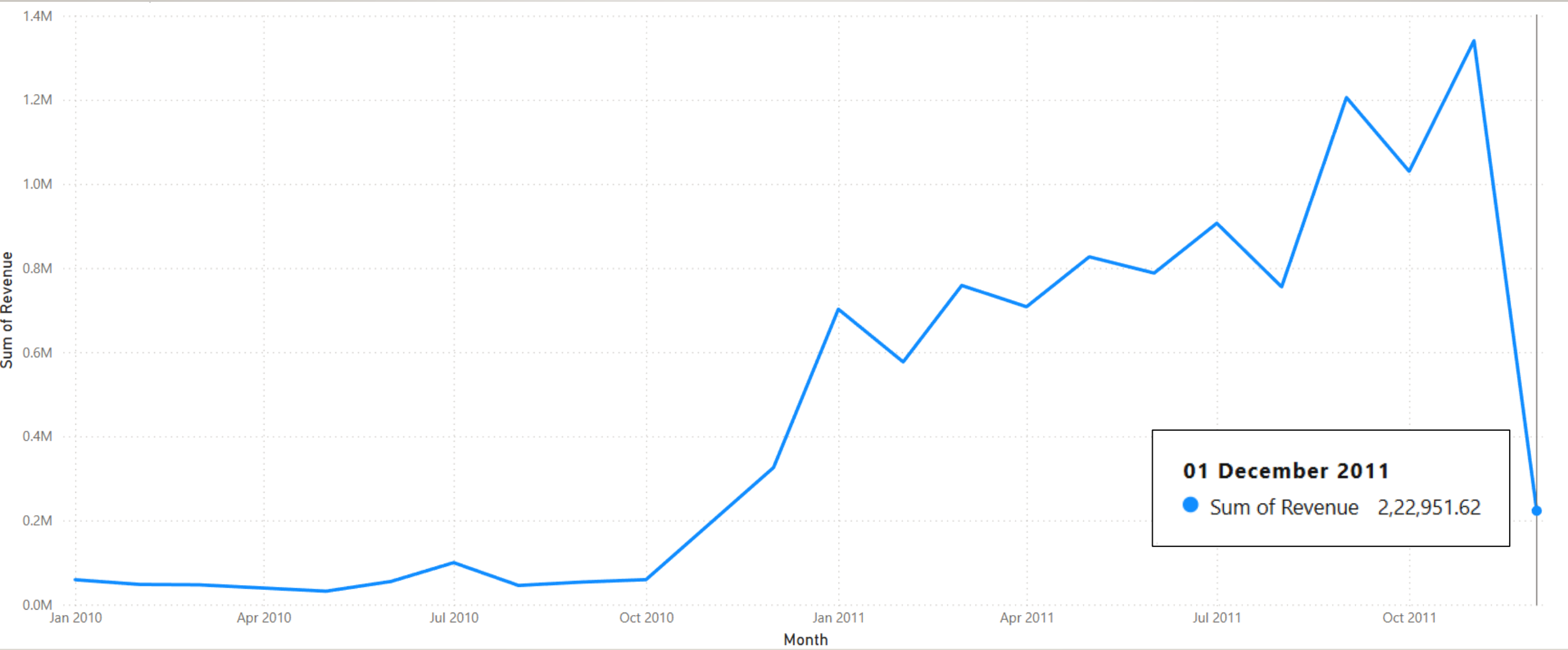
MONTHLY REVENUE TREND ANALYSIS

A monthly revenue trend plot to understand sales patterns over time. The analysis highlights growth through 2011 with a strong peak in November.

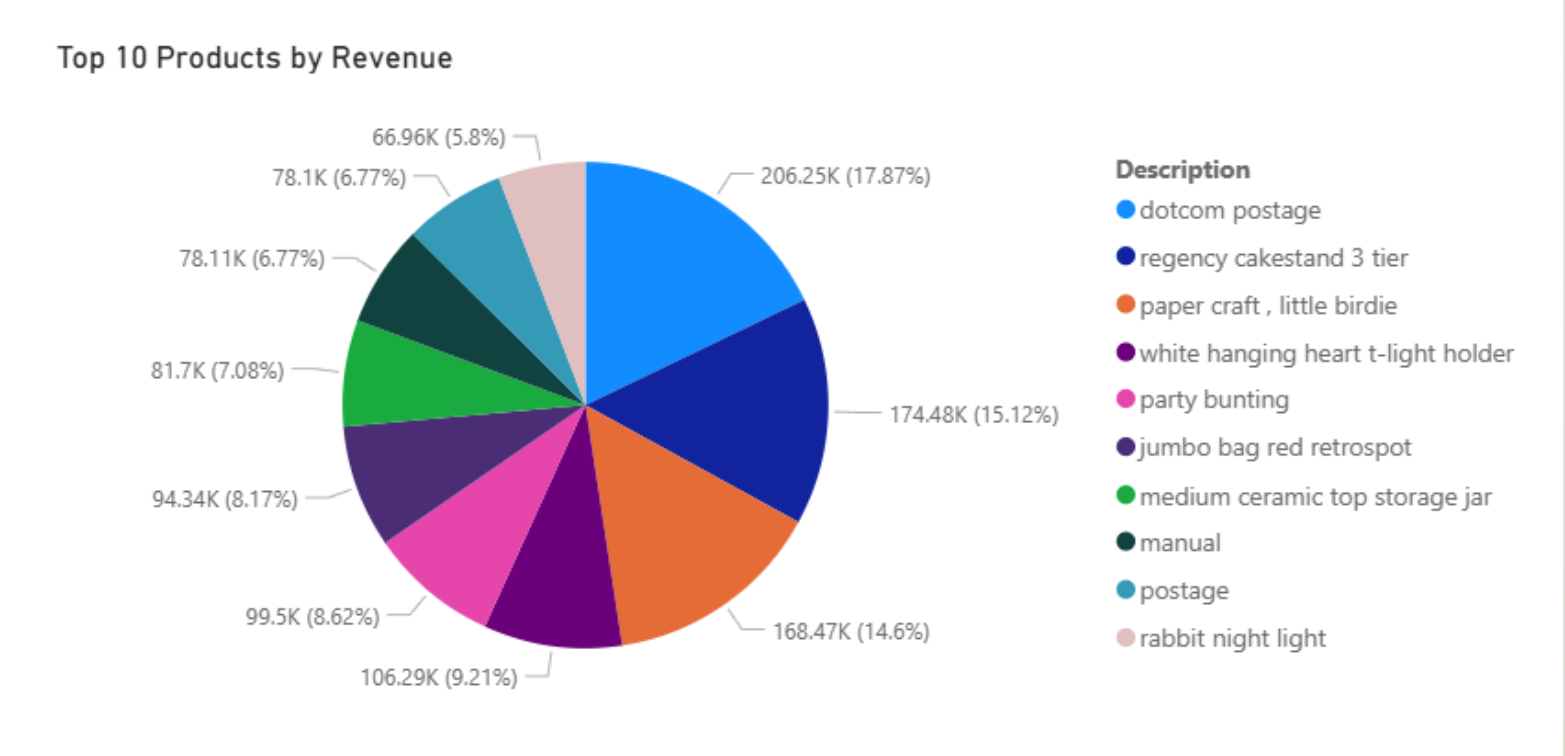



1.4M
1.2M
1.0M
0.8M
0.6M
0.4M
0.2M
0.0M
Sum of Revenue
01 January 2010
Sum of Revenue 58,960.79
Jan 2010
Apr 2010
Jul 2010
Oct 2010
Jan 2011
Apr 2011
Jul 2011
Oct 2011
Month

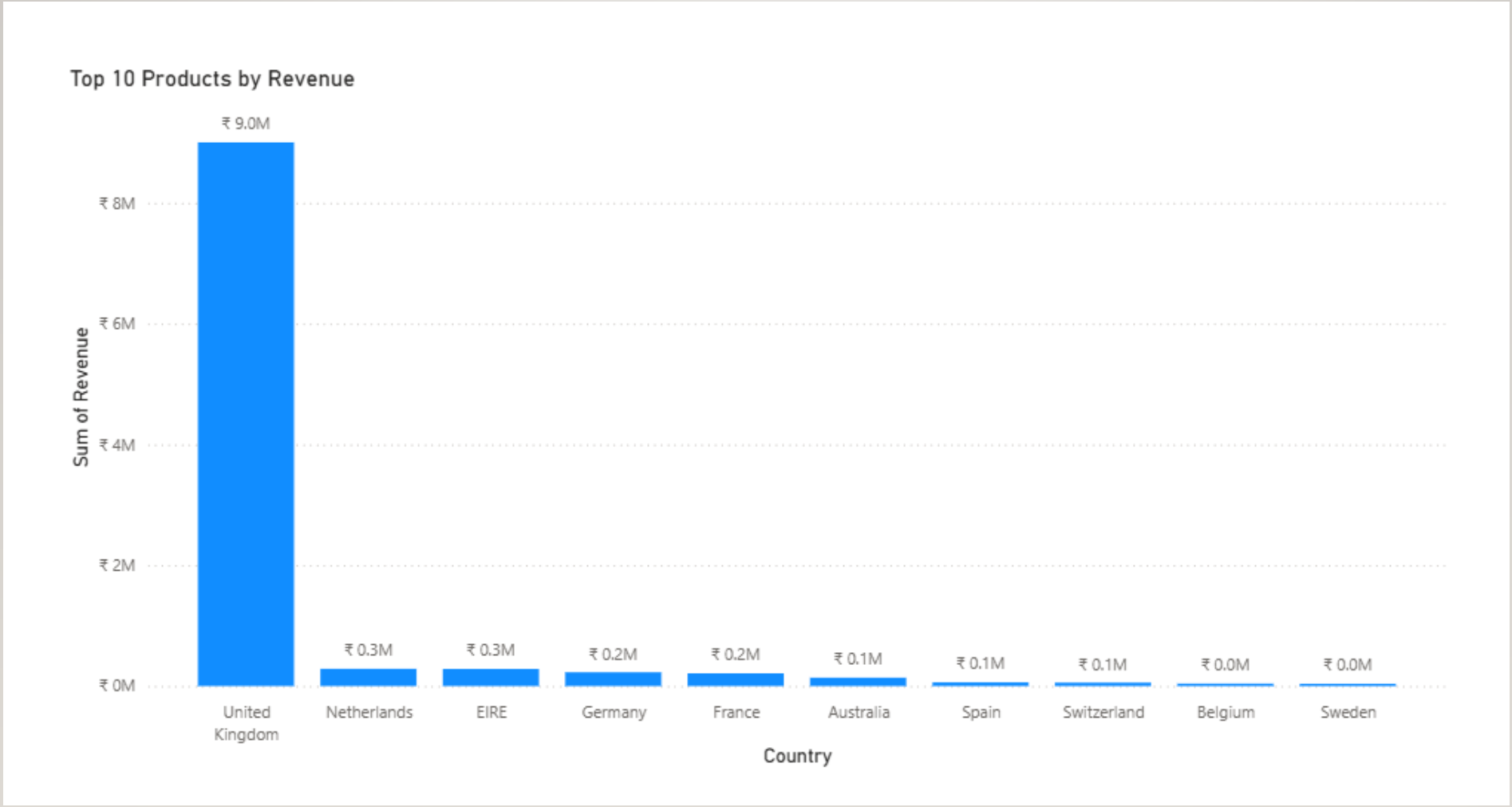

01 November 2011
Sum of Revenue 13,40,417.58
1.4M
1.2M
1.0M
0.8M
0.6M
0.4M
0.2M
0.0M
Jan 2010
Apr 2010
Jul 2010
Oct 2010
Jan 2011
Apr 2011
Jul 2011
Oct 2011
Month


1.4M
1.2M
1.0M
0.8M
0.6M
0.4M
0.2M
0.0M
Sum of Revenue
Jan 2010
Apr 2010
Jul 2010
Oct 2010
Jan 2011
Apr 2011
Jul 2011
Oct 2011
Month
01 December 2011
Sum of Revenue 2,22,951.62

TOP 10 PRODUCTS BY REVENUE



TOP COUNTRIES BY REVENUE



RFM ANALYSIS

What is RFM?

RFM = Recency, Frequency, & Monetary

- Recency (last purchase)
- Frequency (how often they buy)
- Monetary (total spending).

It helps categorize customers into meaningful segments for better marketing and retention strategies.

RFM CALCULATION

To prepare for customer segmentation, RFM values was calculated for each customer:

Recency - Days since the customer's last purchase.
Frequency - Total number of invoices per customer.
Monetary - Total revenue generated by each customer.

Customer-ID	Recency	Frequency	Monetary
17850	30	45	12,500
13047	12	20	4,800
12583	90	10	1,200

Assigning R, F, and M Scores

Each customer's Recency, Frequency, and Monetary values were converted into scores from 1 to 5 using quartiles. Higher scores represent better performance in each category.

CustomerID	Recency	Frequency	Monetary	R-Score	F-Score	M-Score	RFM-Score
12347	41	7	₹ 4,310	4	5	5	455
12356	24	3	₹ 2,811	4	3	5	435
12349	20	1	₹ 1,758	4	1	4	414
12352	74	8	₹ 2,506	3	5	5	355
12348	77	4	₹ 1,797	3	4	4	344

RFM Segmentation

Champions - Most valuable customers, recent buyers, frequent purchases, high spenders.

Loyal Customers - Regular buyers who frequently purchase but may not spend as much as champions.

Potential Loyalists - New or moderately active customers showing signs of becoming loyal.

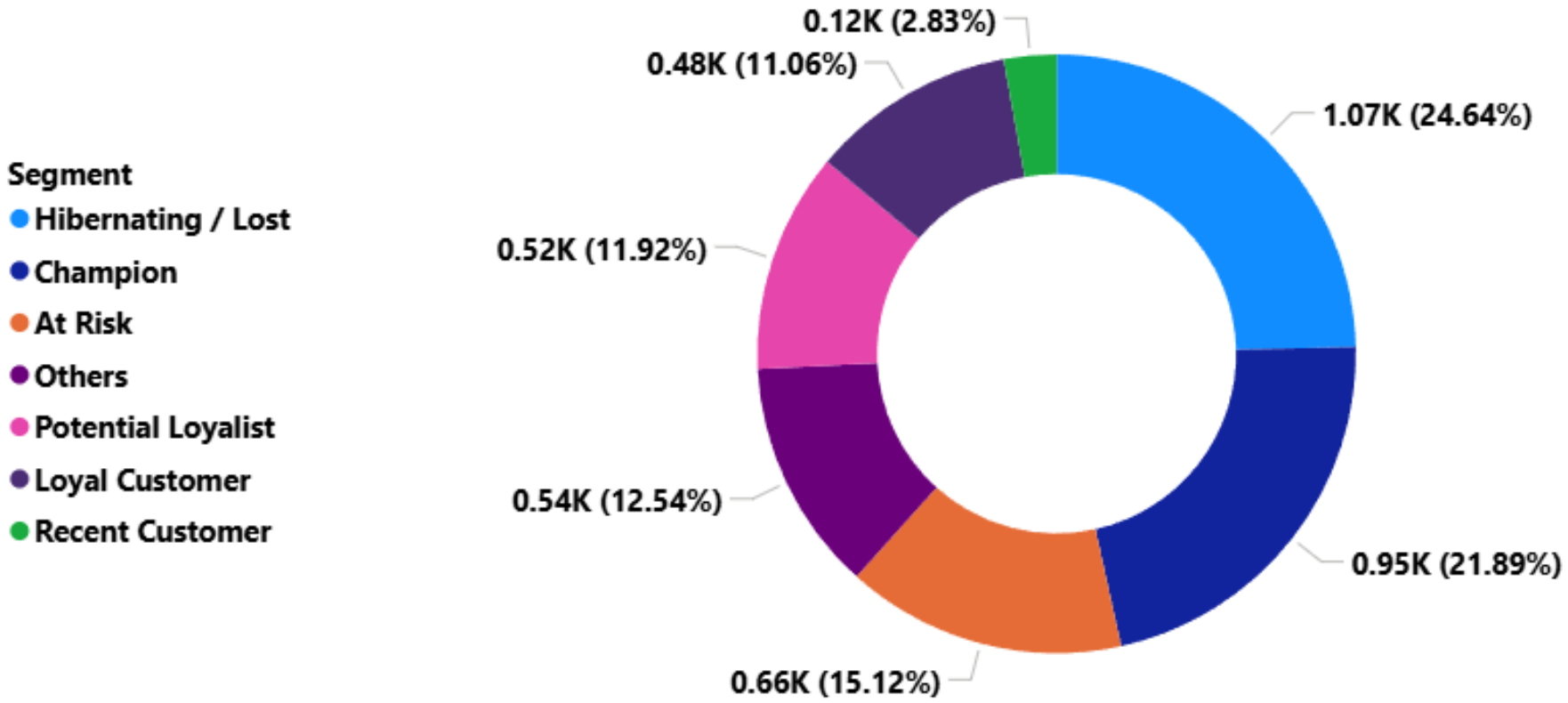
Recent Customers - Made purchases recently but haven’t built loyalty yet.

At Risk - Previously active customers who haven’t purchased recently.

Hibernating / Lost - Very old customers with long inactivity and low engagement.

Others - Customers who don’t strongly fit any specific category.

CustomerID by Segment



Key Insights

Hibernating/Lost is the largest group → needs reactivation campaigns.

Champions make up ~950 key buyers → highest priority to retain.

Potential Loyalists offer strong growth opportunities.

At Risk customers need engagement or discount strategies to prevent churn.

Business Suggestion

Win back Hibernating/Lost Customers - Email campaigns, personalized product suggestions, discount coupons.

Retain Champions with VIP treatment - Early access, premium support, exclusive products, reward points.

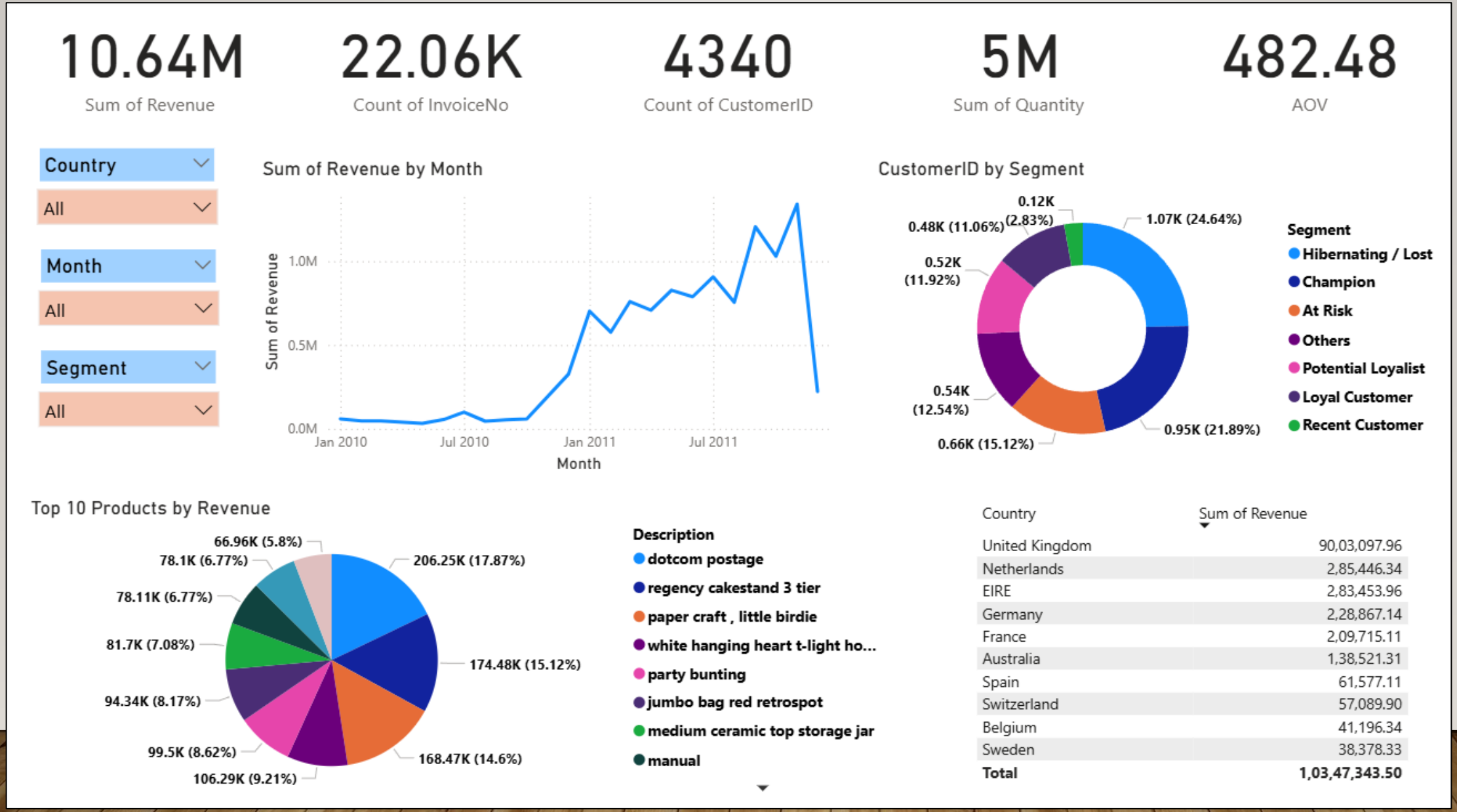
Recover At Risk customers - Reminder emails, special win-back offers, limited-time discounts.

Convert Potential Loyalists into Loyal Customers -Bundle offers, targeted recommendations, loyalty program enrollment.

Strengthen Loyal Customers - Reward their consistency with cashback, referral bonuses.

Engage Recent Customers- Tutorials, thank-you emails, personalized next-purchase suggestions.

Interactive Power BI Dashboard



Conclusion

November 2011 recorded the highest revenue peak across the entire dataset.

The United Kingdom contributed over 85% of total revenue, dominating international markets.

DOTCOM POSTAGE and home décor items contributed heavily to product-level revenue.

Customer activity increased significantly throughout 2011 compared to 2010.

RFM shows a large number of Hibernating customers, indicating opportunities for reactivation.

Champions and Loyal Customers together form a strong revenue foundation.

Potential Loyalists show promising buying patterns that can be nurtured for future growth.

Project Summary

Completed full data cleaning and transformation in Python.

Built KPIs, trends, and segmentation using statistical techniques.

Created end-to-end interactive Power BI dashboard.

Delivered actionable insights for marketing, retention, and revenue growth.

Demonstrated capability in data analysis, visualization, and business understanding.

Thank You