CRM Data Cleaning and Analysis for an Online Programming School

A comprehensive project to improve operational efficiency through in-depth data analysis.

Liliia Yakunina

- **Data Analyst**

- **15 years** of experience in **scientific analysis** (chemistry)

- **Transition** to Data Analytics (2025)

- Strong **analytical** and logical **thinking**

- Professional Principle: *"Data is the key to meaningful decisions."*

- Languages: Ukrainian (native), Russian (native), English (B1), German (B2)

Project Goal

Improving Operational Efficiency of an Online School

Cleaning and analyzing data from the CRM system to enhance the operational efficiency of an online programming school.

Data Preparation

Cleaning data to eliminate noise.

Data Analysis

Researching statistics, trends, campaigns, sales, payments, and geography.

Optimization and Growth

Identifying growth points and optimizing processes using data insights.

Python DA Task: Campaign and Sales Effectiveness

Campaign Effectiveness

Comparison of various campaigns by lead generation and conversion rate.

Marketing Sources

Assessment of the effectiveness of sources (Source) in generating quality leads.

Sales Department Effectiveness

Evaluation of deal owners and campaigns by the number of processed deals, conversion, and sales amount.
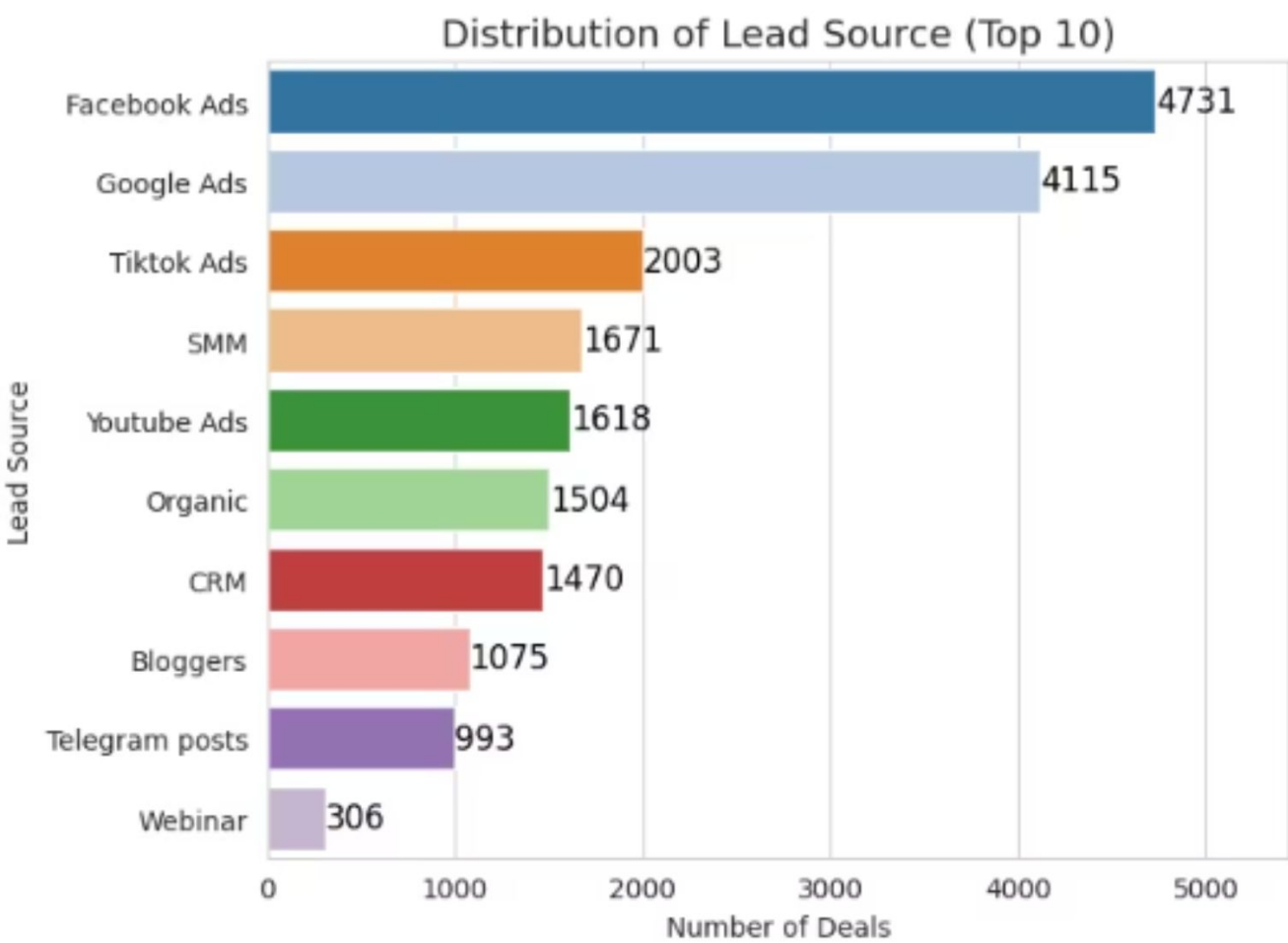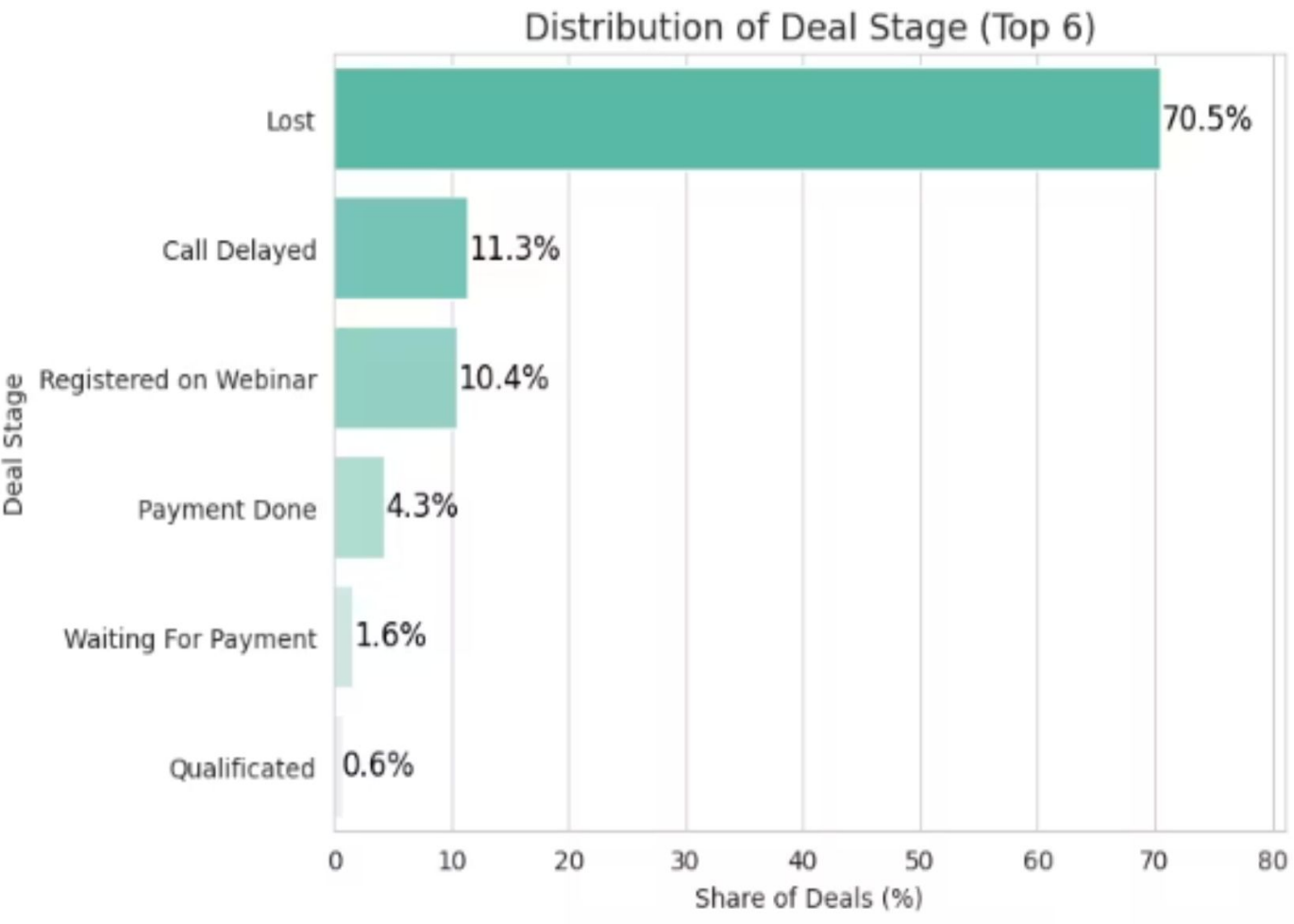
Data Imbalance:

Why Context Matters

Metric	Mean	Median	25th Percentile	75th Percentile
Call Duration	164.8 sec	8 sec	4 sec	97 sec
SLA Response Time	16,736 sec	3,573 sec	0 sec	27,676 sec
Campaign Expenses	7.53 €	0.74 €	0 €	6.16 €

Conclusion:

Our data is highly imbalanced: most metrics are concentrated at **low values**, while rare outliers significantly distort the average. Therefore, for decision-making, we will rely on the **median** and **quartile values**.

Key Insights: Deal Stages and Lead Sources

Distribution of Deal Stage (Top 6)

Deal Stage	Share of Deals (%)
Lost	70.5%
Call Delayed	11.3%
Registered on Webinar	10.4%
Payment Done	4.3%
Waiting For Payment	1.6%
Qualificated	0.6%

Distribution of Lead Source (Top 10)

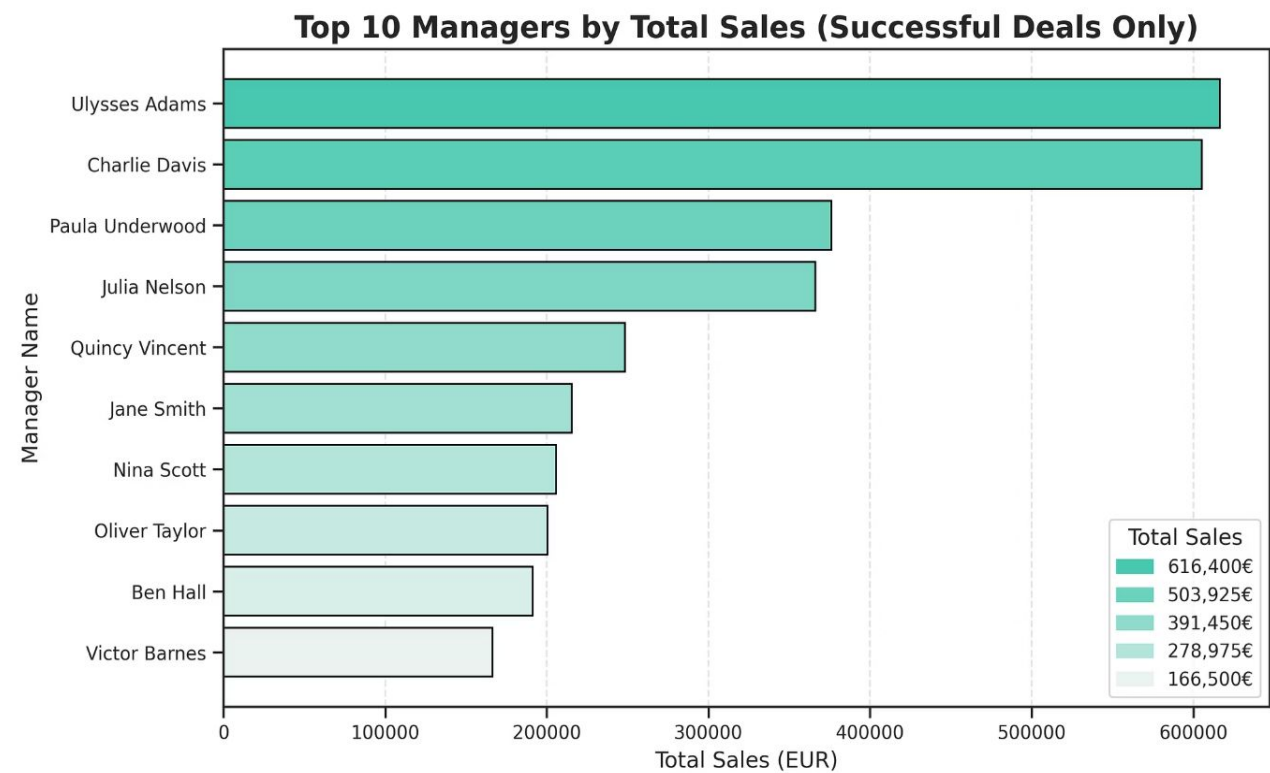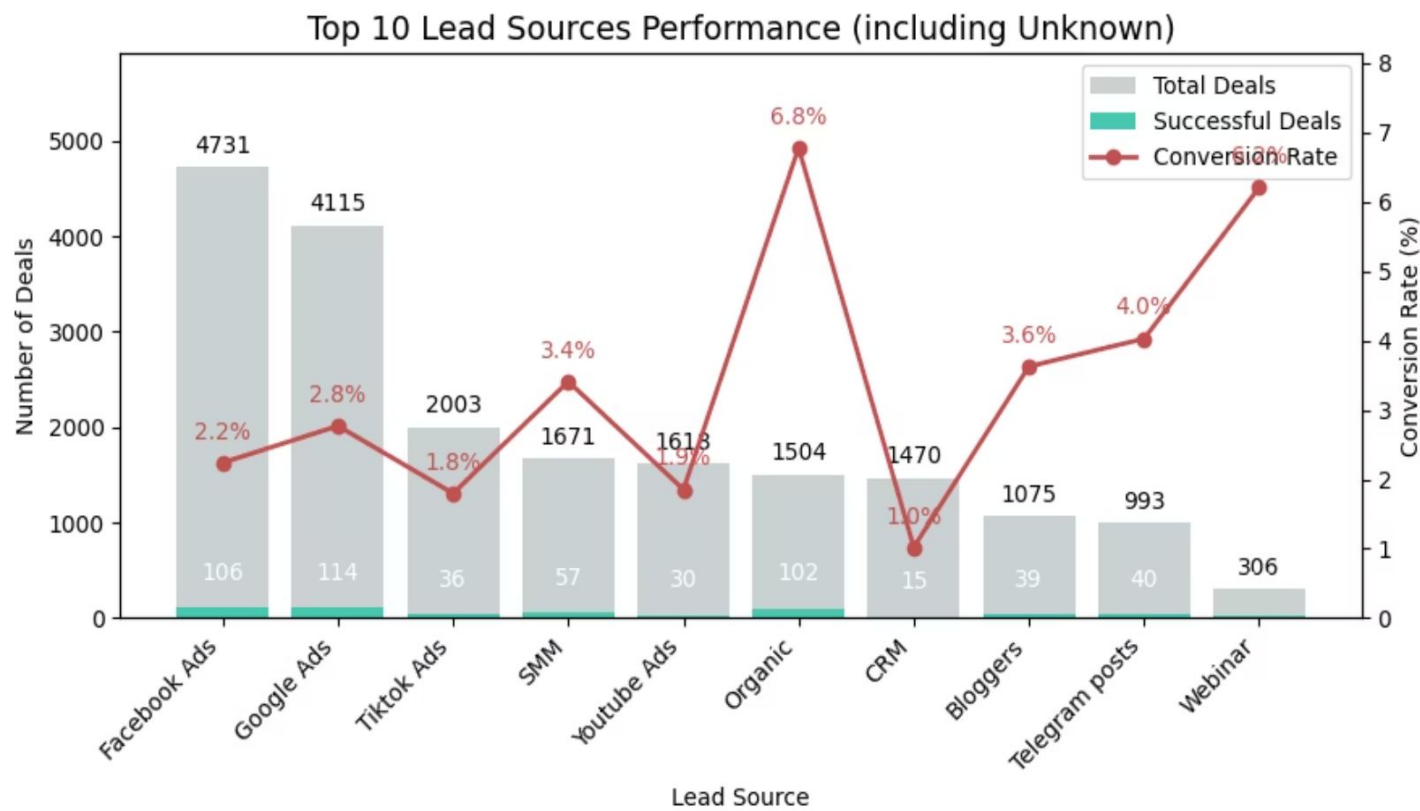| Lead Source	Number of Deals
Facebook Ads	4731
Google Ads	4115
Tiktok Ads	2003
SMM	1671
Youtube Ads	1618
Organic	1504
CRM	1470
Bloggers	1075
Telegram posts	993
Webinar	306

Conclusion:

 As the diagram shows, approximately **70% of deals** are **lost** at various stages, indicating the need to optimize lead nurturing and deal closing processes. The largest share of losses occurs at the "Lost" stage, which requires attention to initial interactions.

Deal Dynamics: how we grow

Number of Deals and Calls Over Time
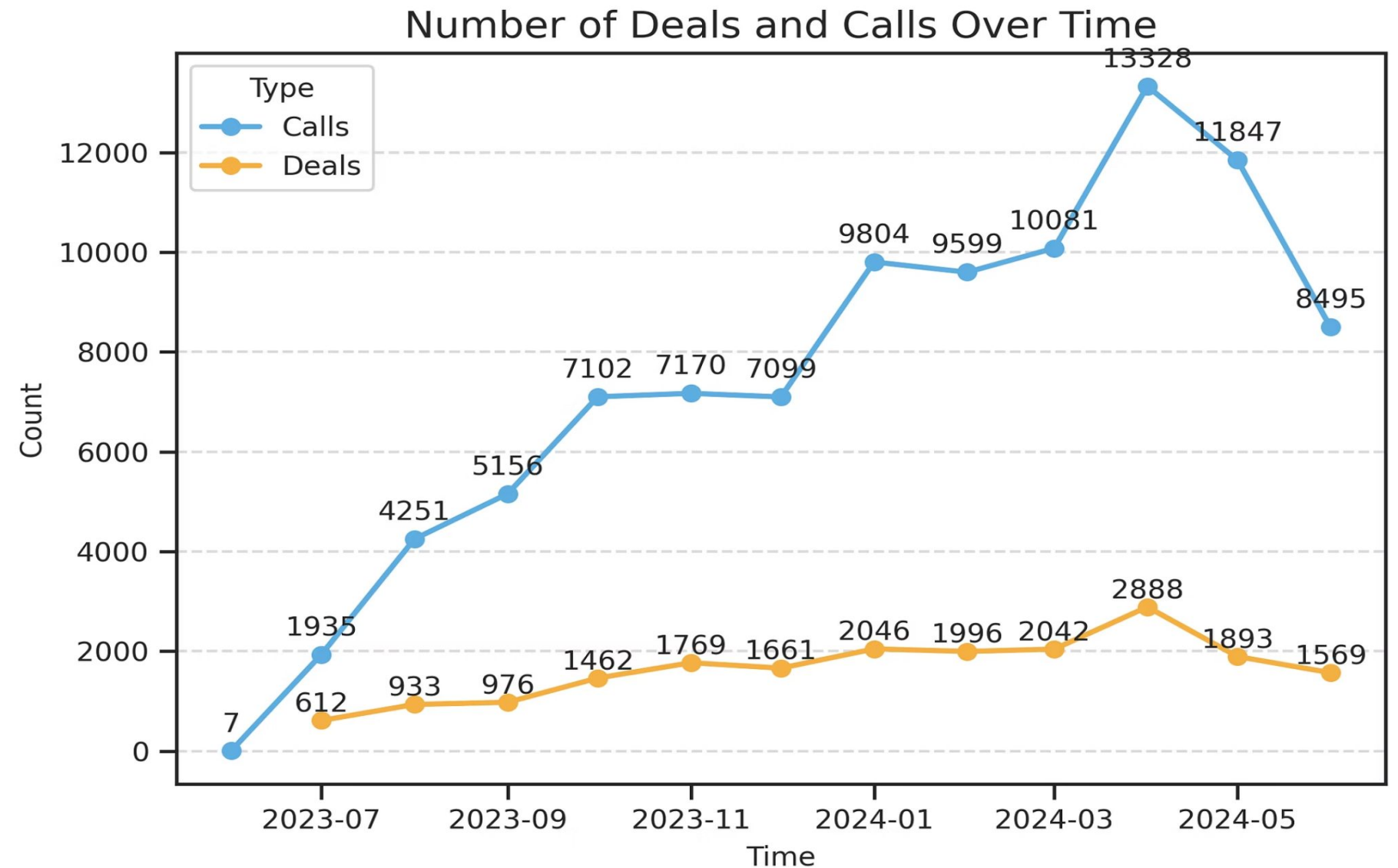
Conclusion: The graph clearly demonstrates a steady **growth** in the number of **deals** and calls from **July 2023** to a peak in **April 2024**. A strong correlation is observed between the number of calls and the number of deals created, which confirms the **high** efficiency of the **sales and marketing departments**. This clear growth is a strong and positive trend.

# Key Success Factors: campaign and manager analysis

Top 10 Lead Sources Performance (including Unknown)

Legend:
- Total Deals
- Successful Deals
- Conversion Rate

Lead Source	Total Deals	Successful Deals	Conversion Rate
Facebook Ads	4731	106	2.2%
Google Ads	4115	114	2.8%
Tiktok Ads	2003	36	1.8%
SMM	1671	57	3.4%
Youtube Ads	1613	30	1.9%
Organic	1504	102	6.8%
CRM	1470	15	1.0%
Bloggers	1075	39	3.6%
Telegram posts	993	40	4.0%
Webinar	306		6.2%

Top 10 Managers by Total Sales (Successful Deals Only)

Managers (top to bottom): Ulysses Adams, Charlie Davis, Paula Underwood, Julia Nelson, Quincy Vincent, Jane Smith, Nina Scott, Oliver Taylor, Ben Hall, Victor Barnes

Total Sales legend:
- 616,400€
- 503,925€
- 391,450€
- 278,975€
- 166,500€

Conclusion:

To increase overall business efficiency, we need to focus on **optimizing advertising expenses** (reallocating the budget towards higher-converting channels such as Telegram and webinars) and re-distributing the workload among managers to ensure an even load.

What We Sell: Product Popularity and Payment Types

Distribution of Successful Deals by Product

■ Digital Marketing   ■ UX/UI Design   ■ Web Development
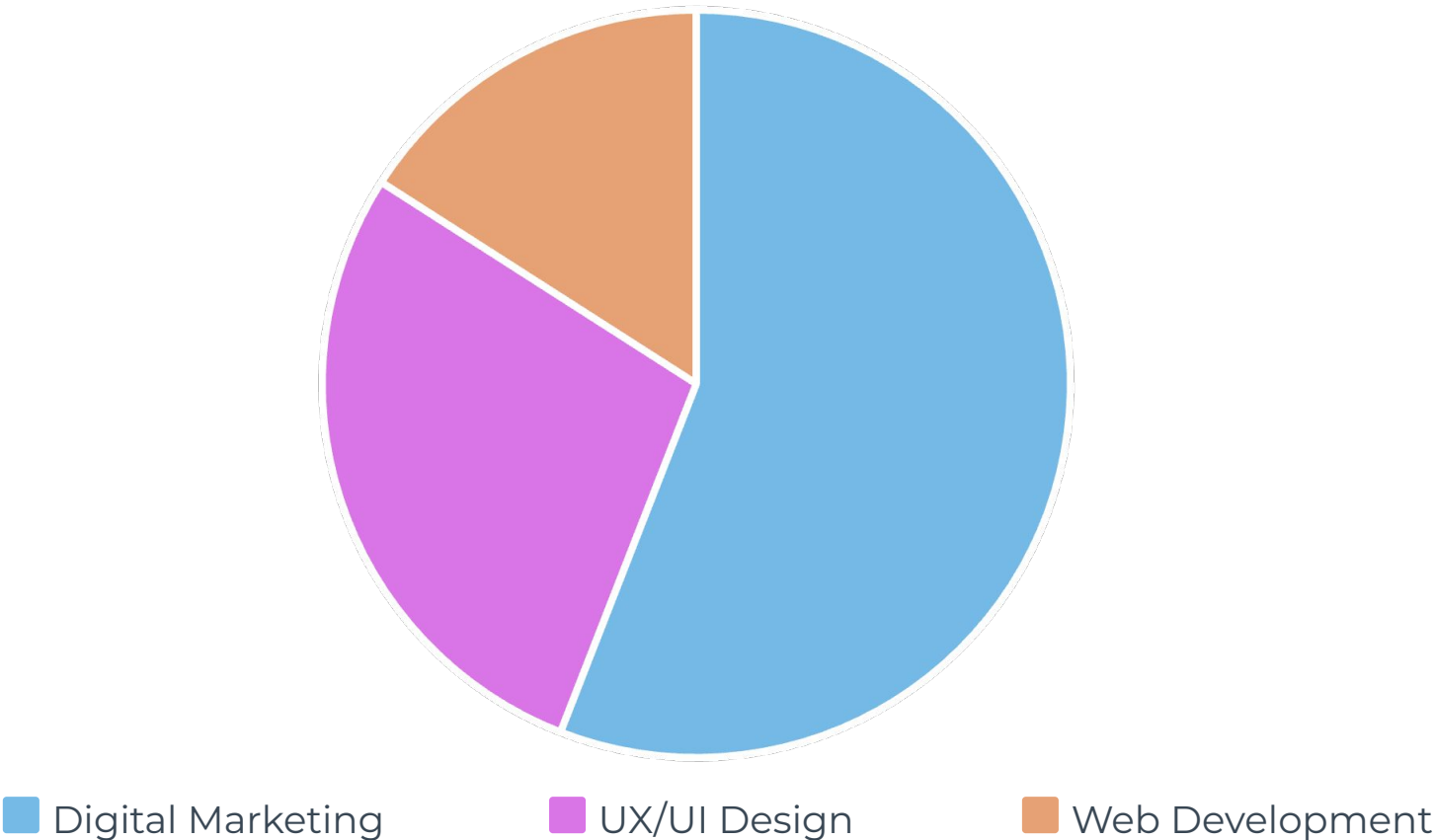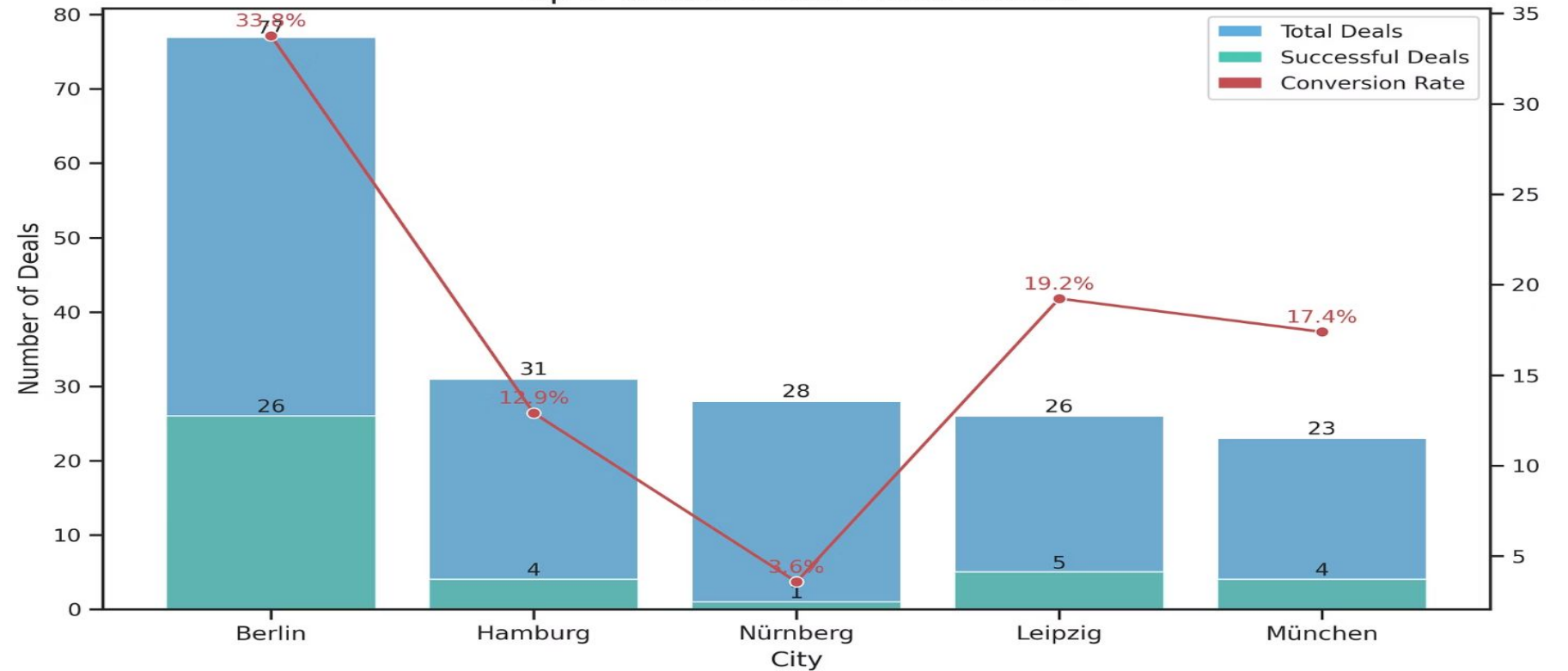
Conclusions:

Digital Marketing is our most popular product, accounting for over **55%** of successful deals.

We found that the **morning study type** shows higher success rates.

Analysis of payment types revealed that almost **98% of the data is missing**. This is a significant gap in data collection that needs to be addressed to allow future analysis of how payment type affects sales.

Deal Geography:
Where Our Clients Are

Conclusions:

Berlin is not only the largest market by number of deals but also shows an **exceptionally high conversion rate (33.8%)**, making it a key city for the business.

Leipzig and **Munich** also demonstrate high conversion rates despite a smaller number of deals. This indicates their high potential for development.

Further analysis shows that **German language proficiency** directly influences deal success: clients with **B1** and **B2** levels are more likely to complete a purchase.

Product Analytics Task

Find a business growth point and formulate a hypothesis for improving the business process to increase metrics.

01

Unit Economics
Calculating unit economics by product.

02

Growth Points
Identifying business growth points based on unit economics.

03

Metrics Tree
Building a metrics tree for the business.

Product Analytics: Hypotheses and Testing

1

Hypothesis Formulation

Defining the product metric to be impacted and formulating hypotheses.

2

Verification Method

Describing the method for testing hypotheses, including conditions for execution.

3

Testing Period

The test should not take more than 2 weeks.

Channel Effectiveness: Who Brings in Profit

Here are the main unit economics metrics for analyzing product effectiveness:

Источник	Общая Выручка (€)	Средний Чек (AOV) (€)	Стоимость Привлечения (CPA/CAC) (€)	ROI (%)
Google Ads	712,400	6,249.12	507.00	1132.56
Facebook Ads	670,100	6,321.70	318.44	1885.20
Telegram posts	31,100	7,775.00	171.51	4433.29
Webinars	121,000	6,368.42	151.27	4110.10
SMM	292,400	5,129.82	127.54	3922

ROI by Lead Source: Investment Efficiency
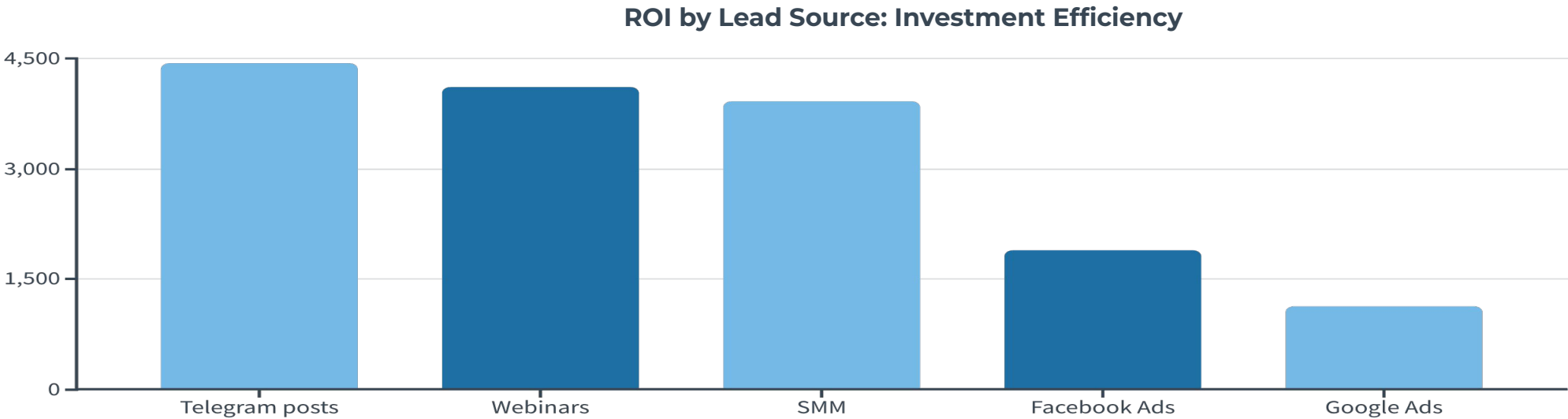
Conclusions:

To maximize profitability, we should reallocate our marketing budget to high-ROI channels like **Telegram** and **Webinars**.

Final Conclusions: An Action Plan for Growth

Key Findings

Growth Dynamics: The business shows stable growth in deal volume, which directly correlates with the sales team's activity.

Core Product: Digital Marketing is the **key produc**t, generating the majority of revenue.

Channel Effectiveness: There is a significant difference in efficiency. Telegram posts and Webinars have the highest ROI, while Google Ads and Facebook Ads drive large volume but with lower profitability.

Funnel Issues: 70% of deals are lost at various stages, and data on reasons for loss and payment types is poorly collected.

Geography & Clients: Berlin is a key market, and clients with **B1/B2 German** language proficiency show a **higher conversion rate.**

Recommendations

Optimize Marketing Budget: Reallocate a portion of the budget from **Google Ads and Facebook Ads** to high-ROI channels like Telegram posts and Webinars.

Improve Sales Team Efficiency: Establish a process for collecting data on lost deals and analyze the workload to ensure a more even distribution among managers.

Enhance Data Quality: Begin collecting complete information on reasons for loss and payment types to enable more accurate future forecasting.

Focus on Key Markets: Direct efforts toward high-conversion regions (e.g., **Berlin**) and to attracting clients with **B1/B2** German language proficiency.

Thank you!

Do you have any questions?