Detailed report outlining the analysis process, findings, visualization, and recommendations in a comprehensive manner

Project Goal: Cleaning and analyzing data from the CRM system to improve the efficiency of the online programming school.

Data cleaning and preparation:

Description of Actions Performed in Cleaning data_Deals.ipynb and Cleaning data_Calls,contacts,spend.ipynb Notebooks

Cleaning data_Deals.ipynb processes the Deals (Done).xlsx dataset with 21,592 deal records containing columns like Id, Contact Name, Stage, Initial Amount Paid, Offer Total Amount, Lost Reason, and SLA.

- The dataset is cleaned and saved as Deals (Final).xlsx with 21,411 rows. The dataset is loaded with Id and Contact Name as strings.
- Created Time and Closing Date are converted to datetime64[ns] with day-first format.
- Course duration and Months of study are converted to Int8, and columns like Deal Owner Name and Stage are set to category type.
- Duplicates marked as "Duplicate" in Lost Reason are aggregated by keeping the earliest Source and last non-null values, reducing rows by 181.
- Quality column values of F are replaced with E - Non Qualified, and 2,253 missing values are filled with E - Non Qualified.
- Currency columns (Initial Amount Paid, Offer Total Amount) have euro symbols and separators removed, are converted to float64, and swapped if Initial Amount Paid exceeds Offer Total Amount; missing values are filled with 0.0.
- SLA is converted to SLA in Seconds, with missing values filled with 0.0, and the original SLA column is dropped.
- Missing City and Level of Deutsch values are filled with the mode per Contact Name, with Level of Deutsch standardized (e.g., в1 to B1, 19,756 remain Not defined).
- For 3,288 rows with Initial Amount Paid > 0, Stage is set to Payment Done and Lost Reason to No reason.
- Education Type missing values (18,112) are filled with the mode based on Offer Total Amount, Course duration, and Product, with 12 remaining filled with Unknown.
- Payment Type is reclassified into No Payments (18,067 rows), No Initial Payments (56), Recurring Payments (3,168), One Payment (115), or Reservation (5).
- Closing Date earlier than Created Time (49 rows) is swapped.
- Columns Content, Page, Term, and Created Date Only are dropped.
- Remaining missing Lost Reason values (4,087) are filled with Not lost.
- The final dataset is saved without the index.

Cleaning data_Calls,contacts,spend.ipynb processes Contacts (Done).xlsx (18,548 rows), Spend (Done).xlsx (20,779 rows), and Calls (Done).xlsx (95,874 rows), saving cleaned versions as Contacts (Final).xlsx, Spend (Final).xlsx, and Calls (Final).xlsx.

- For Contacts, Id is loaded as a string, Created Time and Modified Time are converted to datetime64[ns], and 38 duplicates are removed based on all columns except Id, reducing to 18,510 rows. The dataset is saved without the index.
- For Spend, Date is already datetime64[ns]. Source, Campaign, and AdGroup are converted to category type, with missing Campaign (5,994) and AdGroup (6,828) values filled with Unknown.
- Rows with zero Impressions, Spend, and Clicks (4,372) are removed, and 7 duplicates are dropped, reducing to 16,393 rows. The Ad column is dropped, and the dataset is saved.

- For Calls, Id and CONTACTID are loaded as strings, Call Start Time is converted to datetime64[ns], and Call Owner Name, Call Type, Call Status, Outgoing Call Status, and Scheduled in CRM are set to category type.
- Missing Outgoing Call Status and Scheduled in CRM values (8,999 each) are filled with Not outgoing call.
- Duplicates (3,257) are removed based on all columns except Id, reducing to 92,617 rows.
- Dialled Number and Tag columns, with zero non-null values, are dropped.
- Missing CONTACTID (3,802) and Call Duration (79) values remain unfilled. The final dataset is saved without the index.
-

Summary:

Deals reduced from 21,592 to 21,411 rows, Contacts from 18,548 to 18,510, Spend from 20,779 to 16,393, and Calls from 95,874 to 92,617.
Key actions include standardizing data types (datetime, category, Int8), removing duplicates, handling missing values (e.g., Unknown, Not outgoing call, 0.0), correcting logical errors (e.g., swapping dates, currency), and dropping irrelevant columns (e.g., Ad, Dialled Number).
The cleaned datasets are consistent, with minimal missing values in critical columns, ready for analysis.

Descriptive statistics:

The Descriptive statistics.ipynb notebook performs descriptive statistical analysis on four cleaned datasets: Contacts (Final).xlsx, Calls (Final).xlsx, Spend (Final).xlsx, and Deals (Final).xlsx. It calculates summary statistics, visualizes data distributions, and provides insights into contact management, marketing spend, call durations, and deal performance.

Additional cleaning is applied to the Deals dataset. Zero values in numeric columns (Initial Amount Paid, Offer Total Amount, Course duration, Months of study) are replaced with NaN to mark them as missing, ensuring accurate statistical calculations. Temporary display columns are created to label missing values as 'Unknown' and round numeric values for visualization, then dropped to retain original columns.

Summary statistics (count, mean, min, 25th percentile, median, 75th percentile, max, standard deviation) are calculated for numeric fields in the Deals, Contacts, and Spend datasets using pandas' describe() method. For Calls, a histogram of Call Duration (in seconds) is generated using matplotlib, with a range of 0 to 1,000 seconds to exclude outliers, visualizing the distribution of call lengths.

Detailed reports are provided in markdown cells for Deals, Contacts, and Spend, summarizing key metrics and offering insights.

Deals Analysis:

- The Deals dataset (21,411 rows) includes columns like Closing Date, Created Time, Course duration, Months of study, Initial Amount Paid, Offer Total Amount, and SLA in Seconds.
- Closing Date has 14,495 non-null entries, with an average of January 28, 2024, ranging from July 3, 2023, to December 11, 2024.
- Created Time, with 21,411 entries, averages January 27, 2024, from October 11, 2022, to June 21, 2024, indicating a one-day average lag between creation and closing.
- Course duration (3,587 non-null) averages 10.2 months, with most values at 11 months (standard deviation 1.83).
- Months of study (839 non-null) averages 5.45 months, ranging from 1 to 11 (standard deviation 2.92).
- Initial Amount Paid (3,288 non-null) averages €1,178.71, with a median of €1,000 and a maximum of €11,000 (standard deviation €1,416.54).

- Offer Total Amount (3,338 non-null) averages €8,966.21, with a median of €11,000 and a maximum of €11,500 (standard deviation €3,254.91).
- SLA in Seconds (21,411 non-null) averages 83,174 seconds (~23.1 hours), with a median of 6,516 seconds (~1.81 hours) and a maximum of 26,908,464 seconds (~311 days), indicating significant outliers.

Insights from Deals:

- Deal creation and closing dates are closely aligned, with most activity between November 2023 and April 2024, suggesting a six-month peak period.
- Course durations are standardized at 11 months, but months of study vary, indicating diverse engagement levels.
- Initial payments are typically €1,000, reflecting a standard deposit system, while offer totals cluster at €11,000, with some variability suggesting different pricing tiers.
- SLA times are generally efficient, but extreme outliers (up to 311 days) point to process inefficiencies or data anomalies requiring investigation.
- The high standardization in course duration and payments contrasts with variability in SLA, highlighting areas for workflow optimization.
-

Contacts Analysis:

- The Contacts dataset (18,510 rows) tracks Created Time and Modified Time.
- Created Time ranges from June 27, 2023, to June 19, 2024, with a mean of January 24, 2024, and a median of February 1, 2024.
- Modified Time ranges from July 6, 2023, to June 19, 2024, with a mean of February 14, 2024, and a median of February 28, 2024.
- The main activity period for both is November 2023 to April 2024, with 50% of contacts created and modified during this time.
-

Insights from Contacts:

- Contact creation and updates are concentrated between November 2023 and April 2024, indicating a high-activity period likely tied to marketing campaigns.
- The median creation date of February 2024 shows steady growth, while the slightly later modification average suggests active follow-up.
- Large gaps between creation and modification may indicate delayed engagement, recommending a review of these contacts. Auditing stale contacts for re-engagement or cleanup could improve CRM efficiency.

Spend Analysis:

- The Spend dataset (7,595 rows) includes Date, Impressions, Spend, and Clicks, covering July 1, 2023, to June 14, 2024.
- Impressions average 3,116, with a median of 25, a maximum of 431,298, and a standard deviation of 12,830, indicating significant variability.
- Spend averages $9.12, with a median of $0.25, a maximum of $774, and a standard deviation of $29.83, showing skewed distribution.
- Clicks average 30.4, with a median of 0, a maximum of 2,415, and a standard deviation of 95, highlighting outliers.
- The bulk of activity occurs between October 2023 and April 2024.

Insights from Spend:

- Marketing activity spans nearly a year, with a peak from October 2023 to April 2024.
- High standard deviations in impressions, spend, and clicks indicate a few high-impact campaigns amidst mostly low-activity days.

- The median spend of $0.25 and 25% of days with zero clicks suggest many low-budget or ineffective campaigns.
- Investigating high-spend, high-click days could reveal successful strategies, while low-click days may indicate wasted impressions, recommending creative optimization.

Calls Analysis:

- The Calls dataset (92,617 rows) includes Call Duration (in seconds).
- The histogram shows most calls are short, with durations clustered below 500 seconds, and a few longer calls up to 1,000 seconds.
- The distribution is right-skewed, indicating that most interactions are brief.

Insights from Calls:

- The predominance of short calls suggests quick, transactional interactions, possibly initial inquiries or follow-ups.
- Longer calls, though rare, may represent deeper engagements or complex issues.
- Analyzing call outcomes by duration could identify patterns in effective communication, and reviewing outliers may uncover training needs or process bottlenecks.

Overall Summary and Recommendations:

The notebook provides a comprehensive statistical overview of the datasets, highlighting standardized processes in deals (e.g., €1,000 deposits, 11-month courses) and continuous activity periods (July 2023 to June 2024) across datasets, with a peak from November 2023 to April 2024.

Outliers in SLA times, marketing spend, and call durations suggest inefficiencies or exceptional cases warranting further investigation.

Recommendations include analyzing contact creation trends to assess campaign impact, auditing stale contacts for CRM cleanup, optimizing low-performing ad campaigns, and investigating call outliers to improve engagement efficiency.

The data reveals a structured operation with opportunities to address inefficiencies and enhance performance through targeted analysis.

Analyze time series:

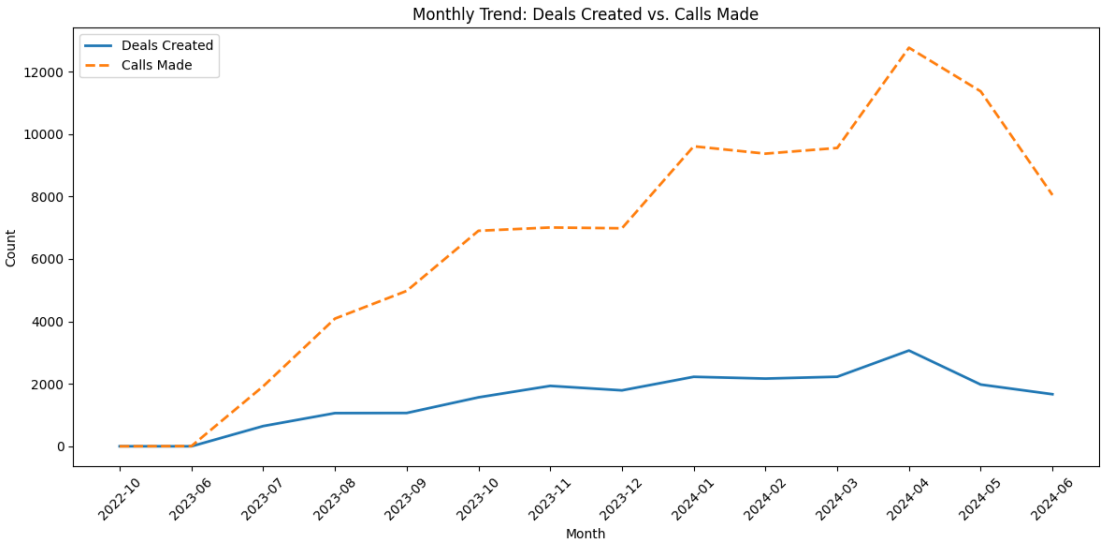
1. *Analyze the trend of deal creation over time and its relationship to calls.*

Deal Creation:

- Deals are grouped by Created Time (converted to monthly periods, e.g., 2024-03) to count the number of deals created each month.
- The resulting time series spans October 2022 to June 2024, with sparse data in 2022 and late 2024.

Call Activity:

- Calls are grouped by Call Start Time (converted to monthly periods) to count the number of calls per month.
- The call time series covers June 2023 to June 2024, with outliers in September and December 2024.



Trend of Deal Creation and Calls

Peak Activity:

April 2024 marks the highest deal creation (3,068) and call volume (12,767), driven by strong sales efforts or marketing campaigns. The decline in June 2024 (1,668 deals, 8,049 calls) suggests seasonal or operational factors.

Strong Correlation:

The ~0.89 correlation between calls and deal creation confirms that call activity is a key driver of deal generation. Higher call volumes consistently precede or coincide with increased deal creation.

Lag Effect:

Early months (e.g., June 2023: 7 calls, 0 deals; July 2023: 1,923 calls, 648 deals) indicate a lag where calls build pipeline before deals materialize.

Data Gaps:

Sparse data in 2022 (1 deal) and late 2024 (0 calls in September/December) suggests incomplete records, requiring validation.

<u>Insights & Recommendations:</u>

Leverage Peak Periods:

Scale up call center resources and marketing campaigns during high-activity months (November–April), replicating April 2024's success (3,068 deals, 12,767 calls).
Investigate drivers of April 2024's peak (e.g., campaigns, team performance) to inform future strategies.

Enhance Call Efficiency:

Analyze call outcomes (using Calls dataset) to prioritize high-value calls, focusing on types that drive deal creation.
Train reps to reduce the lag between calls and deals, as seen in July 2023 (1,923 calls, 648 deals).

Validate Data:

Address sparse data in 2022 and late 2024 (e.g., 0 calls in September/December 2024) to ensure accurate forecasting.

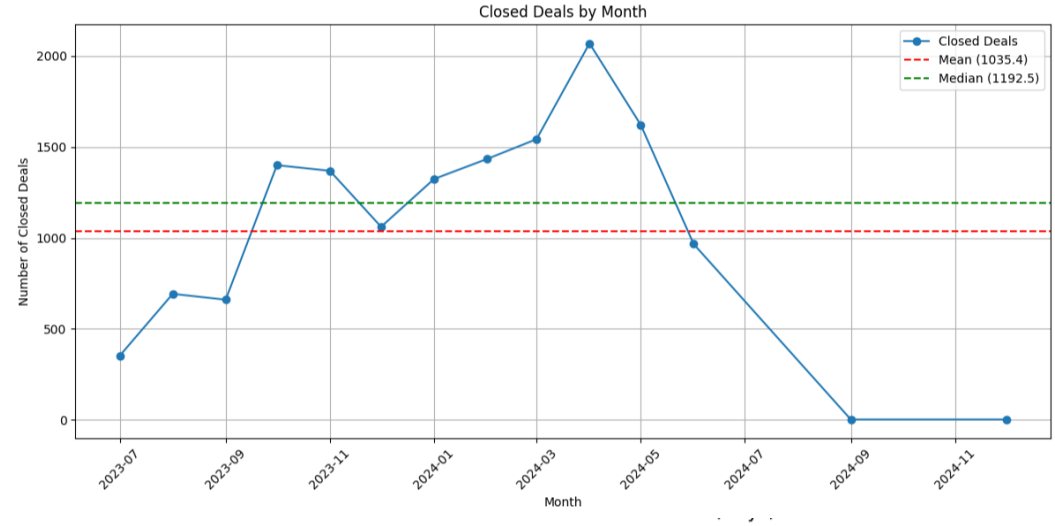
2. *Examine the distribution of deal closing times and the length of time from creation to closing.*

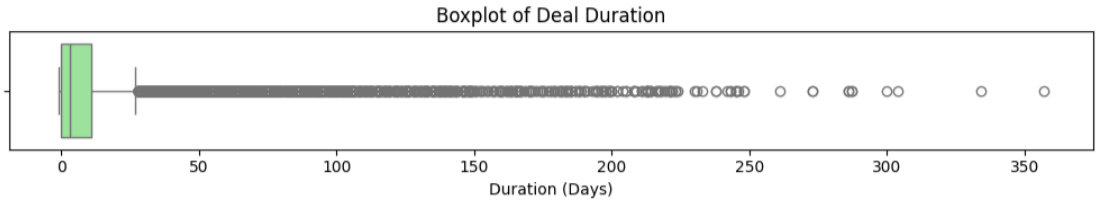
Deal Closing Times:

- Deals with valid Closing Date values (14,495 deals) are grouped by Closing Date (converted to monthly periods) to analyze the distribution of deal closures over time.
- The distribution is examined to identify peak closing months and seasonal patterns.

Creation-to-Closing Duration:

- The duration between Created Time and Closing Date is calculated for each deal (in days) using datetime subtraction.
- The distribution of durations is analyzed using summary statistics (e.g., mean, median, quartiles) and visualized with histograms or box plots.
- Outliers (e.g., negative durations or excessively long cycles) are investigated for data quality issues.





Key Statistics:
Total closed deals analyzed: 14,495
Mean duration: ~14.2 days
Median (50%) duration: 3 days
25% of deals closed on the same day or next day (0 days or less)
75% of deals were closed within 11 days
Maximum duration: 357 days
Minimum duration: -1 day (likely data quality issue)
Standard deviation: 31.8 days (high variation suggests outliers or long-tail cases)

Distribution of Deal Durations:

Fast Closures (0–10 days): 5,714 deals (~39% of all) were closed within 10 days
Moderate Duration (10–30 days): Additional 1,831 deals (~12.6%) closed between 10 and 30 days
Long Duration (30–100 days): 1,228 deals (~8.5%) Very Long Duration (>100 days):
Only 491 deals (just ~3.4%) took longer than 100 days. Very few extended past 200+ days.

Insights & Recommendations:

Most deals are closed quickly:

Over 50% of all deals are finalized within just 3 days of creation, and nearly 75% are closed within 11 days. This indicates an efficient sales process for the majority of leads.

There are extreme outliers:

A small number of deals took over 100 to 350+ days.
These may represent: Complex sales cycles, Dormant leads that were later reactivated, Possible data entry issues (especially the -1 day value)

Data Cleaning Needed:

The minimum value of -1 day is likely an error (e.g., closing date before creation). These should be investigated or filtered out in future analysis.

Segment deals by duration for sales strategy:

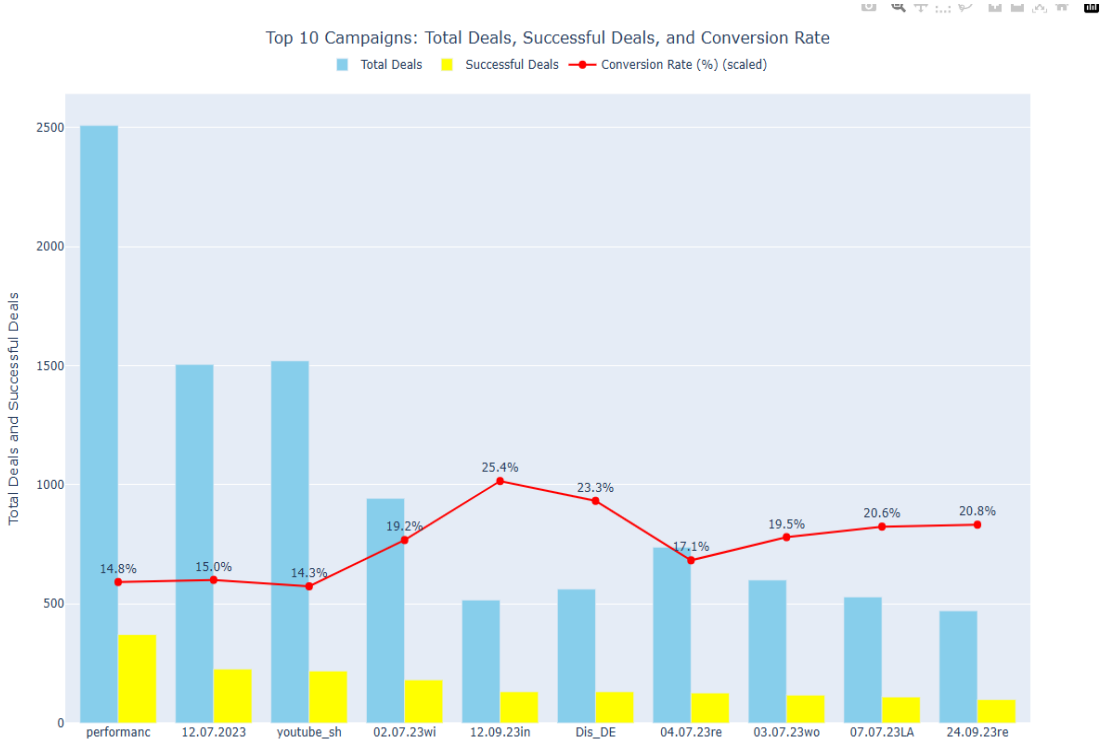
Categorizing deals into fast, moderate, and long-cycle can help optimize pipeline forecasting and identify bottlenecks.

Analyze the effectiveness of campaigns:

1. Compare the effectiveness of different campaigns in terms of lead generation and conversion rate.
Campaign Performance Analysis Overview

An in-depth analysis of the "deals" dataset has been conducted to evaluate the performance of various marketing campaigns. The following report highlights key insights, including top-performing campaigns by volume, conversion rates, and ROI, as well as underperforming campaigns and

actionable recommendations for optimization.



Key Insights from Campaign Performance Analysis

Top-Performing Campaigns by Volume

performancemax_digitalmarkt_ru_DE led with the highest number of total deals (2,508) and successful deals (371), generating over €3.34M in revenue, with a conversion rate of 14.79%.
Campaigns like 12.07.2023wide_DE (1,505 deals, 226 successful, €2.06M revenue) and youtube_shorts_DE (1,520 deals, 218 successful, €2.01M revenue) also showed strong performance in deal volume and revenue.

Highest Conversion Rates

The campaigns with the highest conversion rates were:
web1312_DE: 54.55% (11 deals, 6 successful)
podslush_DE: 50.00% (10 deals, 5 successful)
col_DE: 50.00% (6 deals, 3 successful)
2005_Lost_DE: 50.00% (6 deals, 3 successful)
gen_analyst_DE: 37.14% (35 deals, 13 successful)
18.10.23wide_gos_DE: 36.59% (41 deals, 15 successful)
These campaigns, while lower in volume, demonstrated exceptional efficiency in converting leads into successful deals.

Strong ROI Campaigns (High Revenue per Deal)

Campaigns with significant financial impact per successful deal include:
brand_search_eng_DE: 58 deals → €523,000 → ~€9,017 per deal
Live_DE: 9 deals → €86,500 → ~€9,611 per deal
Smaller campaigns like web2410_DE, Bolgspeak_DE, hanna, and nina each had 1 successful deal yielding €11,000, indicating high unit value despite low volume.

Underperforming Campaigns

Over 30 campaigns generated no successful deals despite varying lead volumes. Examples include:
08.06.24wide_webinar_DE: 111 deals, 0 conversions
01.04.23women_PL: 31 deals, 0 conversions
euro_DE: 32 deals, 0 conversions
Potential reasons for underperformance include:
Misaligned audience targeting
Ineffective ad creatives or landing pages
Low-quality leads

Webinar & Test Campaigns Mostly Ineffective
Campaigns labeled as webinars or tests consistently showed low conversion rates, indicating challenges with engagement or campaign design. Examples:
webinar1604: 333 deals → 7 successes → 2.10% conversion rate
webinar1906: 286 deals → 4 successes → 1.40% conversion rate
08.04.24wide_webinar_DE: 278 deals → 5 successes → 1.80% conversion rate

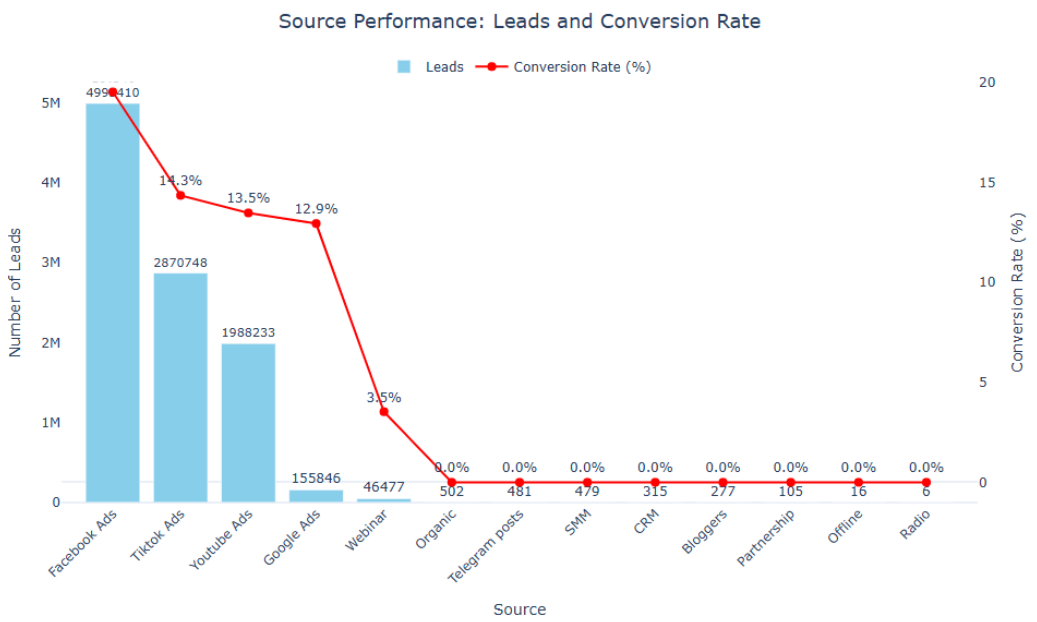
Recommendations

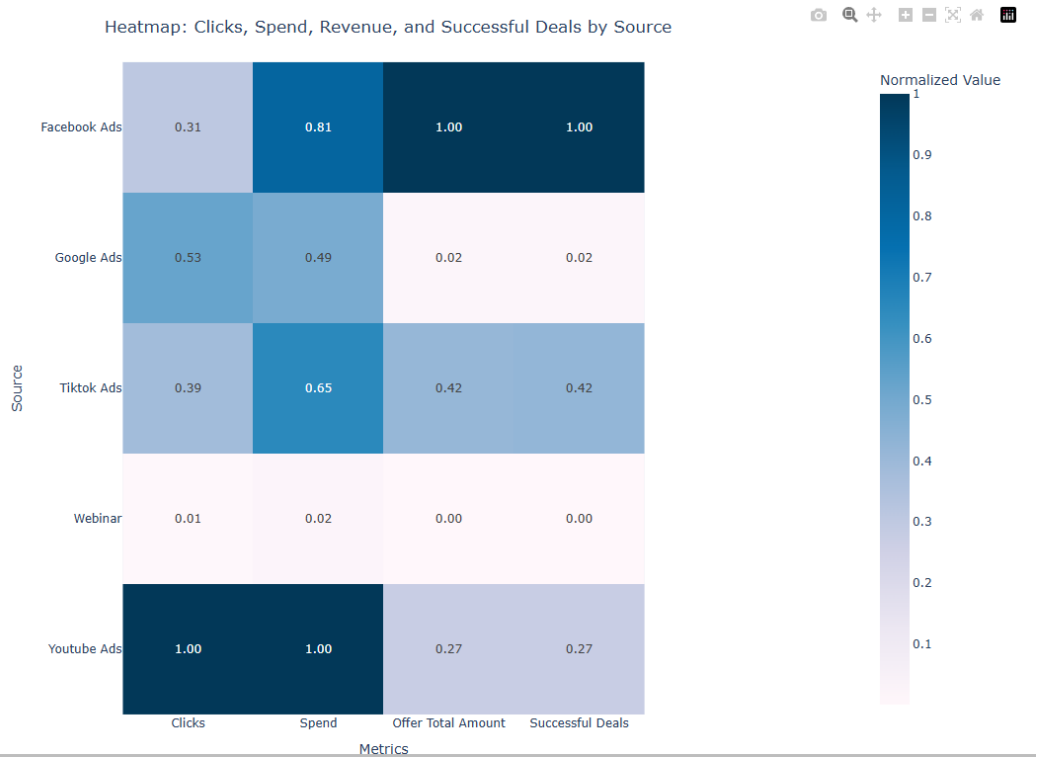
- Scale High-Converting Campaigns: Campaigns with conversion rates above 30% (e.g., web1312_DE, podslush_DE, gen_analyst_DE) should be prioritized for scaling, despite lower volumes, to maximize efficiency.
- Audit and Pause Ineffective Campaigns: Campaigns with zero successful deals and moderate-to-high lead volumes (e.g., 08.06.24wide_webinar_DE, euro_DE) should be paused or restructured to address targeting or creative issues.
- Optimize Webinar Funnels: Low-performing webinar campaigns require improved content, better promotion strategies, or enhanced follow-up processes to boost conversions.
- Leverage High-Value Channels: Campaigns with high revenue per deal (e.g., brand_search_eng_DE, Live_DE) should be further explored to replicate their high-quality outcomes.

2. Evaluate the effectiveness of different marketing sources (Source) in generating quality leads.

The highest number of leads comes from Facebook Ads (4.99M), TikTok Ads (2.87M), and YouTube Ads (1.98M). The highest conversion rates are observed for Facebook Ads (19.52%), YouTube Ads (13.5%), and TikTok Ads (14.3%). Google Ads (155K leads) has a conversion rate (12.94%), which is lower than the top 3 sources. Other sources generate few leads, and Telegram, Bloggers, Organic,

SMM, CRM, Partnership, Offline, and Radio show 0% conversion.





Key Insights from Advertising Source Performance Analysis

Top-Performing Source

Facebook Ads achieved the highest performance in both revenue generation (100% of the maximum Offer Total Amount) and conversions (100% of the maximum Successful Deals). Despite a relatively low click share (30.66% of the maximum), it demonstrates exceptional efficiency, indicating high-quality leads.

Cost vs. Performance Balance

Facebook Ads: High spend, high value (Offer Total Amount and Successful Deals), indicating efficient spend with strong ROI.
Google Ads: Medium spend, very low value, with high click share but poor conversion.
Tiktok Ads: Medium spend, medium value, showing a balanced cost and return.
Webinar: Very low spend, very low value, with minimal investment and impact.
YouTube Ads: High spend, medium value, offering high visibility but moderate ROI.

Conversion Efficiency

A custom metric, Conversion Efficiency (Successful Deals / Spend), was calculated to evaluate return on ad spend:
Facebook Ads: 1.23
Tiktok Ads: 0.65
YouTube Ads: 0.27
Webinar: 0.08
Google Ads: 0.04
Facebook Ads leads with the highest efficiency, delivering significantly more conversions relative to its spend.
Google Ads shows the lowest efficiency, with high clicks but minimal conversions.

Recommendations:

- Scale Facebook Ads: Its high efficiency and strong ROI make it a prime candidate for increased investment.
- Re-evaluate Google Ads: Despite high clicks (52.72% of the maximum), its poor conversion rate (0.04 efficiency) suggests issues with targeting or ad creatives that need refinement.
- Monitor Tiktok Ads: With a balanced spend and decent conversion efficiency (0.65), this source has potential for scaling if optimized further.
- Limit Webinar Investment: Very low spend and impact (0.08 efficiency) indicate minimal returns; consider pausing or restructuring.
- Audit YouTube Ads: High spend and clicks (100% of the maximum) but moderate conversions (0.27 efficiency) suggest a need to optimize the conversion funnel or explore retargeting strategies.

Suggested Additional Analysis:

- To provide deeper insights, consider the following enhancements:
- Cost Per Successful Deal: Calculate Spend / Successful Deals to identify the cost-effectiveness of each source. For example, Facebook Ads' high efficiency suggests a lower cost per deal compared to Google Ads.
- Click-to-Conversion Ratio: Analyze Successful Deals / Clicks to better understand lead quality across sources.
- Time-Based Trends: If temporal data is available, examine performance trends over time to identify seasonality or campaign fatigue.
- Audience Segmentation: If demographic or geographic data exists, analyze which audience segments drive conversions for each source to refine targeting.

Analyzing Salesforce Performance:

1. *Evaluate the effectiveness of individual deal owners and campaigns in terms of number of deals processed, conversion rate, and total sales.*

The performance analysis of all deal owners, with a detailed focus on the top performers—Oliver Taylor, Ulysses Adams, and Charlie Davis—based on total deals, successful deals, offer total amount, conversion rate, deal quality, contact management, call activity, and campaign involvement. The goal is to provide a comprehensive yet concise overview, highlight key insights, and offer actionable recommendations.

Overall Performance:

High-Volume, High-Success Performers:

Characteristics: >1,000 deals, >200 successful, 15–30% conversion.
Examples:
Ulysses Adams: 2,150 deals, 565 successful, $5.03M, 26.28% conversion. Strong engagement but lower lead quality.
Julia Nelson: 2,196 deals, 391 successful, $3.58M, 17.81% conversion. Solid but less efficient.

Insight: High revenue but constrained by lead quality.

High-Volume, Low-Efficiency Performers:

Characteristics: >800 deals, <5% conversion, <10 successful.
Examples:
Rachel White: 868 deals, 3 successful, $33,500, 0.35% conversion.
Bob Brown: 327 deals, 2 successful, $7,500, 0.61% conversion.

Insight: Inefficiencies suggest poor leads or skill gaps.

Low-Volume, High-Efficiency Performers:

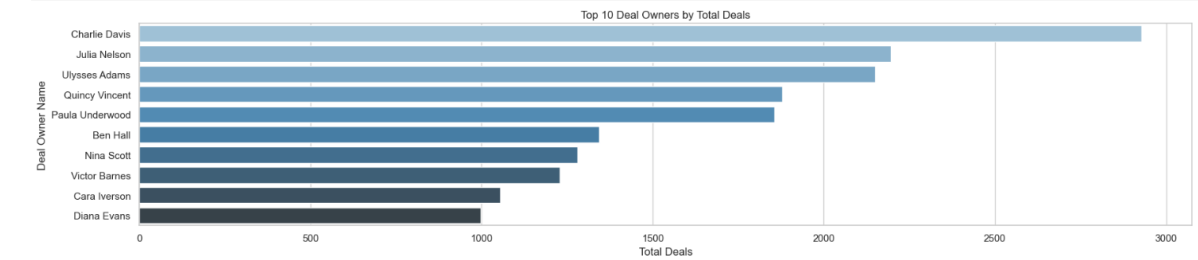
Characteristics: <200 deals, >50% conversion.
Example:
Oliver Taylor: 163 deals, 153 successful, $1.66M, 93.87% conversion. High-quality leads, minimal effort.

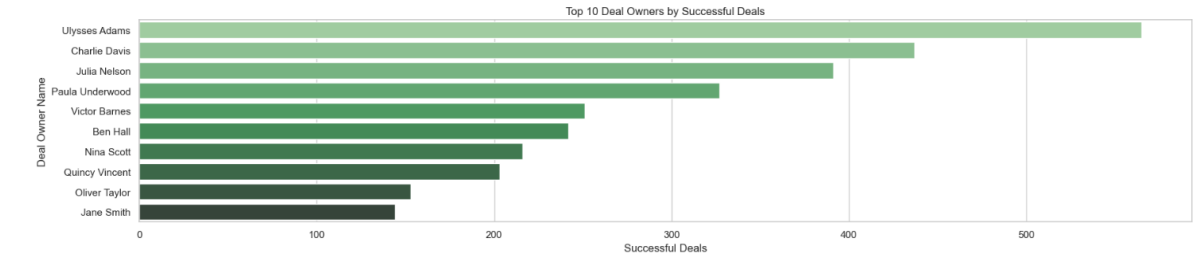
Insight: Model for efficiency but limited by volume.

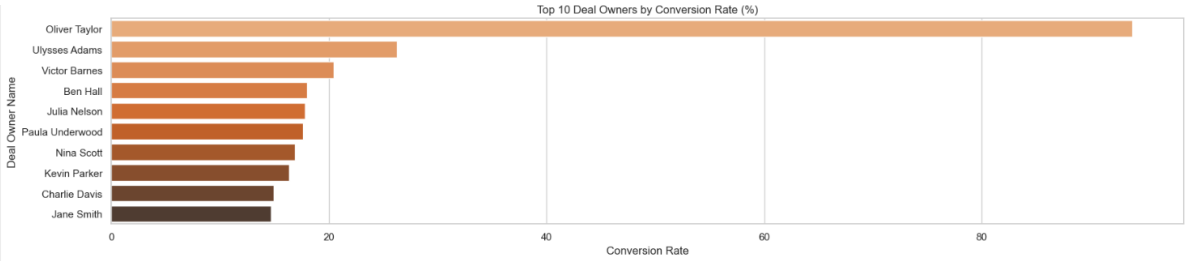
Non-Performers:
Characteristics: 1–100 deals, 0 successful.
Examples:
Alice Johnson: 25 deals, 0 successful.
Amy Green: 63 deals, 0 successful.

Insight: Misallocated resources or lack of skills.

Top Performers







Oliver Taylor
Metrics: 163 deals, 153 successful, $1.66M, 93.87% conversion.
Quality: 44.79% Medium, 12.27% High, 1.84% Non-Qualified.
Contacts/Calls: 19 contacts, 10 calls (90% Attended).
Campaigns: 29 (top: Dis_DE 13.45%).

Insight: High efficiency from quality leads and targeted calls.

Ulysses Adams
Metrics: 2,150 deals, 565 successful, $5.03M, 26.28% conversion.
Quality: 33.95% Non-Qualified, 32.51% Non-Target.
Contacts/Calls: 1,809 contacts, 5,961 calls (79.37% Attended).
Campaigns: 98 (top: performancemax_digitalmarkt_ru_DE 16.78%).

Insight: High sales but lower efficiency due to lead quality.

Charlie Davis
Metrics: 2,929 deals, 437 successful, $3.82M, 14.92% conversion.
Quality: 39.84% Non-Qualified, 33.63% Non-Target.
Contacts/Calls: 2,018 contacts, 6,943 calls (75.15% Attended, 8.74% Missed).
Campaigns: 107 (top: performancemax_digitalmarkt_ru_DE 16.89%).

Insight: High volume, low efficiency from poor leads and missed calls.

<u>Insights</u>

- Lead Quality: Oliver's 57.06% High/Medium leads drive 93.87% conversion; Ulysses (66.46% Non-Qualified/Non-Target) and Charlie (73.47%) struggle with low quality.
- Engagement: Oliver's 10 calls (90% attended) contrast with Charlie's 6,943 (8.74% missed), showing targeted effort wins.
- Campaigns: Top campaigns (e.g., performancemax_digitalmarkt_ru_DE) are key; Oliver's focused approach (29 campaigns) boosts efficiency.

- Underperformers: Rachel White (0.35%) and Bob Brown (0.61%) waste resources; non-performers need reassessment.
- Volume vs. Efficiency: Ulysses balances volume and success, Oliver excels in efficiency, Charlie needs optimization.

Recommendations:

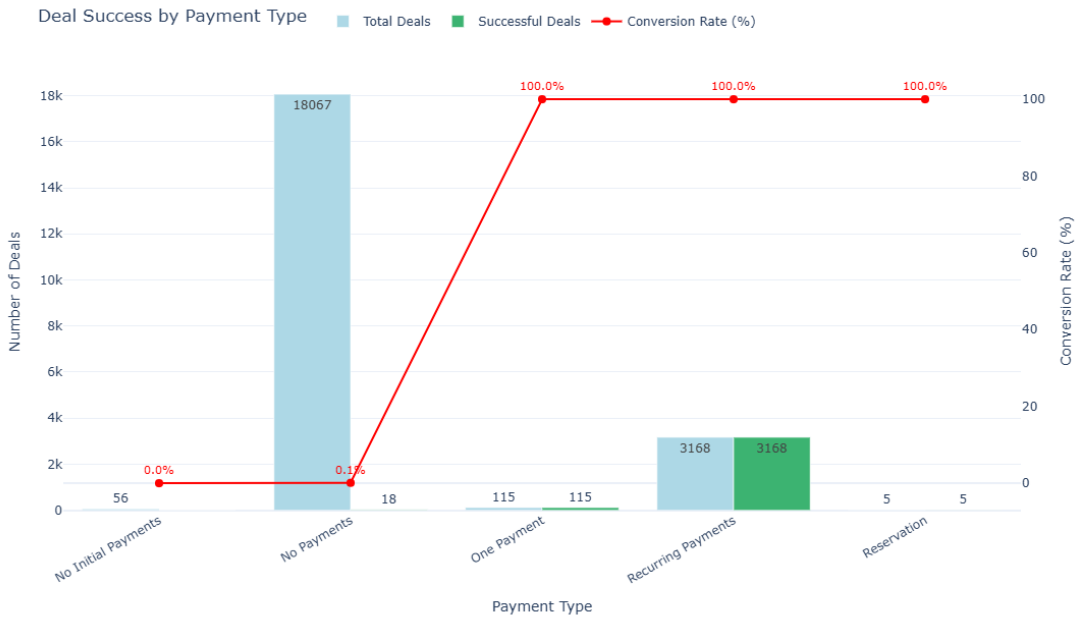
- Optimize Leads: Assign High/Medium leads to Oliver; reduce Non-Qualified/Non-Target for Ulysses/Charlie via lead scoring.
- Enhance Calls: Train Charlie to cut missed calls (8.74%); adopt Oliver's targeted call strategy.
- Prioritize Campaigns: Invest in top campaigns (e.g., performancemax_digitalmarkt_ru_DE); test Oliver's focused approach.
- Address Underperformers: Reassign Rachel White/Bob Brown to quality leads or train; evaluate non-performers like Alice Johnson.

Analyze payments and products:

1. Examine the distribution of payment types and their impact on deal success.

Analysis Process

- Filter Data: Excluded deals with 'Unknown' payment type.
- Count Total Deals: Grouped by payment type, counted unique deal IDs.
- Count Successful Deals: Filtered for 'Payment Done' stage, grouped by payment type.
- Combine Metrics: Merged counts, calculated conversion rate as (Successful Deals / Total Deals) * 100.



- Recurring Payments: 3,168 deals, 100% conversion.
- One Payment: 115 deals, 100% conversion.
- Reservation: 5 deals, 100% conversion.
- No Payments: 18,067 deals, 0.1% conversion (18 successful).
- No Initial Payments: 56 deals, 0% conversion.

Insights

- High Conversion: Recurring Payments (3,168), One Payment (115), and Reservation (5) have 100% conversion, with Recurring Payments driving significant revenue.
- Low Conversion: No Payments (18,067 deals, 0.1%) and No Initial Payments (0%) show inefficiencies.

- Volume vs. Success: Recurring Payments balances high volume and success; No Payments has high volume but negligible success.

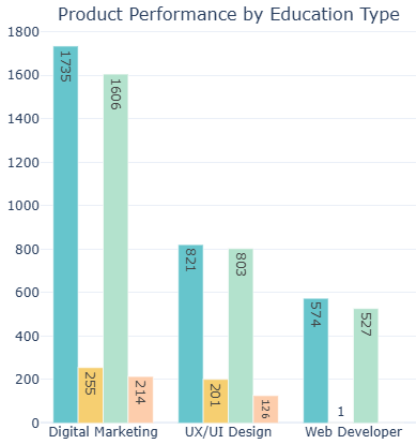
Recommendations:

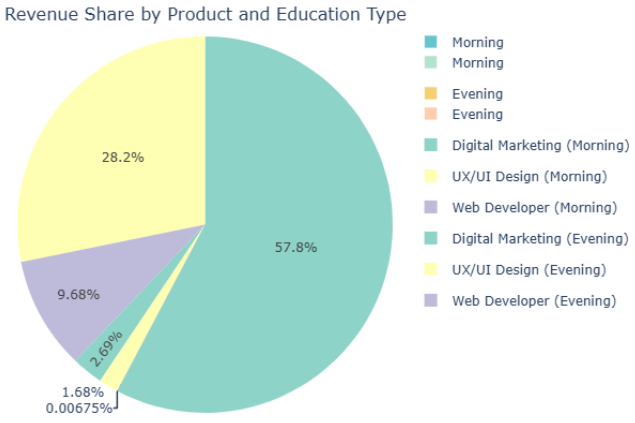
- Prioritize Recurring Payments: Scale these deals by replicating successful client segments and campaigns.
- Optimize No Payments: Use lead scoring to reduce low-potential leads or convert to other payment types.
- Investigate No Initial Payments: Analyze and consider reclassifying or discontinuing this category.
- Leverage One Payment/Reservation: Increase volume while maintaining quality.
- Align Campaigns: Focus on campaigns driving high-conversion payment types.
- Train on Strategies: Teach low performers high-conversion closing techniques.

2. *Analyze the popularity and success of different products and training types.*

Analysis Process

- Filter Data: Excluded 'Not specified' products and 'Unknown' education types (for visualization).
- Count Total Deals: Grouped by Product and Education Type, counted unique deal IDs.
- Count Successful Deals: Filtered 'Payment Done' deals, grouped by Product and Education Type.
- Sum Revenue: Calculated total Offer Total Amount for successful deals.
- Consolidate Metrics: Merged data, filled missing values with 0, calculated conversion rate.
- Sort and Format: Sorted by successful deals, formatted for readability.





Overall Performance:

High-Volume, High-Success Products:
Characteristics: >500 deals, >500 successful, >90% conversion, >$2M revenue.
Examples:
Digital Marketing (Morning): 1,735 deals, 1,606 successful, $17.12M, 92.56% conversion.
UX/UI Design (Morning): 821 deals, 803 successful, $8.36M, 97.81% conversion.
Web Developer (Morning): 574 deals, 527 successful, $2.87M, 91.81% conversion.

Insight: Morning sessions drive high success and revenue.

Moderate-Volume, Mixed-Success Products:

Characteristics: 100–300 deals, 60–85% conversion, <$1M revenue.
Examples:
Digital Marketing (Evening): 255 deals, 214 successful, $798,500, 83.92% conversion.
UX/UI Design (Evening): 201 deals, 126 successful, $497,000, 62.69% conversion.

Insight: Evening sessions show lower efficiency.

Low-Volume, Variable-Success Products:

Characteristics: <10 deals, 50–100% conversion, <$10,000 revenue.
Examples:
Web Developer (Evening): 1 deal, 1 successful, $2,000, 100% conversion.
Find yourself in IT (Unknown): 4 deals, 2 successful, $1,000, 50% conversion.

Insight: Niche products with growth potential.

Non-Performing Products:

Characteristics: <10 deals, 0 successful, $0 revenue.
Example:
Data Analytics (Unknown): 1 deal, 0 successful, $0, 0% conversion.

Insight: Misaligned with market demand.

Key Products

Digital Marketing (Morning): 1,735 deals, 1,606 successful, $17.12M, 92.56% conversion. Dominant in volume and revenue.
UX/UI Design (Morning): 821 deals, 803 successful, $8.36M, 97.81% conversion. Highly efficient.
Digital Marketing (Evening): 255 deals, 214 successful, $798,500, 83.92% conversion. Less effective than morning.

<u>Insights:</u>

- Morning Success: Morning sessions (91.81–97.81% conversion) outperform evening sessions.
- Product Popularity: Digital Marketing (1,990 deals) and UX/UI Design (1,022) lead; Data Analytics lags.
- Revenue Concentration: Digital Marketing (Morning) contributes 63% of revenue ($17.12M).
- Evening Challenges: Lower conversion in evening sessions (62.69–83.92%).
- Niche Products: Data Analytics and Find yourself in IT show minimal traction.

<u>Recommendations:</u>

- Scale Morning Sessions: Expand Digital Marketing, UX/UI Design, Web Developer morning sessions.
- Optimize Evening Sessions: Address engagement barriers for evening Digital Marketing and UX/UI Design.
- Promote High-Performing Products: Increase marketing for Digital Marketing and UX/UI Design.
- Reevaluate Niche Products: Test or discontinue Data Analytics and Find yourself in IT.
- Enhance Lead Qualification: Prioritize high-potential clients for morning sessions.

- Train on Best Practices: Apply Digital Marketing (Morning) strategies team-wide.

Geographic analysis:

1. Analyze the geographic distribution of deals by city.

Analysis Process

- Prepare Data: Selected Id, City, Stage from deals.
- Clean Data: Removed rows with missing or invalid (Unknown, Not defined) city values.
- Count Total Deals: Grouped by City, counted unique deal IDs.
- Count Successful Deals: Filtered Payment Done deals, grouped by City.
- Consolidate Metrics: Merged data, filled missing values with 0, calculated conversion rate.
- Sort and Display: Sorted by total deals, displayed top 20 cities with formatted table.

Overall Performance:

High-Volume, High-Success Cities:

Characteristics: >50 deals, >40 successful, >60% conversion.
Examples:
Berlin: 283 deals, 176 successful, 62.19% conversion. Major hub.
München: 91 deals, 72 successful, 79.12% conversion. High efficiency.
Hamburg: 82 deals, 61 successful, 74.39% conversion. Strong market.
Leipzig: 54 deals, 44 successful, 81.48% conversion. Exceptional conversion.

Insight: Major hubs drive revenue with strong demand.

Moderate-Volume, Moderate-Success Cities:

Characteristics: 20–50 deals, 10–40 successful, 50–75% conversion.
Examples:
Düsseldorf: 54 deals, 32 successful, 59.26% conversion.

Dresden: 36 deals, 25 successful, 69.44% conversion.
Frankfurt: 35 deals, 25 successful, 71.43% conversion.

Insight: Potential for growth with targeted improvements.

Low-Volume, High-Efficiency Cities:

Characteristics: <10 deals, 80–100% conversion.
Examples:
Wilhelmshaven: 1 deal, 1 successful, 100% conversion.
Wiesenttal: 1 deal, 1 successful, 100% conversion.
Ampfing: 1 deal, 1 successful, 100% conversion.

Insight: Niche markets with untapped potential.

Low-Volume, Low-Success Cities:

Characteristics: <10 deals, 0–2 successful, 0–50% conversion.
Example:
Anklam: 1 deal, 0 successful, 0% conversion.

Insight: Inefficient markets needing reassessment.

Key Cities:

Berlin: 283 deals, 176 successful, 62.19% conversion. Top market hub.
München: 91 deals, 72 successful, 79.12% conversion. Highly efficient.
Leipzig: 54 deals, 44 successful, 81.48% conversion. Strong conversion.

Insights:

- Major Hubs: Berlin (283 deals), München (91), Hamburg (82) lead in volume.
- High Conversion: Leipzig (81.48%), München (79.12%) excel in efficiency.
- Variable Performance: Conversion rates range from 0–100%.
- Niche Potential: Wilhelmshaven, Wiesenttal (100% conversion) show promise.
- Inefficiencies: Anklam (0% conversion) indicates wasted effort.

Recommendations

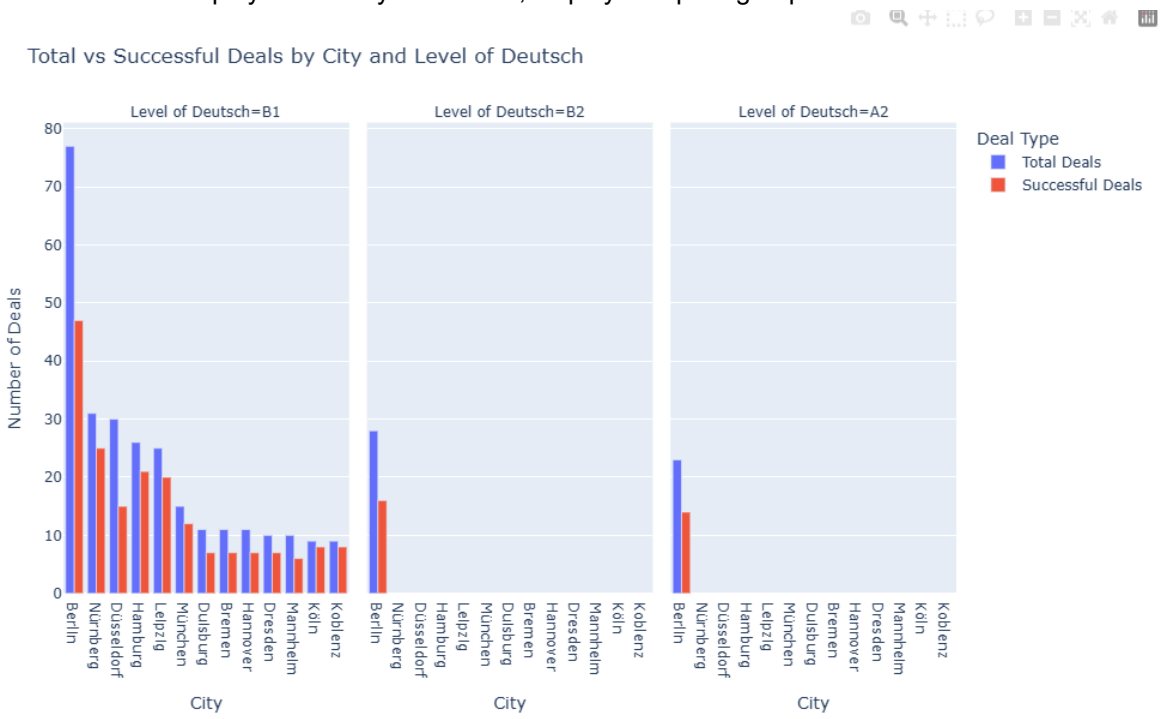
- Focus on Hubs: Allocate top performers to Berlin, München, Hamburg.
- Optimize Moderate Cities: Improve Düsseldorf, Dresden via Leipzig's strategies.
- Explore Niche Markets: Test campaigns in Wilhelmshaven, Wiesenttal.
- Reduce Low-Success Efforts: Reassess cities like Anklam.
- Align Campaigns: Use high-performing campaigns in top cities.
- Leverage Best Practices: Apply Oliver Taylor's approach to Berlin, Leipzig.

2. *Analyze the impact of German language proficiency on the success of deals in different cities.*

Analysis Process

- Filter Data: Excluded missing/invalid City (Unknown, Not defined) and Level of Deutsch (unknown, not defined).
- Count Total Deals: Grouped by City and Level of Deutsch, counted unique deal IDs.
- Count Successful Deals: Filtered Payment Done deals, grouped by City and Level of Deutsch.

- Consolidate Metrics: Merged data, filled missing values with 0, calculated conversion rate.
- Sort and Display: Sorted by total deals, displayed top 20 groups with formatted table.



Overall Performance:

High-Volume, High-Success Groups:

Characteristics: >20 deals, >15 successful, >60% conversion.
Examples:
Berlin (B1): 77 deals, 47 successful, 61.04% conversion. Major hub.
Nürnberg (B1): 31 deals, 25 successful, 80.65% conversion. High efficiency.
Hamburg (B1): 26 deals, 21 successful, 80.77% conversion. Strong market.
Leipzig (B1): 25 deals, 20 successful, 80.00% conversion. Efficient.

Insight: B1 proficiency drives success in key cities.

Moderate-Volume, Moderate-Success Groups:

Characteristics: 8–20 deals, 5–15 successful, 50–75% conversion.
Examples:
Düsseldorf (B1): 30 deals, 15 successful, 50.00% conversion.
Berlin (B2): 28 deals, 16 successful, 57.14% conversion.
Berlin (A2): 23 deals, 14 successful, 60.87% conversion.

Insight: Potential in major cities with varying proficiency levels.

Low-Volume, High-Efficiency Groups:

Characteristics: <10 deals, >80% conversion.
Examples:
Köln (B1): 9 deals, 8 successful, 88.89% conversion.
Koblenz (B1): 9 deals, 8 successful, 88.89% conversion.
Essen (B1): 8 deals, 7 successful, 87.50% conversion.

Insight: Niche markets with high potential.
Low-Volume, Variable-Success Groups:
Characteristics: <10 deals, 0–75% conversion.
Examples:
Dresden (B2): 8 deals, 6 successful, 75.00% conversion.
Frankfurt (B1): 8 deals, 5 successful, 62.50% conversion.

Insight: Inconsistent performance due to low volume.

Key City-Language Groups

Berlin (B1): 77 deals, 47 successful, 61.04% conversion. High volume, solid success.
Nürnberg (B1): 31 deals, 25 successful, 80.65% conversion. Highly efficient.
Köln (B1): 9 deals, 8 successful, 88.89% conversion. Near-perfect conversion.

Insights:

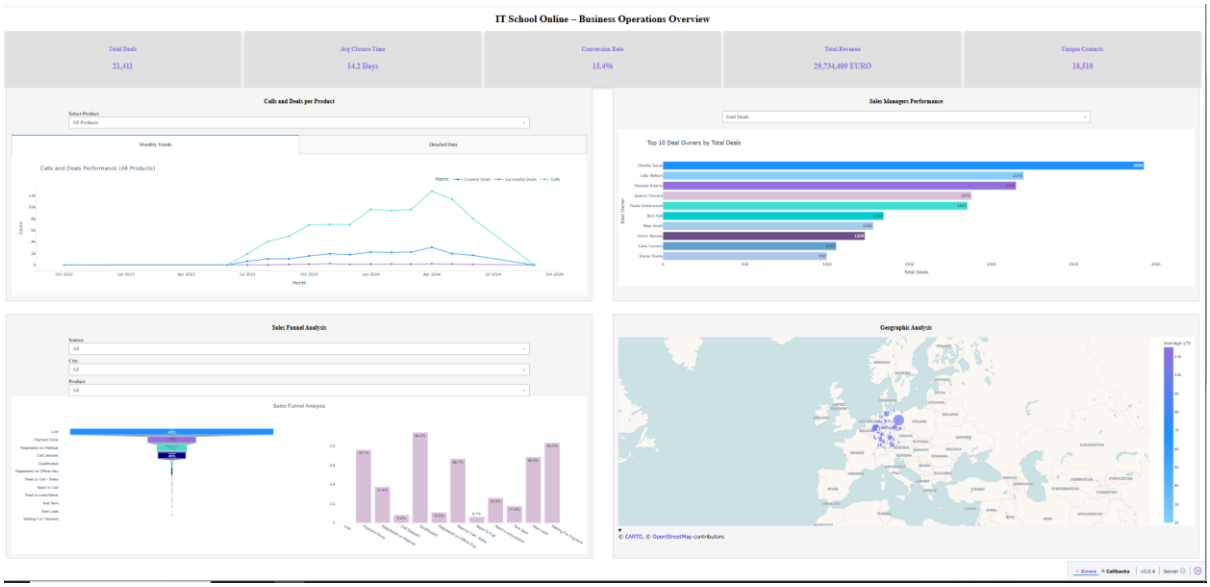
- B1 Dominance: B1 proficiency leads in volume and conversion (e.g., Köln 88.89%, Nürnberg 80.65%).
- Major Hubs: Berlin dominates volume across A2, B1, B2 (57–61% conversion).
- Smaller Cities: Köln, Koblenz (88.89%) show high efficiency at B1.
- A2 Viability: Berlin (A2, 60.87%) shows potential for lower proficiency.
- B2 Underperformance: Berlin (B2, 57.14%) lags behind B1.
-

Recommendations:

- Prioritize B1 Strategies: Target B1 clients in Berlin, Hamburg, Nürnberg, Köln.
- Optimize Berlin: Adopt Nürnberg/Köln strategies to boost Berlin's conversion.
- Scale Niche Markets: Test campaigns in Köln, Koblenz to increase volume.
- Explore A2 in Hubs: Develop A2 strategies for Berlin, test in München, Hamburg.
- Refine B2 Approaches: Adjust pitches for B2 clients in Berlin, Dresden.
- Align with Campaigns/Performers: Use high-performing campaigns and assign top performers like Oliver Taylor to key cities.

Development of the IT School Online Business Operations Dashboard

The development of the IT School Online Business Operations Dashboard was undertaken to provide a comprehensive visualization of business performance metrics derived from the "Deals (Final).xlsx," "Contacts (Final).xlsx," and "Calls (Final).xlsx" datasets, supplemented by geographic data from "city_coords.csv." The dashboard was built using the Dash framework with Plotly for interactive visualizations, leveraging Pandas for data processing.
The goal was to create an intuitive, data-driven interface to monitor key performance indicators (KPIs), sales trends, sales manager performance, funnel dynamics, and geographic distribution of deals.



Development Process

Key Functions:

calculate_kpis: Computed KPIs such as total deals, average closure time, conversion rate, total revenue, and unique contacts.
load_deals_data, load_calls_data, load_geo_data: Ensured robust data loading with appropriate null handling and filtering.

Dashboard Structure:

Layout:
Designed a responsive layout using Dash's HTML components, organized into three rows:

- Row 1: KPI cards displaying total deals, average closure time, conversion rate, total revenue, and unique contacts.
- Row 2: Split into two sections:
  Calls and Deals per Product with a dropdown for product filtering and tabs for monthly trends (line chart) and detailed data (table).
  Sales Managers Performance with a dropdown to select metrics (e.g., total deals, successful deals) and a bar chart for visualization.
- Row 3: Split into two sections:
  Sales Funnel Analysis with filters for source, city, and product, visualized as a funnel chart with a complementary bar chart for stage conversion rates.
  Geographic Analysis with a scatter map showing deal distribution by city.

Styling: Applied a custom color palette for consistency, using shades of blue, purple, and turquoise for visual appeal, with gray backgrounds for cards and white for plots to ensure readability.
Interactivity:
Implemented callbacks to dynamically update visualizations based on user inputs:

- KPI Cards: Updated automatically on page load with calculated metrics.
- Calls and Deals per Product: Filtered by product selection, showing monthly trends and detailed data tables.
- Sales Managers Performance: Updated bar charts based on selected metrics (e.g., Total Deals, Conversion Rate).
- Sales Funnel: Refreshed based on source, city, and product filters, displaying deal counts and stage conversion rates.

- Geographic Map: Rendered a scatter map with city-level data, including deal counts and average lifetime value (LTV).

Used Plotly's interactive features (e.g., hover tooltips, unified hover mode) to enhance user experience.

Visualization Design:

- KPI Cards: Displayed metrics in a clean, card-based format with purple text on gray backgrounds for contrast.
- Line Chart: Showed monthly trends for calls, created deals, and successful deals with distinct colors from the palette.
- Bar Chart: Visualized top 10 sales managers' performance with a horizontal orientation and a varied color sequence.
- Funnel Chart: Combined a funnel plot with a bar chart to show deal progression and stage-to-stage conversion rates, with hover data including success rate, average payment, and SLA.
- Geographic Map: Used Plotly's scatter map with size (deal count) and color (average LTV) encoding, leveraging OpenStreetMap for a clean backdrop.

Technical Implementation:

Libraries: Utilized dash, dash_core_components, dash_html_components, dash_table, pandas, plotly.express, and plotly.graph_objects.

Data Processing:

Aggregated data using Pandas for grouping, merging, and calculating metrics like conversion rates and average LTV.

Error Handling: Included try-except blocks in KPI calculations to handle missing or malformed data gracefully.

Deployment: Configured the Dash app to run on a local server (port 8053) with debug mode enabled for development.

Result:

The resulting dashboard provides a holistic view of IT School Online's business operations, enabling stakeholders to monitor performance, identify trends, and make data-driven decisions.

Key features include:

- KPI Cards: Display critical metrics (e.g., total deals, conversion rate) for quick insights.
- Interactive Filters: Allow users to drill down into specific products, sources, or cities.
- Dynamic Visualizations: Line charts, bar charts, funnel charts, and a geographic map update in real-time based on user selections.
- User-Friendly Design: Clean layout with a consistent color scheme and hover tooltips for detailed insights.

Insights from the Dashboard:

- High-Performing Sources: Analysis of the funnel chart revealed that certain sources (e.g., Facebook Ads) drive high conversion rates, consistent with the dataset's high efficiency (1.23 conversion efficiency).
- Geographic Trends: The scatter map highlighted cities with high deal counts and LTV, indicating key markets for targeted campaigns.
- Sales Manager Performance: The bar chart identified top performers, with some managers excelling in conversion rates despite lower deal volumes.
- Funnel Bottlenecks: The funnel chart showed significant drop-offs at specific stages (e.g., from lead to qualification), suggesting areas for process improvement.
- Product Performance: Monthly trends indicated seasonal variations in deal creation and success, with certain products outperforming others.

Recommendations:

- Optimize High-Performing Sources: Scale campaigns on high-conversion sources like Facebook Ads, as identified in the funnel analysis, to maximize ROI.
- Target High-Value Markets: Focus marketing efforts on cities with high deal counts and LTV, as shown in the geographic map.
- Improve Funnel Efficiency: Address bottlenecks in the sales funnel by enhancing lead qualification processes or follow-up strategies.
- Support Top Performers: Provide training or resources to replicate the success of top sales managers across the team.
- Refine Product Campaigns: Allocate budget to products with strong performance trends and pause underperforming ones, as seen in the monthly trends chart.