Capstone Project

AI-Driven Marketing Channels Optimization and Identification for Maximizing Paid Media ROI - Final Report

1. **Define the Problem Statement:** Identifying which channels to invest at Salesforce Marketing Department is a big challenge. The multi million budget spending requires strong data driven insights to identify which paid advertising channel is more viable to ROI. The goal is to identify where to invest more and identify data backed scalable long term efficiency strategies.

2. **Model Outcomes or Predictions:** This analysis tries to classify and predict an ideal investment methodology and process for advertising channels. The project is based on the CRISP-DM methodology.

3. **Data Acquisition:** Data is representing 3 years of advertising channels data with their hierarchical KPIs such as impressions, clicks, conversions, leads, MQLs, SALs, Opportunities Stage 1 and Stage 2 and last but not least ACV. The dataset has been downloaded from Snowflake table connected to Salesforce CRM. The data is representing 3 years of KPIs downloaded on advertising channels segments level.

4. **Data Preprocessing/Preparation:** For this deliverable I have applied CRISP-DM with the following steps to clearly define step by step this project:
   a. Understand the Data and Import Necessary Libraries
   b. Read in and Explore the Data
   c. Data Visualization and Analysis
   d. Cleaning Data
   e. Data Modeling and Evaluation
   f. Choosing the Best Model - Model Comparison
   g. Scoring and Improving the Model
   h. Providing Best Business Recommendations

5. **Modeling:** I have used the following ML models:
   a. Linear Regression
   b. KNN
   c. SVD (via matrix factorization on platforms)
   d. Clustering (PCA + KMeans)
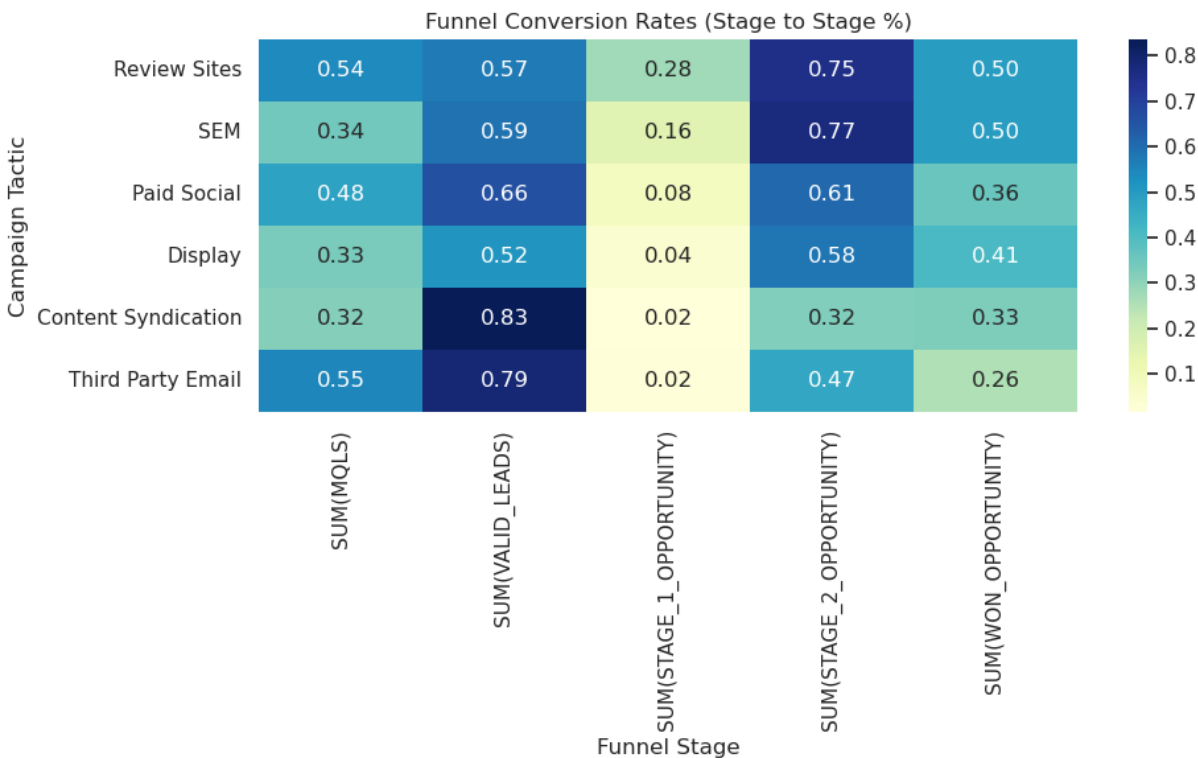   e. Decision Tree or Logistic Regression
   Reasons for selecting the above models for clustering, because we're learning to predict campaign cluster type, a proxy for ROI potential
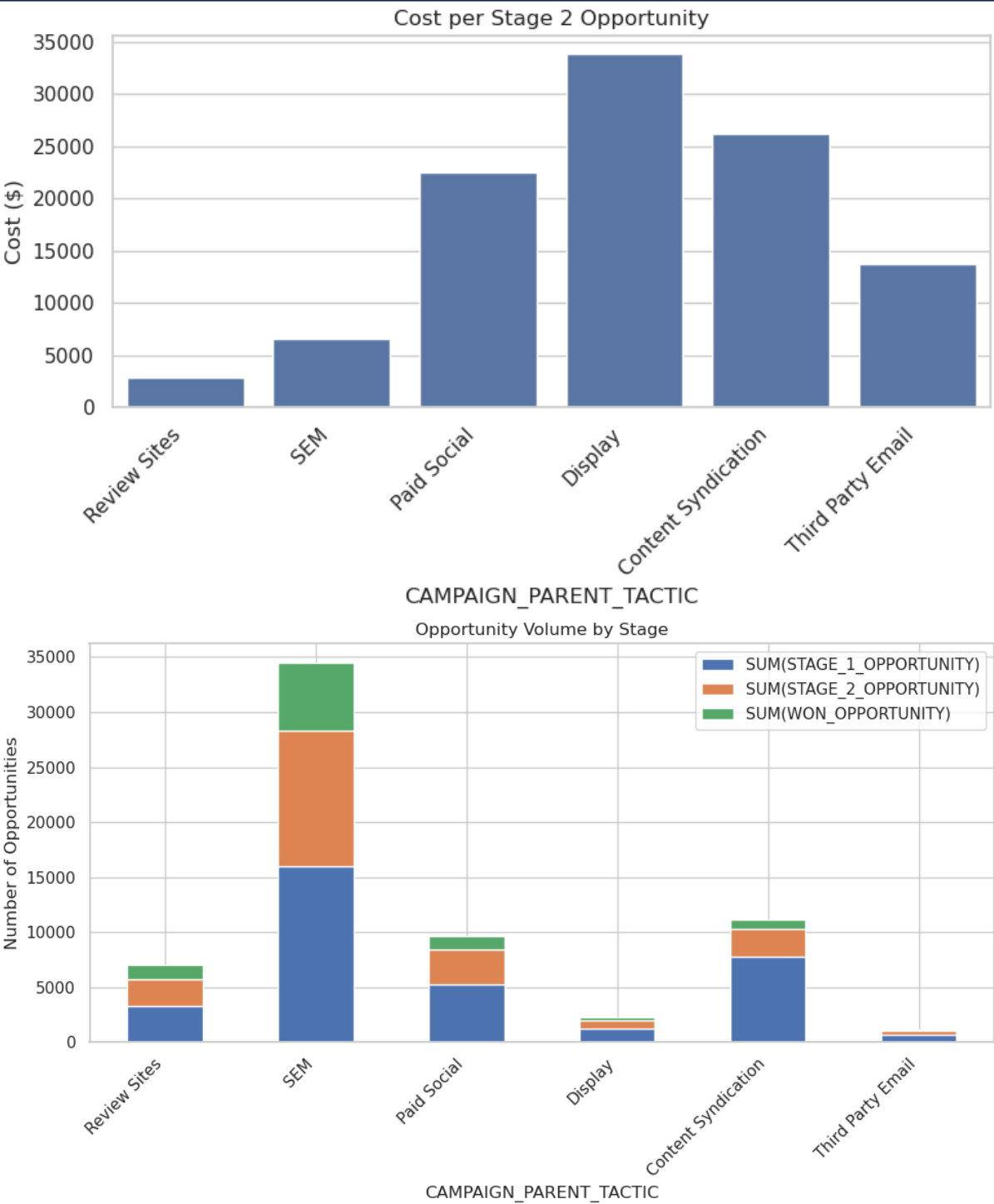   o Logistic Regression: Interpretable weights >> To show feature importance
       o Ridge Classifier: Regularized version of logistic for multicollinearity
       o KNN: Non-parametric, useful as a baseline

6. **Model Evaluation:** I have discovered based on my domain knowledge that ROI data was extremely skewed. Most values are very low (0–40), and a few outliers are much higher, dragging the axis scale but not affecting bin counts. For this reason, I applied a scoring model to improve the model. I built a sales-weighted ROI model by assigning strategic value to funnel stages, then normalizing that against media cost.

This approach:
   a. Rewards pipeline contribution, not just volume
   b. Helps prioritize channels/campaigns that generate real revenue signals
   c. Replaces flat ROI with sales-intelligent ROI

Insights:
- STAGE_1 and STAGE_2 Opportunities are the most predictive features across both models.
- WON_OPPORTUNITY has a weaker signal, possibly due to multicollinearity with STAGE_2.
- MQLS and RESPONSES contribute very little predictive power.
- MEDIA_COST has almost no predictive weight and can likely be dropped for modeling.

Funnel Conversion Rates (Stage to Stage %)

Campaign Tactic	SUM(MQLS)	SUM(VALID_LEADS)	SUM(STAGE_1_OPPORTUNITY)	SUM(STAGE_2_OPPORTUNITY)	SUM(WON_OPPORTUNITY)
Review Sites	0.54	0.57	0.28	0.75	0.50
SEM	0.34	0.59	0.16	0.77	0.50
Paid Social	0.48	0.66	0.08	0.61	0.36
Display	0.33	0.52	0.04	0.58	0.41
Content Syndication	0.32	0.83	0.02	0.32	0.33
Third Party Email	0.55	0.79	0.02	0.47	0.26

Funnel Stage

Cost per Stage 2 Opportunity

Opportunity Volume by Stage

- Funnel Conversion Rates Table: shows how well each campaign tactic is progressing from one funnel stage to the next.
- Funnel Conversion Heatmap: visual representation of the conversion efficiency between stages.
- Cost Efficiency Plot: bar chart of cost per Stage 2 Opportunity, revealing which tactics are more economical in deeper funnel conversion.
- Buying Stage Effectiveness Plot: stacked bar chart showing absolute volume of opportunities at each funnel stage by tactic.

Conclusions:

Model results indicate that campaign clusters are primarily driven by conversion-stage outcomes such as stage 1/2 opportunities and won deals. Early funnel engagement (clicks, impressions) and media cost show negligible influence, suggesting that investment should prioritize campaigns with proven downstream performance rather than just volume metrics.

Final Takeaways:

Final business recommendation summary based on the funnel conversion rates, cost efficiency, and opportunity volume:

Review Sites

Strengths: Strong funnel progression, especially from Stage 2 to Won. Low cost per win (~$1,484). Balanced tactic.
Weaknesses: Moderate top-of-funnel scale.
Recommendation: Keep and cautiously scale. It's efficient and reliably delivers wins. Ideal for targeted investment.

SEM

Strengths: High volume at every stage. Delivers the highest number of won opps (6,150).
Weaknesses: Higher cost per Stage 2 opp (~$2,244). Weak ROI_SALES_FOCUSED.
Recommendation: Keep as a volume engine, but audit keyword/ad set performance to cut waste. Optimize for better bottom-funnel payoff.

Paid Social

Strengths: Very strong Stage 2-to-Won conversion. Low cost per win (~$778).
Weaknesses: Leaky mid-funnel (weaker conversion from Valid Leads to Stage 1). Composite score is modest.
Recommendation: Refine mid-funnel targeting. Great potential if nurtured properly.

Display

Strengths: Reasonable conversion into Stage 2.
Weaknesses: Very few wins (only 306). Inefficient relative to spend. Mid-tier composite performance.
Recommendation: Limit spend or reserve for retargeting only. Reevaluate creative and audience alignment.

Content Syndication

Strengths: Massive top-of-funnel volume.
Weaknesses: Terrible funnel efficiency. High spend with weak win volume. Low composite score.
Recommendation: Aggressively cut or rebuild. Only keep specific vendors showing real opp conversion.
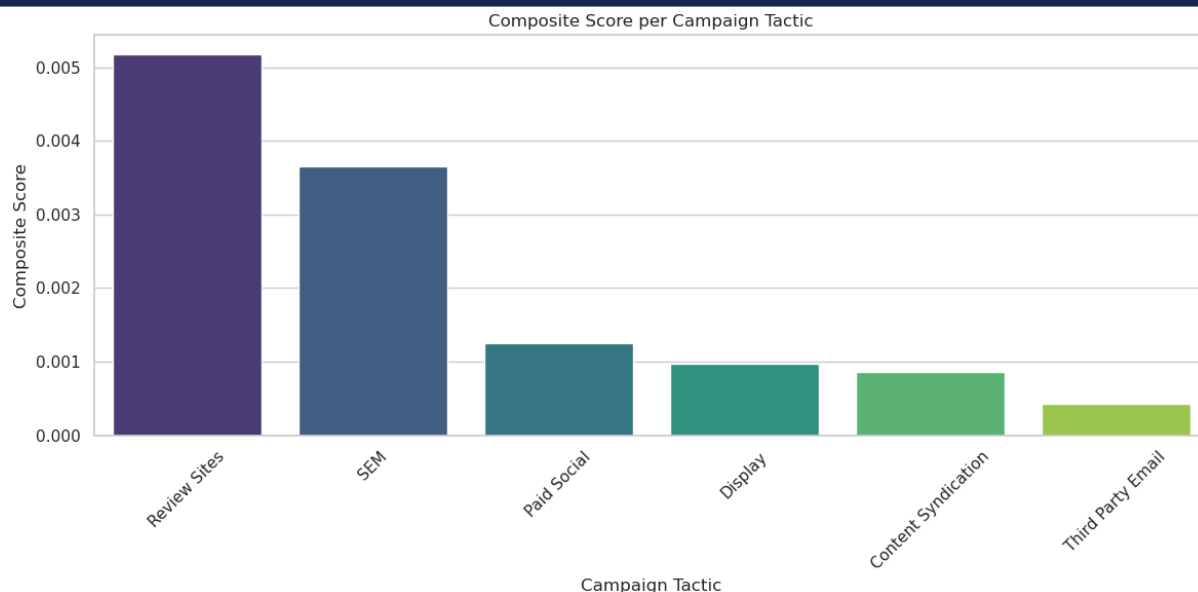
Third Party Email

Strengths: Best log-weighted ROI and lowest cost per win (~$323).
Weaknesses: Extremely weak volume (only 85 wins). Poor Stage 1 and 2 conversion.
Recommendation: Niche tactic. Use only with verified, high-intent lists. Do not scale broadly.

Strategic Takeaways

Composite Score per Campaign Tactic

- Focus investment on Review Sites and SEM for reliable wins and funnel stability.
- Use Paid Social as a scalable low-CAC source — but plug funnel leaks.
- Display and Content Syndication need aggressive cleanup — too costly for current return.
- Third Party Email is only worth running on precision-targeted campaigns.

What We Should Do Next?
- Focus channel optimization efforts on tactics that drive opportunity creation.
- Consider removing weak features or aggregating them into new KPIs (e.g., cost per opp).
- May re-cluster using only high-impact features, or add engineered features like ROI ratios.
    - Adding engineered ROI-based features will both sharpen the model and align directly with my current goal of this project.