

Myren Bobryk-Ozaki  
Case Analysis 2  
MKTG 768  
4/16/2025

## DataSphere Analytics: Segments, Pathways, and Recommendations

### Customer Segmentation

- Utilizing K-means and CRM behaviors, it was determined that 3 segments would be effective as noted by silhouette scores:
  - Silhouette score for k=2: 0.645
  - Silhouette score for k=3: 0.633
  - Silhouette score for k=4: 0.461
  - Silhouette score for k=5: 0.482
  - Silhouette score for k=6: 0.449
  - Silhouette score for k=7: 0.463
- 3 segments determined these:
  - 0: Casual Users** (n=903) - a large number of users with a lot of active days and decent usage, but very low conversion
  - 1: Outliers/Test/Bot Accounts** (n=9) - high visualizations, exports, and dashboards but low active days (avg 1.11) and no conversions
    - This segment was removed from further analysis due to the outlier data, however, I did not want to shift to 2 segments as the outliers might impact the analysis.
  - 2: Power Users** (n=88) - high engagement, feature usage, and the highest lifetime value



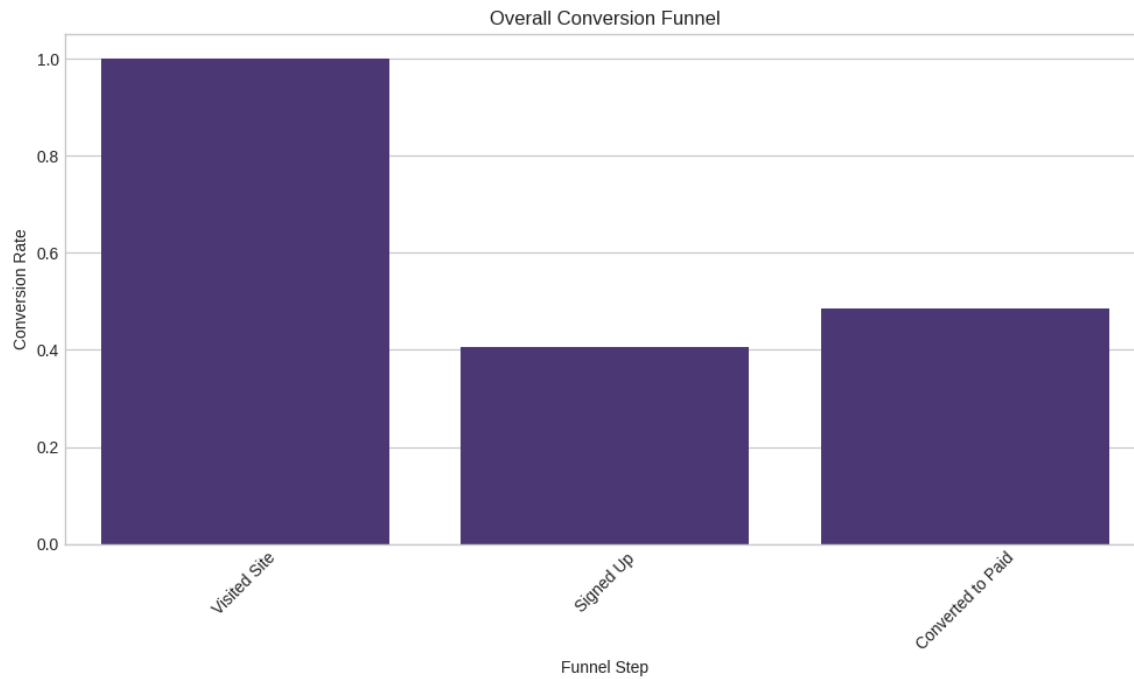
## Conversion Pathways

- The overall funnel was created via website aggregation and merged with the CRM segmentation created in the previous segmentation work.

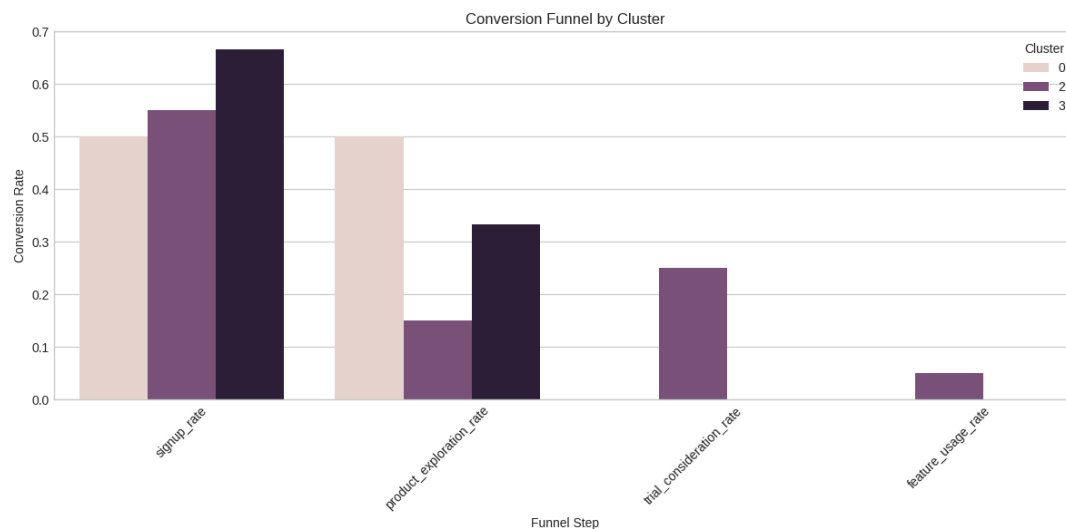
```

user_journey = pd.merge(
    website_agg,
    crm_segmented_3[['user_id', 'cluster', 'user_status', 'acquisition_channel']],
    on='user_id',
    how='left'
)

```

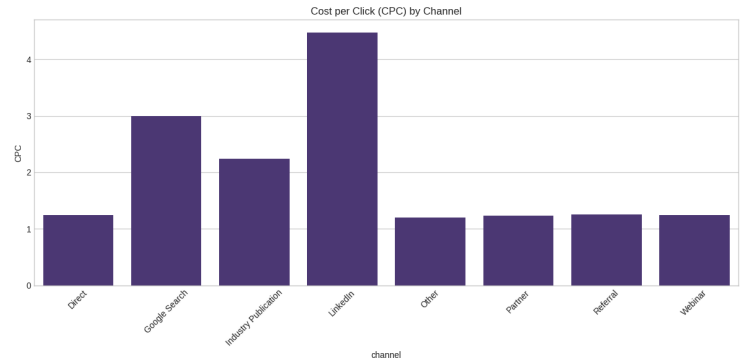


- We are seeing low sign-up rates and moderate conversion rates, however, both of our key segments aren't converting!

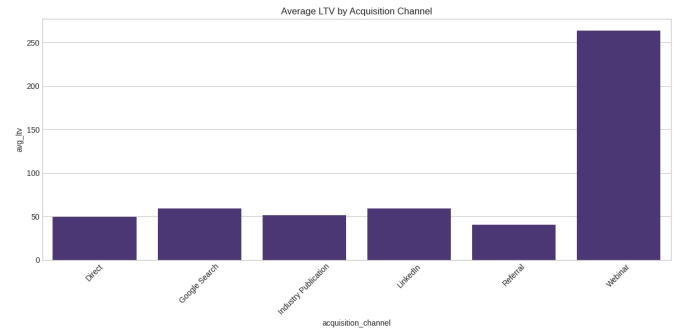
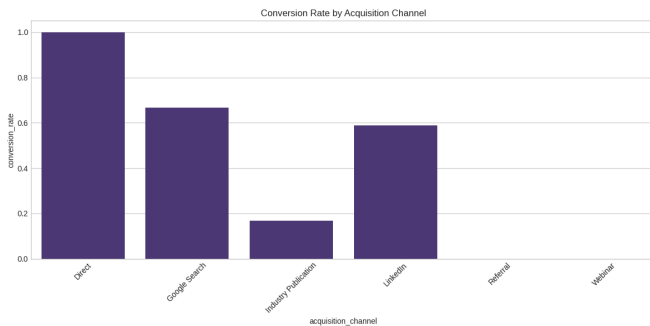


## Channel Evaluation

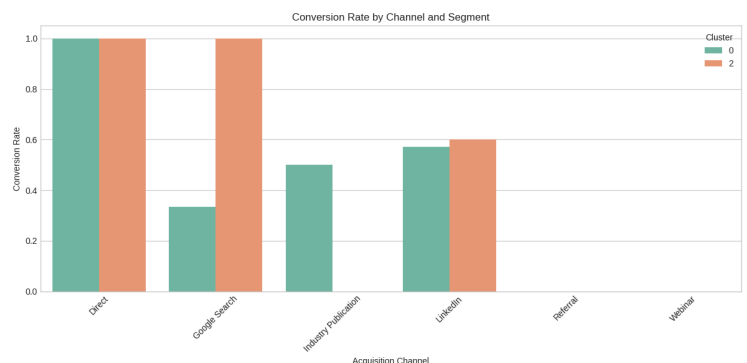
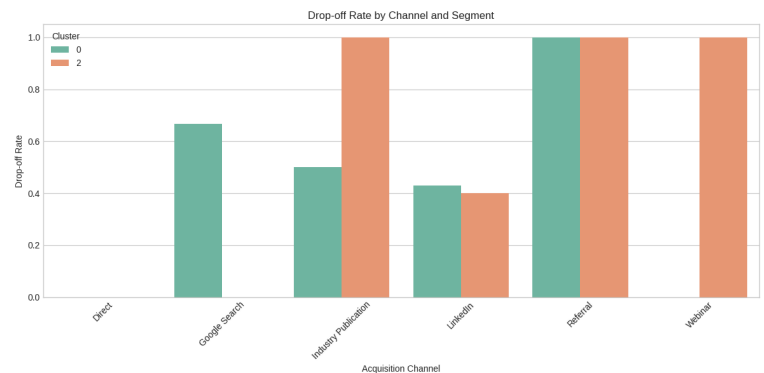
- Notable channel performance metrics are CPC and Cost per Paid Acquisition. Other performance metrics demonstrated similar results across all channels.
  - LinkedIn seems to be the main drain of cost across the channels.



- Combining LTV with the User Journey created earlier in my code allowed for channel analysis – focus remained on conversion rate and average LTV.
  - Highest conversions come from direct, and the highest LTV comes from LinkedIn.



- Drop off and Conversion Rate of the 2 defined customer segments indicate slight differences
  - Casual Users (0) are dropping off the most from webinars, referrals, and industry publications; highest conversions from direct and search
  - Power Users (2) are dropping off from search and referrals; the highest conversions are through direct and LinkedIn



## Recommendations

- Channel Optimization
  - Invest More
    - Direct Traffic - Strong for both segments
      - Double down on SEO, branded content, email reminders, and retargeting
    - Search - Solid for Casual Users, and relatively efficient in cost-per-acquisition
      - More specifically targeted search ads
  - Reduce
    - Referrals & Industry Publications - High drop-off rates for both segments!
      - Audit the quality and relevance of referral partners or content placements
      - Test replacements or pull back the budget
  - Remove
    - LinkedIn
      - High cost and low efficiency, even for Power Users
- Conversion Funnel
  - Bottlenecks:
    - Low sign-up rates overall
    - Power Users aren't converting
  - Funnel Fixes:
    - Optimize Homepage / Key Landing Pages
      - Focus on strong CTAs and feature value/benefit highlighting
    - Pricing Page Engagement
      - Add live chat, short demos, or testimonials
    - Onboarding Flow
    - A/B Test Content for Top Channels
      - Tailor search ads and direct traffic to landing content
- Segment-Specific
  - Casual:
    - Engage with educational content and clear conversion paths
    - Stronger CTA onboarding nudges
  - Power:
    - Rewards program and/or early access programs
    - Trigger-based outreach during trials
    - Feature unlocks or usage-based nudges
    - Dedicated onboarding assistant for high-engagement trial users

### Data Limitations and Considerations

- Very small website sample, which, when combined with CRM data, makes it difficult to parse key elements about website interactions. We need a larger sample size across all the data sets, but particularly the website data.
- Users couldn't be matched across all the data sets, which further reduced analysis capabilities.
- Some values, like LTV, were NaN in the datasets, which were corrected in filtering but may have affected the outcome of the analysis.
- A smaller signup rate than the conversion rate. This indicates an issue with aggregation, which may link back to user match-up issues.