North-Point Software Mailing Analytics

- Proffessor: Dr. Willie Rivers
- Course: CSDA 6010 Analytical Practicum
- Student: Phanidhar Kasuba
- ID: 4249169

Business Problem

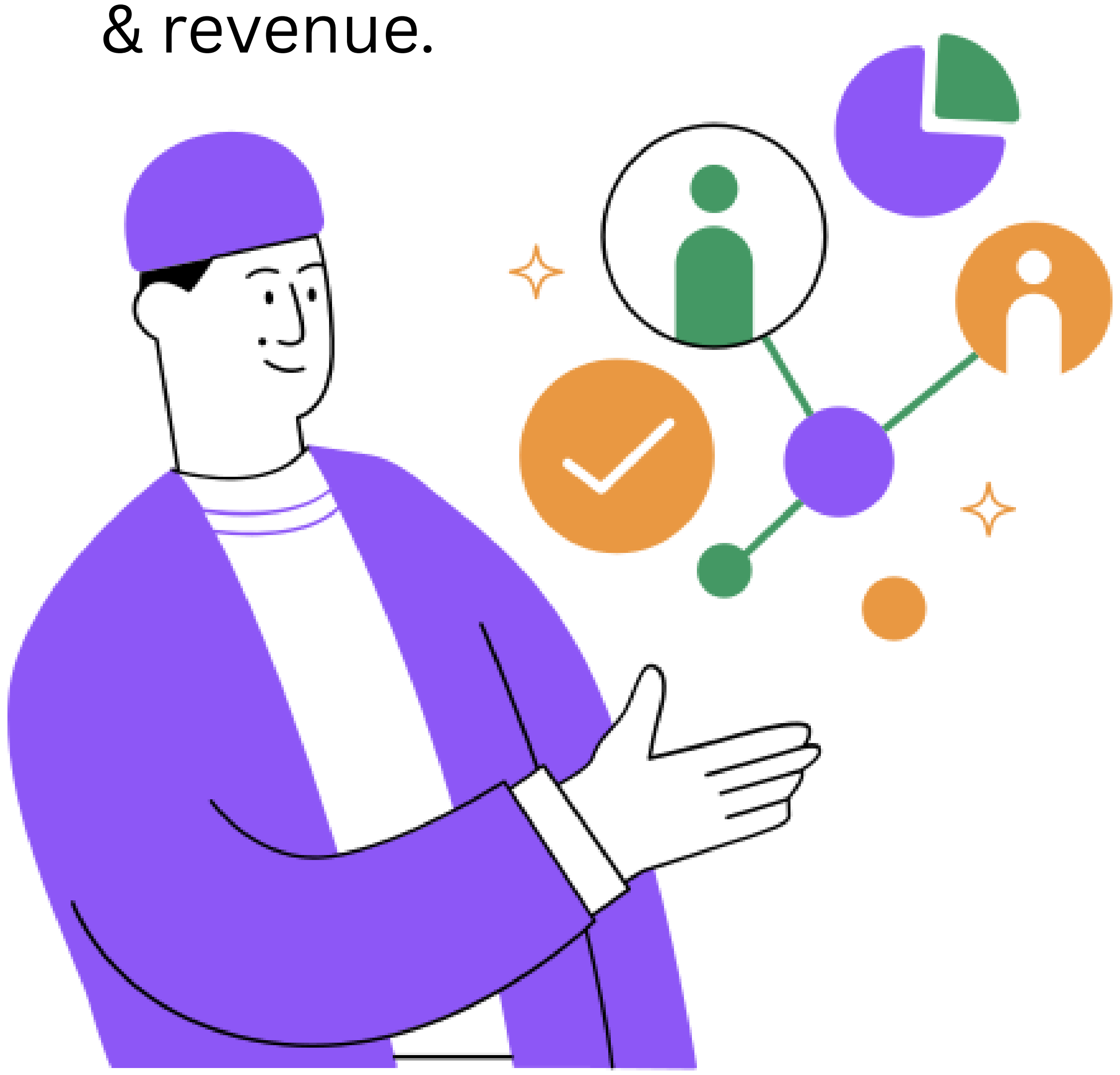
- Need to select 200 k names from 5 M pool for mailing
- Random selection → low response (5.3 %) and high cost
- Which customers are most likely to buy?

Business Goals

- Increase sales & response above pilot baseline (5.3%)
- Cut marketing waste by targeting high-probability buyers
- Increase sales & ROI through data-driven mailing

Analytic Goals

- Predict purchase likelihood (who will buy).
- Estimate expected spending among buyers.
- Segment customers to tailor offers/channels.
- Identify key predictors that drive response & revenue.



Analytic Approach

- Data Preparation & Understanding
- Exploratory Data Analysis
- Model Development(Classification, Regression, Clustering)
- Model Evaluation
- Calibration & Validation

Data Understanding & Quality

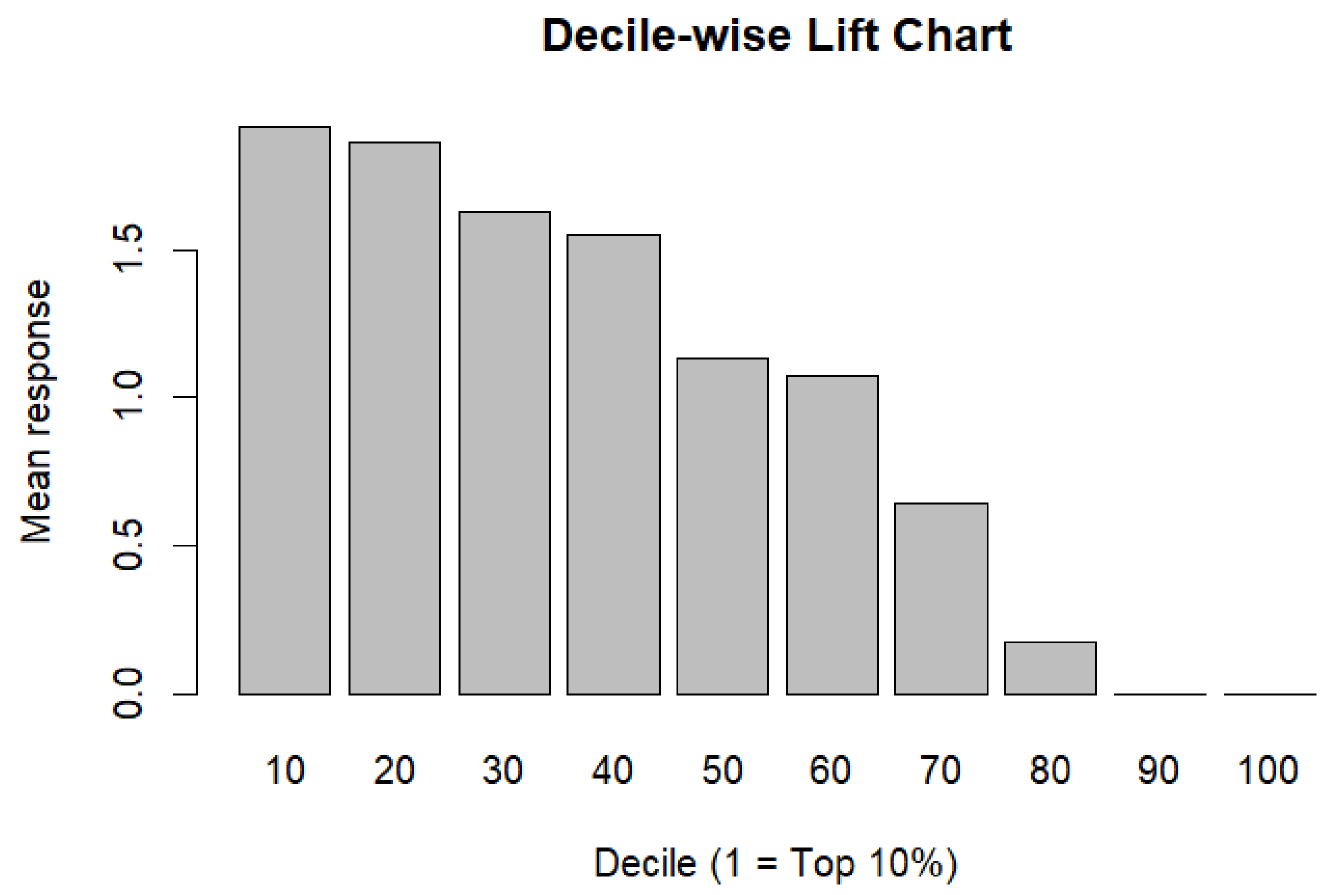
- 2,000 records (1,000 purchasers + 1,000 non-purchasers)
- 25 variables Mix of binary (US, Web Order, Gender) and numeric (Freq, Updates, Spending)
- No missing values detected in R
- Zeros = valid behaviour (not missing)

Classification Models & Lift Analysis

Model	Accuracy	Sensitivity	Specificity
Logistic Regression	79%	77.70%	80.30%
Decision Tree	78.60%	82.00%	75.10%
Random Forest	81.70%	86.30%	77.10%

- Real campaign base rate = 5.3 % response
- Model calibrated to the same base rate (× 0.106) for fair comparison.
- Top 10 % model segment → Lift = 1.9 (2× better than Real campaign base rate)
- Top 30 % → captures around 54 % of buyers with 70 % fewer mail pieces.
- Translates to higher ROI, lower acquisition cost, and less marketing waste.

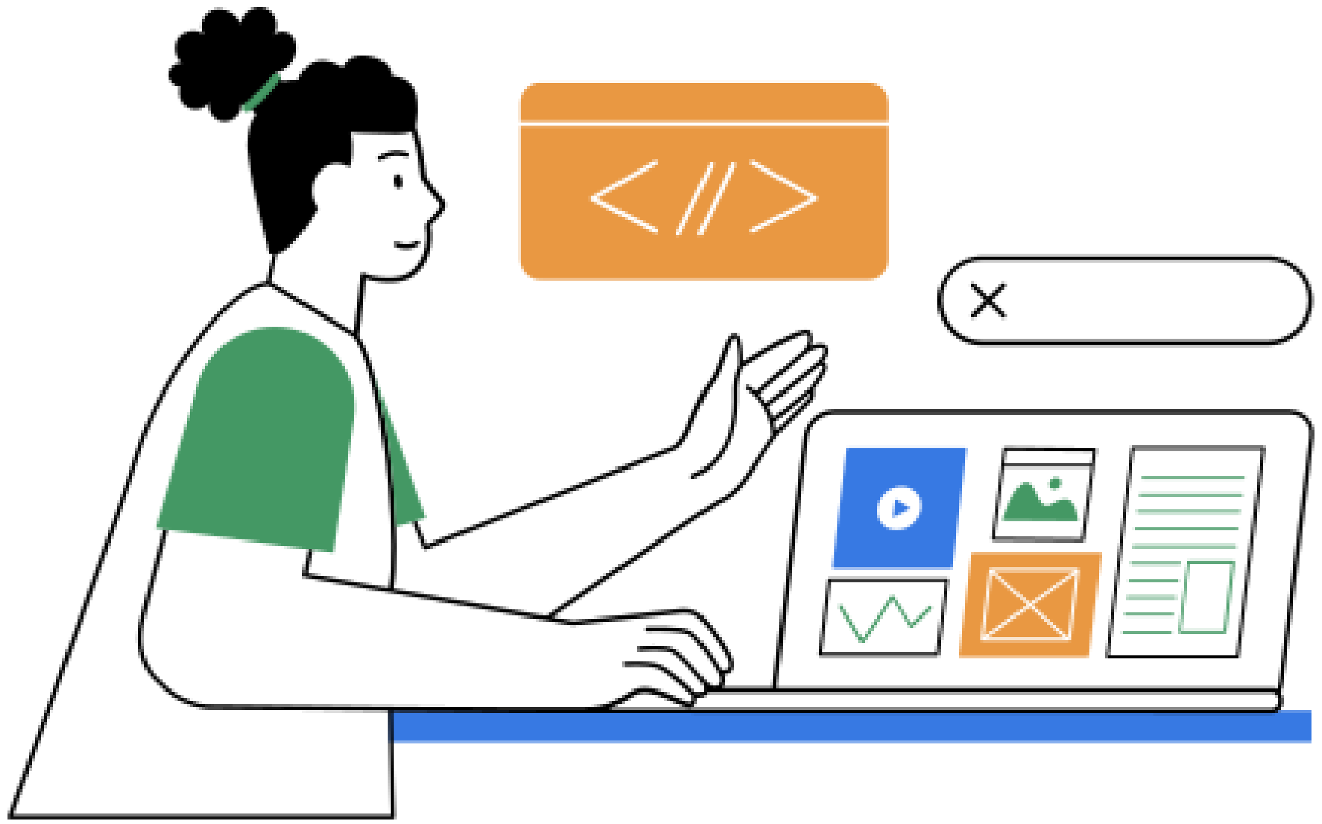
- Random forest captures complex buyer patterns catches 86% of true buyers, enabling data-driven targeting & cost savings.
- Logistic Regression is Interpretable → explains how Freq, Recency, Web Orders affect buying
- Decision Tree is a simple business rule; visual & easy to operationalize in marketing workflows.



Regression Models

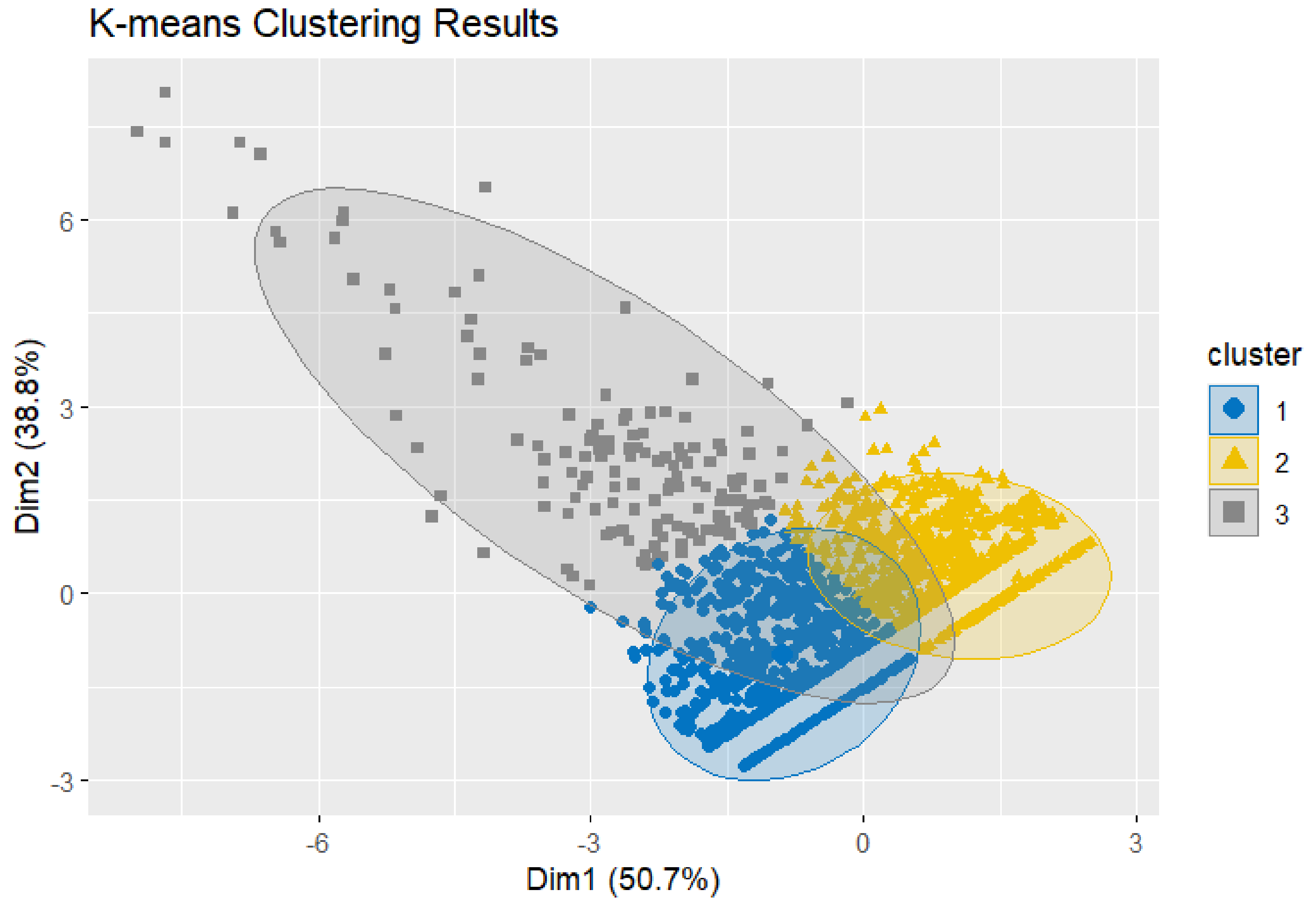
Model	R^2	RMSE
Multiple Linear Regression	0.536	131.27
Random Forest Regression	0.531	140.5

- Stepwise regression provides a simpler, interpretable model with high explanatory power ($R^2 \approx 0.54$), making it practical for forecasting campaign revenue and prioritizing high-spend segments.
- Frequency drives spend: each extra prior purchase ≈ +$95.5 expected spend.
- Recency matters: more recent updates → higher spend.
- Residential addresses spend less: about −$66 on average.
- Best way to use: combine spend prediction with purchase probability to rank expected revenue; allocate bigger offers to high-value tiers, lighter touch to low-value tiers; track ROI by decile.

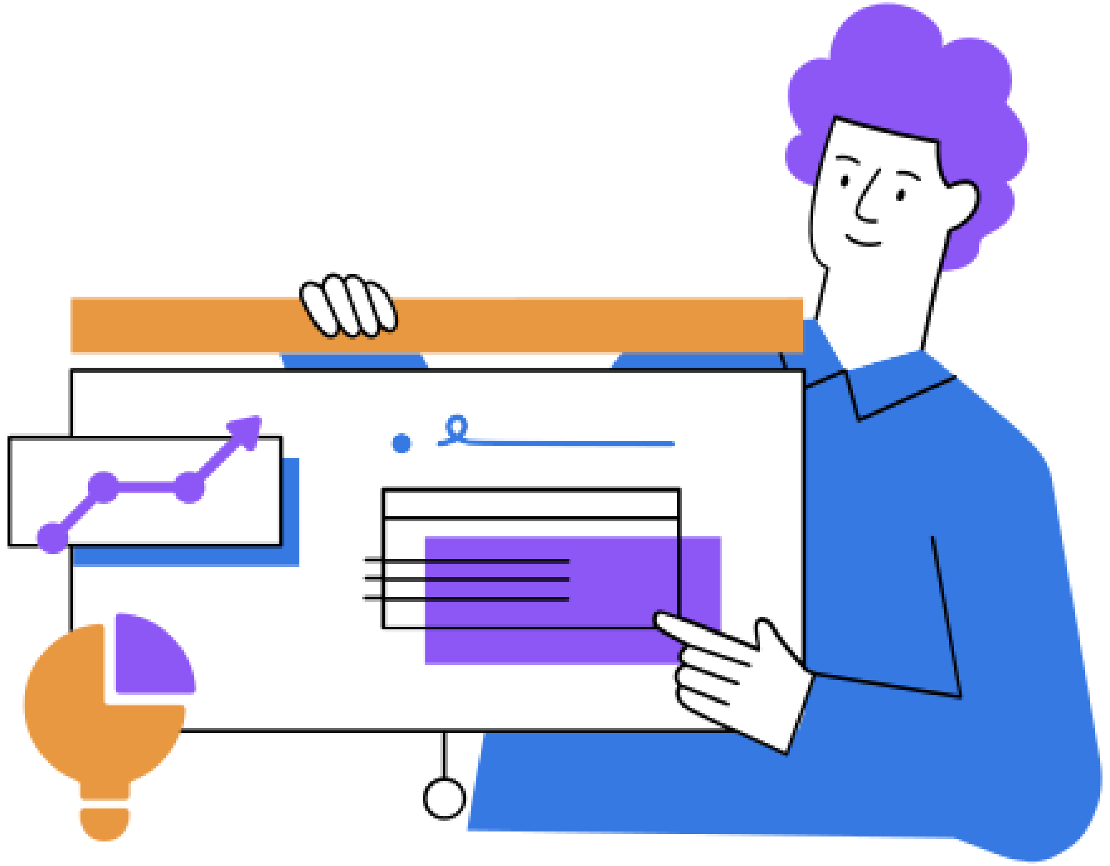


Clustering / Customer Segmentation

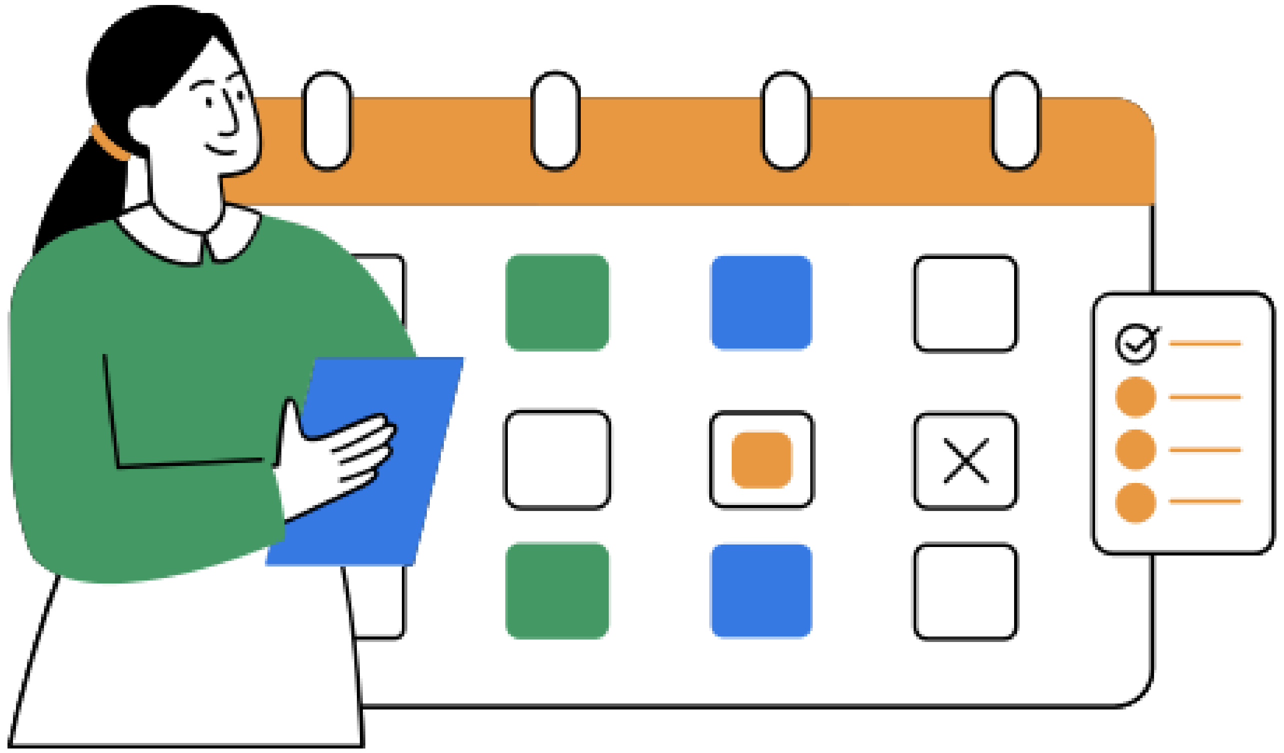
Cluster	Size (n)	Avg. Freq	Avg. Spending ($)	Avg. Last Update (Days)
Cluster 1 – "Warm"	804	1.41	73.9	1,149
Cluster 2 – "Cold"	1,056	0.96	60.8	3,076
Cluster 3 "VIP"	140	4.86	583.3	984



- Three clusters were identified using standardized numeric features.
- Cluster 3 (VIP) is confirming a small but high-value segment driving disproportionate revenue.
- Cluster 1 (Warm) is ideal for standard promotional campaigns. Cluster 2 (Cold) low-cost digital reactivation.
- These clusters enable tiered marketing strategies (e.g., premium offers for VIPs, nurture campaigns for warm leads, re-engagement for cold group).



Business Impact



Smarter Targeting
Predictive modeling improved targeting precision from random selection to data-driven mailing.
Top 30 % of customers captured 54 % of actual buyers.

Higher Response & ROI
Campaign response improved from 5.3 % → approx. 10 % (1.9× lift).
Similar sales achieved with ≈ 70 % fewer mail pieces, cutting cost substantially.

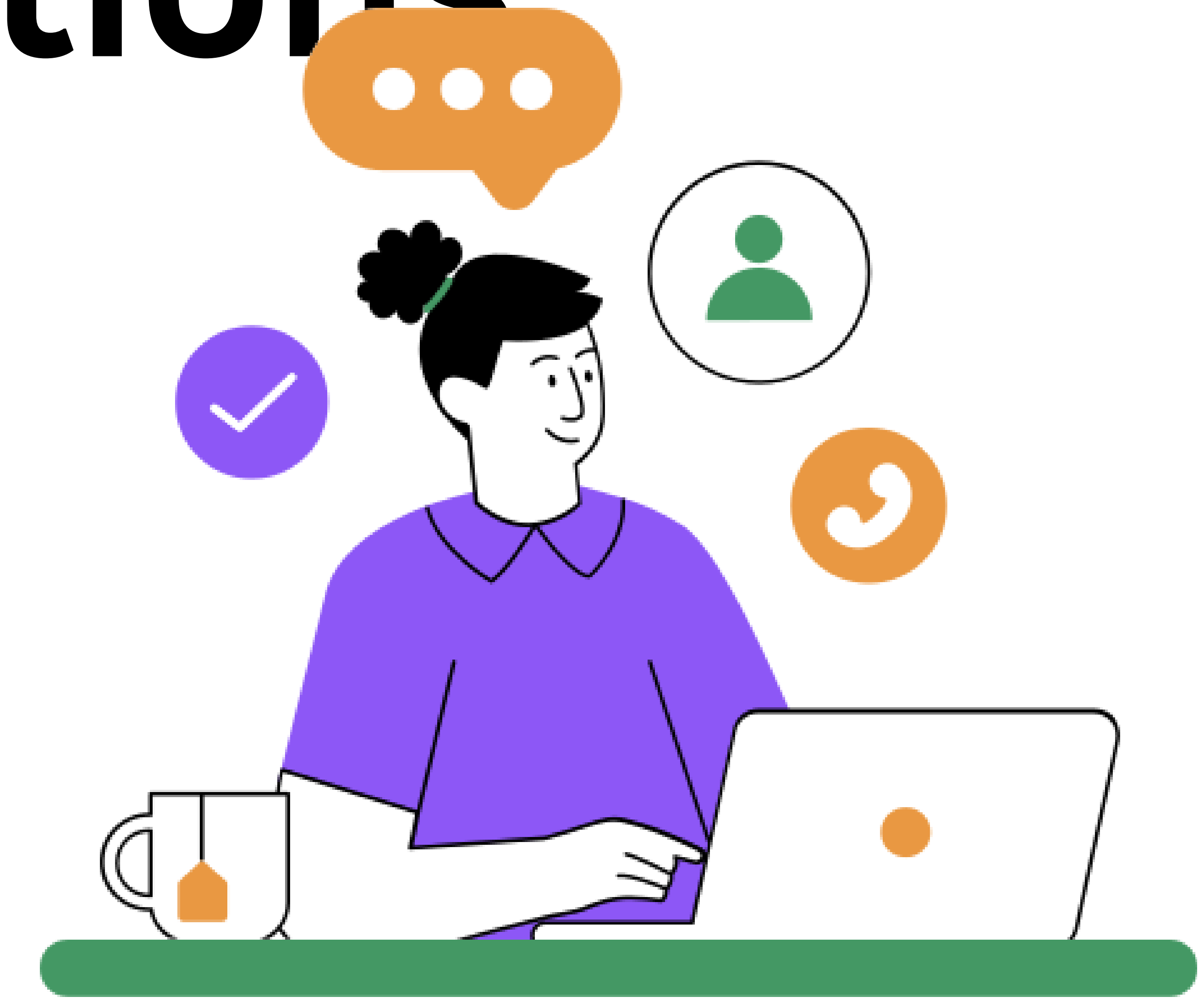
Revenue Optimization
Regression identified high-value buyers
Clustering revealed VIP / Warm / Cold segments for tiered offers and loyalty focus.

$$$

Analytics turned random mailing into a 2× more efficient, ROI-driven marketing strategy.

Conclusion & Recommendations

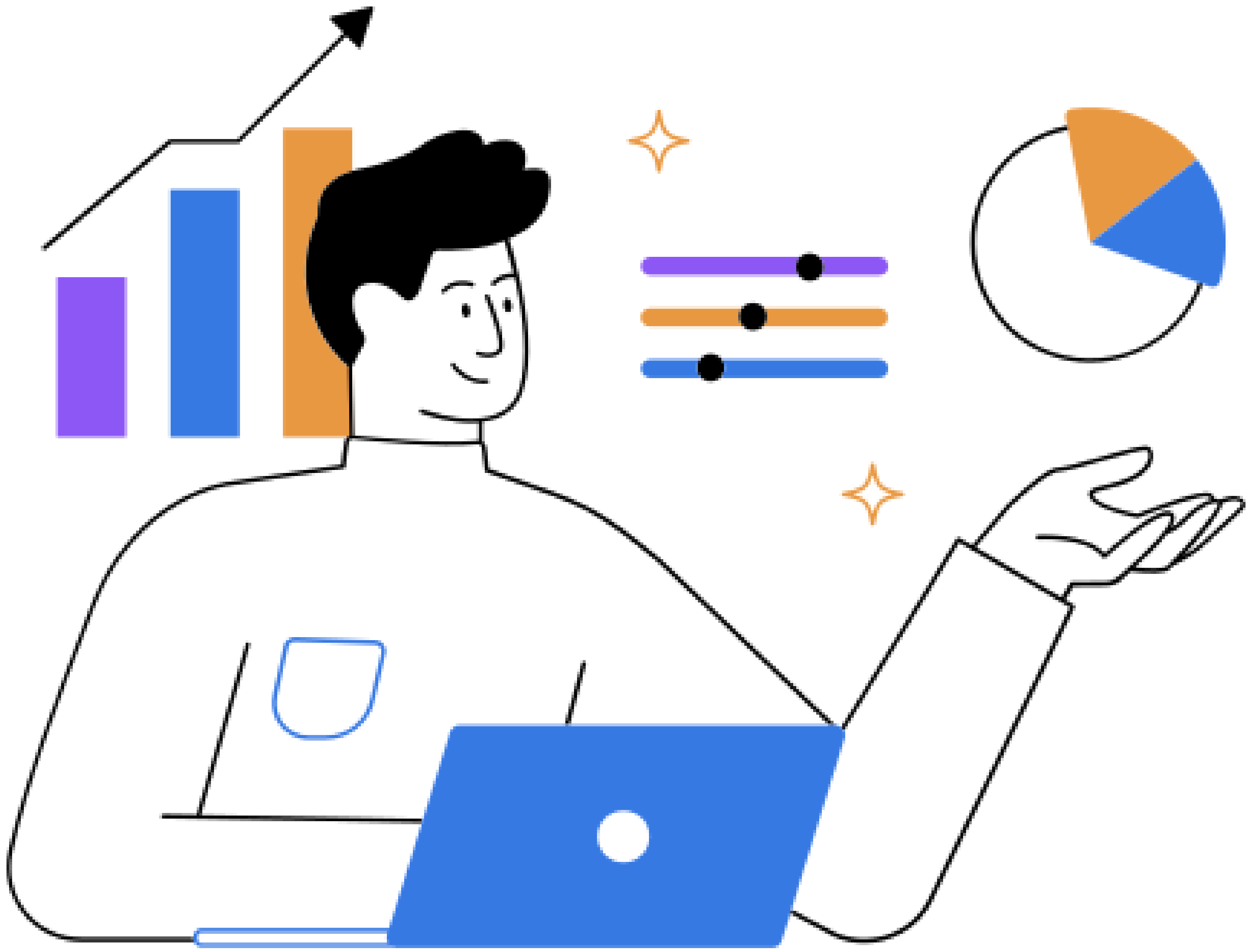
- Predictive analytics turned North-Point's random mailing into a precision-driven campaign system, improving response rate from 5.3 % to around 10 % and delivering around 2 times ROI.
- Models uncovered clear behavioral drivers. Frequency, Recency, and Web Orders, and segmented customers into VIP, Warm, and Cold tiers, allowing marketing teams to tailor outreach and offers intelligently.
- Regression results showed that each additional past purchase increases expected spend by approximately $95, enabling revenue-based prospect ranking and targeted upselling.
- With machine-learning-guided targeting, North-Point can reduce mail volume by 70 % while retaining most of the sales, investing saved costs into personalized digital engagement and loyalty initiatives.



To maintain impact, deploy the predictive models in upcoming campaigns, refresh data quarterly, monitor ROI lift, and automate dashboards for continuous learning and improvement.

Future Work

- Dynamic Mailing Optimization – Build an AI-driven system that continuously updates target lists based on live campaign response.
- Next-Best-Offer Prediction – Extend modeling to recommend what each customer is most likely to buy next.
- Automated Performance Dashboard – Deploy a real-time dashboard to monitor ROI, lift, and customer segments dynamically.



Thank You

Questions?

- Phanidhar Kasuba
- pkasubavenkatana@webster.edu
- Std ID: 4249169