Customer Churn Prediction

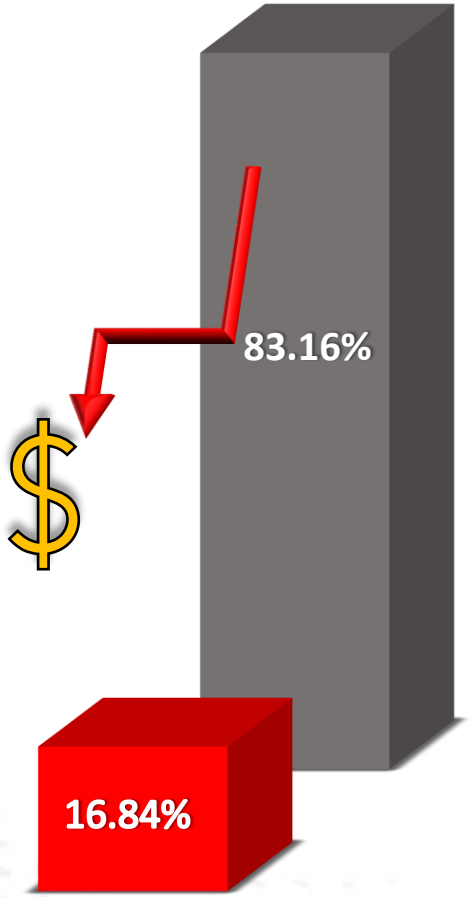
Data-Driven Retention Strategy

Presented By :
Rejin CR
28-03-2025

greatlearning
Power Ahead

The Churn Crisis

Business Problem



Problem:

"16.84% of customers churned yearly → $1.463B lost revenue."

Key constraints:

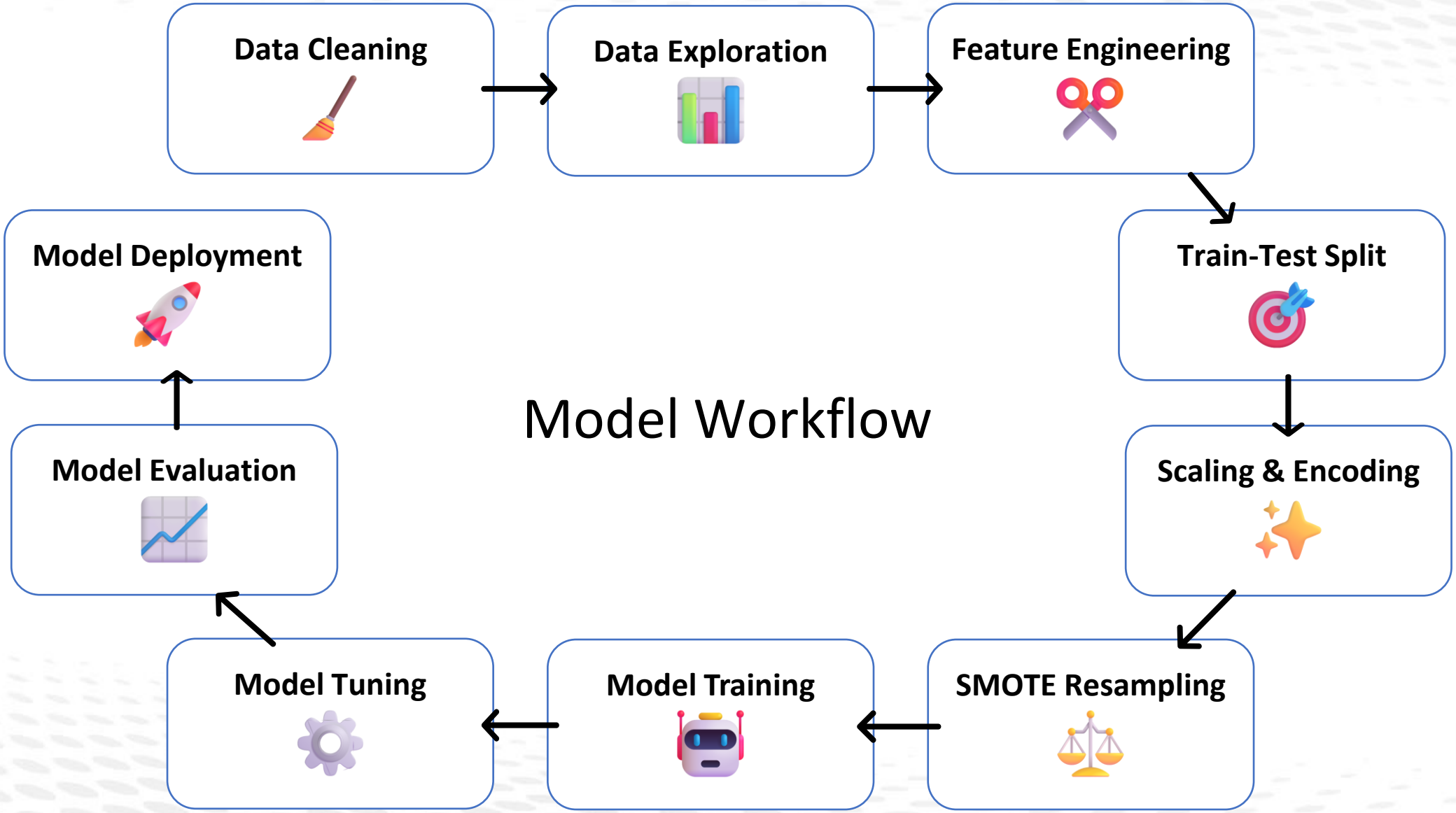
⚖️ *"Class imbalance (1:5 ratio)"*

❌ *"False negatives cost 5x more than false alarms"*

Objective :

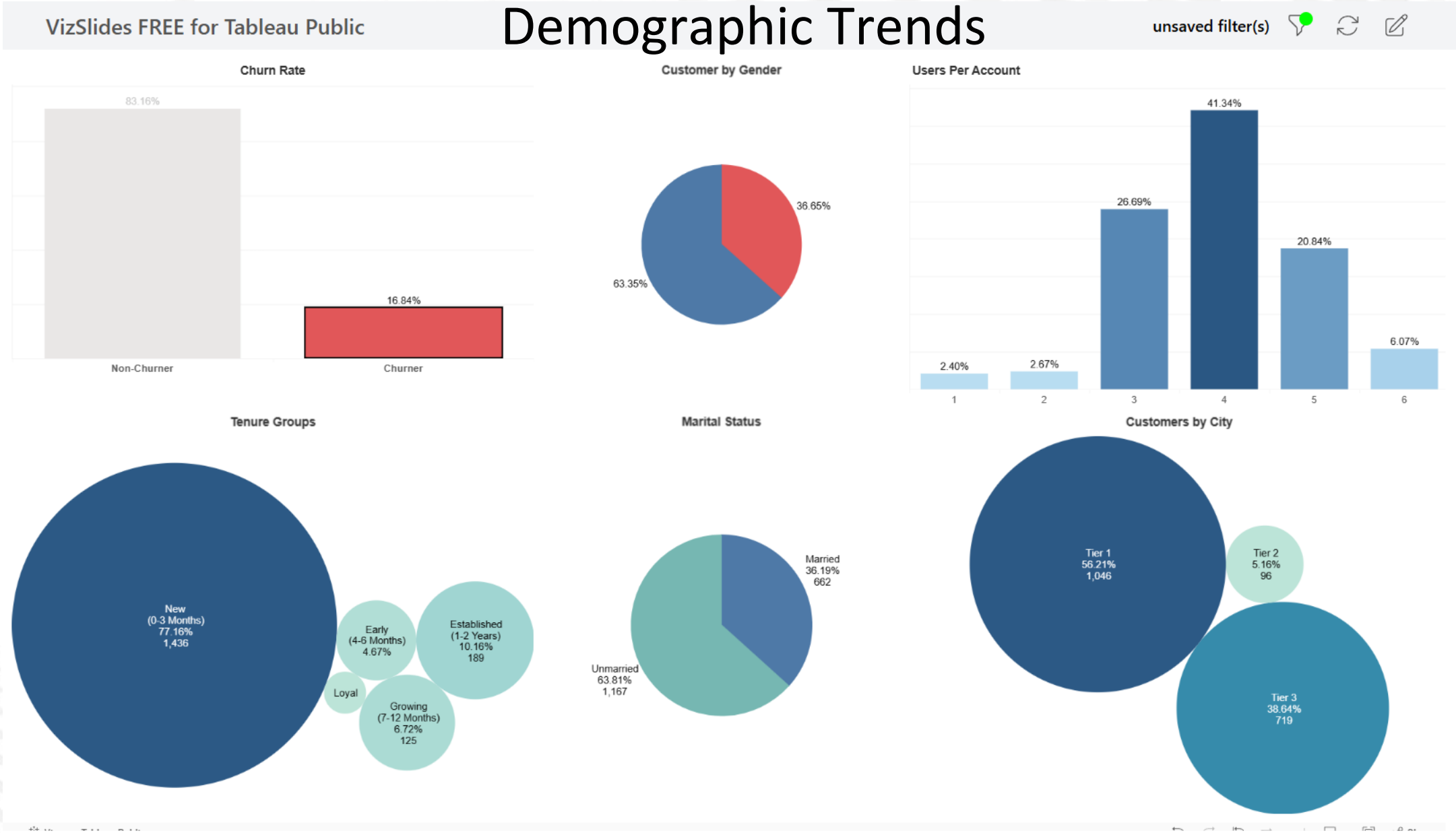
"Maximize Recall to identify ≥90% of potential churners for proactive retention"

Our Data Journey



Analytical Focus

Demographic Trends



Analytical Focus

Behavioural Risk Factors

VizSlides FREE for Tableau Public

Total Revenue generated LY

$8.69B

Avg. Revenue Growth % YoY

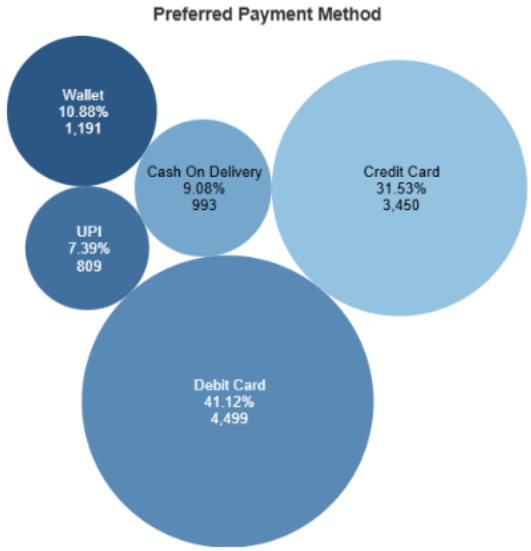
16.19%

Avg. Cashback Earned by User

$196.53

Avg. Coupon Usage

2



City Tier
(Multiple values)

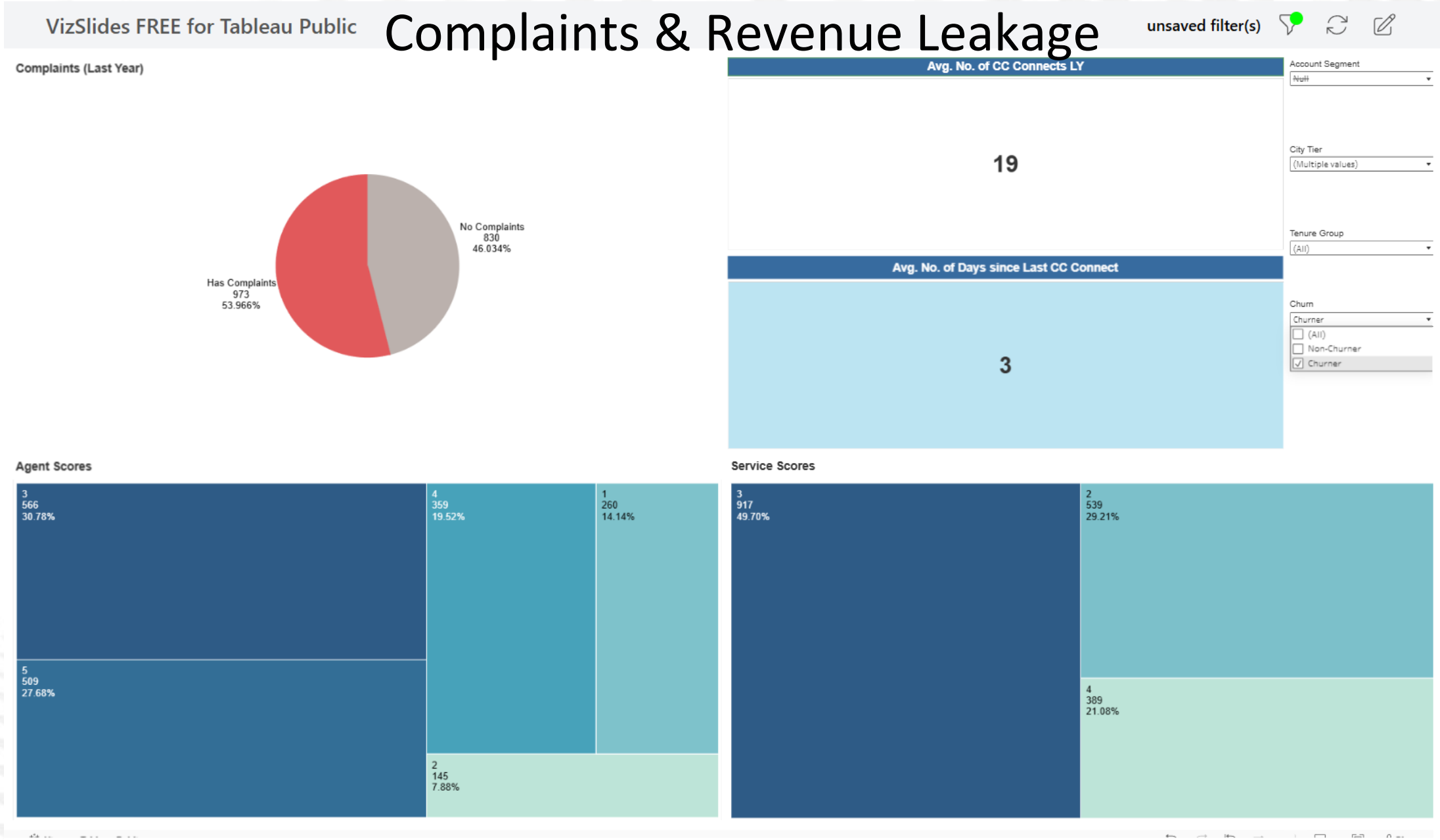
Tenure Group
(All)

Churn
(All)

Account Segment
Null

Analytical Focus

Complaints & Revenue Leakage



Model Showdown

Performance Comparison of Initial Models on Training Set

Model	Accuracy	Precision	Recall	F1-Score	ROC-AUC
DecisionTree	1	1	1	1	1
SVM	0.796281	0.44169	0.792422	0.567217	0.869775
RandomForest	1	1	1	1	1
AdaBoost	0.889259	0.747031	0.518122	0.611868	0.888891
GBM	0.913128	0.849169	0.588962	0.695525	0.942098
XGBoost	0.999445	1	0.996705	0.99835	1
LogisticRegression(statsmodel)	0.854704	0.558026	0.66145	0.605352	0.865096

Performance Comparison of Initial Models on Validation Set

Model	Accuracy	Precision	Recall	F1-Score	ROC-AUC
DecisionTree	0.920644	0.751572	0.788779	0.769726	0.868039
SVM	0.801332	0.448211	0.785479	0.570743	0.868189
RandomForest	0.95838	0.983051	0.765677	0.860853	0.991365
AdaBoost	0.883463	0.704846	0.528053	0.603774	0.889043
GBM	0.897336	0.775701	0.547855	0.642166	0.927184
XGBoost	0.954495	0.907749	0.811881	0.857143	0.986021
LogisticRegression(statsmodel)	0.854704	0.558026	0.66145	0.605352	0.865096

Model Showdown

Performance Comparison of Initial Models on Training Set

	Model	Accuracy	Precision	Recall	F1-Score	ROC-AUC
	DecisionTree	1	1	1	1	1
	SVM	0.796281	0.44169	0.792422	0.567217	0.869775
I	**RandomForest**	**1**	**1**	**1**	**1**	**1**
	AdaBoost	0.889259	0.747031	0.518122	0.611868	0.888891
	GBM	0.913128	0.849169	0.588962	0.695525	0.942098
II	**XGBoost**	**0.999445**	**1**	**0.996705**	**0.99835**	**1**
	LogisticRegression(statsmodel)	0.854704	0.558026	0.66145	0.605352	0.865096

Performance Comparison of Initial Models on Validation Set

	Model	Accuracy	Precision	Recall	F1-Score	ROC-AUC
	DecisionTree	0.920644	0.751572	0.788779	0.769726	0.868039
	SVM	0.801332	0.448211	0.785479	0.570743	0.868189
I	**RandomForest**	**0.95838**	**0.983051**	**0.765677**	**0.860853**	**0.991365**
	AdaBoost	0.883463	0.704846	0.528053	0.603774	0.889043
	GBM	0.897336	0.775701	0.547855	0.642166	0.927184
II	**XGBoost**	**0.954495**	**0.907749**	**0.811881**	**0.857143**	**0.986021**
	LogisticRegression(statsmodel)	0.854704	0.558026	0.66145	0.605352	0.865096

Model Showdown

Performance Comparison of Models on Training Set (After SMOTE)

Model	Accuracy	Precision	Recall	F1-Score	ROC-AUC
DecisionTree	1	1	1	1	1
SVM	0.804322	0.801554	0.808912	0.805216	0.87835
RandomForest	1	1	1	1	1
AdaBoost	0.875167	0.878962	0.87016	0.874539	0.947459
GBM	0.927654	0.934248	0.92006	0.9271	0.979072
XGBoost	0.999666	1	0.999332	0.999666	0.999998
LogisticRegression(statsmodel)	0.796646	0.787761	0.812083	0.799737	0.870985

Performance Comparison of Models on Validation Set (After SMOTE)

Model	Accuracy	Precision	Recall	F1-Score	ROC-AUC
DecisionTree	0.909545	0.708333	0.785479	0.744914	0.860051
SVM	0.801887	0.449057	0.785479	0.571429	0.86887
RandomForest	0.95838	0.895833	0.851485	0.873096	0.986667
AdaBoost	0.846282	0.530374	0.749175	0.621067	0.884551
GBM	0.885683	0.645646	0.709571	0.676101	0.916926
XGBoost	0.955605	0.885813	0.844884	0.864865	0.983593
LogisticRegression(statsmodel)	0.796646	0.787761	0.812083	0.799737	0.870985

Model Showdown

Performance Comparison of Models on Training Set (After SMOTE)

	Model	Accuracy	Precision	Recall	F1-Score	ROC-AUC
	DecisionTree	1	1	1	1	1
II	**SVM**	**0.804322**	**0.801554**	**0.808912**	**0.805216**	**0.87835**
	RandomForest	1	1	1	1	1
	AdaBoost	0.875167	0.878962	0.87016	0.874539	0.947459
	GBM	0.927654	0.934248	0.92006	0.9271	0.979072
I	**XGBoost**	**0.999666**	**1**	**0.999332**	**0.999666**	**0.999998**
	LogisticRegression(statsmodel)	0.796646	0.787761	0.812083	0.799737	0.870985

Performance Comparison of Models on Validation Set (After SMOTE)

	Model	Accuracy	Precision	Recall	F1-Score	ROC-AUC
	DecisionTree	0.909545	0.708333	0.785479	0.744914	0.860051
II	**SVM**	**0.801887**	**0.449057**	**0.785479**	**0.571429**	**0.86887**
	RandomForest	0.95838	0.895833	0.851485	0.873096	0.986667
	AdaBoost	0.846282	0.530374	0.749175	0.621067	0.884551
	GBM	0.885683	0.645646	0.709571	0.676101	0.916926
I	**XGBoost**	**0.955605**	**0.885813**	**0.844884**	**0.864865**	**0.983593**
	LogisticRegression(statsmodel)	0.796646	0.787761	0.812083	0.799737	0.870985

Model Showdown

Performance Comparison of Tuned models on Training Set

Model	Accuracy	Precision	Recall	F1-Score
RandomForestClassifier	1	1	1	1
XGBClassifier	0.977887	0.977648	0.978138	0.977893
LogisticRegression	0.790971	0.77723	0.815754	0.796026
SVC	0.994993	0.992364	0.997664	0.995007

Performance Comparison of Tuned models on Validation Set

Model	Accuracy	Precision	Recall	F1-Score
RandomForestClassifier	0.959489	0.891156	0.864686	0.877722
XGBClassifier	0.924528	0.761755	0.80198	0.78135
LogisticRegression	0.779134	0.418803	0.808581	0.551802
SVC	0.963374	0.881029	0.90429	0.892508

Model Showdown

Performance Comparison of Tuned models on Training Set

Model	Accuracy	Precision	Recall	F1-Score
RandomForestClassifier	**1**	**1**	**1**	**1**
XGBClassifier	0.977887	0.977648	0.978138	0.977893
LogisticRegression	0.790971	0.77723	0.815754	0.796026
SVC	**0.994993**	**0.992364**	**0.997664**	**0.995007**

Performance Comparison of Tuned models on Validation Set

Model	Accuracy	Precision	Recall	F1-Score
RandomForestClassifier	**0.959489**	**0.891156**	**0.864686**	**0.877722**
XGBClassifier	0.924528	0.761755	0.80198	0.78135
LogisticRegression	0.779134	0.418803	0.808581	0.551802
SVC	**0.963374**	**0.881029**	**0.90429**	**0.892508**

Model Showdown

Best Model: SVC

- ***Highest Recall on Validation Set: 0.90429***
- ***Balanced Performance:*** *Strong precision (0.881) and F1-score (0.893)*
- ***Generalization:*** *Small gap between training (0.998) and validation (0.904) recall, suggesting minimal overfitting.*

Backup Model: RandomForest

- ***Second-Highest Recall:*** *0.864686 (better than XGBoost and LogisticRegression).*
- ***Robustness:*** *Handles non-linear patterns well and less prone to overfitting than XGBoost (validation recall drop: 0.978 → 0.802).*
- ***Interpretability:*** *Provides feature importance for business insights.*

Why SVC Won

Technical Edge

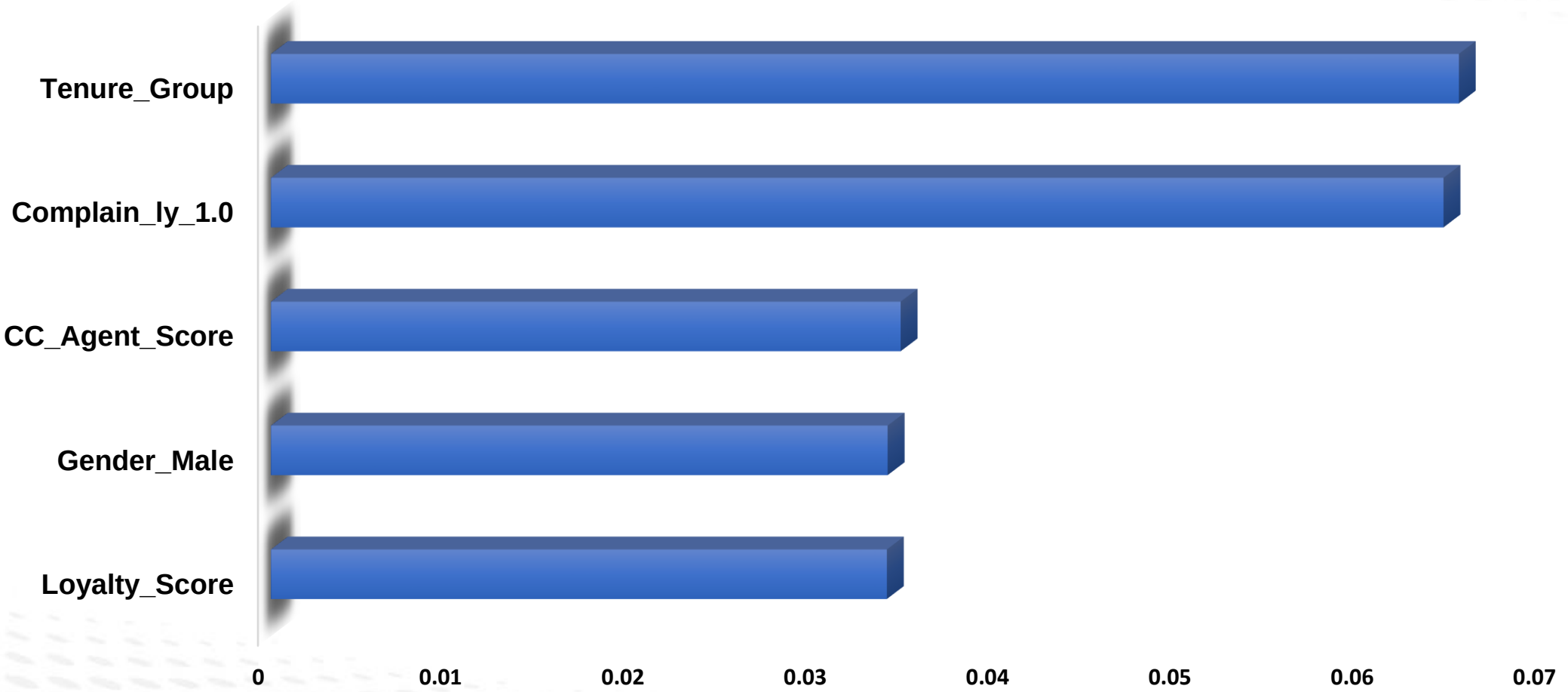
- ***Best at capturing true churners***
  Recall on Validation Set: 0.90429
- ***Strong Precision and F1-score indicating fewer false alarms.***
  Strong precision (0.881) and F1-score (0.893)
- ***Best Generalization and minimal overfitting.***
  Highest Accuracy (0.963374)
  Minimal Difference b/w Train and Val scores(0.089)

SVM Model Performance on Test Set

Model	Accuracy	Precision	Recall	F1-Score
SVC	0.968028	0.911528	**0.897098**	0.904255

Churn Red Flags

Top Features



Action Plan

Recommendations

🚀 "Now" (Quick Wins, High Impact)

1.Launch a "First 90 Days" Retention Program

- *Target: Short-tenure customers (<6 months).*
- *Actions:*
  - *Send personalized welcome kits.*
  - *Offer a "loyalty bonus" (e.g., 10% cashback after 3 months).*

2. 24-Hour Complaint Resolution Pledge

- *Assign a rapid-response team to address complaints flagged in the last year.*
- *Track resolution time and customer satisfaction post-fix.*

3. Boost CC Agent Performance

- *Implement weekly training sessions focused on low-scoring agents (scores ≤2).*
- *Tie bonuses to customer satisfaction metrics.*

Action Plan

Recommendations

⏳ "Next" (Mid-Term, Strategic)

1. Personalized Campaigns for High-Risk Segments

- *Target: Single males in Tier 1 cities.*
- *Actions:*
  - *Curate gender-specific offers (e.g., "Exclusive Male-Only Discounts").*
  - *Partner with local influencers for hyper-local engagement.*

2. Payment Method Optimization

- *Phase out COD for high-churn segments; incentivize UPI/Wallet use with 1% cashback.*

3. Predictive Churn Alerts

- *Integrate model insights into CRM to flag at-risk customers (e.g., low loyalty score + recent complaint).*
- *Auto-trigger retention offers (e.g., "We miss you! Here's 15% off").*

Action Plan

Recommendations

🔮 "Future" (Long-Term, Scalable)

1. Dynamic Pricing

- *Adjust discounts/rewards in real-time based on churn probability.*

2. Gamified Loyalty Program

- *Launch tiered rewards (e.g., "Gold Tier" for consistent spenders) with non-monetary perks (early access, VIP support).*

3. City-Tier Service Hubs

- *Open localized support centers in Tier 1 cities to address geographic churn drivers.*

Q&A Ready

Your Questions

Thank You