Customer Segmentation and Retention Strategies for Winter Sports Tourism

Chiara Pelizza
Davide Tortorella

Today's Agenda

Presentation Outline

1. Introduction & Business Context
2. Data Overview
3. Driving Questions
4. Models Developed: RFM, Churn, Sentiment Analysis
5. Marketing Campaigns
6. Conclusions & Future Perspectives

Business Context

Snowit is a sports and leisure company, focused on skiing services.

This presentation aims to analyze customer behavior and develop data-driven strategies.

Understanding churn and customer value helps maximize retention and revenue.

Data Sources

Our dataset includes over 385 000 orders made by 728 000+ users between 2022 and 2025,

Users & Profiles: customer demographics (age, gender, location, preferences)

Orders & Order Details: transaction history, product types, purchase frequency

Cards: skipass assignment and usage status

Reviews & Labelled Reviews: customer feedback for sentiment analysis

Note: due to privacy reasons, reviews and labelled reviews are decontextualized from the business problem, they were used just to build a NLP model

Weather Data Integration

To enrich our set of variables, we included weather data for each customer's location

Data retrieved via open APIs (Open-Meteo) and mapped to customer zones

Features aggregated over lookback periods prior to reference date:

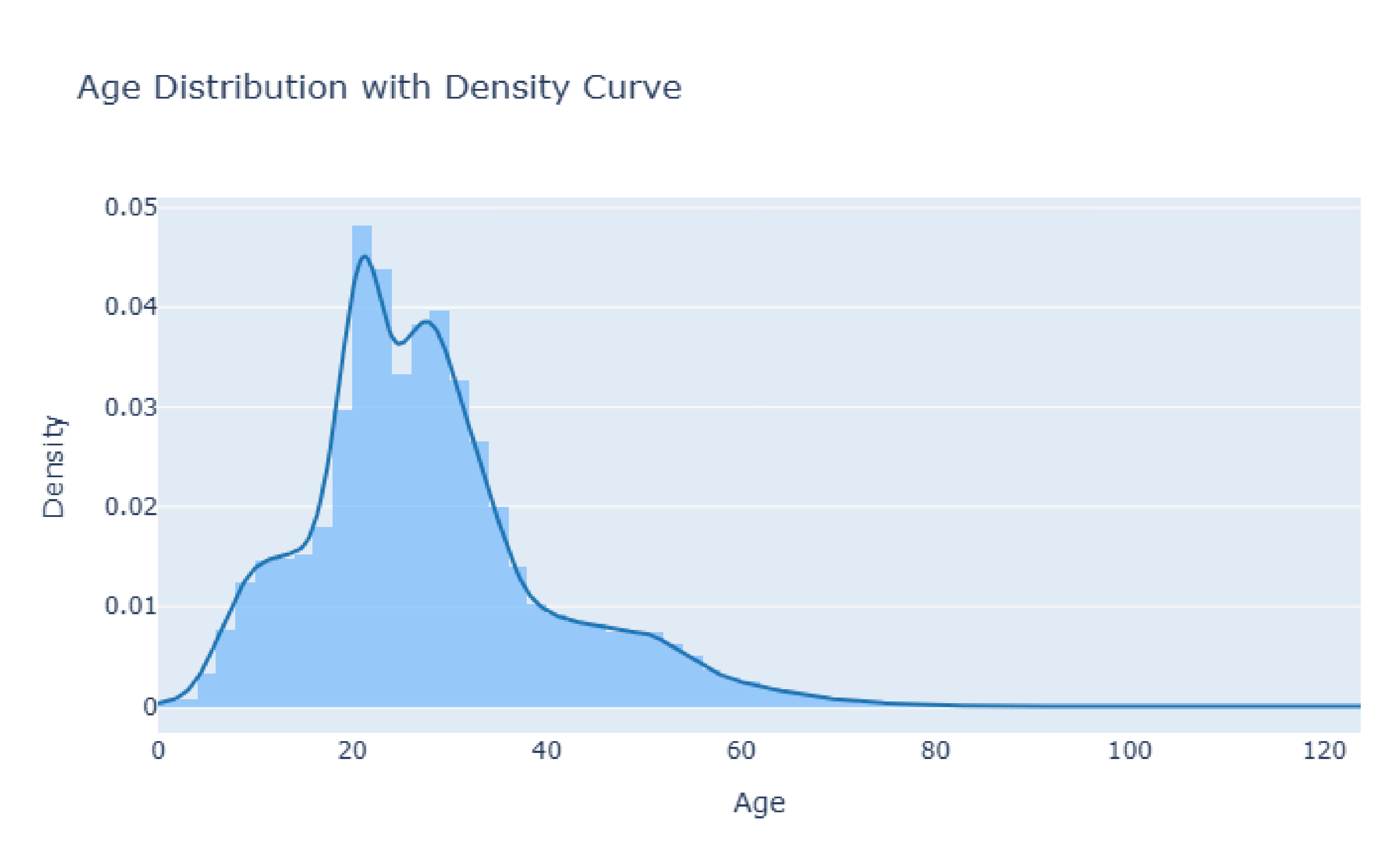
- Total snowfall
- Maximum snow depth
- Average temperature
- Number of snowless days

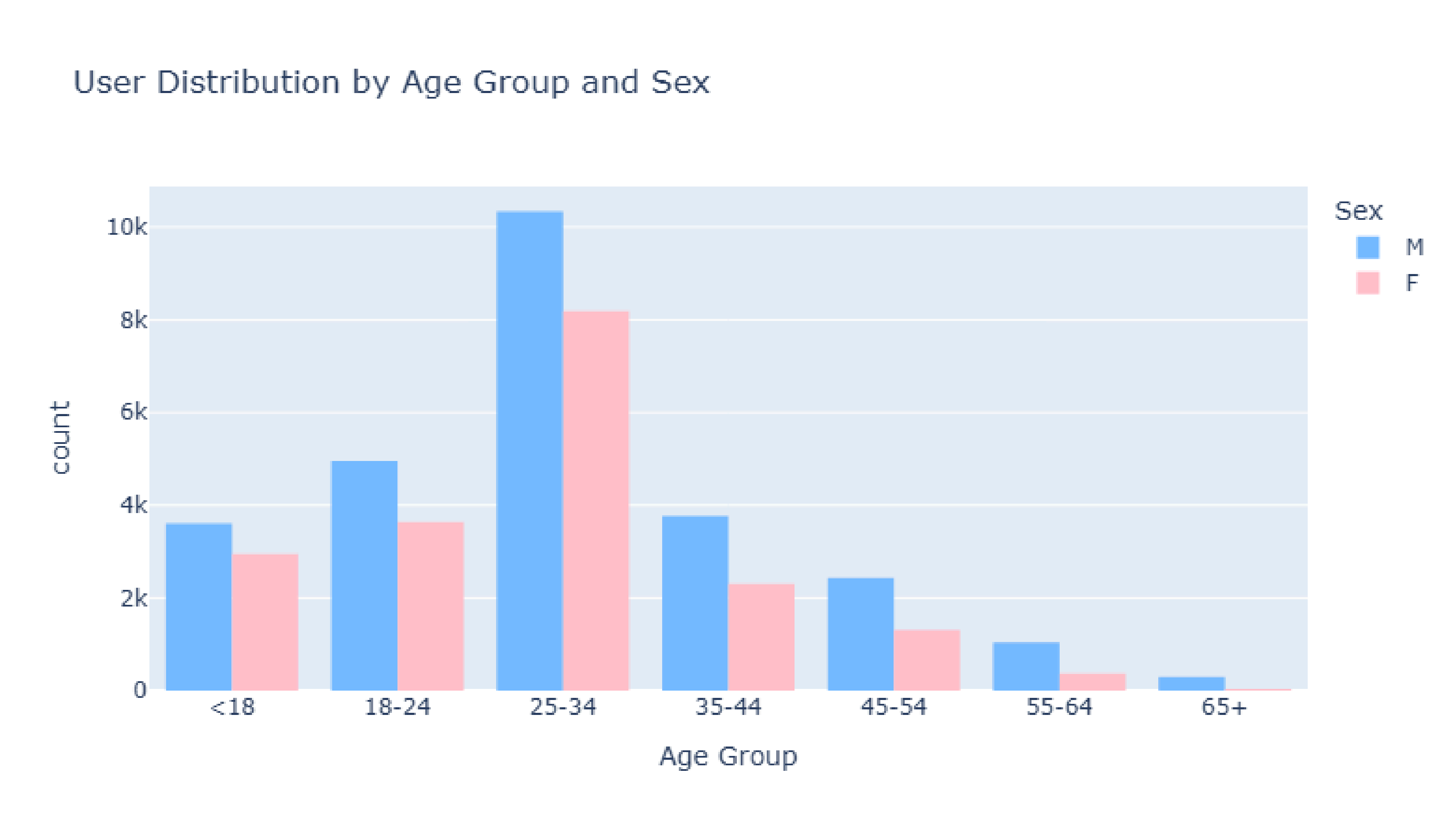
Purpose: capture local weather effects on customer churn

Only used for within-season churn prediction (mid-season campaign)

Data Overview

Customer Demographics

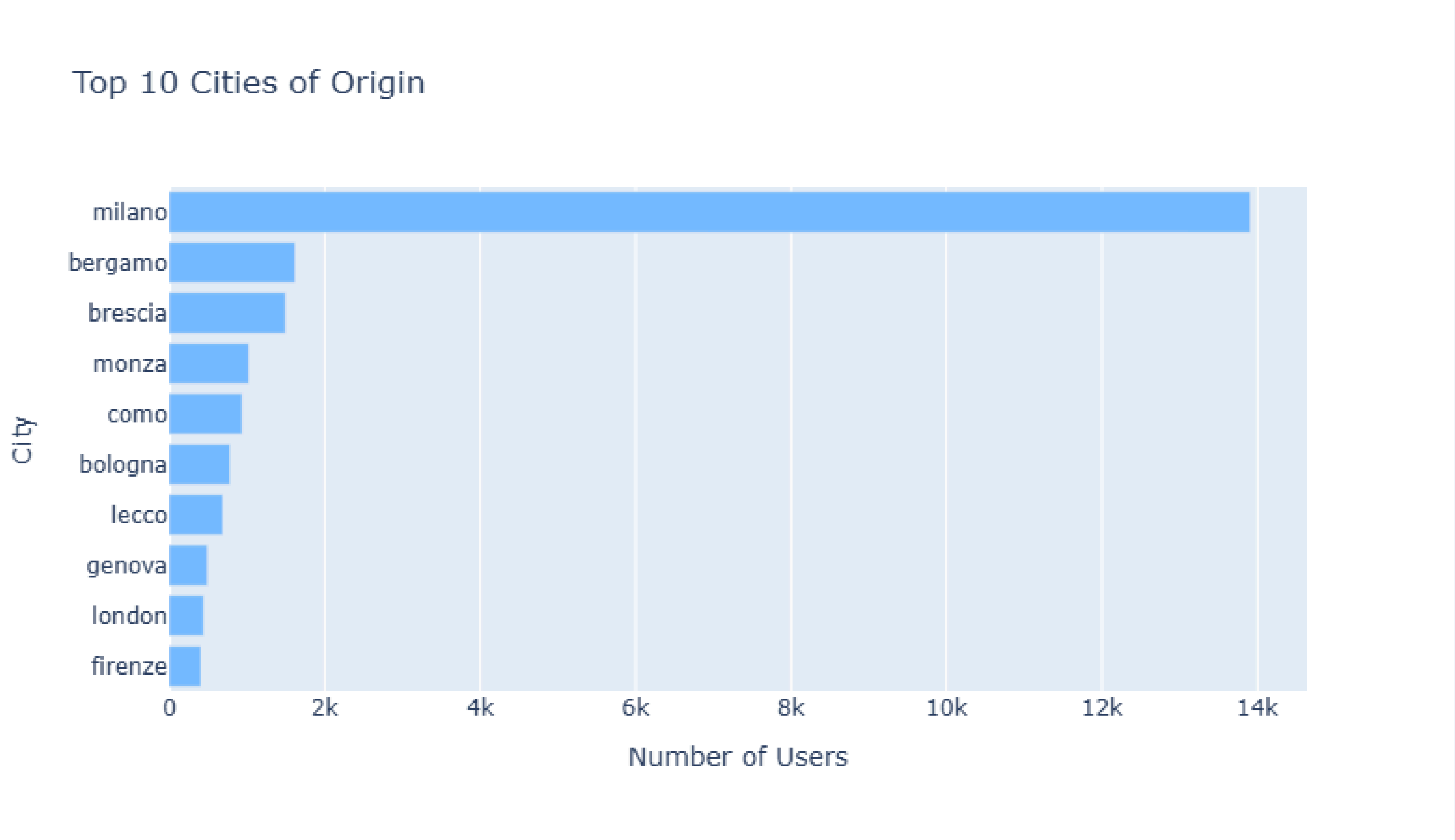


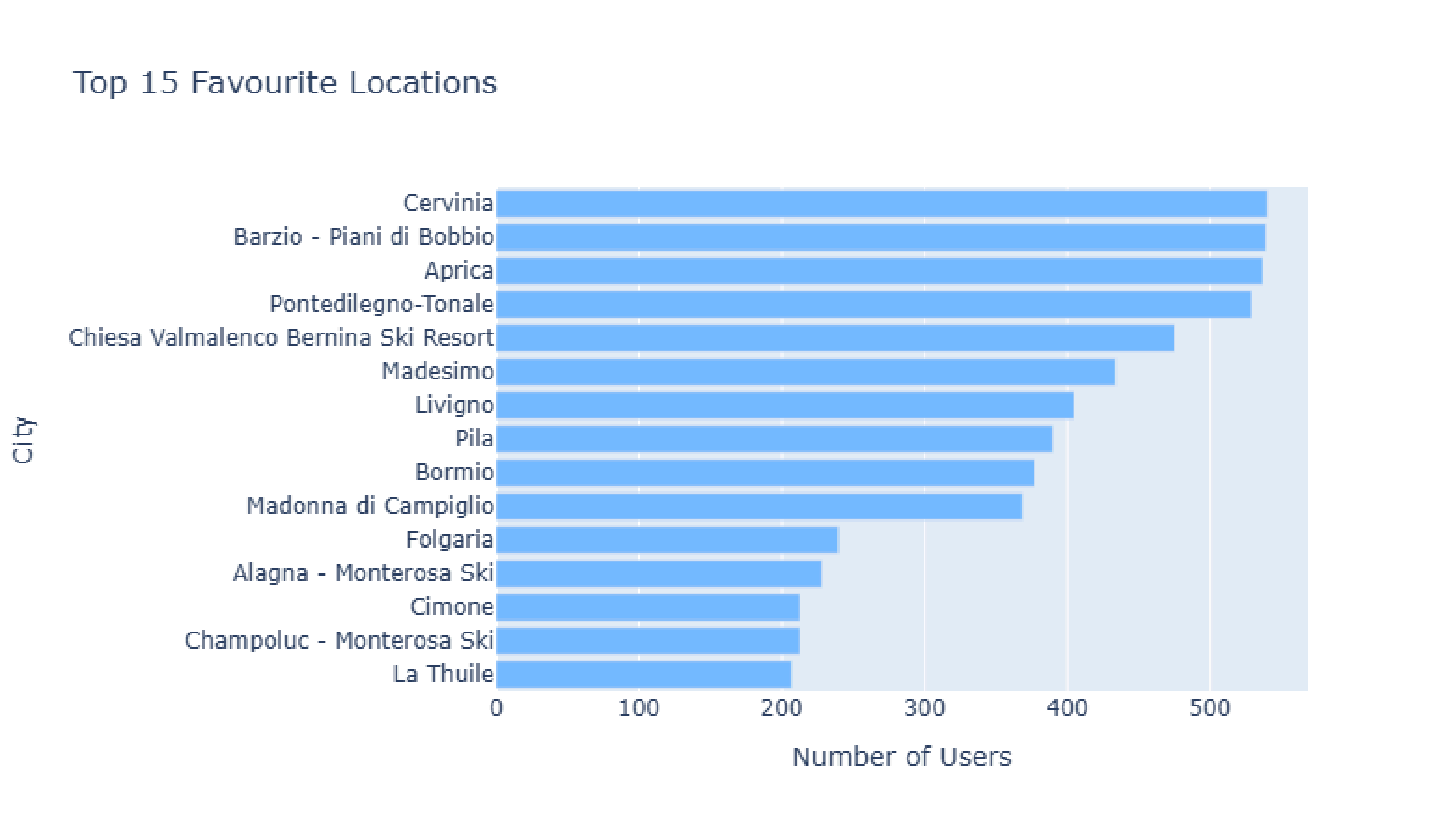


Most users fall in the middle-age range (30–50), with males consistently outnumbering females across all age groups.

Data Overview

Customer Origins & Preferred Vacation Locations

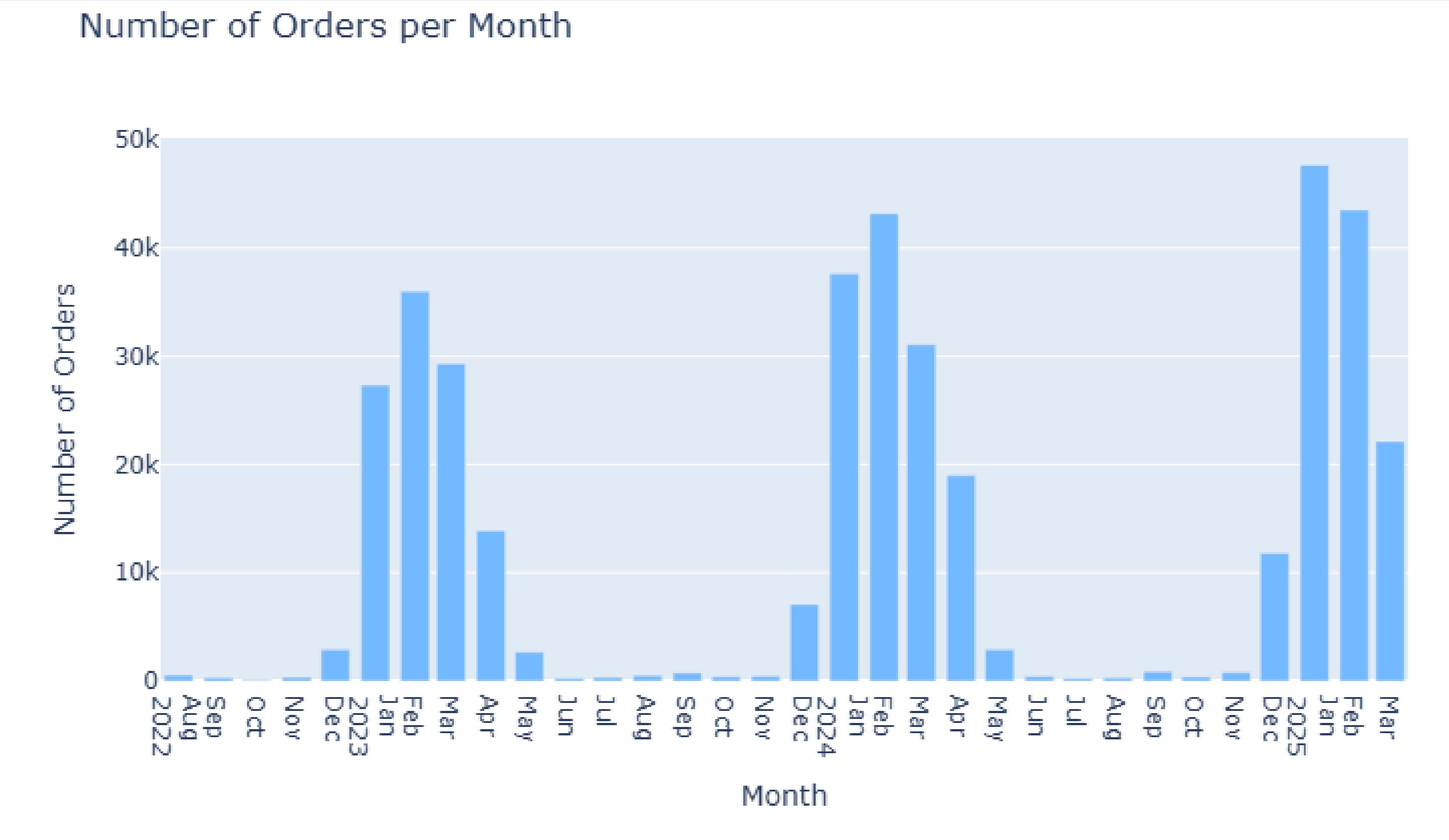


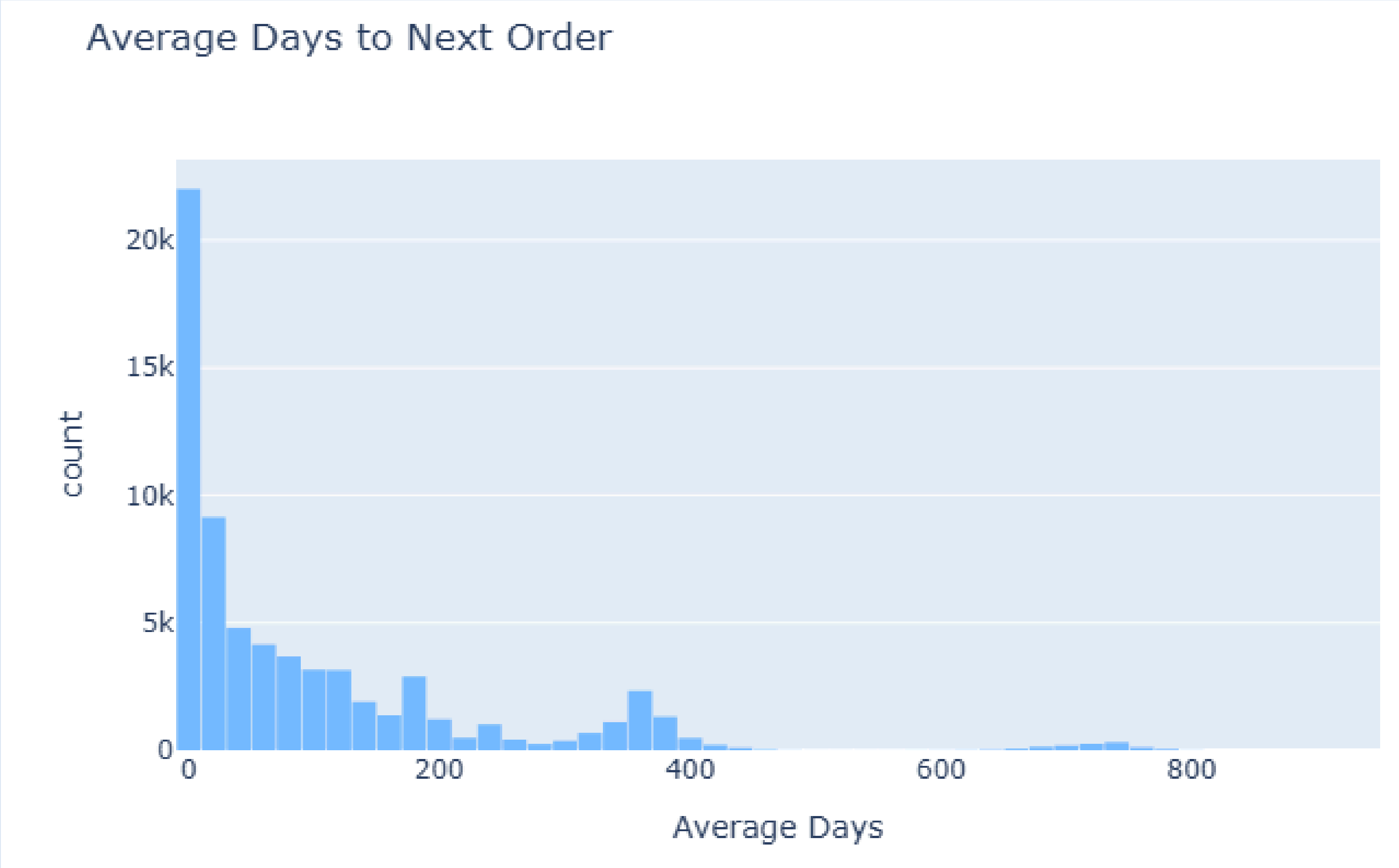


Most customers come from Milan, but top vacation destinations highlight popular ski resorts, guiding targeted marketing and promotions.

Data Overview

Seasonality & Purchase Frequency





Strong winter peaks and short repeat-purchase gaps reveal a clear seasonality that adds complexity to the analysis

Driving Questions

Framing marketing decisions in a seasonal contex

Which high-value customers are at risk of churn?

How should we target them with pre-season and mid-season campaigns to maximize retention and revenue?

Our Approach

RFM Model:
Customer value segmentation

Sentiment Analysis:
Customer feedback evaluation

Churn Model:
Churn propensity prediction

Integrating these models enables data-driven targeting for pre and mid-season campaigns.

RFM Model

RFM segments active customers by purchase behavior to identify top-value clients.

Combine R, F, M scores to obtain the following customer categories:

		RF				
		Top	Leaving Top	Engaged	Leaving	One-timer
MONETARY VALUE	High	Diamond	Gold	Silver	Bronze	Copper
	Medium	Gold	Silver	Bronze	Copper	Tin
	Low	Silver	Bronze	Copper	Tin	Cheap

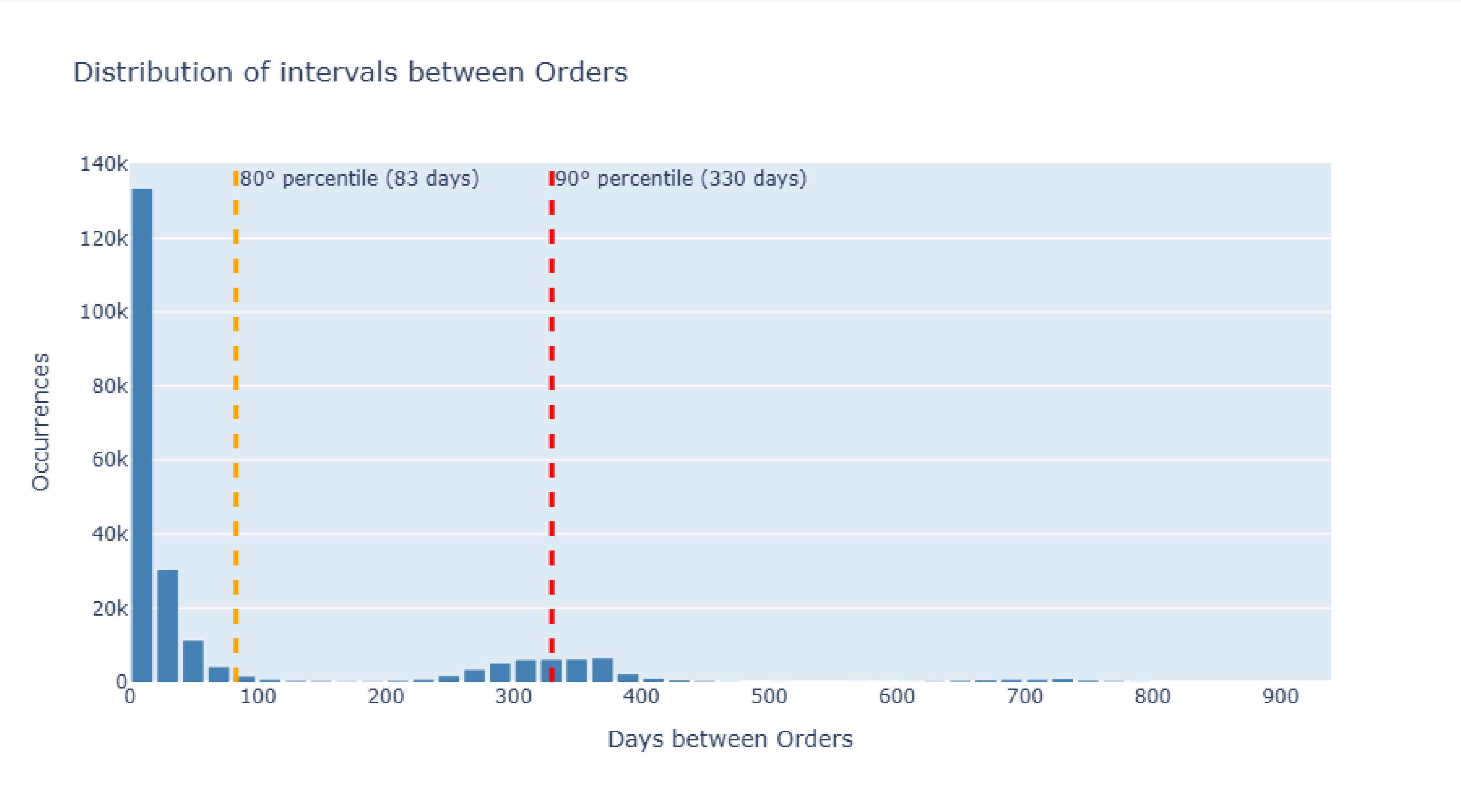
Diamond ➔ Gold ➔ Silver ➔ Bronze ➔ Copper ➔ Tin ➔ Cheap

Definitions:

- Recency (R): Days since last purchase
- Frequency (F): Total number of purchases in the reference period
- Monetary (M): Total amount spent

Snowit RFM Findings

Defining engagement based on recent purchase behavior



Given the strong seasonality of the market, both the 80th percentile (83 days) and the 90th percentile (330 days) of the repurchase curve are relevant thresholds.

The 330-day cutoff is used in the between-season analysis, while the 83-day cutoff is applied in the within-season analysis to identify active/inactive users

Churn Model

Predicts the likelihood that a customer becomes inactive, helping prioritize retention efforts.

For each campaign we define:

- Reference Date: date used to assess customer activity
- Look-back Period: historical timeframe for feature extraction
- Holdout Period: future timeframe for evaluating model predictions

- Churn: customers active in the lookback period but inactive in the holdout period

Churn Modeling Approach

Algorithms Evaluated:

✓ **Logistic Regression:** Interpretable linear model, fast to train and easy to explain

✓ **Decision Tree**: Transparent rule-based splits, intuitive for business users

✓ **Random Forest**: Robust to noise, captures non-linear relationships and feature interactions

✓ **Gradient Boosting**: Sequential trees that learn from errors, strong predictive accuracy

✓ **XGBoost**: Optimized gradient boosting with regularization for top performance on large datasets

We identify **Gradient Boosting** as the best-performing model based on the following key evaluation metrics

Model Evaluation Metrics

Key Metrics:

- ✓ **Precision**: proportion of predicted churners who actually churn
- ✓ **Recall**: proportion of actual churners correctly identified
- ✓ **F1 Score**: balance between precision and recall, useful for imbalanced datasets
- ✓ **ROC-AUC**: ability to distinguish between churners and non-churners

Note: due to class imbalance, accuracy was not considered a reliable evaluation metric

These metrics help identify models that are both reliable and actionable for targeting potential churners.

Sentiment Analysis

Supervised text analytics technique that categorizes opinions expressed in text as **positive**, **negative** or **neutral**

Text Input → Processing & Transformation → Model Analysis → Insights

Turning feedback from reviews, surveys and social media into actionable knowledge helps businesses identify trends, issues and opportunities

Sentiment Analysis Approach

Algorithms Evaluated:

- ✓ **Logistic Regression:** simple, interpretable linear model
- ✓ **Support Vector Machines (SVM)**: effective for high-dimensional text data
- ✓ **Naive Bayes**: probabilistic model, works well for word frequencies

We select **Logistic Regression** as the best-performing model on the labelled data using the same evaluation metrics previously applied for the churn analysis.

After applying the model, the unlabelled dataset was classified into **65.932** positive reviews, **20.705** neutral reviews, and **6.792** negative reviews

Pre-Season Marketing Campaign

Objective: Prevent churn and maximize pre-season revenue.

Trigger: Once per year, before the winter season (e.g. Oct 1st).

Target:
- High-value clients from the previous season (Diamond, Gold, Silver)
- Customers with high probability of churn (>0.80)

Call to Action: Renew or upgrade ski pass before the season begins.

Offer
- 5 % discount on a single skipass.
- 10 % discount on the season pass.

Rationale:
- Re-engages customers most at risk of inactivity.
- Maximizes pre-season revenue and secures repeat visits.

Model Setup

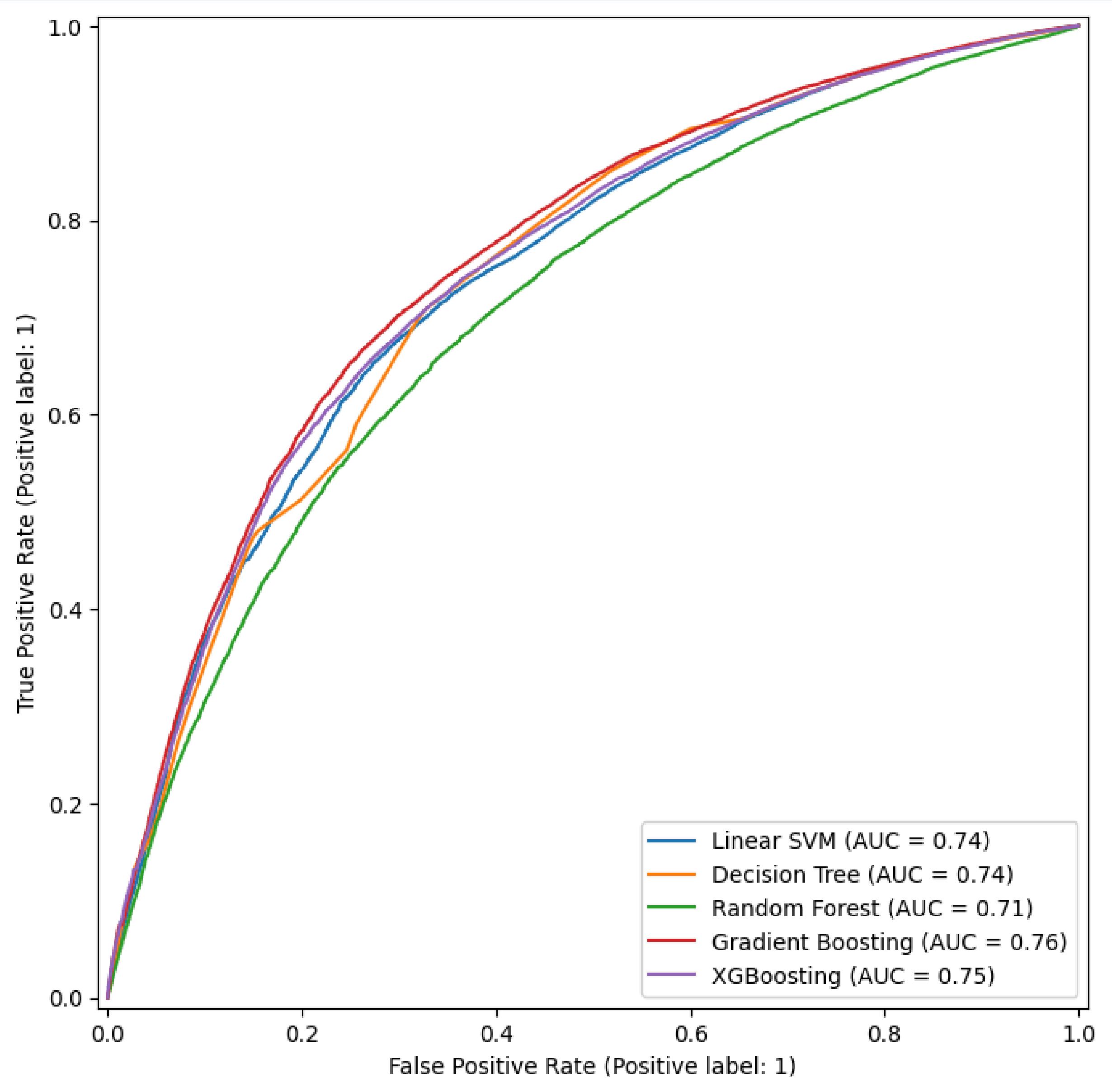
Reference date: before the season start (2023-10-01, 2024-10-01)

Lookback period: previous year

Holdout window: following year

Feature Engineering

- Removed highly correlated features
- Scaled monetary features
- Created aggregate and temporal features



Selected Model: Gradient Boosting, trained on 22-23, tested on 23-24

2025/2026 Predictions

Using the selected Gradient Boosting model, we analyzed **66.149** active clients in the 2024-25 season and predicted **54.023** churners for 2025-26.

Through RFM analysis, **15.048** top-value clients were identified

Target audience of **3.021** clients based on churn probability and customer value.

Campaign's estimated effects

- Expected clients retained (assuming 20% response rate): **604**
- Estimated campaign cost (€): **10.734** (discounts + operational + false positives)
- Estimated benefit (€): **297.155** (revenue from prevented churn)
- Net impact (€): ≈ **286.420**

Mid-Season Marketing Campaign

Objective: Increase spending and strengthen loyalty during the active ski season

Trigger: Once per season, at mid-season (e.g. Feb 15th)

Target:
- High-value clients from the first half of the season (Diamond, Gold)
- Customers with high probability of churn (>0.80)
- Positive and Neutral reviewers

Call to Action: Additional skipass purchases before the season ends

Offer: 5 % discount on all skipasses purchased for the remainder of the season

Rationale:
- Rewards loyal and high-value clients
- Promotes repeat purchases and positive word-of-mouth

Model Setup

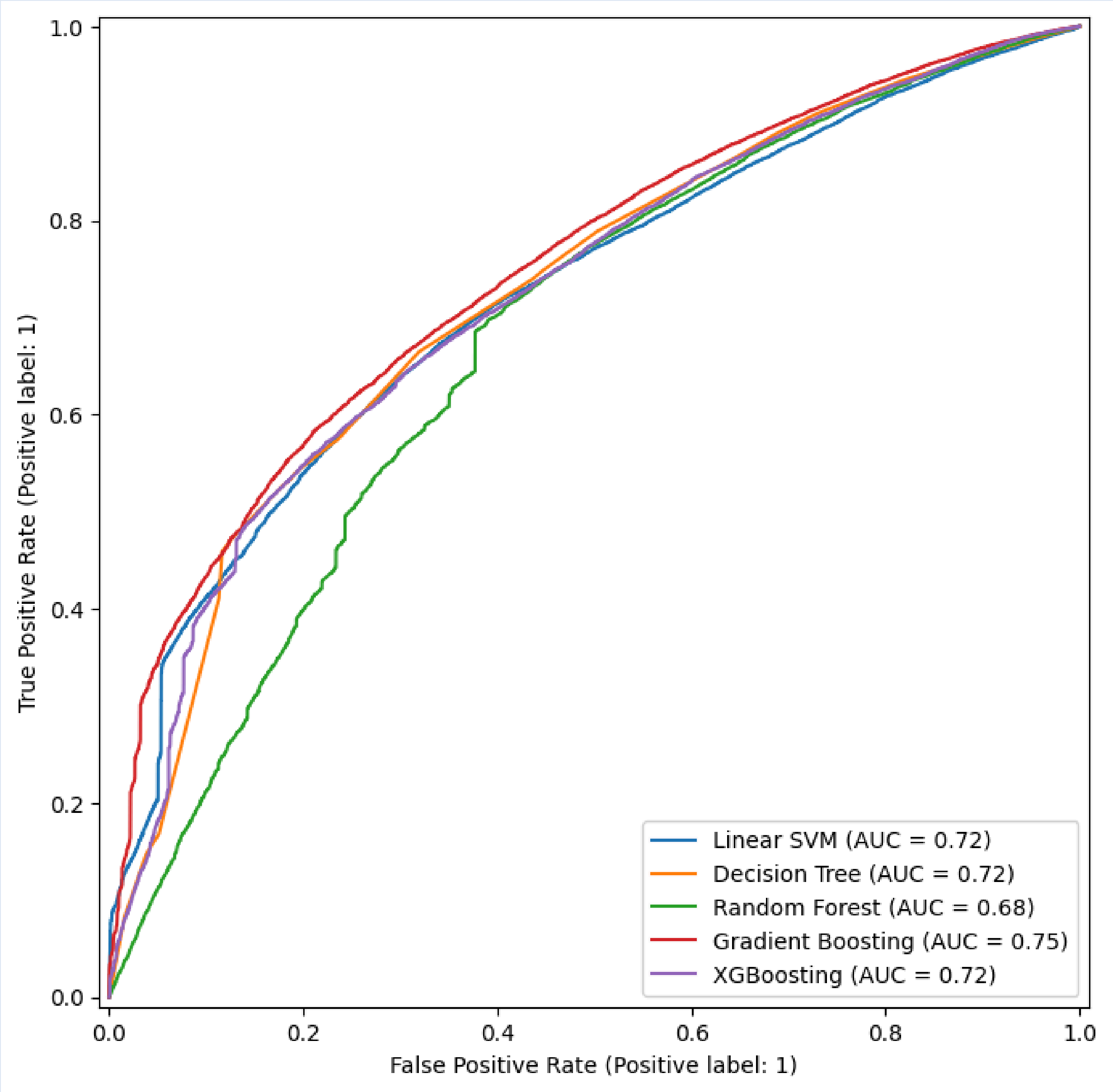
Reference date: mid-season (2023-02-15, 2024-02-15)

Lookback period: 3 months (season's start)

Holdout window: 4 months (season's end)

Feature Engineering
- same as Pre-Season
- weather data

Selected Model: Gradient Boosting
trained on 22-23/23-24, tested on 24-25



2024/2025 Mid-Season Predictions

Using the selected Gradient Boosting model, we analyzed **62.229** active clients at mid-season 2024/25 and predicted **59.526** of them to churn in the second half

Target audience of **1.486** clients based on churn probability, customer value and positive (and neutral) reviews.

Campaign's estimated effects

- Expected clients retained (assuming 15% response rate): **222**
- Estimated campaign cost (€): **4.187** (discounts + operational + false positives)
- Estimated benefit (€): **148.968** (revenue from prevented churn)
- Net impact (€): ≈ **144.781**

Note: due to the natural absence of data from the first half of the 2025/2026 season (the project was carried out in September 2025), we simulated the campaign in the first half of the previous season.

Conclusions

Both the between-season and the mid-season campaign are expected to increase profits of the company.

The estimated net impacts are approximately of **286.420€** and **144.781€**

These results demonstrate the effectiveness of data-driven marketing, combining churn prediction, RFM segmentation and sentiment analysis.

Future Developments

Refine predictions with real-time behavioral and weather data.

Include reviews from real customers who actually mention this company.

Test alternative incentives (bundles, loyalty programs, cross-selling)