CASE STUDY

How Rento Perú identified high-probability renters and achieved 24% conversion in targeted campaigns

Peru's leading peer-to-peer car-sharing marketplace used Rockfish synthetic booking data to train a customer-propensity model that, in its first production campaign, delivered a high-precision call list the commercial team is now actively working.

> *"*
>
> *The generation of synthetic data via the Rockfish platform allowed us to address the constraints imposed by limited historical data availability, facilitating the development of a more accurate predictive model. This approach enhanced the precision of customer targeting and resulted in conversion rates substantially exceeding the organic baseline.*
>
> **— Giulio Marchena Sekli,** CTO, Rento Perú

24%	76%	2.4×
Same-month booking conversion in the Top 5% priority segment — vs. 0.65% organic in the broader base	Customers in the Top 5% list who either booked for May or expressed intent for June or July	Lift over a random call list in the Top 5% segment, validated out-of-time

At a glance

Customer	Rento Perú — Peru's leading peer-to-peer car-rental marketplace
Data environment	Relational production database; ~2 years of usable booking history
Use case	Synthetic booking data to train a customer-propensity gradient-boosted classifier model for bi-weekly rental campaigns

CASE STUDY

The customer

Rento is an AI-enabled, technology-first mobility access platform that unlocks car rental for first-time users by removing barriers, decentralizing supply, and creating new travel possibilities beyond traditional rental markets.

Production data lives in a relational database covering reservations, customers, vehicles, pricing, and trip history. The company's next phase is to flip its growth model from reactive to proactive — reaching out to likely renters with personalized car suggestions via WhatsApp and phone calls, before they ever start searching.

The challenge

To run proactive outreach at all, Rento needed a customer-propensity model — a system that scores each customer by their likelihood to rent in the coming weeks — that could refresh every two weeks and rank customers so the call-center could focus on the highest-value leads.

The problem was data. Three constraints made it hard to train such a model on real data alone:

- **Only about two years of usable booking history.** Not enough seasonal cycles to train a stable propensity model that generalizes to future months.
- **Heavy class imbalance.** Customers who rent in any given month are a small minority of the active base. Without correction, a model trained on real bookings collapses to predicting “no rental” for everyone.
- **Sparse signal where it mattered most.** Recurring renters — the customers Rento's outreach campaigns are designed for — made up only a small share of the historical booking base, leaving the model with thin signal exactly where it needed to learn most.

The solution

Rento partnered with Rockfish to generate synthetic booking data that preserved the statistical patterns of real customer behavior while expanding the volume and rebalancing the segments their real data didn't cover.

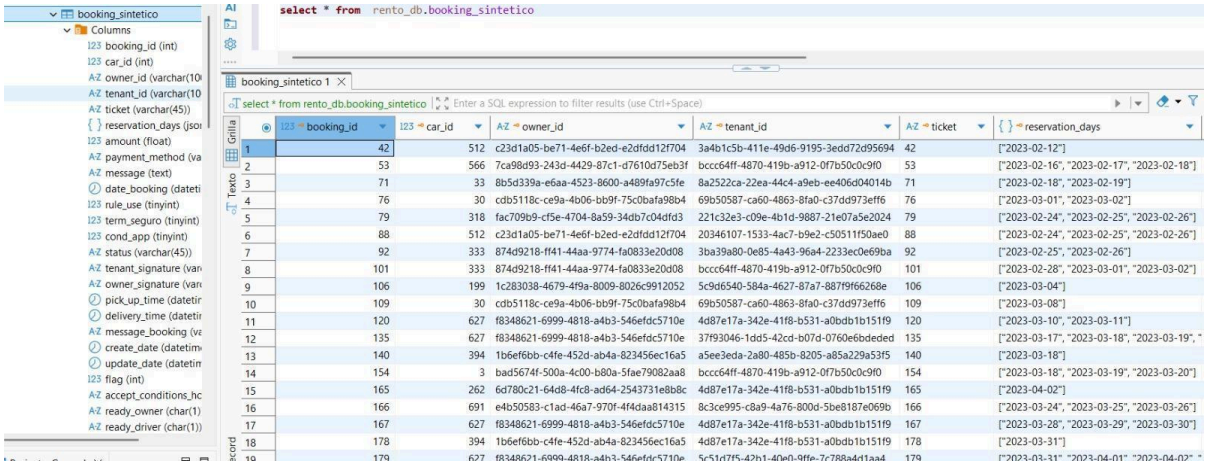
Generation, validation, targeted augmentation

Rockfish designed the generation approach from Rento's database schema alone — no sensitive customer data left Rento's environment in that phase. Once the approach was validated, Rento connected to the Rockfish API and generated over 12,500 synthetic booking records spanning

2023 through 2026 — the same temporal window as Rento's real data, but with denser coverage and rebalanced segments.

Rento validated the output against the real distribution: marginal distributions, pairwise correlations, and descriptive statistics across the columns that mattered for the downstream model. The synthetic output landed in a dedicated table within Rento's environment that the training pipeline reads directly.

Rockfish's Scenario Studio was used to boost coverage of the segments the real data was thinnest on — particularly recurring-customer behavior — so the training set wasn't just larger, it was more evenly covered where it mattered.



The table in Rento's environment — over 12,500 synthetic booking records, with the full schema preserved (reservation_days as JSON arrays, payment_method, status, regional travel fields, and more) so the training pipeline can consume it the same way it would consume real bookings.

The synthetic bookings preserve Rento's real customer and vehicle linkages, while the booking-level behavior — reservation dates, amounts, payment method, status — is synthesized.

This design — preserving the real customer-vehicle graph while synthesizing the booking events on top of it — lets the trained model score real customers in production, rather than only learning patterns about hypothetical ones.

The privacy benefit comes from not duplicating real booking events into downstream training environments and notebooks; the modeling benefit comes from preserving the customer-vehicle graph the propensity signal lives on.

Training pipeline

The downstream model is a gradient-boosted classifier trained on the synthetic-booking table.Engineered features include rental frequency over 1, 3, and 6 months (rolling), lifetime

rental count, recency, and customer-level monetary signals. A class weight of 20× was applied to the positive class to compensate for the natural rental imbalance.

Training data ran from November 2023 through June 2025, with July–September 2025 held out for validation and October–December 2025 held out as an out-of-time test set — the cleanest possible check that the model would generalize to future months.

The results

Model performance (out-of-time test)

Across train, validation, and out-of-time test, the model showed remarkable stability — a hallmark of synthetic-trained models that haven't memorized their inputs:

- AUC: 0.677 (train) → 0.656 (validation) → 0.663 (out-of-time test)
- Gap between validation and out-of-time test: less than 1%. The model is not overfit.

Segment performance against the out-of-time test set shows the model is doing what a propensity model needs to do — separate the high-value tail of the customer base from everyone else:

Segment	Precision	Recall	Lift vs. random
Top 1%	60.14%	2.80%	2.88×
Top 5%	50.21%	12.0%	2.41×
Top 10%	46.15%	22.1%	2.21×
Top 15%	42.47%	30.5%	2.04×

Interpretation: at the Top 5% threshold — a list large enough to be operationally meaningful for a call-center campaign — roughly half of contacted customers convert in the test data, and the model is 2.4× more efficient than a random call list pulled from the same customer base. Even at the Top 15% threshold, the lift holds at 2.0×.

First production campaign (April 2026 → May 2026 bookings)

Rento scored its active customer base of 7,484 travelers in April 2026 and used the model output to build a prioritized outreach list for the commercial team to contact via WhatsApp and phone calls. The top 1.6% of customers by score (120 individuals across the top five percentile

bands) were contacted directly. The remaining 7,364 customers were left to organic channels as a comparison group.

Score group	Contacted	Booked	Future intent	Still thinking	Declined	Booking rate
Top 1%	2	1	1	0	0	**50%**
Top 5%	17	4	9	0	4	**24%**
Top 10%	21	3	13	0	5	**14%**
Top 15%	34	4	17	6	7	**12%**
Top 20%	46	5	19	10	12	**11%**
Rest (~98% of base)	7,364	48	83	21	45	**0.65%**

Three things to notice in this table:

- **The Top 5% segment is the operational anchor.** From the 17 customers the model flagged as priority contacts in the Top 5% band, 4 made reservations for May — the model's prediction window — and another 6 expressed interest in booking for June. That's a 24% same-month conversion rate with a strong pipeline behind it. Group 1 had higher precision but a small sample (n=2), so it reads as a directional signal rather than a headline.
- **Engagement, not just same-month bookings.** In the Top 5% band, 13 of the 17 contacted customers either booked for May or expressed intent for June or July. That's a 76% near-term engagement rate on a tightly-scored list — pipeline the call-center can keep working.
- **The base rate is the comparison point.** Organic conversion in the un-prioritized group ran at 0.65%. The model's Top 5% segment converted at 24%. That's the value of putting Rockfish synthetic data underneath a propensity model on a thin real-data base.

What's next

Rento and Rockfish are now operationalizing the pipeline so it runs as steady-state infrastructure rather than a one-time experiment:

- **Continuous training.** As Rento's real booking data grows month over month, the Rockfish generator can be incrementally refreshed so the synthetic distribution keeps tracking real-world shifts.

- **Bi-weekly scoring cadence.** Aligning model refresh and inference with Rento's call-center planning cycle.
- **Threshold tuning by campaign type.** Top 1% for high-touch concierge outreach; the F1-optimal threshold (0.758, recall 54%, precision 32%) for broader growth campaigns.
- **Expanding the use case.** The same synthetic-data foundation can support Rento's WhatsApp recommendation agent, vehicle-class affinity modeling, and dynamic-pricing experiments — without each new model waiting on years of additional real-world data.

About Rento Perú

Rento Perú is Peru's peer-to-peer car-sharing marketplace, connecting vehicle owners with renters in a trust-managed, fully-insured platform. Learn more at rento.pe.

About Rockfish

Rockfish Data, built on foundational research from Carnegie Mellon University, generates high-coverage, labeled datasets purpose-built for evaluating AI models and data analytics agents in production. Unlike generic benchmarks, Rockfish systematically covers edge cases, rare scenarios and domain-specific failure modes — giving teams the evaluation infrastructure to measure accuracy and reliability before deployment. Learn more at rockfish.ai.