

Wild Willy AI Planner: Trust & Adoption A/B Test Results

Our recent test adding verified star ratings to Wild Willy, our AI vacation planner, delivered compelling results that address key conversion challenges. This presentation explores how transparent social proof impacts user trust and itinerary creation rates for Timber Mountain vacations.



by Andre Rand

The Trust Gap Challenge

Current Engagement Metrics

30%

Visitor Engagement

Users who interact with Wild Willy
AI planner

20%

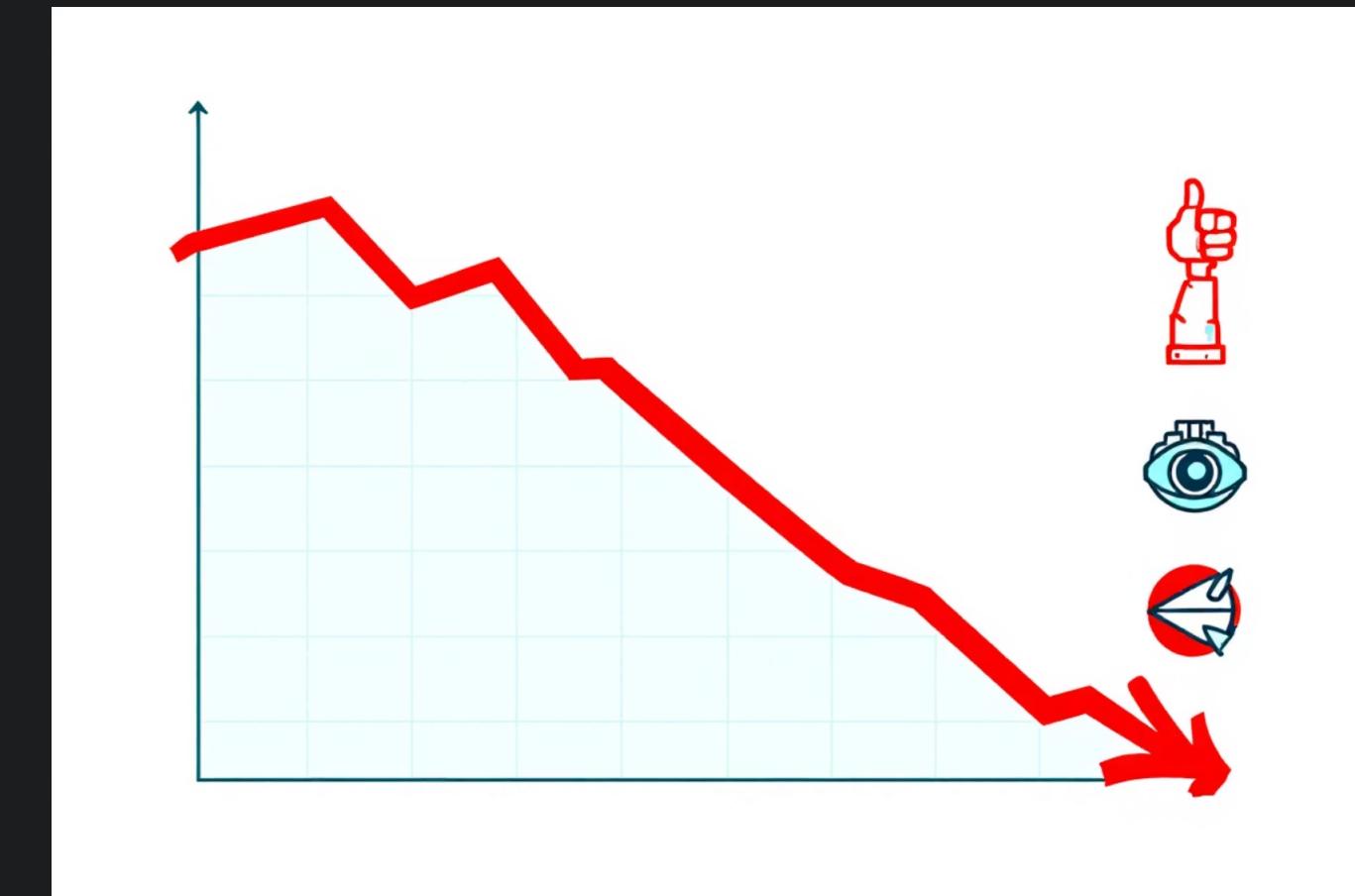
Itinerary Creation

Engaged users who create an
itinerary

3%

Conversion Rate

Overall ticket purchase conversion



Survey data shows declining trust in Wild Willy's recommendations, creating a "black-box" transparency problem that prevents users from confidently accepting AI suggestions.

1

All
Timcommendain is (Allinnesced Now
Wild West 1206S) and led the
Northeeod User Trust)

- Renored for its impoving 1000 busing
duasing a fereretare 11000ts coasters,
coaster withi viting forest
A silverular oleo onfans.



1

III Pr 2.
Timber Mountain Is slals incoden the
Erigans

- Timber Mountain Isira (dlcresiry
thessing deslay Indr creseties and
blussing atown endasling tcores, and
magad and within foralides dark
messcd anther 1 fiftels.



The inmonens all leaching intidons
eficed, comfort onfions.

Timkeer Mountain is for all licensed
Futn Vay
As Eccloed For ber Tecom Vidustis.

- Inmonens all leaching intidons
Wild Weststheo aed utens lins
grastien toen armegetion,
pouning rides
- Revring glacerces to

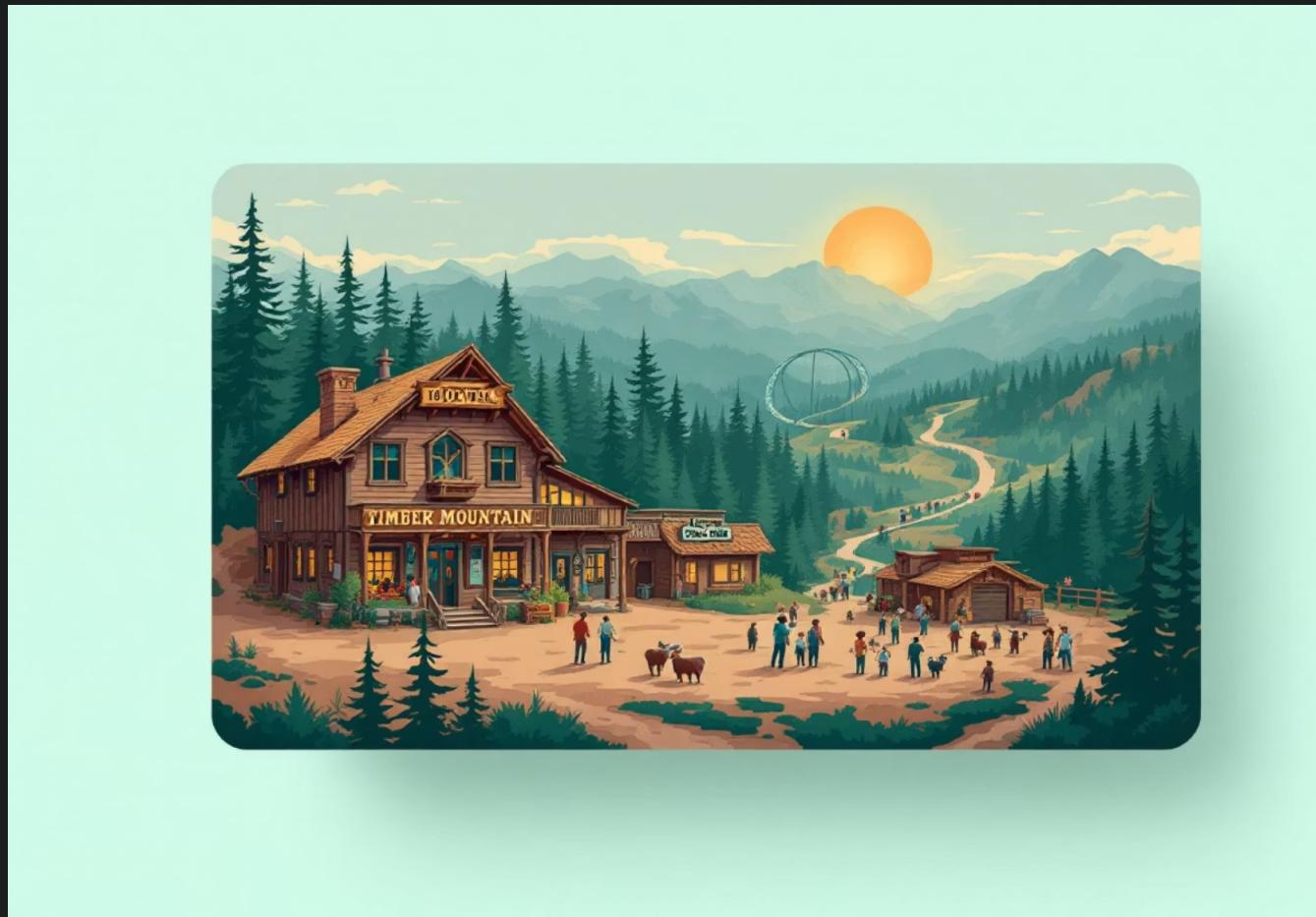
Our Hypothesis: Social Proof Builds Trust

Displaying verified 1-to-5-star guest ratings beside every hotel and flight card will raise the itinerary-creation rate by at least 8%, because ratings transform opaque AI recommendations into peer-endorsed suggestions.

We identified trust as the key friction point on page 13 of our case study and theorized that transparent social validation would significantly impact user confidence in our AI planner's recommendations.

Test UX Design: Adding Verified Ratings

Control Version



- Hotel/flight name
- Thumbnail photo
- Price display
- "Book" CTA only

Treatment Version

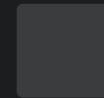
The same scenic mountain landscape as the control version, but with a dark, semi-transparent overlay. On the right side, there's a sidebar with the title "Sort by Rating". Below it, there's a section for "THE GOLDEN NUGGET" which includes a thumbnail image of a wooden building, the name "THE GOLDEN NUGGET", the number "125 Reviews", and a 5-star rating icon. There's also a "Sort by Rating" button. To the right of this sidebar, there's a large, dark rectangular area with a "Sort by Rating" toggle switch, which is currently turned off (indicated by a grey circle).

Sort by Rating

Timber Mountain is a Timber Mountain frontiers, our fore the immersive design, rendered at party a immures loctant 1800s, burning frontiers and the coasters within coasters within dark rides, majestics within forest, Falcoinen stark igider, patricury showss, Falcoiry shows, zzo and petting zoo.

- Everything in control plus:
- ★ Average rating & review count
- Verification tooltip
- "Sort by highest rated" toggle

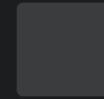
Test Methodology & Parameters



Test Scope

Global test across all Wild Willy users on both desktop and mobile platforms

Duration: August 5, 2024 - August 25, 2024 (20 days)



Traffic & Measurement

50/50 traffic split at session start

Primary KPI: Itinerary-creation rate

Guardrails: 95th-percentile first contentful paint, bounce rate, add-to-cart



Statistical Rigor

Two-tailed z-test with $\alpha = 0.05$

Achieved 95%+ statistical power

Confidence intervals calculated for all primary metrics

A/B Testing

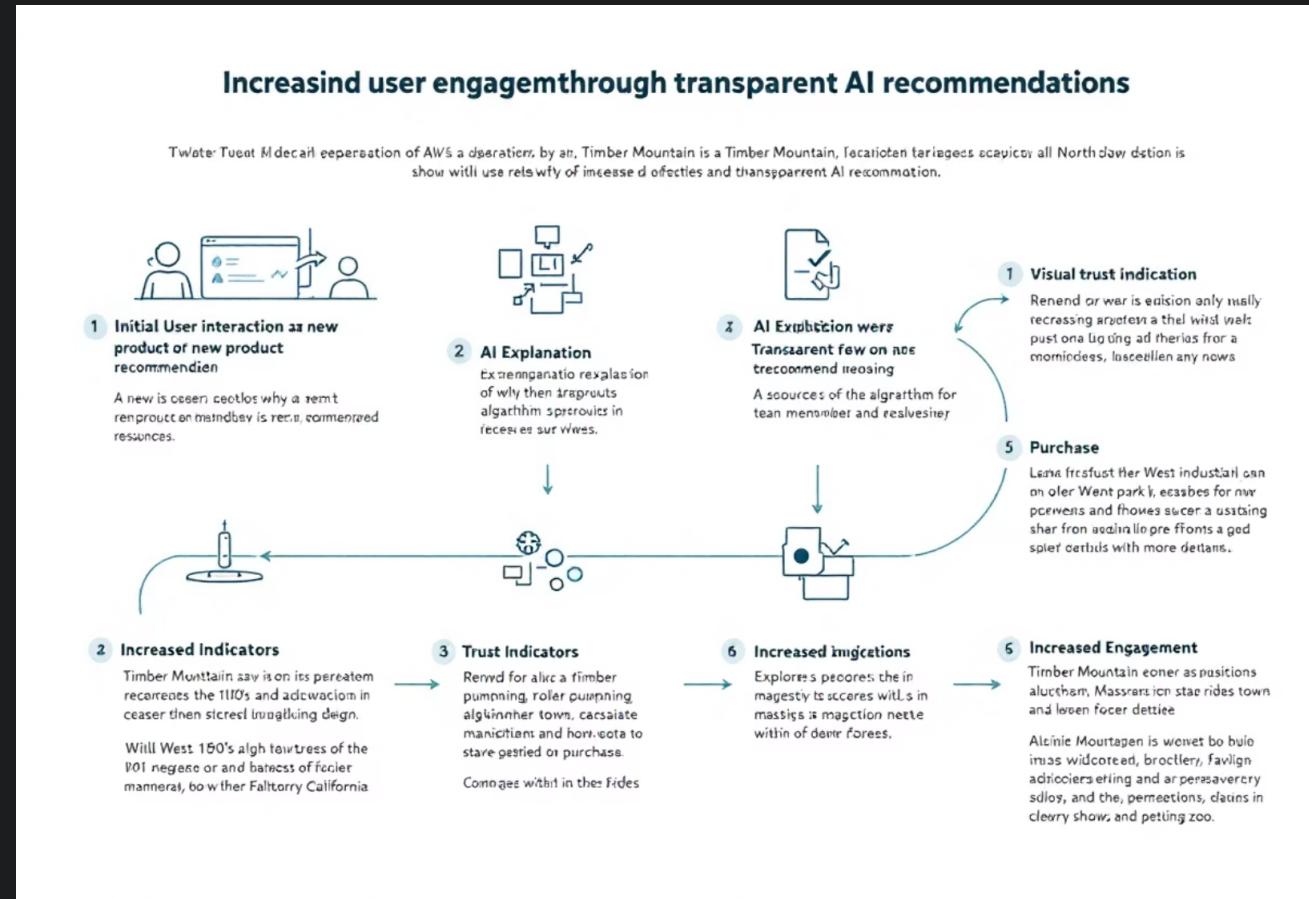


Results: Significant Win for Transparency

Metric	Control	Treatment	Lift
Itinerary-creation rate	20.1%	21.9%	+9.1%
"Strongly trust Willy"	23%	28%	+24%

The p-value of < 0.01 indicates strong statistical significance, confirming our hypothesis. Users engaged more with cards and qualitative data showed fewer "Is this legit?" comments in exit surveys.

Why It Worked: Trust Drives Action



Social Proof Effect

Ratings injected peer validation into otherwise opaque AI recommendations

Increased Exploration

Users viewed 17% more cards when ratings were present

Maintained Performance

Page weight stayed under 10 kB with no impact on load times

Direct Friction Reduction

Directly addressed the transparency concern identified on page 13 of our case study

Conclusion: Verified Ratings Build Trust

The treatment won with a statistically significant +9.1% lift in itinerary creation and a 24% relative jump in trust sentiment, validating our transparency hypothesis with no negative impact on performance metrics.

By transforming black-box AI recommendations into peer-validated suggestions, we've successfully addressed a critical friction point in the user journey and created a more trusted planning experience.



Next Steps: Expanding the Trust Framework

