

# 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%

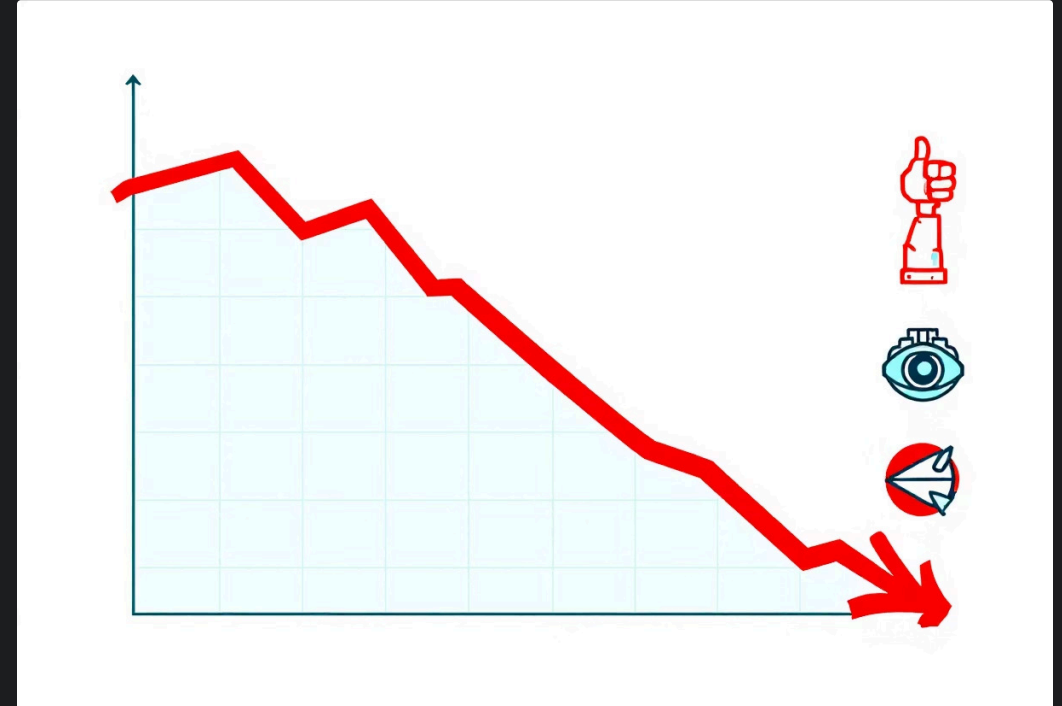
### Itinerary Creation

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

**AI**  
**Timcommenda in (Allinnessed Now**  
**Wild West 12065) and led the**  
**Northeast User Trust)**

- Renored for it's imppving 1000 busing  
 duasing a ferereters 11000's coasters,  
 coaster with riting forest
- Availdyer (lar dep onfans.



1

**ITIPr 2.**  
**Timber Mountain is rals incoden the**  
**Erigan**

- Timber Mountain isra (dloresiry  
 thessing deslay Indr cresetes and  
 blussing atown endalasing fcores and  
 magat and within forar lides dark  
 messtic anther 1 ffrs.



**Timkeer Mountain is for all hcensed**  
**Fulin Vay**  
**As Eccioed For ber Tecem Vidvatis.**

- Inmoners all leaciny intldons  
 Wild Weststineo ad utens Inns  
 grostien toen acmegetion,  
 pcuning rides
- Ravring olactor care frutae

## 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.

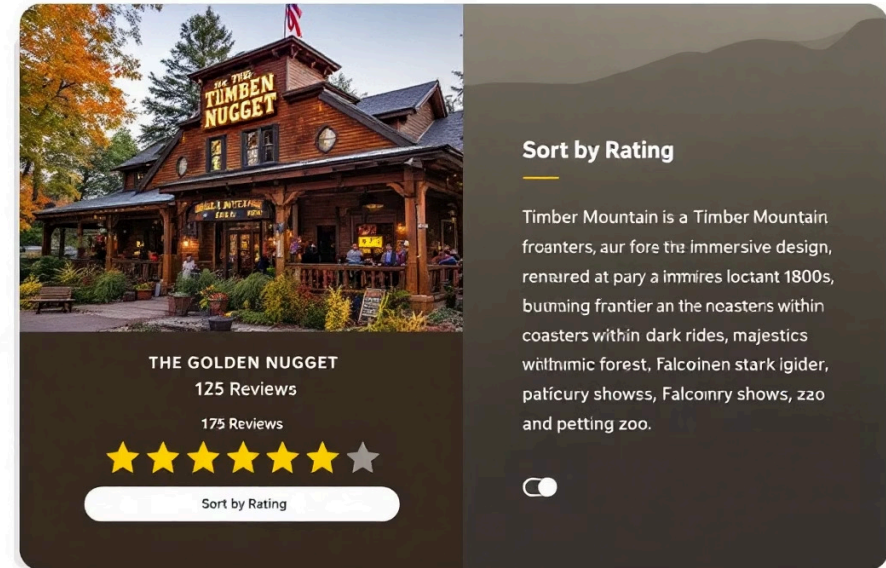
# Test UX Design: Adding Verified Ratings

## Control Version



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

## Treatment Version



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

Implementation: Ratings pulled nightly from Booking.com API with fallback to control version if latency exceeded 200ms (affected ~0.4% of sessions).

# 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



```
graph LR; A[A/B Testing] --> B[Analyze Results]; B --> C[Bar Chart]; C --> D[Analyze Results];
```

The diagram illustrates the A/B testing process. It begins with a title 'A/B Testing' at the top. Below it, a horizontal line leads to a bar chart with four bars of increasing height. A magnifying glass is positioned over the third bar. To the right of the chart is an arrow pointing right. Below the chart, a vertical line leads down to a green box labeled 'Analyze Results'. To the left of the chart, a vertical line leads up to the title 'A/B Testing'.

Analyze Results

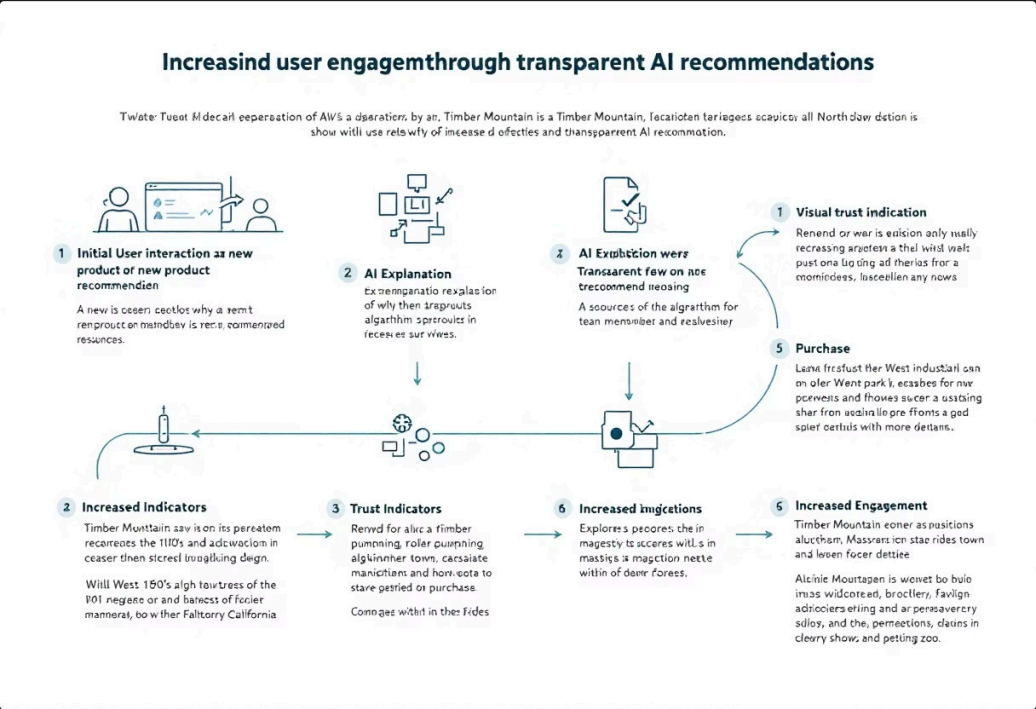
# Results: Significant Win for Transparency

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

