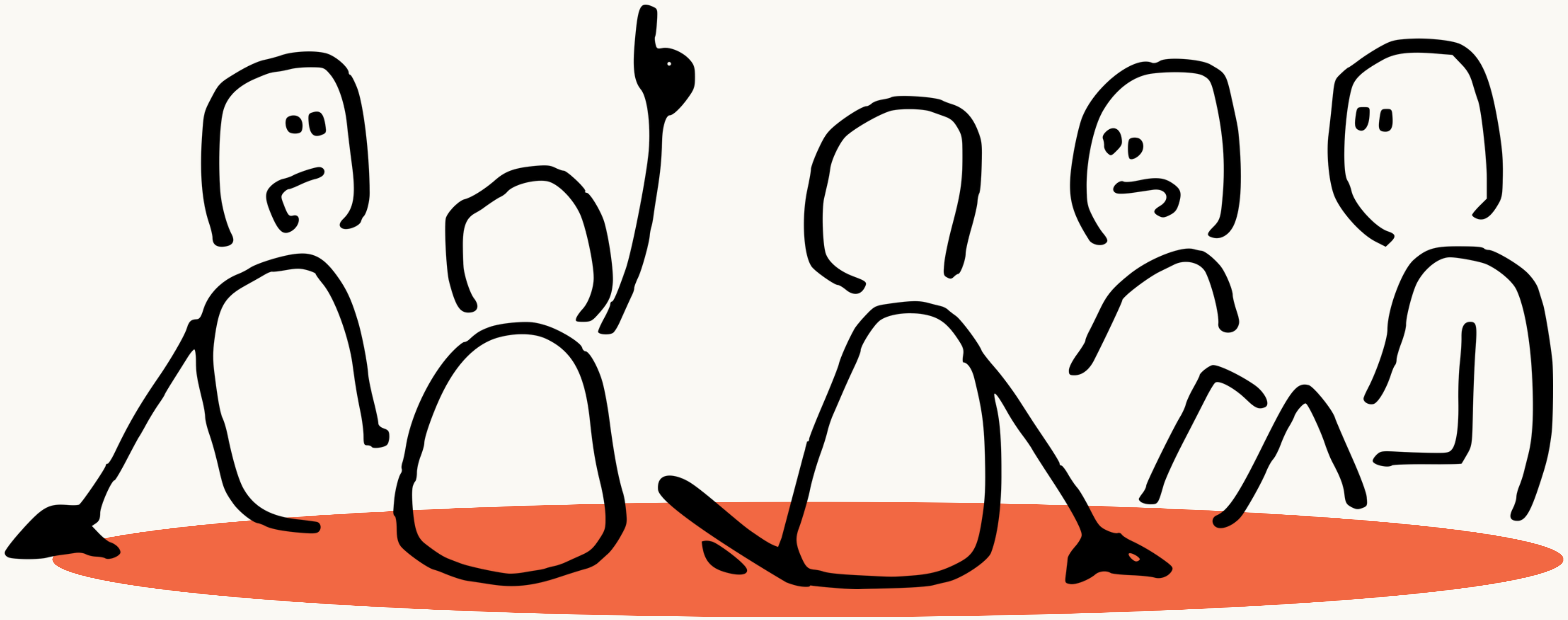
CASE STUDY : WEBSITE USER BEHAVIOR RESEARCH



Presented by
Nuriman Altybayeva

PwC Advisory Test Task Solution

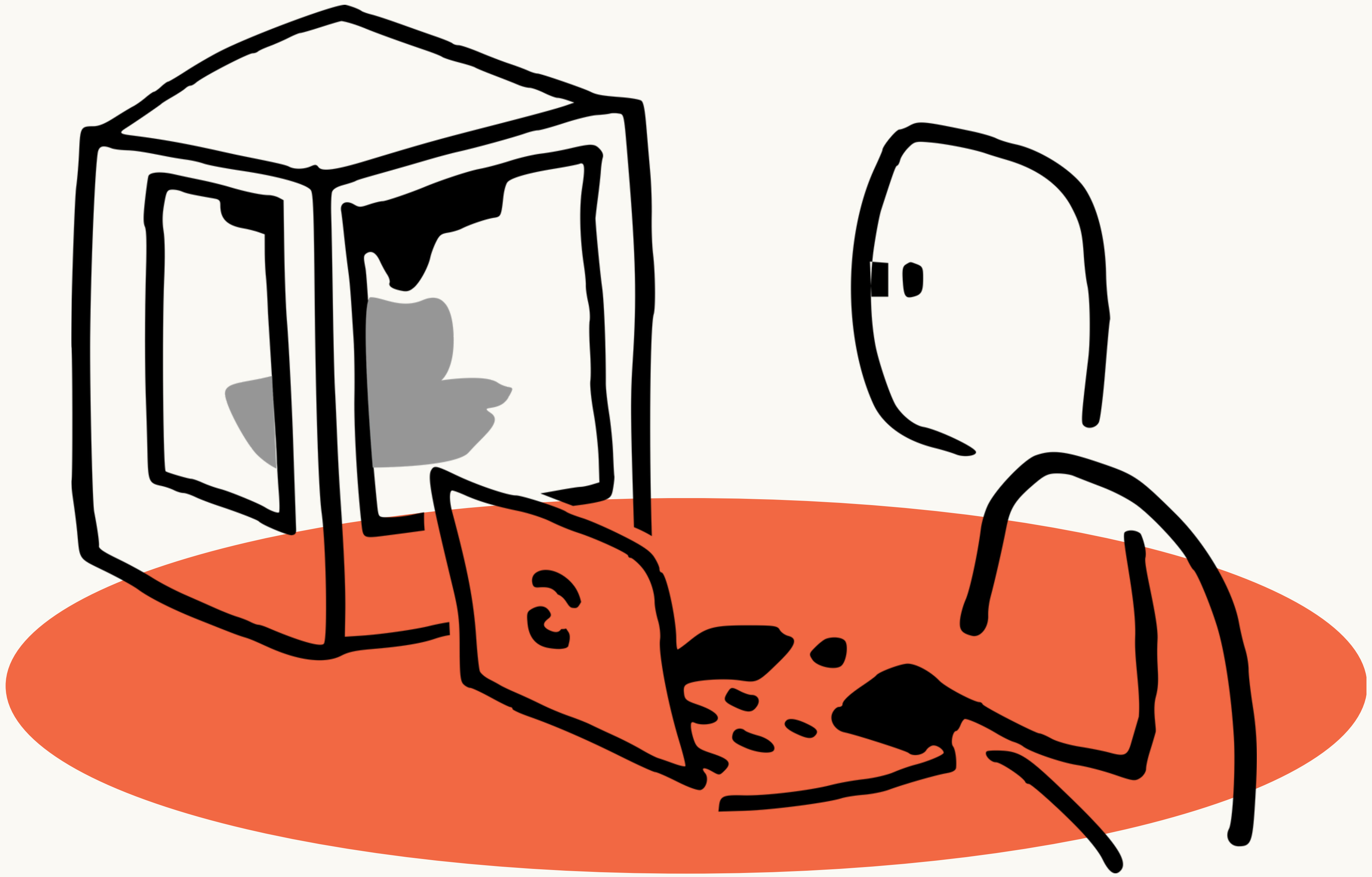
23th November 2025

Main Goals

- to understand what types of users come to the site, how they interact with the content, and how their conversions differ.

We conducted a clusterization of user behavior over 3 weeks and identified four stable groups.
What we were looking for:

- patterns of behavior
- barriers to purchase
- points where personalization can increase conversion



User Clusters Overview

Cluster 1: Passive Users

65,463 users (~84%)

- AddToCart ~3%
- Purchase ~0%
- Minimum activity: 1-3 events on average
- Rarely reach the product page
- Often come "accidentally" or without a clear intention.

Cluster 2: Hot Buyers

2,004 users (~2.5%)

- AddToCart: 81%
- Purchase: 99.95% (!!)
- Actively clicking on products
- Often go to the page of a specific model.
- Show the most predictable and "clean" buying behavior.

Cluster 3: Researchers

7,696 users (~10%)

- AddToCart ~32%
- Purchase ~0%
- Actively go through the catalog
- Study different products characteristics are compared, categories are often changed
- The paradox: they're involved, but they're not buying.

Cluster 4: Indecisive Buyers

289 users (~0.4%)

- Very high activity inside the site
- AddToCart: 76%
- Purchase: 46%
- Half of the users "fall off" at the final stage. This is the most valuable group for personalization. They just don't have the confidence to click “Pay”.

Conversions and Statistics

Conversion to adding to cart:

- 3% (cluster 1 — low interest)
- up to 81% (cluster 2 — confident buyers)

Conversion to purchase:

- 0% (clusters 1 and 3)
- up to 99.9% (cluster 2)

Statistics (z-test for proportions):

- p-value for all comparisons < 0.001
- differences in behavior between clusters are not accidental
- the groups are statistically independent and meaningful

It means: Clusters can be used in segmentation, retargeting, and personal recommendations — predictability is high.

Personalization by Segment (Recommendations)

Cluster 1: Passive Users

Goal:
To engage, not sell right away
Tools:
- Personal banners by the most popular categories
- Mini-collections on the main page: *"Here's what they're looking for now"*
- Trend recommendations
- Simple onboarding: hints, structured catalog

Cluster 2: Hot Buyers

Goal:
To simplify the purchase to 1-2 clicks
Tools:
- Fast payment methods
- Auto-completion of data
- Shortened checkout
- The “Popular” add-ons (cross-sell) block
- Saving the last viewed product

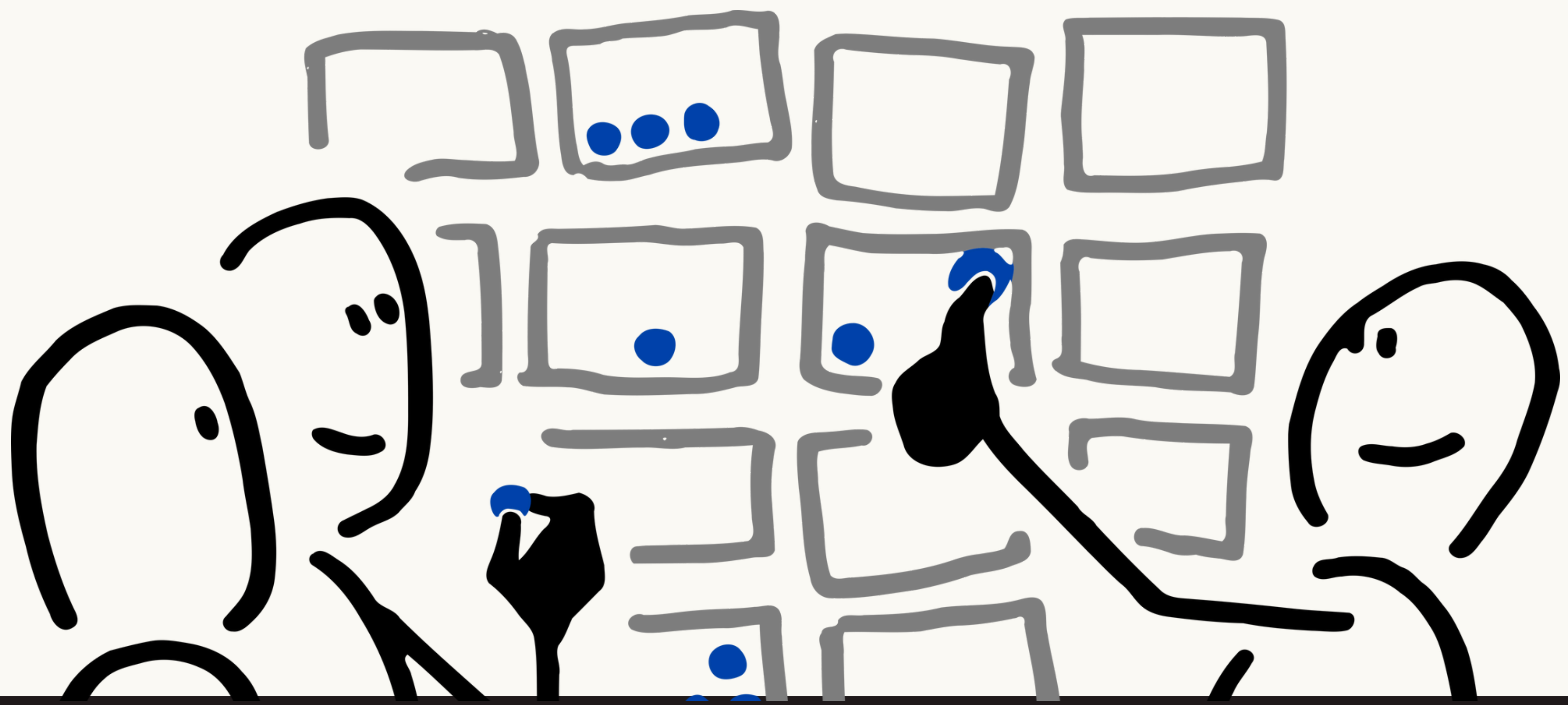
Cluster 3: Researchers

Goal:
To increase trust and reduce the moment of doubt
Tools:
- Reviews, ratings, video reviews
- Model comparison
- Section "This product is often viewed"
- Cards with extended descriptions
- Personalized collections based on interests

Cluster 4: Indecisive Buyers

Goal:
To return and put the squeeze on the purchase
Tools:
- Email/push about an abandoned shopping cart
- Temporary discounts
- Highlighting benefits: free shipping, warranty
- Social proof: "This product has been purchased 34 times in the last hour"
- Reminders when logging in again

User Clustering & Behavioral Insights



Approach

User activity logs represent event sequences → vectorized using:

- event frequency features (Bag-of-Events),
- event fractions,
- engagement metrics (evt_unique_count),
- key funnel actions (catalog, cart, checkout).

Tested k from 2 to 10 using **Elbow** and **Silhouette** methods. The selected k provides the clearest separation and interpretable behavioral segments

Key Findings

User base splits into clear behavioral groups:

- **Researchers** — large number of views, high event diversity, low conversion.
- **Indecisive Buyers** — frequent cart interactions but often drop before checkout.
- **Hot Buyers** — strong presence of checkout-related actions, high conversion rate.
- **Passive Users** — minimal interaction, low engagement.

Conversion differences between clusters are statistically significant (z-test, $p < 0.001$).

Purchase Prediction & Key Drivers

Model Performance

Results (*Model - ROC-AUC*):

- **Logistic Regression** — 0.977
- **Gradient Boosting** — 0.976
- **Random Forest** — 0.975
- **HistGradientBoosting** — 0.973

Consistently high performance → feature set clearly separates pre-purchase windows from non-purchase ones.

Top Predictive Features

- **Checkout_PaymentDelivery** (count & fraction) — strongest indicator of purchase intent.
- **CheckoutInstallment** — high predictive power for imminent checkout.
- **cart / add_to_cart / buyButton actions** — represent the final conversion steps.
- **evt_unique_count** — separates "explorers" from "ready-to-buy" users.

Conclusion

- User behavior can be effectively segmented and predicted.
- Checkout-related actions dominate the conversion signals.
- The predictive model can be used for early detection of high-intent users (20 steps before purchase) and targeted personalization.