Decoding E-Commerce data

Unlocking Customer Insights and Propelling Sales

Let’s talk data

The business

E-commerce platform offering a range of 5 products, users can reach the products via different paths through **5** different categories. As a business we track the paths our customers take via our website to.

Click stream_df

- IP Address: Location from where the site was accessed
- Timestamp: Precise time of each page load
- User Session ID: Unique identifier for each visitor session
- URL: The specific web page accessed
- Purchase Flag: Indicates if a transaction occurred
- Error Flag: Highlights any page loading issues
- Geographic Data: City, State, and Country of access

(421.2k records for 16k sessions)

Data processing

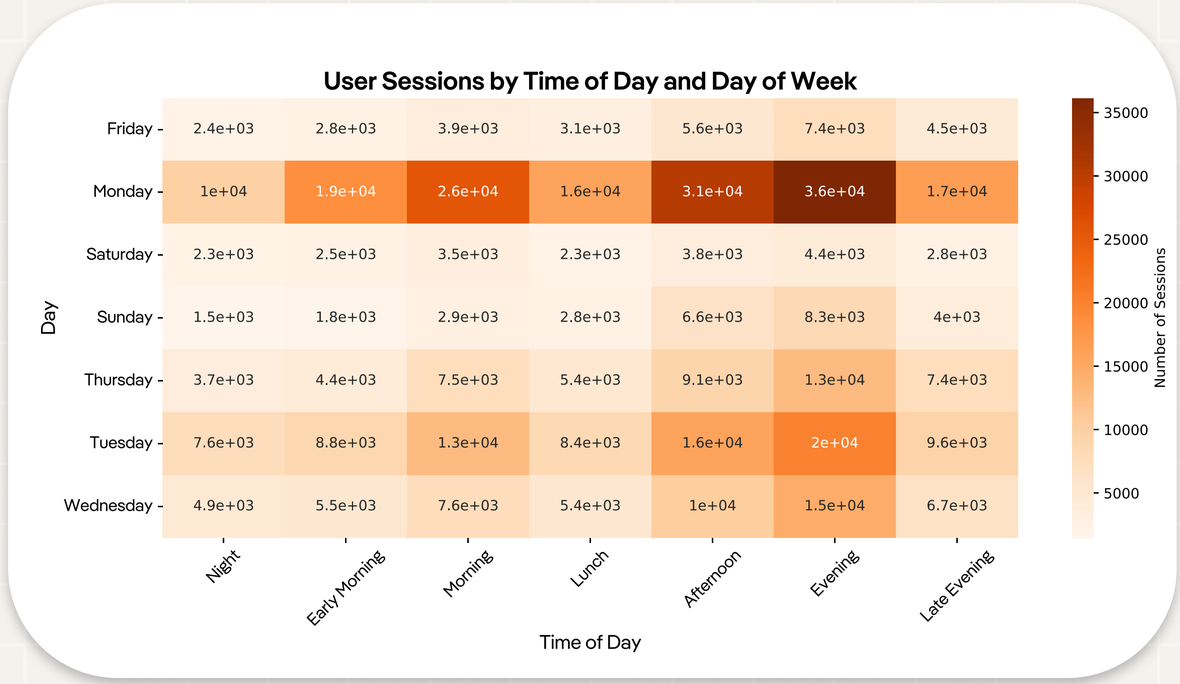
Processing

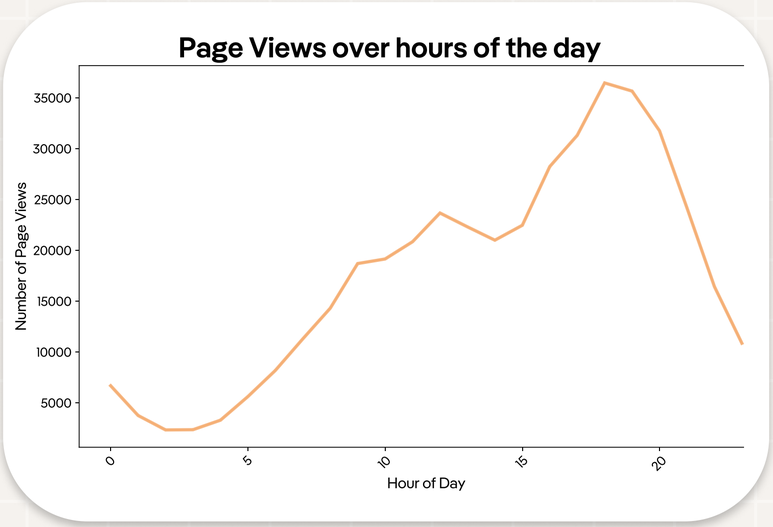
Add the url details to the product df, in order to merge our data with the clickstream data via url.

Removed any NaN from the user data (38k)

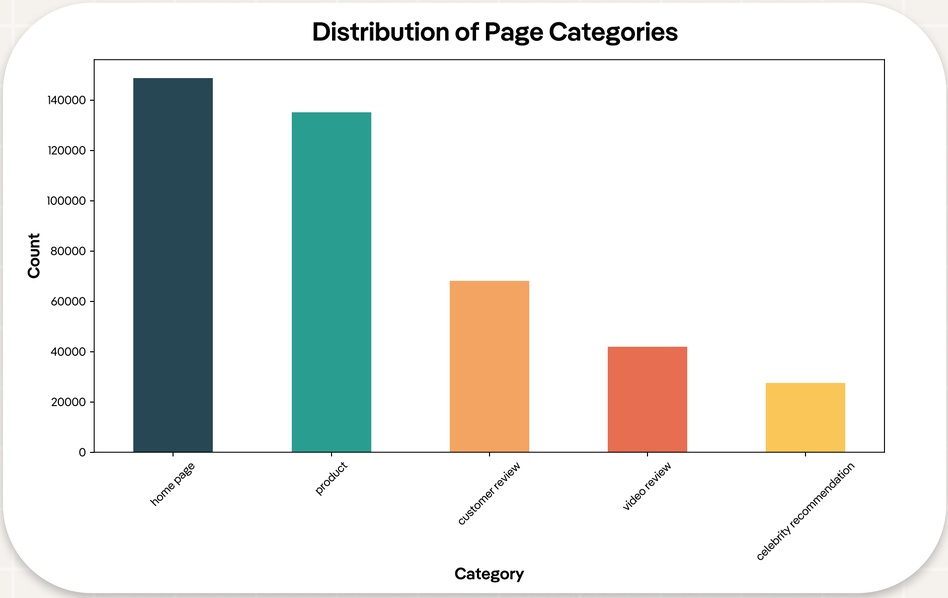
Let’s dive right in

Most users come on the website Monday evening, then the rest of the week the sessions start to drop





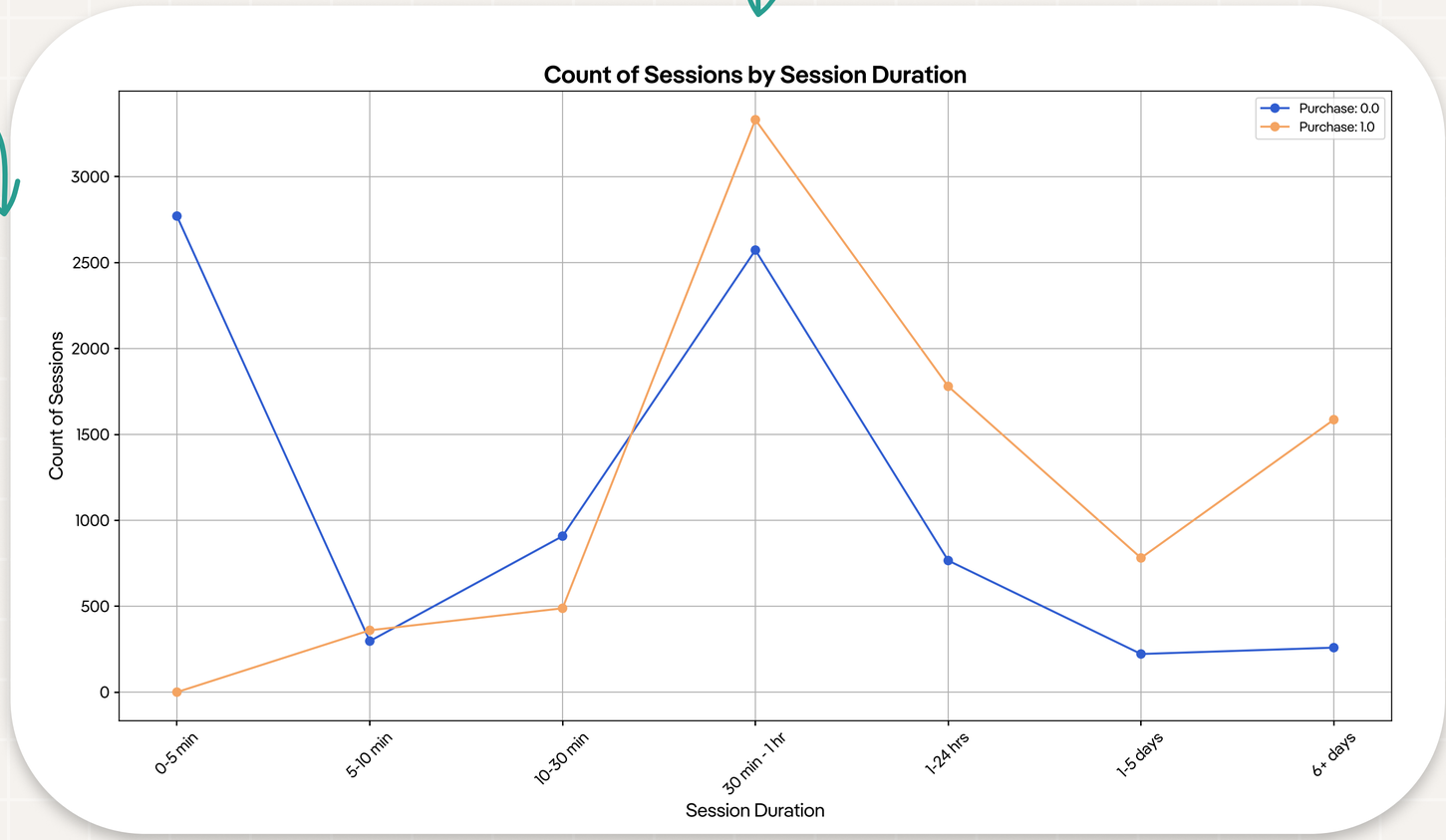
Purchases happen when a customer lands on the product page, but how do they get there? Which are the most viewed pages?



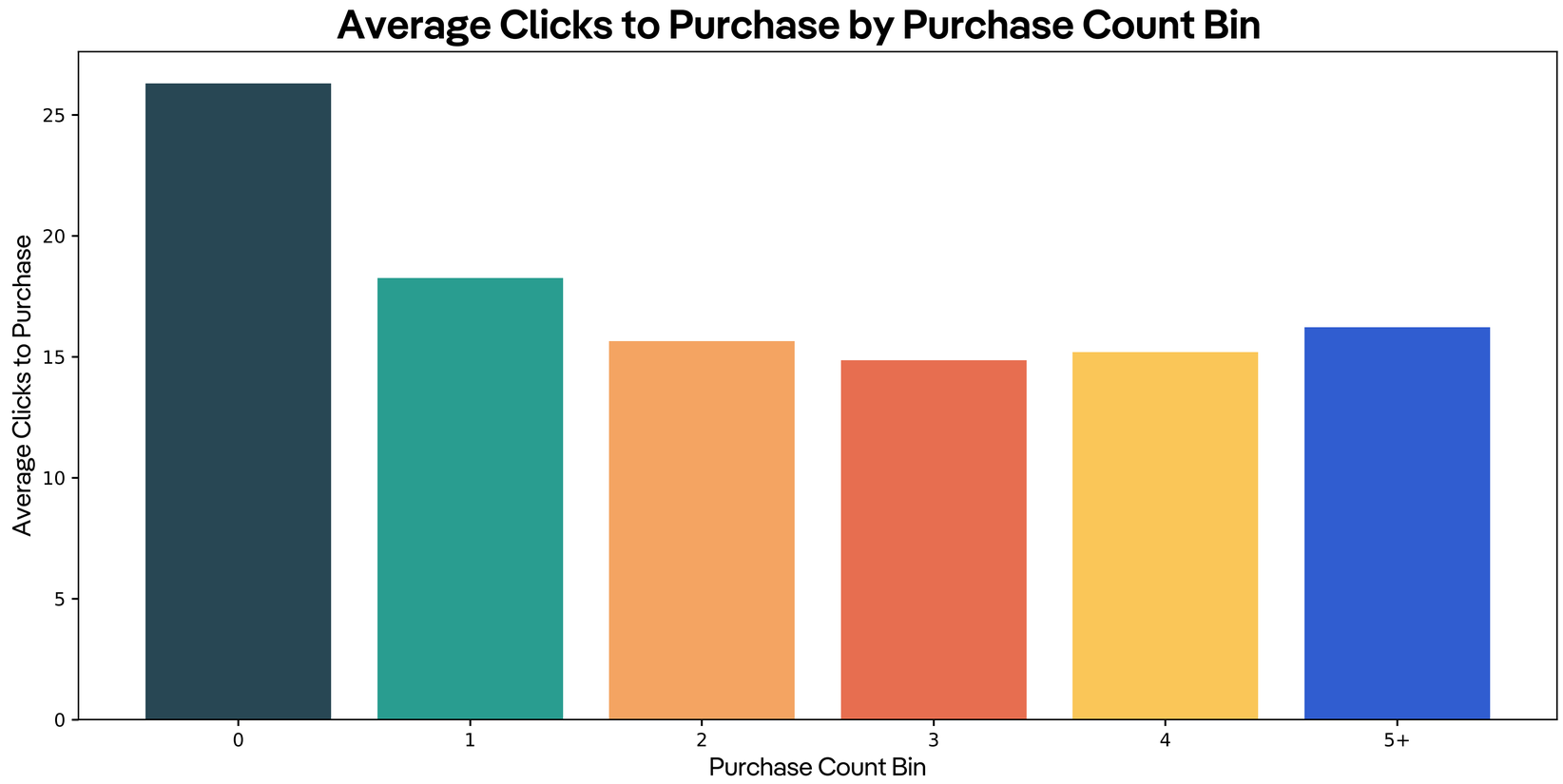
How much time are users spending on the website prior to purchase?

Users tend to make purchases at the 30 minute to 1hr mark

Users who do not make any purchase spend the least time on the site exiting almost immediately



How many clicks are users taking prior to purchase?



- Average Clicks Prior to Exit: 26.13
- Average Clicks Prior to Purchase: 18.84

Attribution modelling: How does the last step influence conversion? 🤔

When looking at the purchases that happened based on the last click prior to the purchase we see that customer review has a higher impact.

Product	Customer Review	Video Review	Celebrity Review
4001	3.31%	1.98%	1.12%
4002	3.00%	1.99%	1.09%
4003	3.18%	1.91%	1.11%
4004	2.96%	1.88%	1.13%
4005	2.89%	1.75%	1.00%

These cells have been highlighted in yellow as we do not have video or celebrity endorsements for the products the customers ended up purchasing. Rather the user browsed multiple product reviews prior to choosing which product to buy.

In terms of last click we calculated the conversion rate as:
Purchases of product x based on specific review / Total of purchases

We chose to focus on time decay as a measure to obtain a more complete picture

Product	Customer Review	Video Review	Celebrity Review
4001	3.20%	1.89%	1.06%
4002	2.87%	1.91%	1.02%
4003	3.04%	1.84%	1.08%
4004	2.82%	1.79%	1.08%
4005	2.75%	1.65%	0.95%

In terms of time decay we calculated the conversion rate as:
Purchases influenced by review, weighted by influence over time. / Total of purchases

Here, our focus shifts to time decay analysis to evaluate the cumulative effect of different reviews on conversion. Time decay gives a deeper understanding of the gradual build-up of influence each review type has over time, leading up to a purchase.

Game theory and regression give us our final touchpoint

Shapley helps us understand how different factors (called "touchpoints") influence a customer's decision to buy your product. They tell us how important each one was in convincing the customer to make that purchase.

Customer review
1.13%

Celebrity recommendation
1.10%

Video Review
1.01%

The bump in practice...

We then perform a linear regression analysis to determine if a purchase occurred, summing the occurrences of each review type to assess how their prevalence impacts the likelihood of a purchase

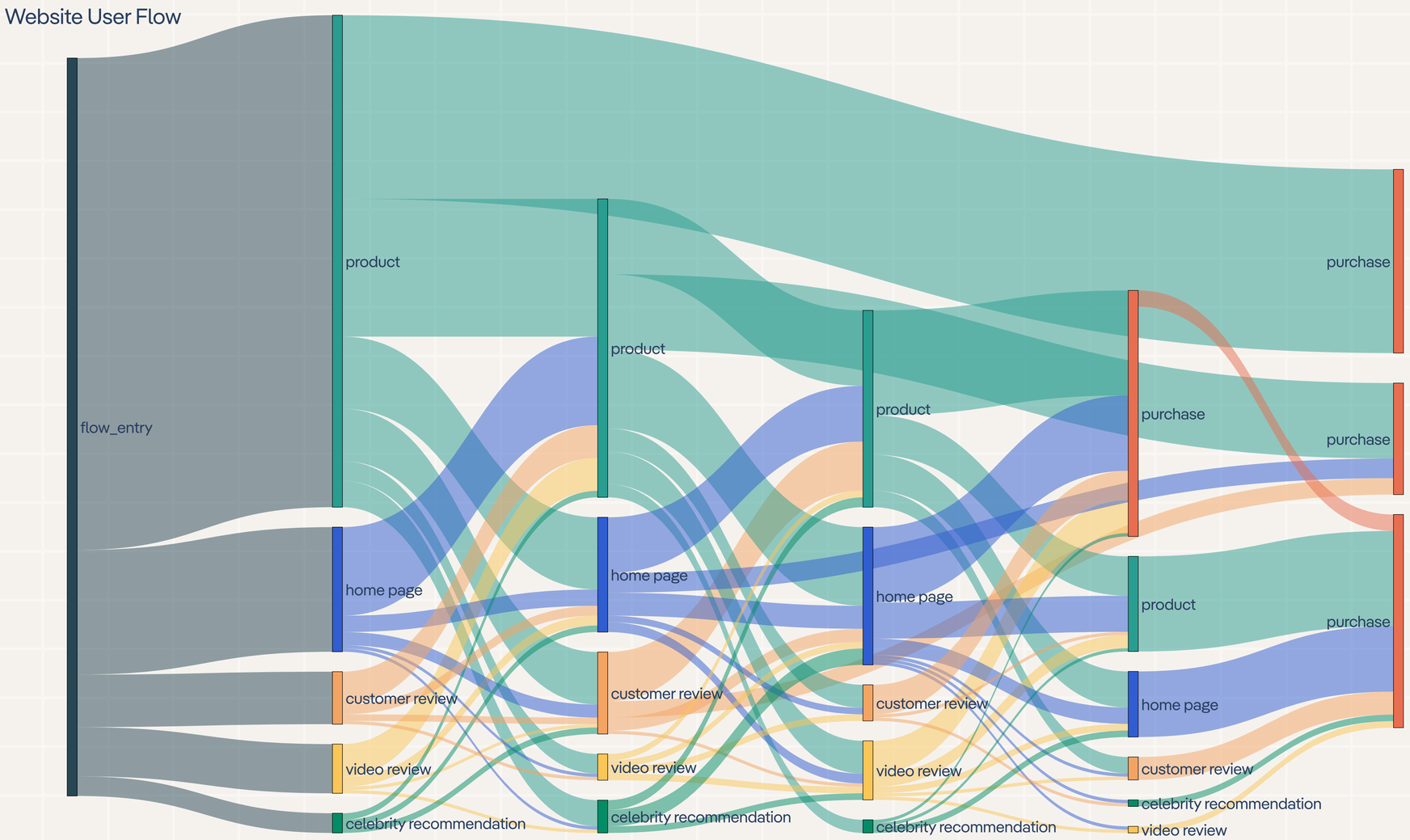
***Purchase** = 0.318 + 0.029 Customer reviews + 0.028 Celebrity reviews + 0.011 Video reviews*

Review Type	Coefficient	Revenue Increase per Customer Session	Total Increase for 1,000 Sessions	Total Revenue for 1,000 Sessions
Customer Reviews	0.029	£ 2.90	£ 2,900	£ 102,900
Celebrity Reviews	0.028	£ 2.80	£ 2,800	£ 102,800
Video Reviews	0.011	£ 1.10	£ 1,100	£ 101,100
Customer + Celebrity	0.057	£ 5.70	£ 5,700	£ 105,700
Customer + Video	0.040	£ 4.00	£ 4,000	£ 104,000
Celebrity + Video	0.039	£ 3.90	£ 3,900	£ 103,900
All Three Reviews	0.068	£ 6.80	£ 6,800	£ 106,800

5.18 increase in effectiveness of celebrity and customer combined compared to video.

1.34 increased effectiveness of celebrity and video reviews compared to customer

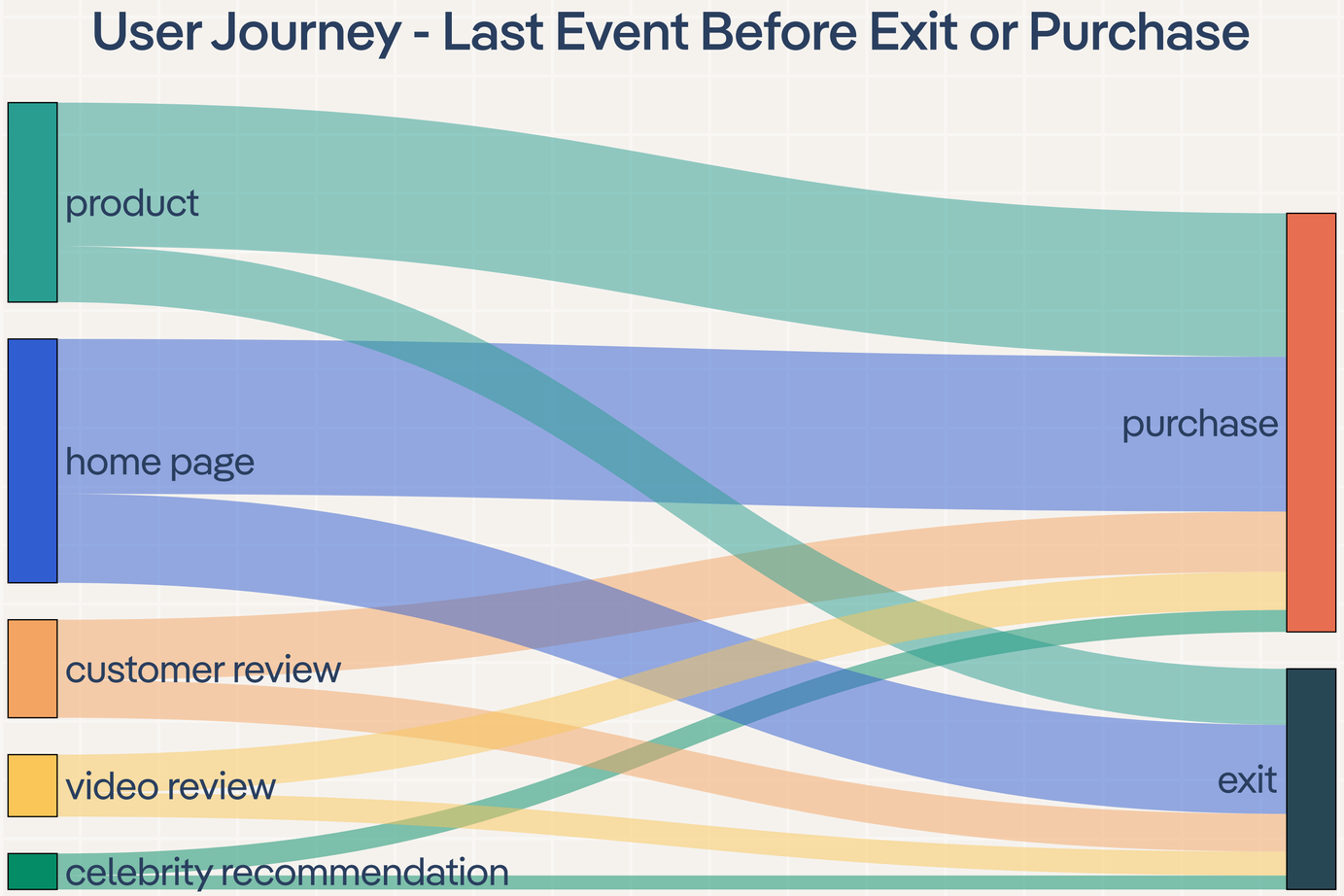
What kind of paths do customers that end up purchasing take (first 6 screens)?



Several customer browse through different pages within the homepage, or from review to product

What about Last Touch?

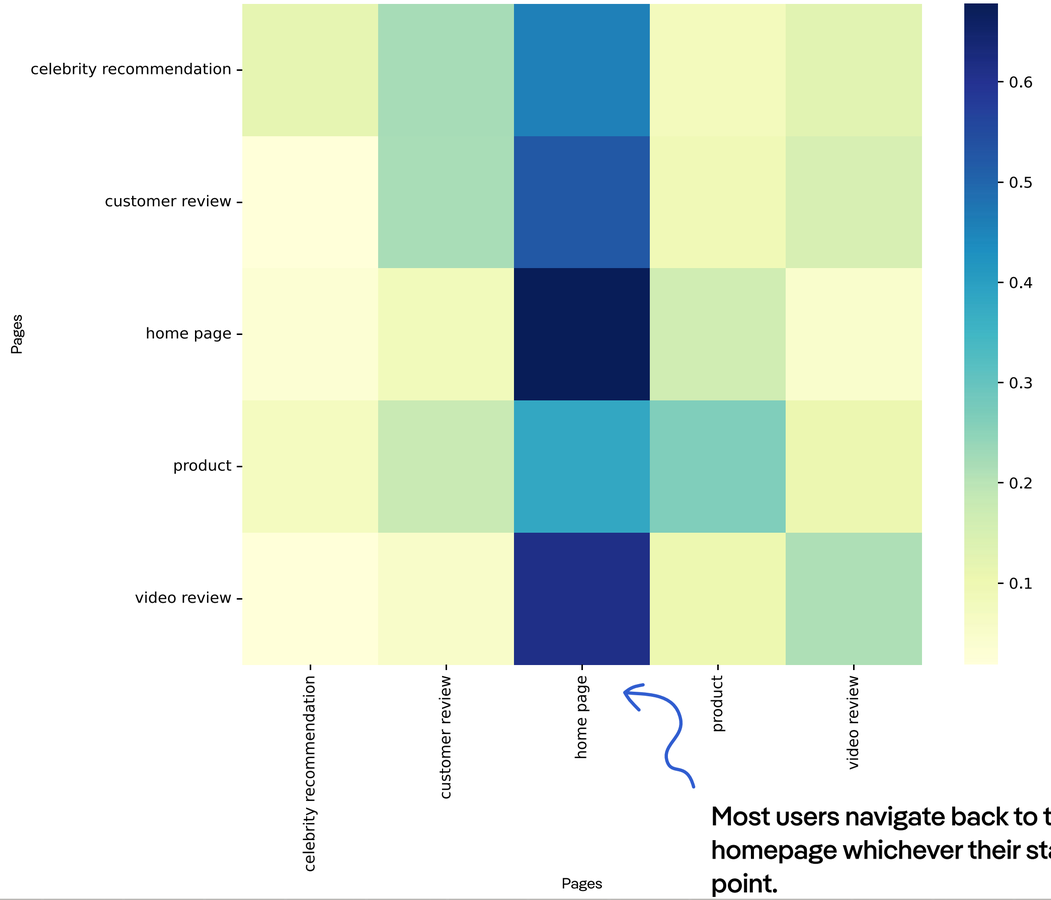
It shows also that many customers are already aware of their intention to make a purchase when they land on the homepage.



Most users that exit do so from the homepage. It highlights a lot of users drop from the flow from the landing page.

Identifying trends in clicking behaviour 🕵️

The matrix shows the likelihood of different pages being visited consecutively within a user's session

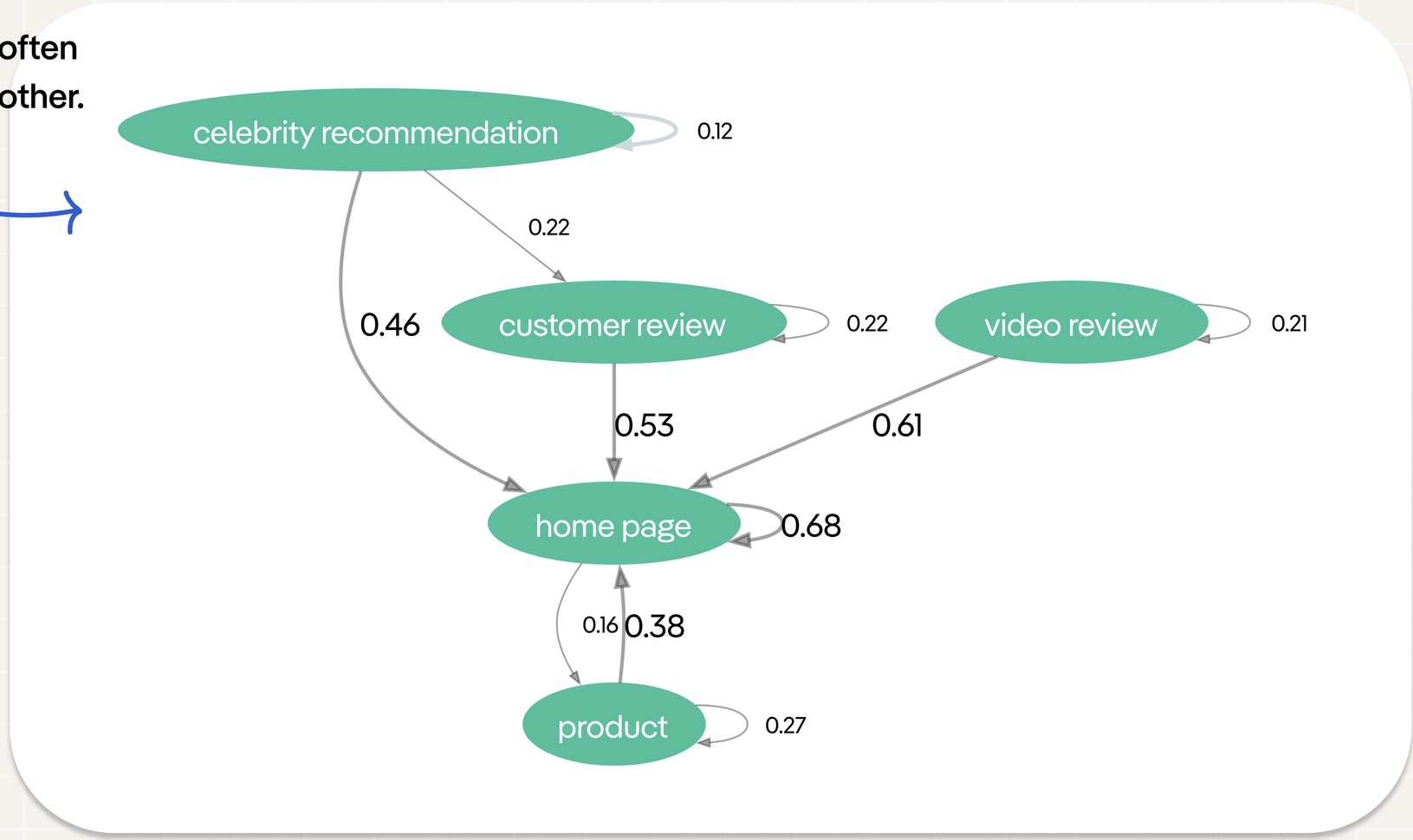


Customers also jump from one product to the other. Or review to review.

Most users navigate back to the homepage whichever their starting point.

Network Path Analysis of Customer Clickstream Behavior

Customers transition often from one review to another.



This is just a start

Conclusion

- Customer reviews and celebrity recommendations have the biggest impact on conversion rates and thus revenue, based on the attribution modelling and regression analysis
- The Markov chain analysis shows celebrity recommendations and customer reviews are key pages users visit before purchasing
- Longer session durations (10-30 mins) have higher conversion rates compared to very short (<5 min) or very long (>1 day) sessions
- Most sessions occur in the afternoon and evening hours

Limitations

- Doesn't factor in impact of pricing, promotions, inventory, etc. on purchasing behavior
- Lack of information in regards to actual revenue data and had to work with assumptions
- Limited visibility into behavior after a user exits the site (e.g. if they come back later to purchase)

Immediate reccomentations

- Highlight customer reviews and celebrity recommendations prominently on product pages and in marketing
- Focus ad spend and promotions on afternoon/evening hours when site traffic is highest
- Implement abandoned cart emails to re-engage users who viewed but didn't purchase

Reccomendations for the future

- A/B test different placements of reviews/celebrity recommendations to optimise
- User research/surveys to understand why users are exiting without purchasing
- Incorporate pricing, promotion, inventory data into propensity model
- Use predictive modelling to identify and proactively engage users likely to churn

Thank you

Now time for questions 🙋‍♀️