TRAVELTIDE SLIDE STORY

Slide 1

Hi there, welcome to our presentation. Today, we'll discuss the results of our feature engineering efforts, focusing on establishing an effective perks system designed to enhance customer retention and encourage repeat usage of our services.

Slide 2

As you know, TravelTide is now focused on gaining a deeper understanding of customer behavior to address customer retention issues. Elena, our newly appointed Head of Marketing, is leading the initiative to design and implement a personalized rewards program aimed at improving customer loyalty.

To create an effective rewards program, Elena is collaborating closely with our data team. She emphasizes that the perks that customers value most, such as free cancellations or free checked bag, are expected to boost customer returns. The strategy involves identifying distinct customer segments and aligning them with tailored travel perks, ensuring the program resonates with different customer groups.

Our project commenced with an exploratory data analysis, leveraging SQL and Tableau to delve into the initial data. We developed relevant metrics, followed by the creation of indices and a segmentation model in Excel. Finally, we built interactive dashboards in Tableau, providing a comprehensive view of the data and insights, including geographic analysis, to guide decision-making.

This presentation will walk you through the steps taken, from data exploration to the development of a targeted rewards program, highlighting how these efforts are poised to enhance customer retention at TravelTide.

Slide 3

Let's begin with the characteristics of our user cohort. Following Elena's recommendation, we selected customers who used the service after the Christmas holidays and had more than 7 sessions. This resulted in a cohort of 5,998 people, representing just 0.6% of TravelTide's entire user base. The cohort is predominantly from the US, accounting for 83%, with the remaining users based in Canada. Notably, around 21% of the cohort is concentrated in three cities: New York, Los Angeles, and Toronto.

Slide 4

Before diving into feature engineering, we analyzed the cohort's demographics and key consumer behavior patterns. Several features raised red flags. Firstly, there's a significant skew towards female users, who make up 88% of the cohort. We're not sure if this structure is representative of the entire customer base but it is certainly the case of class imbalance, which can affect the model's outcomes. Secondly, the travel patterns are highly unusual: most flights are either domestic within the US or between the US and Canada, with truly international flights accounting for just 3% of total flights. Notably, in Canada, flights to the US account for 79% of the total flights and thus are more common than domestic flights, which account for just 18%. Put it simply, five times more travel occurs between, say, Toronto and New York than between Toronto and other Canadian cities. Lastly, the profiles of US and Canadian customers are surprisingly similar, leading us to abandon the idea of developing separate predictive models for each country. These findings suggest that we're working with a highly biased sample that may not represent the broader TravelTide customer base. However, this is not unexpected since the cohort was selected based on browsing intensity rather than a representative random sample.

Slide 5

To extract effective predictive features out of the dataset, we expanded our analysis beyond browsing activity to include conversion rates (the ratio of bookings to total sessions) and customer lifetime value (CLV). Note that our 'lifetime' metric is confined to the period we have data for. Initially, we aimed to find correlations between browsing intensity, conversion rates, and CLV, broken down by age groups.

However, since individual browsing metrics didn't fully capture browsing intensity across age groups, we developed an integrated Browsing Activity Index, which scales and combines all three metrics. We then compared this index with conversion rates and CLV across age groups.

Key takeaway: Browsing intensity does not directly impact conversion rates or CLV when analyzed by age group. Older customers (60+ years) exhibit higher browsing intensity than middle-aged groups (30-59 years) but have significantly lower conversion rates and CLV. This raises an important question: Was it plausible to use browsing session count as the primary criterion for cohort selection?

Slide 6

Yet we can't say the same about different customer behavioral groups, and to be able to say this, we need to establish them by extracting and transforming more features. Our goal here is to identify distinct segments that can be targeted with each perk from Elena's list.

But first, we need to address a crucial aspect of TravelTide's business structure, at least within this cohort. As shown in the chart, around 75% of bookings combine both flights and hotel stays. It's logical to assume that 'combined bookers' will form the largest behavioral segment, absorbing significant chunks of both flight deal hunters and hotel deal hunters.

It's also important to note that, even when breaking down these combined bookings, flight spending consistently outweighs hotel spending. This occurs even if people might be more interested in hotel savings. Generally, flights have higher discount rates and a larger proportion of bookings under discount, meaning they not only account for the lion's share of booking spend but also offer the most discount value to customers. This dynamic is likely to influence preferences for flight-centric perks, especially for those who make combined bookings.

Slide 7

Now, let's dive into our model. We've developed six hypothetical behavioral segments that would be the primary targets for our six perks: Flight Bargain Hunters, Hotel Bargain Hunters, Hotel-Only Bookers, Combined Bookers, Heavy Packers, and Flexible Travelers. While we won't go into detail about each segment here, you can refer to our extended report for a deeper understanding of their demographic profiles and consumer behaviors.

In analyzing the data, we've assigned distinguishing features to each segment. However, at this stage, these segments are not mutually exclusive. Our next step is to refine them using a series of indices, which should help us differentiate these groups more clearly, at least in theory.

Slide 8

Here's the schematic representation of our model using the Index Ranking method. As illustrated, we first normalized the selected metrics to ensure they were on the same scale before calculating the indices. Once the indices were calculated, we ranked users in descending order based on their respective index. The next step involved ranking them from MIN to MAX across indices, assigning the MIN as the definitive index. In cases where multiple MIN ranks appeared across the index columns, the first MIN rank was selected as the appropriate one.

At the output stage, we unexpectedly identified seven segments instead of six, including a group we call 'Passive Deal Seekers' (8% of the total). These users, despite showing high browsing activity (8+ sessions), do not convert into active customers by making bookings. As anticipated, Combined Bookers (39%) constitute the largest segment, which includes many potential flight-centric bargain seekers, leaving the pure Flight Bargain Hunters segment relatively small. Interestingly, Hotel Bargain Bookers emerged as the second largest segment, accounting for 20% of the cohort.

Slide 9

And voilà, the next chart confirms that browsing activity does not directly influence conversion rates or value creation across customer segments, either. For instance, despite having the highest browsing activity index, Flexible Travelers exhibit one of the lowest conversion rates. Conversely, Combined Bookers are nearly twice as 'browsing intensive' as Hotel Bargain Hunters, yet the latter show higher conversion rates and CLV. Lastly, the Passive Deal Seekers spend significant time browsing without converting or adding value. The key takeaway: browsing activity, particularly the number of sessions, is not an effective criterion for defining a representative user cohort.

Slide 10

Let's get to our conclusions and recommendations. First, we've established that browsing activity does not necessarily lead to conversion. Relying solely on this metric for cohort selection risks excluding users with high conversion rates and CLV who may have fewer than 8 browsing sessions. Therefore, it's crucial to incorporate additional features that align more closely with real demographic and business structure when defining segments.

Second, within our setup cohort, we've identified a segment of Passive Deal Seekers who, despite high browsing activity, do not convert or create value. These users should be excluded from the rewards program. However, they could be targeted with a different incentive program, such as an "Exclusive discount on the first booking," to stimulate conversion.

Third, to validate the relevance and effectiveness of targeted perks, we recommend conducting A/B testing. The success of these perks should be measured using key metrics like conversion rates and customer value.

And finally, the next logical step after our feature engineering efforts is to develop a full-fledged predictive machine learning model. This model will be enhanced with new data from the trial reward campaign and can significantly improve not only targeted marketing (such as reward campaigns), but also customer insights, retention, and resource allocation.

Slide 11

Thank you for your attention today. As we conclude, I'd like to remind you to refer to the full report for more detailed information on the demographic analysis of our customer segments, as well as deeper insights into the consumer behavior patterns we've uncovered.

Additionally, a comprehensive Excel-based Index model for the entire cohort is available for your review. This model allows you to follow the complete feature engineering journey—from key account metrics through benchmark scaling and index calculation, to self-generated perk recommendations for each customer account within the cohort.

I also encourage you to explore our interactive dashboards. These dashboards are designed as ready-to-use business intelligence tools, enabling you to view all key benchmarks in their interconnections. We're particularly excited about our geographic approach to cohort analysis, offering a granular view of user behavior at the city level. This localization allows for even more personalized and meticulously designed perks. For instance, you can easily identify which of your customers live in a specific town, where they typically fly, how long they stay, how much they spend, and many other details that can refine your marketing strategies.

All these additional resources are accessible via the links provided here.

Thank you once again, and I look forward to receiving your feedback and discussing any questions you may have.