

Characterizing Intent Using Customer Journey: a Sequential and Graphical Model Approach

PROBLEM STATEMENT

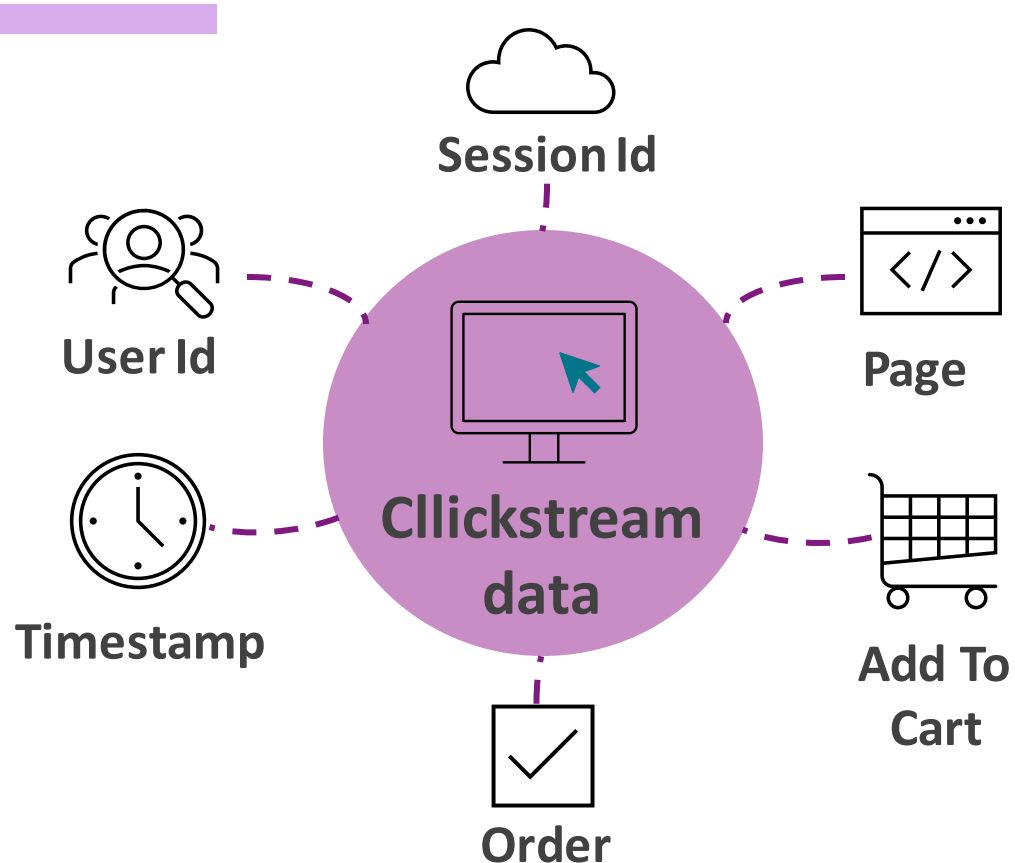


What are the different personas in Wayfair customer base?

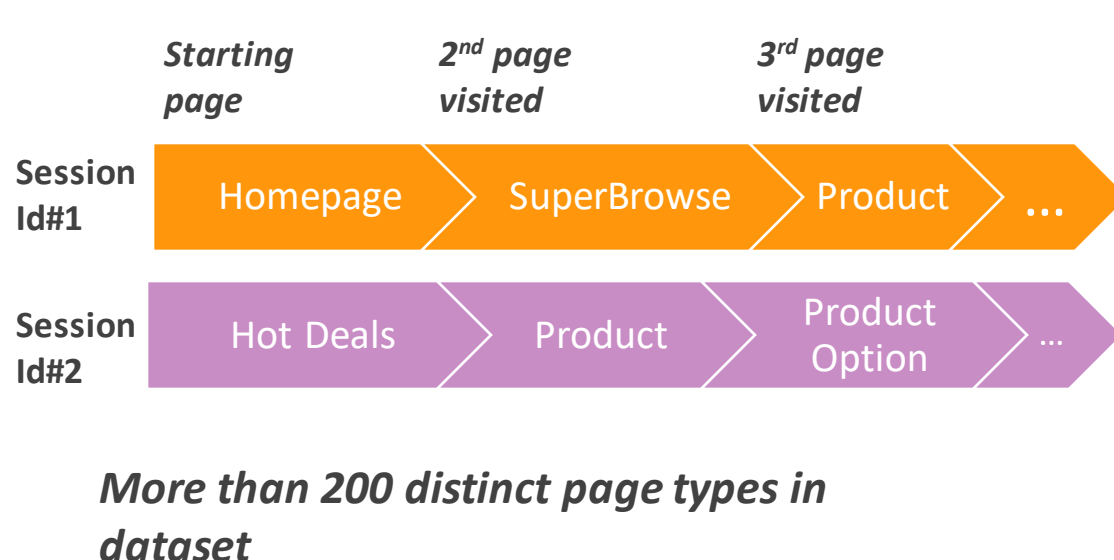
Identifying and clustering customer behaviors is a first step towards personalizing the customer journey.

Wayfair clickstream data logs every action customers take on the website and has the potential to provide insight into consumers behaviors.

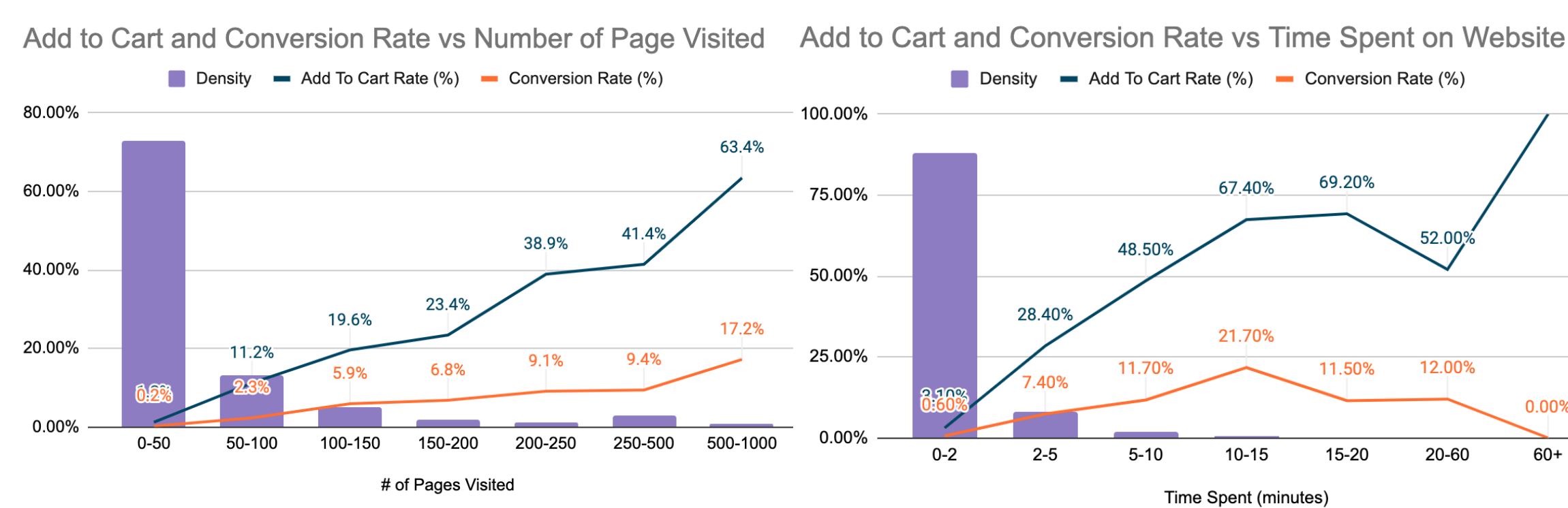
DATASET



Starting from individual clickstream data, we **reconstruct the sequence of pages** visited by a customer during a given visit



EXPLORATORY ANALYSIS

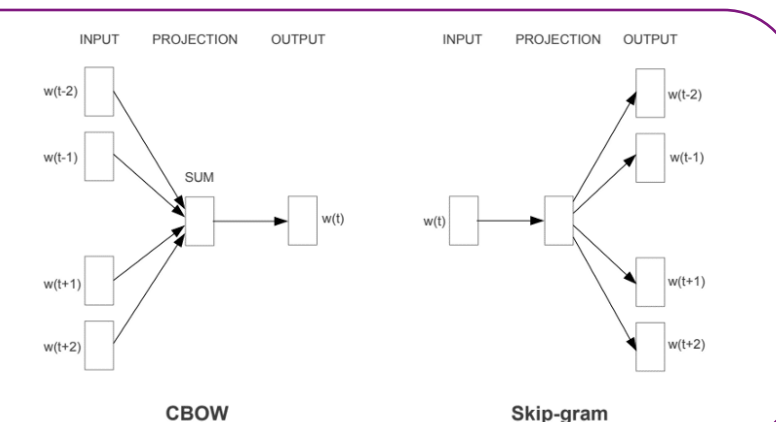


METHODOLOGY

Our method takes into account both the **sequential aspect** and the **graphical aspect** of the data to build an embedding for each session. We then perform **unsupervised clustering** algorithms to obtain the final clusters.

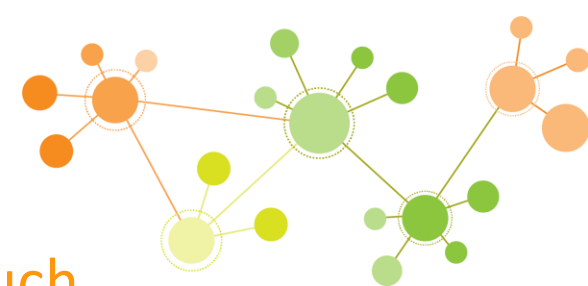
Leveraging sequential aspect of data

A sequence of pages can be regarded as a sentence. We train a **Word2Vec model** in order to obtain **embeddings** for every single page in our dataset. From there, we compute a **session embedding** using a **TF-IDF averaging**.



Leveraging graphical aspect of data

The journey followed by a customer during a visit can be seen as a graph. **Nodes** represent the **pages** visited, and **edges** represent the **connections** between pages, carrying transition probability in their weight. Building a **graph** for each **session**, we can then extract features such as **density**, **centrality** or **betweenness** to characterize the session.

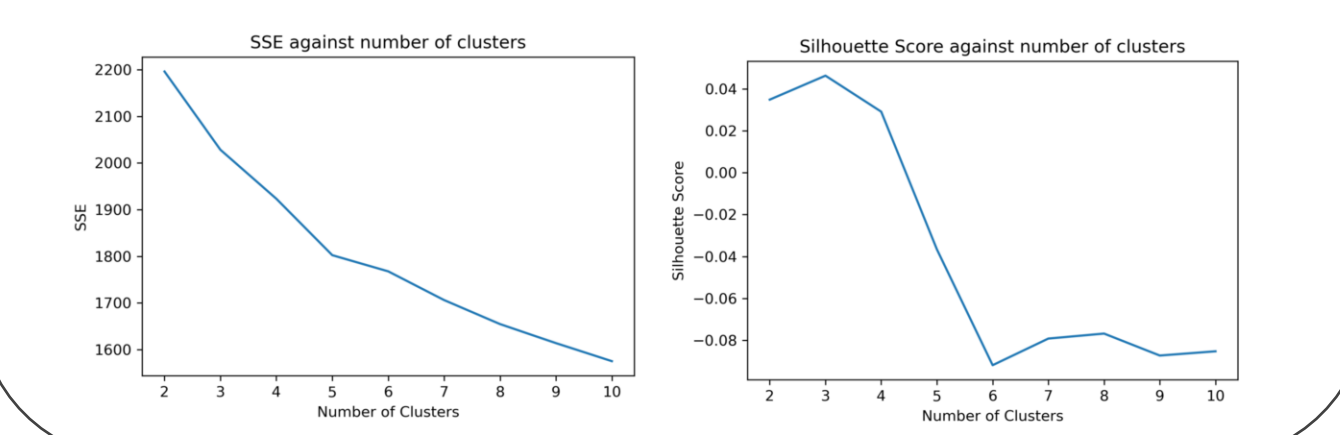


We concatenate **NLP trained embeddings**

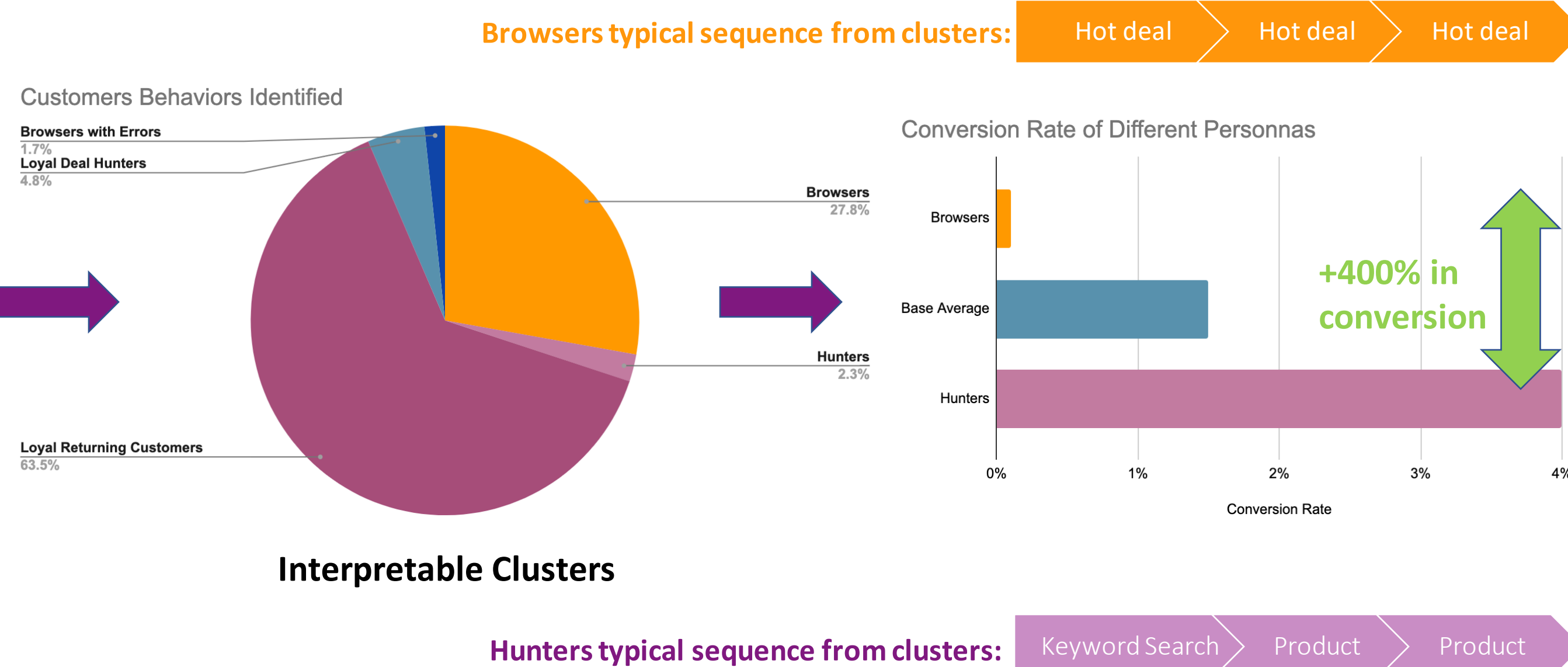
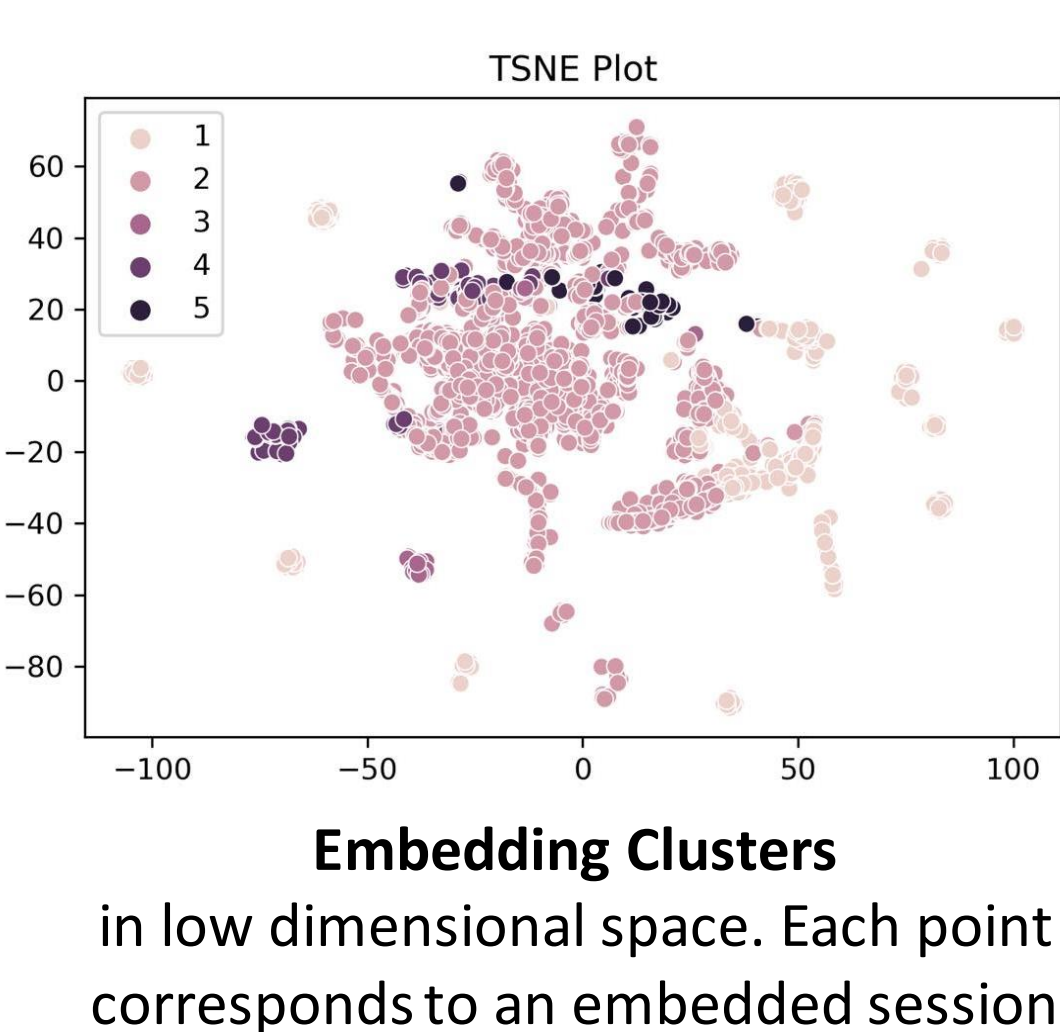
and **graphical features**

Unsupervised Clustering

We use K-Means algorithm to cluster the embeddings obtained. We figure out the optimal number of clusters by analyzing the evolution of Sum of Squared Errors (SSE) and Silhouette coefficient as the number of clusters increase. We find that the optimal number of clusters is 5.



RESULTS



We observe **strong discrimination between clusters** which links back to the **business intuition**:

- Different **conversion rates**
- Different **Add-To-Cart rates**
- Different **typical subsequences**
- Different **channel distribution**
- Different **visitor types**

IMPACT & NEXT STEPS

- We **developed a framework** that takes minimum feature engineering and **combines graphical and sequential nature of clickstream data** to identify customer's behaviors on the website.
- Our method translates in **identifying** very clearly **5 interpretable clusters**, namely **hunters, browsers, loyal deal hunters, loyal returning customers** and **browsers with errors**.
- These results can be immediately used to **study shifts in customer profile over time** and add knowledge to Wayfair database.

- By improving customer journey for specific personas, **post purchase touchpoints can be reduced by 10X**, and **reduction of customer service cost and increased customer return rate can translate overall into \$100M additional revenues**
- The **insights** from the model will be **used to do back-testing and AB test** different summarized insights from our work.
- On a technical standpoint, **model improvement** can be considered by integrating **new features about customers, catalogue specific details on pages** and other graphical features that can be used to improve clustering results.