Traveltide Customer Segmentation Analysis (Summary)

E-booking startup **TravelTide** is a hot new player in the online travel industry. It has experienced steady growth since it was founded at the tail end of the COVID-19 pandemic (2021-04) on the strength of its data aggregation and search technology, which is best in class. Customer feedback has shown - and industry analysts agree - that **TravelTide** customers have access to the largest travel inventory in the e-booking space. Following the startup playbook, **TravelTide** has maintained a hyper-focus on building an unfair advantage along a limited number of dimensions - in this case, building the biggest travel inventory and making it easily searchable. Because of this narrow focus, certain aspects of the **TravelTide** customer experience are underdeveloped, resulting in poor customer retention.

To keep more customers a **reward system** shall be applied!

Since Customer and their behavior differ tremendously, we need to **segment** them first to provide them the **best fitting rewards**!

Rewards we can offer**:**

- Free Meal
- Free Bag
- No cancellation fees
- Exclusive discounts
- Free flight + one hotel night

Steps of Analysis

1. Data Exploration and Cleaning
2. Feature Engineering
3. Segmentation
4. Perk assignment

INFORMATION:

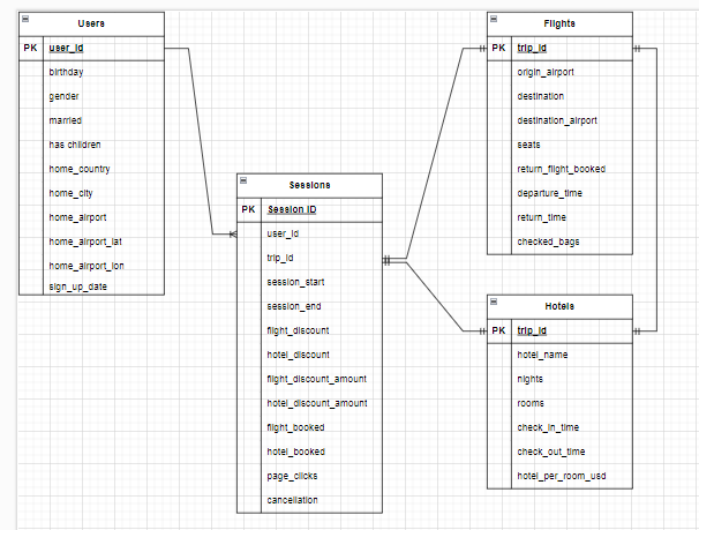
*The whole Analysis was performed in **Google Collab**!*

FULLY DOCUMENTED ANALYSIS:

https://colab.research.google.com/drive/1rlyWdrRorVldK20oZ2KM7ZSmvUNuiage?usp=sharing

1. Data Exploration and Cleaning

To understand the Data, the Schema of the **TravelTide** Database got visualized. There were one Dimension-table (**Users**) and three fact tables (**Sessions**, **Hotels**, **Flight**) The Session table Connects all tables with each other and had the Most entries (5408062).



After the broader exploration of the Database, the single tables got inspected. Following adjustments were made:

- **Converting of datatypes** (dates in string format to datetime format)
- **Splitting all of string-columns** which hold more than one information (Session ID, Trip ID, Hotel Brand, Hotel City)

In the Hotels-table one columns ("**Nights**") showed **untrustworthy** Data since it included some negative- as well as null-Values. The Values got **imputed** by using the "**return time**" as an estimator of the spend Nights. If there was no return time, then the value 1 was taken in place. There were no other outliers found in other columns which needed to be dealt with.

After the exploration and cleaning process all Tables got **joined** together in one **Sessions-Table**

This Main-Table got also filtered for **recen**t and **active** users.

Criteria given by management:

- **More than 6 Session**
- **After 04.01.2023**

2. Feature Engineering

Since we want to segment on a **user-level,** user-aggregated measures need to be made first. We already had 10 Features in the original User Table, but they just hold small amounts of behavioral information. Following Features got generated from the **Main-Table** to gather insights on behavioral information (*For deeper insights in calculation see linked Collab file):*

Number of Sessions, AVG Session Length, Total Cancellations, AVG Page Clicks, Total Trips, Total booked Flights, Total booked Hotels, Number of discounted Flights, Number of discounted Hotels, Total Spent: Flights, Total Spent: Hotels, Total Spent: Overall, AVG Seats per Trip, AVG Bags per Trip, AVG Nights in Hotel, AVG Rooms in Hotel, Total Last Minute Flights, AVG Trip Duration, Total Flights Distance, Favorite Destination (Mode), Discounted Flight User (Mode), Discounted Hotel User (Mode), Favorite Hotel Brand (Mode), Conversion Rate, Activeness Score, Spending per Trip, Percent of Return Flights, Percent of Last-Minute Trips

After their calculation, the new metrics got **joined** to the filtered (**active** and **recent) Users-Table** to continue with the Segmentation process.

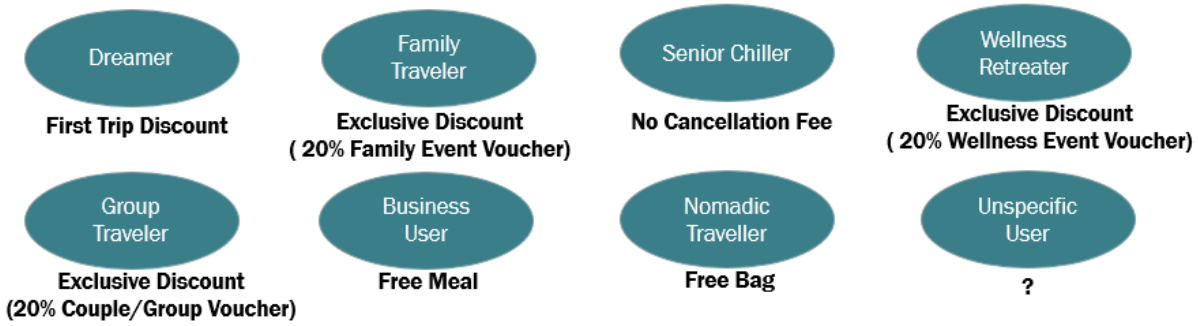
3. Segmentation

For the **Segmentation-process** two **different approaches** got used and combined in the end.

1. **Decision Tree** (Psychological theory driven)
2. **Machine Learning** (KMeans)

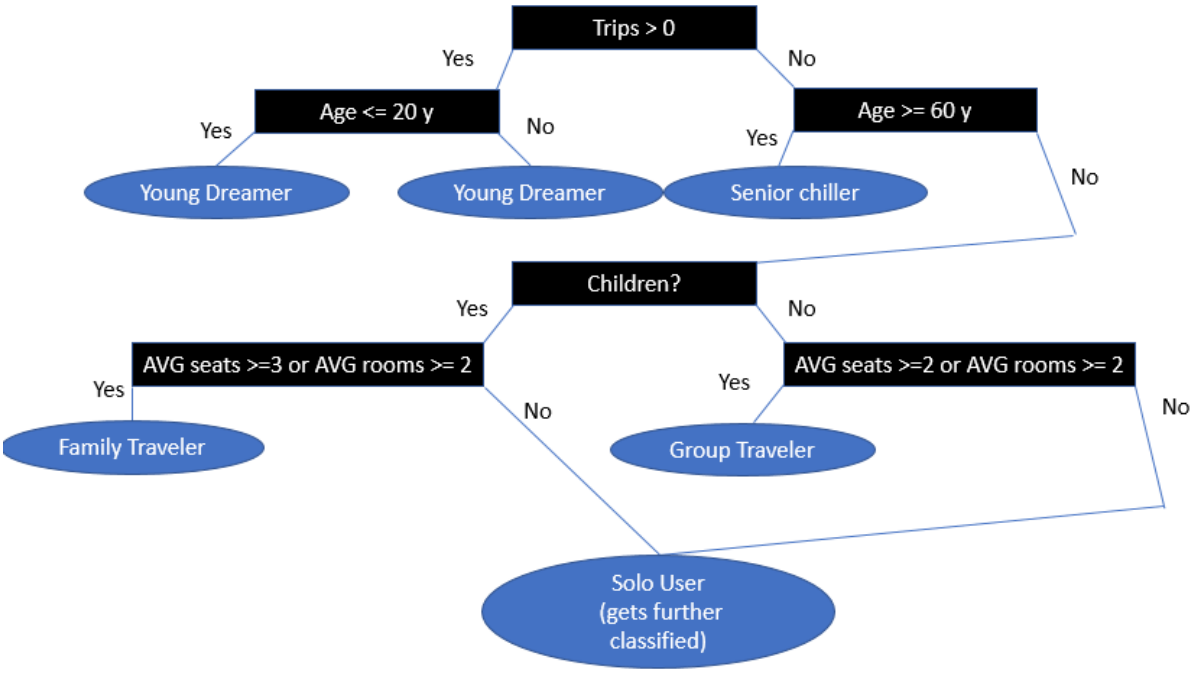
Decision Tree Approach

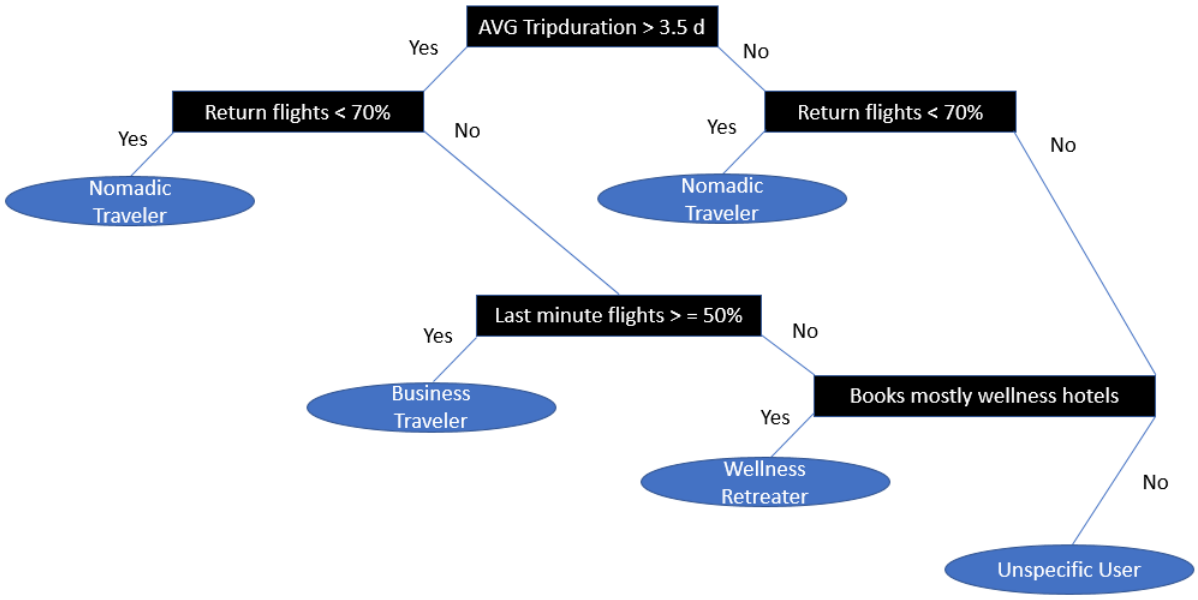
Using of **domain knowledge** to assume which segments of users are most likely and which perks are appropriate due to their attributes:



How to **Assign** the Segments?

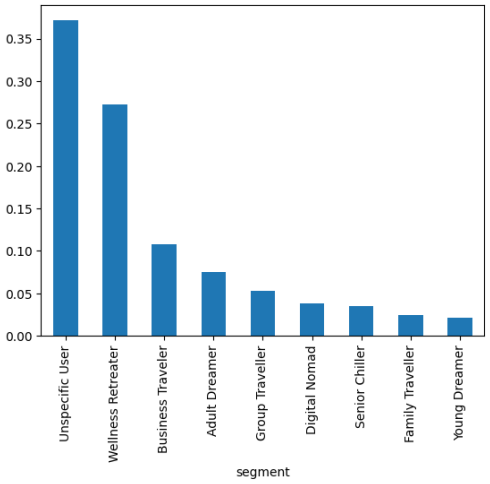
Use of specific empirically and theoretically derived Thresholds to get a clean separation:





Findings of Distribution Tree Approach

- **37%** of Users are **unspecific**
- Due to very **strict segmentation**
- **ML** approach will **supplement** this



Machine Learning Approach (KMeans)

Finds the mathematically correct segmentation based on distances calculated from pre-defined metrics. The Metrics which got chosen for the Algorithm were:

Has Children, Age, Total Sessions, Total Cancellations, Total Trips, Total booked Flights, Total booked Hotels, Total Spent: lights, Total Spent: Hotels, Avg Seats, Avg Rooms, Conversion Rate, Spending per Trip, Percent of Return Flights, Percent of Last-Minute Trips

5 Groups got identified:

The Low-Budgeter **Perk:** 5% Trip Discount

Median	C	T	nF	nH	$F	$H	S	R	CR %	$T	%RF	%LM
Overall	0	2	2	2	713.7	1156.0	1	1	28.5	822.0	100	50.0
Low-Budgeter	0	2	1	1	435.5	720.0	1	1	28.5	767.2	100	50.0

C – Total Cancellations, T – Total Trips, nF – Number of Flights, nF – Number of Hotels, $F – Spending on Flights, $H – Spending on Hotels, S – Seats, R – Rooms, CR% – Conversion Rate in %, $T – Spending per Trip, %RF – Percent of Return Flights, %LM – Percent of Last-Minute Flights

The Not-Alone-Vacationer **Perk:** Free Bag

Median	C	T	nF	nH	$F	$H	S	R	CR %	$T	%RF	%LM
Overall	0	2	2	2	713.7	1156.0	1	1	28.5	822.0	100	50.0
Not-Alone-Vacationer	0	2	1	2	1109.2	2652.0	2	2	28.5	2214.7	100	33.3

C – Total Cancellations, T – Total Trips, nF – Number of Flights, nF – Number of Hotels, $F – Spending on Flights, $H – Spending on Hotels, S – Seats, R – Rooms, CR% – Conversion Rate in %, $T – Spending per Trip, %RF – Percent of Return Flights, %LM – Percent of Last-Minute Flights

The Veteran User **Perk:** Free Flight and Hotel Night on 5th Trip

Median	C	T	nF	nH	$F	$H	S	R	CR %	$T	%RF	%LM
Overall	0	2	2	2	713.7	1156.0	1	1	28.5	822.0	100	50.0
Veteran User	0	4	3	3	1267.8	1923.5	1	1	50.0	895.0	100	50.0

C – Total Cancellations, T – Total Trips, nF – Number of Flights, nF – Number of Hotels, $F – Spending on Flights, $H – Spending on Hotels, S – Seats, R – Rooms, CR% – Conversion Rate in %, $T – Spending per Trip, %RF – Percent of Return Flights, %LM – Percent of Last-Minute Flights

The Canceller **Perk:** No Cancellation Fee

Median	C	T	nF	nH	$F	$H	S	R	CR %	$T	%RF	%LM
Overall	0	2	2	2	713.7	1156.0	1	1	28.5	822.0	100	50.0
Group	1	2	2	2	601.2	979.2	1	1	28.5	844.0	100	50.0

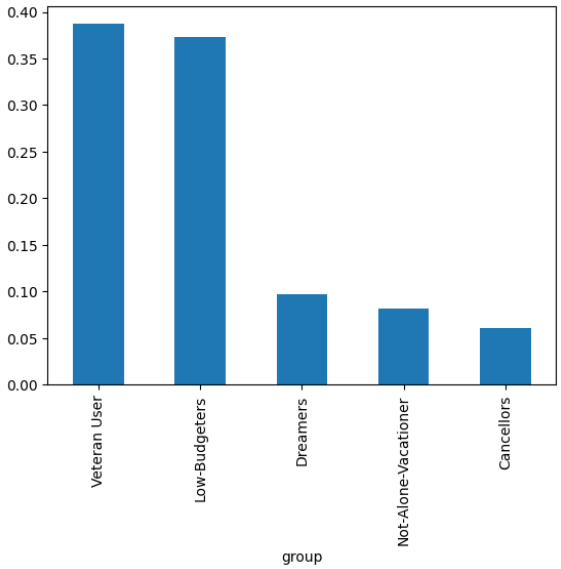
C – Total Cancellations, T – Total Trips, nF – Number of Flights, nF – Number of Hotels, $F – Spending on Flights, $H – Spending on Hotels, S – Seats, R – Rooms, CR% – Conversion Rate in %, $T – Spending per Trip, %RF – Percent of Return Flights, %LM – Percent of Last-Minute Flights

The Dreamer **Perk:** See Decision Tree

Pre-defined group, not included in the ML- Algorithm!

Never Booked a Trip.

- Biggest Group is Veteran Users
- All people get assigned to a group



Cross Validation

segment / group	Adult Dreamer	Business Traveler	Digital Nomad	Family Traveller	Group Traveller	Senior Chiller	Unspecific User	Wellness Retreater	Young Dreamer
Cancellers	0	198	57	39	75	58	624	420	0
Dreamers	1811	0	0	0	0	0	0	0	523
Low-Budgeters	0	1428	295	91	240	555	3769	2612	0
Not-Alone-Vacationer	0	5	32	462	868	149	249	190	0
Veteran User	0	970	532	10	92	81	4296	3331	0

4. Perk Assignment

segment group	Adult Dreamer	Business Traveler	Digital Nomad	Family Traveller	Group Traveller	Senior Chiller	Unspecific User	Wellness Retreater	Young Dreamer
Cancellers		No Cancellation Fee OR Free Meal	No Cancellation Fee or Free Bag	20% Family Event Discounts or No Cancellation Fee	20% Group/Couple Event Discounts or No Cancell...	No Cancellation Fee	No Cancellation Fee	20% Wellness Voucher or No Cancellation Fee	
Dreamers	20 % Discount on first trip								30 % Discount on first trip
Low-Budgeters		5% Trip Discount OR Free Meal	5% Trip Discount or Free Bag	20% Family Event Discounts or 5% Trip Discount	20% Group/Couple Event Discounts or 5% Trip Di...	No Cancellation Fee or 5% Trip Discount	5% Trip Discount	20% Wellness Voucher or 5% Trip Discount	
Not-Alone-Vacationer		Free Flight and Hotel Night (5th flight) OR Fr...	Free Flight and Hotel Night (5th flight)	Free Flight and Hotel Night (5th flight)	Free Flight and Hotel Night (5th flight)	Free Flight and Hotel Night (5th flight)	Free Flight and Hotel Night (5th flight)	Free Flight and Hotel Night (5th flight)	
Veteran User		Free Bag OR Free Meal	Free Bag	20% Family Event Discounts or Free Bag	20% Group/Couple Event Discounts or Free Bag	No Cancellation Fee or Free Bag	Free Bag	20% Wellness Voucher or Free Bag	

Take Home

- This was just a first approach on a limited data foundation!
- For the Users, two different segments can be applied since there are two approaches which have different angles of view
- To know which of the two works better in general – AB-testing should be applied in the future