Executive Summary “Customer Segmentation"

TravelTide

TravelTide has maintained a hyper-focus on building the biggest travel inventory and making it easily searchable. But certain aspects of the TravelTide customer experience are underdeveloped, resulting in poor customer retention.

The objective of the project is to design and execute a personalized **rewards program** that keeps customers returning to the TravelTide platform.

The methodologies for data exploration, metric development, and client segmentation integrated SQL, Tableau, and spreadsheets. SQL was used to calculate key performance indicators and uncover behavioral patterns, while Tableau enabled dynamic segmentation through Ranking, visual analysis and calculated fields. This approach supported the identification of distinct client groups and the assignment of potential perks based on performance insights.

The Key Metrics we used to build a segmentation around our customers´ behavior which would potentially match the perks proposed by head of Marketing are:

Metrics Used for Ranking

User ID	Avg Km Flown	Money Spent Flight	Money Spent Hotel	Number Of Nights
94883	1,451	864	230	2
101486	996	190	2,452	9
101961	1,367	1,238	2,798	22
106907	12,797	27,804	8,514	11
118043	2,505	2,339	6,443	24
120851	2,260	385	2,002	11
125845	2,862	1,647	864	7
133058	2,887	424	0	0
149058	1,960	1,919	6,274	30
152583	2,266	368	639	3
153982	1,344	657	1,302	7
160754	528	94	642	3
164522	1.808	312	3.262	14

- *Average Kilometers Flown (how many kilometers in average did our customers flown in total, calculated with home Latitud and Longitud towards destination latitud and longitud)*
- *Money Spent in Hotels (how much money was spent in total only in Hotels, calculated with base fare amount minus discount)*
- *Money Spent in Flights (how much money was spent in total only in Flights, calculated with base fare amount minus discount)*
- *Total number of Nights stayed (how many number of nights did our customers stayed in the trip)*

Recommendations for Segmentation and Perk assignment

- ☆ Potencial New Customer - 10% discount first trip
- ☆ Heavy Luggage Customer - Free checked bag
- ☆ Long-Stay Customer - Free Hotel Night
- ☆ Flyer Customer - 15% discount next round flight
- ☆ Luxury Dinner Customer - Free Hotel Dinner

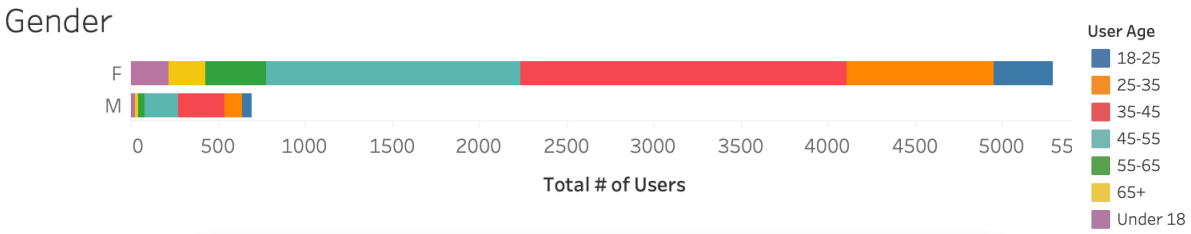
Methodology

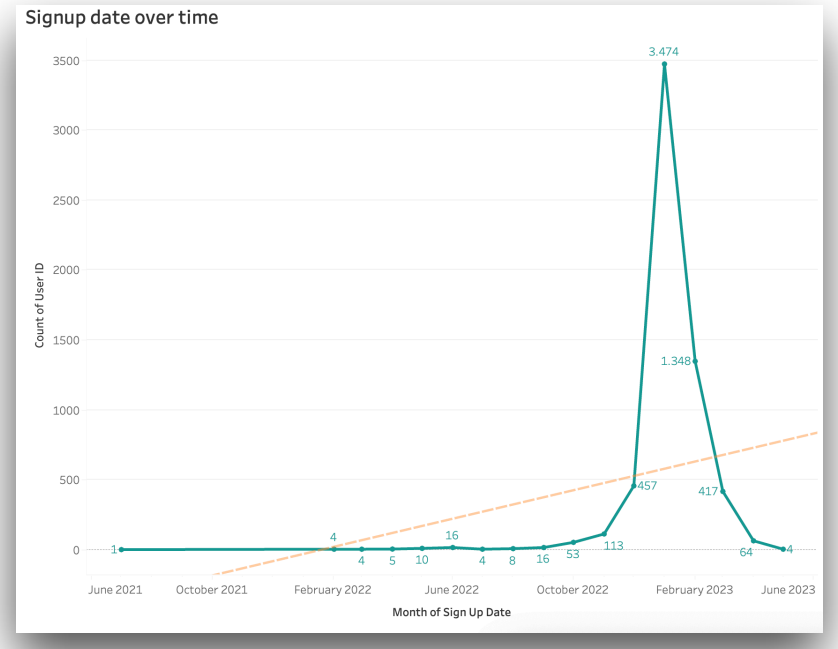
Data Exploration

To explore and analyze the dataset, Structured Query Language (SQL) was employed as the primary tool for data extraction, transformation, and aggregation. The process began by querying raw tables to understand data structures, identify key relationships, and filter relevant records. SQL functions such as JOIN were used in order to join all tables provided “Users”, “Hotels”, “Flights” and “Sessions”. We analyzed the Age Ranking for our customers, as well as all demographic data that was provided (Gender, Has Children, Married, etc.).

We saw a spike of signup date in 2023, which was later utilized for filtering the Data in order to not have many outliers. Date used: 2023-01-04. As well as filtered out users with less than 7 sessions created in the platform.

After applying the above mentioned filters we cleaned the Data with a correction in the number of nights and number of rooms summed for each user. We noticed inaccurate data related to the previous dimensions, and made the necessary corrections using check in and out time, and number of nights only for hotels not cancelled.





SQL Methodology for Devising Metrics

SQL was used to define and calculate key performance metrics essential to the project's analytical objectives. This involved creating derived fields using conditional logic (`CASE WHEN`), aggregating measures with functions such as `COUNT()`, `AVG()`, and `SUM()`, and applying `GROUP BY` clauses to evaluate performance across different dimensions (e.g., user level, time periods, or trip metrics). CTEs were created to manipulate the data an were employed to normalize and compare user behavior or outcomes. These custom metrics provided a structured and repeatable framework for measuring engagement, performance, and trends, forming the foundation for subsequent ranking, segmentation, and visualization tasks.

```sql
SELECT
  u_id,
  AVG(nights_new) as avg_nights,
  AVG(checked_bags) as avg_bags,
  AVG(seats) as avg_seats,
  SUM(seats) as total_seats,
  SUM(new_rooms) as count_rooms,
  SUM((1-new_flight_discount) * base_fare_usd) AS money_spent_flight,
  SUM(hotel_per_room_usd * new_rooms * nights_new * (1 - new_hotel_discount)) as money_spent_hotel,
  AVG(haversine_distance(
        home_airport_lat,
        home_airport_lon,
        destination_airport_lat,
        destination_airport_lon
      )) as avg_km_flown,
  COUNT(DISTINCT trip_id_sessions) as num_trips,
  COUNT(CASE WHEN flight_discount_amount is not null
        OR hotel_discount_amount is not null
```

Save Run

u_id	to...	sign_up_date	seconds	user_age	gender	has_children	married	home_country	home_city	avg_seat
23557	8	2021-07-22	76.625000	55-65	F	false	true	usa	new york	(NULL
94883	8	2022-02-07	67.750000	45-55	F	false	true	usa	kansas city	1.500(
101486	8	2022-02-17	122.250000	45-55	F	true	true	usa	tacoma	1.000(
101961	8	2022-02-17	117.750000	35-45	F	false	true	usa	boston	1.000(
106907	8	2022-02-24	758.915066	35-45	F	true	true	usa	miami	6.000(
118043	8	2022-03-10	153.125000	45-55	F	true	false	usa	los angeles	2.000(
120851	8	2022-03-13	51.750000	45-55	F	true	true	usa	santa ana	1.000(
125845	9	2022-03-19	131.777778	35-45	F	false	true	usa	el paso	1.333
133058	8	2022-03-27	68.125000	18-25	F	false	false	usa	austin	1.000(
149058	8	2022-04-14	131.750000	45-55	F	true	false	usa	birmingham	1.200(

5998 0 affected 41.01 s Download

Metrics Used for Ranking

User ID	Avg Km Flown	Money Spent Flight	Money Spent Hotel	Number Of Nights
23557	0	0	3,671	20
94883	1,451	864	230	2
101486	996	190	2,452	9
101961	1,367	1,238	2,798	22
106907	12,797	27,804	8,514	11
118043	2,505	2,339	6,443	24

-The output was a CSV with a UserLevel DB (with 5998 rows) cleaned and ready to upload in Tableau for accurate visualizations.

Tableau Methodology for Customer Segmentation

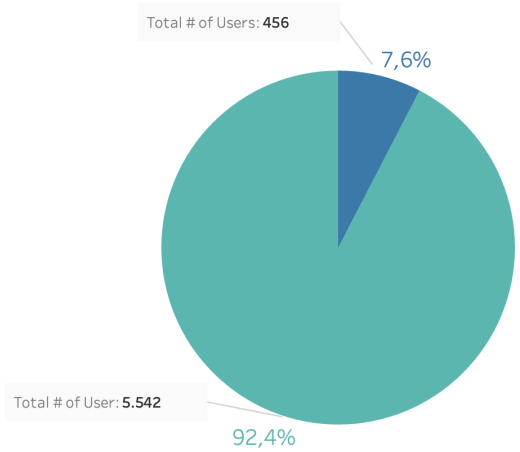
In the third stage of the project, Tableau was utilized to develop dynamic customer segmentations based on behavioral and performance metrics derived from prior SQL analysis. Using calculated fields, ranking logic, and set membership, customers were grouped into distinct segments reflecting patterns such as activity level, engagement, or metric-based performance. Tableau's interactive features—such as filters, and set actions—enabled flexible exploration of these segments, allowing for quick comparisons and visual analysis of segment-specific trends. This ranking segmentation approach supported strategic decision-making by highlighting key differences between customer groups in a clear, data-driven manner.

Ranking segmentation

User ID	Rank by MS Hotel	Rank by MS Flights	Rank by Num of Nig..	Rank by Km Fl..	MIN rank
23557	0,185426046	1	0,122061031	1	0,122061031
94883	0,867933967	0,502584626	0,875104219	0,655994664	0,502584626
101486	0,330998833	0,816741704	0,515924629	0,774553944	0,330998833
101961	0,278305820	0,336334834	0,089711522	0,683675171	0,089711522
106907	0,022011006	0,000333500	0,413540103	0,009671502	0,000333500
118043	0,054027014	0,110221778	0,065532766	0,273470068	0,054027014
120851	0,406703352	0,740370185	0,413540103	0,350341838	0,350341838
125845	0,686009672	0,218275805	0,626813407	0,194930799	0,194930799
133058	1	0,721193930	1	0,188594297	0,188594297
149058	0,059363015	0,164082041	0,024512256	0,467567117	0,024512256
152583	0,753710188	0,748207437	0,834917459	0,348841087	0,348841087
153982	0,567950642	0,610305153	0,626813407	0,690511923	0,567950642
160754	0,752709688	0,847757212	0,834917459	0,840253460	0,752709688
164522	0,219609805	0,772719693	0,280807070	0,526263132	0,219609805
167852	1	1	1	1	1
171470	1	0,004835751	1	0,003835251	0,003835251
174997	0,769051192	0,057362014	0,737868934	0,066366517	0,057362014
175032	0,220943805	0,386026347	0,323495081	0,675504419	0,220943805
181157	0,097215274	0,456228114	0,413540103	0,725362681	0,097215274
182191	1	0,668000667	1	0,581457395	0,581457395

Overall Ranking

User ID	
23557	Free Hotel Night
94883	15% discount next round flight
101486	Free Hotel Dinner
101961	Free Hotel Night
106907	15% discount next round flight
118043	Free Hotel Dinner
120851	Free checked bag
125845	Free checked bag
133058	Free checked bag
149058	Free Hotel Night
152583	Free checked bag
153982	Free Hotel Dinner
160754	Free Hotel Dinner
164522	Free Hotel Dinner
167852	10% discount first trip



Potencial new customers in relation with total # of users

Conclusion

This project successfully leveraged SQL for the development of key performance metrics and Tableau for the visual segmentation of clients based on behavioral patterns. Through this combined methodology, five distinct customer segments were identified: *Potential New Customer* (offered a 10% discount on their first trip), *Heavy Luggage Customer* (eligible for a free checked bag), *Long-Stay Customer* (awarded a free hotel night), *Flyer Customer* (granted a 15% discount on their next round trip), and *Luxury Dinner Customer* (offered a complimentary hotel dinner). These targeted perks aim to enhance customer engagement, encourage loyalty, and personalize the travel experience based on data-driven insights.