# Customer Segmentation

Experiment: Customer segmentation based on shopping behavior

Author: Neha Varshney | Date:October 2023

Abstract

Customer Segmentation is the process of dividing the customers into groups based on common characteristics in the context of a very specific business initiative.

Purchase behaviour-based segmentation means to distinguish how different customers behave, with social proof, taking into account the trends and processes of how they ultimately purchase.

It can also help us to understand: How different customers decide and approach the purchase decision.

INTRODUCTION

The idea behind this project is to deal with segmentation based on shopping behavior. This form of behavioral segmentation provides insight into the buying stage that your customer might be in, their role in the purchasing process, the obstacles they are facing, the incentives they are most likely to respond to and much more.

From a business perspective, customer segmentation divides a customer database into groups to provide a tailored customer experience - rather than one size fits all. The business value of creating such groups is the ability to implement specific business strategies, including:

* Customized promotions (e.g. special offers, discounts)
* Personalized communications (e.g. targeted marketing emails)
* Specific account policies (e.g. credit line)

The main objective of this project is to retain and add value to existing customers with a marketing strategy built on a solid understanding of customer behavior.

The project featured three stages:

1. Exploring the data using SQL and form a plan for further analysis.
2. Make calculations related to the business context and then segment the customer behavior data with statistical and visual techniques.
3. Present the customer segmentation via slides.

METHODOLOGY

The methods used to deal with segmentation based on shopping behavior according to the three stages of project are following:

* SQL(Structured Query Language) - It is used for data exploration. With the TravelTide dataset, it was time to explore the data provided. It is also used for calculating behavioral metrics that map to TravelTide rewards program perks.
* Microsoft Excel (MS Excel) - It is used for calculating the behavioral metrics. Using Excel, I calculated the average and mode of the metrics.
* Tableau - It is used for segmenting the customers and visualizing the customer segmentation results.
* Microsoft Powerpoint - Powerpoint is used for presenting the customer segmentation Results.

KEY FINDINGS

Metrics can be set to track the performance of businesses from various perspectives. Many of the metrics tracked the interaction between business strategy and customer behavior at an aggregate level.

We defined and calculated the behavioral metrics at the level of individual customers.

The observations made using behavioral metrics in SQL and excel are for all the users whose session started on and after 2023-01-04 and have more than 7 page clicks are as follows:

1. There were 5998 customers who booked the flights under the discount.
2. There were 5998 customers who booked the hotels under the discount.
3. The amount of flight discount on average for customers is between 0.05 - 0.55 after rounding.
4. The amount of hotel discount on average for customers is between 0.05 - 0.45 after rounding.
5. The conversion rate of customers is 38.22
6. The amount of dollars saved per kilometer on average traveled is 0.046132085
7. The most frequent average number of page clicks per customer is 0.10952381
8. The most frequent average session time per customer is 0.31040418

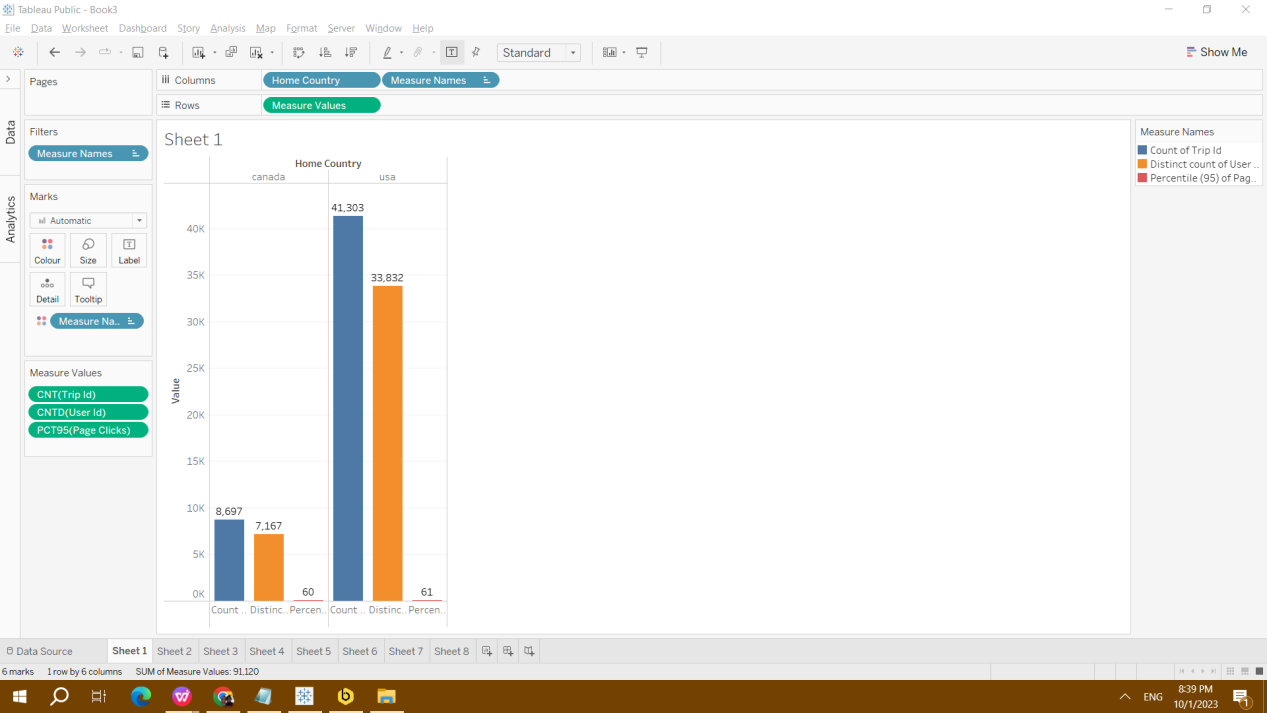
EXPLORATORY DATA ANALYSIS

Once the exploration of data is being done, it’s time to find the nuances of customer behavior and intent, segment customers searches by trips, that is, the process of dividing trips into groups based on common characteristics.

There are important factors to characterize the “trips”. Some of those are as follows:

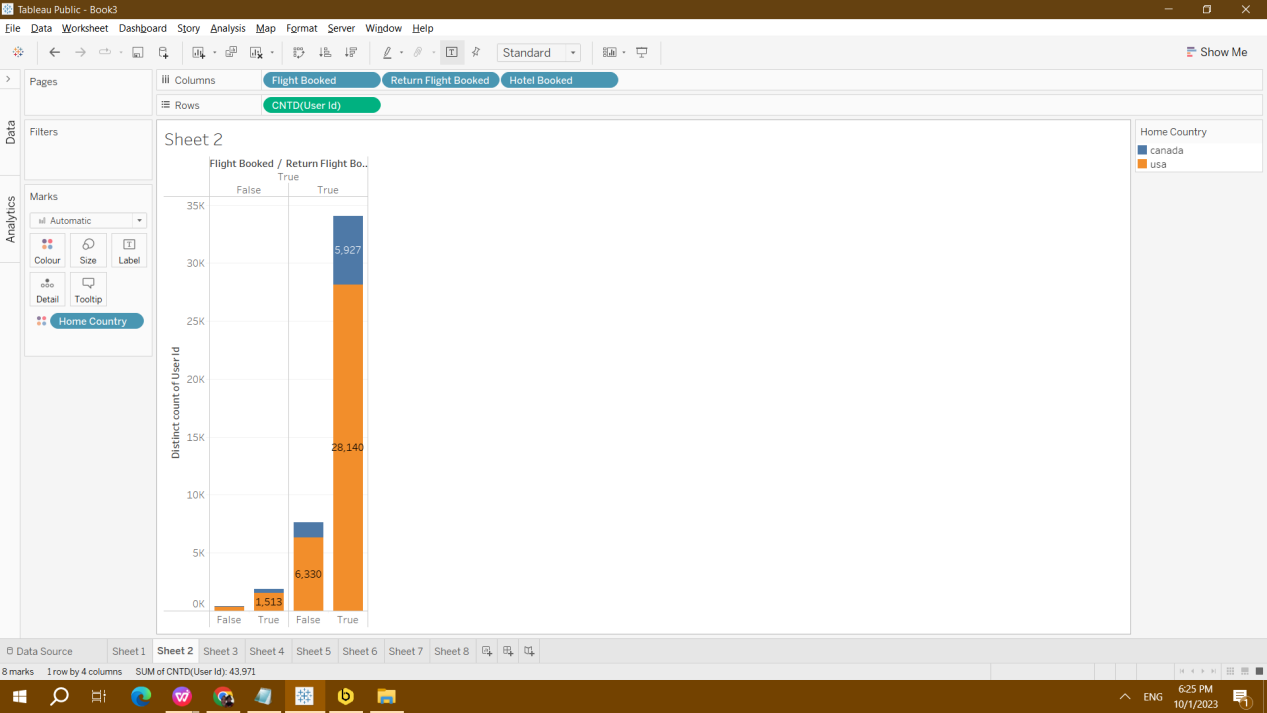
Most of the trips were searched by the customers whose home country were United States and Canada.

About 41,303 trips were from United States with 61% of Percentile(95) of page clicks and about 8,697 trips were from Canade with 60% of Percentile(95) of page clicks.

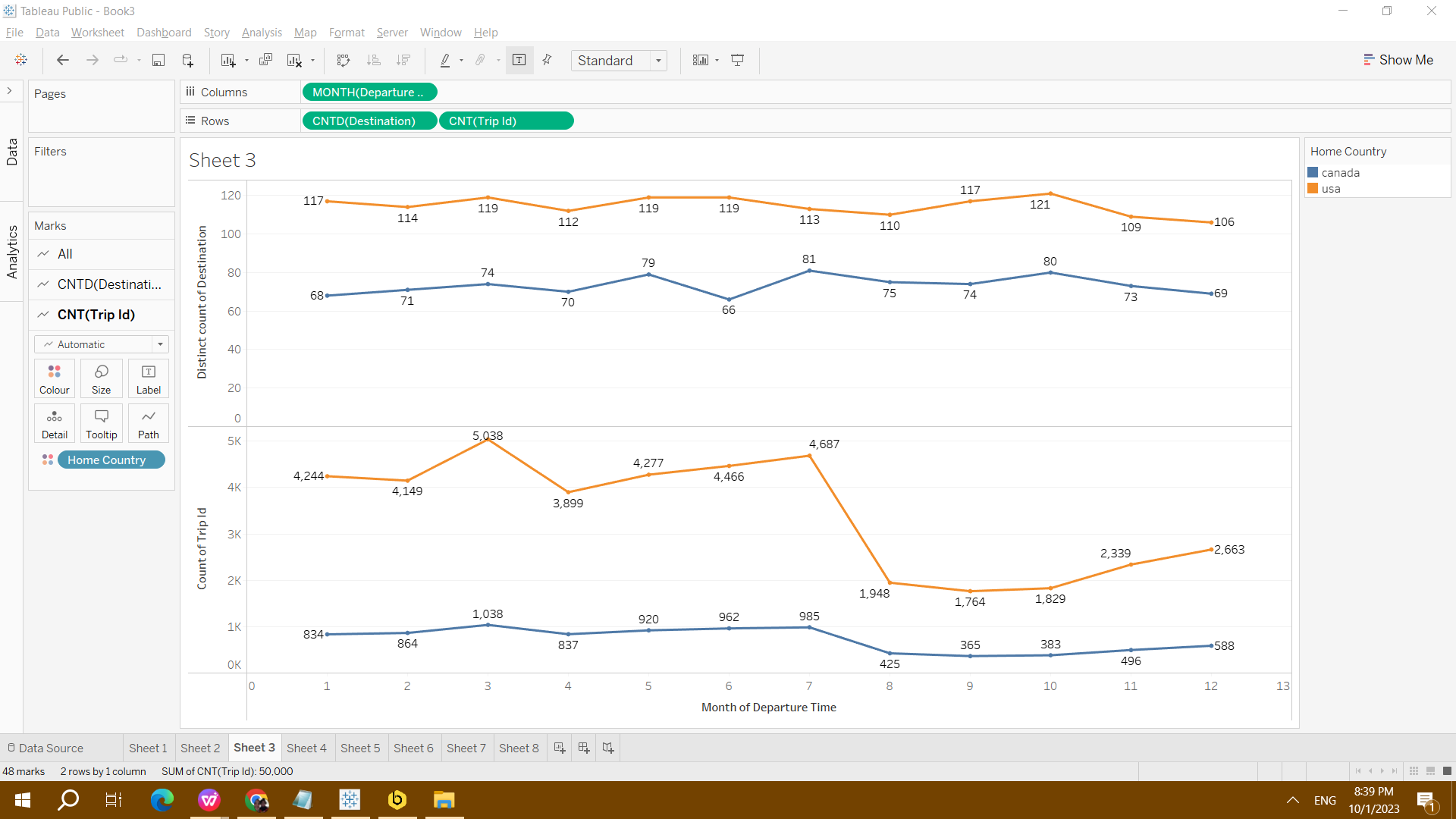


I wanted to see how many customers have booked one-way trips and round about trips from the United States and Canada.

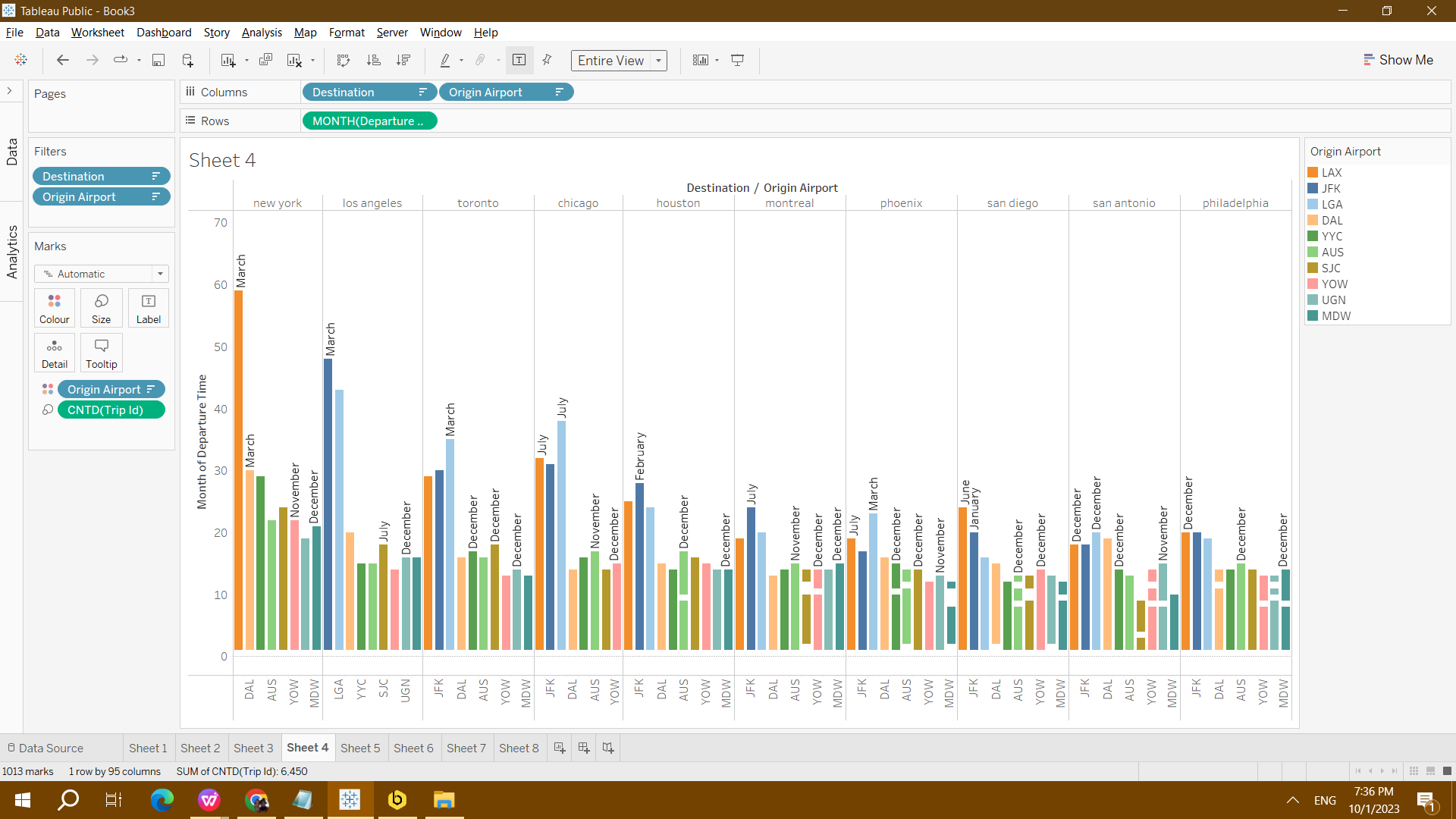
28,140 customers from United States and 5,927 customers from Canada booked round about trips with hotel bookings. 1,513 customers from United States and 362 customers from Canada booked one way trips without the return flights being booked.



Next, I wanted to see how many trips were taken to the destination during which month of the year. There are 140 destinations. Most trips from USA and Canada were in March and July.



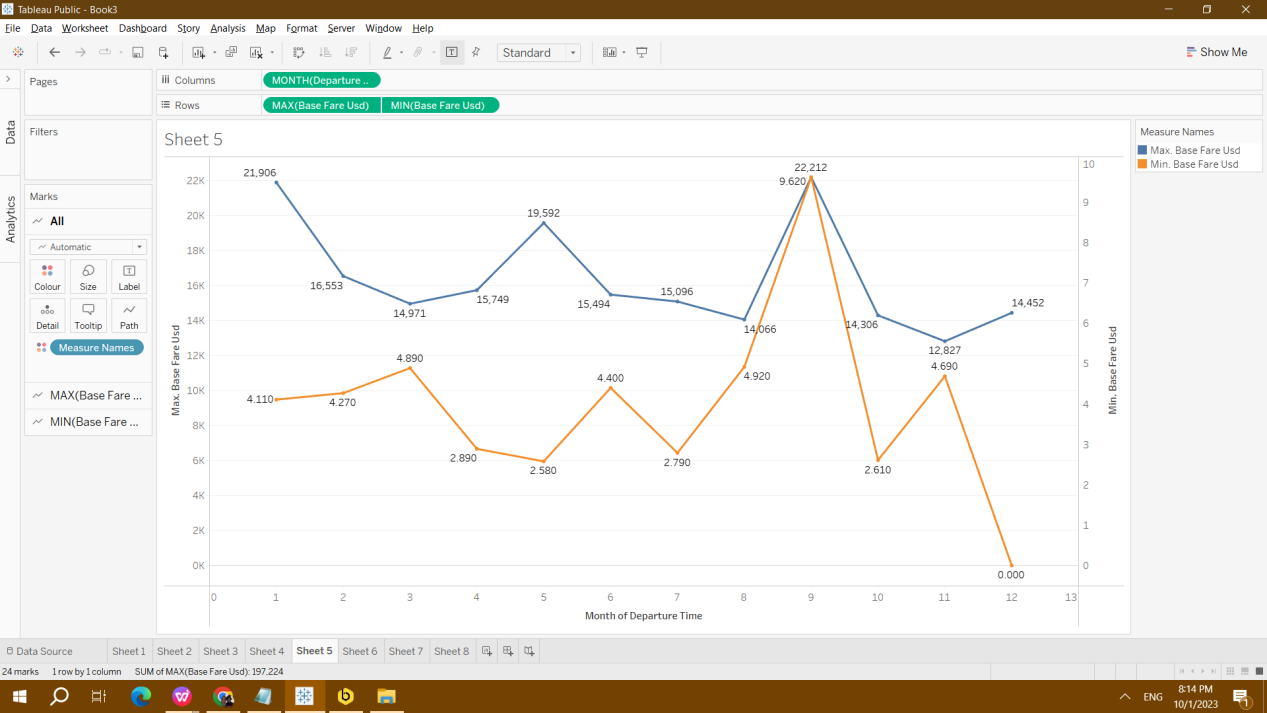
The data set is largely dominated by trip destination within the United States.



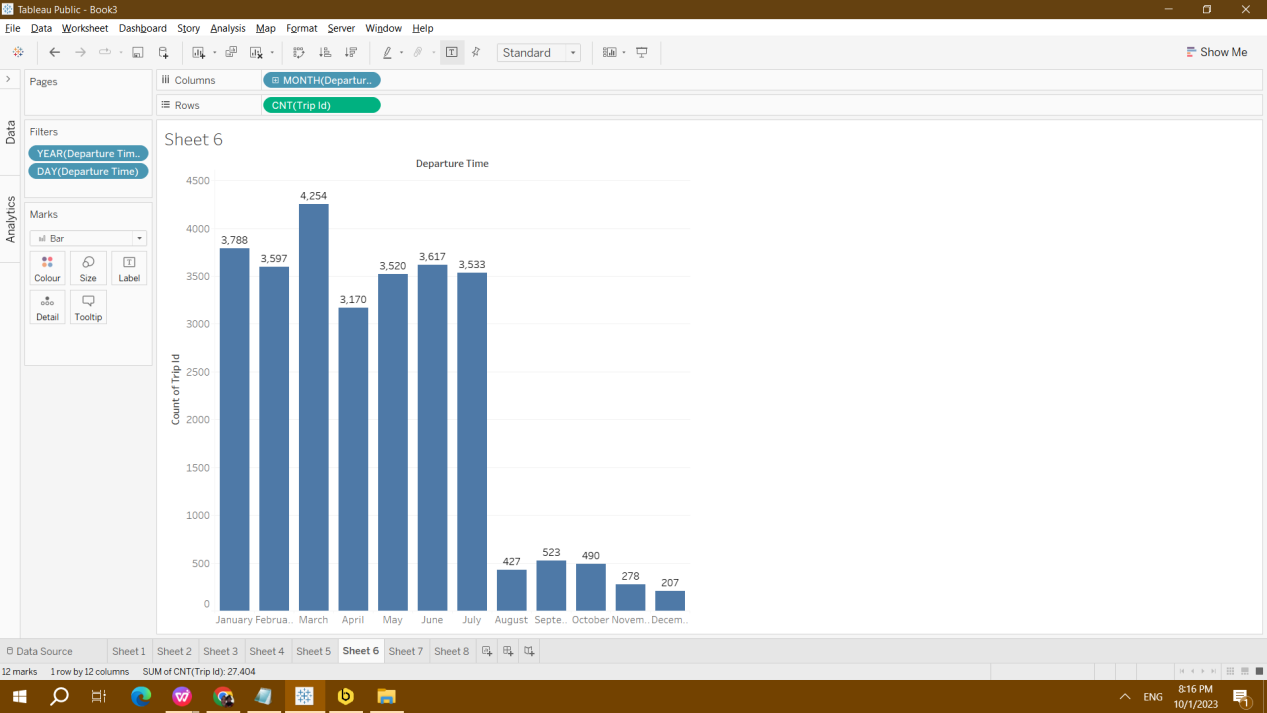
8 of these top 10 searched routes were travelling within the United States.

The Base Fare Usd are high in the month of January, May and September. The Base Fare Usd is highest in September.

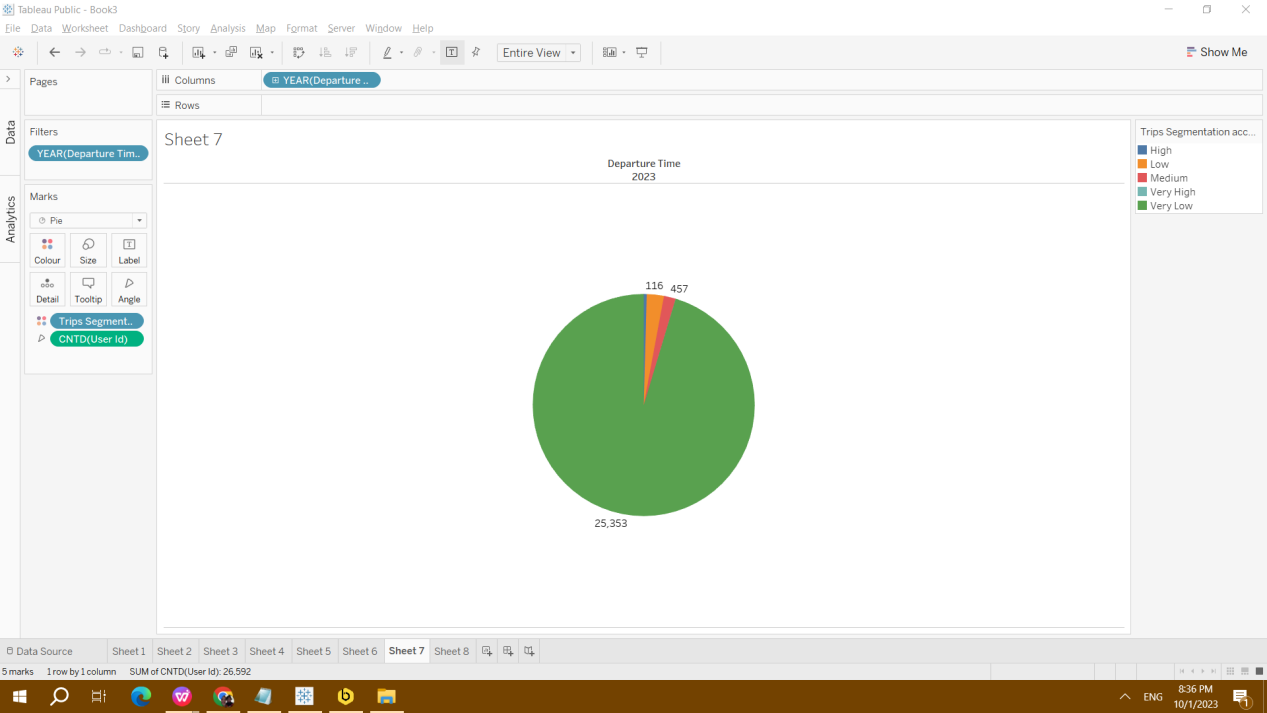
The Base Fare Usd is low in the month of April, May, July and October.



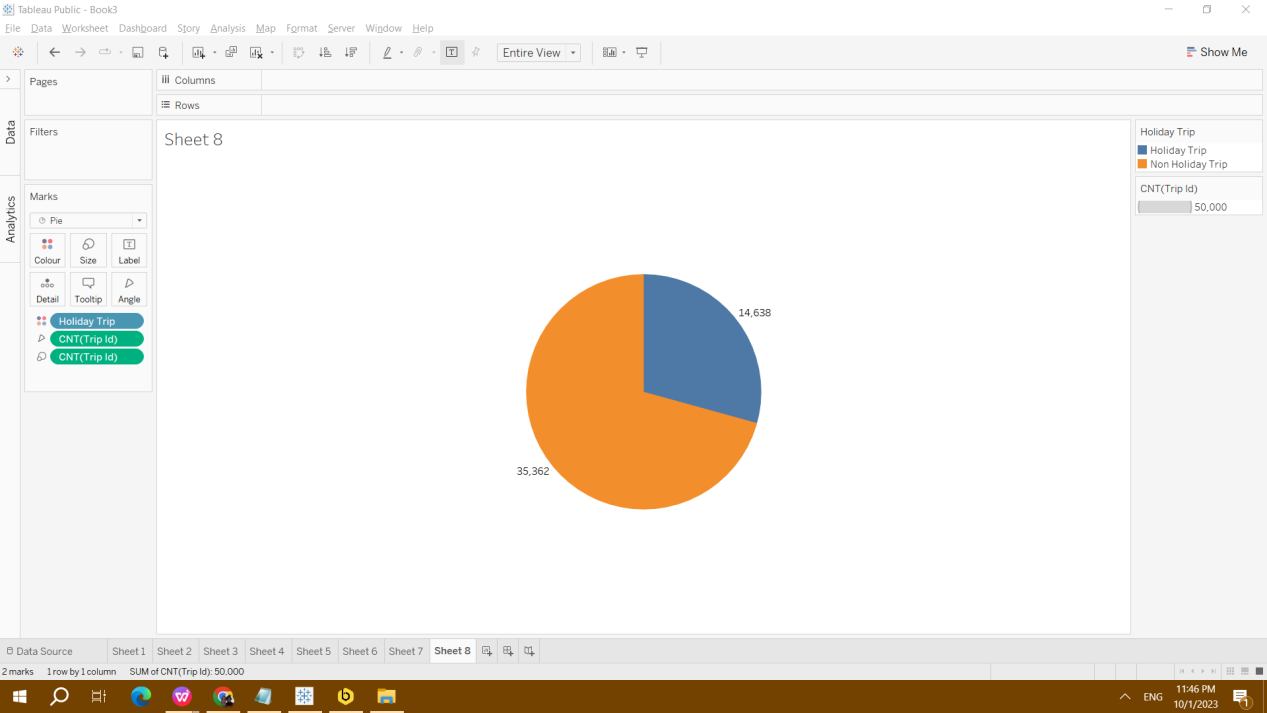
Most of the trips are planned in the month of January, March and during the summer months of May, June and July.



I also segmented trips according to price for the year 2023. There are 25,353 customers that are placed in ‘Very Low’ category. There are 116 customers that are placed in ‘High’ category.



I segmented trips according to Holidays. Most of the trips are booked in non holiday season.



35,362 customers booked trips in non holiday season while 14,638 customers booked in the holiday season.

RECOMMENDATIONS

The business will be able to provide an optimal experience for its customers based on their intent by using above knowledge to create more effective content or recommendations.

My conclusion is based on the following:

There are 52 weeks in an year.

Total days for holiday period are 52\*2=104

104+3=107

On Average I.e. (14,638/107) = 136.803 customers are travelling in the holiday period

For non-holiday season, there are 365-107=258

On Average i.e. (35362/258) = 137.062 customers are travelling in the non-holiday period.

So, the customers are not booking on the basis of base fare. If the customers get decent base fare in holiday period as well, we can retain the customers.

APPENDIX

SQL Queries

Query to get user-level analysis dataset

select users.user\_id,sessions.trip\_id,users.home\_airport, flights.origin\_airport,flights.destination\_airport,users.home\_country,flights.destination,

users.home\_airport\_lat,users.home\_airport\_lon,flights.destination\_airport\_lat,flights.destination\_airport\_lon, sessions.session\_start,

sessions.session\_end,sessions.page\_clicks,sessions.flight\_discount,sessions.hotel\_discount,sessions.flight\_booked,sessions.hotel\_booked,

flights.return\_flight\_booked,flights.base\_fare\_usd,flights.destination,flights.departure\_time, flights.trip\_airline,flights.departure\_time

from users

join sessions

on users.user\_id = sessions.user\_id

join flights

on sessions.trip\_id = flights.trip\_id

Imported into excel for statistical calculations

Imported into Tableau for visualizations.

Analysis Files

Excel for statistical calculations

Tableau for visualizations

[CustomerSegmentation | Tableau Public](https://public.tableau.com/app/profile/neha.varshney/viz/CustomerSegmentation_16962343222130/Dashboard1?publish=yes)

Loom for presentation

<https://www.loom.com/share/33b253ad3cee45dd979060f3c919f6a2?sid=b61e3062-0829-4f30-9ec4-0d504d6c139e>