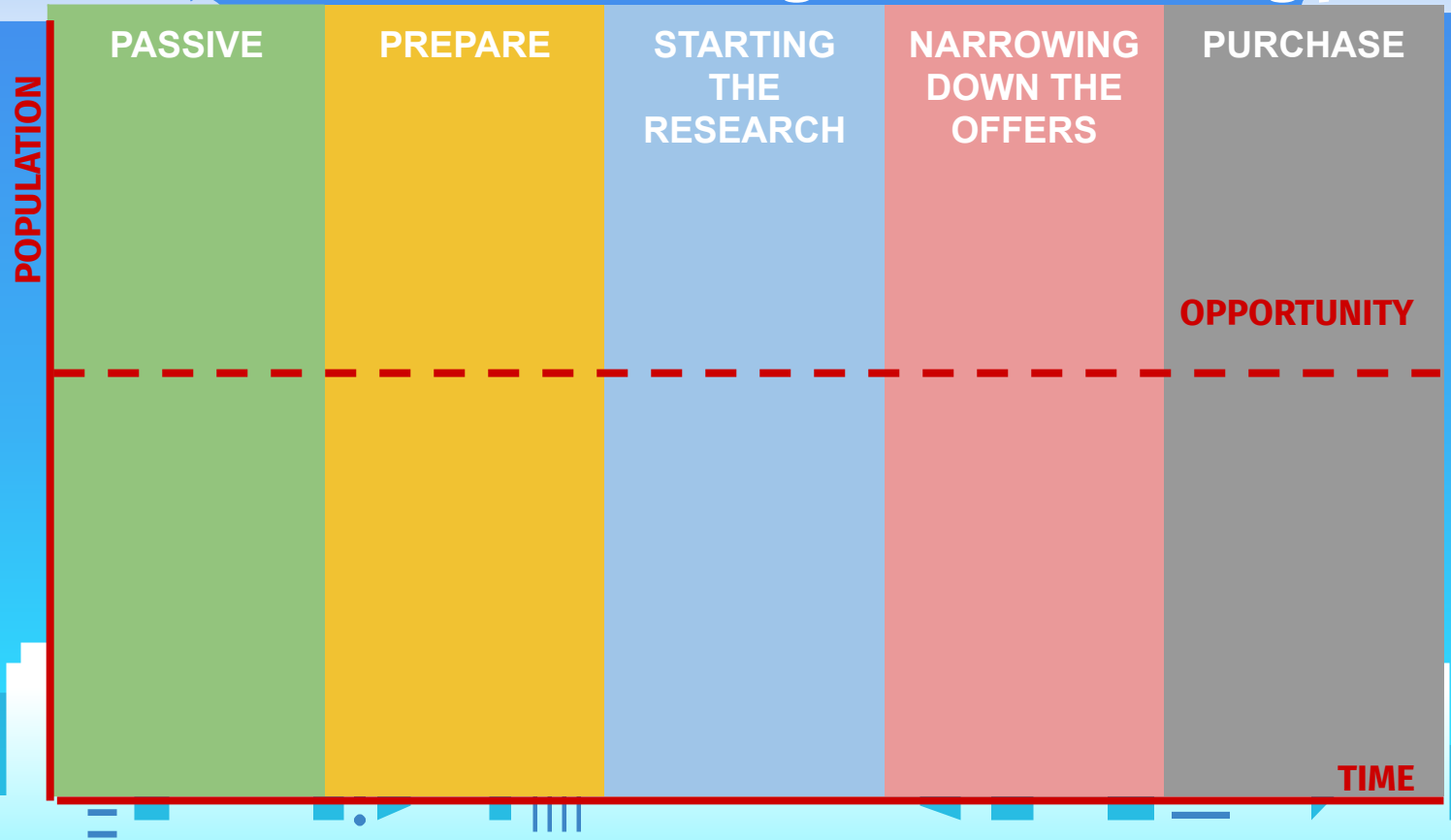
Understanding Consumer Behavior: Expedia data

STATS 140SL S2019: Annie Choi, Albert Na, Breanna Qin, Chris Wan, Kathy Fu, Kristee Song, Tiffaney Pi

A Brief Overview

- **What is Expedia?**
  - Expedia is a global travel tech company that specializes in helping someone from start to finish manage their travel plans. They provide resources for airplane tickets, hotel bookings, car rentals, and different bundles associated with traveling.
- **Datasets:**
  - Data: Information on hotel bookings
    - Dimensions: **10,884,539** observations **x 27** variables
  - Dest: Information on destinations
    - Dimensions: **36,407** observations x **144** variables
- **Our Goal:**
  - Understand **consumer tendencies** and how Expedia can optimize this information to better implement **advertisements** and package deals.

Consumers at different stages in the booking process



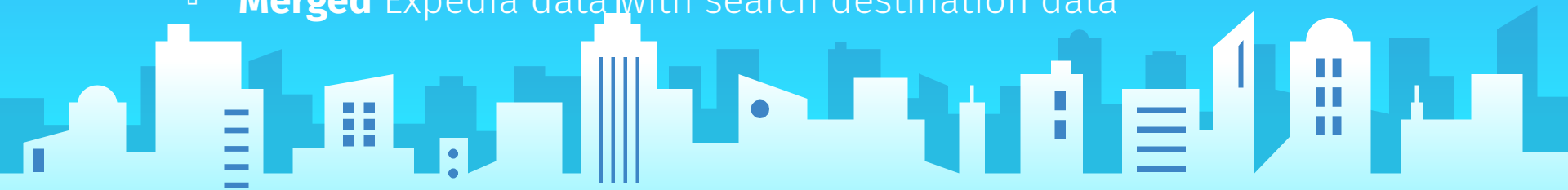
Research Question:

How can we better understand American consumer habits on Expedia and what insights can we draw from this? (to boost advertising)

Data Preparation

We are only concerned with users who have successfully booked at least one hotel with Expedia. We specifically are looking at users booking from the US and have travel plans to somewhere in the US.

- Filter out any users who never booked a hotel with Expedia
- Filter out international hotels
- Only keep unique city destinations for each individual user
- Remove rows with NA values in the check-in/check-out date
- Only look at successful bookings
- **Merged** Expedia data with search destination data

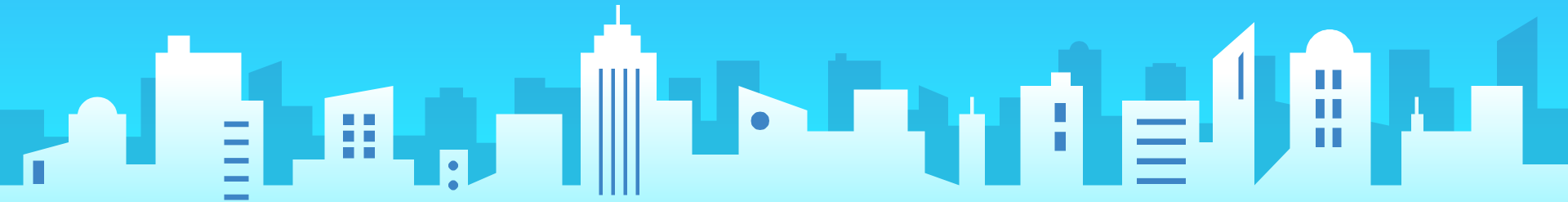


Creating new variables

`Num_hotels:` Number of hotel searches the consumer looked at before booking

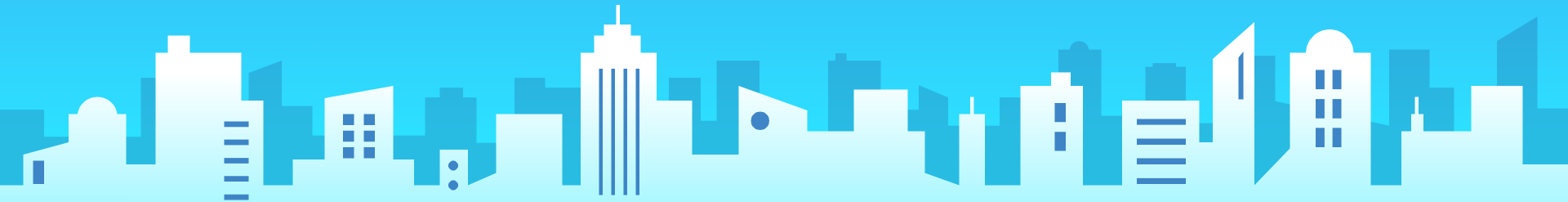
`Time_spent:` Time (in days) between search date and booking date

`Time_advance:` Time (in days) between booking date and check-in date

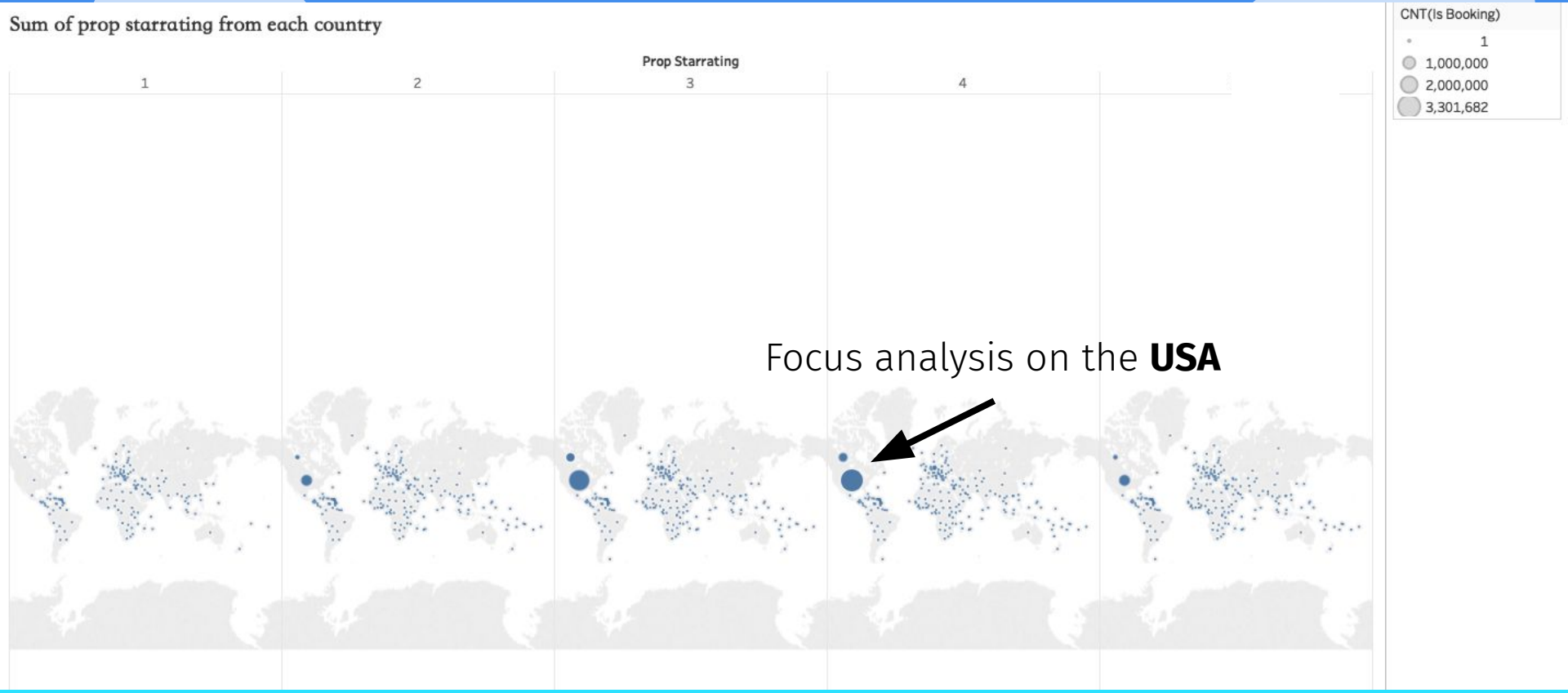


Overview:

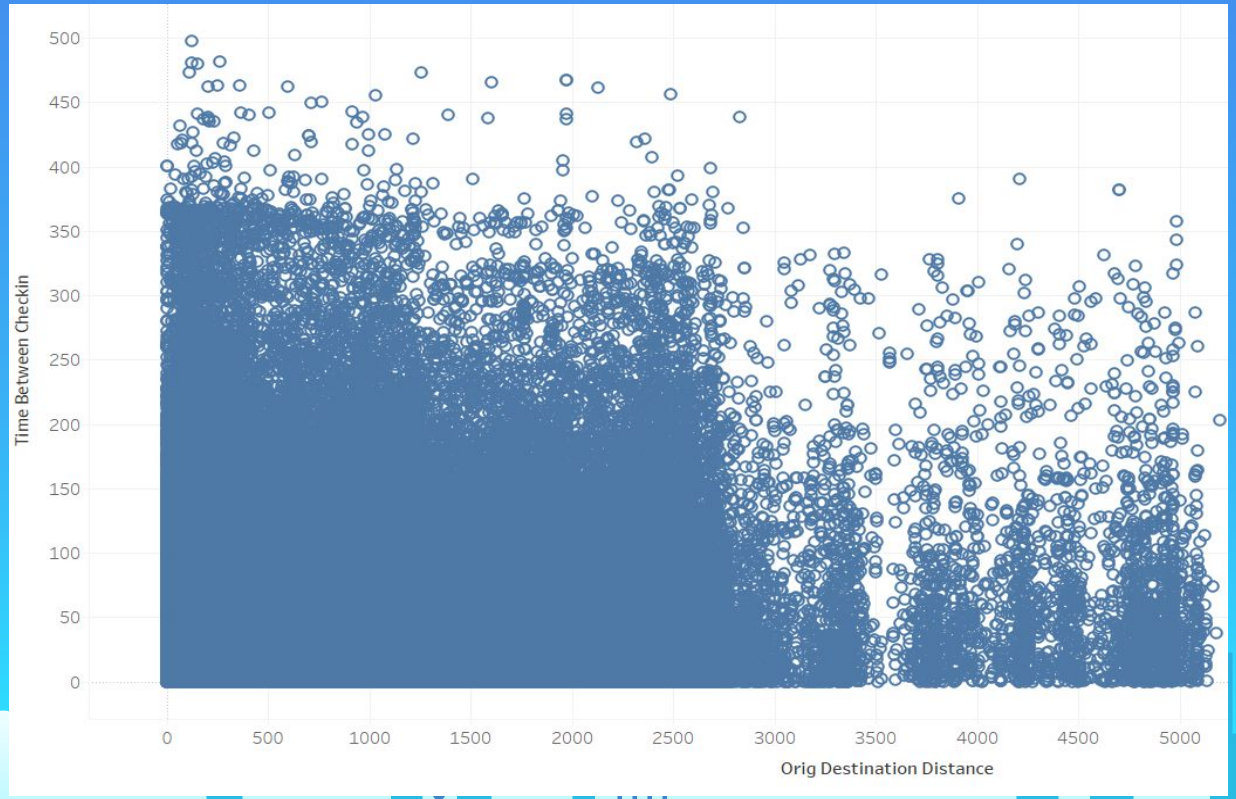
1. Exploratory Analysis
2. Mobile Users & Package Bookings
3. Booking time between first search & booking date
4. Booking time between booking date & check in date
5. US Consumer Locations vs Destinations [maps]
6. Booking Channels (i.e.: Trip Advisor, Google search, Trivago, etc.)



EDA: Where to start?

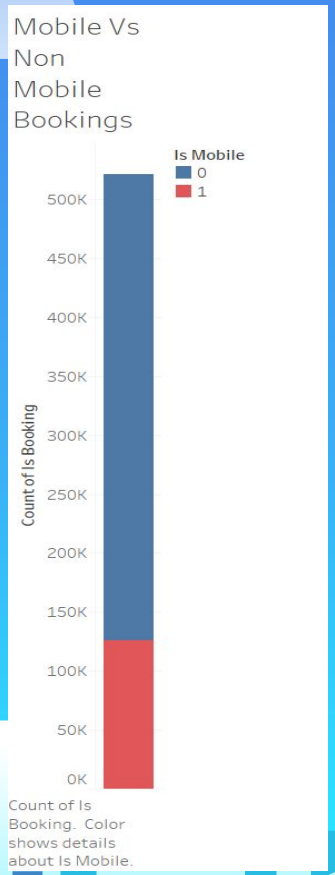


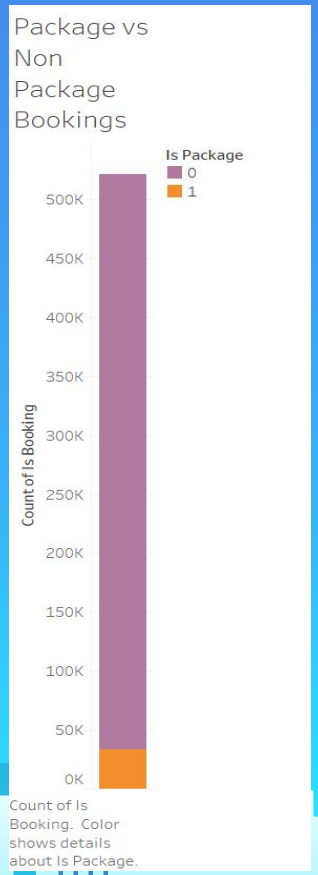
EDA: Distance from destination and early booking



Users booking hotels at further destinations tend to book closer to check in date

EDA: Price & booking across the year





- Moderate mobile user base - beneficial to update website and enhance mobile friendly web experience to increase bookings
- Small proportion of bookings come from packages - less focus on selling in package

Types of **Mobile** Booking Users

```
heavyresearchers <- successful[successful$num_searches>8,]
nonresearchers <- successful[successful$num_searches==1,]
```

```
Heavy researchers non-mobile versus mobile bookings
         0         1
0.7572592 0.2427408
Non-researchers non-mobile versus mobile bookings
         0         1
0.866519 0.133481
```

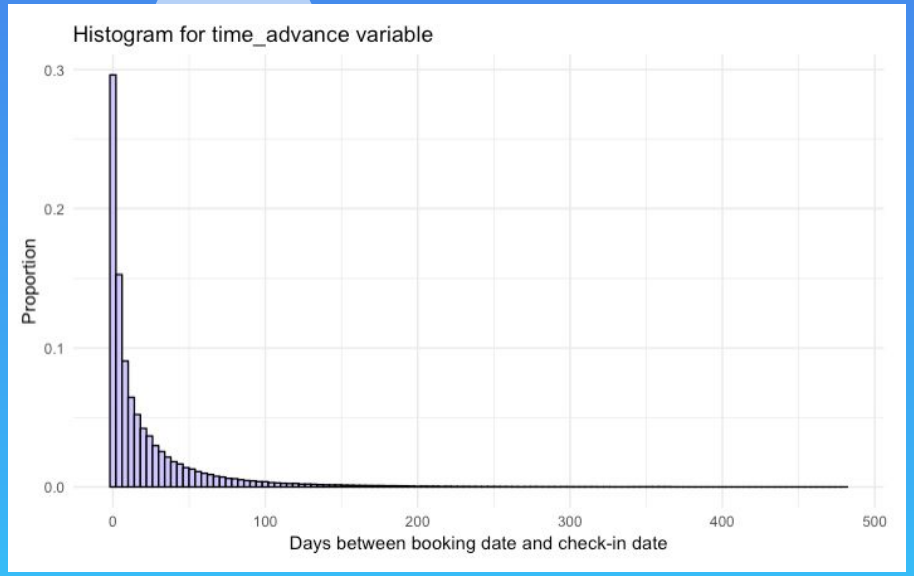
```
earlybirds <- successful[successful$time_advance > 14,]
lastminute <- successful[successful$time_advance==0,]
```

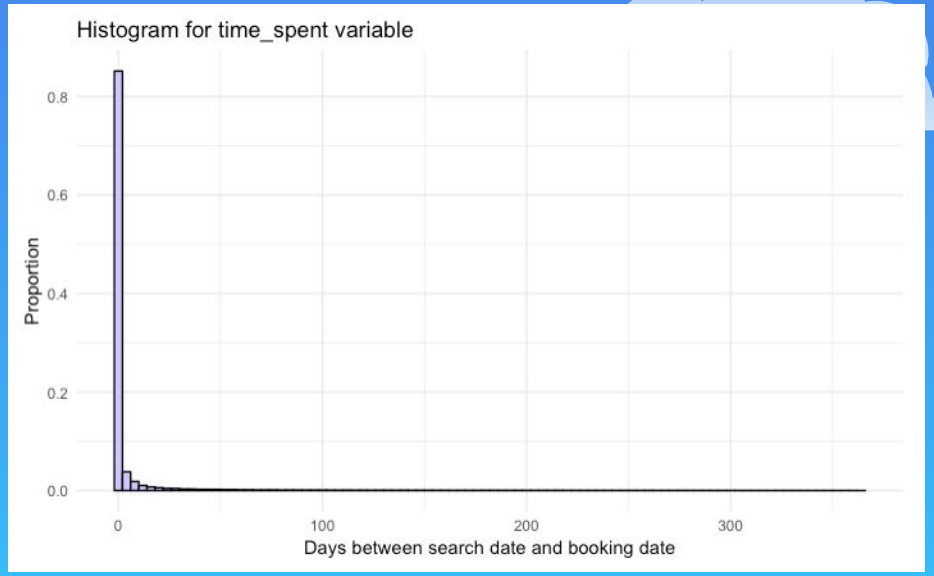
```
Early bookers non-mobile versus mobile bookings
         0         1
0.830649 0.169351
Same-day bookers non-mobile versus mobile bookings
         0         1
0.4257627 0.5742373
```

Types of Users booking **package deals**

```
Early bookers non-packaged versus packaged bookings
         0         1
0.8891085 0.1108915
Same-day bookers non-packaged versus packaged bookings
          0           1
0.994416696 0.005583304
Heavy researchers non-packaged versus packaged bookings
          0          1
0.91697237 0.08302763
Non-researchers non-packaged versus packaged bookings
         0         1
0.6602049 0.3397951
```

num_hotels		time_spent		time_advance	
Min.	1.000	Min.	0.0000	Min.	0.000
1st Qu.	2.000	1st Qu.	0.0035	1st Qu.	1.186
Median	3.000	Median	0.0083	Median	8.089
Mean	4.639	Mean	8.0658	Mean	23.455
3rd Qu.	5.000	3rd Qu.	0.0528	3rd Qu.	27.771
Max.	270.000	Max.	364.1493	Max.	481.452

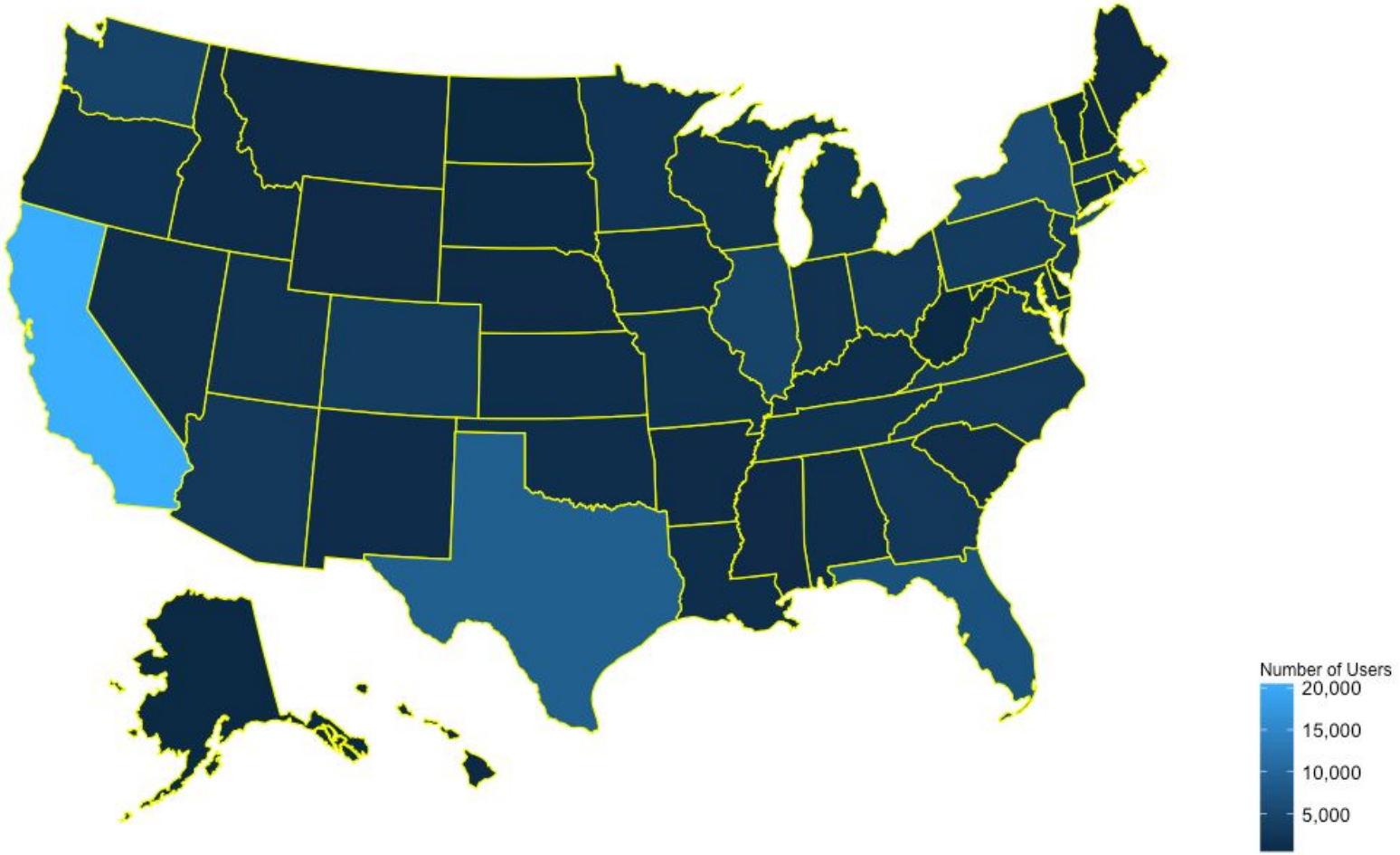




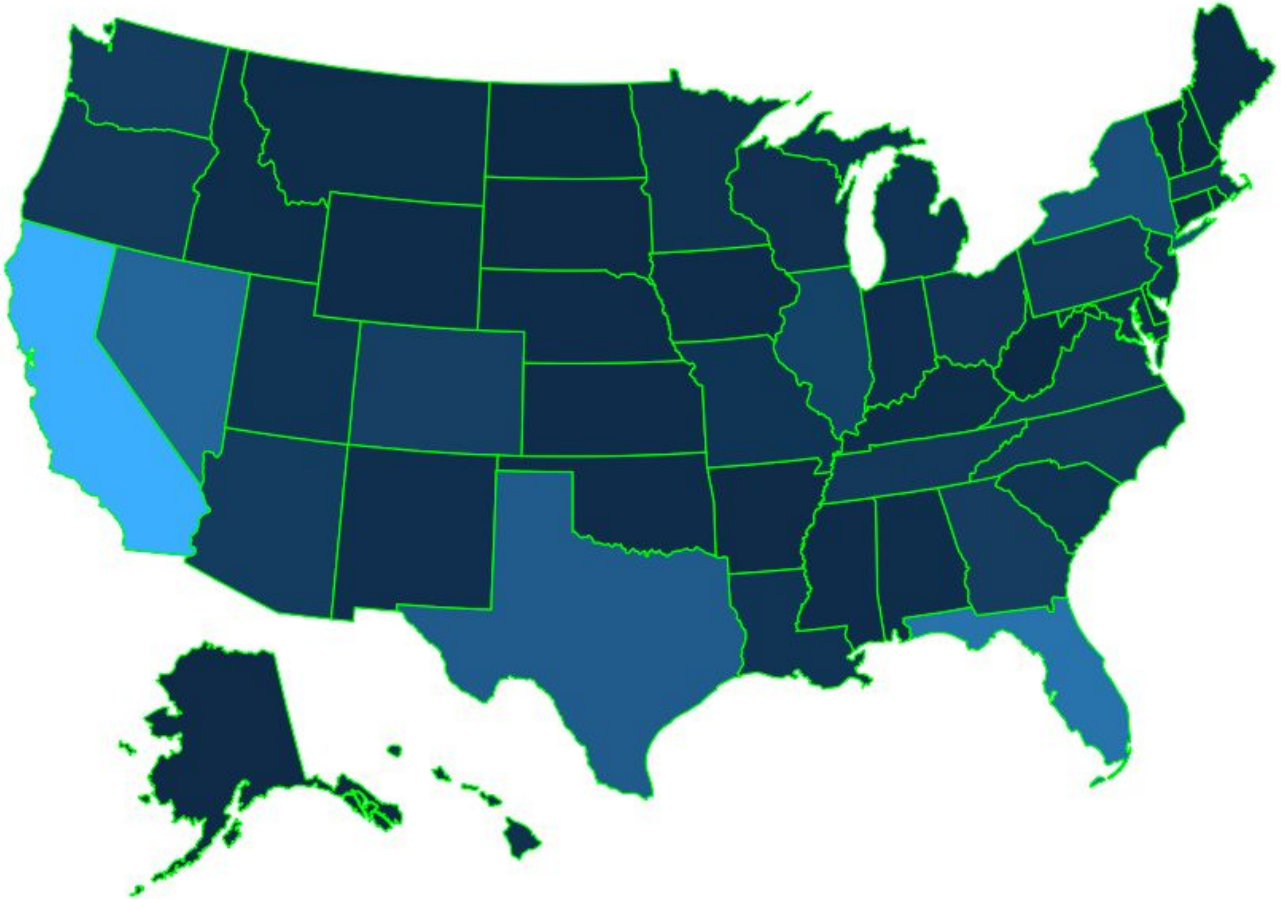
Based on the summary statistics, the data for **time_spent** and **time_advance** is highly right skewed.

We see that with most successful bookings, users start considering to book only days in advance to their hotel stay

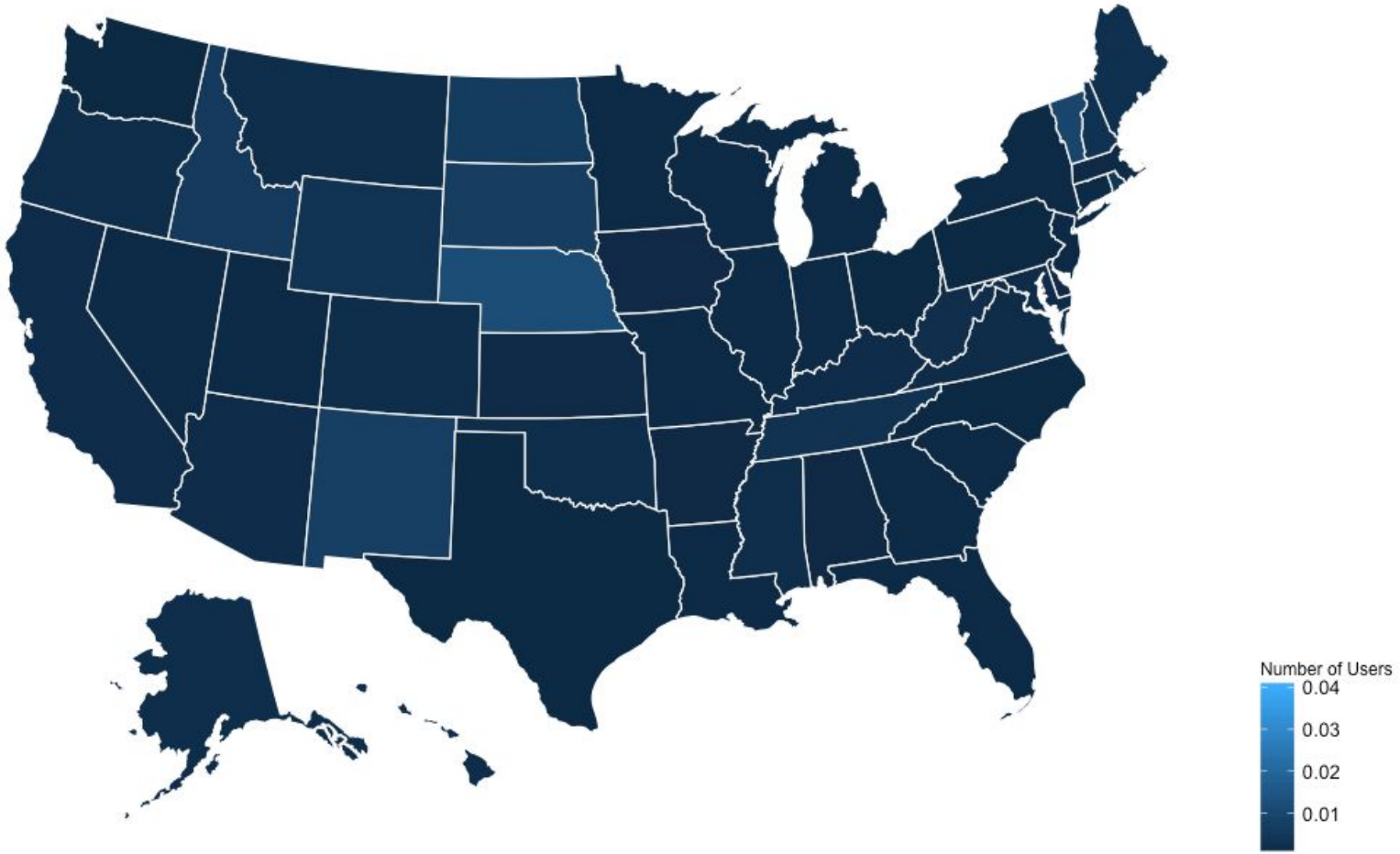
Expedia User Locations in the United States



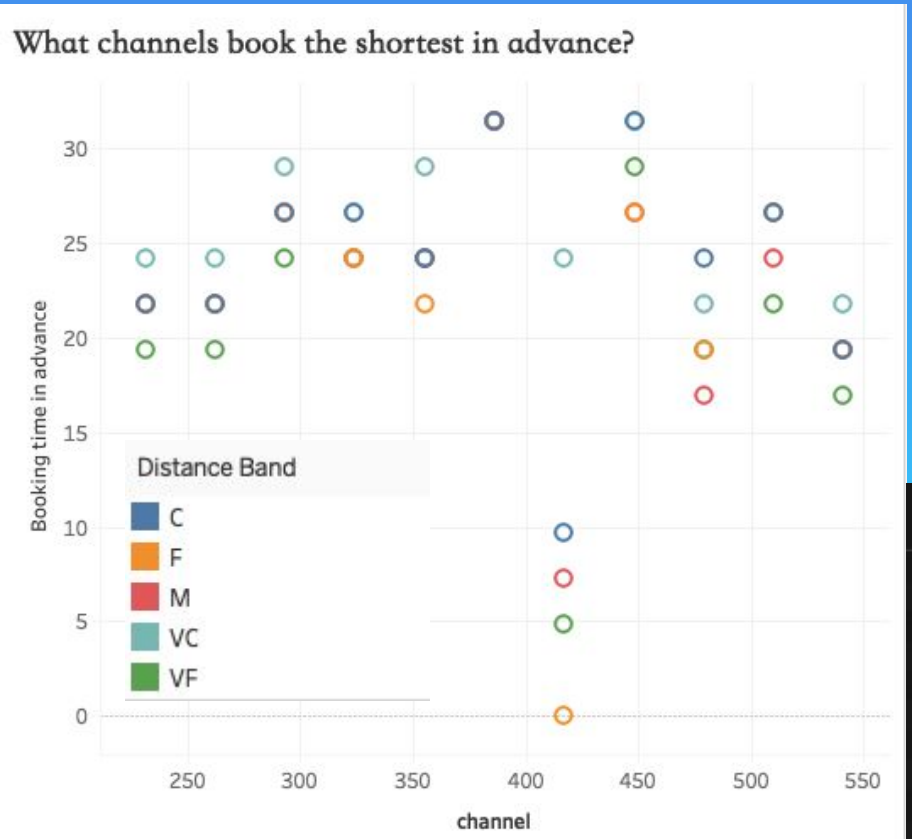
Expedia User Travel Destinations in the United States



Expedia User Travel Destinations in the United States as Proportions of Population



Booking Channels

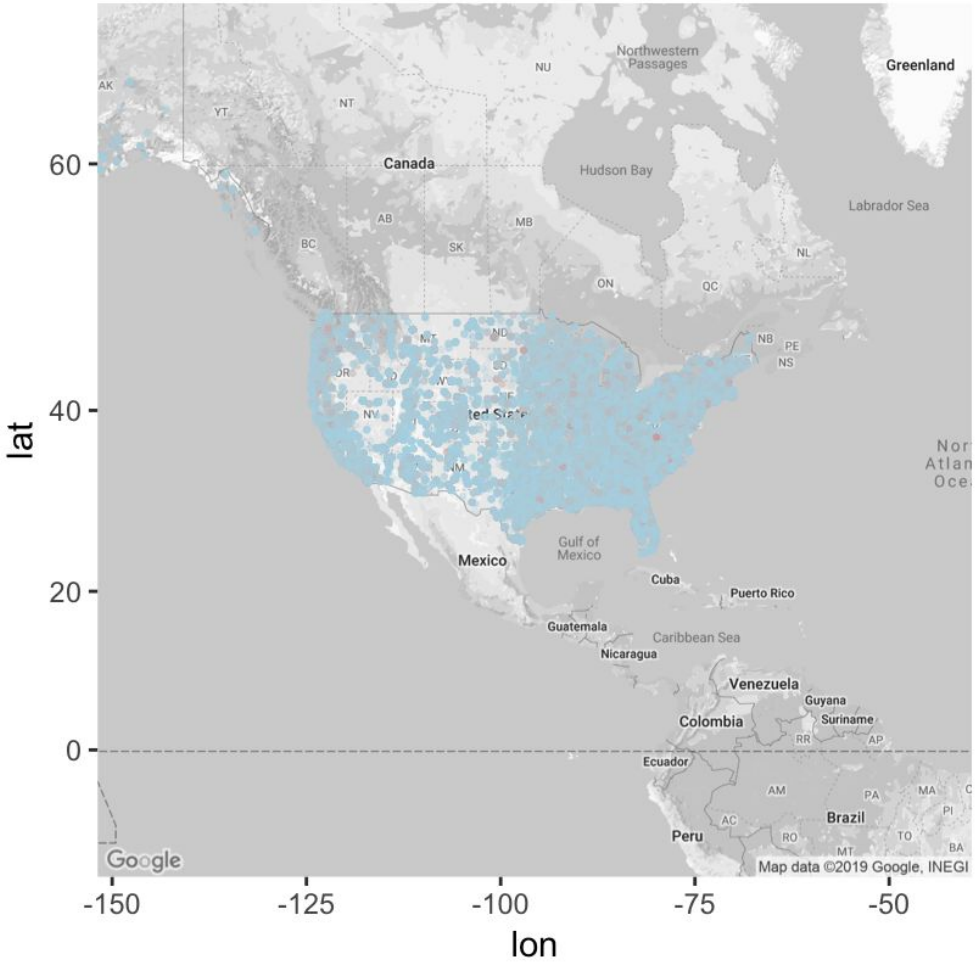


- Similarly, channel **417** indicates a significantly **shorter** mean of booking times in advance of their departure date than in all other channels. The color legend reinforces the fact that the farthest and very farthest distances from the user to the destination yield an **inverse relationship** to booking time in advance.

channel <int>	distance_band <chr>	mean(time_advance) <dbl>
231	C	23.014137
231	F	22.128391
231	M	22.286832
231	VC	25.303353
231	VF	20.569482
262	C	23.709100
262	F	22.378506
262	M	22.316923
262	VC	25.025475
262	VF	20.758868

channel <int>	distance_band <chr>	mean(time_advance) <dbl>
417	C	10.958763
417	F	2.198422
417	M	7.808185
417	VC	24.760880
417	VF	6.778869
448	C	31.878614
448	F	27.598924
448	M	28.204929
448	VC	33.570449
448	VF	29.774518

Average Time Booked In Advance



as.numeric(time_advance)

Time Advance

300

200

100

0

Conclusion

Conclusion

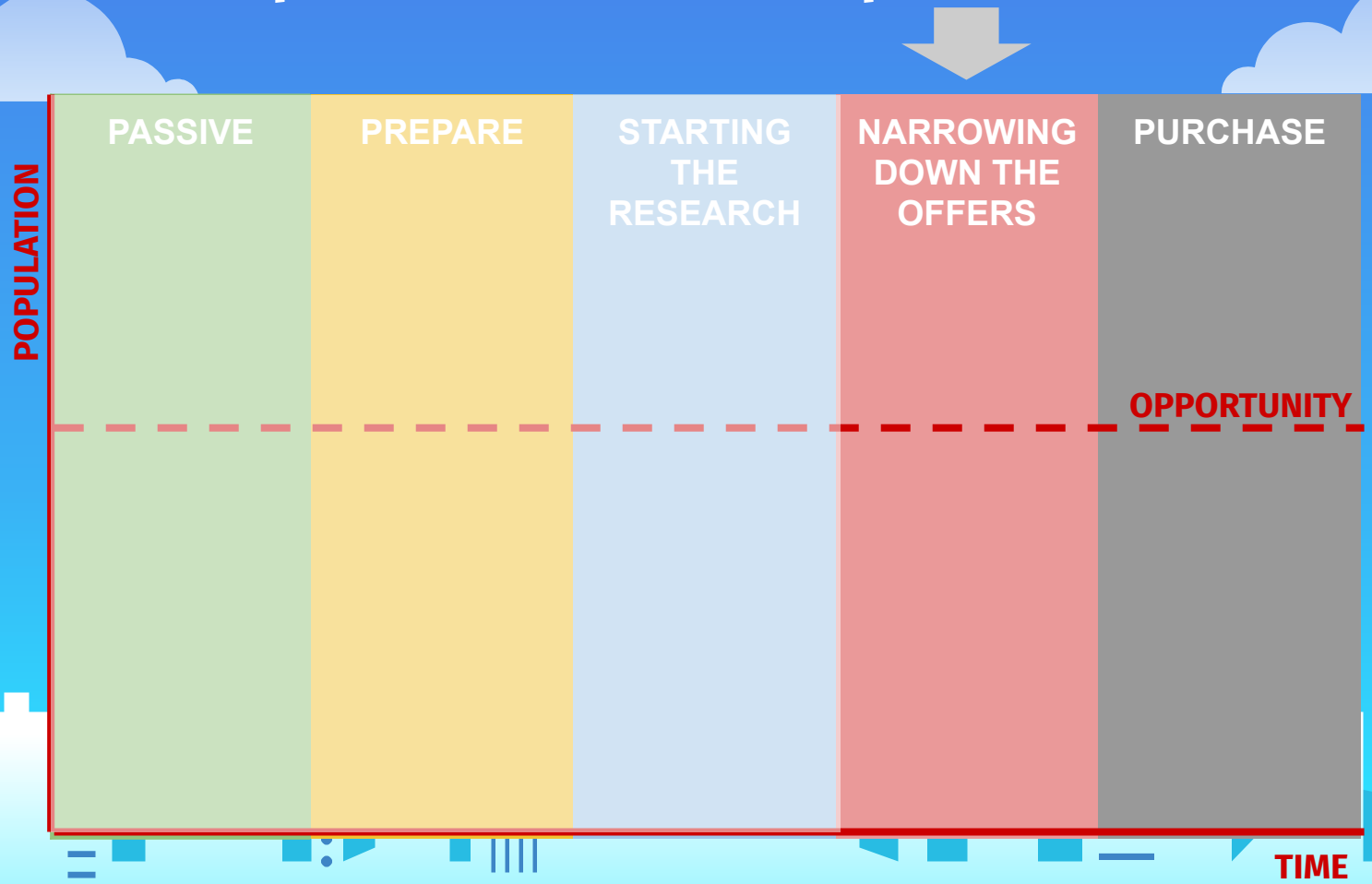
US users traveling domestically who are further from their destination point book closer in advance to their trip. This is odd and could mean one of the following things:

- People use Expedia for spontaneous trips
- People who procrastinated and forgot to book in advance use Expedia because it gives the most last minute options and/or best last minute deals.
- Business travelers who have to travel with short notices use Expedia

distance_band <chr>	mean(time_advance) <dbl>
Very Close	25.82114
Close	23.90776
Medium	22.91191
Far	22.67991
Very Far	21.17142

distance_band <chr>	median(time_advance) <dbl>
Very Close	9.602083
Close	8.417014
Medium	7.550000
Far	7.371528
Very Far	6.516319

Expedia users enter the purchase funnel



Areas of Interest:

→ Improve **mobile** user experience for those who use Expedia to browse through hotels and those who are making last-minute purchases

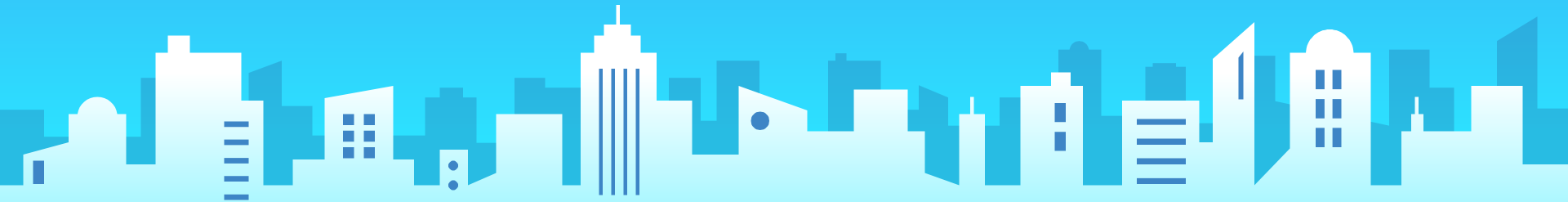
→ Seasonal **summer** ads

→ Target CA, Texas, Florida users

→ Advertise various CA cities/locations as a premier travel destinations in the USA.

Limitations and room for modifications

- With more time, we can perform predictions to see if consumers will book or not.
- Dataset had 10,000,000+ observations, which made the run-time really long.
- The channel IDs were kept private, so we didn't know specifically how people were arriving at the Expedia site.



Thank you!