

BAX421 Data Management Project

Yingxuan(Selina) Bian



Table of Contents

1. Database Introduction
2. Business Questions
3. Results & Insights



1. Database Introduction



Database Overview

Dataset

“Airbnb New User Bookings”

- New users on Airbnb can book a place to stay in 3400+ cities across 190+ countries.
- 5 data files originally in this dataset, which are *age_gender_bkts*, *countries*, *sessions*, *test_users*, *train_users*.

Tables

6 Tables in Total

- ▶  *action*
- ▶  *age_gender_bucket*
- ▶  *countries*
- ▶  *sessions*
- ▶  *users_behavior*
- ▶  *users_profile*

2. Business Questions



Question List

1. How many users signed up from each affiliate channel, and what is the percentage distribution?
2. What is the total number of users associated with each device type in 2010?
3. What are the Top 3 actions performed by users based on action frequency?
4. What are the Top 3 country destinations, and how many users belong to each of these destinations?
5. For each action type, what are their average session duration?
6. What is the gender distribution of users for each destination?
7. What are the average ages of users of various devices?
8. Which destination country has the highest population contribution (in thousands) for each gender and age bucket?

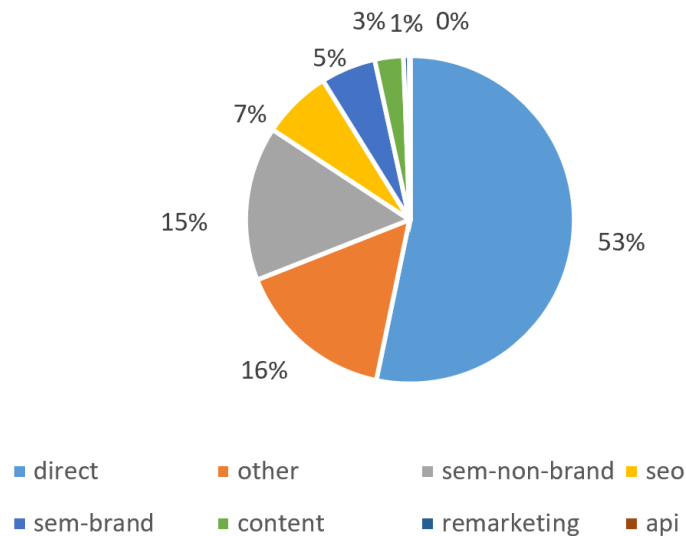
3. Results & Insights

Question 1

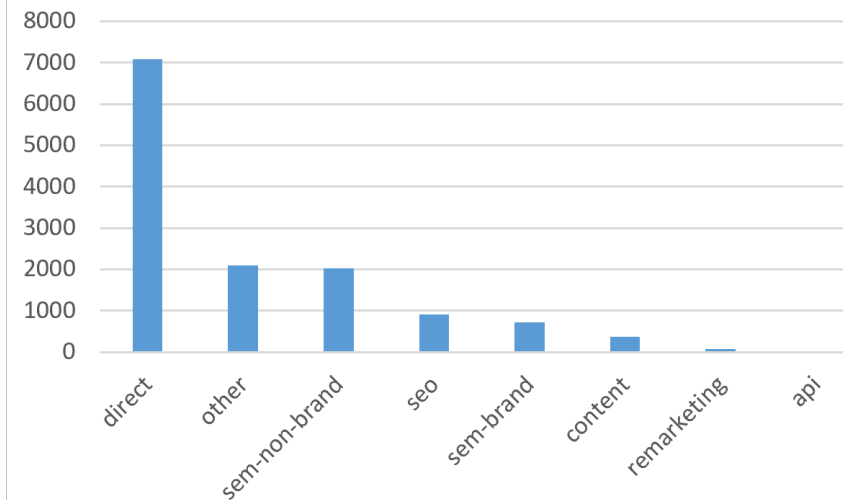
How many users signed up from each affiliate channel, and what is the percentage distribution?

- “Direct” is still the major channel that worth keep investing in (e.g. Ads, Marketing)
- “SEM” and “SEO” are emerging channels that worth paying attention

Percentage of Users for Each Affiliate Channel



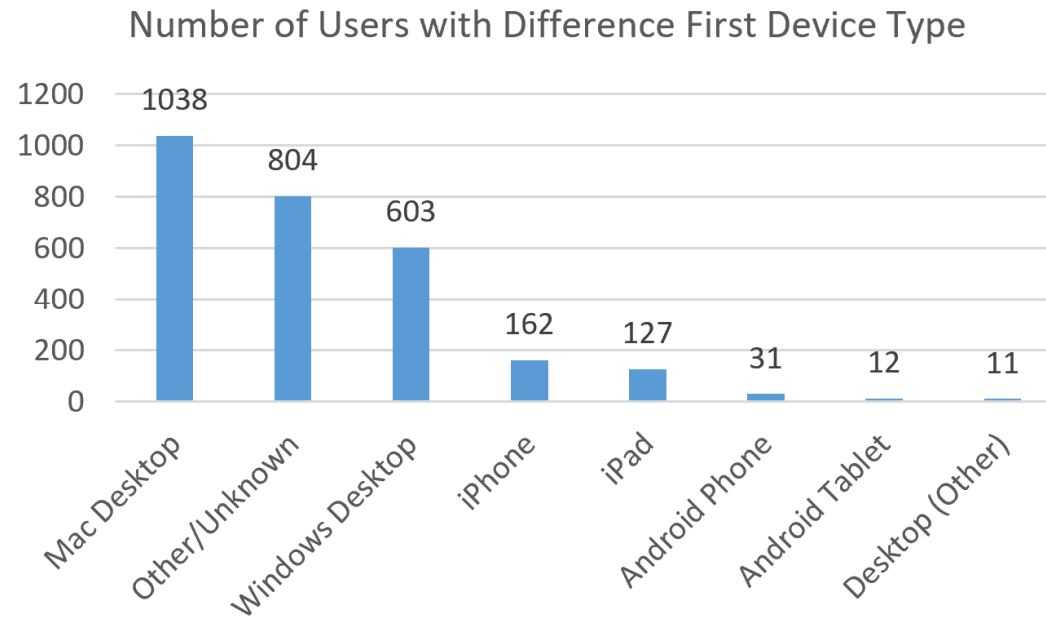
Number of Users for Each Affiliate Channel



Question 2

What is the total number of users associated with each device type in 2010?

- Desktops (Including Mac and Windows) are popular device back in 2010.
- IOS users are much more than Andriod users.



Question 3

What are the Top 3 actions performed by users based on action frequency?

- “Searching” is the most frequent action for new users
- “Personalize” Interface and “Show” the results of searching are also popular

action	action_count
search_results	6117
personalize	6005
show	5437

Question 4

What are the Top 5 country destinations, and how many users belong to each of these destinations?

- US take the majority and rank the 1st. It may be because Airbnb is an App/platform originated from US.
- France (FR), Spain (ES), Italy (IT) follow the rank. They are popular tourist destinations in Europe, aligning with people's common sense.

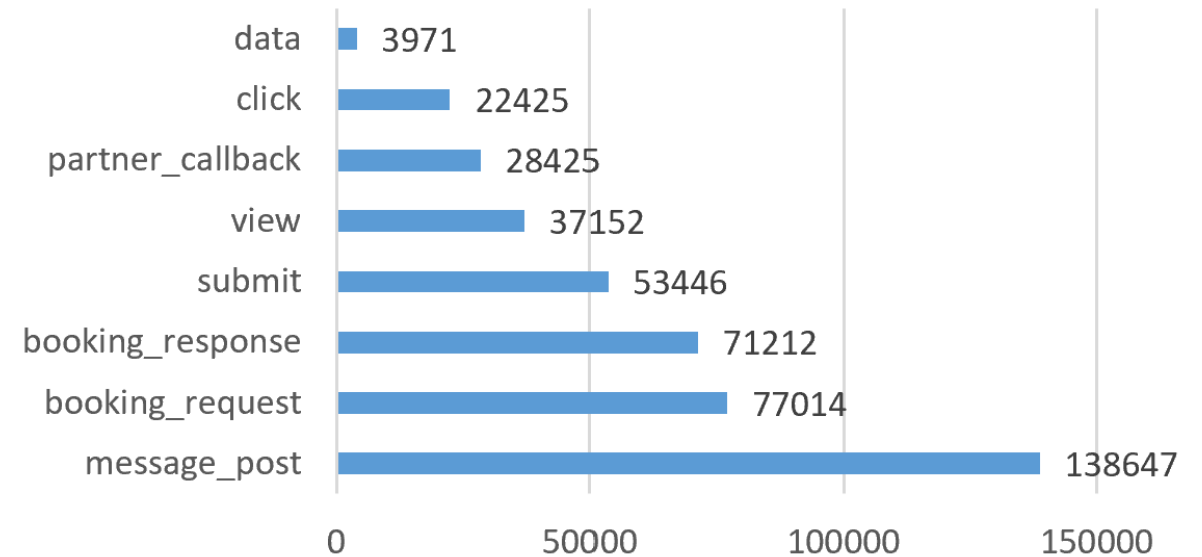
country_destination	user_count
US	5261
FR	558
other	553
ES	225
IT	216

Question 5

For each action type, what are their average session duration?

- “message_post” stands out among all Action Types
- Booking related ones (Booking_response, booking_request) rank the next
- “data” has the least session duration

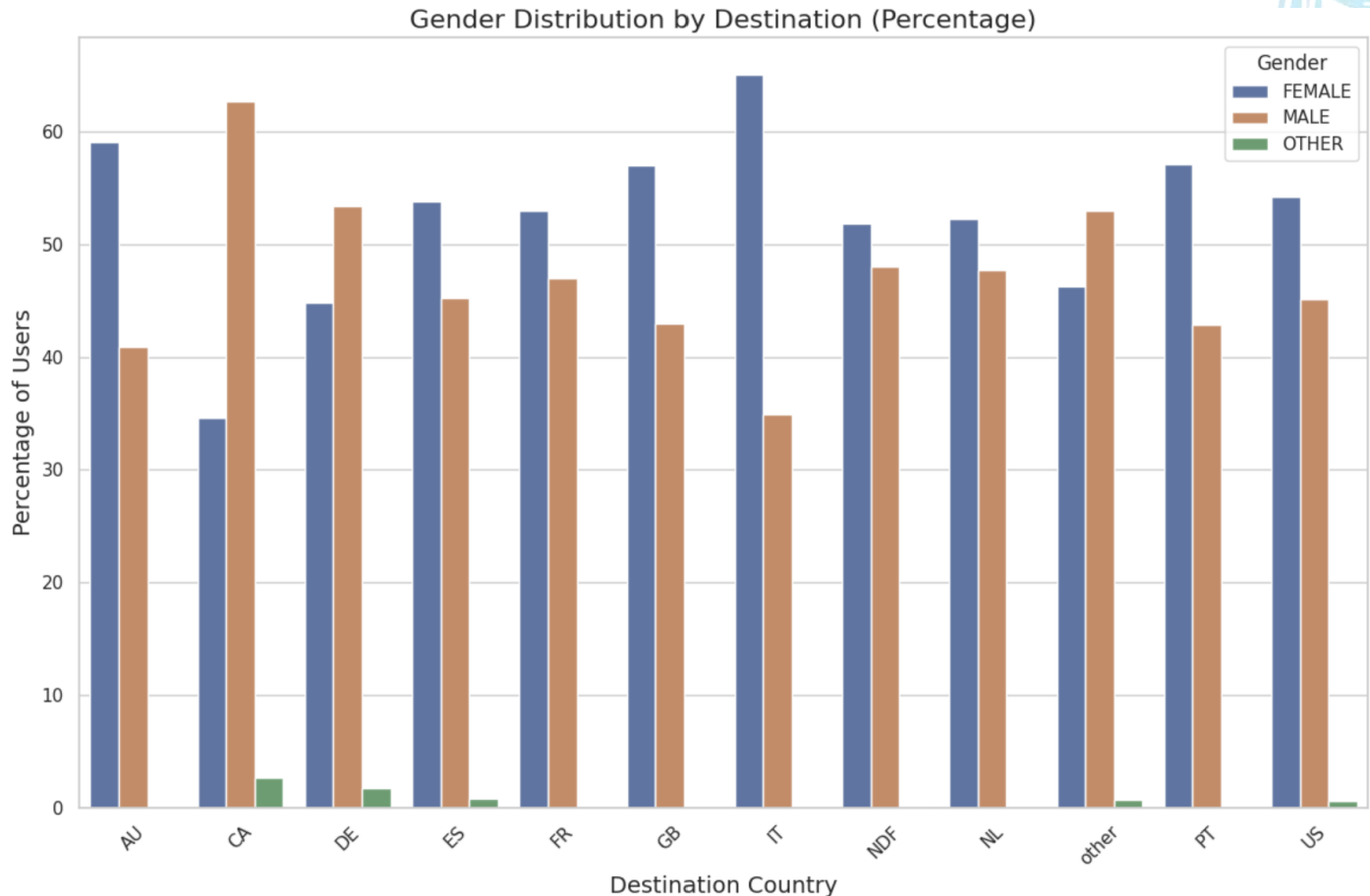
Average Session Duration of Action Types



Question 6

What is the gender distribution of users for each destination?

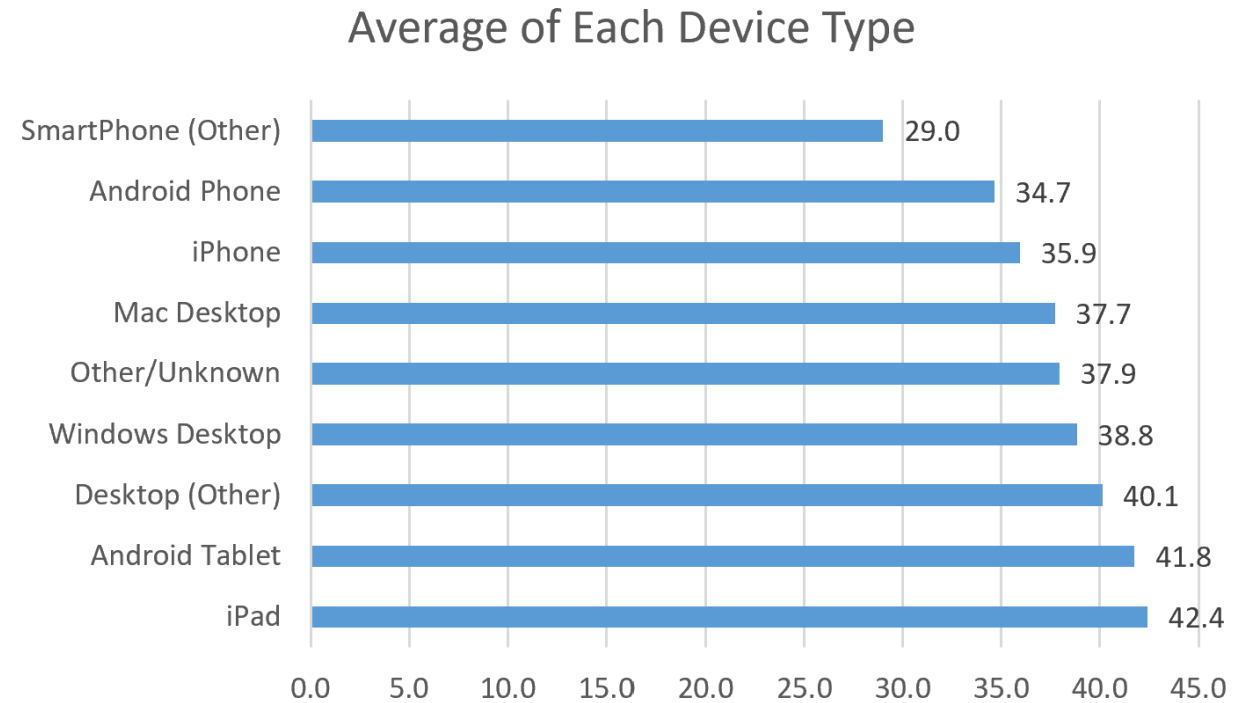
- Among 12 destinations, 9 of them has more female users, and rest 3 of them has more male users.
- Indicate female users might have higher intention and are more willing to book on Airbnb.
- Outstanding value: Canada (CA) has twice male users then female; Italy (IT) has twice female users then male



Question 7

What are the average ages of users of various devices?

- Average Age are in range between 34 to 42
- Younger the user, higher the willingness on using mobile phone
- Average age of users using Tablet accede 40. It might be due to larger screen compare with mobile phone and its convenience compare with Desktop



Question 8

Which destination country has the highest population contribution (in thousands) for each gender and age bucket?

- The destination country is still US, aligning with the result of previous question.
- Top 5 age bucket goes with young people (20-29) and middle-aged people (50-59).
- Young people are more willing to try new things; middle-aged people are at retirement age having more time for tour.

country_destination	age_bucket	gender	total_population
US	20-24	male	11601
US	50-54	female	11413
US	25-29	male	11385
US	55-59	female	11264
US	20-24	female	11094

Summary

Business Wise:

- At early stage of an App (the dataset is at 2010-2011), user data are very important, and worth explore, which can provide insights of their habit and choices.
- Do marketing and iterate product design according to demographic and behavior data of users.

Data Management Wise:

- Be careful and deal with “NA”, “unknown”, blank, and other type of anomaly data before diving into query writing and data analysis.
- Making ERD and Tables according to the goal can be helpful.

Thanks!

