BI

Mini Project

Airbnb

Table of Contents

[1 Introduction 4](#_Toc384331217)

[1.1 Setup Checklist for Mini Project 4](#_Toc384331218)

[1.2 Instructions 4](#_Toc384331219)

[2 Problem Statement 5](#_Toc384331220)

[2.1 Objective 5](#_Toc384331221)

[2.2 Abstract of the project 5](#_Toc384331222)

[2.3 Technology used: 5](#_Toc384331223)

[3 Implementation in BI LOT 6](#_Toc384331224)

[3.1 Guidelines on the functionality to be built: 7](#_Toc384331226)

# Introduction

This document outlines a mini project for the BI LOT. The project is to predict which country a new user's first booking destination will be. This document contains the work flow of the system and gives guidelines on how to build the functionality gradually in each of the course modules of the BI LOT.

## Setup Checklist for Mini Project

Minimum System Requirements

* Intel Pentium 90 or higher (P166 recommended)
* Microsoft Windows 2010 or above.
* Memory: 4GB of RAM (64MB or more recommended)
* Internet Explorer 10.0 or higher
* Oracle 11g/ SQL Server 2012 or higher
* Informatica 10.2.0 / SSIS / Data Stage / Abinitio
* Tableau 10 or higher / SSRS / PowerBI / Qlikview

## Instructions

* The code modules in the mini project should follow all the coding standards.
* Create a directory by your name in drive **<drive>**. In this directory, create a subdirectory **MiniProject**. Store your Project here.
* You can refer to your course material.
* You may also look up the help provided in the BI docs and documentation provided in respective tools.

# Problem Statement

## Objective

Airbnb Analysis based on the destinations.

## Abstract of the project

The project aim is to predict which country a new user's first booking destination will be. All the users in this dataset are from the USA. There are 12 possible outcomes of the destination country: 'US', 'FR', 'CA', 'GB', 'ES', 'IT', 'PT', 'NL','DE', 'AU', 'NDF' (no destination found), and 'other'. Please note that 'NDF' is different from 'other' because 'other' means there was a booking, but is to a country not included in the list, while 'NDF' means there wasn't a booking.

Technology used:

* Oracle 11g / SQL Server (Database)
* Informatica power center / SSIS / Data Stage / Abinitio(ETL Tool)
* Tableau Desktop / SSRS / PowerBI / Qlikview (Reporting Tool)

# Implementation in BI LOT

## Summary of the functionality to be built:

The participants need to develop the Airbnb Analysis to build the functionality incrementally in each of the course modules of BI LOT using one of the above ETL and Reporting tool.

## Guidelines on the functionality to be built:

**Project flow:**

Heterogeneous Sources i.e. operational data (Flat files)

EXTRACT

TRANSFORM

LOAD

Data warehouse (Oracle)

Reports Based On Tableau

R

E

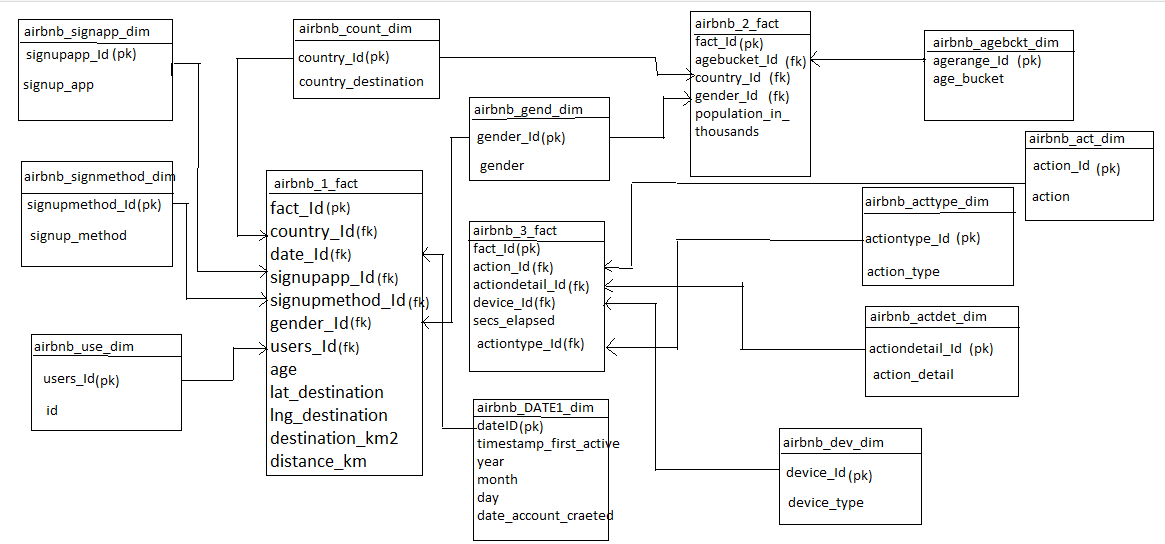
P

O

R

T

**Schema Design:**



**NOTE:**

The datatype/length for the Dimension/Fact table attributes can be changed as required. Additional fields can be added, if required.

* Airbnb\_act\_dim

This dimension contains actions which is done by users.

This oracle table comprises of the following Fields:

* action\_id(PK)
* action

The action\_ID is a running sequence and unique number to be generated for all the actions.

* Airbnb\_acttype\_dim

This dimension contains the type of action done by user and the oracle table comprises of the following Fields:

* actiontype\_id(PK)
* action\_type
* airbnb\_actdet\_dim

This dimension contains details of action and the oracle table comprises of the following Fields:

* actiondetail\_id(PK)
* action\_detail
* airbnb\_dev\_dim

This dimension contains details of device type and the oracle table comprises of the following Fields:

* device\_id(PK)
* device\_type
* airbnb\_agebckt\_dim

This dimension contains age range of users with their unique id and the oracle table comprises of the following Fields:

* agebucket\_id
* age\_bucket
* airbnb\_country\_dim

This dimension contains the details of country destinations and oracle table comprises of the following Fields:

* country\_id(PK)
* country\_destination
* airbnb\_gender\_dim

This dimension contains the details of gender and oracle table comprises of the following Fields:

* gender\_id(PK)
* gender
* airbnb\_date\_dim

This dimension contains the details of country destinations and oracle table comprises of the following Fields:

* date\_id(PK)
* timestamp
* year
* month
* day
* date\_account\_created
* airbnb\_signupapp\_dim

This dimension contains the details of signing and oracle table comprises of the following Fields:

* signupapp\_id(PK)
* signup\_app
* airbnb\_signupmethod\_dim

This dimension contains the details of signup method and oracle table comprises of the following Fields:

* signupmethod\_id(PK)
* signup\_method
* airbnb\_user\_dim

This dimension contains the details of user id and oracle table comprises of the following Fields:

* users\_id(PK)
* id
* airbnb\_1\_fact

This fact contains details of id’s and measuring units. Oracle table comprises of the following Fields:

* fact\_id(PK)
* country\_id
* date\_id
* signupapp\_id
* signupmethod\_id
* gender\_id
* users\_id
* age
* lat\_destination
* long\_destination
* destination\_km
* destination\_km2
* airbnb\_2\_fact

This fact contains details of measuring units and oracle table comprises of the following Fields:

* fact\_id(PK)
* agebucket\_id
* country\_id
* gender\_id
* population\_in\_thousands
* airbnb\_3\_fact

This is a Fact table that would contain all the measuring units

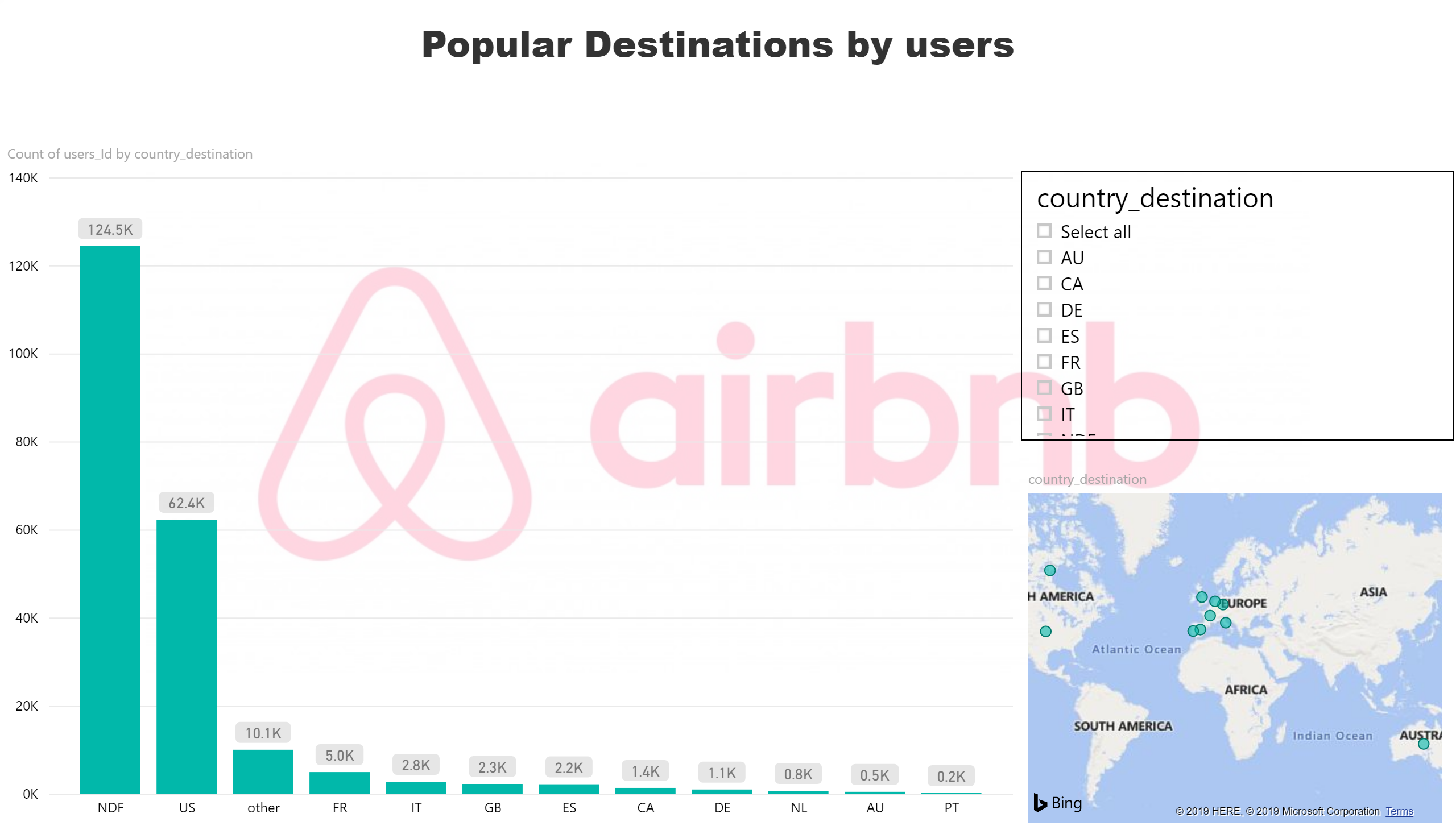
* Fact\_id
* action\_id(FK)
* actiondetail\_id(FK)
* actiontype\_id(FK)
* device\_id(FK)
* secs\_elapsed

**Data Transformation for data warehouse:**

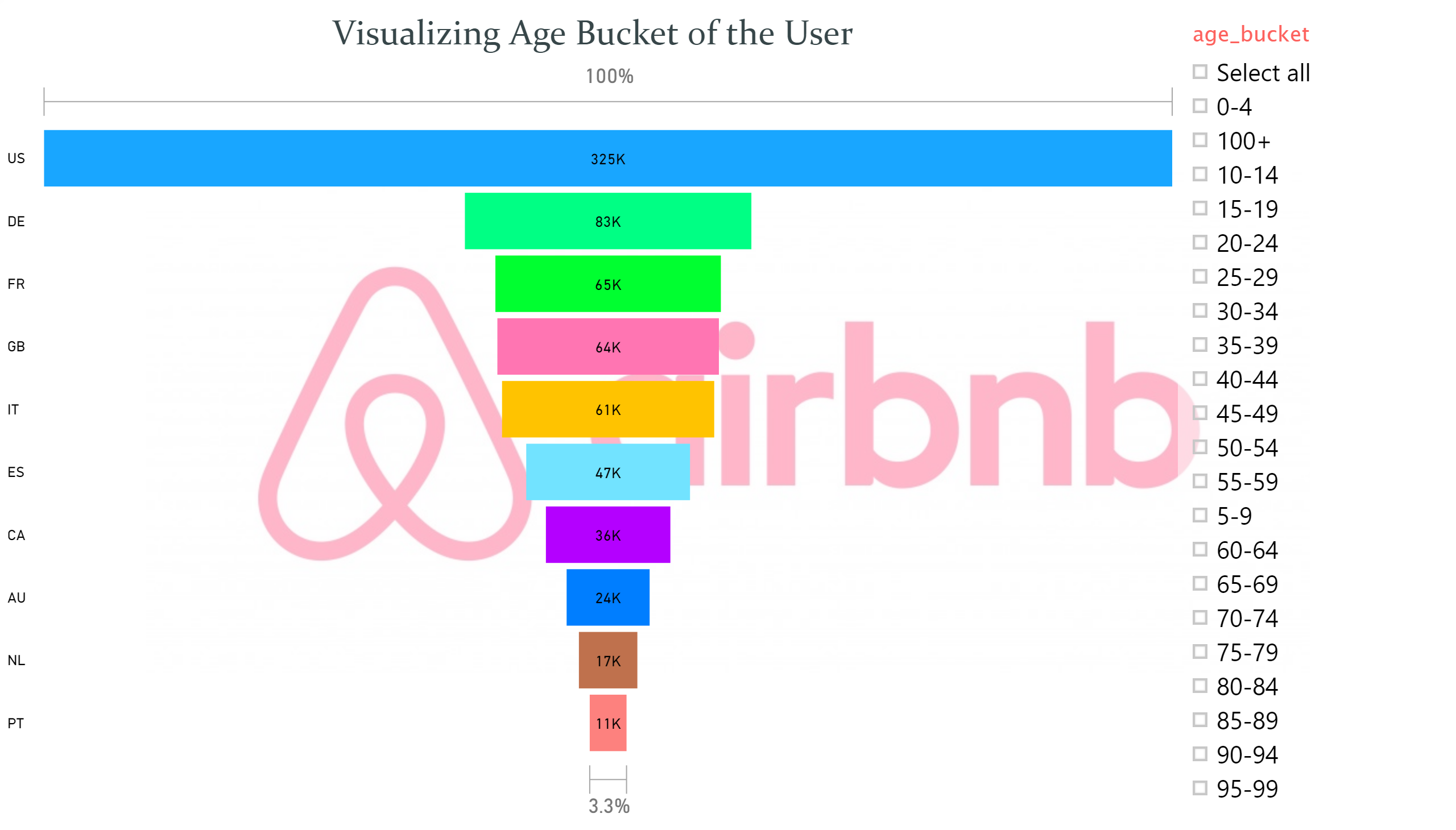
* Load the Data into Dimension tables using the CSV files provided.
* Load the data into Fact table.
  + Points to ponder.
    - Validate the foreign key values before loading.
    - Fact\_id is the primary key.
    - The fact table should store the data pertaining to all the survey details.
    - The measures that do not have data in the CSV file should be given default value NULL in the fact table while loading the data.

**Sample Reports:**

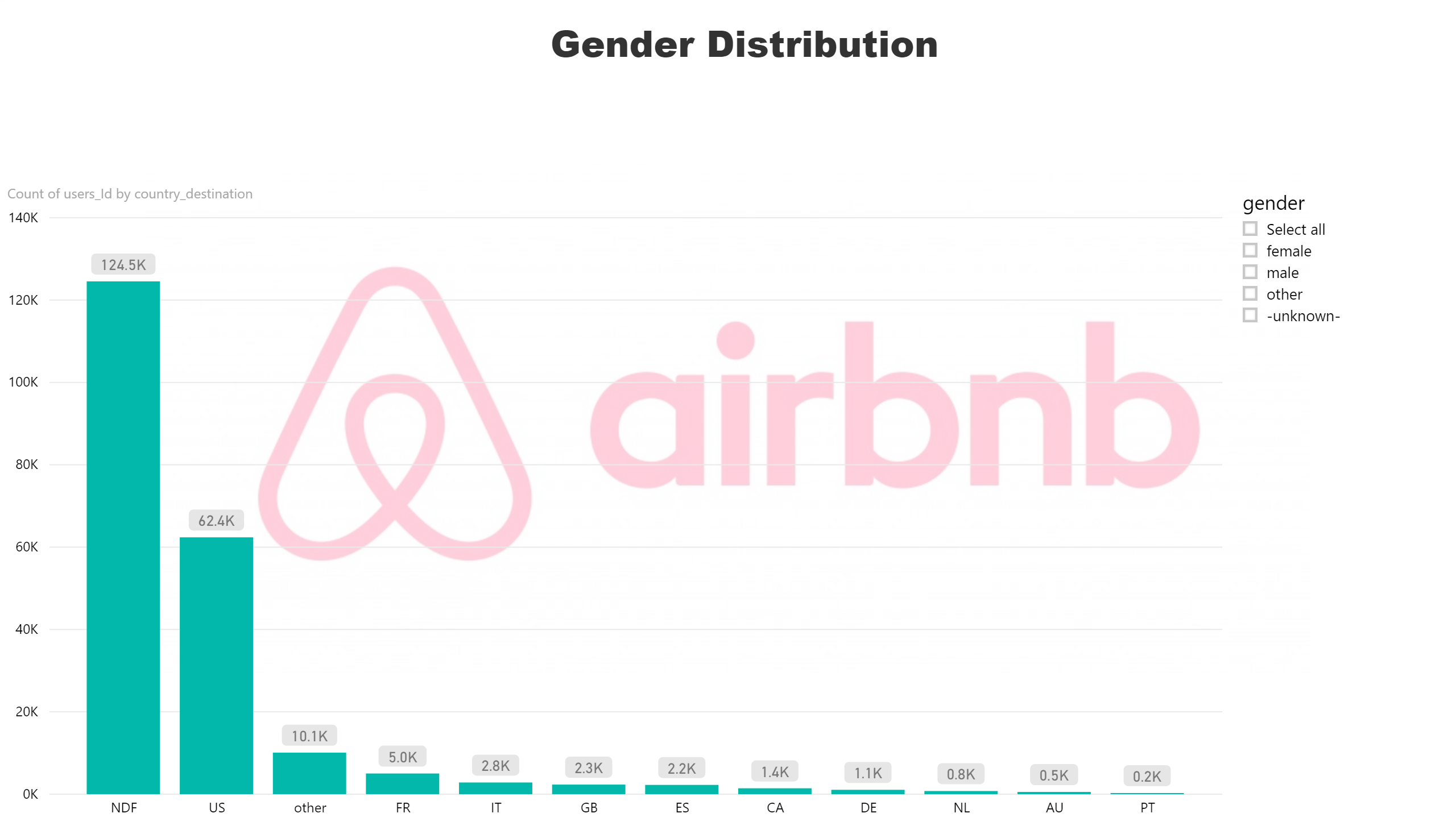
1. Popular destinations among users.

[](https://app.powerbi.com/reports/1e2c9760-e41a-4d69-9d62-557b2f0be06c/ReportSection?pbi_source=PowerPoint)

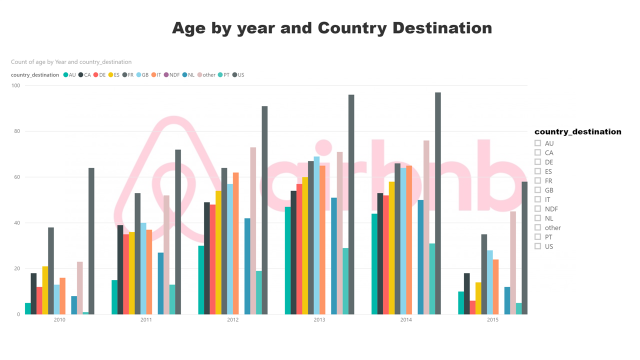
1. Visualizing the age bucket of the users.

[](https://app.powerbi.com/reports/1e2c9760-e41a-4d69-9d62-557b2f0be06c/ReportSection4fbec07681d7fa819157?pbi_source=PowerPoint)

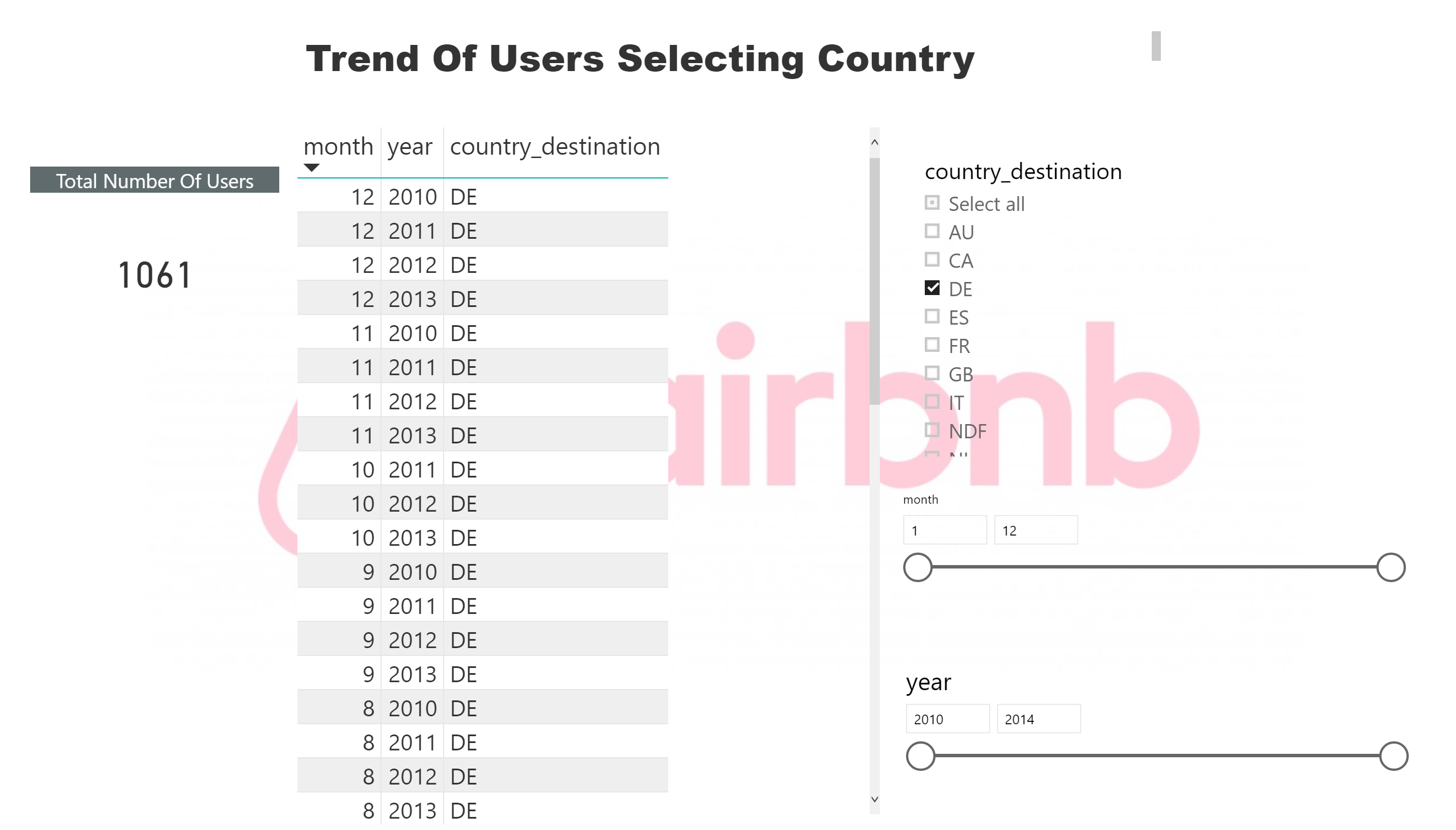
1. Visualizing gender distribution of the users.

[](https://app.powerbi.com/reports/1e2c9760-e41a-4d69-9d62-557b2f0be06c/ReportSectionba7bc506aedcefd52bb2?pbi_source=PowerPoint)

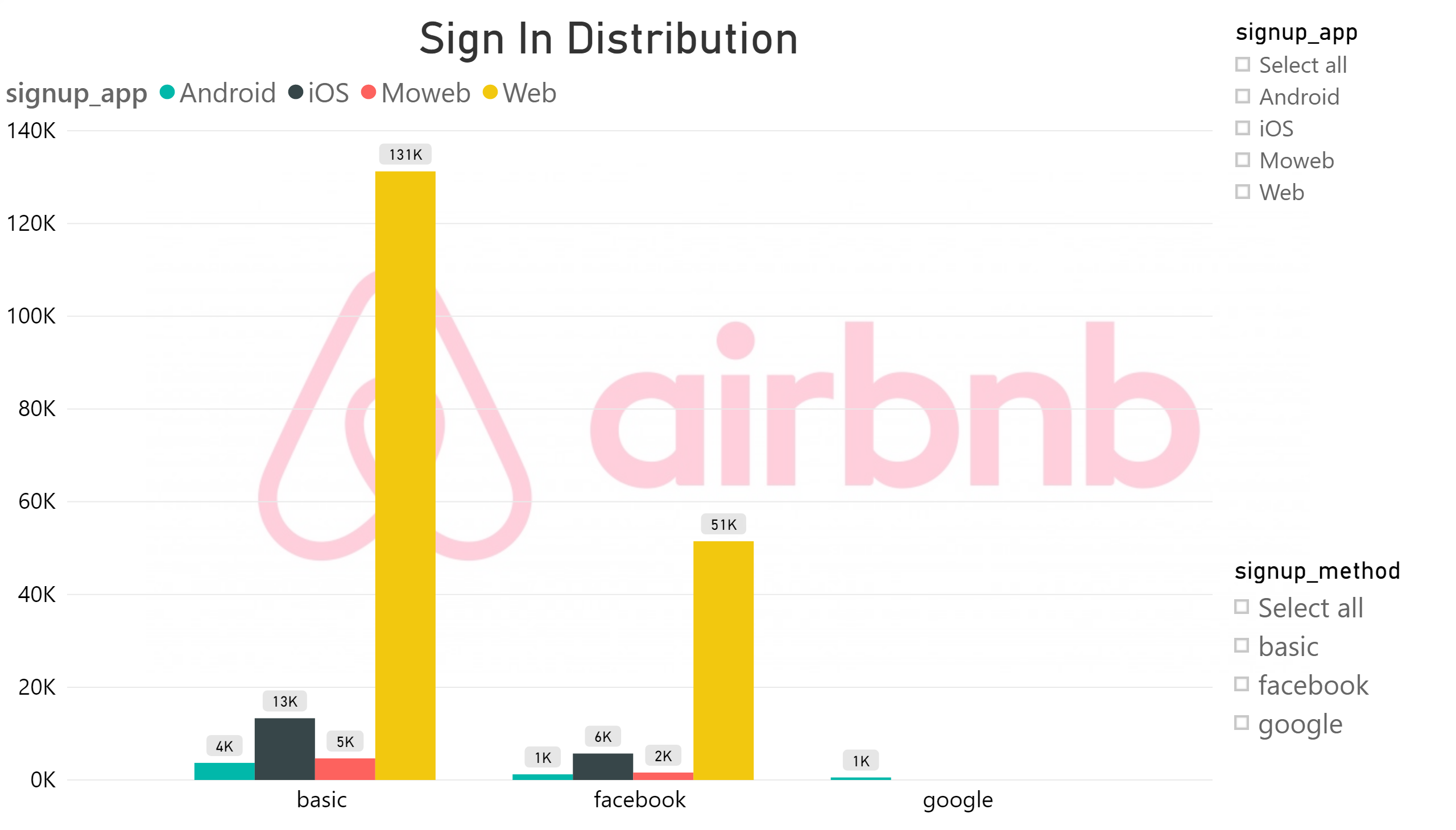
1. Choice of destination countries varies across age groups.



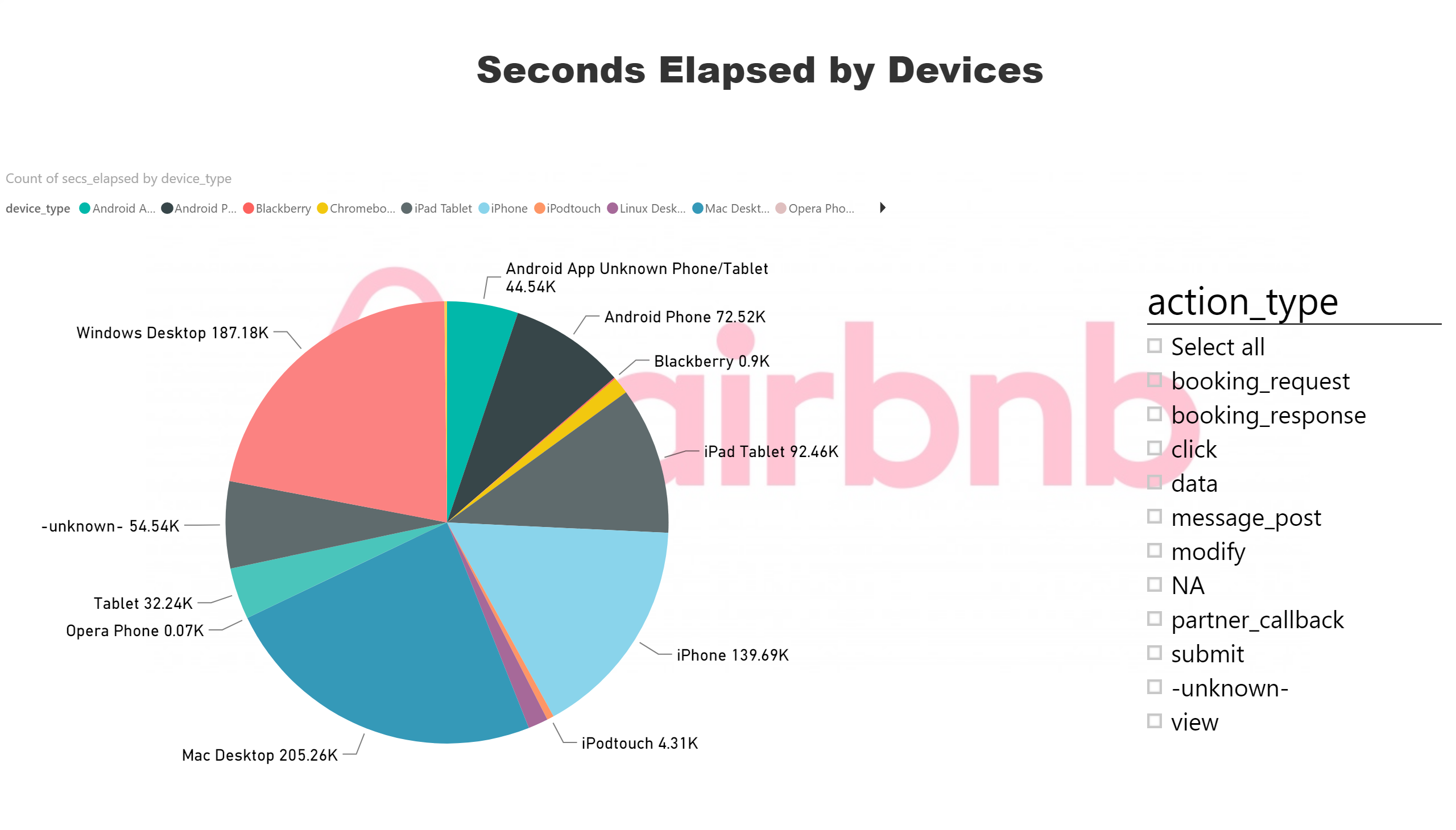
1. Choice of destination countries depending on date

[](https://app.powerbi.com/reports/1e2c9760-e41a-4d69-9d62-557b2f0be06c/ReportSection6abba1dfa6cdd9e17f62?pbi_source=PowerPoint)

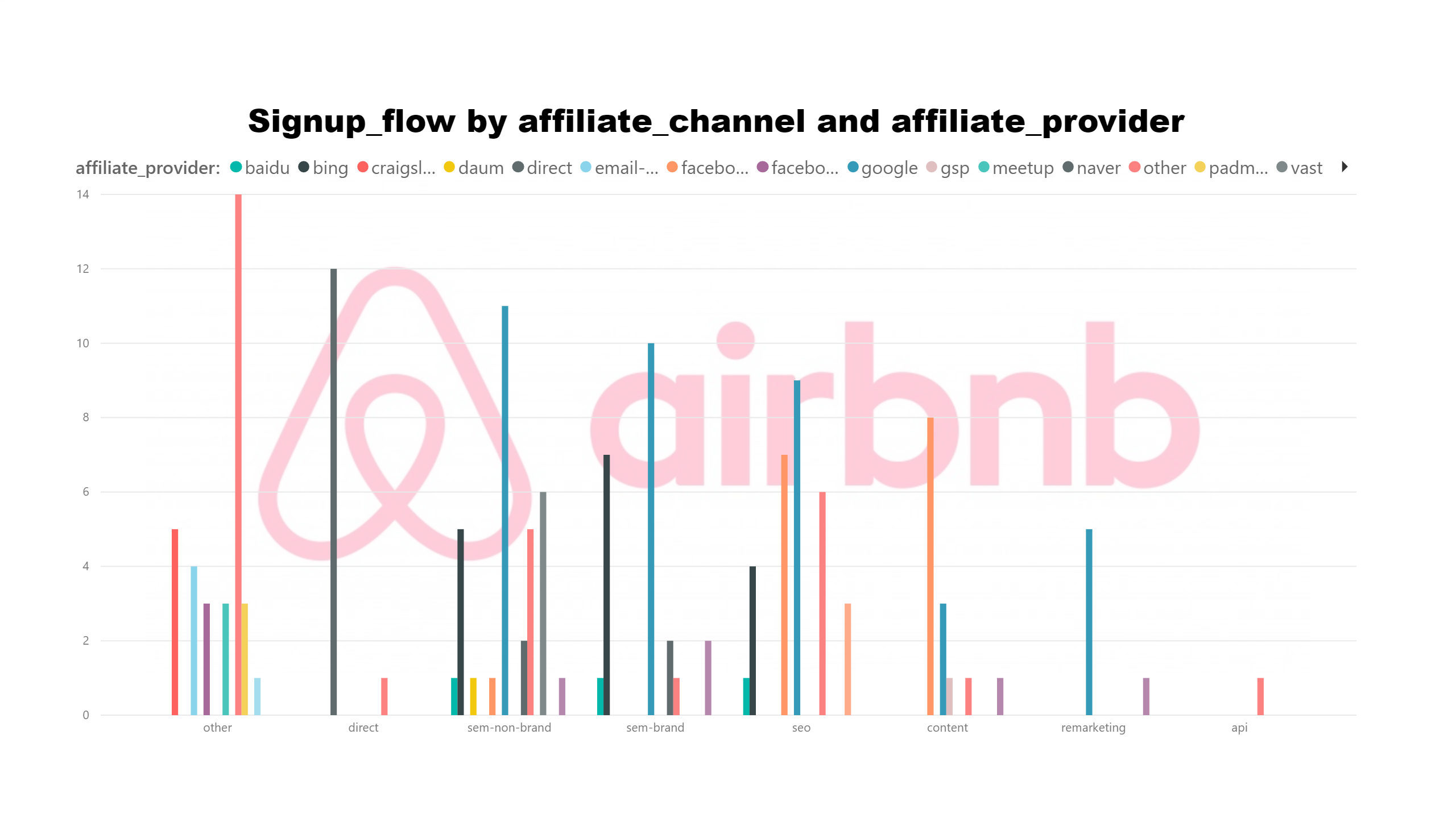
1. Sign-In method distribution.

[](https://app.powerbi.com/reports/1e2c9760-e41a-4d69-9d62-557b2f0be06c/ReportSection83a930b53d44b487c5fe?pbi_source=PowerPoint)

1. Seconds elapsed by various devices

[](https://app.powerbi.com/reports/1e2c9760-e41a-4d69-9d62-557b2f0be06c/ReportSectione3cc6147369dab8246d2?pbi_source=PowerPoint)

1. Signup Flow by affiliate providers and affiliate channels.

[](https://app.powerbi.com/reports/1e2c9760-e41a-4d69-9d62-557b2f0be06c/ReportSection398518a046ac433896bf?pbi_source=PowerPoint)