Looker E-commerce Dataset

Data source : Kaggle

Cloud Technology project by

Hui Yi Huang

Sara Canu

Agenda:

0. **Dataset overview**

1. **Users' differentiation**:
   Find the biggest proportion of users in gender, age, country, and traffic.

   Users' registration:
   Find out when the most users created accounts and the difference in the average amount of user accounts created between each month between 2020-2023.

2. **Cart Events:**
   Find out which traffic source causes the most cart events, dividing users between age and gender; Find out the difference in the number of cart events between each month in 2020-2023.

3. **The conversion rate of purchase events:**
   Find the conversion rate and the difference in average monthly conversion rates for each traffic source.

4. **Orders and Profits**:
   Find the profit and profit margin in countries and the proportion of users' age.

0. Overview

Users ['**id**', 'first_name', 'last_name', 'email', 'age', 'gender', 'state', 'street_address', 'postal_code', 'city', 'country', 'latitude', 'longitude', 'traffic_source', 'created_at']

Orders ['**order_id**', **'user_id'**, 'status', **'gender'**, 'created_at', 'returned_at', 'shipped_at', 'delivered_at', 'num_of_item']

Events ['**id**', **'user_id'**, 'sequence_number', 'session_id', 'created_at', 'ip_address', 'city', 'state', 'postal_code', 'browser', 'traffic_source', 'uri', 'event_type']

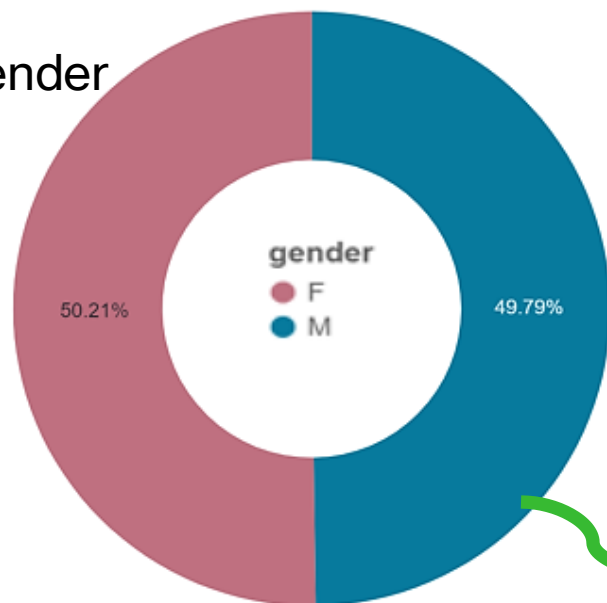Distribution_centers ['**id**', 'name', 'latitude', 'longitude']

Products ['**id**', 'cost', 'category', 'name', 'brand', 'retail_price', 'department', 'sku', **'distribution_center_id'**]

Inventory_items ['**id**', **'product_id'**, 'created_at', 'sold_at', **'cost'**, **'product_category'**, **'product_name'**, **'product_brand'**, **'product_retail_price'**, **'product_department'**, **'product_sku'**, **'product_distribution_center_id'**]
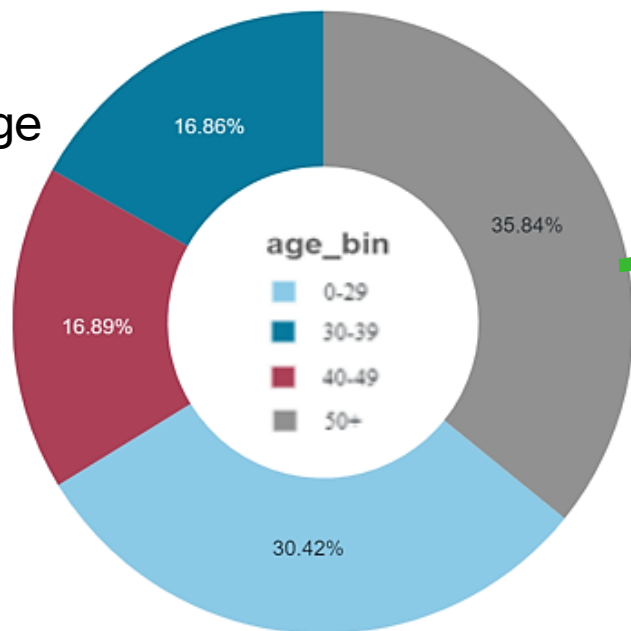
Order_items ['**id**', **'order_id'**, **'user_id'**, **'product_id'**, **'inventory_item_id'**, 'status', 'created_at', 'shipped_at', 'delivered_at', 'returned_at', 'sale_price']
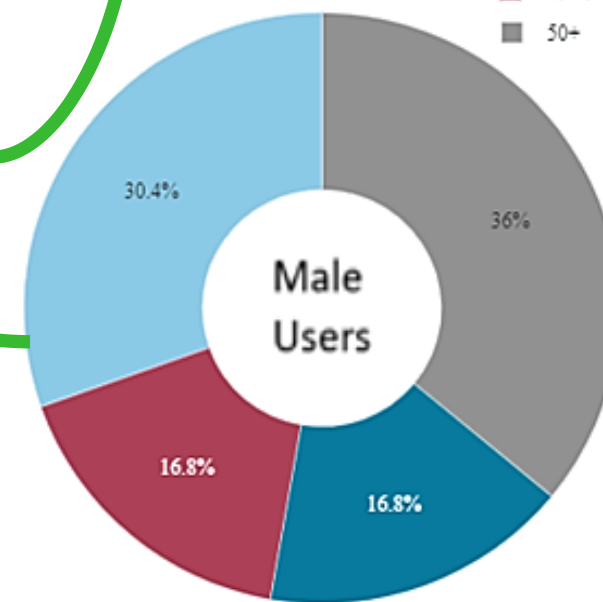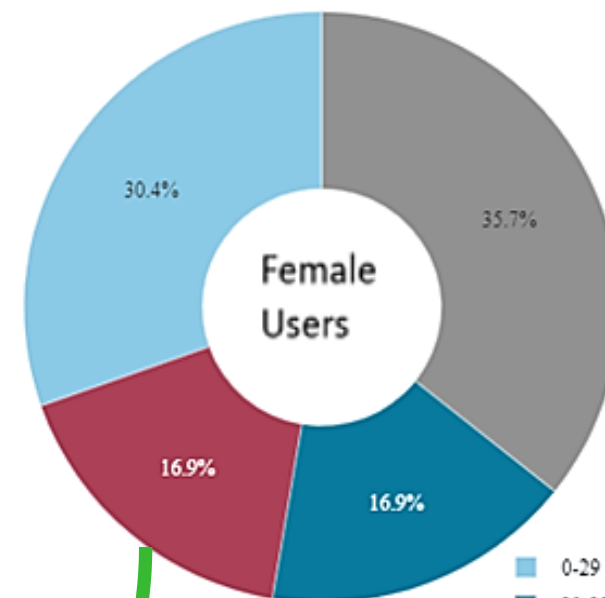
Gender

Age

Q:
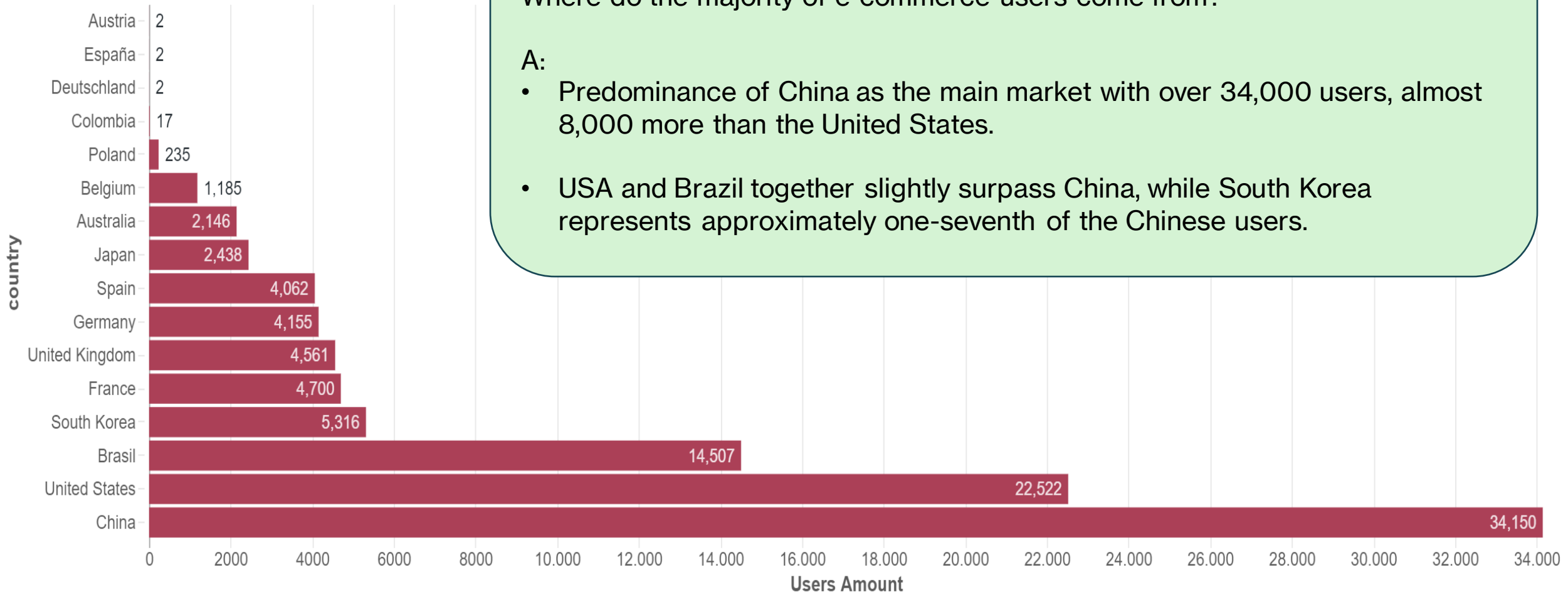Is there any difference in the usage of the E-commerce by gender and age?

A:
- There is an almost even split in E-commerce usage between genders.

- Two distinct age groups stand out those under 30 and those over 50.

- It's interesting to note that the gender division also follows these same age groupings.

Age division per gender

Amount of users per country

Q:
Where do the majority of e-commerce users come from?

A:
- Predominance of China as the main market with over 34,000 users, almost 8,000 more than the United States.

- USA and Brazil together slightly surpass China, while South Korea represents approximately one-seventh of the Chinese users.

| country	Users Amount
Austria	2
España	2
Deutschland	2
Colombia	17
Poland	235
Belgium	1,185
Australia	2,146
Japan	2,438
Spain	4,062
Germany	4,155
United Kingdom	4,561
France	4,700
South Korea	5,316
Brasil	14,507
United States	22,522
China	34,150

Count of traffic sources

Q:
What are the primary traffic sources through which users access our e-commerce platform?

A:
- Evident disparity among traffic sources, with 'search' overwhelmingly dominating at over 70,000 accesses, indicating the significance of search engines in driving traffic.

- Following observe 'organic' traffic with just over 15,000 accesses.

- Other notable sources include 'Facebook' with 5,800 accesses, 'email' with 4,900 accesses, and 'display' with 4,000 accesses.
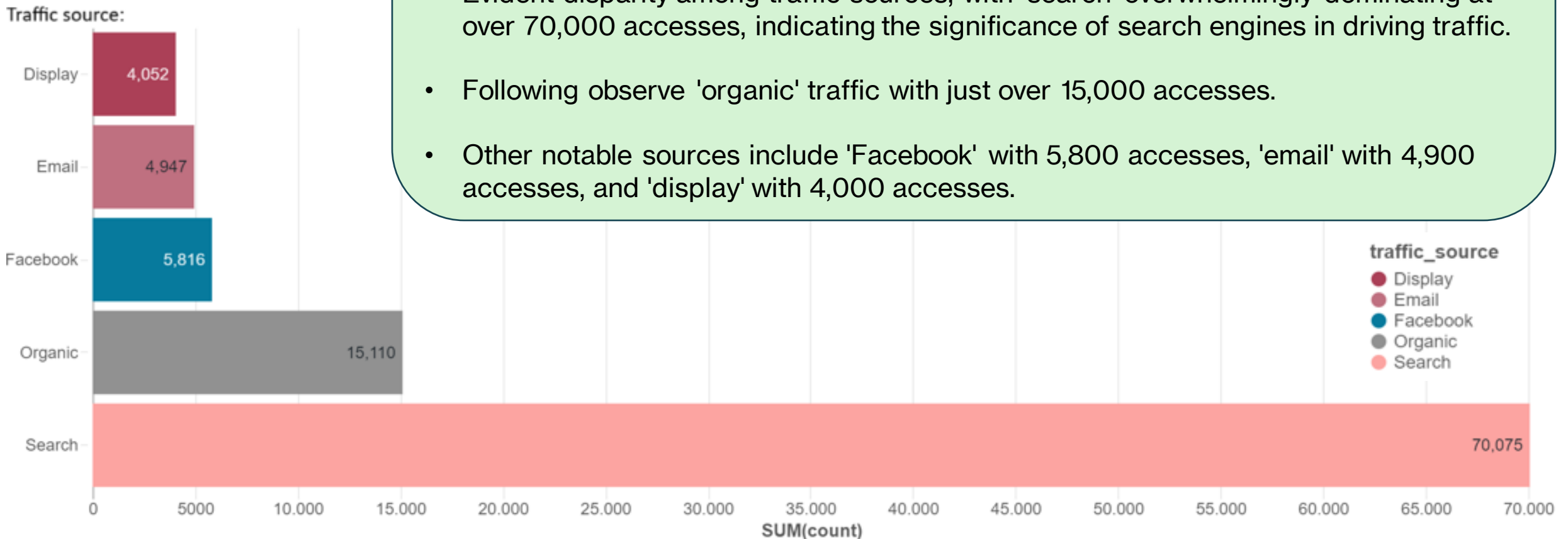
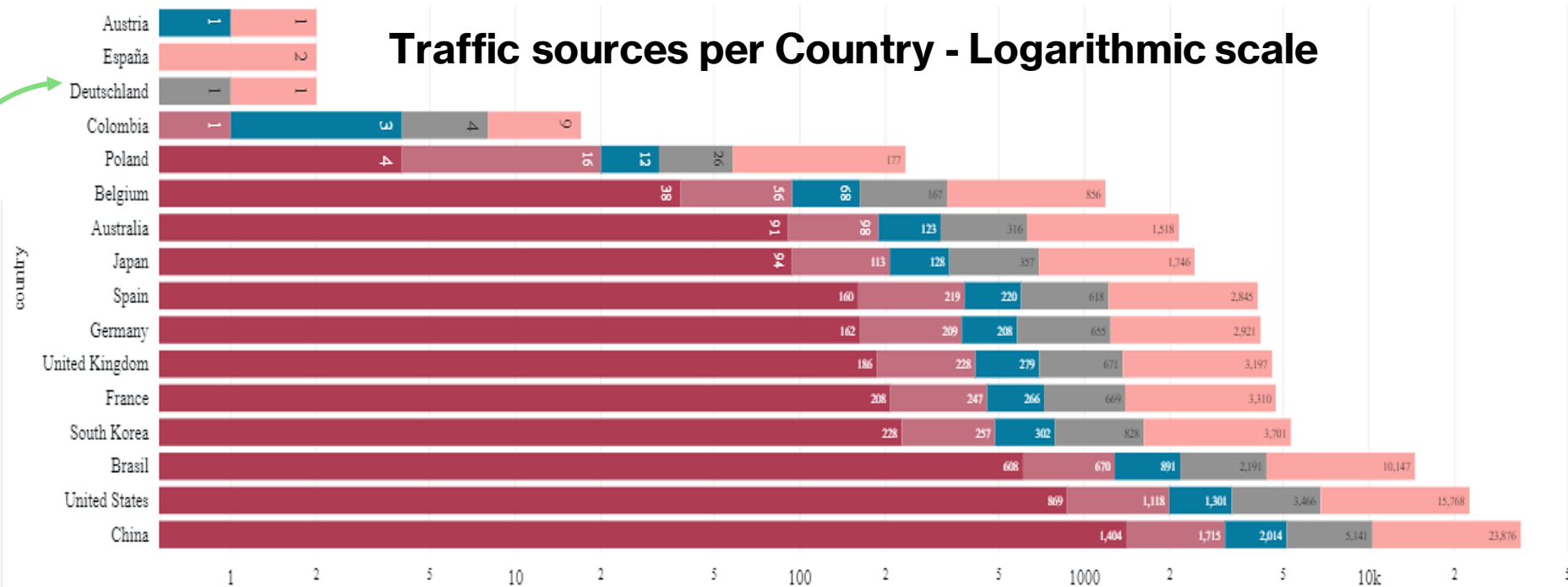**Traffic sources per Country – Linear scale**
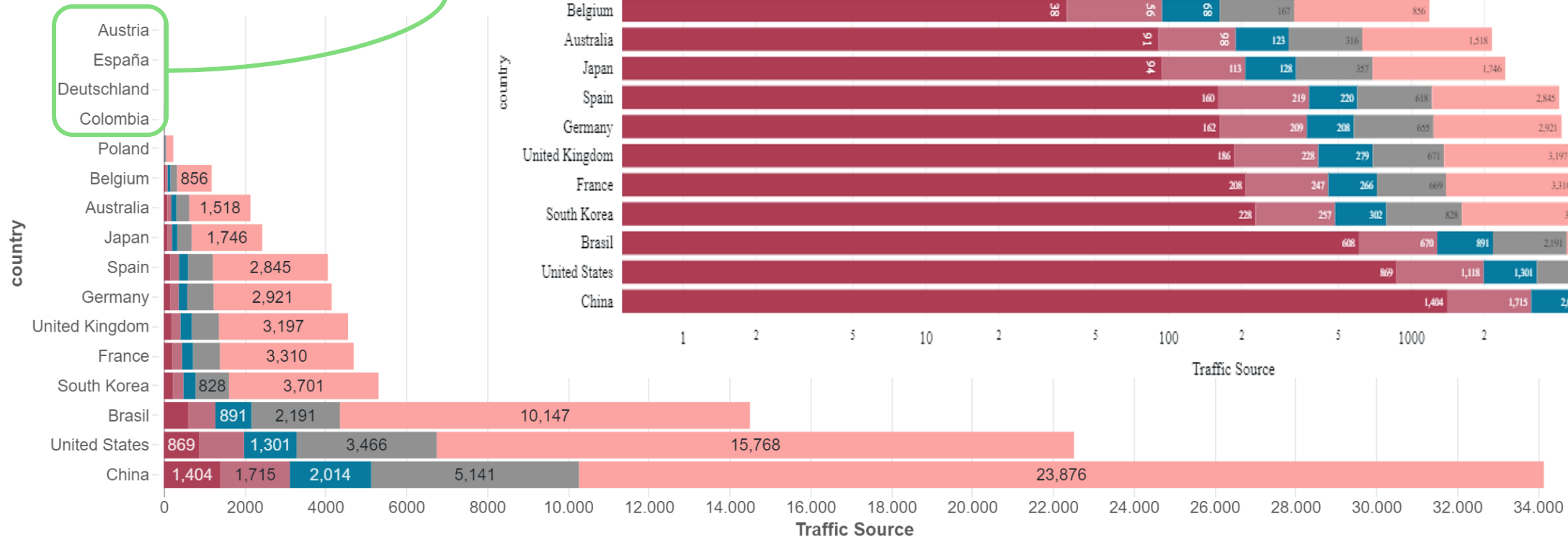
Traffic sources per Country - Logarithmic scale

traffic_source
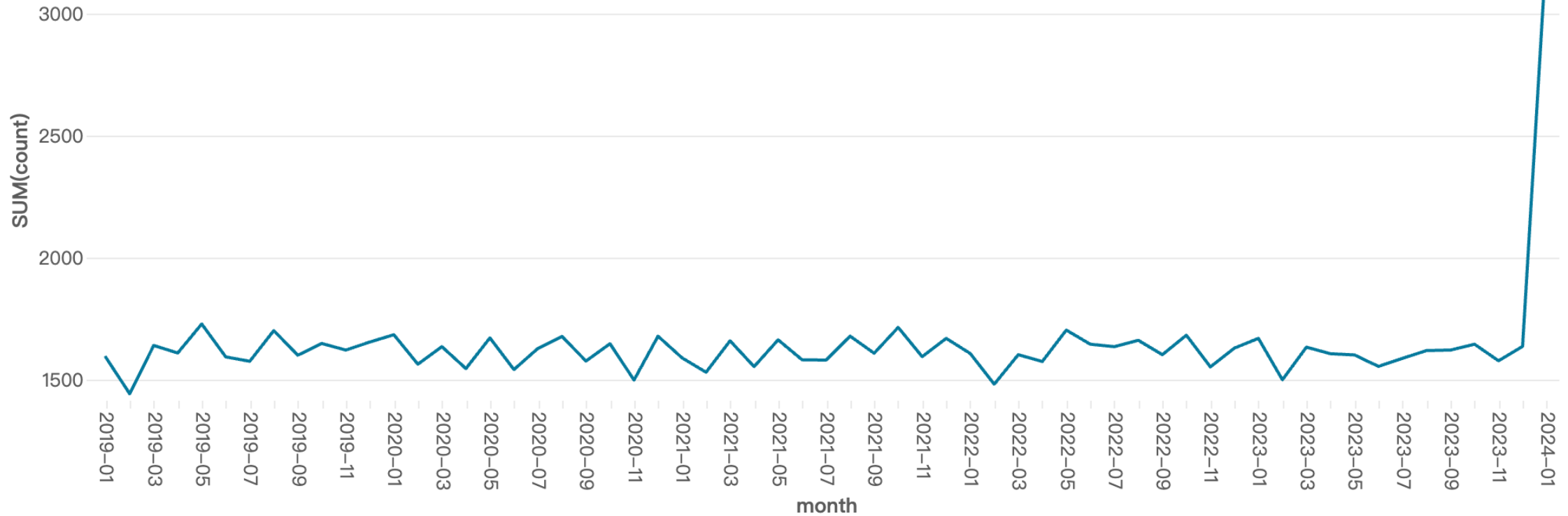- Display
- Email
- Facebook
- Organic
- Search

Q: Does the variation in the distribution of traffic sources across different countries of our users suggest anything?
A:
- The entire website traffic from Spain comes from search engines.
- Austria and Deutschland are characterised by the usage of only two types of traffic sources.

Users' registration per month until 2024-01

Q: When did most of the users create their accounts?

A:
- The number of user registrations follows a relatively consistent pattern until the end of 2023, with monthly fluctuations remaining within a certain range.

```
The month with the maximum count is: 2024-01 with count: 3276
```
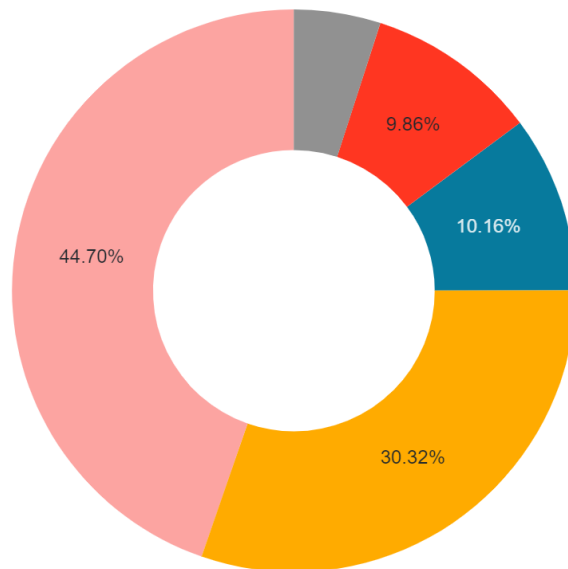
Users' registration per each year

Q:
Is there any difference in the average amount of user accounts created between 2020-2023 per month?

A:
- Appearance of a significant difference in the average amount of user accounts created between 2020 and 2023 for each February with a very low peak.
- Great low peak difference also in November 2020.

Cart event in traffic source

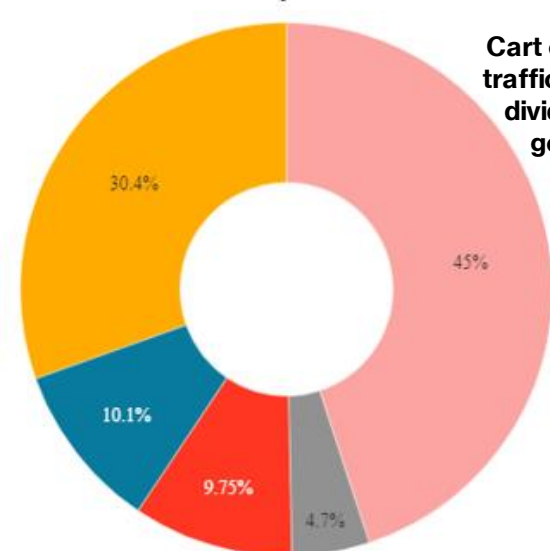traffic_source
- Email
- Adwords
- Facebook
- YouTube
- Organic

44.70%
9.86%
10.16%
30.32%

Cart events in traffic sources divided per gender

F
45%
30.4%
10.1%
9.75%
4.7%

M
44.4%
30.2%
10.2%
9.96%
5.22%

Q:
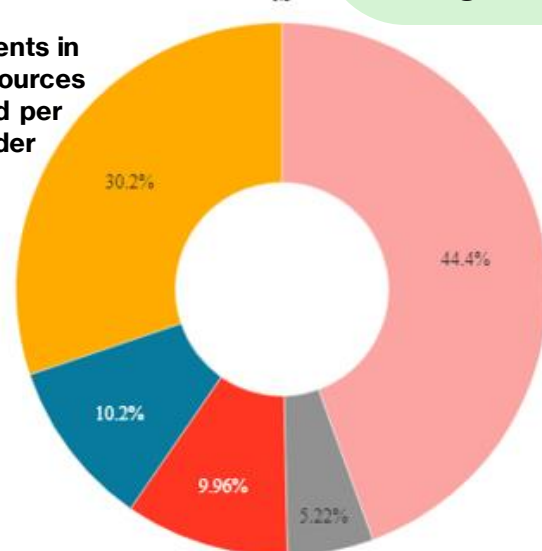Find out which traffic source causes the most cart events, dividing users between age and gender.

A:
- Email is the predominant traffic source, accounting for approximately 45% of all cart addition events.

- AdWords ranks second at 30%.

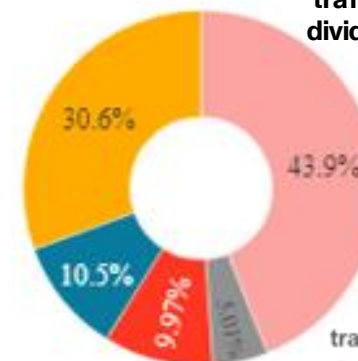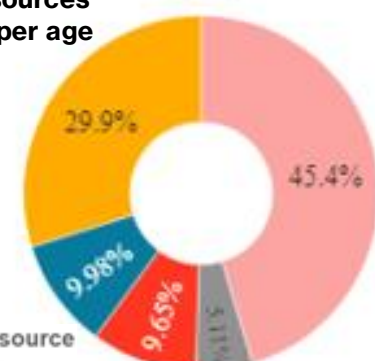- Facebook and YouTube each contribute around 10%. The remaining traffic is generated organically.

- These proportions remain consistent when considering segmentation by gender and age of users.
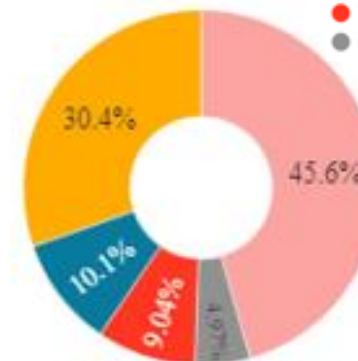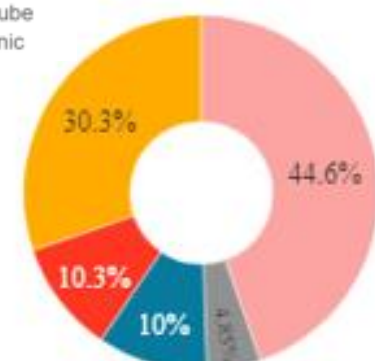
Cart events in traffic sources divided per age

0-29
43.9%
30.6%
10.5%
9.97%

30-39
45.4%
29.9%
9.98%
9.65%

traffic_source
- Email
- Adwords
- Facebook
- YouTube
- Organic

40-49
45.6%
30.4%
10.1%
9.04%

50+
44.6%
30.3%
10.3%
10%
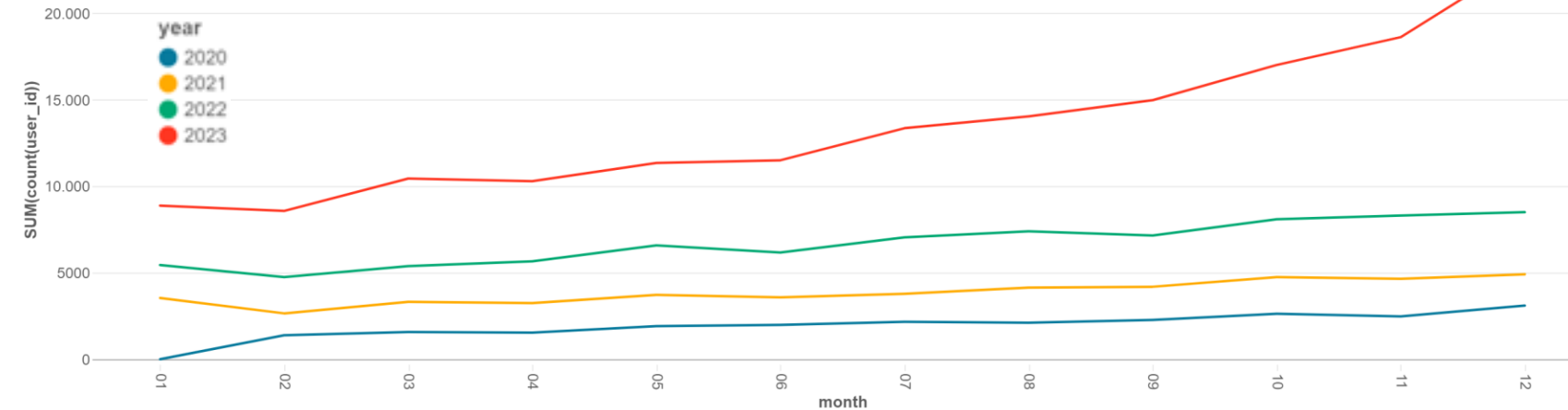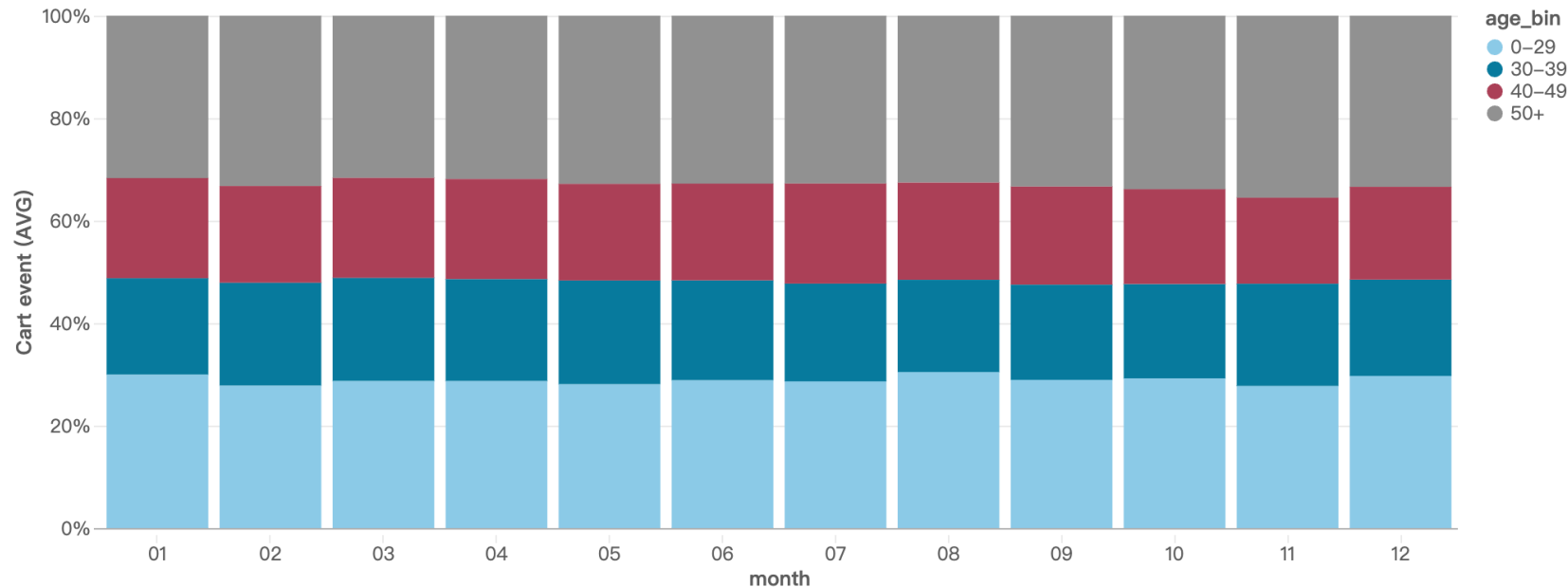
Cart events per year


Cart events proportion by age

Q:
Is there any difference in the number of cart events between each month in 2020-2023?

A:
- From 2020 to 2022 the number of cart events follows a relatively consistent pattern throughout the year, with minimal variations between months.

- In 2023 a significant shift, with linear growth in the quantity of cart events for most of the year with a significant peak at the end.

- There was little difference in age proportions from month to month.

Conversion rate AVG growth
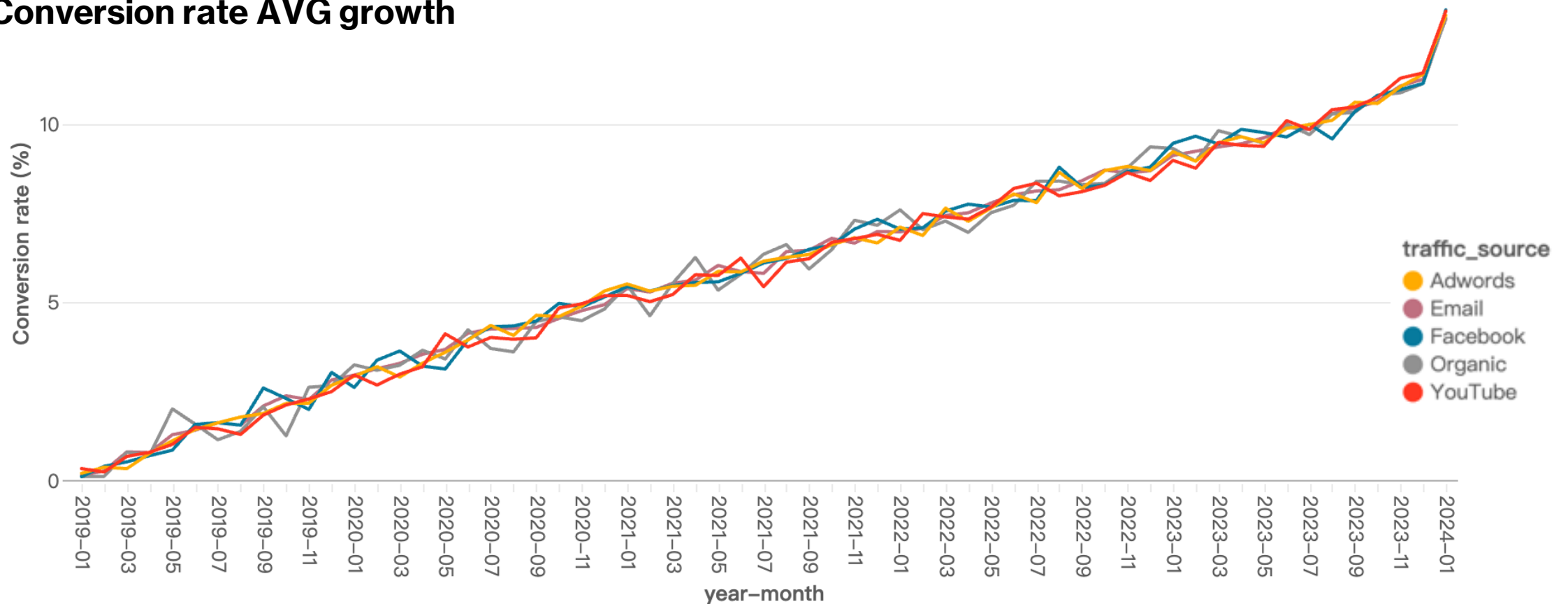
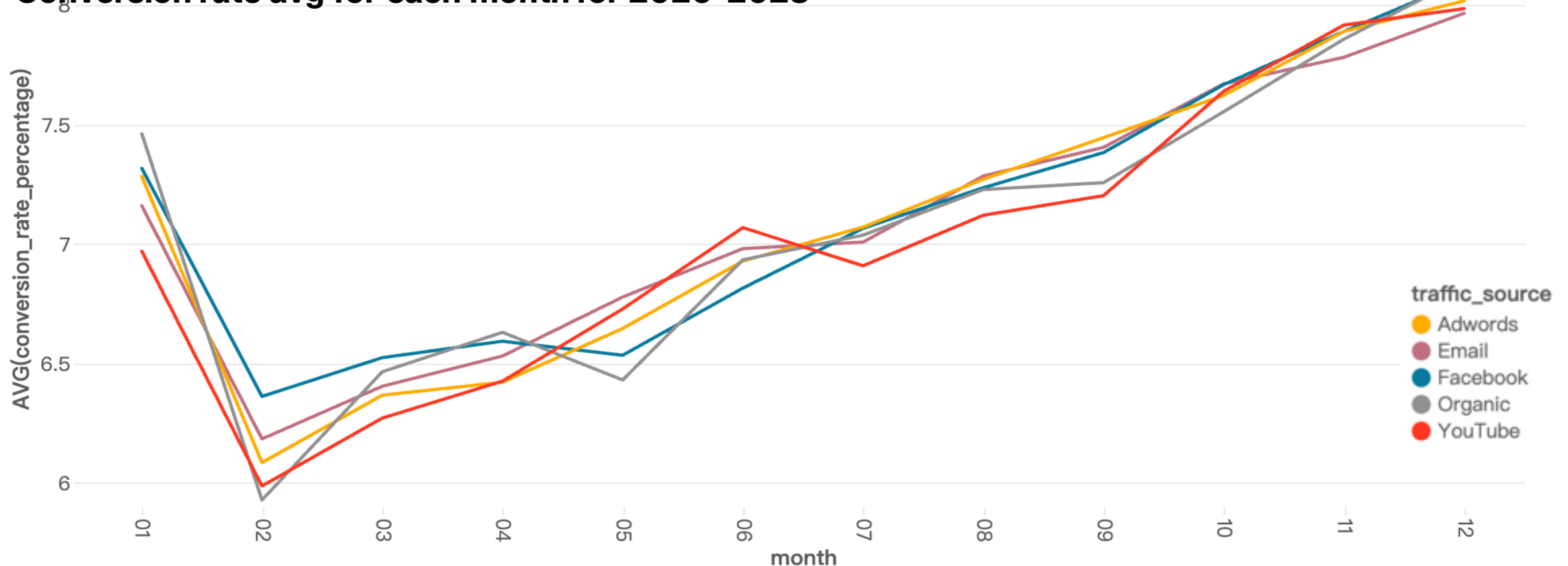Q: How effective is traffic-induced purchases?
A:
- It has been growing since the beginning (2019-01).
- Especially in 2024-01, the conversion rate is the highest at 13.21%.

The date with the highest conversion rate is 2024-01, with a conversion rate of 13.21%.

**Conversion rate avg for each month for 2020-2023**

Q: Which month has an unusually low traffic conversion rate?
A:
- According to the mature data between 2020-2023, the lowest traffic conversion rate is in February.
- We dug into the differences between sources and found that no matter which source had the lowest conversion rate in February

Profit and Sale Price by country
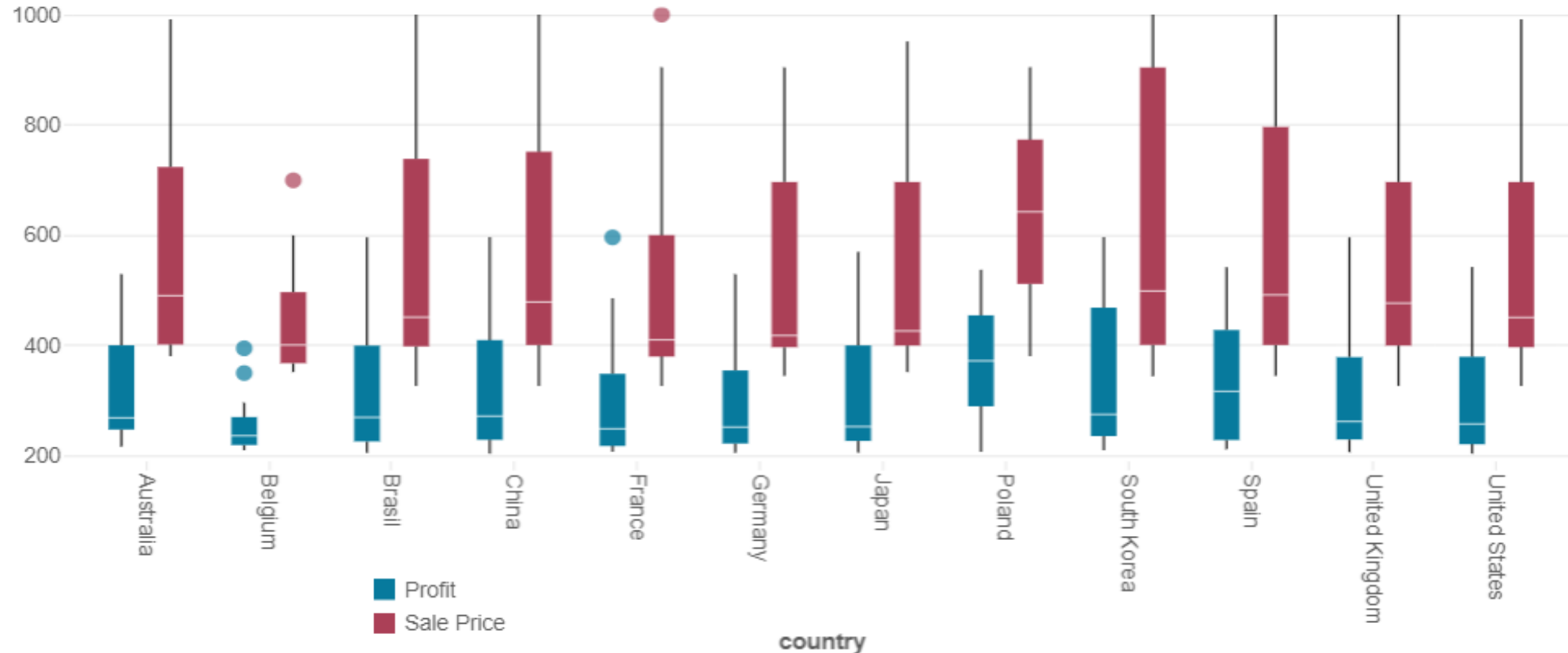
Q: In which country does we have to use a specific strategy to make more profits?
A:
- South Korea. There's a longer box which means a dispersed range of sales profit and prices, potentially due to factors like intense market competition or demand fluctuations.
- Belgium. Most of the orders are in lower price and profit than other countries, only few of the orders are as high as others. We can treat these users in different strategy.

4. Orders and Profits (2/2)

Profit by country and age
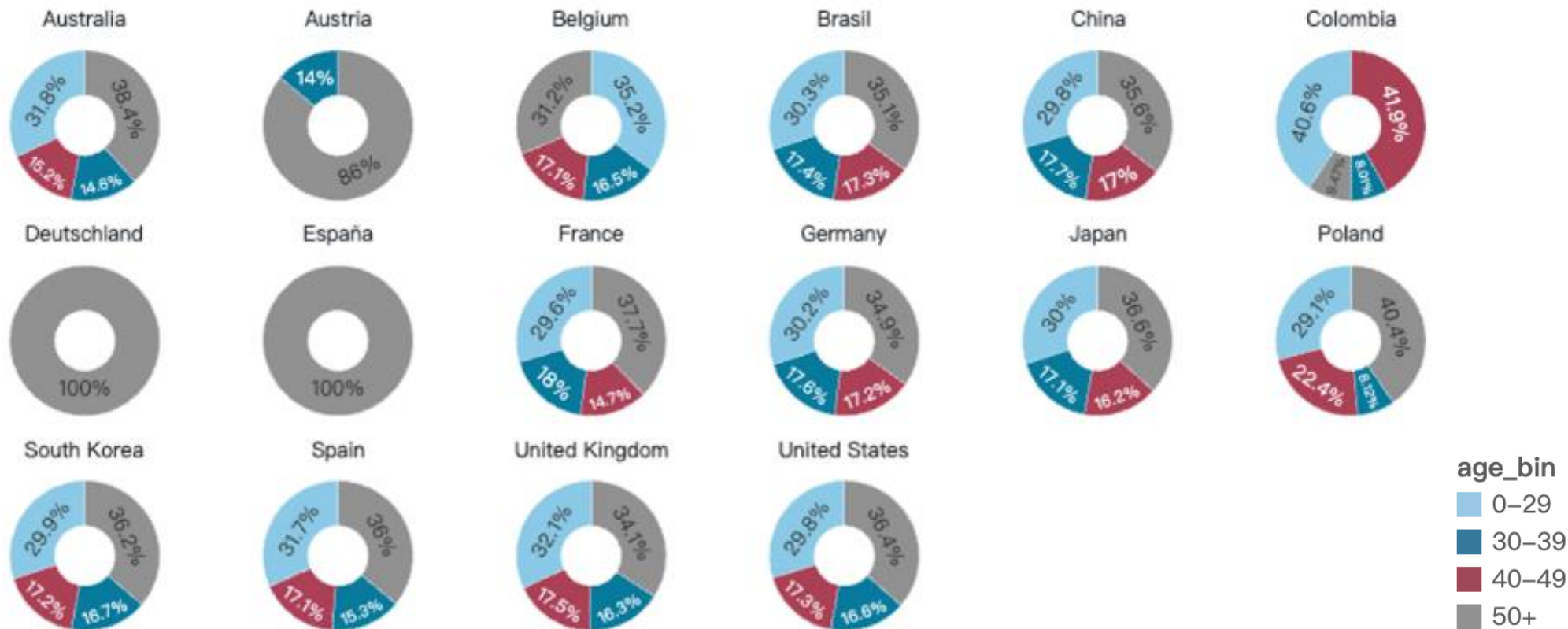
Q: In which country the users are in a specific age proportion?
A:
- Deutschland and España: The users are all over 50 years old.
- Austria: 86% of users are over 50 years old, and 14% are between 30 and 39 years old.
- Colombia: The only country where more than 40% of users are aged 40 to 49.

Conclusion

This ecommerce data set from the Looker team provides a large amount of information.

In a limited time, we tried to point out key insights such as users, events, profits, traffic source conversions.

However, there are different dimensions of information waiting to be mined.

We are excited to dig out more valuable insights after this project!

By students **Hui Yi Huang** and **Sara Canu.**