{Propulsion}

How to differentiate?

Expedia

Mathieu Stremsdoerfer

July 11, 2018

Your profile and background information.

UP TO NOW

- INSEAD MBA in Marketing. ESCP Master in Economics, Technology & Org.
- International experience in digital, analytics, market research
- Serial entrepreneur and intrapreneur.
- Head of pricing and market research at Syngenta Lawn & Garden division

NOW:

- Student at Propulsion

FUTURE

- Connect data science with business

The business model

38% Direct
27% Referrals
Trivago, Tripadvisor
15% Google search
9% Paid search
7% Mail
2% Social

Expedia

Revenue drivers:
1 Booking rate
2 Package
3 ADR: average daily rate
4 LOS: length of stay
5 Services

Flight booking
$ ~3%

Hotel booking
$$ 10-15%

Car rental
$ 5%

Services
$ 5%

The challenge

IT IS GETTING CROWDED OUT THERE…

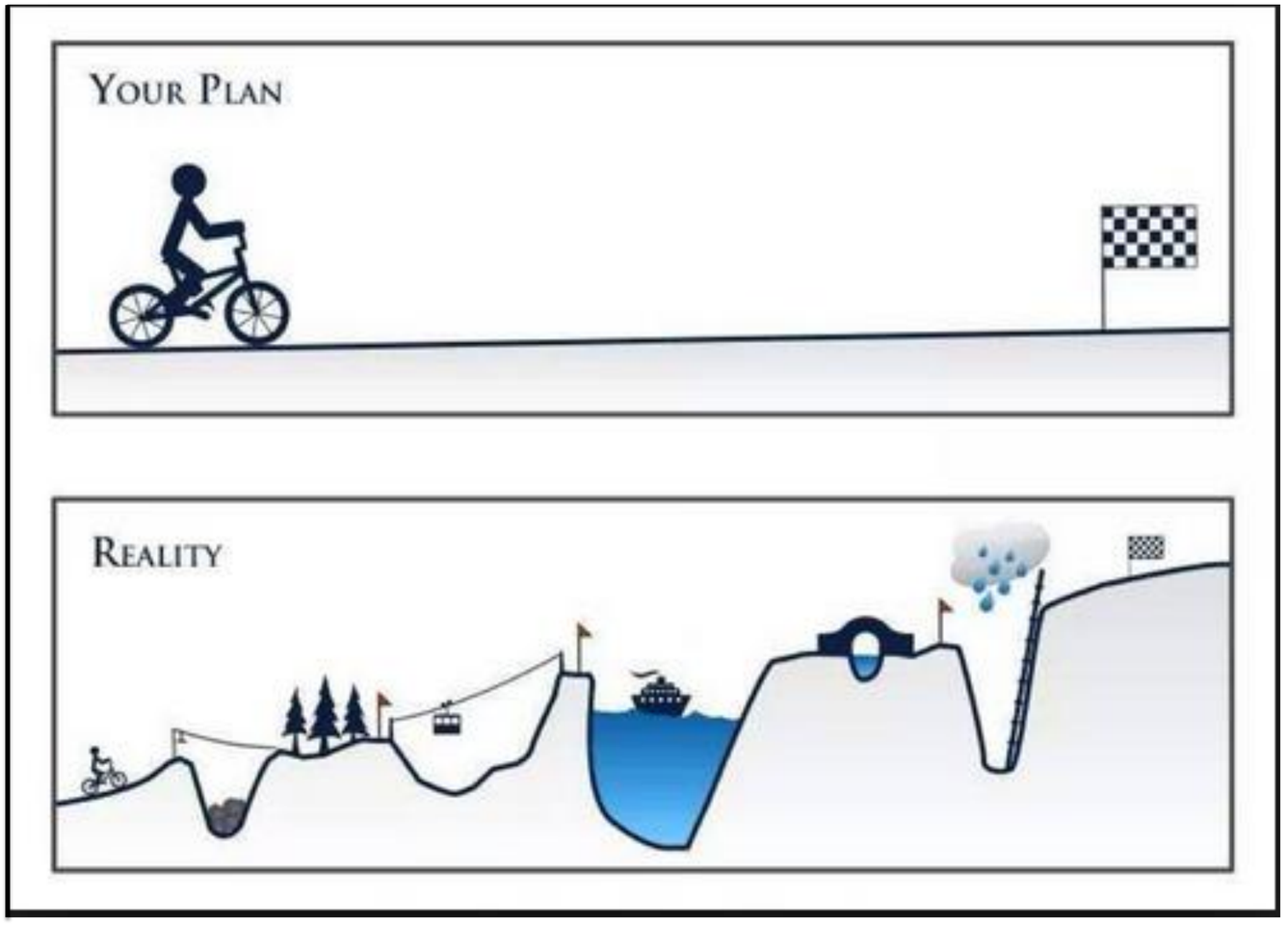
Ctrip Google Expedia airbnb tripadvisor Booking.com

GOAL:

“Segment hotel booking customers, identify drivers of booking and try to predict booking vs churn”

Scope: US travellers who travel in the US and Rest of the World (ROW)

Approach



Approach

1st week

Understand the market

Understand the data 

Generate business questions
Create plan

2nd week

Create features, target variable
Clean data as needed, get third party data
Explore new data
Analyze correlations

Learn selenium

Reinstall Anaconda

Learn virtual environments

3rd week

Create visualization

Time consuming

Learn new tools

2 days left!!! ZEN

Create booking predictions

Expedia

Add your property | Account | My Lists | Support | Español | 简体中文

Feature creation: build them from the sites entries

Flights | Hotels | Bundle and Save | Cars | Cruises | Things to Do | Vacation Rentals | Discover

Going to
New York, New York

Check-in
07/14/2018

Check-out
07/18/2018

Travelers
2 Adults, 0 Children, 1 Room

Add a flight Add a car

Search

Examples:

Length Of Stay (LOS) = Date check-out – date check-in

Booking window (BW) = Date check-in – date searched

Business: = search (1 adult – 0 children during work days and hours) & travel (Monday- Friday)

“**Bleisure”:** search (1 adult – 0 children during work days and hours) & travel (Monday-Sunday)

TOTAL: 35 new features in 9 categories

Definition: key binary variables

- **Booking:** for a specific user session,
  Booking is 1 [hotel booked] or 0 [hotel not booked]

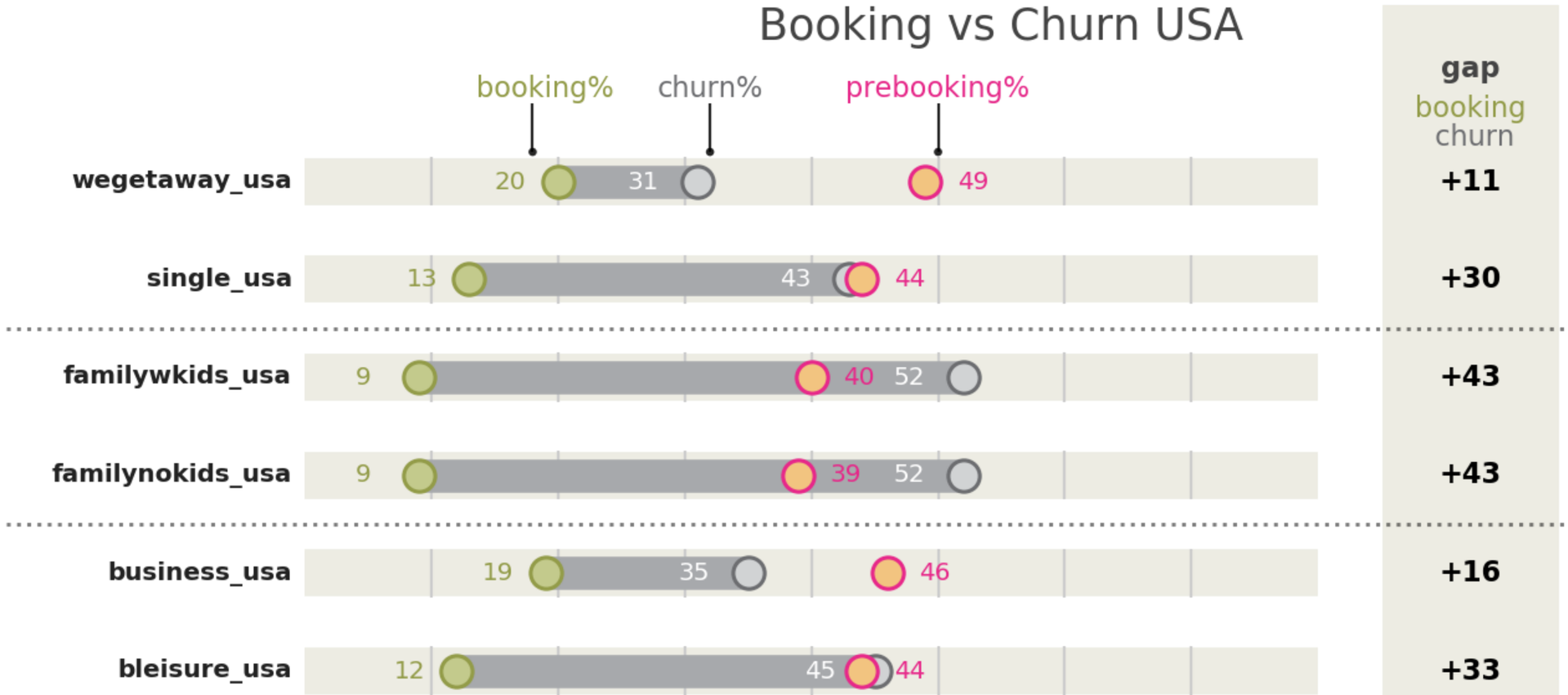
- **Churn:** for a specific user ID, destination (season)
  Churn is 1 [sum of Booking = 0] else 0

- **Prebooking:** for a specific user session (season)**:**
  neither a churn nor a booking ~a visit that will ultimately lead to a booking

Prebooking = 1 - booking - churn

Realized with plotline, ggplot

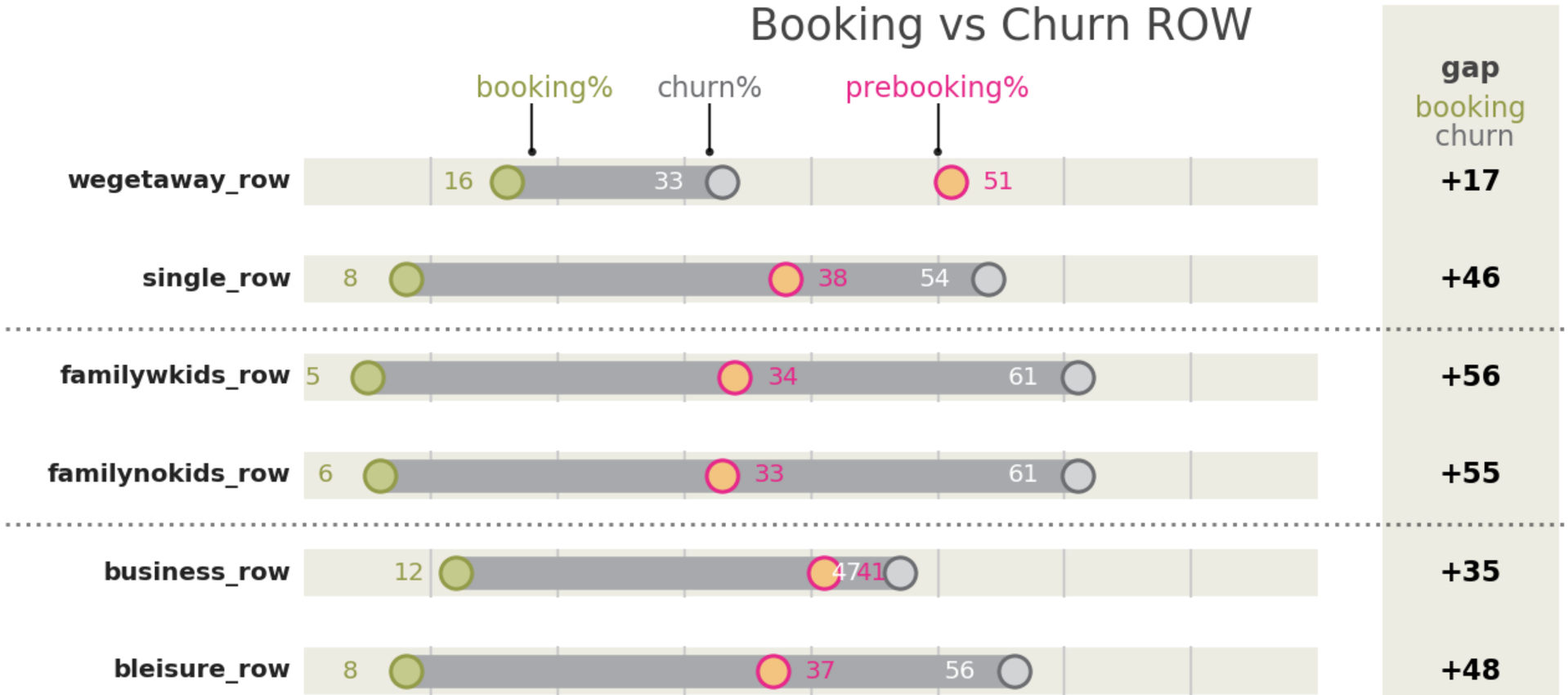
User segments observations



#1: User segments: could merge families into one. Family underperform in booking %
#2: Booking positively correlated with prebooking

Realized with plotline, ggplot

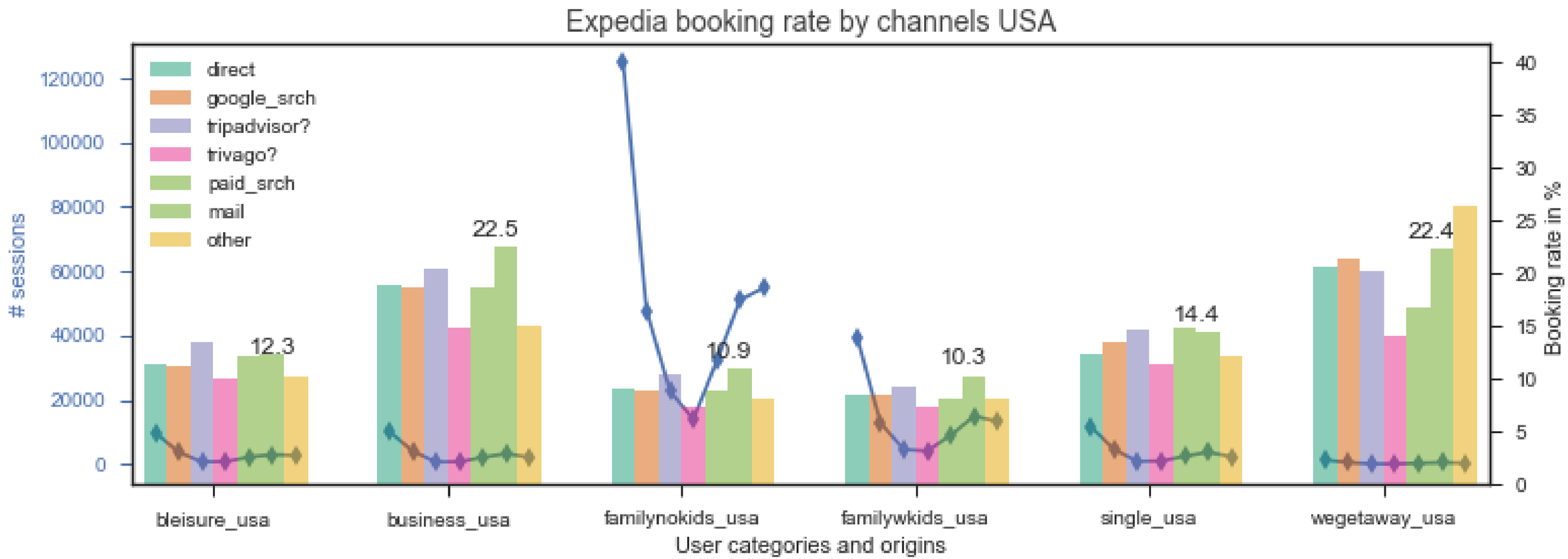
User segments observations



3: Same trends as USA, but lower booking rates overall

Realized with plotline, ggplot

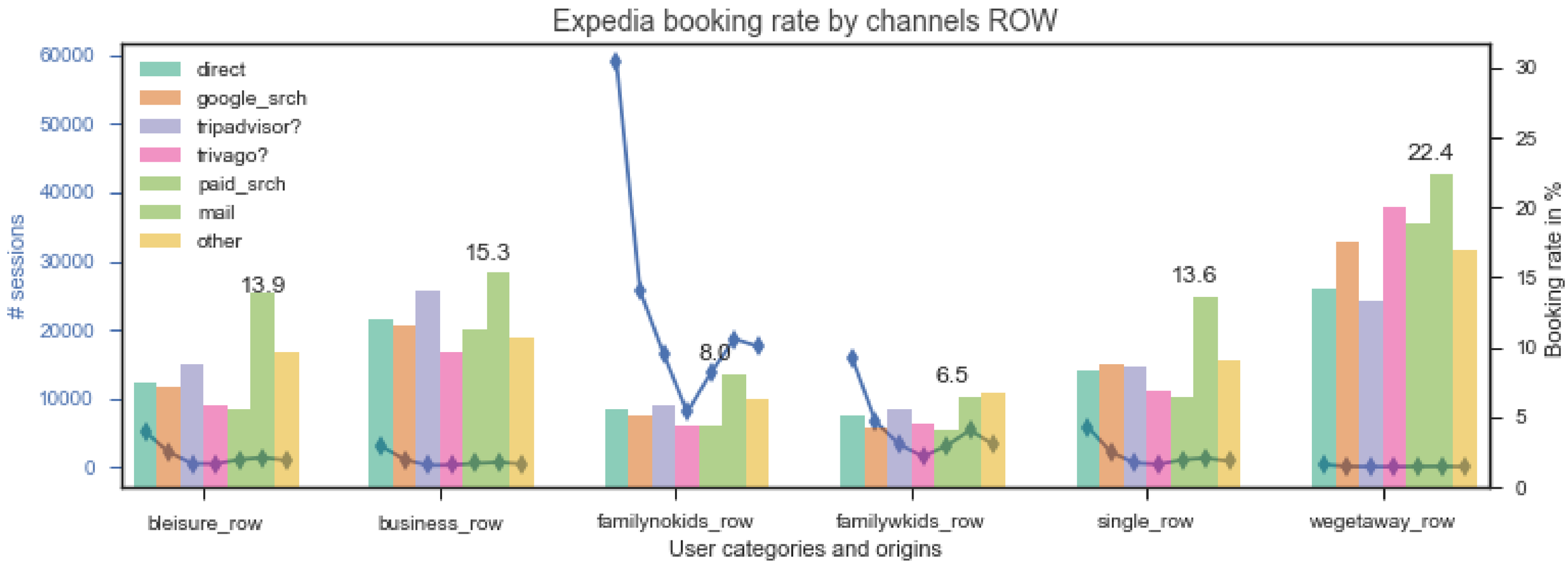
Booking rate by channels and user segments



4: Channels: Volume (#1 direct, # 2: Google) ; Booking rate (#1 mail)
5: User segments: Volume (# 1 family no kids, # 2 family w kids) ; BR (#1 We getaways)

Realized with matplotlib

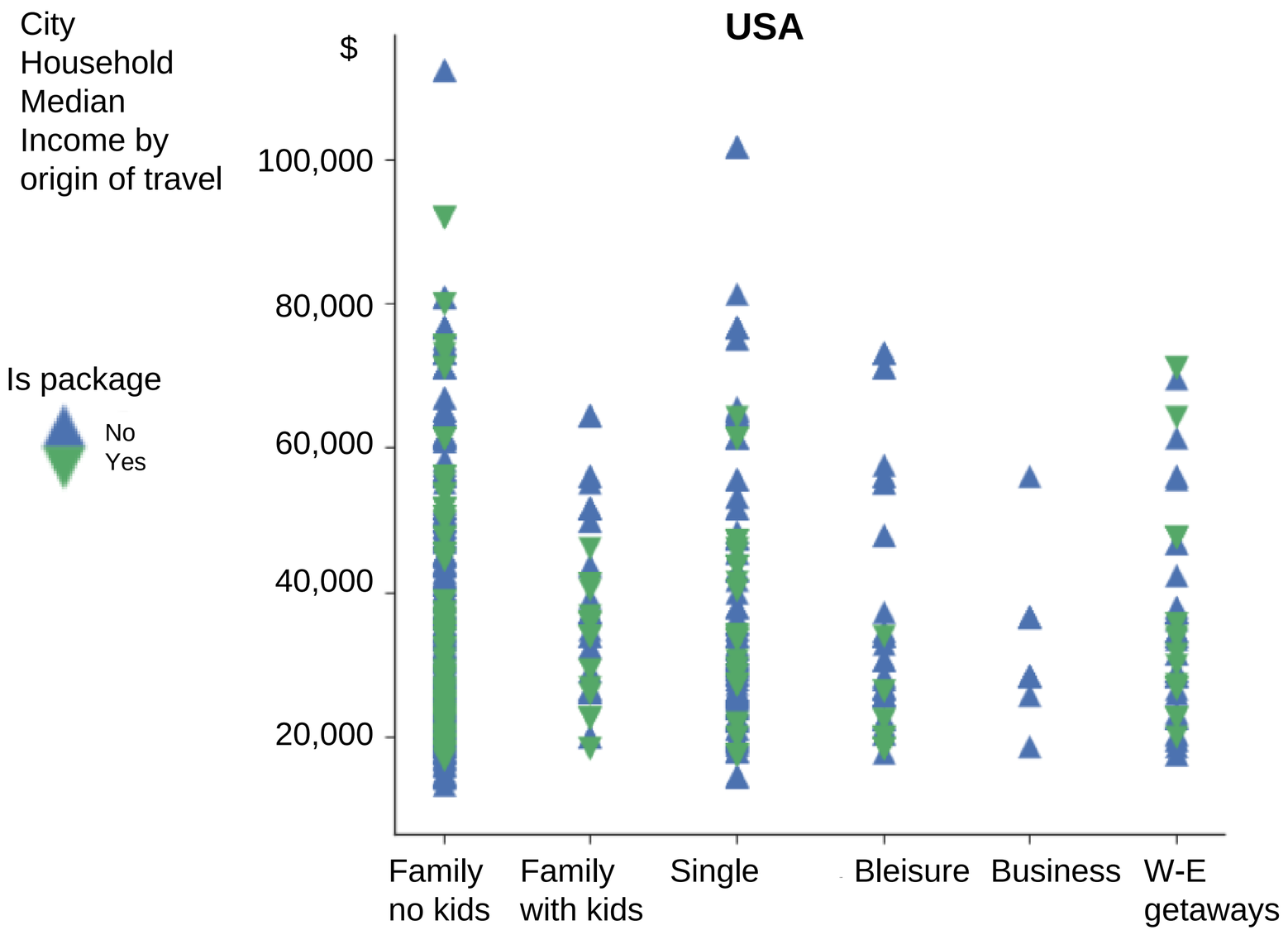
Booking rate by channels and user segments



6 traffic ~ ½ of USA
7: similar trends as USA

Realized with matplotlib

Income sensitivity versus packages

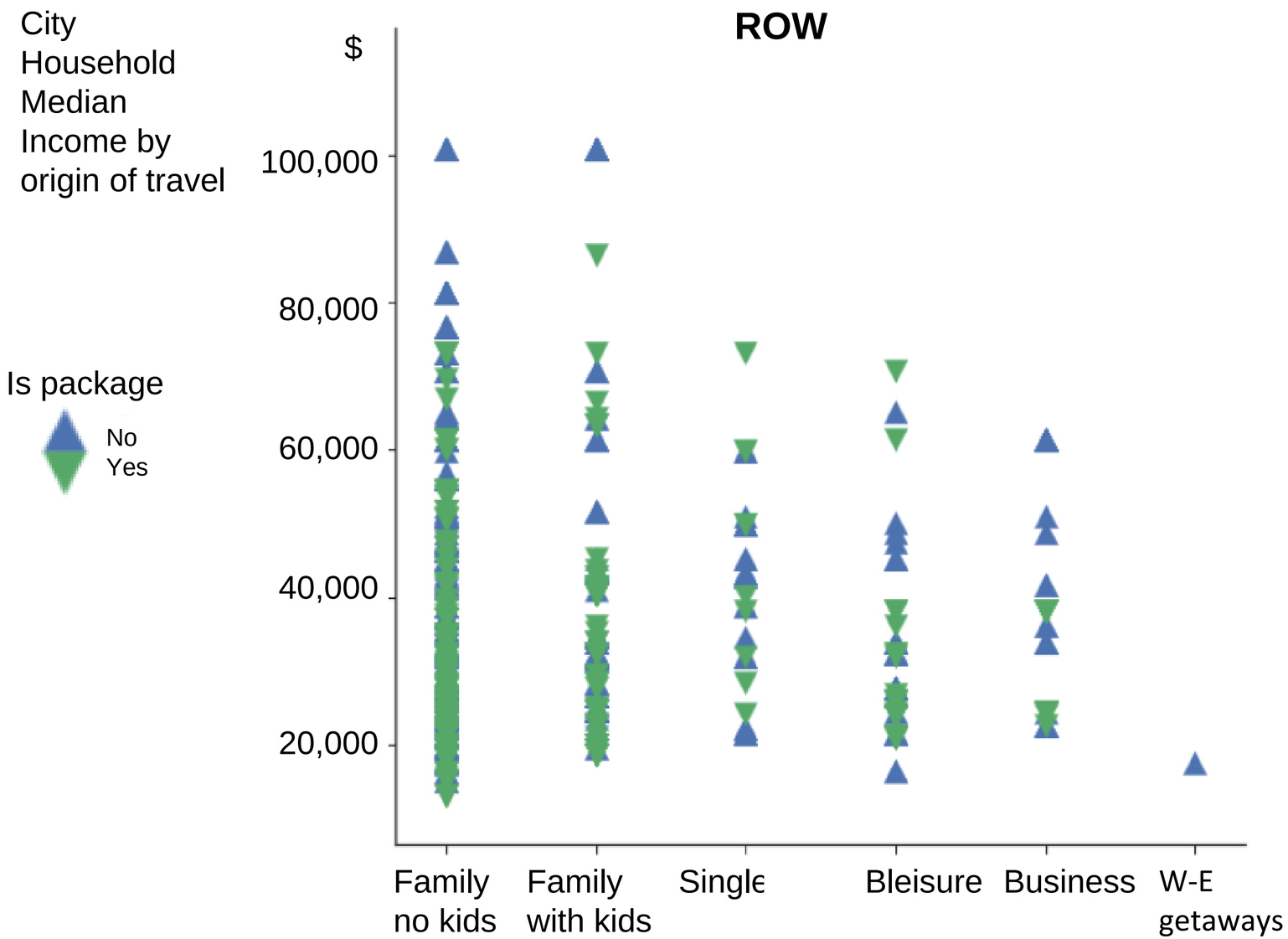


8: Clear opportunities for packages for budget travelers

Opportunity to geo-localize campaigns based on city Household income

Realized with Seaborn using a random sample with household median income data scrapped

Income sensitivity versus packages

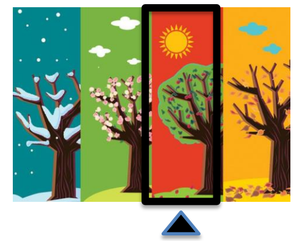


9: similar findings for international travellers

Note:
Same correlation not found with hotel stars and price bands

Realized with Seaborn using a random sample with household median income data scrapped

Seasonality effect



Total: 168160
Direct 44%

10: campaign overspend?

Tripadvisor 16%

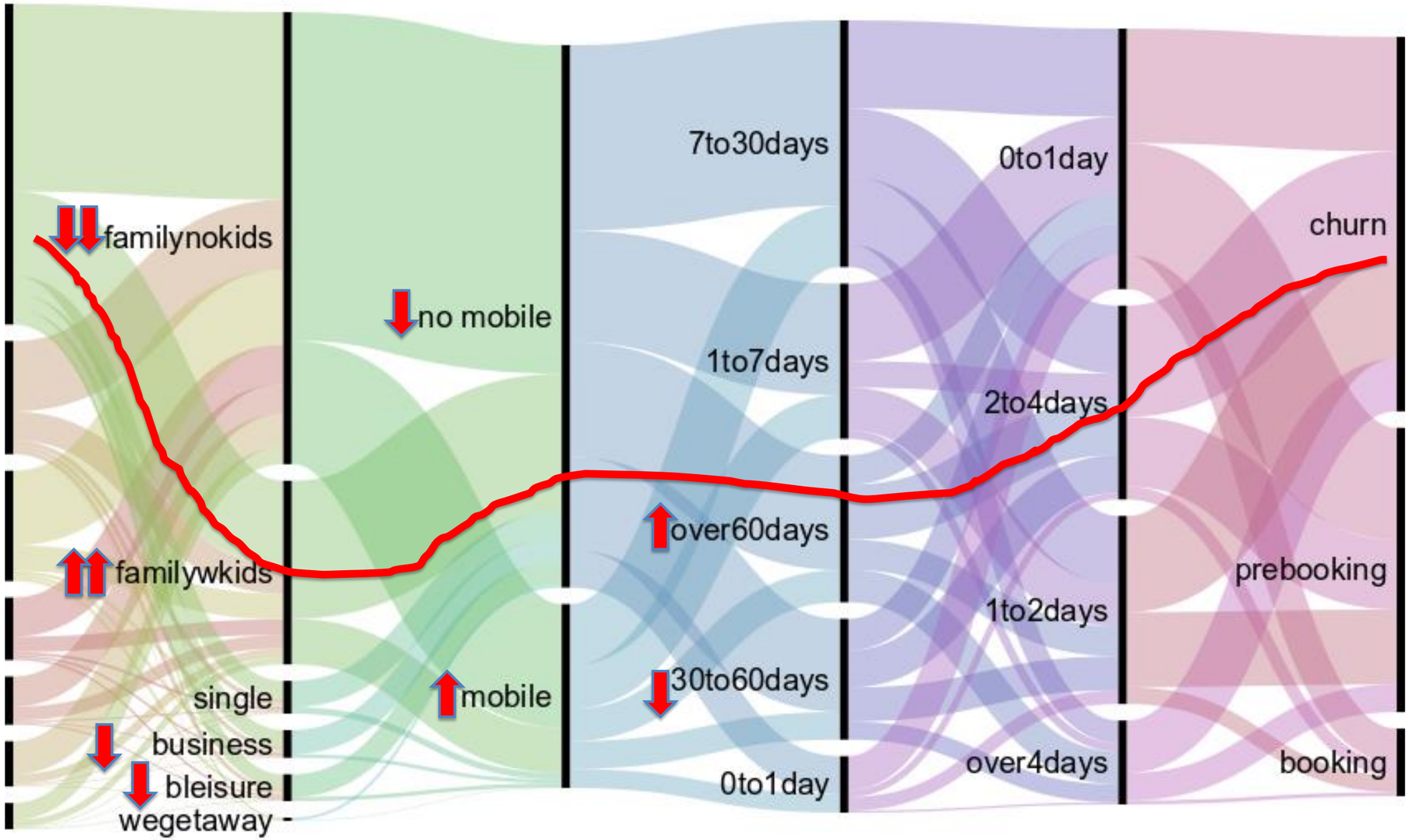
Trivago 15%

Paid search 9%

Google 7%

mail 6%

other 4%



Realized with rawgraphs

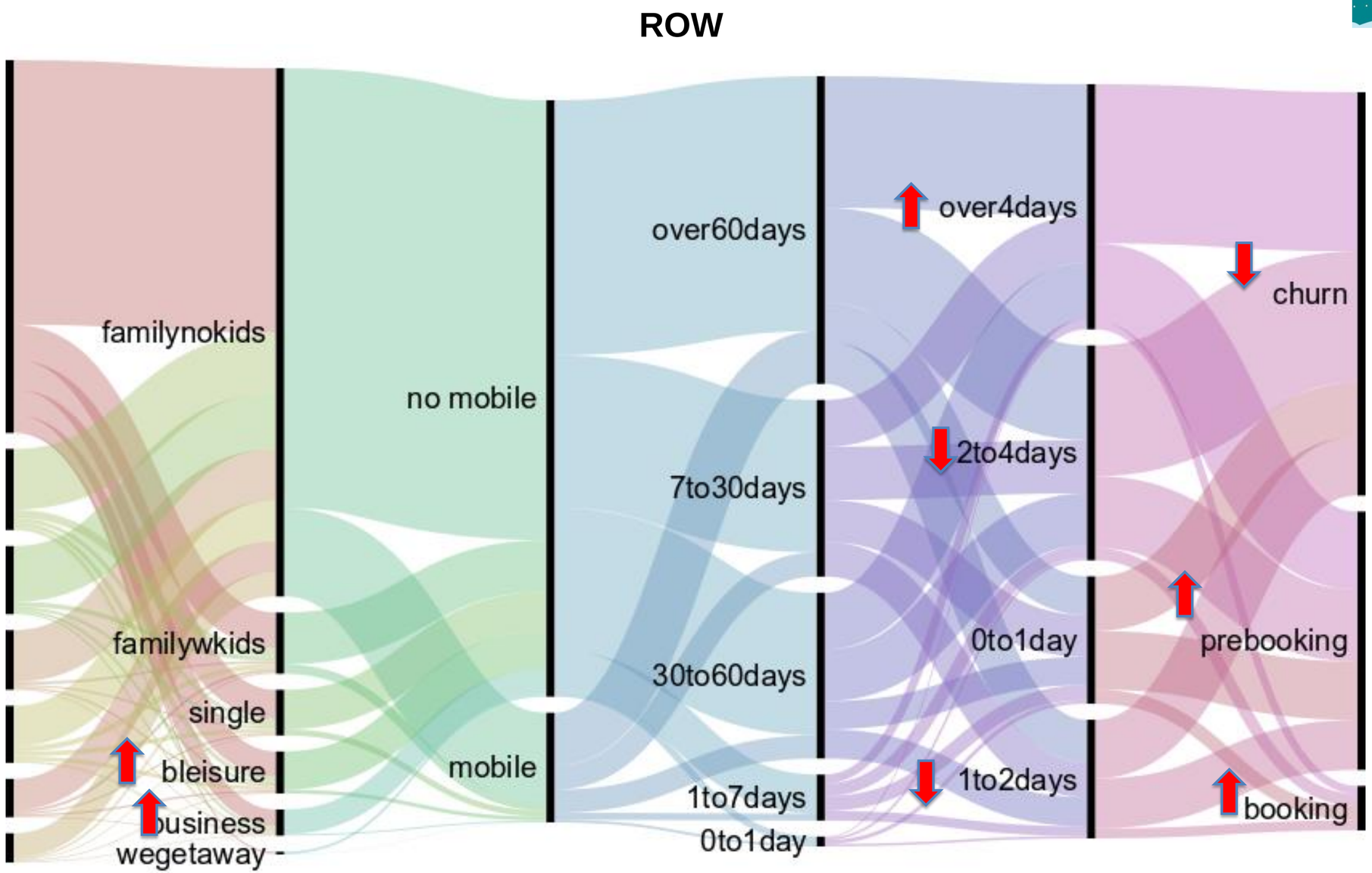
Seasonality effect

Realized with rawgraphs

5. Seasonality effect

Total: 55295
Direct: 53%

12: campaign overspend?



Tripadvisor: 11%

Trivago: 10%

Mail: 8%

Paid: 8%
Google: 5%

Other:4%

Realized with rawgraphs

Hotel booking predictions (a work in progress)

Methodological Approach

Business problem
We want to predict if a user will **churn, prebook, or book in a given seas**on

Definition of the problem
A supervised –**multi class** classification problem with imbalanced classes and mixed variables (binary, numeric, categorical) with time series data

Preparation
- Divide dataset into **four seasons**
- Split train and test

Feature selection (kstratified folds, Man Whitney , collinearity, ridge regression)

Modelling: Logistic regression, KNN, Random Forest, Naïve Bayes, SVM, Gradient Boost,

Conclusions:

Quick win:

- Define user categories and session features per season to better understand needs
- Geo target campaigns based on IP/income – user categories
- Increase mobile offerings for last minute bookings

Tactical:

- Investigate Expedia seasonal campaign spent for possible optimization
- **Use booking/churn/prebooking prediction to identify “real churn” and offer help and attractive offers**

Strategic:

- Investigate root causes of low prebooking and BR for family and ROW

ex: is Expedia hotel offering adequate for families?

ex: should Expedia develop planning tools for ROW travelers?

Outlook

To go further:

- Complete multiclass classification
- Inquire user category stability over time and user loyalty
- Price data (quantitative) and multi year would be helpful for predictions

Perspective: marketing themes also applicable to other businesses:

- Fit for purpose segmentation
- Traffic drivers - channels
- Conversion efficiency: churn vs booking vs prebooking
- Price sensitivity - geo localization
- Seasonality
- Local vs international
- Campaign spend efficiency

Thank you!

EVEN THE

GREATEST WAS

ONCE A

BEGINNER. DON'T

BE AFRAID

TO TAKE THAT

FIRST STEP.

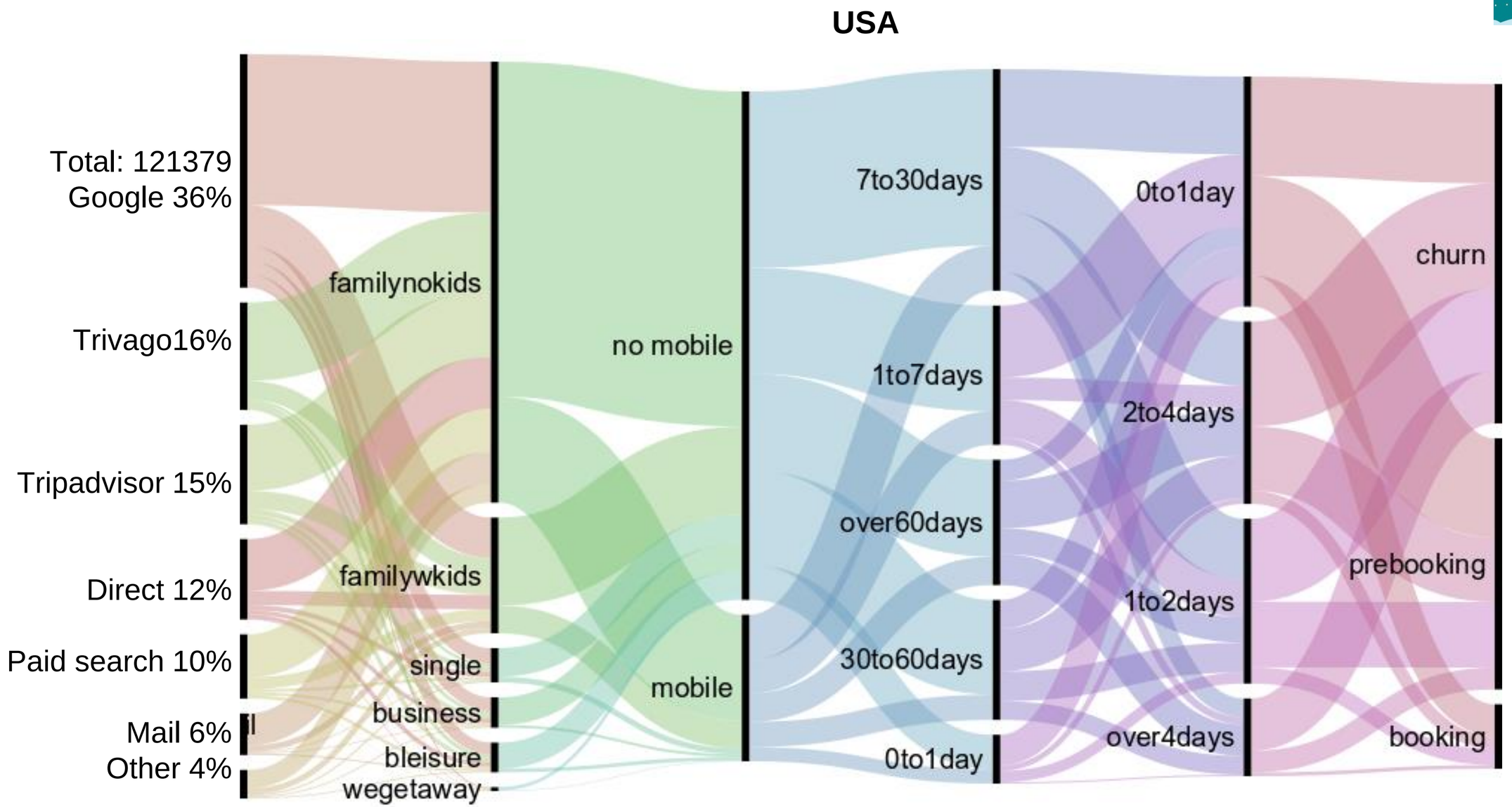
Mathieu Stremsdoerfer

mstrems@gmail.com

www.linkedin.com/in/mstrems/

APPENDIX

Seasonality effect

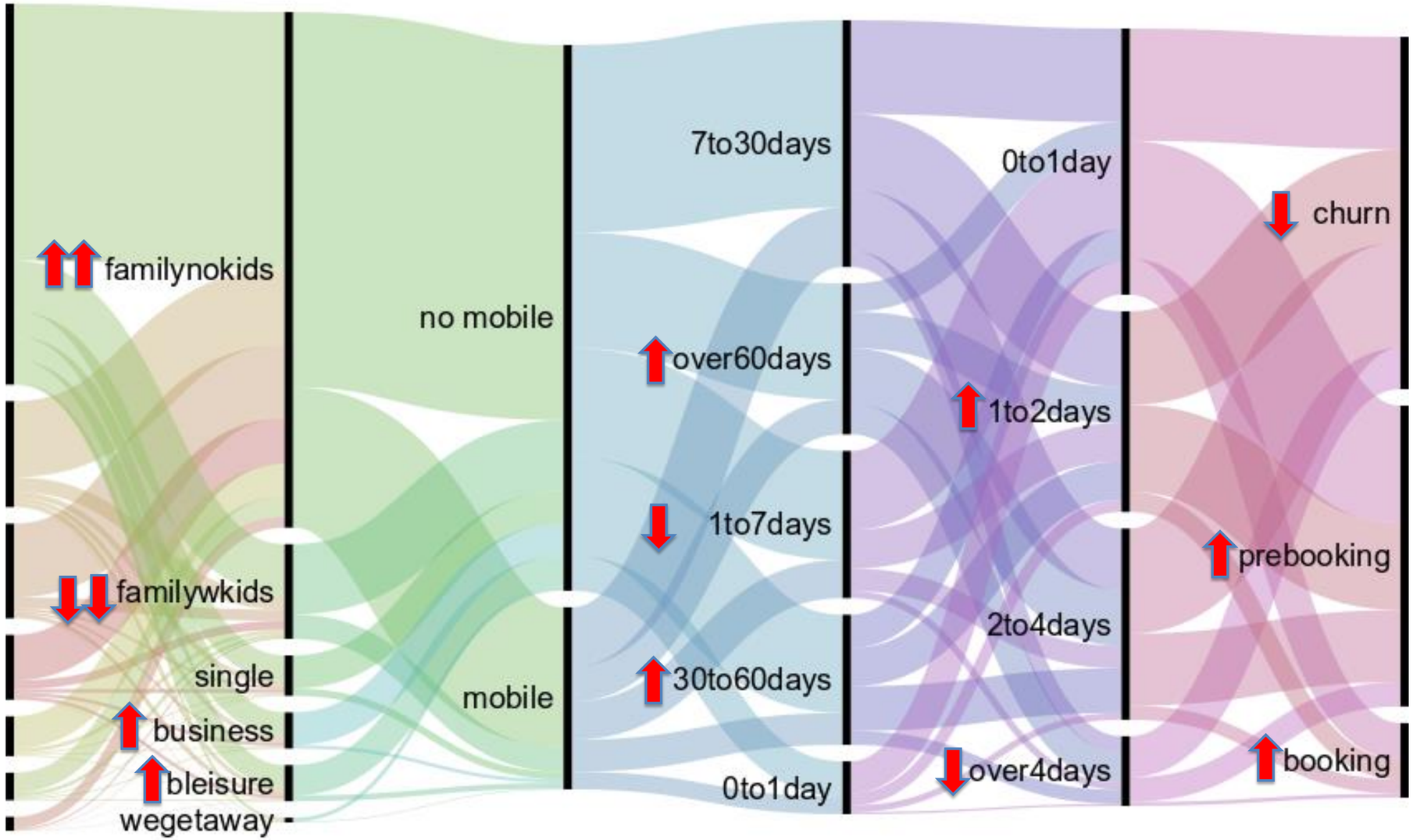


Realized with rawgraphs

Seasonality effect

Total 127683
Direct 52%

10: campaign overspend?



Tripadvisor 15%

Trivago 13%

Paid search 9%

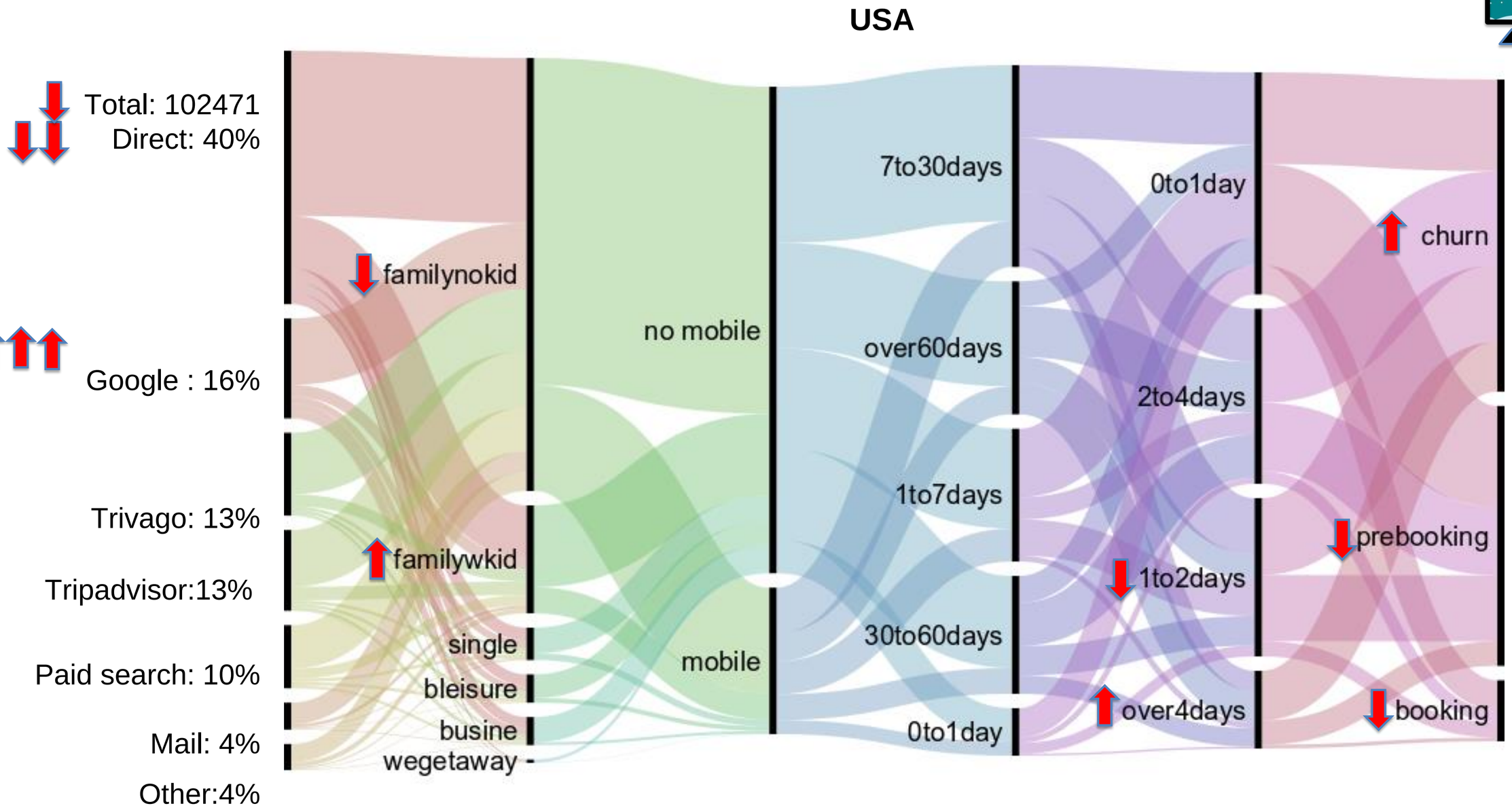
Mail 5%

Other 4%

Google 2 %

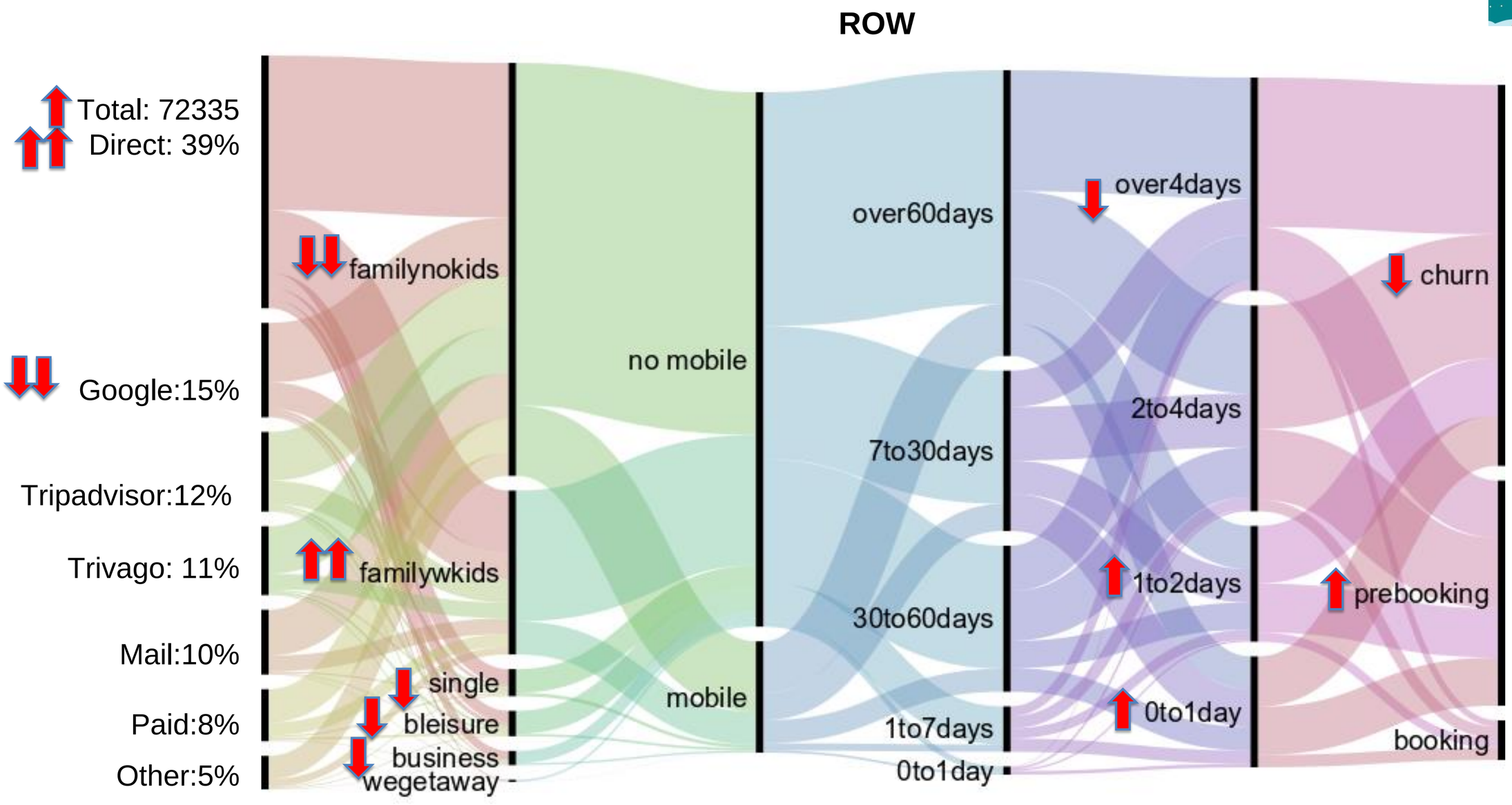
Realized with rawgraphs

Seasonality effect

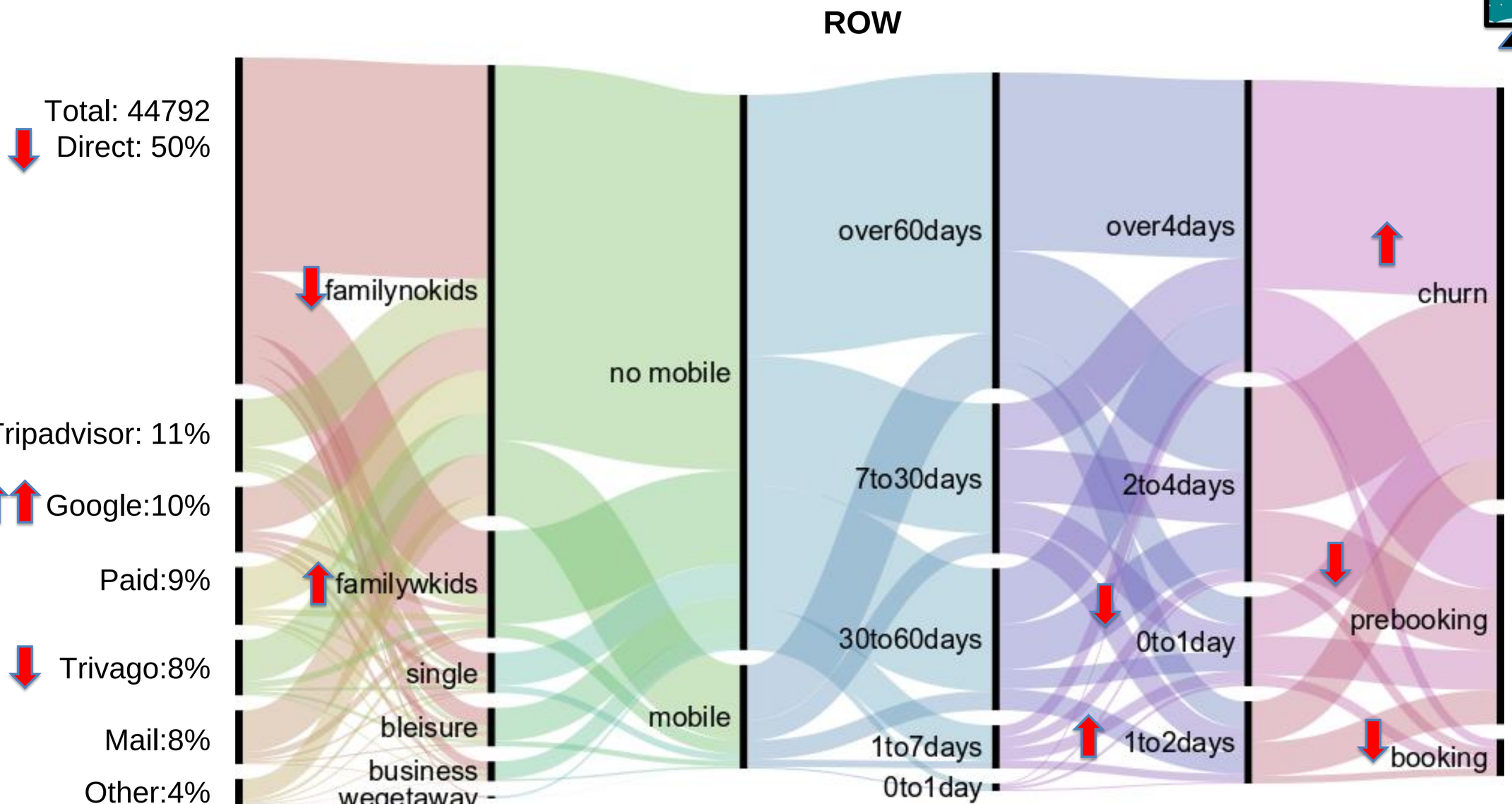


Realized with rawgraphs

Seasonality effect



Seasonality effect



Realized with rawgraphs