Codebook for dataset

Autogenerated data summary from dataMaid

2019-12-08 15:50:56

Data report overview

The dataset examined has the following dimensions:

Feature	Result
Number of observations	275547
Number of variables	15

Codebook summary table

Label	Variable	Class	# unique values	Missing	Description
	date_account_created	Date	1726	0.00 %	the date of account creation
	timestamp_first_active	numeric	275547	0.00 %	timestamp of the first activity
	date_first_booking	Date	1977	67.73 %	date of first booking
	gender	character	4	0.00 %	the gender of the user
	age	numeric	86	42.41 %	the age of the user
	signup_method	character	4	0.00 %	the sign up method of the user
	signup_flow	numeric	18	0.00 %	the page a user came to signup up from
	language	character	26	0.00 %	international language preference
	affiliate_channel	character	8	0.00 %	what kind of paid marketing
	affiliate_provider	character	18	0.00 %	where the marketing is e.g. google, craigslist, other
	signup_app	character	4	0.00 %	the signup app of the user
	first_device_type	character	9	0.00 %	the device type used in user's first activity
	first_browser	character	55	0.00 %	the browser used in user's first activity
	country_destination	character	13	0.00 %	the destination country where user booked the place to stay
	book	factor	3	0.00 %	the booking status of user

Variable list

date_account_created

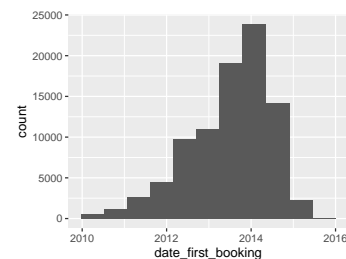| Feature	Result
Variable type	Date
Number of missing obs.	0 (0 %)
Number of unique values	1726
Mode	"2014-07-23"
Min. and max.	2010-01-01; 2014-09-30
1st and 3rd quartiles	2013-04-03; 2014-06-18
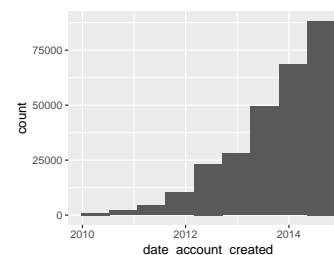
timestamp_first_active

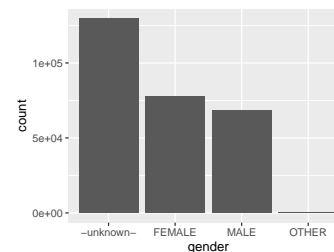| Feature	Result
Variable type	numeric
Number of missing obs.	0 (0 %)
Number of unique values	275547
Median	20140103192726
1st and 3rd quartiles	20130403195140; 20140618184764
Min. and max.	20090319043255; 20140930235901
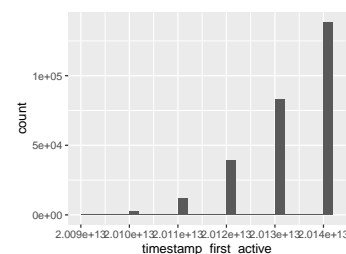
date_first_booking

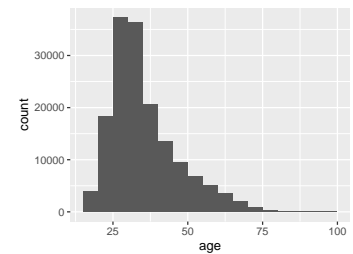| Feature	Result
Variable type	Date
Number of missing obs.	186639 (67.73 %)
Number of unique values	1976
Mode	"2014-05-22"
Min. and max.	2010-01-02; 2015-06-29
1st and 3rd quartiles	2012-12-02; 2014-04-04

gender

Feature	Result
Variable type	character
Number of missing obs.	0 (0 %)
Number of unique values	4
Mode	"-unknown-"

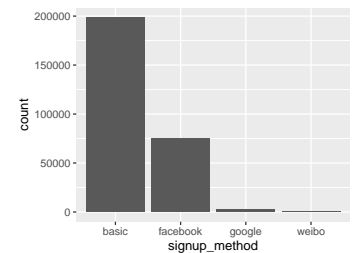- Observed factor levels: "-unknown-", "FEMALE", "MALE", "OTHER".

age

Feature	Result
Variable type	numeric
Number of missing obs.	116866 (42.41 %)
Number of unique values	85
Median	33
1st and 3rd quartiles	28; 41
Min. and max.	15; 99

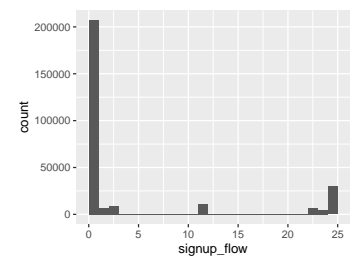## signup_method

Feature	Result
Variable type	character
Number of missing obs.	0 (0 %)
Number of unique values	4
Mode	"basic"

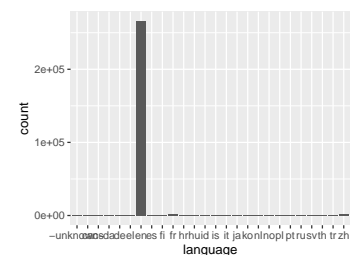- Observed factor levels: "basic", "facebook", "google", "weibo".

signup_flow

Feature	Result
Variable type	numeric
Number of missing obs.	0 (0 %)
Number of unique values	18
Median	0
1st and 3rd quartiles	0; 1
Min. and max.	0; 25

language

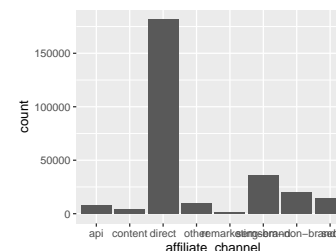| Feature	Result
Variable type	character
Number of missing obs.	0 (0 %)
Number of unique values	26
Mode	"en"

- Observed factor levels: "-unknown-", "ca", "cs", "da", "de", "el", "en", "es", "fi", "fr", "hr", "hu", "id", "is", "it", "ja", "ko", "nl", "no", "pl", "pt", "ru", "sv", "th", "tr", "zh".
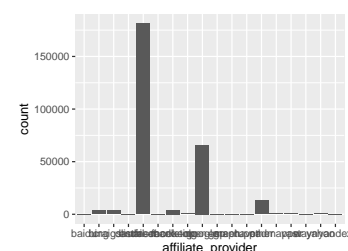
affiliate_channel

Feature	Result
Variable type	character
Number of missing obs.	0 (0 %)
Number of unique values	8
Mode	"direct"

- Observed factor levels: "api", "content", "direct", "other", "remarketing", "sem-brand", "sem-non-brand", "seo".
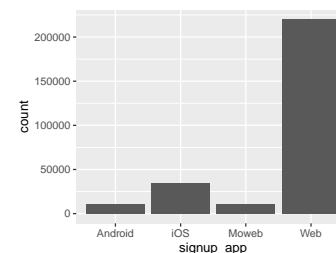
affiliate_provider

Feature	Result
Variable type	character
Number of missing obs.	0 (0 %)
Number of unique values	18
Mode	"direct"

- Observed factor levels: "baidu", "bing", "craigslist", "daum", "direct", "email-marketing", "facebook", "facebook-open-graph", "google", "gsp", "meetup", "naver", "other", "padmapper", "vast", "wayn", "yahoo", "yandex".
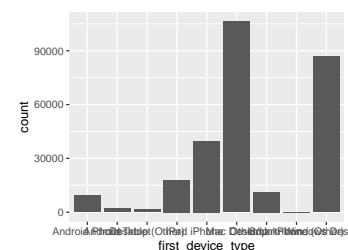
signup_app

Feature	Result
Variable type	character
Number of missing obs.	0 (0 %)
Number of unique values	4
Mode	"Web"

- Observed factor levels: "Android", "iOS", "Moweb", "Web".

first_device_type

Feature	Result
Variable type	character
Number of missing obs.	0 (0 %)
Number of unique values	9
Mode	"Mac Desktop"

- Observed factor levels: "Android Phone", "Android Tablet", "Desktop (Other)", "iPad", "iPhone", "Mac Desktop", "Other/Unknown", "SmartPhone (Other)", "Windows Desktop".
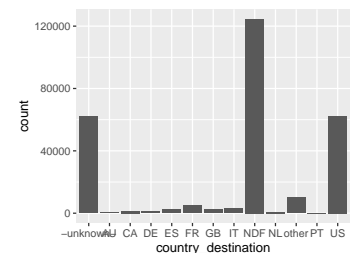
first_browser

Feature	Result
Variable type	character
Number of missing obs.	0 (0 %)
Number of unique values	55
Mode	"Chrome"
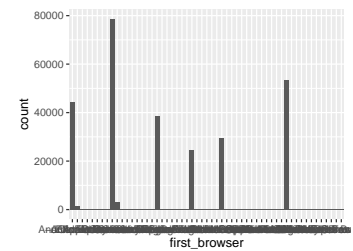
- Observed factor levels: "-unknown-", "Android Browser", "AOL Explorer", "Apple Mail", "Arora", "Avant Browser", "BlackBerry Browser", "Camino", "Chrome", "Chrome Mobile", "Chromium", "CometBird", "Comodo Dragon", "Conkeror", "CoolNovo", "Crazy Browser", "Epic", "Firefox", "Flock", "Google Earth", "Googlebot", "IBrowse", "IceDragon", "IceWeasel", "IE", "IE Mobile", "Iron", "Kindle Browser", "Maxthon", "Mobile Firefox", "Mobile Safari", "Mozilla", "NetNewsWire", "Nintendo Browser", "OmniWeb", "Opera", "Opera Mini", "Opera Mobile", "Outlook 2007", "Pale Moon", "Palm Pre web browser", "PS Vita browser", "RockMelt", "Safari", "SeaMonkey", "Silk", "SiteKiosk", "SlimBrowser", "Sogou Explorer", "Stainless", "TenFourFox", "TheWorld Browser", "UC Browser", "wOSBrowser", "Yandex.Browser".
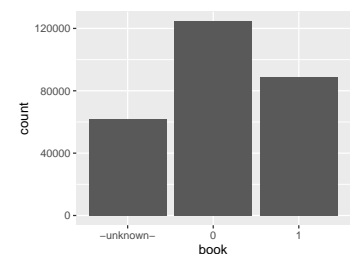
country_destination

Feature	Result
Variable type	character
Number of missing obs.	0 (0 %)
Number of unique values	13
Mode	"NDF"

- Observed factor levels: "-unknown-", "AU", "CA", "DE", "ES", "FR", "GB", "IT", "NDF", "NL", "other", "PT", "US".

book

Feature	Result
Variable type	factor
Number of missing obs.	0 (0 %)
Number of unique values	3
Mode	"0"
Reference category	-unknown-

- Observed factor levels: "-unknown-", "0", "1".

Report generation information:

- Created by Junhua Tan (username: `JunhuaTan`).

- Report creation time: Sun Dec 08 2019 15:51:29

- Report was run from directory: /Users/JunhuaTan/Desktop/College/FA2019/STATS301/Final Project

- dataMaid v1.3.2 [Pkg: 2019-07-27 from CRAN (R 3.6.0)]

- R version 3.6.1 (2019-07-05).

- Platform: x86_64-apple-darwin15.6.0 (64-bit)(macOS Catalina 10.15).

- Function call: `makeDataReport(data = dataset, mode = c("summarize", "visualize", "check"), smartNum = FALSE, file = "codebook_dataset.Rmd", replace = TRUE, checks = list(character = "showAllFactorLevels", factor = "showAllFactorLevels", labelled = "showAllFactorLevels", haven_labelled = "showAllFactorLevels", numeric = NULL, integer = NULL, logical = NULL, Date = NULL), listChecks = FALSE, maxProbVals = Inf, codebook = TRUE, reportTitle = "Codebook for dataset")`