

Feature Mart Data Dictionary



The Mobilewalla Feature Mart is a collection of sophisticated, highly predictive features that help data scientists improve the feature engineering process and build more accurate and predictive machine learning models. Features in the Mobilewalla Feature Mart fit into nice broad characteristic categories:

App Engagement

Features related to the apps being used on the device such as most seen app category, number of distinct apps used by the device, number of premium apps used by the device etc.

2 Behavior

Features based on behavior and location-based data including an engagement score with luxury brands and behavior related to different types of POIs like travel, transportation, retail, food, and business services.

3 Carrier

Features based on device carrier like signal distribution by telco type, last seen cellular carrier and most seen WiFi carrier.

4 Demographics

Features like user age and gender.

5 Device Engagement

Device profiling based on usage time like night riders, commuters, early risers, weekday, and weekend engagements.

6 Device Mobility

Features like area of mobility, average distance travelled in a day, total number of distinct locations visited by device, home location, work location, commute distance etc.

7 Householding

Features focused on related devices and device attributes for users at that same location.

8 Phone

Features like phone model, phone age, device type, platform etc.

9 Relationship Network

Features like work and social engagement and relationships, etc.

Custom Features:

Apart from the features listed in our data dictionary, Mobilewalla can also generate custom features based on specific customer requirements.

The Mobilewalla Feature Mart has been built and tested based on data aggregated from multiple data sources over anextended period andis refreshed every month based on data aggregated for the past 90 days.



Data Dictionary for Feature Mart

For better understanding of Mobilewalla features, the dictionary has been split in categories. We have all the features mapped to IFA (identifier for advertisers) which is the device ID or MAID (mobile advertiser ID).

App Engagement Features

Field Name	Data type	Feature Definition	Example
norm_distinct_app	Double	Normalized distinct apps for devices in the country	0.103448276
most_seen_ appcategory	Array	App category with most received mobile signals	[SOCIAL]
apps_used	Array	Array of bundle IDs/ASN observed for the device	[bubbleshooter.orig, cn.wps.moffice_eng, com.lenovo. anyshare.gps, com. parallel.space.lite]
norm_paid_app	Double	Normalized paid app engagement score for devices in the country	0
norm_avail_app	Double	Normalized engagement score for all apps found in devices in the country	0.004975124
norm_premium_app	Double	Normalized engagement score with the premium apps for devices in the country	0
global_books_and_ reference	Decimal	Device engagement across 38 global app categories	0
global_business	Decimal		0
global_education	Decimal		0
global_entertainment	Decimal		0
global_finance	Decimal		0
global_food_and_drink	Decimal		0
global_games	Decimal		0.10102
global_health_and_ fitness	Decimal		0
global_lifestyle	Decimal		0
global_maps_and_ navigation	Decimal		0
global_medical	Decimal		0
global_music	Decimal		0
global_news_and_	Decimal		0

magazines



Field Name	Data type	Feature Definition	Example
global_photography	Decimal	Device engagement across 38 global app categories	0
global_productivity	Decimal		0
global_shopping	Decimal		0
global_social	Decimal		0
global_sports	Decimal		0
global_tools	Decimal		0
global_travel	Decimal		0
global_weather	Decimal		0
global_game_action	Decimal		0.03439
global_game_ adventure	Decimal		0
global_game_arcade	Decimal		0
global_game_board	Decimal		0.034392
global_game_card	Decimal		0
global_game_casino	Decimal		0
global_game_casual	Decimal		0
global_game_ educational	Decimal		0
global_game_music	Decimal		0
global_game_puzzle	Decimal		0
global_game_racing	Decimal		0
global_game_role_ playing	Decimal		0.0243
global_game_ simulation	Decimal		0
global_game_sports	Decimal		0
global_game_strategy	Decimal		0
global_game_trivia	Decimal		0.034
global_game_word	Decimal		0



2 Behavior Features

Field Name	Data type	Feature Definition	Example
norm_premium_ behavioral_seg_ engagement	Double	Normalized engagement score with the luxury segments for devices in the country	0.071428571
norm_distinct_loc	Double	Normalized distinct location for devices in the country	0.269230769
home_zipcode	Array	Zip code for location where highest count of mobile signals distribution is recorded for the home part of the day	[190002]
work_zipcode	Array	Zip code for location where highest count of mobile signals distribution is recorded for the work part of the day	[190002]
most_seen_zipcode	Array	Zip code for location where highest count of mobile signals distribution is recorded for the day	[190002]
automotive_engagement	Double	Automobile POIs engagement score for the device	0
business_engagement	Double	Business services POIs engagement score for the device	0
food_and_beverage_ engagement	Double	Food Services POIs engagement score for the device	0
retail_engagement	Double	Retail POIs engagement score for the device	0
travel_engagement	Double	Travel POIs engagement score for the device	0
norm_premi- um_brand_engagement score	Double	Normalized High value brand engagement scores for the devices in the country	0
norm_premium_brands	Double	Normalized premium brand visit score for the devices in the country	0
norm_premium_brand days	Double	Cumulative number of days a device seen on premium brand locations normalized for a country	0



Field Name	Data type	Feature Definition	Example
cdl_geohash	String	Geohash for the latitude-longi- tude where work hours of the day have maximum signals. We observed all locations for a device & filter location which have been seen for "N" days or more for a country. These locations are then computed for getting common day locations	ttvprjtg
cdl_city	String	City of location where highest count of mobile signals distribution are recorded for the work part of the day. We observed all locations for a device & filter location which have been seen for "N" days or more for a country. These locations are then computed for getting common day locations	Kulgam
cdl_state	String	State of location where highest count of mobile signals distribu- tion are recorded for the work part of the day. We observed all locations for a device & filter location which have been seen for "N" days or more for a country. These locations are then computed for getting common day locations	JК
cel_geohash	String	Geohash for the latitude-longi- tude where home hours of the day have maximum signals. We observed all locations for a device & filter location which have been seen for "N" days or more for a country. These locations are then computed for getting common evening locations	ttvprjtg
cel_city	String	City of location where highest count of mobile signals distribu- tion are recorded for the home part of the day. We observed all locations for a device & filter location which have been seen for "N" days or more for a country. These locations are then computed for getting common evening locations.	Kulgam
cel_state	String	State of location where highest count of mobile signals distribu- tion are recorded for the home part of the day. We observed all locations for a device & filter location which have been seen for "N" days or more for a country. These locations are then computed for getting common evening locations.	JК



Field Name	Data type	Feature Definition	Example
cel_zipcode	Array	Zip code of location where highest count of mobile signals distribution are recorded for the home part of the day. We observed all locations for a device & filter location which have been seen for "N" days or more for a country. These locations are then computed for getting common evening locations	[180001]
cdl_zipcode	Array	Zip code of location where highest count of mobile signals distribution are recorded for the work part of the day. We observed all locations for a device & filter location which have been seen for "N" days or more for a country. These locations are then computed for getting common day locations	[180001]
most_seen_location	String	Geohash for the latitude-longi- tude having maximum signals. We observed all locations for a device & filter location which have been seen for "N" days or more for a country. These locations are then computed for getting most viewed location	ttvprjtg
most_seen_city	String	City for location where highest count of mobile signals distribu- tion is recorded for the day. We observed all locations for a device & filter location which have been seen for "N" days or more for a country. These locations are then computed for getting most viewed locations	Kulgam
most_seen_state	String	State for location where highest count of mobile signals distribu- tion is recorded for the day. We observed all locations for a device & filter location which have been seen for "N" days or more for a country. These locations are then computed for getting most viewed locations	JK
cl_most_seen_zipcode	Array	Zip code for location where highest count of mobile signals distribution are recorded for the day. We observed all locations for a device & filter location which have been seen for "N" days or more for a country. These locations are then computed for getting most viewed locations	[180001]



Field Name	Data type	Feature Definition	Example
norm_travelled_coun- tries	Double	Normalized travel countries for devices in the country	0
norm_travelled_days	Double	Normalized travel days for devices in the country	0
norm_premi- um_poi_seg_engagement	Double	Normalized visit score to the premium POIs for devices in the country. High Value POIs are locations for luxury engagement like premium brand stores, golf clubs, 5-star hotels etc. This is a curated list of all luxury engage- ment POIs for the country.	0.024390244

3 Carrier Features

Field Name	Data type	Feature Definition	Example
wifi_carrier	Array	Most seen wifi carrier (whole day)	[Airtel Broadband]
wifi_engagement	Double	Mobile signals distribution by connection type where connection type is equal towifi	0.0298
cellular_engagement	Double	Mobile signals distribution by connection type where connection type is equal to cellular	0.9702
cellular_carrier	String	Last seen cellular carrier	Airtel
sec_cellular_carrier	String	Second current cellular carrier name for the device [in dual sim devices]	Jio
previous_cellular_carrier	String	Immediate Previous cellular carrier name for the device before current carriers	Vodafone
business_engagement	Double	Business services POIs engagement score for the device	0
all_previous_cellular_car- rier	Array	Array of All Previous Cellular Carrier Names	[Vodafone, Idea]



Field Name	Data type	Feature Definition	Example
norm_cellular_carrier tenure	Double	Country normalized tenure of the current cellular carrier of the device	0.3333
norm_sec_cellular_carri- er_tenure	Double	Country normalized tenure of the second current cellular carrier of the device [in dual sim devices]	0.1098191
norm_all_cellular_carri- ers_median_tenure	Double	Country normalized median Tenure of the device across all cellular carriers	0.04878
norm_previous_cellular carrier_tenure	Double	Country normalized tenure of the Immediate previous cellular carrier of the device	0.23365
norm_total_cellular_carri- ers_seen	Double	Country normalized count for all carriers seen for the device	0.14954

4 Demographics Features

Field Name	Data type	Feature Definition	Example
gender_prediction	Struct	Gender of the Device. Possi- ble values are "1" meaning male and "0" meaning female.	{gender_predic- tion=0, gender_confi- dence_bucket=hc, gender_confidence score=0.7991588088 80887}meaning female.
gender_confidence_buck- et	String	Confidence bucket for gender [Self-declared(SD), High(HC), Medium(MC) & Low (LC)]	hc
age_bucket_prediction	Struct	Age Bucket of the Devices. Value is received as Bucket- Code where each bucket code means the respective (Bucket Label) [1(18-24); 2(25-34); 3(35-44); 4(45-54); 5(55+)]	{age_bucket_predic- tion=3, age_bucket_ confidence_bucket=hc, age_bucket_ confidence_score=0.8}
age_bucket_confidence bucket	String	Confidence bucket for age group [Self-declared(SD), High[(HC) , Medium(MC) & Low (LC)]	sd



5 Device Engagement Features

Field Name	Data type	Feature Definition	Example
norm_signals	Double	Normalized value of signal count for all devices.	0.032572232
active_day_percentage	Double	Active days % in the observa- tion period/device period	0.35294
commute_hour_engage- ment	Array	Mobile signals distribution for the commuting hours of the day [7AM-10AM AND 5PM-8PM]	0.15625
early_hour_engagement	Array	Mobile signals distribution for the early hours of the day[4AM-7AM]	0.01563
night_hour_engagement	Array	Mobile signals distribution for the late hours of the day. [11PM-4AM]	0
other_hour_engagement	Double	Mobile signals distribution for day besides commuter hours, night-rider hours, early-riser hours or work time hours	0.1875
work_hour_engagement	Double	Mobile signals distribution for the work hours of the day [10AM-5PM]	0.64063
norm_daily_engage- ment_score	Double	Normalized average mobile signals count per active day for the country	0.098194985
norm_device_period	Double	Normalized device observation period of the devices in the country	0.054158607
norm_inactive_period	Double	Normalized inactive period of the devices in the country	0.280898876
norm_num_days	Double	Normalized active engage- ment period of the devices in the country	0.019554956
weekday_engagement	Double	Mobile signals distribution for the weekdays [Monday-Thurs- day]	0.61538
weekend_engagement	Double	Mobile signals distribution for the weekends [Friday-Sun- day]	0.38462



6 Device Mobility Features

Field Name	Data type	Feature Definition	Example
norm_rog	Double	Normalized radius of gyration for the country	0.116353685
norm_average_daily_dis- tance_m	Double	Normalized average daily distance for devices in the country	0.18878704
norm_daily_commute	Double	Normalized daily commute for devices in the country	0
norm_average_daily_ac- tive_hr	Double	Normalized daily average active hours for devices in the country	0.02173913
norm_median_daily_ac- tive_hr	Double	Normalized daily median active hours for devices in the country	0



Field Name	Data type	Feature Definition	Example
platform	String	Platform used on device as {IOS, ANDROID or NIL}	ANDROID
year_released	String	Year of Release for the model	2017
device_category	String	Category of device as {SMART- PHONE, TABLET or NIL}	SMART PHONE
device_manufacturer	String	Device Manufacturer / Phone brand / Vendor	Орро
device_model	String	Model number for the device.	CPH1609
norm_phone_model_age	Double	Normalized device model age of the devices in the country	0.2
norm_phone_price	Double	Normalized price of the devices in the country	0.036971709

Household Features 8

Field Name	Data type	Feature Definition	Example
household_size	Long	Number of IFAs seen during household engagement hours. Household engagement is defined as mobile signals received during daily time between 8PM - 6AM on week- days, and all hours on week- ends for devices seen together at the same location, having same IP address or within 50m radius of devices having the same [7-digit precision] geohash accounting for the frequency of "seen together" in the device household scores.	2
household_phone_price	Double	Weighted average of prices for all the devices seen in house- hold size	14566.75
household_phone_age	Double	Weighted average of all phone ages seen for all devices in household size	1.5



9 Relationship Network Features

Field Name	Data type	Feature Definition	Example
social_size	Long	Number of IFAs seen during social engagement hours. Social engagement hours are defined as mobile signals received daily between 6PM-10AM on weekdays, and all hour's weekend for the devices seen together at the same time also accounting for the frequency of seen together in the device social scores.	5
social_phone_price	Float	Weighted average of prices for all the devices seen in social size.	19508.154
social_phone_age	Float	Weighted average of all phone ages seen for all devices in social size.	2.0682657
work_size	Long	Number of IFAs seen with that particular IFA during work engage- ment hours. Work engagement is defined as mobile signals received during daily time between 8AM–6PM on weekdays for devices seen together at the same time; also accounting for the frequency of seen together in the device work scores.	3



Field Name	Data type	Feature Definition	Example
work_phone_price	Float	Weighted average of prices for all the devices seen in work size	26622.924
work_phone_age	Float	Weighted average of all phone ages seen for all devices in work size	2.6643028

About Mobilewalla

Mobilewalla is a global leader in consumer intelligence solutions, leveraging the industry's most robust data set and deep artificial intelligence expertise to better understand, model and predict customer behavior. With rich insights into consumer behavior, our proprietary data and AI solutions help organizations make more informed business decisions and effectively acquire, understand, and retain their most valuable customers. To learn more please visit www.mobilewalla.com.