Habidatum Data Dictionary

The below descriptions outline the parameters of the key metrics offered by Habidatum in the package of Location Risk Score and underlying components. All metrics are attached to space and time and reflect socio-economic trends of areas locating real estate properties of various types.

Travel time data

A set of metrics providing information on the size and availability of critical services within a set time interval by car. Travel time data is based on the modeling of the fastest travel paths between any two points on a road network, taking into account the road type, maximum speed and traffic distribution throughout the day. Updated quarterly.

Data Field	**Data Type**	**Comments**
Accessible area size	Integer	The size of the accessible area (in square meters) for a set time interval (5-60 minutes) by driving. *Assumption: the smaller the accessible area, and the less the amount of critical services is available within the isochrones — the higher the risk of the locations may be (the thresholds are based on each metric's frequency distribution).*
Number of critical utility providers in the accessible area	Integer	Number of professional damage response providers (fire stations, police stations, repair services, etc) available within a set time interval by car. *Example: fire stations coverage risk — can the building be accessed by the fire truck within X minutes?*

Mobility data

Shows buildings' or grid cells/clusters' visitation patterns occupancy/vacancy and foot and car traffic patterns around them. On building level, the metrics allows defining types of buildings, like buildings with regular visits, recently vacant buildings, buildings, vacant for some time. Updated monthly (can be more frequent).

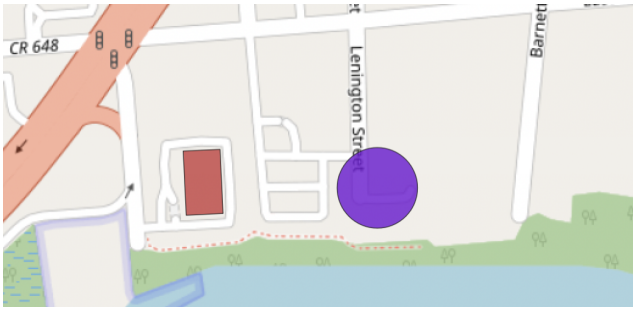
For mobility metrics, we suggest looking at both extremes of the metric values distribution: relatively high values signaling higher exposure of population within and around the properties to possible risks; relatively low values might indicate lower attendance with corresponding risks. Additionally, low recovery rates might indicate the out-of-normal operation patterns. Numerical thresholds are based on each metric's frequency distribution.

Source of data: GPS based location data from national mobile application data providers. Data comes from a variety of applications, such as navigation apps, radios, dating applications, weather, and many more. When a user installs an app, they are often asked to share their location data with the company which provides the app and other companies who are partners with the app publisher. Users can opt in to location sharing (or choose not to opt in). When they opt in, then their phone collects data and shares it with the publisher companies. Mobility metrics are the result of aggregation of this data.

Location:

There are two different types of location units used across mobility datasets:

1. **Continuous grid**: a set of square polygons, covering the entire globe (side size is 50 meters or larger). Data can be provided within two systems of encoding:
    - Geohash[1] (level 7, 153m);
    - Quad tiles[2] (zoom level 19, up to 50 meters).
2. **Property-centered geofences:** a set of polygons of custom shape, corresponding to particular addresses.
    - **Building footprint:** Building polygons come from Microsoft Bing building footprints dataset. This is a national dataset, providing the polygons for buildings across the United States based on satellite imagery segmentation using neural networks processing. We extend it with OpenStreetMap building footprints dataset, to provide more precise geofencing in dense urban environments, where a block of buildings can be presented as a single building in Microsoft Bing building footprints dataset.
    - **Property buffer:** a circular buffer of an industry-specific size is used as an alternative to building footprint. A circular buffer provides comparability of measurements across different locations, representing not only the mobility within the property, but also surrounding location.



    - **Land parcel:** an area, listed as the land of the lot where the property stands. The data covers the vast majority of the US properties.
    - **Building trade area:** (area bounded by an isochrone: points accessible within equal time (5-60 minutes) from the building using one of 3 modes of travel — driving, walking, public transit).

There are two base measurements in the mobility metrics:

- Number of *visitors* of location per period;
- Number of *visits* of location per period.

[1]Geohash http://geohash.org/site/tips.html#format
[2]Quad tiles https://medium.com/habidatum/infrastructure-behind-the-location-risk-score-4e09ec157dcd

A single visitor is the presence of a unique device identifier that made a visit to the property geofence at any day or number of days within a month. A unit of location for mobility metrics stands at the top priority, as it geofences the properties and gets the corresponding mobility signals. Locations can be uniform across all properties or tailored for each individual property. Here, we use both building footprints.
The difference between Visits and Visitors lies in how multiple occurrences of the same device within the month is accounted for:

- In *Visits* metrics, a device is counted as many times, as many different days they have been at the property. (For example, a person that have been at the property for 5 days in a month is counted 5 times);
- In *Visitors* metrics, a device is counted once per period (a month). (For example, a person that has been at the property for 5 days in a month is counted once).

Number of visits is always equal or greater than the number of visitors, as each visit is made by a certain visitor at some day in that month.

We provide both measured and approximated measurements — see methodology behind approximation below:
Mobility data covers a fraction of the country's population, 7% on average. This number varies from region to region (New York City vs Dallas) and from period to period (January 2019 vs September 2019). In order to compensate for these differences and provide comparable mobility estimates across the national portfolio of properties over a long period of time, the approximation mechanisms are taken into account using the following formula:

$$Visits_{estimated,\,region,\,period} = Visits_{observed,\,region,\,period} * PopRatio_{region,\,period}, \text{ where}$$

$$PopRatio_{region,\,period} = \frac{PopulationACS_{region,\,period}}{PopulationMobilityData_{region,\,period}}$$

For each period (a month) and for each region (regions defined by the Bureau of Economic Analysis. Each region consists of a single or multiple MSAs) the approximation ratio is estimated as the ratio of the region's population, reported by American Community Survey (ACS-5 year estimates) and the total population of devices with home locations reported within the region.

Segments. All mobility metrics can be divided into multiple segments:

1. Dwell time spent by users in a location, segmented into 10 minute time intervals.
2. Location visitation status:
    a. Crosser — any user that was seen at a location;
    b. Visitor — user spending at least 2 minutes at a location;
    c. Resident — user spending most of their time during night hours at a location. Residency location is updated weekly, based on the last few weeks of location information;
    d. Worker — user for whom the location is the most frequent during working hours (and is different from the residency location). Work location is updated weekly, based on the last few weeks of location information. Custom worker segments (shift workers) can be derived based on dwell time and additional filters.

3. Visitation origin — the origin location that the user left before visiting the location: state, county and census tract of the origin, as well as the distance from the home of the visitor to the location, split to tens of kilometers (within 10 km, 10-20 km, etc.). Can be used to determine out of state visitors[3].

Mobility data is updated monthly or quarterly and can be aggregated to the following temporal units, producing individual metrics like the ones described in the table below:

Data Field	**Data Type**	**Comments[4]**
Total number of visitors per month	Integer	Total number of unique visitors per month at a location. Provides information on the total volume of people seen at a particular month. The same device[5] seen at multiple days is only counted once per month.
Average number of visitors per day	Integer	Average number of visitors per day at a location: an average of the count of unique users per each day in a month. Provides information on the average volume of people seen per day at a particular month.
Median number of visitors per day	Integer	Median number of visitors per day at a location: a median of the count of unique users per each day in a month. Provides information on the average volume of people seen per day at a particular month.
Per hour distribution of the number of users over [0, 23] hour interval	Integer	The number of visitors at a given hour, for each hour in the day. Calculated as total number of visitors at a particular hour (from [0, 23] hour interval) across all days within a month.
Per hour distribution of total dwell time[6] of users over [0, 23] hour interval	Integer	Total dwell time is a sum of the dwell time of each visitor within an hour. Calculated as total dwell time at a given hour at a particular hour (from [0,23] hour interval) across all days within a month.

The following combinations and derivatives of mobility metrics can be computed and provided in addition to the above:

- Distribution and thresholds: histograms show amplitude of the metric values — whether there is a high variance across property locations or if properties share a similar range of

[3] Based on the origins on visits and census data, aggregated socio-demographic parameters of visitos may be added, like Household Income, Age, Gender, Education level, Marital status.
[4] As mentioned above, mobility metrics go through an approximation process. So, all the described metrics can be offered in measured (initial) and approximated values.
[5] In case visits are measured instead of visitors, as described above, the same device is counted as many times, as many different days they have been at the property during the month.
[6] As suggested in the section describing data segments, dwell times can also be used to segment visits and visitors. The examples of dwell time segments: long visits (3 hour +), out & about (1-3 hours), short visits (less than 1 hour) etc. (customizable).

values; potential threshold for the division of property locations (“good” and “bad”) according to the metric

- Relative and absolute deltas of the metric between two periods (for example, %YoY change of total number of users per month), and change of distribution of locations/buildings by metric value buckets in each time period (see Figure 2 below)
- Rankings of the metrics (for example, ranking of the total number of users per month compared to locations in a county/state/nation/locations with the same type of property located in it)
- Ratios of metrics in different location types (for example, the ratio of users per month at a building footprint to user per month in 5 minutes walking trade area)
- Ratios of the metrics (for example, Total number of users per month divided by Average number of users per day — a high number indicating a high turnover (e.g. retail location) with a low number indicating a lower turnover (e.g. office); or, relative “density” metric derivative with visits or visitors divided by the area of the property geofence (building footprint or circular buffer) — per acre of the property geofence)
- Temporal patterns identification as separate features (for example, presence of a morning and evening peak at a location, or a uniform hourly distribution of visitors throughout the day)

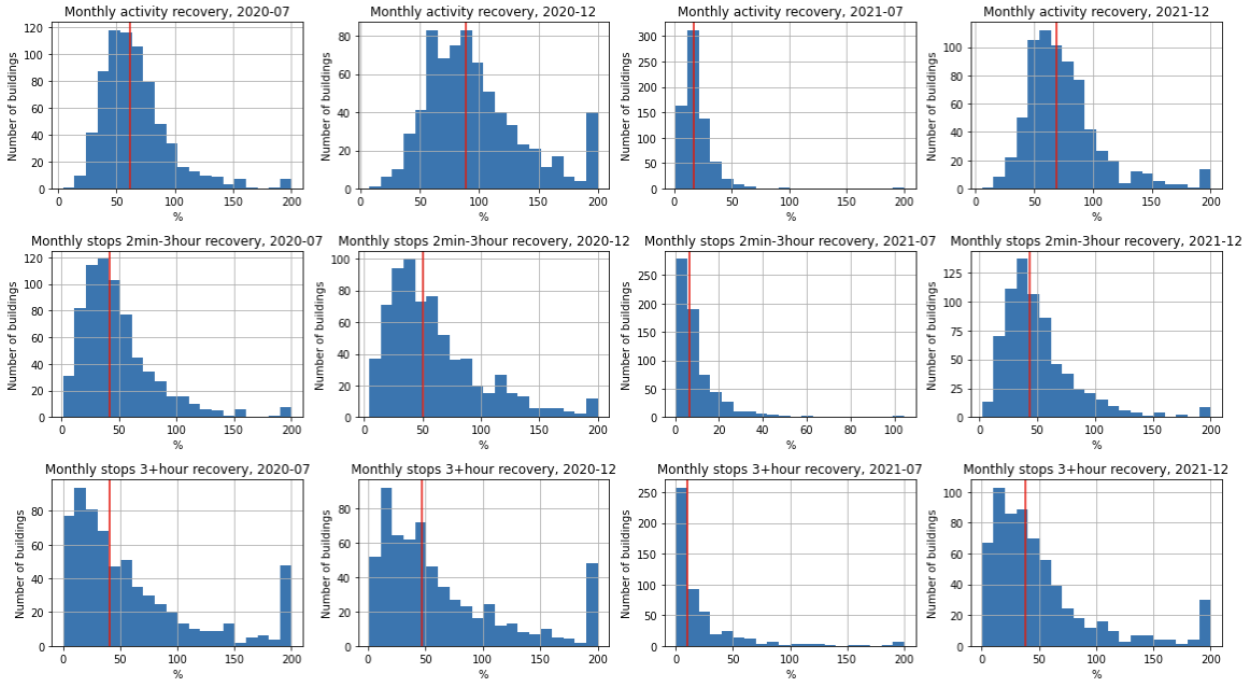


Figure. Histograms of mobility recovery for 4 periods: July 2020, December 2020, July 2021, December 2021. Separate row for each metric: Monthly activity recovery, Short stops (2min-3hour dwell time) recovery, Long stops (3+hour dwell time) recovery. Median value for a particular metric and month is shown with a vertical red line. Note that values are capped at 200%, producing a bump at the left of the charts.

Location Risk Score (Habidatum LRS)

Habidatum LRS is a metric showing the relative standing of locations based on the amount of daily human activity and businesses in and around them. The underlying terms and methodology are described in more detail below and in the annex. Habidatum LRS is a relative metric, ranging from 1 (highest score) to 10 (lowest score) with two benchmarks: Regional (MSA) and National (across all MSAs in the USA).

Location: A cluster of neighboring grid cells (shape).

Habidatum LRS is updated quarterly and produces the following individual metrics (described in more detail in the below table):

- Sub-scores: 10, including Local Score of the Target Cluster and 9 Catchment Areas Scores for 3 Travel Modes and 3 Travel Times; x2, measured for two types of Reference Areas (Regional we suggest to focus on, and National);
- Mixes: 27, combinations of sub-scores and the resulting integral score they provide ; x2, measured for two types of Reference Areas (Regional we suggest to focus on, and National);
    - Each component of the dataset can be a criterion for classification and a signal by itself. For example, the signaling metrics may be the presence of strong Walking or Public Transport Catchment Area Sub-Scores.
- The amount of Mixes resulting in the best possible integral score for the Target Location (the more best Mixes, the more is the commercial resiliency of the property) and the difference between the best and the worst scores for the Mixes.

Data Field	**Data Type**	**Comments**
Location Risk Score	Float	The combined Location Score of areas in close proximity to the building (trade area) is scored relative to other trade areas in the selected geography. Configurable by travel mode (walking, public transit, driving) and by travel times (10, 20, 30 minutes), with suggested configuration for retail, office, industrial, warehouse property types.
Local score	Integer	The rank of a location's commercial potential relative to other locations in selected geography considering its centrality (visitor count, business diversity). The ranking goes from 1 to 10, 1 being the highest, 10 being the lowest score.
Catchment attraction score: driving 10 min	Integer	Combination of active traffic (count of people), business diversity and infrastructure in given transport accessibilities. Configurable by travel mode (walking, public transit, driving) and by travel time (10, 20, 30
Catchment attraction score: driving 20 min		

Data Field	Data Type	Comments
Catchment attraction score: driving 30 min		minutes, etc). The indicator is a sum of scores for all locations (centrality clusters) in a given isochrone. The ranking goes from 1 to 10, 1 being the highest, 10 being the lowest score.
Catchment attraction score: walking 10 min		
Catchment attraction score: walking 20 min		
Catchment attraction score: walking 30 min		
Catchment attraction score: public transit 10 min		
Catchment attraction score: public transit 20 min		
Catchment attraction score: public transit 30 min		
Number of best scoring Habidatum LRS mixes, number_of_best_mixes	Float	Number of Habidatum LRS mixes that produce the same best score (from 1 to 27).
Delta between the best scoring and worst scoring Habidatum LRS mix for the location, best_worst_mix_delta	Float	The absolute value of the difference between the best score and the worst score (out of all 27 possible values for all Habidatum LRS mixes). The larger the delta the bigger the difference.

The following combinations and derivatives of Location Risk Score metrics can be computed and provided:

- Relative and absolute deltas of any metric between two periods (for example, %YoY change of Location Risk Score);
- Variance of the Location Risk Score subscores (all 1s vs a mix of 1s and 5s);
- Difference between the Location Risk Score and the Location Risk Score of the neighboring locations;
- Interaction between the Location Risk Score and rent price: boolean flag for the rent being within the expected rent levels for a given Location Risk score;
- Difference between the Location Risk Score for a particular asset type vs the average of all Location Risk Score components.

POI (Points of Interest)

A set of metrics providing information about the quantity and quality of business and administrative locations (bars, beauty shops, courts, schools, fire stations, etc). Metrics can be provided for a particular address as a parcel, a grid cell (side size is 50 meters or larger), a building footprint or the building trade area. Updated quarterly.

Data Field	Data Type	Comments
Number of POI	Integer	Total number of POI, as well as the number of each category of POI (e.g. bars, beauty shops, courts, schools, fire stations, etc).
Diversity of POI	Integer	Number of different categories of POI at a location.
Combinations of POI	Boolean	Presence of a particular combination of frequent POI (for example, coffee shop + bookstore). Based on the frequent pattern mining of POI pairs in the State of the property.

Inclusive Growth Score™

Inclusive Growth Score is a public-access, web-based service that enables users to learn about measures of inclusion and growth within census tracts across the United States, developed by Mastercard, Habidatum and data partners. It is formed by a set of metrics measuring Growth and Inclusion of each census tract in the US. 18 metrics across 3 pillars — Economy, Place and Community. Metrics are provided as percentile rankings — from 1 to 100, with 3 benchmarks: National, State, and UrbanCode (benchmarks against the census tracts with the same level of urbanization). Underlying metrics values are also provided (% growth YoY for Growth metrics, % share for Inclusion metrics. The only exceptions are three metrics based on Mastercard data — only ranks are provided for them). Data covers 2017-2021 years with some exceptions. Detailed methodology provided here. Updated annually, except 5 metrics (marked in the table).

Data Field	Comments
Acres of park land	Percentage of designated tract land area that is park land Source: Trust for Public Land, PAD-US
Affordable housing	Percentage of renter and owner-occupied housing units where monthly costs are lower than 30 percent of income Source: ACS 5-Year
Commercial diversity	Updated quarterly. Percentage of industries represented Source: POI Provider
Early education enrollment	Percentage of population under the age of five enrolled in early education programs Source: ACS 5-Year
Female above poverty	Percentage of females living above the poverty Source: ACS 5-Year
Gini coefficient	Gini coefficient of income inequality (lower coefficient denotes lower inequality) represented through a percentage

	Source: ACS 5-Year
Growth in spending per capita	Updated quarterly. Rank value only. Percentage growth of average spend per account based on anonymized and aggregated indexed transaction data Source: Mastercard GeoInsights
Health insurance coverage	Percentage of the eligible population with health insurance coverage Source: ACS 5-Year
Internet access	Percentage of households with an internet subscription Source: ACS 5-Year
Labor market engagement index	Index representing the combined employment, labor force participation, and percentage with bachelor's degree Source: ACS 5-Year
Minority/women owned businesses	Updated quarterly. Percentage of minority or women-owned businesses out of all businesses Source: Commercial Data Provider
Net occupancy	Percentage growth in population of renter and owner-occupied housing units Source: ACS 5-Year
New businesses	Updated quarterly. Rank value only. Percentage growth of net new businesses based on anonymized and aggregated location data Source: Mastercard Places
Personal income	Percentage growth of per capita income Source: ACS 5-Year
Residential real estate value	Percentage growth of value of residential real estate Source: ACS 5-Year
Small business loans	Percentage growth of the number of small business loans Source: FFIEC
Spend growth	Updated quarterly. Rank value only. Percentage growth of spending based on anonymized and aggregated indexed transaction data Source: Mastercard GeoInsights
Travel time to work	Percentage of workers with travel time to work under 35 minutes Source: ACS 5-Year

Satellite-based / Natural environment metrics

A variety of metrics based on the satellite imagery data and public data sources.

Data Field	**Data Type**	**Comments**

Tree/shrub proximity from buildings	Float	Metric indicating how close to a property there are trees and how many of them. Can be an indicator of possibility for trees falling on the property during storms and/or fire risk during droughts in hotter climates.
Water bodies proximity	Float	Metric indicating the proximity to the closest body of water, segmented by type (river, lake, sea).
Green accessibility index	Float	A metric based on proximity plus volume of trees in that proximity. Indicates how easy it is to access nature from the property (parks with trees of various sizes; forests). Opposed to calculating proximity of parks / green spaces, this index shows true green spaces and not «grass-only» parks.
Ground elevation	Integer	Ground elevation (on Earth's surface) in meters. The elevation dataset is limited to latitudes from 56ºS to 60ºN.
Green view index	Float	Percent of tree canopy on the street next to the property. Indicates well-being of the neighborhood and is health-related for residential properties.

Property level data

A set of parameters describing a particular property/properties located in the area.

Data Field	**Data Type**	**Comments**
Average rent per square ft per year	Float	Metric, indicating average rent per square foot for on-the-market properties at the location. Applicable for trade area or grid cell location types. No historical data, only available for the date of the inquiry.
Building footprint area	Integer	Area of the building, based on the footprint of the building.
Year built	Integer	The year the property was built.
Building height, floors	Integer	Number of floors in the building.

Location and date properties

Attributes, providing the identification of the property and its location.

Data Field	**Data Type**	**Comments**

LocationID	Integer	The unique identifier of the location.
Date	ISO-8601 formatted date string	The date when the measurement of the score was done.
Location address	String	Optional, the address of the location.
LocationType	String	Parcel/Footprint/Grid cell/Trade area.
Location Centroid Latitude	Float	The latitude of the location center.
Location Centroid Longitude	Float	The longitude of the location center.
Geometry	Polygon	Only in the GeoFile. This field is missing in the tabular data or the API.

Annex. Habidatum LRS methodology

Geographic units or types of areas

Habidatum LRS is built for addresses, or locations targeted for scoring (called **Target Locations**). It depends on three types of areas surrounding these Target Locations:

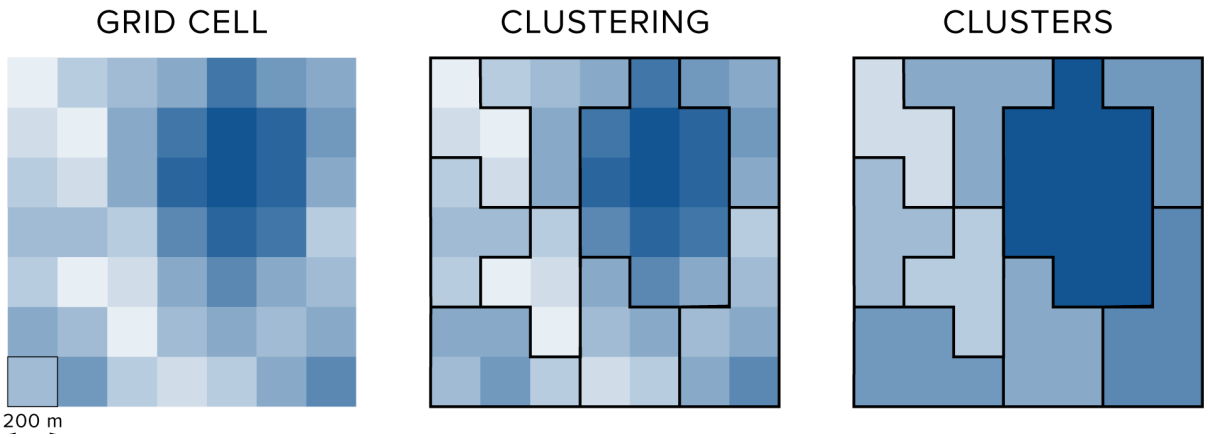
- **Local Cluster**
    - can be either **Target Cluster** — the Local Cluster, corresponding to the Target Location;
    - or a **Reference Cluster(s)** — all other local clusters in the **Catchment** or **Reference Areas**,
- **Catchment Area(s),**
- **Reference Area**.

Local Cluster — immediate vicinity of Target Location spots typically linked to single address; has clear boundaries formed by Clustering algorithm that combines adjacent Grid Cells into single Local Cluster, where Grid Cells and Clustering algorithm are defined as:

- **Grid Cells** — square parcels of land *(size depends on the type of location, based on the population density, more below)*, serving as building blocks of Local Clusters corresponding to benchmark risk metrics; form a uniform grid that covers all geographical areas without gaps.
- **Clustering** — algorithm that combines multiple neighboring Grid Cells to form Local Clusters; determines local centers among Grid Cells and attaches neighboring Grid Cells to them.

Local Clusters may have quite different sizes ranging from single Grid Cell to couple dozen Grid Cells. Thriving downtown areas usually consist of larger Local Clusters, while distant peripheral Locations are represented by smaller Local Clusters.

Grid Cells size. As Habidatum LRS was set to be computed all around the United States, we've proposed using different zoom-level clusters for territories with completely different features: urban and rural. We made large Local Clusters in the low-density areas so that they can cover the whole buildings (underlying grid cell side is 800 meters) and smaller clusters in high-density urban areas (underlying grid cell side is 200 meters). The urban/rural division is based on the population density of the corresponding census tract — it was empirically defined that the population density threshold for urban census tracts stands at approximately 100 people per square kilometer.



Catchment Area — broader area surrounding Local Cluster; represents a set of all Grid Cells accessible by particular Travel Mode within particular Travel Time (from center of corresponding Cluster) defined as:

- **Travel Mode** — one of 3 ways person can reach locality
    - **Walking**;
    - **Public Transit** — any combination of available public transportation;
    - **Driving** — traveling by car (Travel Time estimation accounts for impact of traffic).
- **Travel Time** — longest time it can take to reach locality for it to be considered part of Catchment Area; measured in minutes, can range from 10 to 60 in 5 minute increments**.**
- **Isochrone** *("isos khronos" mean equal time in Greek)* — border line of Catchment Area; its configuration depends on particular combination of Travel Mode and Travel Time that characterize shortest trip from any point of Isochrone to center point of corresponding Cluster that can be reached with such combination. Typically, faster Travel Modes combined with higher Travel Time values produce Isochrones that are more distant from Cluster center point.

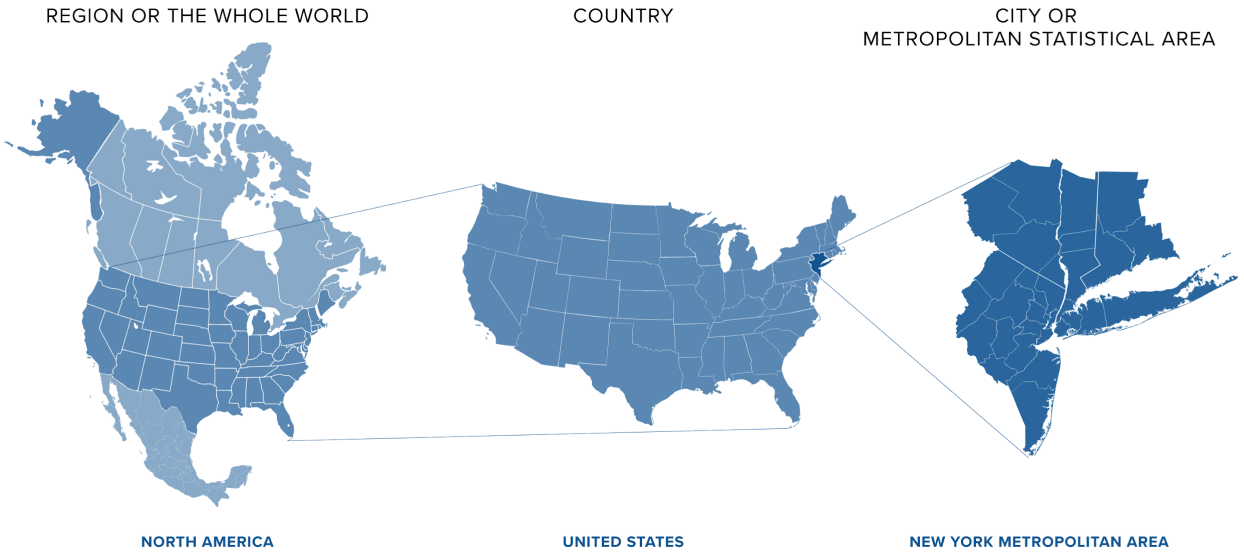
Each Local Cluster has multiple Catchment Areas around it produced by different combinations of Travel Modes and Travel Time values. Local Clusters with good transport accessibility have broader Catchment Areas (more distant from such Locations) than ones with poor transport accessibility. Size of Catchment Area does not solely represent its characteristic metric since its density characteristics are comparably important. More on this in the Underlying Metrics segment.

Reference Area — territory that contains Local Clusters with Habidatum LRS reflecting relative standing of such locations with respect to all underlying metrics that comprise Habidatum LRS-measure and define its ranking quality. Reference Areas fall into several categories according to implicit "hierarchy":

- **City** or **Metropolitan Statistical Area, Region** (e.g. New York City or New York metropolitan area) – higher level categories such as Region (e.g. State or Set of States) can be added as sub-country level;
- **Nation** (e.g United States) or set of countries (US and Canada or US and Western Europe);
- **Continent** (e.g North America), set of continents (North America and Western Europe), or **Global Domain**.

Two Reference Areas are used: ***National*** (all reference clusters in the United States) and **Regional** (regions defined by the Bureau of Economic Analysis. Each region consists of a single or multiple MSAs).

Local Clusters can have different Habidatum LRS in different Reference Areas. It is possible to have a strong score (typically, 2-3 range) when the corresponding Reference Area covers a small city. Yet at Country level the same Local Cluster may get a weaker score (e.g. 5-7 range). Typically, this score will get weaker in broader Reference Areas since underlying Local Cluster will be benchmarked against a greater number of Reference Locations with proportionally higher chance of occurrence for locations with better quality and, hence, better scores.



Underlying metrics

Habidatum LRS is based on ranking Local Clusters and Catchment Areas around them based on the following set of metrics.

- **Grid Cell Attraction** — uses aggregate monthly Grid Cell visitors count as indicator of Grid Cell related human activity (mobility) implied by daily attendance component of general visitors' mobility patterns and businesses in **Grid Cell**. The metric is an aggregate of 3 sub-metrics of the grid cell:
    1. **Monthly activity** — number of people registered in Grid Cell monthly, based on aggregated anonymous mobile phone activity data;

2. **Attractors Density** — number of Commercial and Non-commercial (Cultural & Recreational) **objects** (includes types of commercial real estate as well as goods, services, or jobs offered by its tenant(s), e.g. financial, medical, restaurants, etc.) located within Grid Cell;
3. **Attractors Diversity** — number of different Commercial and Non-commercial (Cultural & Recreational) **types** (includes types of commercial real estate as well as goods, services, or jobs offered by its tenant(s), e.g. financial, medical, restaurants, etc.) located within Grid Cell.

Cluster Attraction metric can vary significantly for different Grid Cell Local Clusters — in thriving central districts they may include multiple Grid Cells, each with dozens of commercial and noncommercial objects of multiple types and hundreds of visitors per day, whereas some peripheral Local Clusters may consist of only one Grid Cell with few visitors per day and single commercial object. Task of achieving uniform measurement efficiently mapped to the original non-uniform spectrum of Cluster Attraction] prompts its conversion into logarithmic scale. More on this in Description of Methodology.

- **Cluster Attraction** — core component metric underlying Habidatum LRS and its ranking methodology that is calculated separately for each Local Cluster. as sum of all corresponding underlying Grid Cell Attraction values for all Grid Cells which form a particular Local Cluster.

- **Catchment Attraction** — sum of all **Cluster Attraction** values of all Clusters within Catchment Area. The total number of Catchment Attraction values corresponding to a single Local Cluster equals the total number of Catchment Areas around that Local Cluster implied by combinations of Travel Modes and Travel Time values.

Larger Catchment Areas reflect better transport accessibility of Local Cluster and thus imply distance one can travel via selected Travel Mode within travel time) contain more Local Clusters. This does not automatically imply that the accessible centrality is the same for two localities with accessible areas of the similar size. In the case of a busy downtown area this local marker is filled with localities of high centrality and their aggregate is a large accessible centrality, while in the case of a distant business cluster the same number of localities in the local marker can have very low centrality values and result in a low accessible centrality metric for that locality

Ranking process

Each Local Cluster within the corresponding Reference Area has a set of metrics: single Cluster Attraction value and multiple Catchment Attraction values. Each metric individually gets ranked and thus converted into scores. Ultimately, Habidatum LRS is calculated as an aggregate of some underlying ranking scores.

Aggregation of multiple ranking scores into Habidatum LRS

Habidatum LRS scale corresponds to a continuous interval from 1 (best, lowest risk) to 10 (worst, highest risk). Habidatum LRS has decimal value format rounded to several digits in necessary intermediate calculations and to single digit after decimal point for its resulting value

(e.g. 1.1 or 2.7); resulting score for each Local Cluster is given by average aggregate of several contributing sub-scores (*more on averaging subscores in the Habidatum LRS mixes section)*; where each contributing sub-score is produced by benchmarking underlying sub-metric corresponding to that Local Cluster against all similar sub-metrics for all Local Clusters within chosen Reference Area (region or nation); each sub-score reflects particular level of risk associated with Local Clusters.

Contributing sub-scores and associated risk levels are following:

- **Local Score** — reflects relative standing and in this context relative risk (according to Habidatum LRS-scale) of a particular Local Cluster resulting from ranking Catchment Attraction values of all Local Clusters in the Reference Area. Indicates risk of Local Clusters without direct contribution of areas around it.
- **Catchment Attraction Scores** — one per Catchment Area (e.g. for Walking Mode within 20 minutes Travel Time) for all Catchment Areas corresponding to each particular Local Cluster; indicate risk levels associated with broader areas around corresponding Local Cluster.

We calculate each sub-score independently — each of them is scored in a similar manner — from 1 (best) to 10 (worst), 10 integer values, and then average the selected set of sub-score to get a final Habidatum LRS. The selection of appropriate accessible scores is called an Habidatum LRS mix (*more in the Habidatum LRS mixes section)* — more on the selection of the sub-scores later.

Sub-score ranking process

Let's review the ranking process of Local Score ranking (the same process is applied to each Catchment Area Score). The local cluster with highest **Cluster Attraction** value obviously gets the highest rank (1), while a set of local clusters with the lowest **Cluster Attraction** value get the lowest rank (10).

How do we assign the rank to all the continuum of localities between these two extremes?
There can be two approaches to this:

- to divide into ranks with *equal number of clusters (e.g 10 ranks with 100 clusters each)*
- **to divide into ranks with equal range of metric spread *(e.g 10 ranks with* Cluster Attraction *spread of 100 in them)***

For Habidatum LRS the latter approach is applied. The full range of cluster attraction values gets divided into 10 equal bins, and each local cluster is assigned the rank based on the bin where its cluster attraction value falls into.

Example: let's say that in a reference area of New York MSA, the lowest cluster attraction value is 1 and the cluster attraction value is 10001. The bin size in this example is (10001 - 1)/10 = 1000. Hence, any local cluster with a cluster attraction value from 1 to 1001 gets a rank 10, from 1001 to 2001 gets a score of 9, and any local cluster with a cluster attraction value between 9001 and 10001 gets a score of 1.

This ranking process results in a variable number of clusters with the same rank across different catchment areas, as the number of local clusters in each bin (rank) is not predefined.

Habidatum LRS mix

Habidatum LRS mix is the selection of a Travel Time for each Travel Mode, producing a set of Catchment Areas and corresponding Catchment Attraction Scores (3 scores, one per Travel Mode). These 3 Catchment Attraction Scores together with the Local Score produce a single Habidatum LRS for a particular Habidatum LRS mix.

Habidatum LRS mix is defined by 3 parameters:

- Travel time for driving Travel Mode: 10, 20 or 30 minutes;
- Travel time for walking Travel Mode: 10, 20 or 30 minutes;
- Travel time for public transit Travel Mode: 10, 20 or 30 minutes;

Overall, this produces 3 in the 3 power of combinations, 27 Habidatum LRS mixes in total. An example of a mix is (Walking 10, Driving 20, Public Transit 20). To calculate the corresponding Habidatum LRS, we take an average of the 4 scores. See example in Table 1 below.

Table 1: example of the calculation of the Location Rsk Score for a Habidatum LRS mix of (Walking: 10, Driving: 10, Public Transit: 20).

Habidatum LRS	**Local score**	**Catchment area score: Walking 10 min**	**Catchment area score: Driving 10 min**	**Catchment area score: Public transit 20 min**
2.25 = (2+3+2+2)/4	2	3	2	2
3.5 = (5+4+2+2)/4	5	4	2	3