

About data

The project utilizes several datasets provided in the VTB Data Fusion competition. All data was readily available and downloaded without issues:

1. transactions.parquet (971 MB)
2. target.parquet (870 KB)
3. hexses.target.lst (27 KB)
4. hexses.data.lst (130 KB)
5. moscow.parquet (193,9 MB)

Dataset	Size	Description	Key Columns
<code>transactions.parquet</code>	971 MB	Customer transaction data. Lat/Lon derived from H3 index (<code>h3_09</code>). <code>h3_09, customer_id, datetime_id, sum</code>	
<code>target.parquet</code>	870 KB	Target variable (ATM usage). Lat/Lon derived from <code>h3_09</code> .	<code>h3_09, customer_id</code>
<code>hexses_target.lst</code>	27 KB	ATM locations (H3 indexes). Lat/Lon derived.	<code>h3_09 → lat, lon</code>
<code>hexses_data.lst</code>	130 KB	Transaction locations (H3 indexes). Lat/Lon derived.	<code>h3_09 → lat, lon</code>
<code>moscow.parquet</code>	193,9 MB	OpenStreetMap data (includes raw Lat/Lon + POI tags).	<code>lat, lon, tag</code>

Transactions Cash withdrawals Locations Moscow

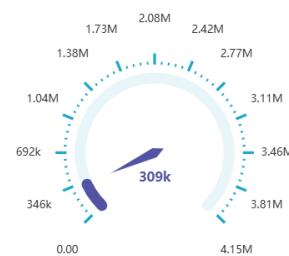
Cash withdrawals

Target variable data (ATM usage). Similar to transactions, this dataset initially only contains the `h3_09` column for location, with lat and lon derived subsequently.

Cash withdrawals dataset example

customer_id	h3_09
1	8911aa4cc2ffff
4	8911aa01d4ffff
4	8911aa4a565bffff
4	8911aa4a45853ffff
4	8911aa4e167ffff
4	8911aa7b453ffff
5	8911aa71b2bffff
9	8911aa615b7ffff
9	8911aa7abc7ffff
11	8911aa7a117ffff

Withdraw Rate



Transaction Locations

`hexses_data.lst` (130 KB)

Attribute	Description
Records	8,154 transaction locations
Original Data	H3 indexes (<code>h3_09</code>) only
Derived Columns	<code>lat, lon</code> calculated from H3
Key Insight	Customers' transaction hotspots (e.g., markets, residential areas).