

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
transactions.parquet	971 MB	Customer transaction data. Lat/Lon derived from H3 index (<code>h3_09</code>). <code>h3_09</code> , <code>customer_id</code> , <code>datetime_id</code> , <code>sum</code>	
target.parquet	870 KB	Target variable (ATM usage). Lat/Lon derived from <code>h3_09</code> .	<code>h3_09</code> , <code>customer_id</code>
hexses.target.lst	27 KB	ATM locations (H3 indexes). Lat/Lon derived.	<code>h3_09</code> → <code>lat</code> , <code>lon</code>
hexses.data.lst	130 KB	Transaction locations (H3 indexes). Lat/Lon derived.	<code>h3_09</code> → <code>lat</code> , <code>lon</code>
moscow.parquet	193.9 MB	OpenStreetMap data (includes raw Lat/Lon + POI tags).	<code>lat</code> , <code>lon</code> , <code>tag</code>

Transactions Cash withdrawals [Locations](#) Moscow

ATM Locations

[hexses.target.lst](#) (27 KB)

Attribute	Description
Records	1,657 ATM locations
Original Data	H3 indexes (<code>h3_09</code>) only
Derived Columns	<code>lat</code> (float64), <code>lon</code> (float64) via H3 library
Granularity	H3 resolution level 9 (1 km ² hexagons)
Use Case	Mapping ATM network coverage and cash demand zones.

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</code> , <code>lon</code> calculated from H3
Key Insight	Customers' transaction hotspots (e.g., markets, residential areas).