

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](#)

Moscow

OpenStreetMap data for a specific region of Moscow. This is the only dataset that originally includes lat and lon columns. It also contains a tag column with rich, nested information about points of interest.

Moscow Dataset

[moscow.parquet](#) (193.9 MB)

Feature Details

Unique Advantage Only dataset with raw `lat` / `lon` coordinates (not H3-derived).

Rich nested metadata in `tag` column:

- Examples: `shop=supermarket`, `amenity=school`, `tourism=museum`

Coverage

Specific Moscow region with 193.9MB of granular OpenStreetMap data. Used to enrich transaction/ATM data with:

Integration

- Demographic context
- Infrastructure proximity

Transaction Locations

[hexses_data.lst](#) (130 KB)

Attribute Description
Records 8,154 transaction locations
Original Data H3 indexes (`h3_09`) only

Derived Columns `lat`, `lon` calculated from H3

Key Insight Customers' transaction hotspots (e.g., markets, residential areas).