

## 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</code> , <code>customer_id</code> , <code>datetime_id</code> , <code>sum</code>	
<code>target.parquet</code>	870 KB	Target variable (ATM usage). Lat/Lon derived from <code>h3_09</code> .	<code>h3_09</code> , <code>customer_id</code>
<code>hexses.target.lst</code>	27 KB	ATM locations (H3 indexes). Lat/Lon derived.	<code>h3_09</code> → <code>lat</code> , <code>lon</code>
<code>hexses.data.lst</code>	130 KB	Transaction locations (H3 indexes). Lat/Lon derived.	<code>h3_09</code> → <code>lat</code> , <code>lon</code>
<code>moscow.parquet</code>	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

### Transactions

Customer transaction data. Crucially, this dataset does not initially contain lat and lon columns. These are derived later using the h3\_09 (H3 index) column and the h3 library's cell\_to\_latlong function.

Total number of transactions

4,150,912

Total number of transactions

### [Team3] Transaction dataset example

transaction_pk	customer_id	count	sum	avg	min	max	std	count_distinct	h3_09	datetime_id	mcc_code
0	1	1	3346.65	3346.65	3346.65	3346.65	N/A	1	8911aa4c62ffff	3	13
1	4	1	450	450	450	450	N/A	1	8911aa7b503fff	3	8
2	5	10	11035.69	1103.569	59	3620.18	1190.5303	6	8911aa63623ffff	3	13
3	9	2	628	314	295	333	26.870058	2	8911aa48577ffff	2	5
4	11	1	4155	4155	4155	4155	N/A	1	8911aa78297ffff	2	10
5	12	2	94	47	47	47	0	2	8911aa78dc7ffff	2	9
6	13	25	9089.44	363.5776	176.99	707.96	162.45757	2	8911aa4ec93ffff	3	0
7	15	10	1400	140	140	140	0	1	8911aa6a9a4c3ffff	1	8
8	16	1	10998	10998	10998	10998	N/A	1	891181b6507ffff	2	10
9	17	2	2760	1380	1380	1380	0	2	8911aa7ab0ffff	2	8

### 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).