**Lab report #8-9**

**Sadovskaya Veronika**

GitHub: <https://github.com/sdveronika/DataMola22>

**Task 1 (LW8) - Extraction Description**

Full Extraction:

The data is extracted completely from the source system. Because this extraction reflects all the data currently available on the source system, there's no need to keep track of changes to the data source since the last successful extraction. The source data will be provided as-is and no additional logical information (for example, timestamps) is necessary on the source site.

Offline Extraction:

The data is not extracted directly from the source system but is staged explicitly outside the original source system. The data already has an existing structure (for example, redo logs, archive logs or transportable tablespaces) or was created by an extraction routine.  ​

**Task 1 (LW9) - Transportation Description**

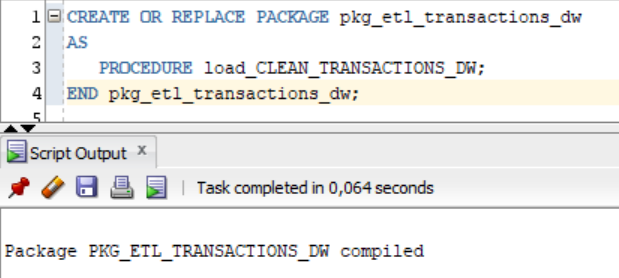
Transportation Using Flat Files​

The most common method for transporting data is by the transfer of flat files, using mechanisms such as FTP or other remote file system access protocols. Data is unloaded or exported from the source system into flat files using techniques discussed in "Extraction in Data Warehouses", and is then transported to the target platform using FTP or similar mechanisms.​

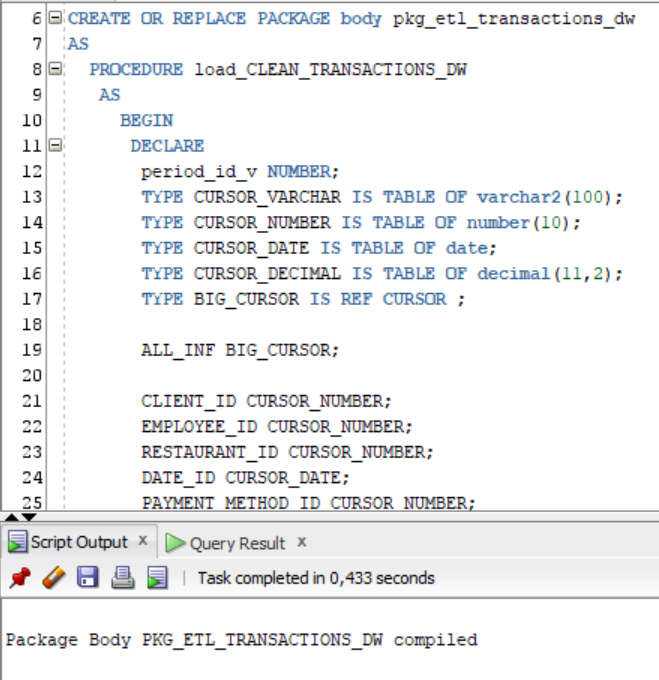
Because source systems and data warehouses often use different operating systems and database systems, using flat files is often the simplest way to exchange data between heterogeneous systems with minimal transformations. However, even when transporting data between homogeneous systems, flat files are often the most efficient and most easy-to-manage mechanism for data transfer.

**Task 2 - Prepare Table of Facts to DW Layer**

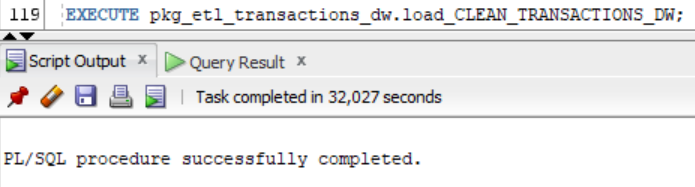
Create package pkg\_etl\_transactions\_dw:



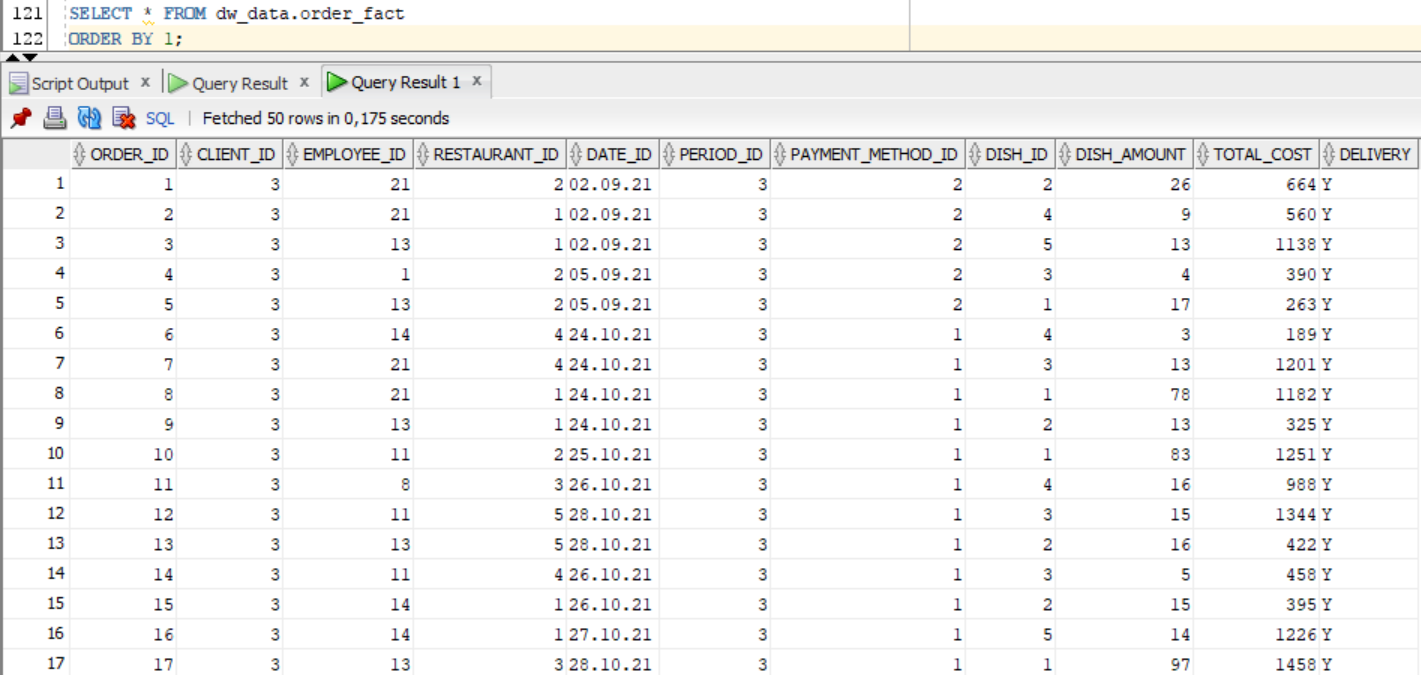
Create procedure load\_clean\_transactions\_dw in package pkg\_etl\_ transactions\_dw:



Execute procedure load\_clean\_transactions\_dw from package pkg\_etl\_ transactions\_dw:



Select from table dw\_data.order\_fact:

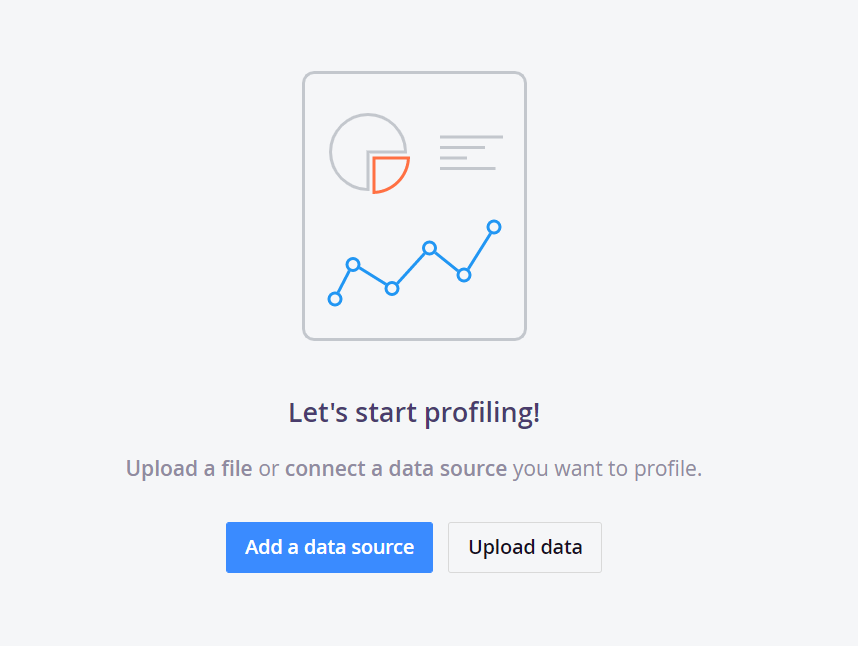


**Getting to know the site** [**https://app.ataccama.com/catalog**](https://app.ataccama.com/catalog)

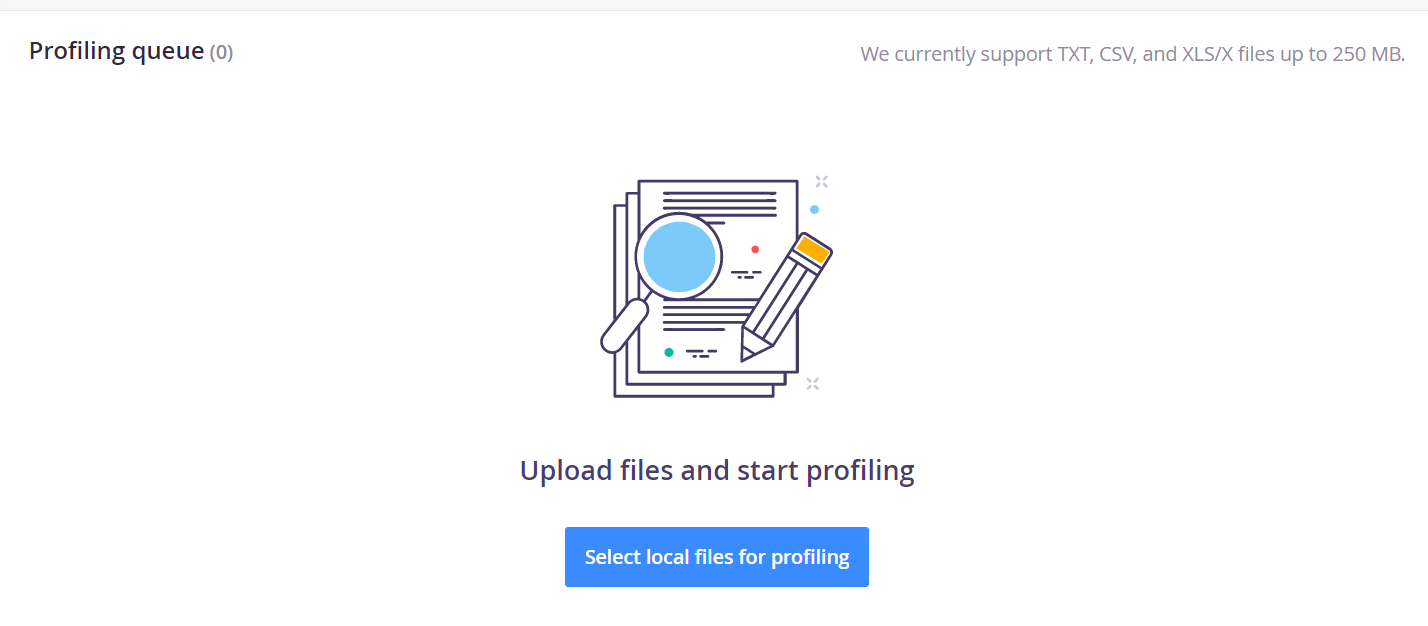
To begin with, we register on the site and choose to analyze our data.

For analysis, I chose my fact table. I copied 2000 rows from my spreadsheet and put them into an excel document.

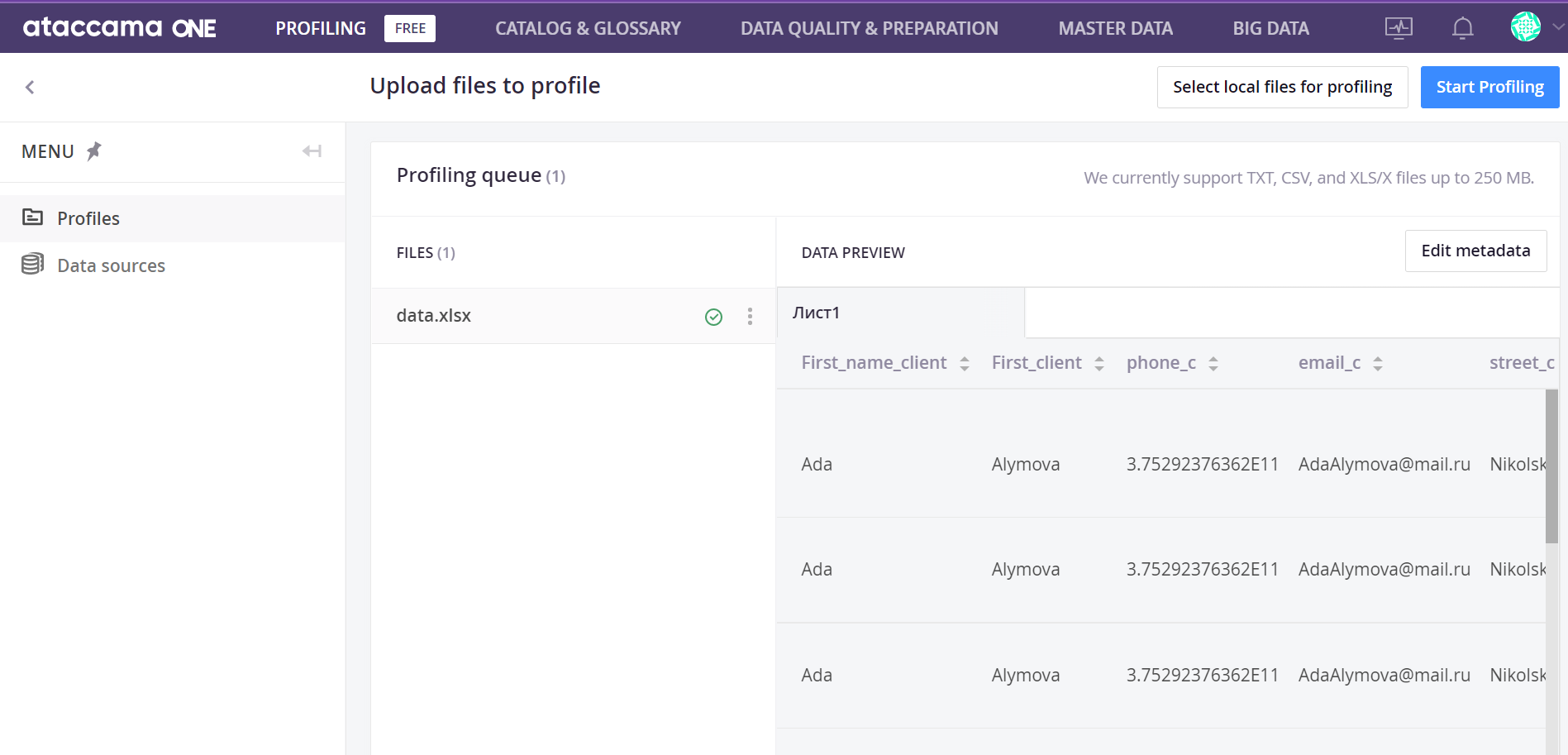
On the site, choose to add your data:

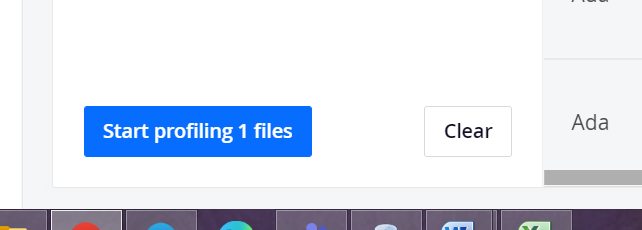


Adding an Excel document with our data:

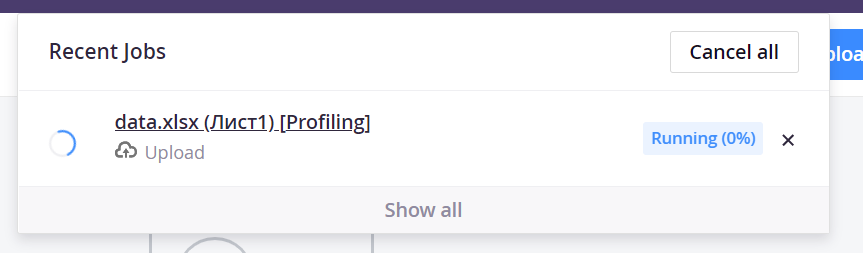


When our data is loaded, we select ‘start profiling 1 files’:

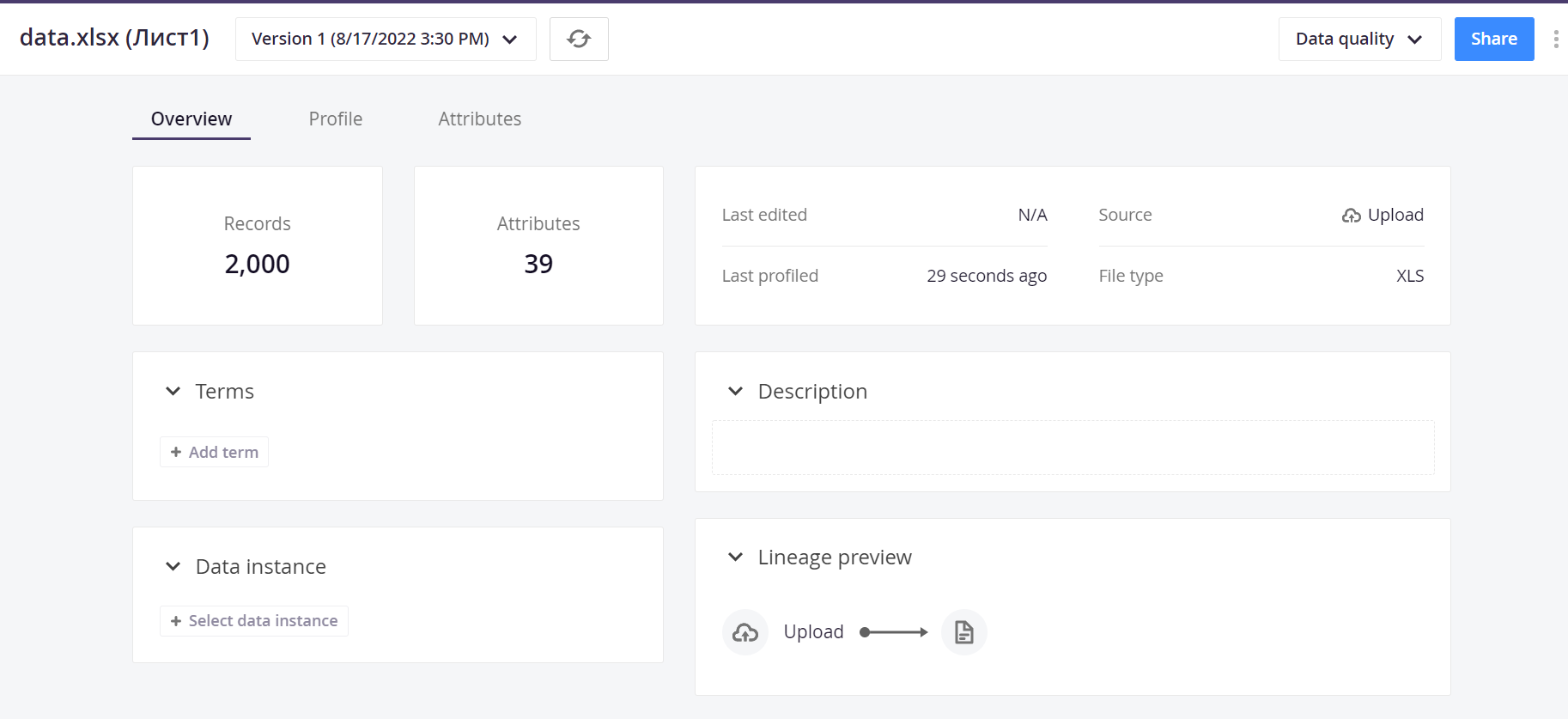




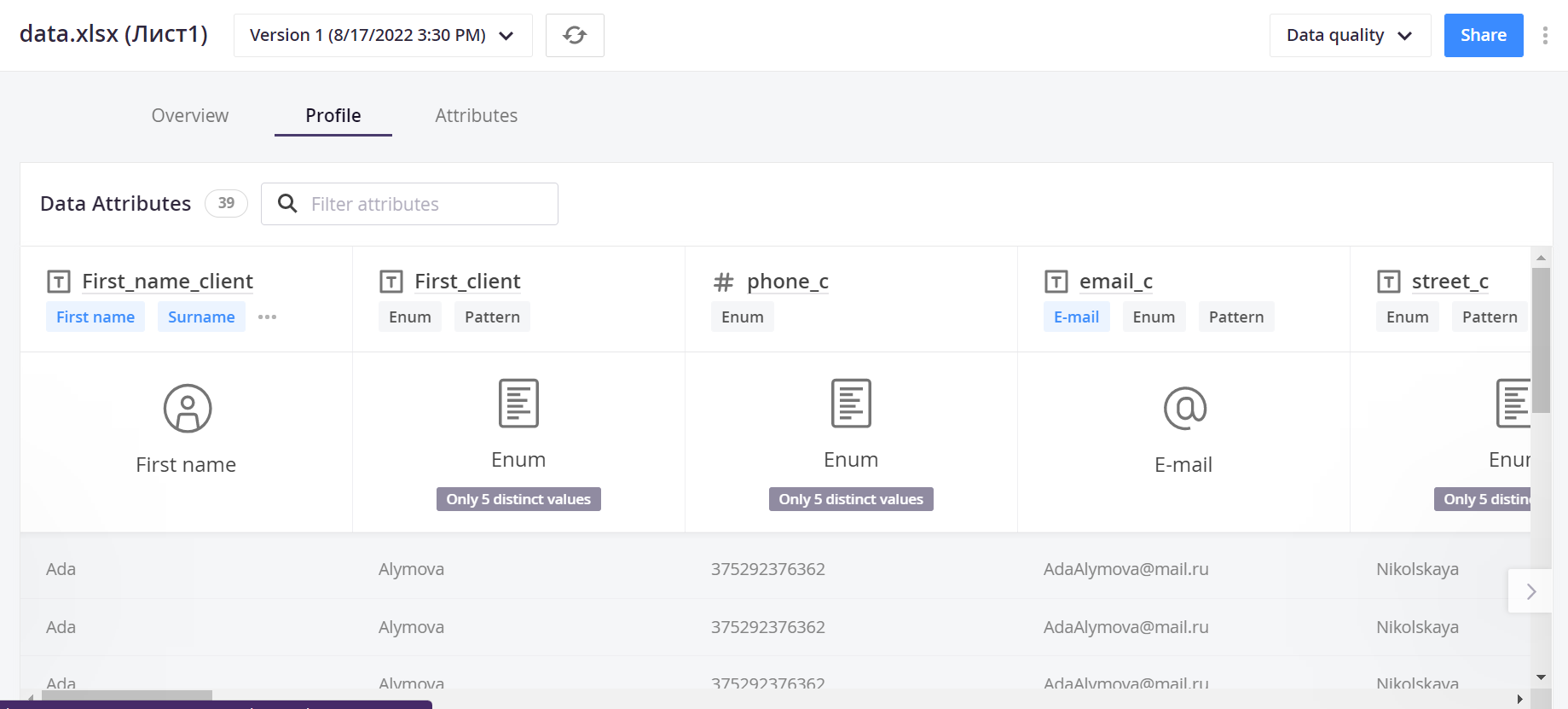
We are waiting for the processing of our data:



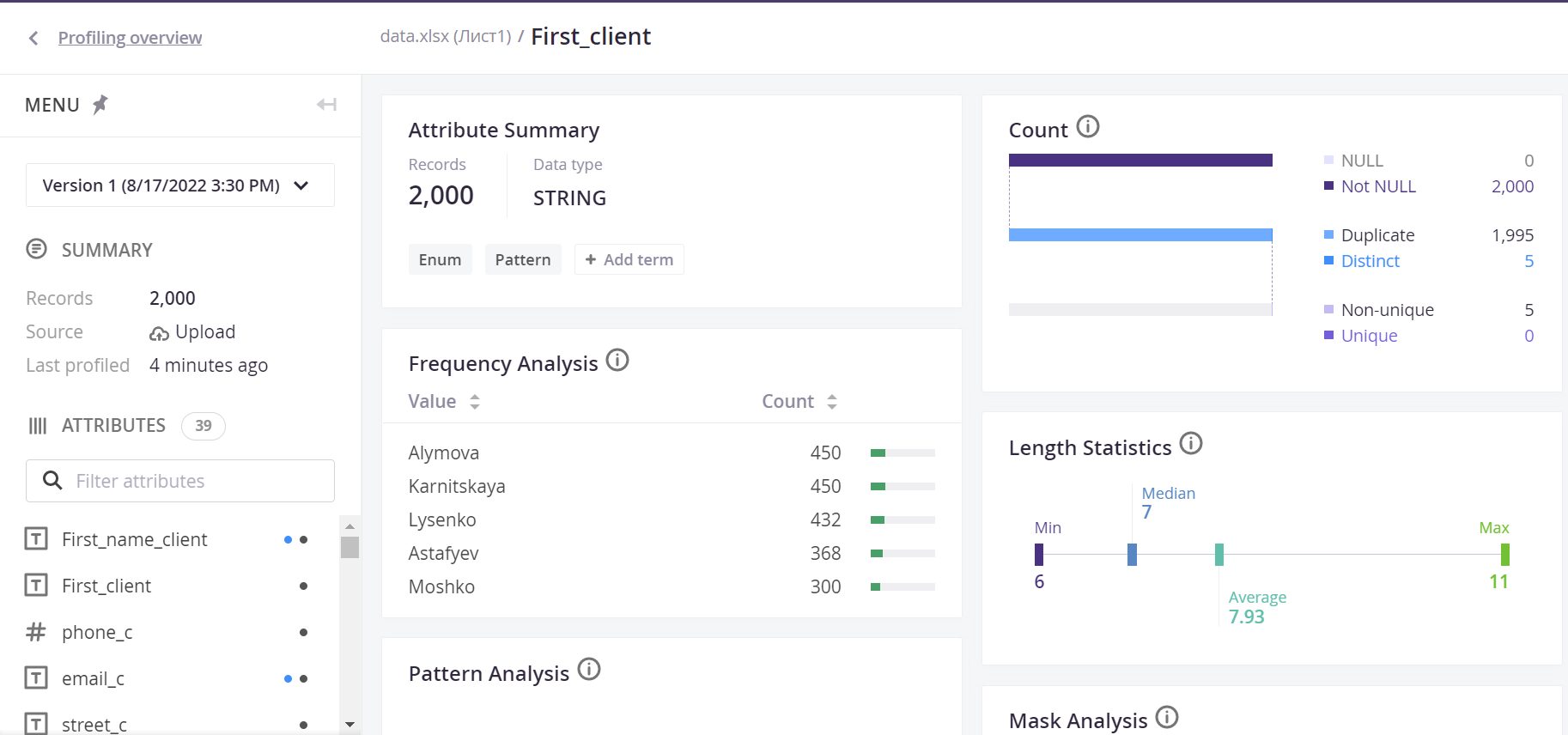
After processing the data, the following page opens, which contains several sections:



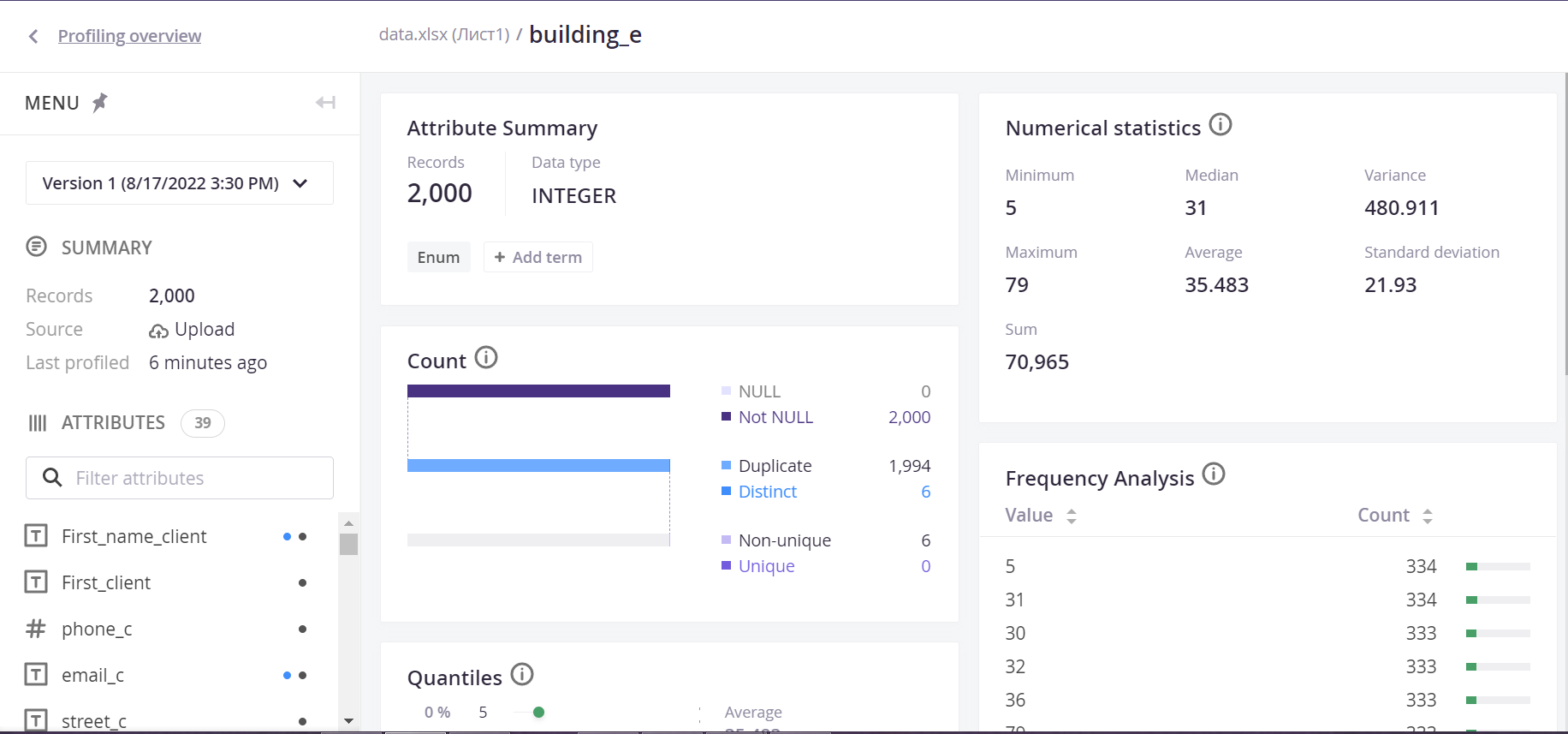
Переходим в раздел ‘profile’:



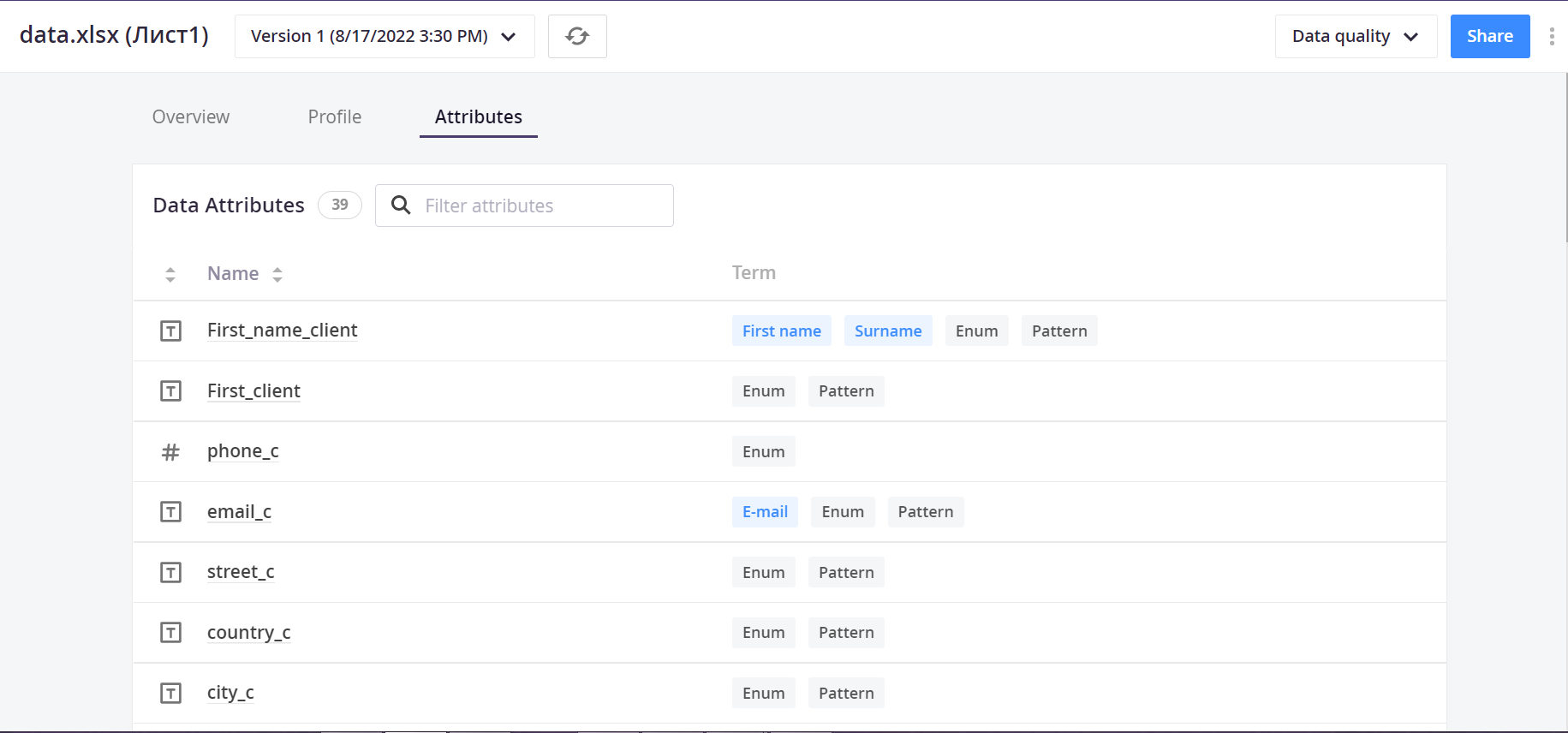
Analysis for each column of our table:



Building\_e column analysis:



The attributes section provides the following analysis data:

****