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:

```
1  CREATE OR REPLACE PACKAGE pkg_etl_transactions_dw
2  AS
3     PROCEDURE load_CLEAN_TRANSACTIONS_DW;
4  END pkg_etl_transactions_dw;
5
```

Script Output

Task completed in 0,064 seconds

```
Package PKG_ETL_TRANSACTIONS_DW compiled
```

Create procedure load_clean_transactions_dw in package pkg_etl_transactions_dw:

```
6  CREATE OR REPLACE PACKAGE body pkg_etl_transactions_dw
7  AS
8    PROCEDURE load_CLEAN_TRANSACTIONS_DW
9     AS
10      BEGIN
11       DECLARE
12        period_id_v NUMBER;
13        TYPE CURSOR_VARCHAR IS TABLE OF varchar2(100);
14        TYPE CURSOR_NUMBER IS TABLE OF number(10);
15        TYPE CURSOR_DATE IS TABLE OF date;
16        TYPE CURSOR_DECIMAL IS TABLE OF decimal(11,2);
17        TYPE BIG_CURSOR IS REF CURSOR ;
18
19        ALL_INF BIG_CURSOR;
20
21        CLIENT_ID CURSOR_NUMBER;
22        EMPLOYEE_ID CURSOR_NUMBER;
23        RESTAURANT_ID CURSOR_NUMBER;
24        DATE_ID CURSOR_DATE;
25        PAYMENT_METHOD_ID CURSOR_NUMBER;
```

Script Output | Query Result | Task completed in 0,433 seconds

```
Package Body PKG_ETL_TRANSACTIONS_DW compiled
```

Execute procedure load_clean_transactions_dw from package pkg_etl_transactions_dw:

```
119  EXECUTE pkg_etl_transactions_dw.load_CLEAN_TRANSACTIONS_DW;
```

Script Output | Query Result | Task completed in 32,027 seconds

```
PL/SQL procedure successfully completed.
```

Select from table dw_data.order_fact:

```
121  SELECT * FROM dw_data.order_fact
122  ORDER BY 1;
```

Script Output | Query Result | Query Result 1 | SQL | Fetched 50 rows in 0,175 seconds

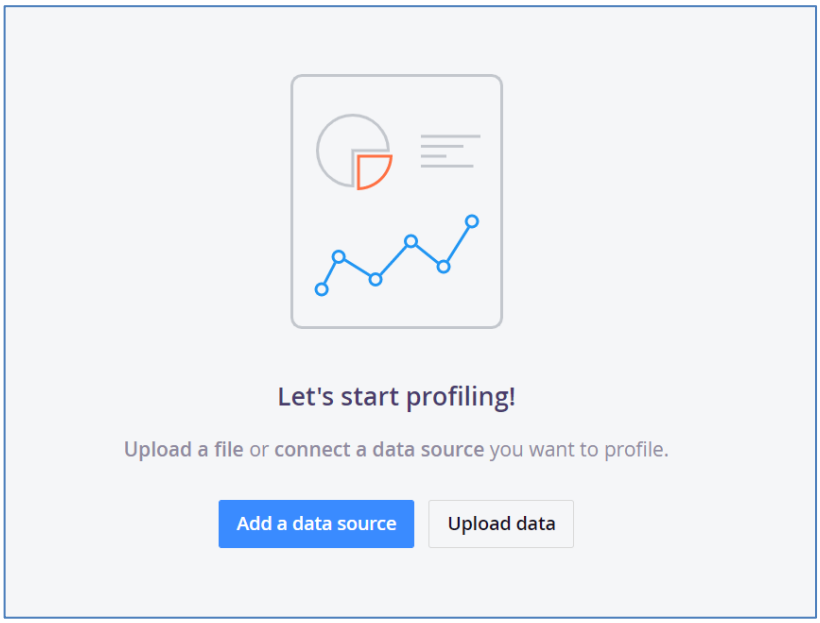
	ORDER_ID	CLIENT_ID	EMPLOYEE_ID	RESTAURANT_ID	DATE_ID	PERIOD_ID	PAYMENT_METHOD_ID	DISH_ID	DISH_AMOUNT	TOTAL_COST	DELIVERY
1	1	3	21	2	02.09.21	3	2	2	26	664	Y
2	2	3	21	1	02.09.21	3	2	4	9	560	Y
3	3	3	13	1	02.09.21	3	2	5	13	1138	Y
4	4	3	1	2	05.09.21	3	2	3	4	390	Y
5	5	3	13	2	05.09.21	3	2	1	17	263	Y
6	6	3	14	4	24.10.21	3	1	4	3	189	Y
7	7	3	21	4	24.10.21	3	1	3	13	1201	Y
8	8	3	21	1	24.10.21	3	1	1	78	1182	Y
9	9	3	13	1	24.10.21	3	1	2	13	325	Y
10	10	3	11	2	25.10.21	3	1	1	83	1251	Y
11	11	3	8	3	26.10.21	3	1	4	16	988	Y
12	12	3	11	5	28.10.21	3	1	3	15	1344	Y
13	13	3	13	5	28.10.21	3	1	2	16	422	Y
14	14	3	11	4	26.10.21	3	1	3	5	458	Y
15	15	3	14	1	26.10.21	3	1	2	15	395	Y
16	16	3	14	1	27.10.21	3	1	5	14	1226	Y
17	17	3	13	3	28.10.21	3	1	1	97	1458	Y

Getting to know the site https://app.ataccama.com/catalog

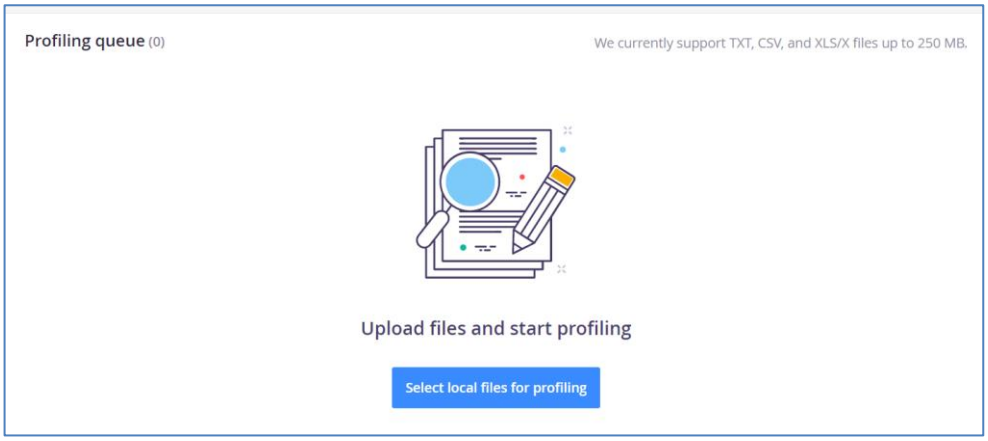
To begin with, we register on the site and choose to analyze our data.

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

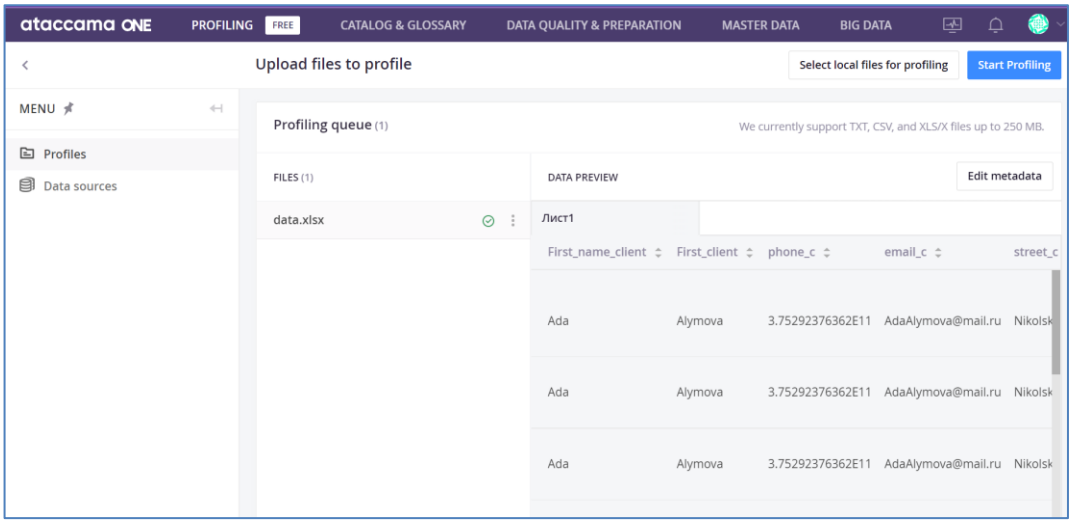
On the site, choose to add your data:

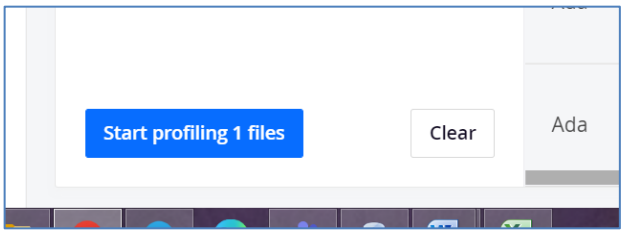


Adding an Excel document with our data:

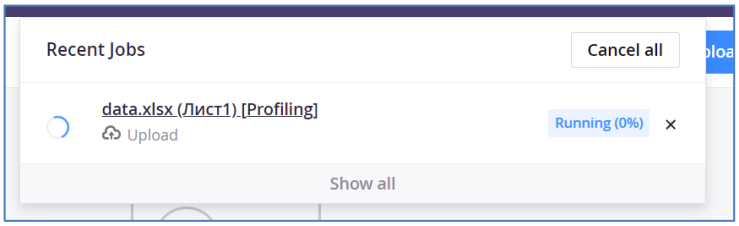


When our data is loaded, we select ‘start profiling 1 files’:

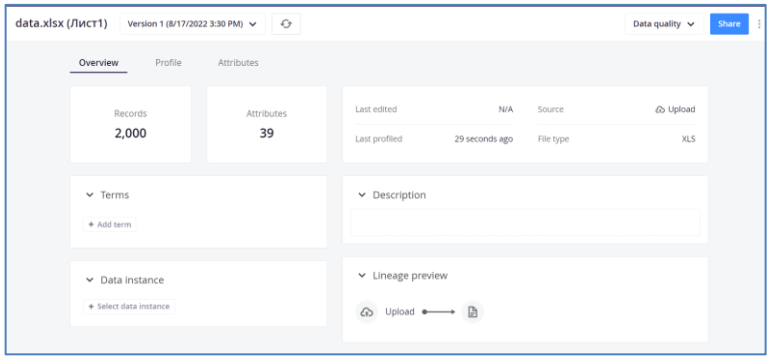




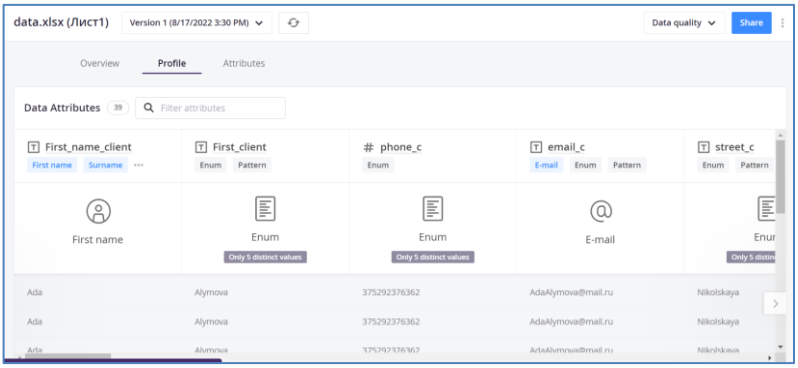
We are waiting for the processing of our data:



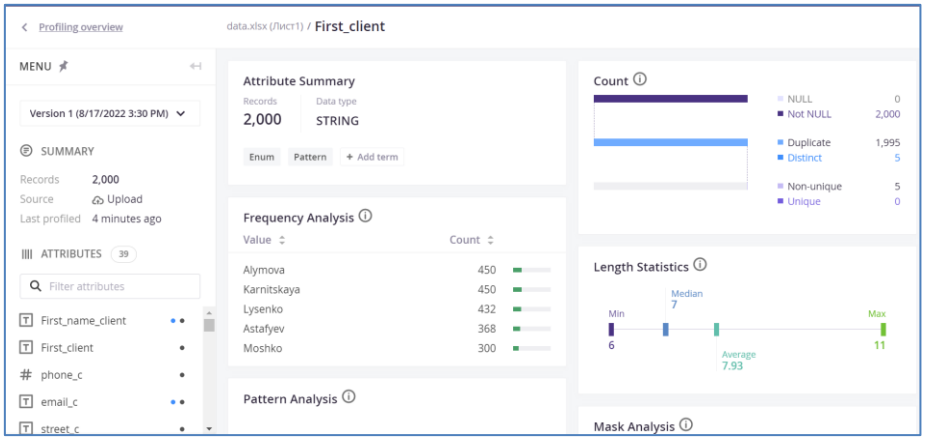
After processing the data, the following page opens, which contains several sections:



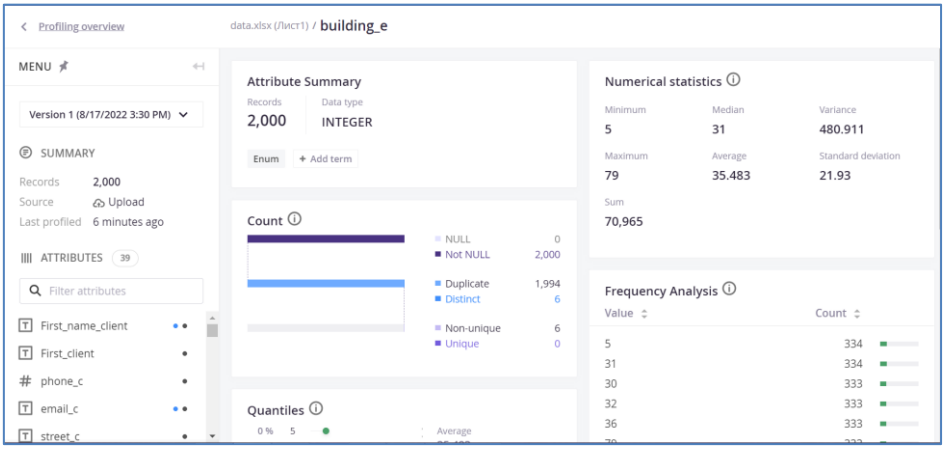
Переходим в раздел ‘profile’:



Analysis for each column of our table:



Building_e column analysis:



The attributes section provides the following analysis data: