Ростелеком
ИТ школа

ИТОГОВАЯ РАБОТА

Название программы	**«Анализ данных в Low-code платформах»**
Группа обучения	**«ПИ ППС»**
Срок обучения	**«с 10.10.2025 по 30.11.2025»**
Дорофеева Виктория Ивановна	
Название темы	**Исследовательский анализ онлайн-продажи и доставки еды с помощью платформы Knime и BI-аналитики посредством Yandex DataLens**

Москва 2025 г.

Ссылки на ресурсы

Весь проект находится на GitHub

https://github.com/VDorofey/Project_Low_code_DorofeevaVI/blob/main/README.md

Данные взяты с Kaggle

https://www.kaggle.com/datasets/sudarshan24byte/online-food-dataset

Ссылка на проект в Knime

https://github.com/VDorofey/Project_Low_code_DorofeevaVI/blob/main/Online_food_project_DorofeevaVI.knwf

Ссылка на ЯндексDataLens
https://datalens.yandex/yjs4kkxdah5yj

Исходные данные кейса

Online Food Order Dataset

Набор данных содержит информацию, собранную с онлайн-платформы для заказа еды за определенный период времени. Он охватывает различные атрибуты, связанные с родом занятий, размером семьи, отзывами и т. д.

Цель: Этот набор данных можно использовать для изучения взаимосвязи между демографическими/локационными факторами и поведением при заказе еды онлайн, анализа отзывов клиентов для улучшения качества обслуживания и потенциального прогнозирования предпочтений или поведения клиентов на основе демографических и локационных атрибутов.

Attributes (атрибуты):

Demographic Information (демографическая информация):

1. Age: Age of the customer (Возраст: Возраст клиента).
2. Gender: Gender of the customer (Пол: Пол клиента).
3. Marital Status: Marital status of the customer (Семейное положение: Семейное положение клиента).
4. Occupation: Occupation of the customer (Род занятий: Род занятий клиента).

5. Monthly Income: Monthly income of the customer (Ежемесячный доход: Ежемесячный доход клиента).
6. Educational Qualifications: Educational qualifications of the customer (Образовательная квалификация: Образовательная квалификация клиента).
7. Family Size: Number of individuals in the customer's family (Размер семьи: Количество человек в семье клиента).

Location Information (информация о местоположении):

8. Latitude: Latitude of the customer's location (Широта: Широта местоположения клиента).
9. Longitude: Longitude of the customer's location (Долгота: Долгота местоположения клиента).
10. Pin Code: Pin code of the customer's location (PIN-код: PIN-код местоположения клиента).

Order Details (сведения о заказе):

11. Output: Current status of the order (e.g., pending, confirmed, delivered) (Выходные данные: Текущий статус заказа (например, ожидающий, подтвержденный, доставленный)).
12. Feedback: Feedback provided by the customer after receiving the order (Обратная связь: обратная связь, предоставленная клиентом после получения заказа).

Предобработка данных и исследовательский анализ

Исходные данные содержат 388 строк и 13 признаков. Пропусков нет. Много объектных данных, которые придется обработать перед МО.

Исследования проводились на платформе KNIME. Опишем этапы исследования:CSV Reader и первичный анализ таблицы.

File Table - 8:1 - CSV Reader

File Edit Hilite Navigation View

Table "default" - Rows: 388 Spec - Columns: 13 Properties Flow Variables

Row ID	Age	Gender	Marital ...	Occupa...	Monthl...	Educati...	Family ...	latitude	longitude	Pin code	Output	Feedback	Column12
Row0	20	Female	Single	Student	No Income	Post Graduate	4	12.977	77.599	560001	Yes	Positive	Yes
Row1	24	Female	Single	Student	Below Rs.10...	Graduate	3	12.977	77.577	560009	Yes	Positive	Yes
Row2	22	Male	Single	Student	Below Rs.10...	Post Graduate	3	12.955	77.659	560017	Yes	Negative	Yes
Row3	22	Female	Single	Student	No Income	Graduate	6	12.947	77.562	560019	Yes	Positive	Yes
Row4	22	Male	Single	Student	Below Rs.10...	Post Graduate	4	12.985	77.553	560010	Yes	Positive	Yes
Row5	27	Female	Married	Employee	More than 5...	Post Graduate	2	12.93	77.685	560103	Yes	Positive	Yes
Row6	22	Male	Single	Student	No Income	Graduate	3	12.977	77.577	560009	Yes	Positive	Yes
Row7	24	Female	Single	Student	No Income	Post Graduate	3	12.983	77.613	560042	Yes	Positive	Yes
Row8	23	Female	Single	Student	No Income	Post Graduate	2	12.977	77.599	560001	Yes	Positive	Yes
Row9	23	Female	Single	Student	No Income	Post Graduate	4	12.985	77.708	560048	Yes	Positive	Yes
Row10	22	Female	Single	Student	No Income	Post Graduate	5	12.985	77.553	560010	Yes	Positive	Yes
Row11	23	Male	Single	Student	Below Rs.10...	Post Graduate	2	12.977	77.577	560009	Yes	Negative	Yes
Row12	23	Male	Single	Student	No Income	Post Graduate	5	12.899	77.576	560078	Yes	Positive	Yes
Row13	21	Male	Single	Student	No Income	Graduate	4	12.977	77.577	560009	Yes	Positive	Yes
Row14	23	Female	Single	Self Employ...	10001 to 25...	Post Graduate	5	12.944	77.574	560004	Yes	Positive	Yes
Row15	24	Female	Single	Student	No Income	Post Graduate	6	12.889	77.64	560068	Yes	Positive	Yes
Row16	28	Female	Single	Employee	25001 to 50...	Post Graduate	2	12.978	77.641	560038	Yes	Positive	Yes
Row17	23	Female	Single	Student	No Income	Graduate	3	12.982	77.626	560008	Yes	Negative	Yes
Row18	25	Male	Single	Student	No Income	Graduate	4	12.899	77.576	560078	Yes	Negative	Yes
Row19	21	Female	Single	Student	Below Rs.10...	Post Graduate	1	12.978	77.641	560038	Yes	Positive	Yes
Row20	24	Male	Single	Student	No Income	Post Graduate	3	12.977	77.577	560009	Yes	Positive	Yes
Row21	22	Male	Single	Student	No Income	Post Graduate	4	13.03	77.605	560032	Yes	Positive	Yes
Row22	22	Female	Single	Student	No Income	Graduate	4	12.998	77.641	560033	Yes	Positive	Yes
Row23	23	Male	Single	Student	No Income	Graduate	4	12.992	77.563	560021	Yes	Positive	Yes
Row24	21	Male	Single	Student	Below Rs.10...	Post Graduate	3	12.931	77.543	560085	Yes	Positive	Yes
Row25	25	Male	Single	Student	No Income	Post Graduate	3	12.982	77.626	560008	Yes	Positive	Yes
Row26	22	Female	Single	Student	No Income	Post Graduate	5	12.935	77.558	560050	Yes	Positive	Yes
Row27	22	Male	Single	Student	No Income	Post Graduate	3	12.915	77.513	560098	Yes	Positive	Yes
Row28	23	Female	Single	Employee	10001 to 25...	Graduate	3	12.985	77.708	560048	Yes	Positive	Yes
Row29	22	Male	Single	Student	Below Rs.10...	Post Graduate	4	13.002	77.571	560003	Yes	Positive	Yes
Row30	22	Female	Single	Employee	10001 to 25...	Graduate	5	12.97	77.75	560066	Yes	Positive	Yes
Row31	22	Male	Single	Student	No Income	Post Graduate	4	12.978	77.641	560038	Yes	Positive	Yes
Row32	25	Male	Married	Employee	More than 5...	Ph.D	4	12.926	77.622	560034	Yes	Positive	Yes
Row33	22	Female	Single	Student	10001 to 25...	Post Graduate	5	12.985	77.553	560010	Yes	Positive	Yes
Row34	22	Female	Single	Student	No Income	Post Graduate	2	12.912	77.645	560102	Yes	Positive	Yes
Row35	25	Male	Single	Student	10001 to 25...	Post Graduate	3	12.931	77.543	560085	Yes	Positive	Yes
Row36	25	Male	Single	Student	No Income	Post Graduate	5	12.977	77.577	560009	No	Positive	No
Row37	32	Female	Prefer not t...	House wife	No Income	Graduate	5	12.982	77.626	560008	Yes	Negative	Yes

2. Воспользуемся Python Script для отдельных действий над таблицей, а именно для перевода гео данных в более удобный формат для дальнейшей визуализации

Dialog - 8:2 - Python Script

Python Script

```
1  import pandas as pd
2  import knime.scripting.io as knio
3
4  df = knio.input_tables[0].to_pandas()
5  # This example script simply outputs the node's input table.
6  # Выводим информацию в консоль (Output Console) KNIME
7  print("--- Информация о DataFrame (df.info()): ---")
8  df.info()
9  print("-----------------------------------------------")
10
11 df['geo_point'] = df.apply(lambda row: [row['latitude'], row['longitude']], axis=1)
12 knio.output_tables[0] = knio.Table.from_pandas(df)
13
```

Ask K-AI | Run selected lines | Run all

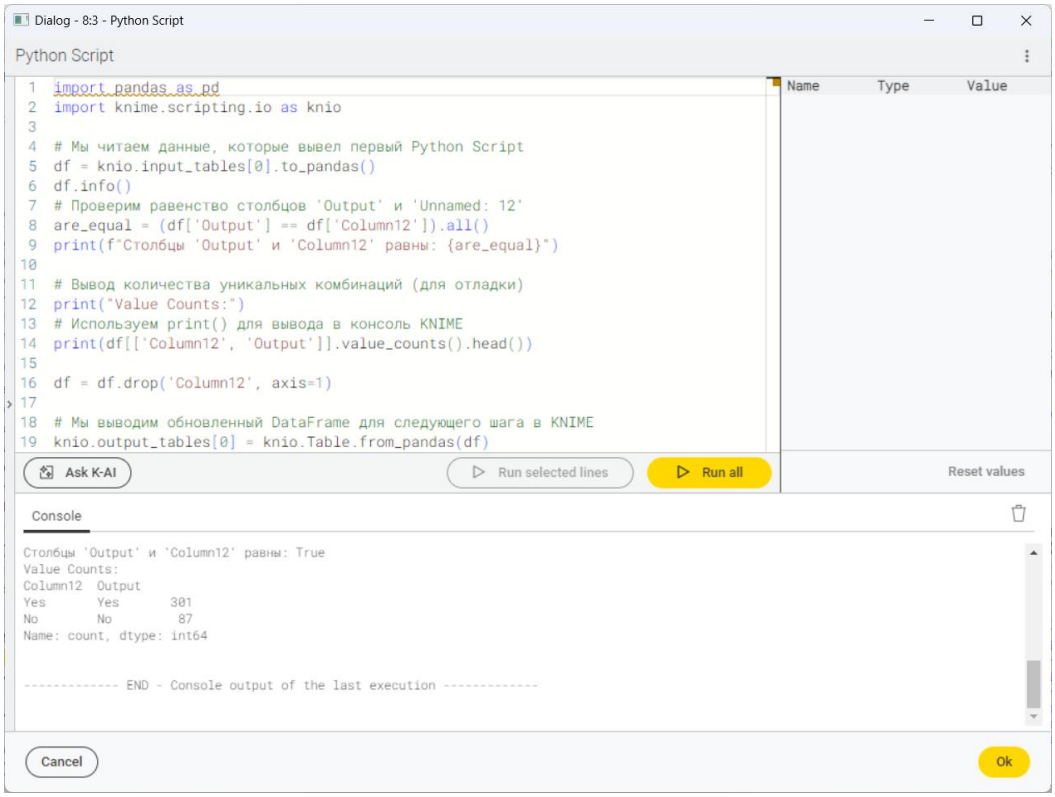
Console

```
 6   Family size          388 non-null    int32
 7   latitude             388 non-null    float64
 8   longitude            388 non-null    float64
 9   Pin code             388 non-null    int32
 10  Output               388 non-null    string
 11  Feedback             388 non-null    string
 12  Column12             388 non-null    string
dtypes: float64(2), int32(3), string(8)
memory usage: 37.9 KB
-----------------------------------------------


------------ END - Console output of the last execution ------------
```

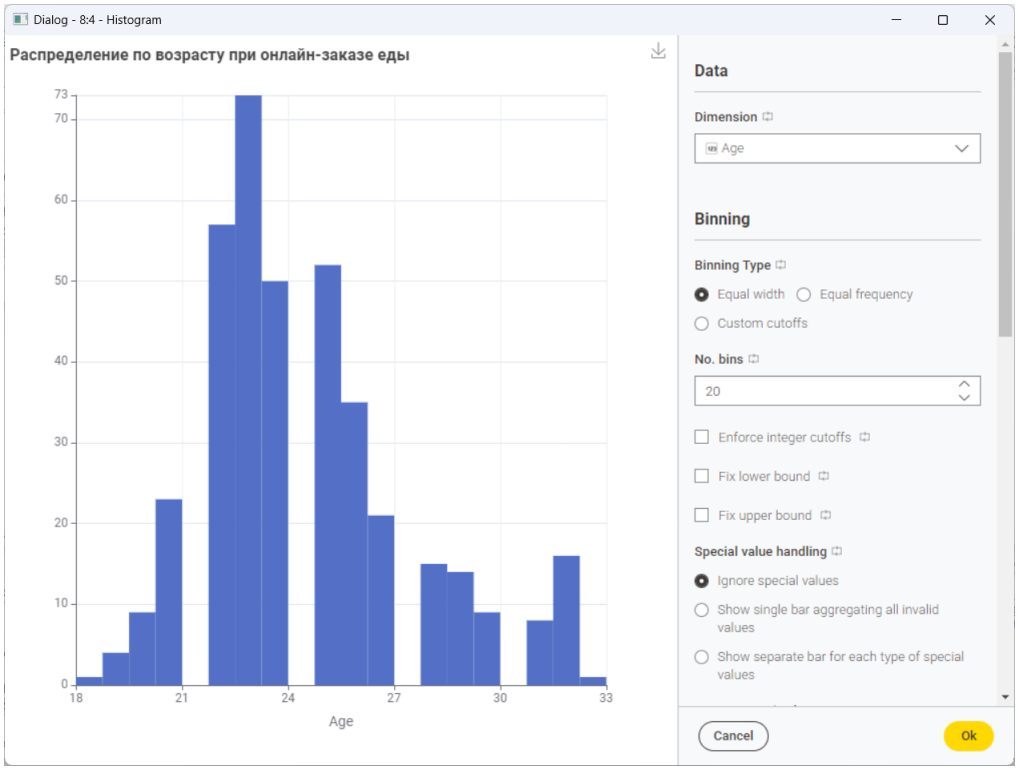
Cancel | Ok

А также воспользуемся Python Script для сравнения двух столбцов с последующим удалением одного из них в силу полного совпадения.

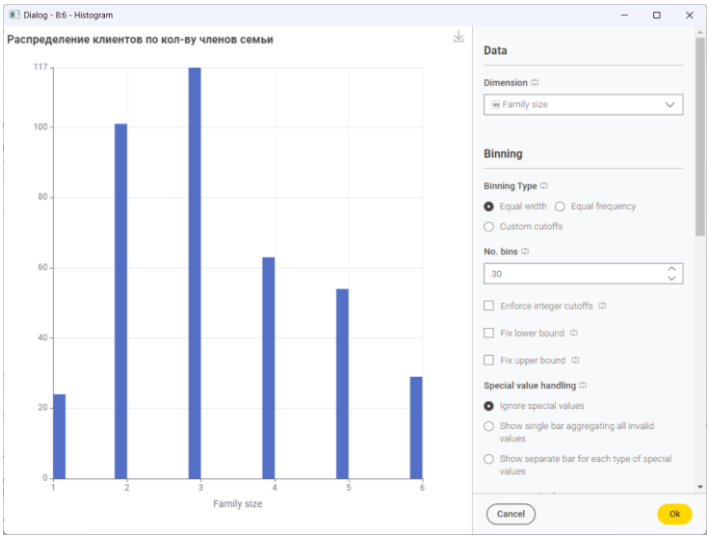


3. Исследуем данные и визуализируем полученные результаты

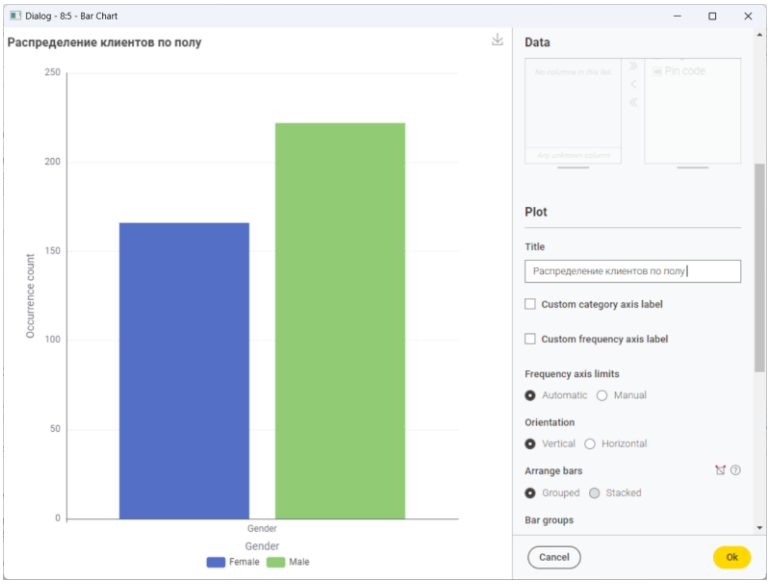
3.1 с помощью Histogram визуализируем по возрасту онлайн заказы еды



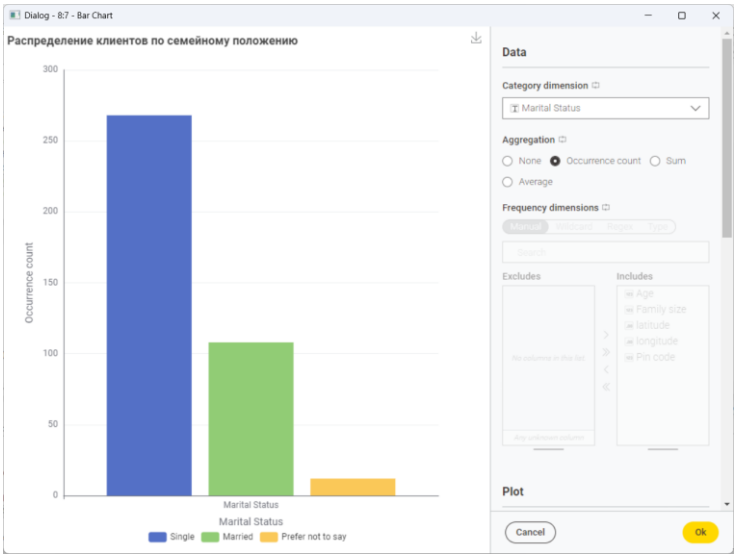
И распределение клиентов по составу семьи (количество членов семьи)



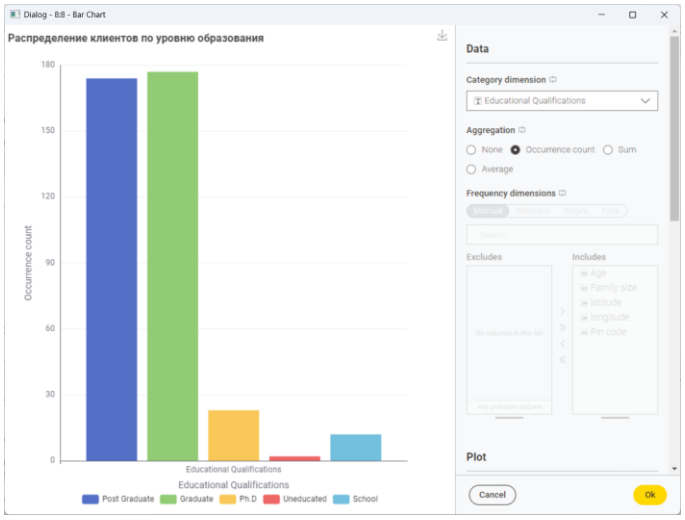
3.2 Посредством Bar Chart построим распредеоение клиентов по полу



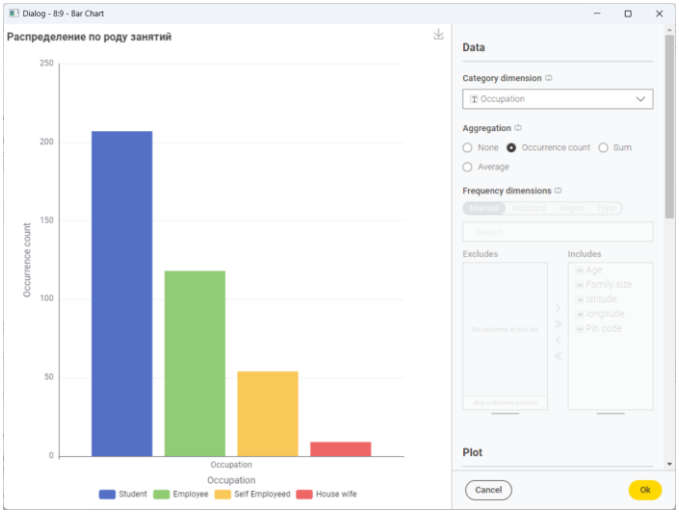
По семейному положению



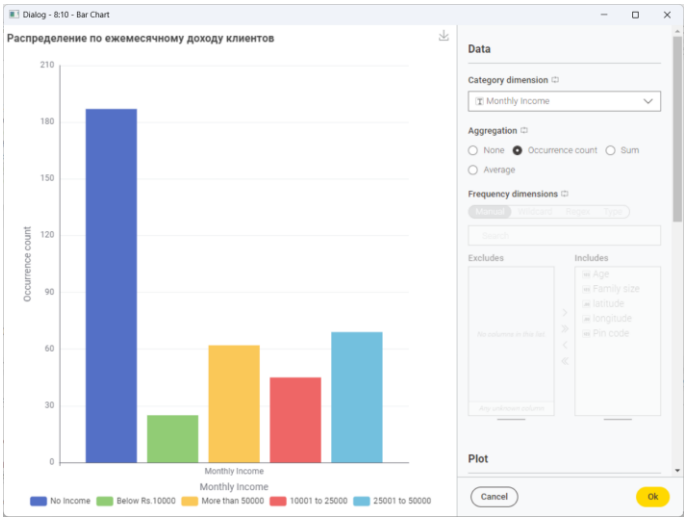
По уровню образования



По роду занятий



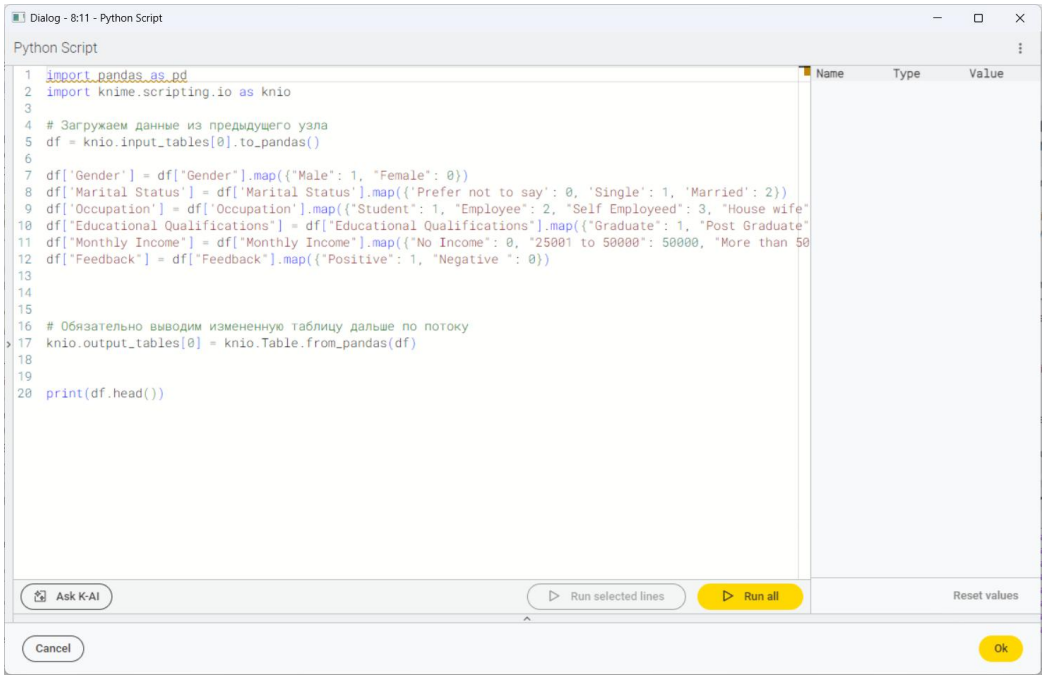
И по ежемесячному доходу клиентов



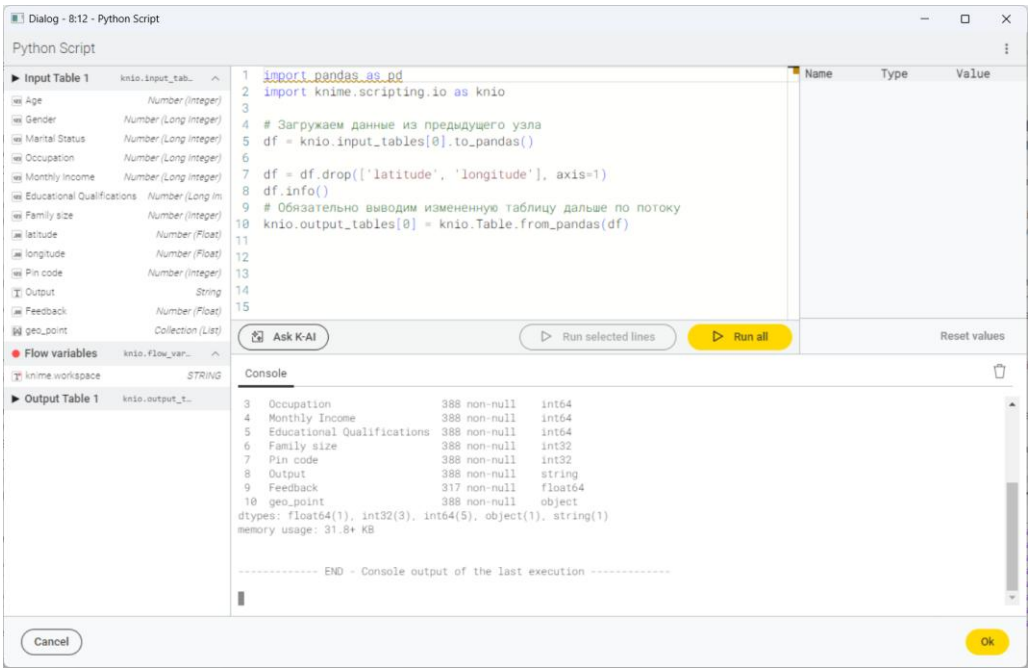
Прогнозирование данных по онлайн покупкам еды с помощью МО (решение задачи классификации)

Распишем последовательно все этапы прогнозирования данных:

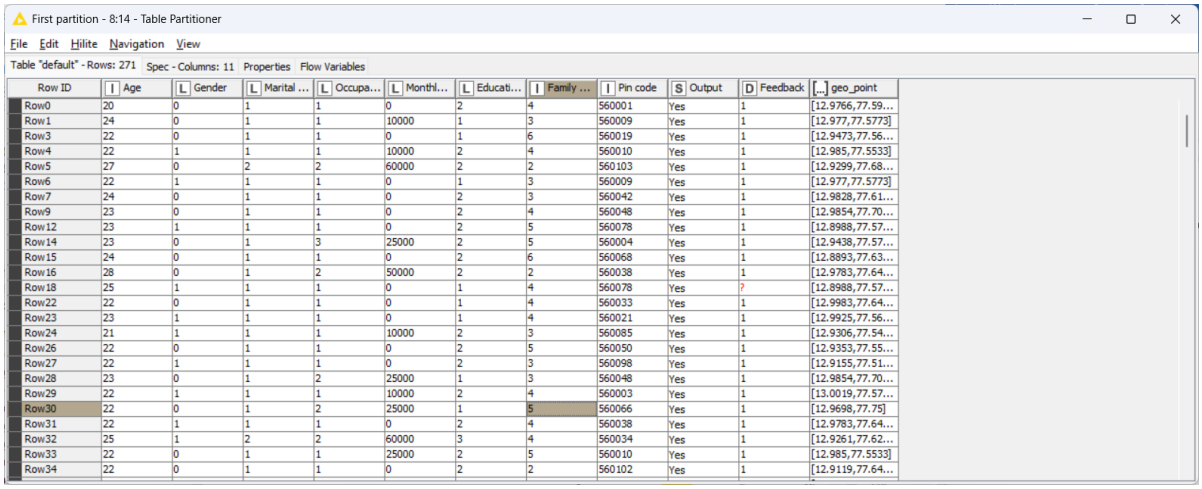
1. Воспользуемся Python Script для перевода строковых данных в числовые (за исключением таргетного признака ‘Output’



2. В предверии МО удалилим признаки, которые касаются географических данных



3. С помощью Table Partitioner разделяем данные на тренировочные и тестовые с учетом 70/30 (включая ytrain и ytest)
   Имеем тренировочные данные:

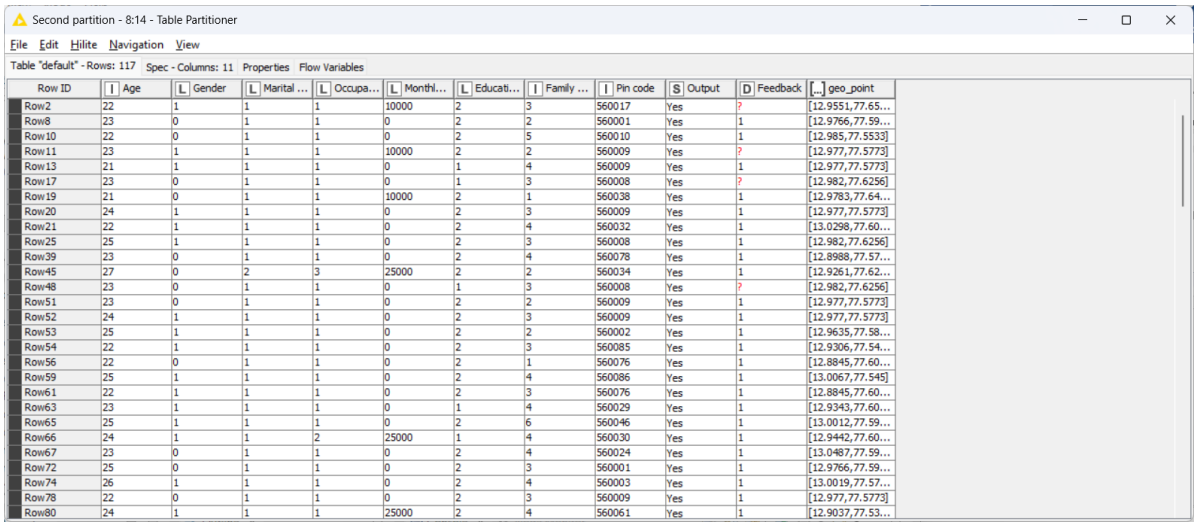

First partition - 8:14 - Table Partitioner

File Edit Hilite Navigation View

Table "default" - Rows: 271 Spec - Columns: 11 Properties Flow Variables

Row ID	Age	Gender	Marital ...	Occupa...	Monthl...	Educati...	Family ...	Pin code	Output	Feedback	geo_point
Row0	20	0	1	1	0	2	4	560001	Yes	1	[12.9766,77.59...
Row1	24	0	1	1	10000	1	3	560009	Yes	1	[12.977,77.5773]
Row3	22	0	1	1	0	1	6	560019	Yes	1	[12.9473,77.56...
Row4	22	1	1	1	10000	2	4	560010	Yes	1	[12.985,77.5533]
Row5	27	0	2	2	60000	2	2	560103	Yes	1	[12.9299,77.68...
Row6	22	1	1	1	0	1	3	560009	Yes	1	[12.977,77.5773]
Row7	24	0	1	1	0	2	3	560042	Yes	1	[12.9828,77.61...
Row9	23	0	1	1	0	2	4	560048	Yes	1	[12.9854,77.70...
Row12	23	1	1	1	0	2	5	560078	Yes	1	[12.8988,77.57...
Row14	23	0	1	3	25000	2	5	560004	Yes	1	[12.9438,77.57...
Row15	24	0	1	1	0	2	6	560068	Yes	1	[12.8893,77.63...
Row16	28	0	1	2	50000	2	2	560038	Yes	1	[12.9783,77.64...
Row18	25	1	1	1	0	1	4	560078	Yes	?	[12.8988,77.57...
Row22	22	0	1	1	0	1	4	560033	Yes	1	[12.9983,77.64...
Row23	23	1	1	1	0	1	4	560021	Yes	1	[12.9925,77.56...
Row24	21	1	1	1	10000	2	3	560085	Yes	1	[12.9306,77.54...
Row26	22	0	1	1	0	2	5	560050	Yes	1	[12.9353,77.55...
Row27	22	1	1	1	0	2	3	560098	Yes	1	[12.9155,77.51...
Row28	23	0	1	2	25000	1	3	560048	Yes	1	[12.9854,77.70...
Row29	22	1	1	1	10000	2	4	560003	Yes	1	[13.0019,77.57...
Row30	22	0	1	2	25000	1	5	560066	Yes	1	[12.9698,77.75]
Row31	22	1	1	1	0	2	4	560038	Yes	1	[12.9783,77.64...
Row32	25	1	2	2	60000	3	4	560034	Yes	1	[12.9261,77.62...
Row33	22	0	1	1	25000	2	5	560010	Yes	1	[12.985,77.5533]
Row34	22	0	1	1	0	2	2	560102	Yes	1	[12.9119,77.64...

И тестовые данные:

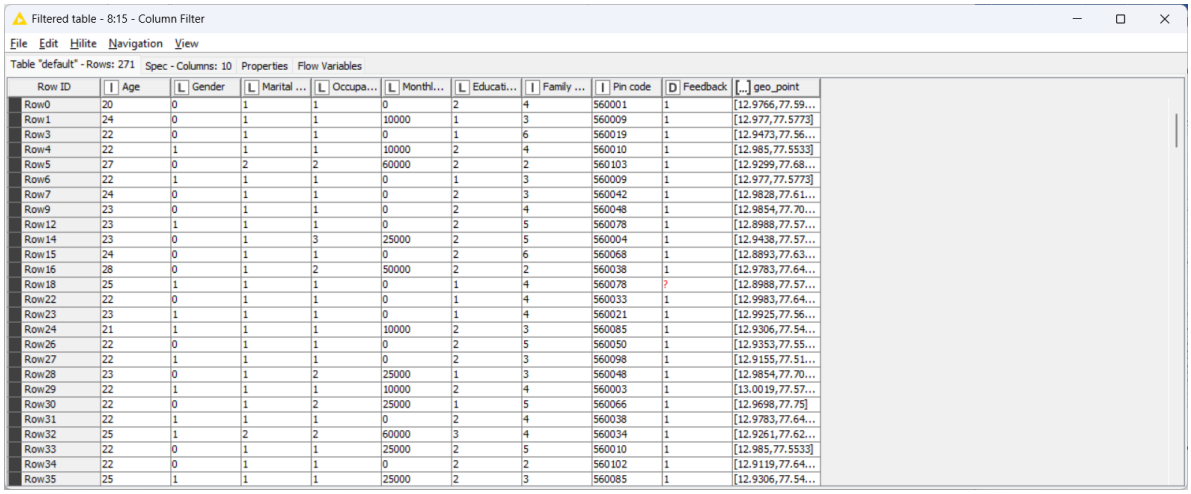

Second partition - 8:14 - Table Partitioner

File Edit Hilite Navigation View

Table "default" - Rows: 117 Spec - Columns: 11 Properties Flow Variables

Row ID	Age	Gender	Marital ...	Occupa...	Monthl...	Educati...	Family ...	Pin code	Output	Feedback	geo_point
Row2	22	1	1	1	10000	2	3	560017	Yes	?	[12.9551,77.65...
Row8	23	0	1	1	0	2	2	560001	Yes	1	[12.9766,77.59...
Row10	22	0	1	1	0	2	5	560010	Yes	1	[12.985,77.5533]
Row11	23	1	1	1	10000	2	2	560009	Yes	?	[12.977,77.5773]
Row13	21	1	1	1	0	1	4	560009	Yes	1	[12.977,77.5773]
Row17	23	0	1	1	0	1	3	560008	Yes	?	[12.982,77.6256]
Row19	21	0	1	1	10000	2	1	560038	Yes	1	[12.9783,77.64...
Row20	24	1	1	1	0	2	3	560009	Yes	1	[12.977,77.5773]
Row21	22	1	1	1	0	2	4	560032	Yes	1	[13.0298,77.60...
Row25	25	1	1	1	0	2	3	560008	Yes	1	[12.982,77.6256]
Row39	23	0	1	1	0	2	4	560078	Yes	1	[12.8988,77.57...
Row45	27	0	2	3	25000	2	2	560034	Yes	1	[12.9261,77.62...
Row48	23	0	1	1	0	1	3	560008	Yes	?	[12.982,77.6256]
Row51	23	0	1	1	0	2	2	560009	Yes	1	[12.977,77.5773]
Row52	24	1	1	1	0	2	3	560009	Yes	1	[12.977,77.5773]
Row53	25	1	1	1	0	2	2	560002	Yes	1	[12.9635,77.58...
Row54	22	1	1	1	0	2	3	560085	Yes	1	[12.9306,77.54...
Row56	22	0	1	1	0	2	1	560076	Yes	1	[12.8845,77.60...
Row59	25	1	1	1	0	2	4	560086	Yes	1	[13.0067,77.545]
Row61	22	1	1	1	0	2	3	560076	Yes	1	[12.8845,77.60...
Row63	23	1	1	1	0	1	4	560029	Yes	1	[12.9343,77.60...
Row65	25	1	1	1	0	2	6	560046	Yes	1	[13.0012,77.59...
Row66	24	1	1	2	25000	1	4	560030	Yes	1	[12.9442,77.60...
Row67	23	0	1	1	0	2	4	560024	Yes	1	[13.0487,77.59...
Row72	25	0	1	1	0	2	3	560001	Yes	1	[12.9766,77.59...
Row74	26	1	1	1	0	2	4	560003	Yes	1	[13.0019,77.57...
Row78	22	0	1	1	0	2	3	560009	Yes	1	[12.977,77.5773]
Row80	24	1	1	1	25000	2	4	560061	Yes	1	[12.9037,77.53...

4. Далее с помощью Column Filter выделяем Xtrain, ytrain, Xtest, ytest/
   Покажем это только для тренировочных данных:
   Xtrain

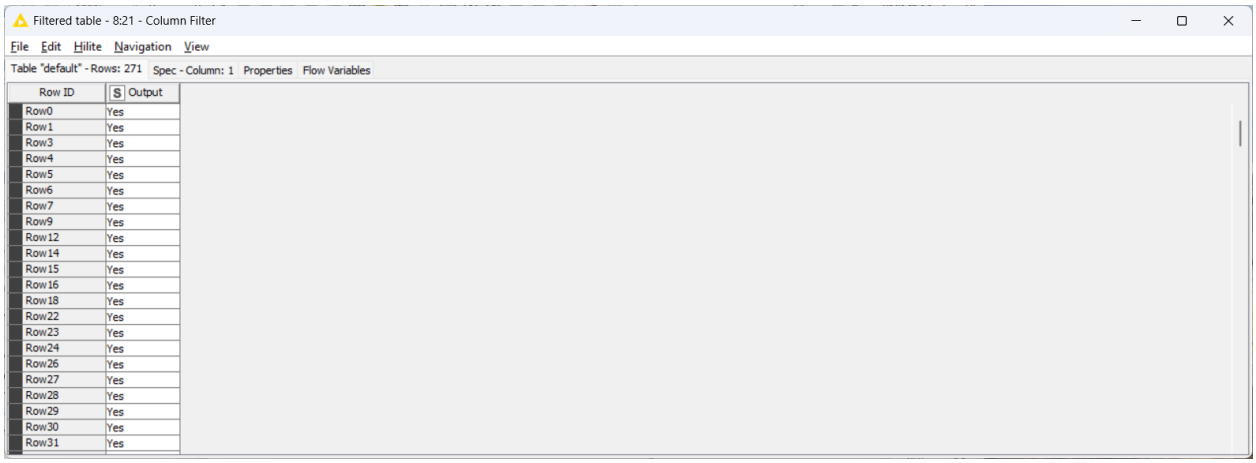

Filtered table - 8:15 - Column Filter

File Edit Hilite Navigation View

Table "default" - Rows: 271 Spec - Columns: 10 Properties Flow Variables

Row ID	Age	Gender	Marital ...	Occupa...	Monthl...	Educati...	Family ...	Pin code	Feedback	geo_point
Row0	20	0	1	1	0	2	4	560001	1	[12.9766,77.59...
Row1	24	0	1	1	10000	1	3	560009	1	[12.977,77.5773]
Row3	22	0	1	1	0	1	6	560019	1	[12.9473,77.56...
Row4	22	1	1	1	10000	2	4	560010	1	[12.985,77.5533]
Row5	27	0	2	2	60000	2	2	560103	1	[12.9299,77.68...
Row6	22	1	1	1	0	1	3	560009	1	[12.977,77.5773]
Row7	24	0	1	1	0	2	3	560042	1	[12.9828,77.61...
Row9	23	0	1	1	0	2	4	560048	1	[12.9854,77.70...
Row12	23	1	1	1	0	2	5	560078	1	[12.8988,77.57...
Row14	23	0	1	3	25000	2	5	560004	1	[12.9438,77.57...
Row15	24	0	1	1	0	2	6	560068	1	[12.8893,77.63...
Row16	28	0	1	2	50000	2	2	560038	1	[12.9783,77.64...
Row18	25	1	1	1	0	1	4	560078	?	[12.8988,77.57...
Row22	22	0	1	1	0	1	4	560033	1	[12.9983,77.64...
Row23	23	1	1	1	0	1	4	560021	1	[12.9925,77.56...
Row24	21	1	1	1	10000	2	3	560085	1	[12.9306,77.54...
Row26	22	0	1	1	0	2	5	560050	1	[12.9353,77.55...
Row27	22	1	1	1	0	2	3	560098	1	[12.9155,77.51...
Row28	23	0	1	2	25000	1	3	560048	1	[12.9854,77.70...
Row29	22	1	1	1	10000	2	4	560003	1	[13.0019,77.57...
Row30	22	0	1	2	25000	1	5	560066	1	[12.9698,77.75]
Row31	22	1	1	1	0	2	4	560038	1	[12.9783,77.64...
Row32	25	1	2	2	60000	3	4	560034	1	[12.9261,77.62...
Row33	22	0	1	1	25000	2	5	560010	1	[12.985,77.5533]
Row34	22	0	1	1	0	2	2	560102	1	[12.9119,77.64...
Row35	25	1	1	1	25000	2	3	560085	1	[12.9306,77.54...

Ytrain

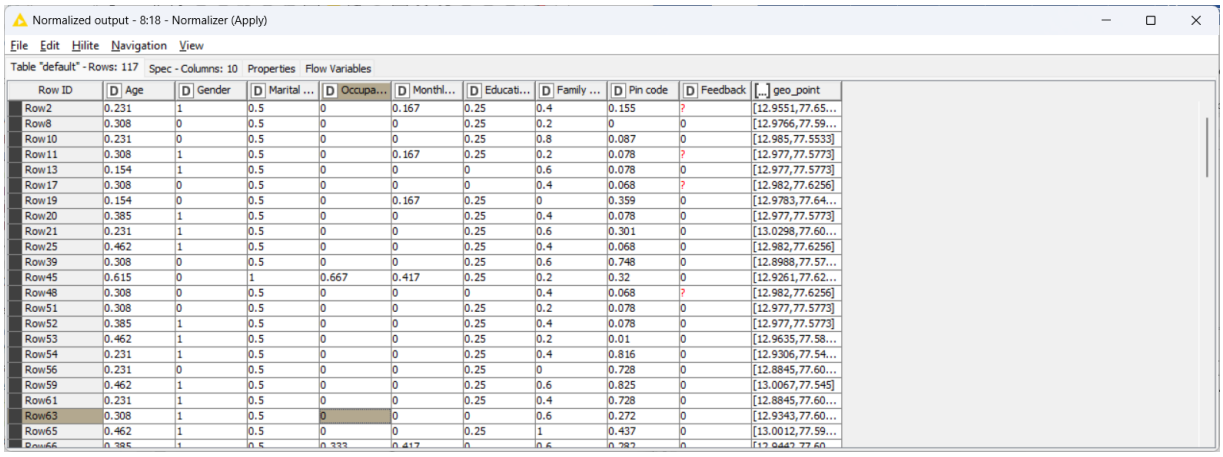

Filtered table - 8:21 - Column Filter

File Edit Hilite Navigation View

Table "default" - Rows: 271 Spec - Column: 1 Properties Flow Variables

Row ID	S Output
Row0	Yes
Row1	Yes
Row3	Yes
Row4	Yes
Row5	Yes
Row6	Yes
Row7	Yes
Row9	Yes
Row12	Yes
Row14	Yes
Row15	Yes
Row16	Yes
Row18	Yes
Row22	Yes
Row23	Yes
Row24	Yes
Row26	Yes
Row27	Yes
Row28	Yes
Row29	Yes
Row30	Yes
Row31	Yes

5. Затем масштабируем данные Xtrain с помощью Normalizer (обучаем с помощью fit) и применяем результат обучения к Xtest с помощью Normalizer (Apply) Получаем для тестовых данных

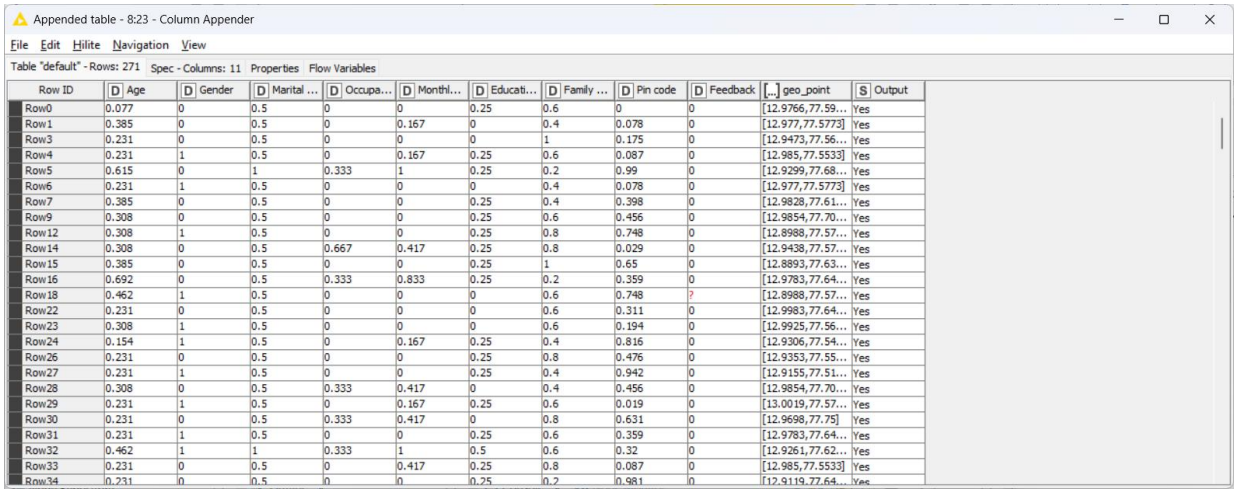

Normalized output - 8:18 - Normalizer (Apply)

File Edit Hilite Navigation View

Table "default" - Rows: 117 Spec - Columns: 10 Properties Flow Variables

Row ID	D Age	D Gender	D Marital ...	D Occupa...	D Monthl...	D Educati...	D Family ...	D Pin code	D Feedback	[...] geo_point
Row2	0.231	1	0.5	0	0.167	0.25	0.4	0.155	?	[12.9551,77.65...
Row8	0.308	0	0.5	0	0	0.25	0.2	0	0	[12.9766,77.59...
Row10	0.231	0	0.5	0	0	0.25	0.8	0.087	0	[12.985,77.5533]
Row11	0.308	1	0.5	0	0.167	0.25	0.2	0.078	?	[12.977,77.5773]
Row13	0.154	1	0.5	0	0	0	0.6	0.078	0	[12.977,77.5773]
Row17	0.308	0	0.5	0	0	0	0.4	0.068	?	[12.982,77.6256]
Row19	0.154	0	0.5	0	0.167	0.25	0	0.359	0	[12.9783,77.64...
Row20	0.385	1	0.5	0	0	0.25	0.4	0.078	0	[12.977,77.5773]
Row21	0.231	1	0.5	0	0	0.25	0.6	0.301	0	[13.0298,77.60...
Row25	0.462	1	0.5	0	0	0.25	0.4	0.068	0	[12.982,77.6256]
Row39	0.308	0	0.5	0	0	0.25	0.6	0.748	0	[12.8988,77.57...
Row45	0.615	0	1	0.667	0.417	0.25	0.2	0.32	0	[12.9261,77.62...
Row48	0.308	0	0.5	0	0	0	0.4	0.068	?	[12.982,77.6256]
Row51	0.308	0	0.5	0	0	0.25	0.2	0.078	0	[12.977,77.5773]
Row52	0.385	1	0.5	0	0	0.25	0.4	0.078	0	[12.977,77.5773]
Row53	0.462	1	0.5	0	0	0.25	0.2	0.01	0	[12.9635,77.58...
Row54	0.231	1	0.5	0	0	0.25	0.4	0.816	0	[12.9306,77.54...
Row56	0.231	0	0.5	0	0	0.25	0	0.728	0	[12.8845,77.60...
Row59	0.462	1	0.5	0	0	0.25	0.6	0.825	0	[13.0067,77.545]
Row61	0.231	1	0.5	0	0	0.25	0.4	0.728	0	[12.8845,77.60...
Row63	0.308	1	0.5	0	0	0	0.6	0.272	0	[12.9343,77.60...
Row65	0.462	1	0.5	0	0	0.25	1	0.437	0	[13.0012,77.59...

6. Далее с помощью Column Appender объединяем тренировочные данные в одну таблицу (Xtrain после нормировки и ytrain)


Appended table - 8:23 - Column Appender

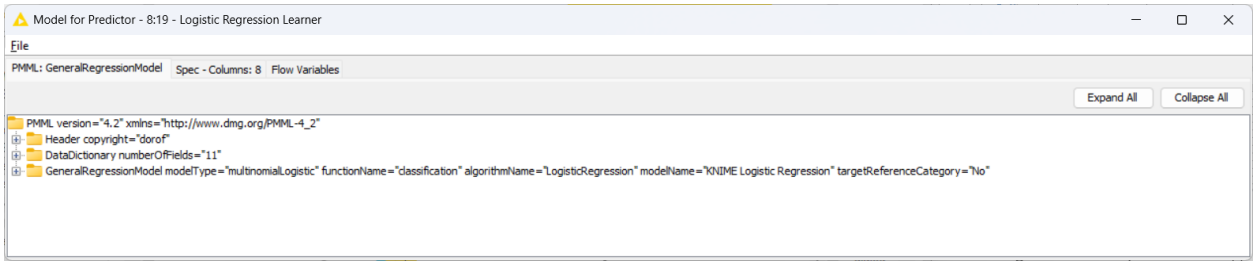
File Edit Hilite Navigation View

Table "default" - Rows: 271 Spec - Columns: 11 Properties Flow Variables

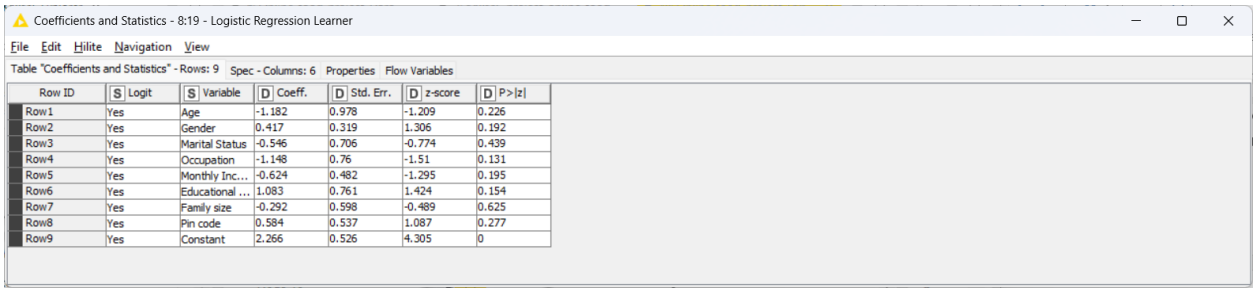
Row ID	D Age	D Gender	D Marital ...	D Occupa...	D Monthl...	D Educati...	D Family ...	D Pin code	D Feedback	[...] geo_point	S Output
Row0	0.077	0	0.5	0	0	0.25	0.6	0	0	[12.9766,77.59...	Yes
Row1	0.385	0	0.5	0	0.167	0	0.4	0.078	0	[12.977,77.5773]	Yes
Row3	0.231	0	0.5	0	0	0	1	0.175	0	[12.9473,77.56...	Yes
Row4	0.231	1	0.5	0	0.167	0.25	0.6	0.087	0	[12.985,77.5533]	Yes
Row5	0.615	0	1	0.333	1	0.25	0.2	0.99	0	[12.9299,77.68...	Yes
Row6	0.231	1	0.5	0	0	0	0.4	0.078	0	[12.977,77.5773]	Yes
Row7	0.385	0	0.5	0	0	0.25	0.4	0.398	0	[12.9828,77.61...	Yes
Row9	0.308	0	0.5	0	0	0.25	0.6	0.456	0	[12.9854,77.70...	Yes
Row12	0.308	1	0.5	0	0	0.25	0.8	0.748	0	[12.8988,77.57...	Yes
Row14	0.308	0	0.5	0.667	0.417	0.25	0.8	0.029	0	[12.9438,77.57...	Yes
Row15	0.385	0	0.5	0	0	0.25	1	0.65	0	[12.8893,77.63...	Yes
Row16	0.692	0	0.5	0.333	0.833	0.25	0.2	0.359	0	[12.9783,77.64...	Yes
Row18	0.462	1	0.5	0	0	0	0.6	0.748	?	[12.8988,77.57...	Yes
Row22	0.231	0	0.5	0	0	0	0.6	0.311	0	[12.9983,77.64...	Yes
Row23	0.308	1	0.5	0	0	0	0.6	0.194	0	[12.9925,77.56...	Yes
Row24	0.154	1	0.5	0	0.167	0.25	0.4	0.816	0	[12.9306,77.54...	Yes
Row26	0.231	0	0.5	0	0	0.25	0.8	0.476	0	[12.9353,77.55...	Yes
Row27	0.231	1	0.5	0	0	0.25	0.4	0.942	0	[12.9155,77.51...	Yes
Row28	0.308	0	0.5	0.333	0.417	0	0.4	0.456	0	[12.9854,77.70...	Yes
Row29	0.231	1	0.5	0	0.167	0.25	0.6	0.019	0	[13.0019,77.57...	Yes
Row30	0.231	0	0.5	0.333	0.417	0	0.8	0.631	0	[12.9698,77.75]	Yes
Row31	0.231	1	0.5	0	0	0.25	0.6	0.359	0	[12.9783,77.64...	Yes
Row32	0.462	1	1	0.333	1	0.5	0.6	0.32	0	[12.9261,77.62...	Yes
Row33	0.231	0	0.5	0	0.417	0.25	0.8	0.087	0	[12.985,77.5533]	Yes

7. После этого можно переходить к процессу обучения модели Logistic Regression Learner и получить на выходе три компонента , которые дает Knime:

Данные модели для предсказания



Коэффициенты модели и статистику

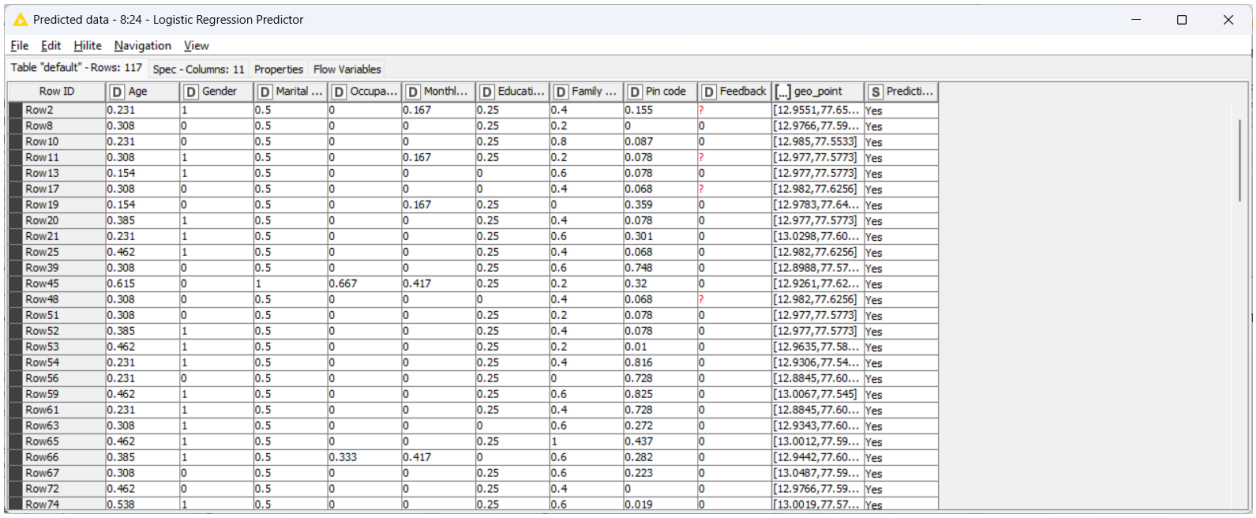


Row ID	Logit	Variable	Coeff.	Std. Err.	z-score	P>|z|
Row1	Yes	Age	-1.182	0.978	-1.209	0.226
Row2	Yes	Gender	0.417	0.319	1.306	0.192
Row3	Yes	Marital Status	-0.546	0.706	-0.774	0.439
Row4	Yes	Occupation	-1.148	0.76	-1.51	0.131
Row5	Yes	Monthly Inc...	-0.624	0.482	-1.295	0.195
Row6	Yes	Educational ...	1.083	0.761	1.424	0.154
Row7	Yes	Family size	-0.292	0.598	-0.489	0.625
Row8	Yes	Pin code	0.584	0.537	1.087	0.277
Row9	Yes	Constant	2.266	0.526	4.305	0

И свойства модели

Row ID	Iterations	LogLikel...
Row0	78	-129.318

8. Далее с помощью Logistic Regression Predictor на основе Xtest и обученной модели получаем прогноз для тестовых данных:



Row ID	Age	Gender	Marital ...	Occupa...	Monthl...	Educati...	Family ...	Pin code	Feedback	geo_point	Predicti...
Row2	0.231	1	0.5	0	0.167	0.25	0.4	0.155	?	[12.9551,77.65...	Yes
Row8	0.308	0	0.5	0	0	0.25	0.2	0	0	[12.9766,77.59...	Yes
Row10	0.231	0	0.5	0	0	0.25	0.8	0.087	0	[12.985,77.5533]	Yes
Row11	0.308	1	0.5	0	0.167	0.25	0.2	0.078	?	[12.977,77.5773]	Yes
Row13	0.154	1	0.5	0	0	0	0.6	0.078	0	[12.977,77.5773]	Yes
Row17	0.308	0	0.5	0	0	0	0.4	0.068	?	[12.982,77.6256]	Yes
Row19	0.154	0	0.5	0	0.167	0.25	0	0.359	0	[12.9783,77.64...	Yes
Row20	0.385	1	0.5	0	0	0.25	0.4	0.078	0	[12.977,77.5773]	Yes
Row21	0.231	1	0.5	0	0	0.25	0.6	0.301	0	[13.0298,77.60...	Yes
Row25	0.462	1	0.5	0	0	0.25	0.4	0.068	0	[12.982,77.6256]	Yes
Row39	0.308	0	0.5	0	0	0.25	0.6	0.748	0	[12.8988,77.57...	Yes
Row45	0.615	0	1	0.667	0.417	0.25	0.2	0.32	0	[12.9261,77.62...	Yes
Row48	0.308	0	0.5	0	0	0	0.4	0.068	?	[12.982,77.6256]	Yes
Row51	0.308	0	0.5	0	0	0.25	0.2	0.078	0	[12.977,77.5773]	Yes
Row52	0.385	1	0.5	0	0	0.25	0.4	0.078	0	[12.977,77.5773]	Yes
Row53	0.462	1	0.5	0	0	0.25	0.2	0.01	0	[12.9635,77.58...	Yes
Row54	0.231	1	0.5	0	0	0.25	0.4	0.816	0	[12.9306,77.54...	Yes
Row56	0.231	0	0.5	0	0	0.25	0	0.728	0	[12.8845,77.60...	Yes
Row59	0.462	1	0.5	0	0	0.25	0.6	0.825	0	[13.0067,77.545]	Yes
Row61	0.231	1	0.5	0	0	0.25	0.4	0.728	0	[12.8845,77.60...	Yes
Row63	0.308	1	0.5	0	0	0	0.6	0.272	0	[12.9343,77.60...	Yes
Row65	0.462	1	0.5	0	0	0.25	1	0.437	0	[13.0012,77.59...	Yes
Row66	0.385	1	0.5	0.333	0.417	0	0.6	0.282	0	[12.9442,77.60...	Yes
Row67	0.308	0	0.5	0	0	0.25	0.6	0.223	0	[13.0487,77.59...	Yes
Row72	0.462	0	0.5	0	0	0.25	0.4	0	0	[12.9766,77.59...	Yes
Row74	0.538	1	0.5	0	0	0.25	0.6	0.019	0	[13.0019,77.57...	Yes

9. Объединяем с помощью Column Appender тестовые данные и предсказанные данные в одну таблицу

Appended table - 8:26 - Column Appender

File Edit Hilite Navigation View

Table "default" - Rows: 117 Spec - Columns: 12 Properties Flow Variables

Row ID	Age	Gender	Marital ...	Occupa...	Monthl...	Educati...	Family ...	Pin code	Feedback	geo_point	Predicti...	Output
Row2	0.231	1	0.5	0	0.167	0.25	0.4	0.155	?	[12.9551,77.65...	Yes	Yes
Row8	0.308	0	0.5	0	0	0.25	0.2	0	0	[12.9766,77.59...	Yes	Yes
Row10	0.231	0	0.5	0	0	0.25	0.8	0.087	0	[12.985,77.5533]	Yes	Yes
Row11	0.308	1	0.5	0	0.167	0.25	0.2	0.078	?	[12.977,77.5773]	Yes	Yes
Row13	0.154	1	0.5	0	0	0	0.6	0.078	0	[12.977,77.5773]	Yes	Yes
Row17	0.308	0	0.5	0	0	0	0.4	0.068	?	[12.982,77.6256]	Yes	Yes
Row19	0.154	0	0.5	0	0.167	0.25	0	0.359	0	[12.9783,77.64...	Yes	Yes
Row20	0.385	1	0.5	0	0	0.25	0.4	0.078	0	[12.977,77.5773]	Yes	Yes
Row21	0.231	1	0.5	0	0	0.25	0.6	0.301	0	[13.0298,77.60...	Yes	Yes
Row25	0.462	1	0.5	0	0	0.25	0.4	0.068	0	[12.982,77.6256]	Yes	Yes
Row39	0.308	0	0.5	0	0	0.25	0.6	0.748	0	[12.8988,77.57...	Yes	Yes
Row45	0.615	0	1	0.667	0.417	0.25	0.2	0.32	0	[12.9261,77.62...	Yes	Yes
Row48	0.308	0	0.5	0	0	0	0.4	0.068	?	[12.982,77.6256]	Yes	Yes
Row51	0.308	0	0.5	0	0	0.25	0.2	0.078	0	[12.977,77.5773]	Yes	Yes
Row52	0.385	1	0.5	0	0	0.25	0.4	0.078	0	[12.977,77.5773]	Yes	Yes
Row53	0.462	1	0.5	0	0	0.25	0.2	0.01	0	[12.9635,77.58...	Yes	Yes
Row54	0.231	1	0.5	0	0	0.25	0.4	0.816	0	[12.9306,77.54...	Yes	Yes
Row56	0.231	0	0.5	0	0	0.25	0	0.728	0	[12.8845,77.60...	Yes	Yes
Row59	0.462	1	0.5	0	0	0.25	0.6	0.825	0	[13.0067,77.545]	Yes	Yes
Row61	0.231	1	0.5	0	0	0.25	0.4	0.728	0	[12.8845,77.60...	Yes	Yes
Row63	0.308	1	0.5	0	0	0	0.6	0.272	0	[12.9343,77.60...	Yes	Yes
Row65	0.462	1	0.5	0	0	0.25	1	0.437	0	[13.0012,77.59...	Yes	Yes
Row66	0.385	1	0.5	0.333	0.417	0	0.6	0.282	0	[12.9442,77.60...	Yes	Yes
Row67	0.308	0	0.5	0	0	0.25	0.6	0.223	0	[13.0487,77.59...	Yes	Yes
Row72	0.462	0	0.5	0	0	0.25	0.4	0	0	[12.9766,77.59...	Yes	Yes

10. И с помощью Scorer в финале получаем точность предсказания

Confusion matrix

Confusion matrix - 8:25 - Scorer

File Edit Hilite Navigation View

Table "spec_name" - Rows: 2 Spec - Columns: 2 Properties Flow Variables

Row ID	Yes	No
Yes	88	25
No	3	1

Accuracy statistics

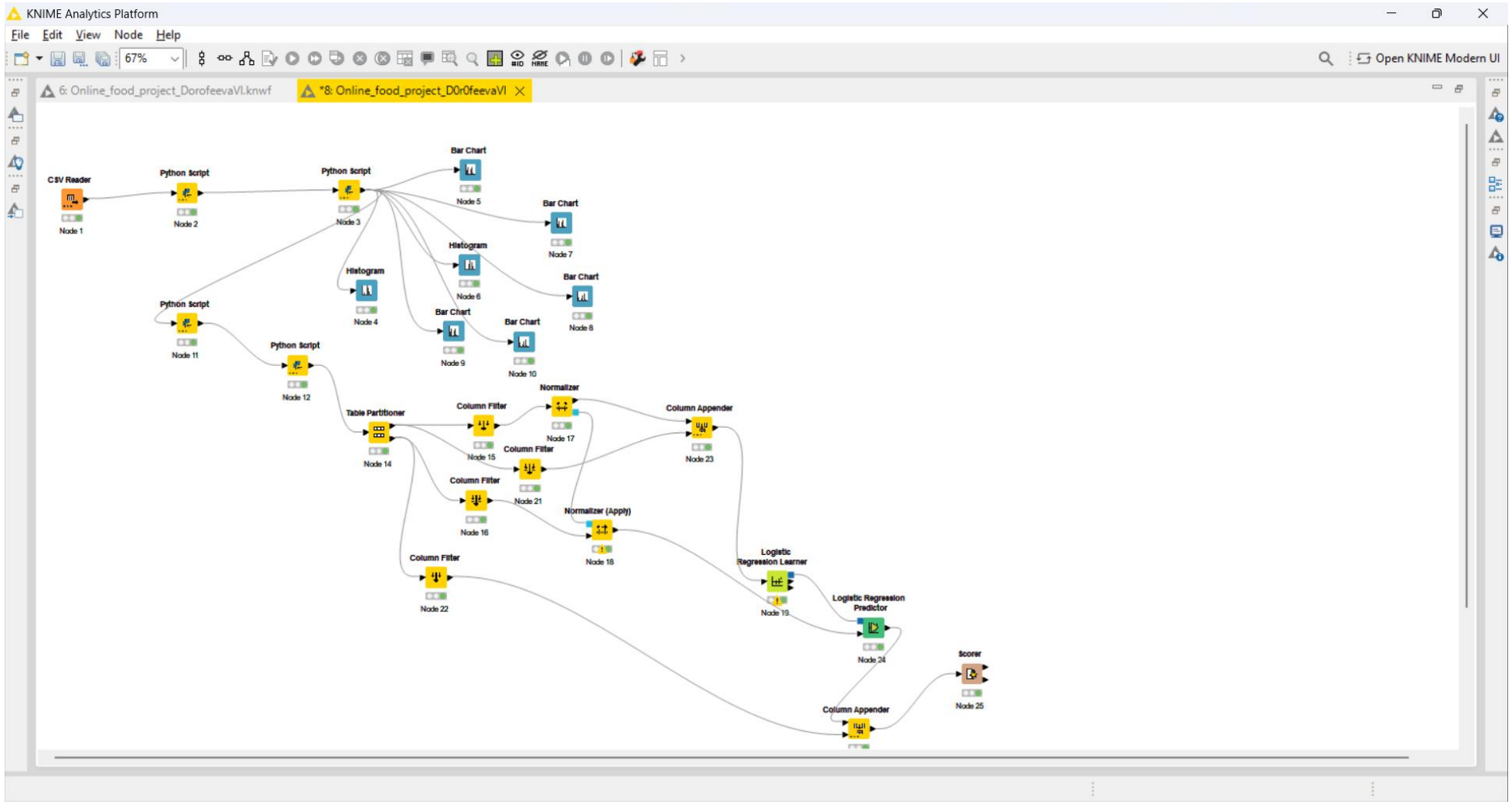
Accuracy statistics - 8:25 - Scorer

File Edit Hilite Navigation View

Table "default" - Rows: 3 Spec - Columns: 11 Properties Flow Variables

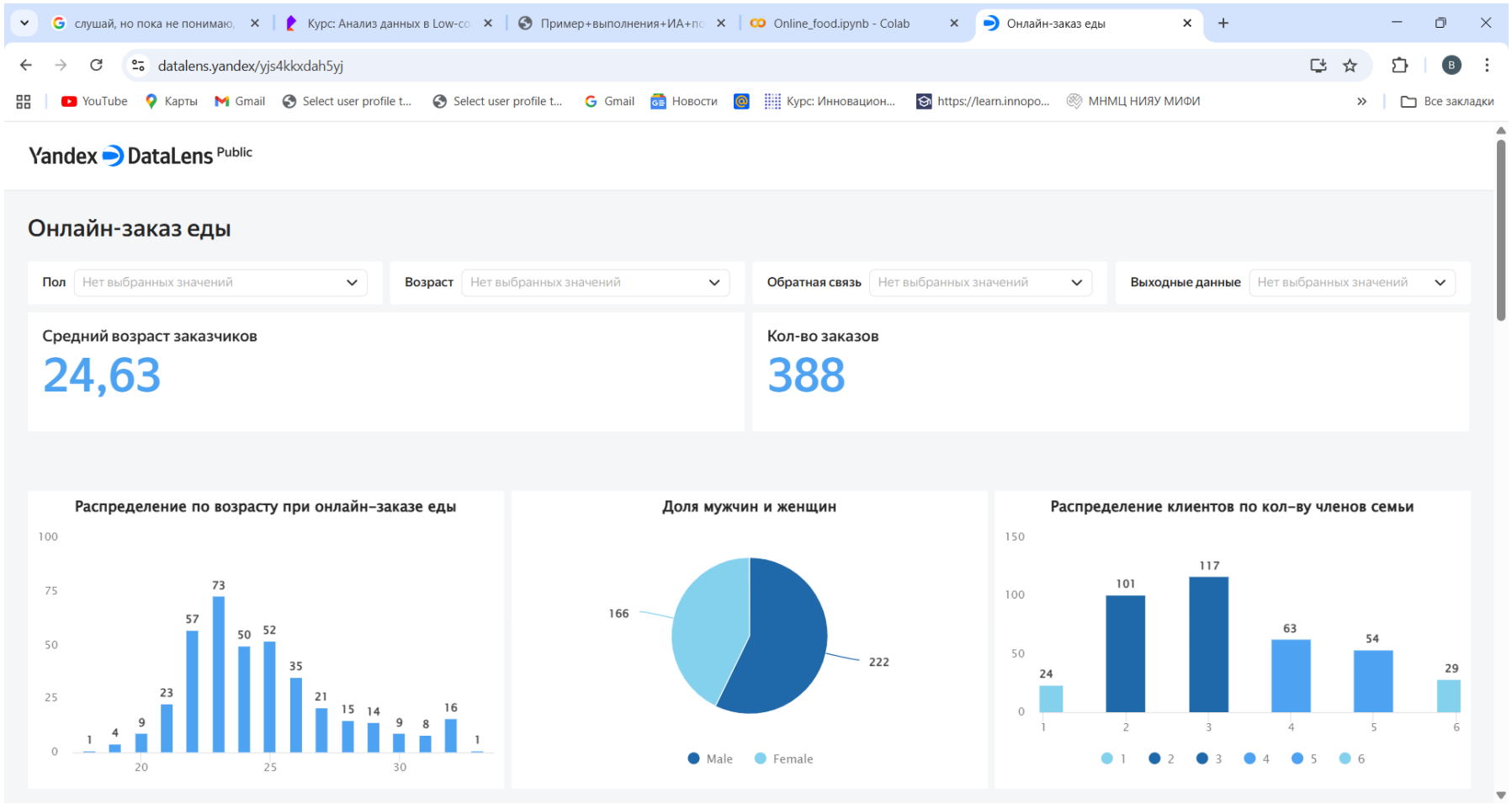
Row ID	TruePo...	FalsePo...	TrueNe...	FalseN...	Recall	Precision	Sensitivity	Specificity	F-meas...	Accuracy	Cohen'...
Yes	88	3	1	25	0.779	0.967	0.779	0.25	0.863	?	?
No	1	25	88	3	0.25	0.038	0.25	0.779	0.067	?	?
Overall	?	?	?	?	?	?	?	?	?	0.761	0.008

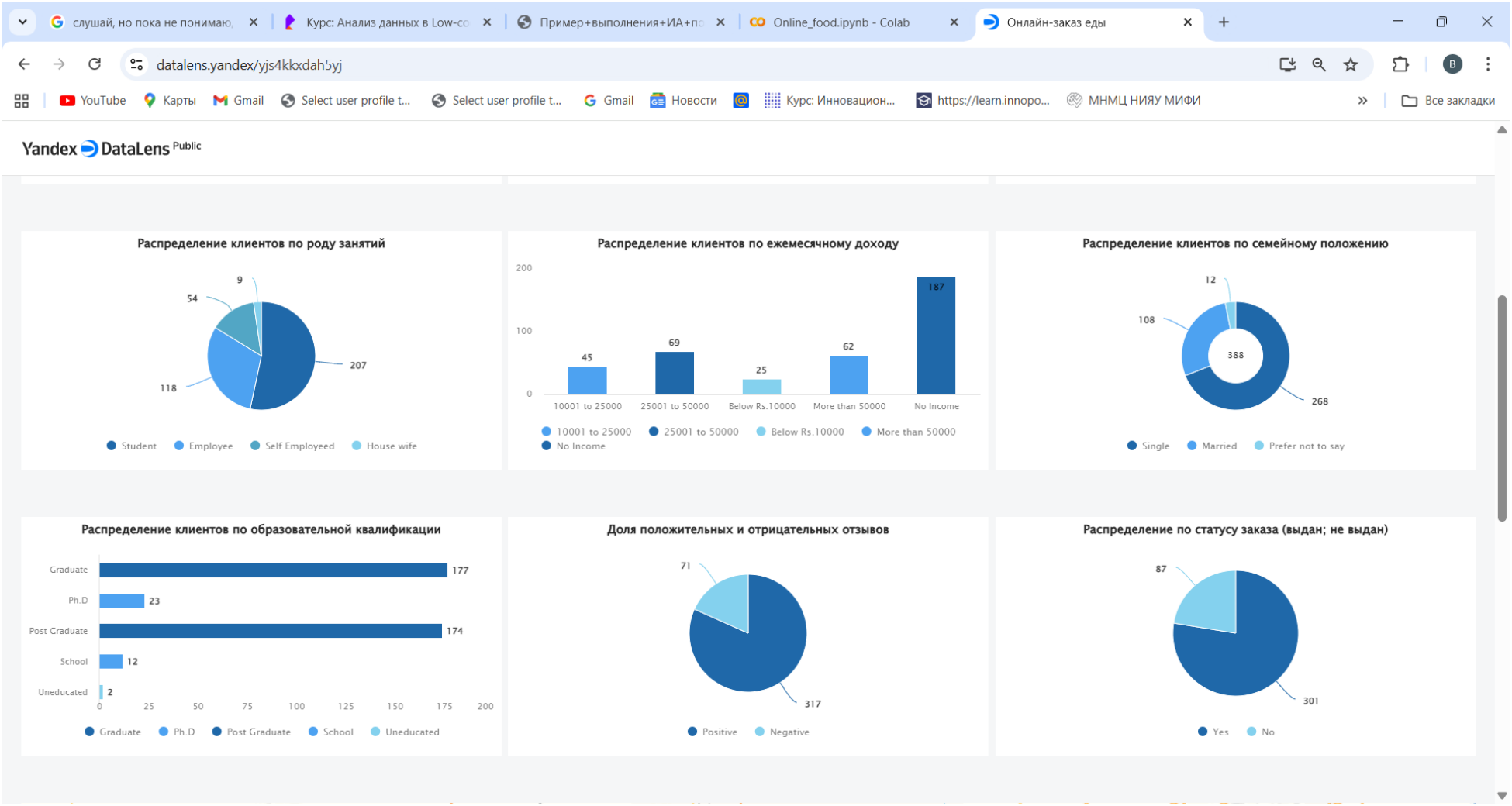
Имеем общую схему исследования:



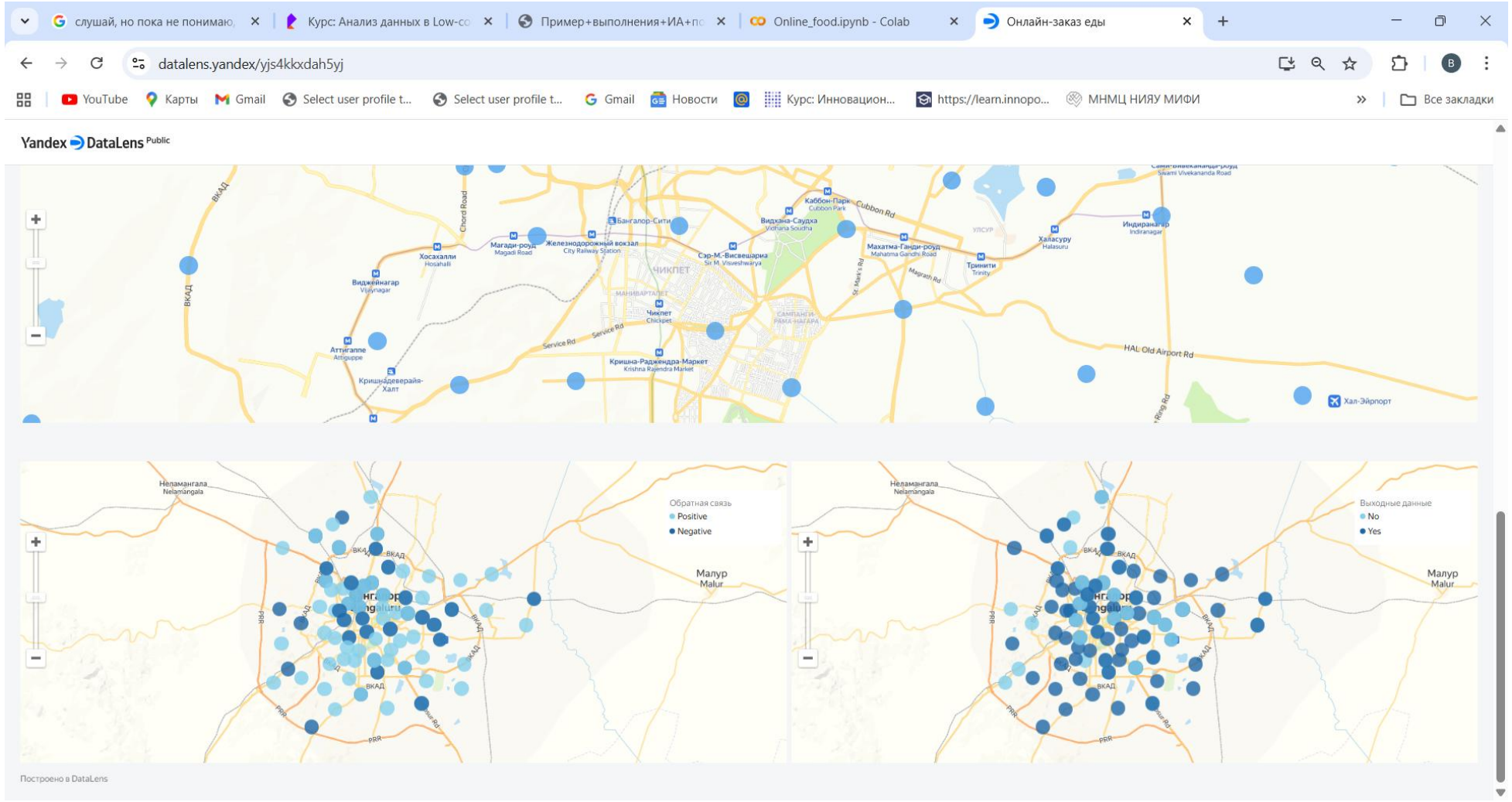
Платформы BI (обязательная часть)
Визуализация в ЯндексDataLens

Проведя исследовательский анализ данных, можно визуально представить его выводы с применением **ЯндексDataLens**
Построены карточки , отражающие исследовательский анализ:




слушай, но пока не понимаю,
Курс: Анализ данных в Low-co
Пример+выполнения+ИА+по
Online_food.ipynb - Colab
Онлайн-заказ еды
datalens.yandex/yjs4kkxdah5yj
YouTube
Карты
Gmail
Select user profile t...
Select user profile t...
Gmail
Новости
Курс: Инновацион...
https://learn.innopo...
МНМЦ НИЯУ МИФИ
Все закладки
Yandex DataLens Public
Распределение клиентов по роду занятий
9
54
207
118
Student
Employee
Self Employeed
House wife
Распределение клиентов по ежемесячному доходу
200
100
0
45
69
25
62
187
10001 to 25000
25001 to 50000
Below Rs.10000
More than 50000
No Income
Распределение клиентов по семейному положению
12
108
388
268
Single
Married
Prefer not to say
Распределение клиентов по образовательной квалификации
Graduate
177
Ph.D
23
Post Graduate
174
School
12
Uneducated
2
0
25
50
75
100
125
150
175
200
Доля положительных и отрицательных отзывов
71
317
Positive
Negative
Распределение по статусу заказа (выдан; не выдан)
87
301
Yes
No

А так же выполнена визуализация с помощью геоданных



Заключение

Большинство людей, заказывающих еду, в возрасте от 23 до 26 лет, что говорит о том, что среди клиентов больше студентов, недавних выпускников или начинающих специалистов из-за ограниченного времени или доступа к кухне. Размер семьи обычно невелик - от 2 до 3 человек, так как заказывающие в основном молодые люди.

Значения широты и долготы тесно сгруппированы в районе 12,97 северной широты или 77,60 восточной долготы, при этом минимальная разница указывает на заказы из определенного городского региона. Пин-коды принадлежат Индии, поскольку они состоят из 6 цифр.

Мужчины заказали больше, чем женщины, что предсказуемо, в целом. Одинокие клиенты, как правило, заказывают больше, чем женатые.

Что касается уровня образования, то большинство клиентов получили степень бакалавра или магистра. Это говорит о том, что сервис привлекает лиц с более высоким уровнем образования.

Домохозяйки обычно не заказывают еду. В основном студенты и сотрудники, живущие вдали от дома, чаще покупают еду онлайн. Работающие люди также заказывают еду в небольших количествах. Это говорит о том, что они предпочитают заказывать еду, а не готовить, чтобы сэкономить время.

Люди, у которых нет дохода или они не имеют стабильного дохода, более склонны покупать продукты питания онлайн.

На основе платформы Knime было выполненно прогнозирование данных онлайн продаж (решение задачи классификации) и резульаты показали доввольно высокую точность (0,76).