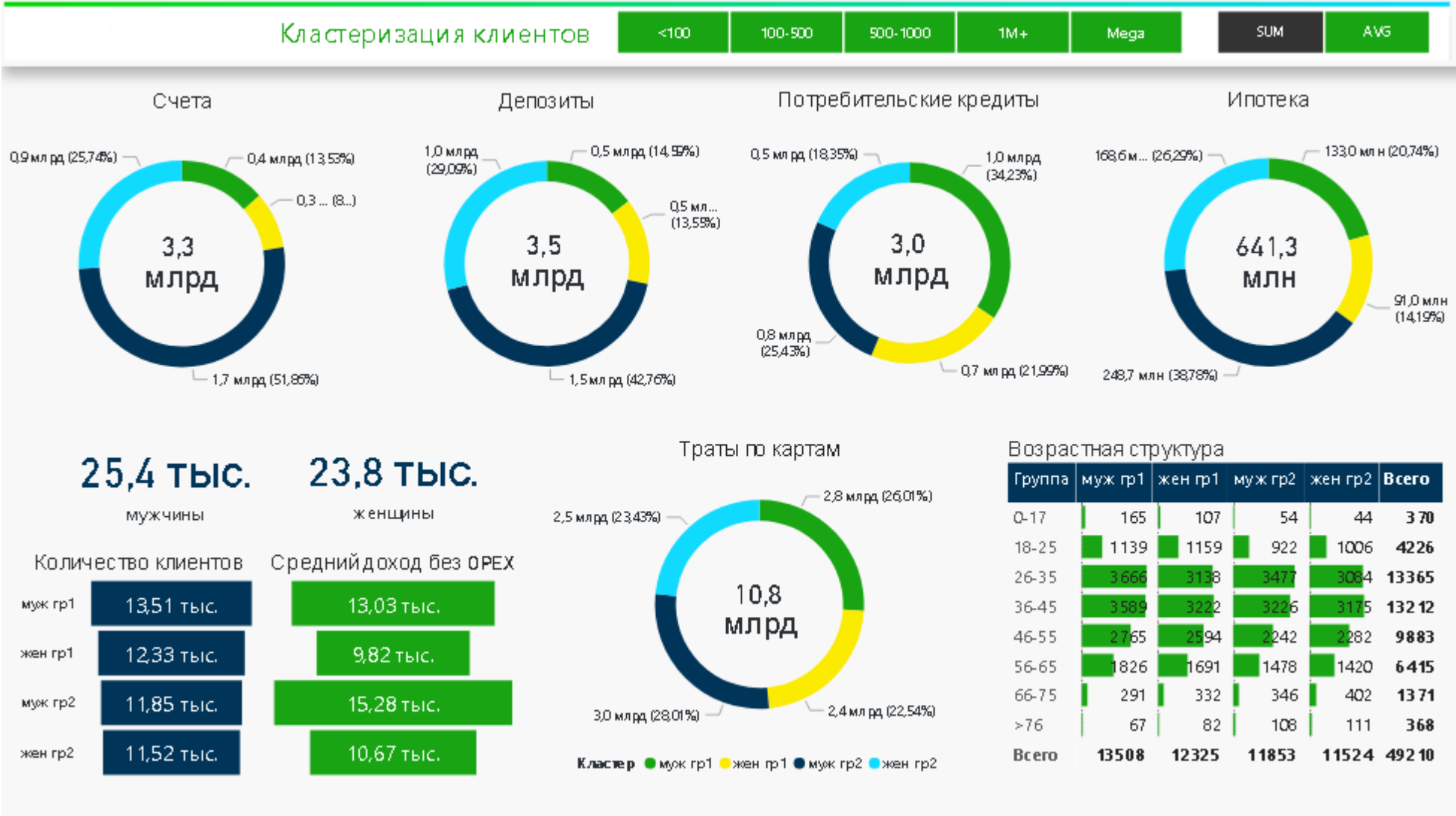
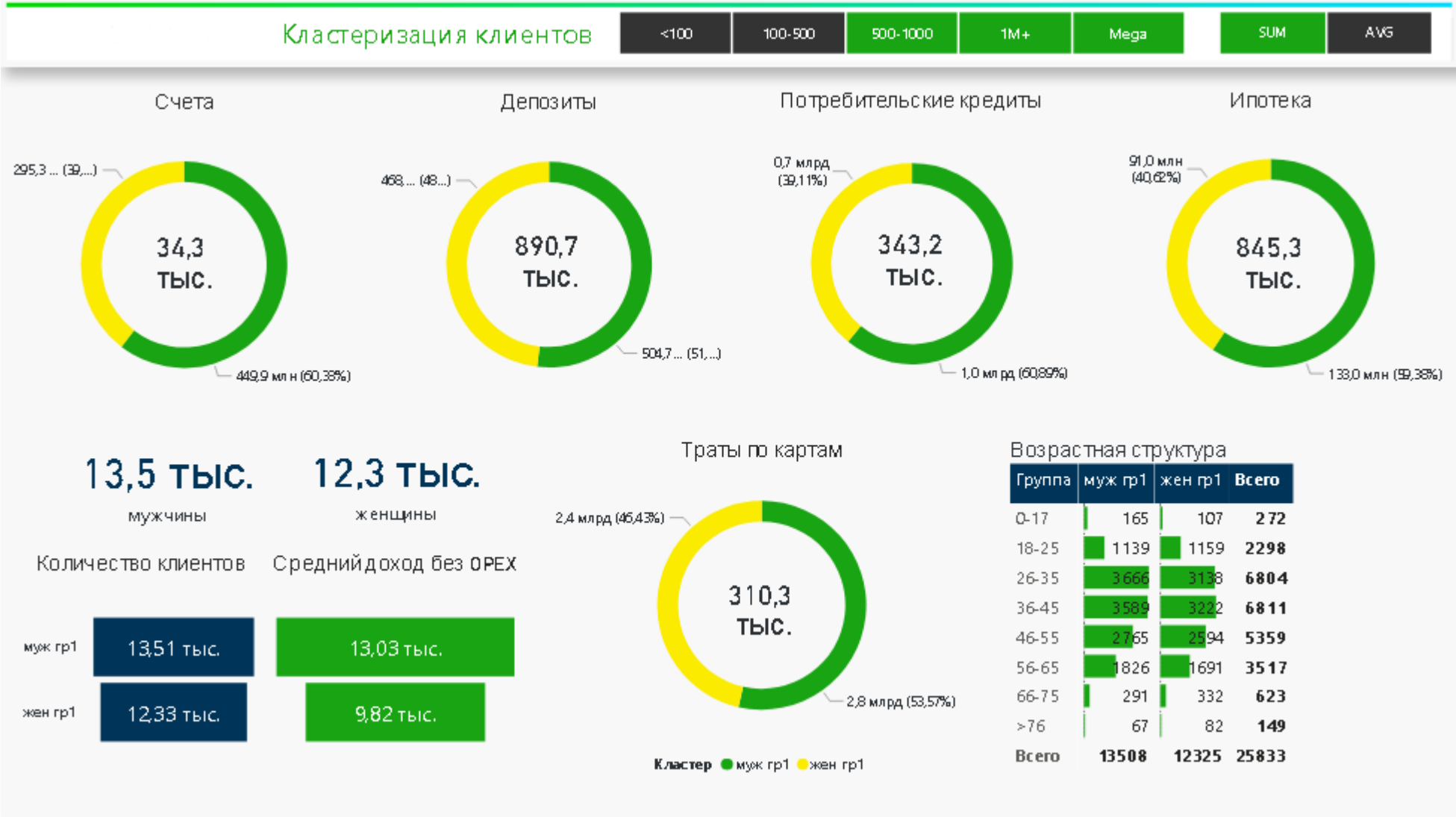


Для проведения аналитического исследования были предоставлены данные в виде датасета формата csv. Исходная информация обрабатывалась посредством JupyterLab. Из исходников удалены записи, для которых не был указан пол клиента. Также удалены строки с невалидными клиентами (бизнес-активность по всем целевым полям равно 0). Для кластеризации было решено применить алгоритм машинного обучения без учителя – kMeans. Перед началом обучения все текстовые поля были приведены к числовым эквивалентам. Далее числовые колонки подверглись нормализации. Расчетным путем установлено, что для метода kMeans оптимальное количество кластеров – 4. Указанная модель была обучена, метки кластеров записаны в отдельный столбец. Финальный датафрейм сохранен в датасет формата csv для последующего анализа цифр на платформе BI.



В качестве платформы BI был выбран Power BI Desktop. Автоматически все клиенты банка были разделены на четыре кластера: мужчины и женщины в городах “<100” и “100-500” (группа 1) и мужчины и женщины в городах “500-1000”, “1М+” и “Мega” (группа 2).



Расчет средних величин по анализируемым категориям проведен с условием: в подсчете участвуют только клиенты с ненулевыми записями по указанным полям. Иначе получается как в анекдоте: “Сосед ест мясо, а я капусту, в среднем мы едим голубцы”. Конец ознакомительного фрагмента)))

