**Документация по системе**

На этом чемпионате мне предстояло разработать модель машинного обучения, предлагающую лучшую следующую покупку клиенту банка, для увеличения доли покупок с данной карты в общей структуре трат клиента и повышения лояльности.

Предложение должно удовлетворять следующим основным критериям:

- предложение должно быть интересно и релевантно клиенту (если он ходит в самые бюджетные продуктовые магазины – то получает предложения на самую бюджетную аптеку, а не дорогую; если пользуется онлайн магазинами еды – то предложение на онлайн магазин одежды и пр.)

- предложение обязательно должно расширять интенсивность использования карты клиентом (увеличивать среднее число и объем транзакций в месяц). То есть клиент через предложение должен научиться либо платить в новых товарных категориях, либо платить чаще, либо тратить больше.

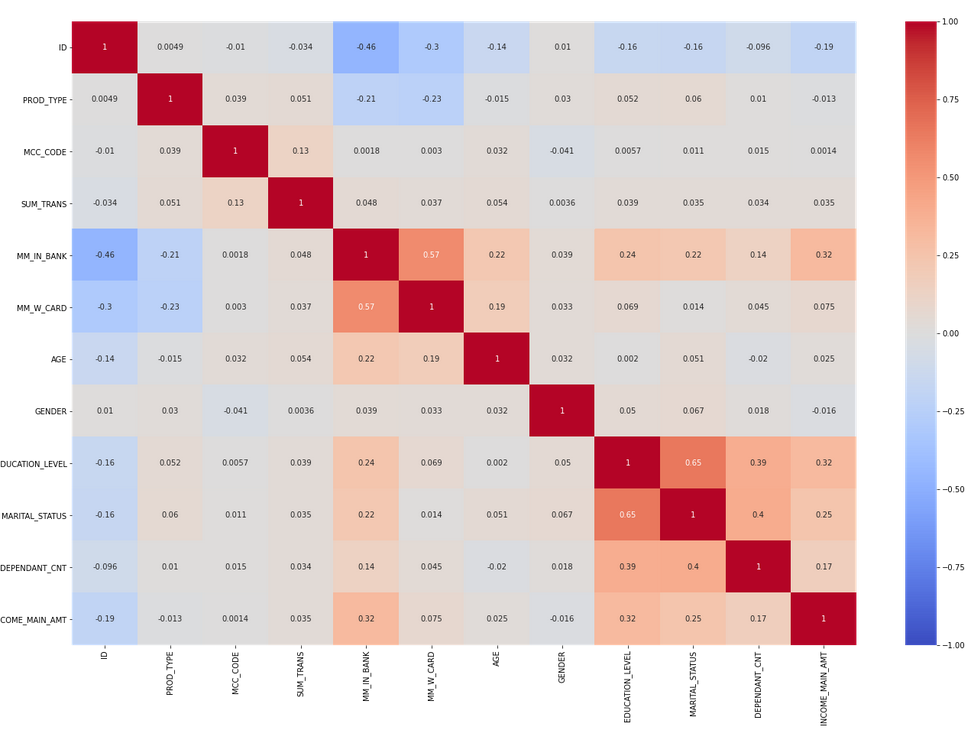
**Сессия 1:**

**Основная задача** первой сессии заключалась в парсинге, обработке данных и выделение значимых атрибутов.

**Решения:**

Для работы с данными, а так же их визуализаций использовались библиотеки pandas и matplotlib (иногда он дополнялся библиотекой seaborn)

Сам парсинг осуществлялся при помощи метода read\_csv и объединялся при помощи метода merge.

После парсинга и обработки данных, надо было обозначить коррелирующие атрибуты, это было реализованно при помощи тепловой карты 

**В итоге** 1 сессии получился следующий дата сет

Изображение выглядит как текст

Автоматически созданное описание

Он был сохранен для последующей работы.

**Сессия 2:**

**Основная задача** во 2 сессии это кластеризироватьи визуализировать зависимости данных.

**Решения:**

Данные были обработаны посредством скалирования и метода главных компонент, после чего передавались в алгоритм кластеризации.

Изображение выглядит как текст

Автоматически созданное описание

Изображение выглядит как текст

Автоматически созданное описание

Алгоритм кластеризации был выбран KMeans, так как он оказался лучше остальных по результатам метрик

Результаты его метрик следующие:

calinski\_harabasz\_score - 252590.28939990335

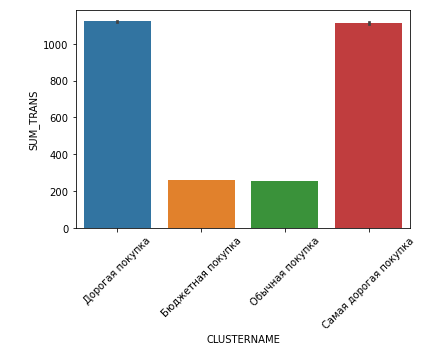
davies\_bouldin\_score - 0.7439965044874522

Изображение выглядит как текст

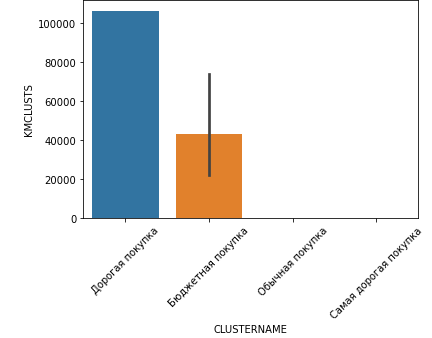
Автоматически созданное описание

Наименование кластеров происходило в зависимости от визуализации

Ниже представлена визуализация зависимости суммы транзакции к кластеру



Ниже представлена визуализация зависимости популярности кластера к самому кластеру



**В итоге** был сохранен дата сет с новым атрибутом, содержащим метки (кластеры).

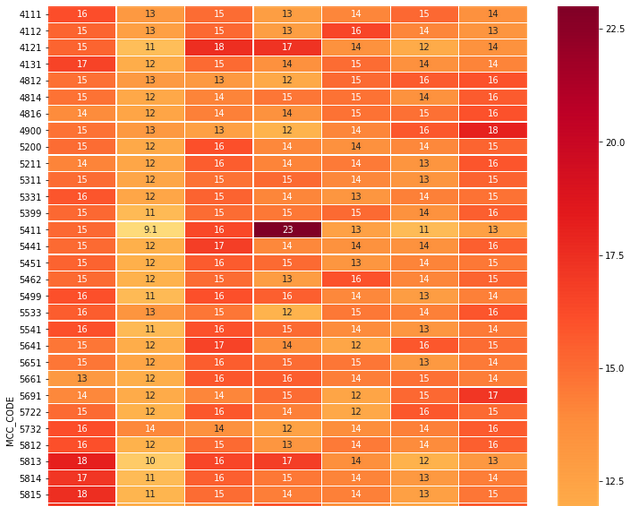
**Сессия 3:**

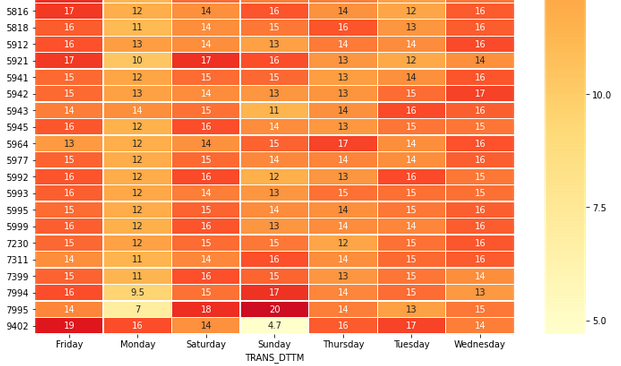
**Основная задача** в 3 сессии — это визуализация определенных атрибутов (категории, суммы покупок и предпочтения пользователей по покупкам)и классификация данных.

**Решения:**

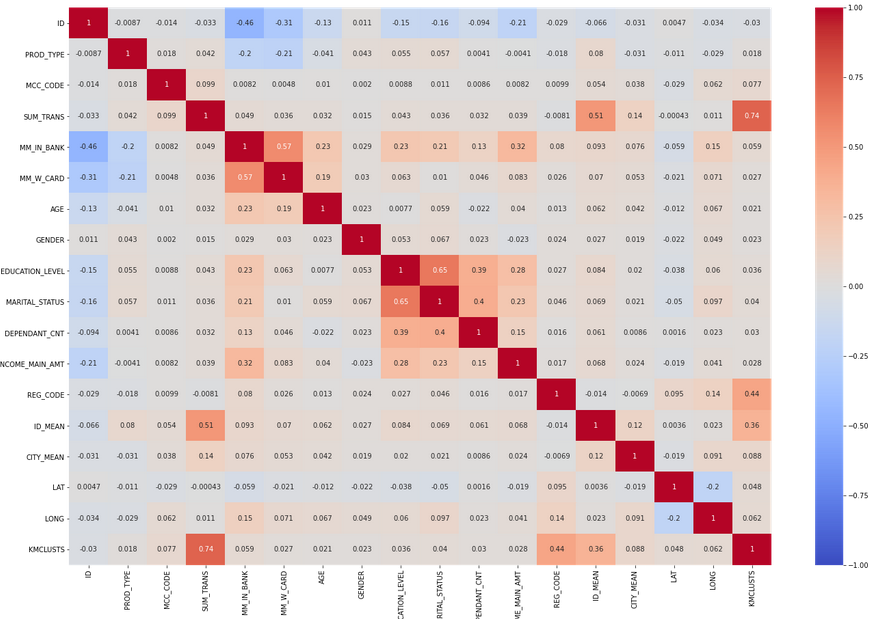
Для визуализации я использовал heatmap, pivot\_tabel, catplot и regplot, так как они наглядно отображают нужный мне результат, а именно влияние атрибутов на категории, суммы покупок и предпочтения пользователей по покупкам

**Pivot\_tabel**



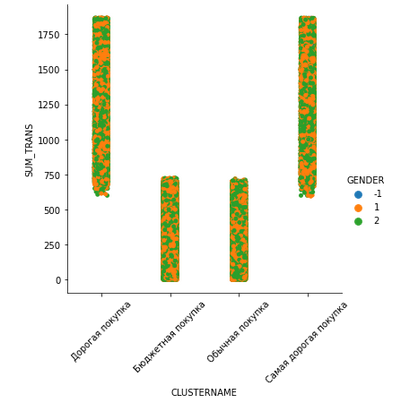
 По сводной таблице выше можно увидеть, что самые нагруженные дни в общем являются пятница и воскресенье, а самой нагруженной категорией продавца является 5411

#### Heatmap



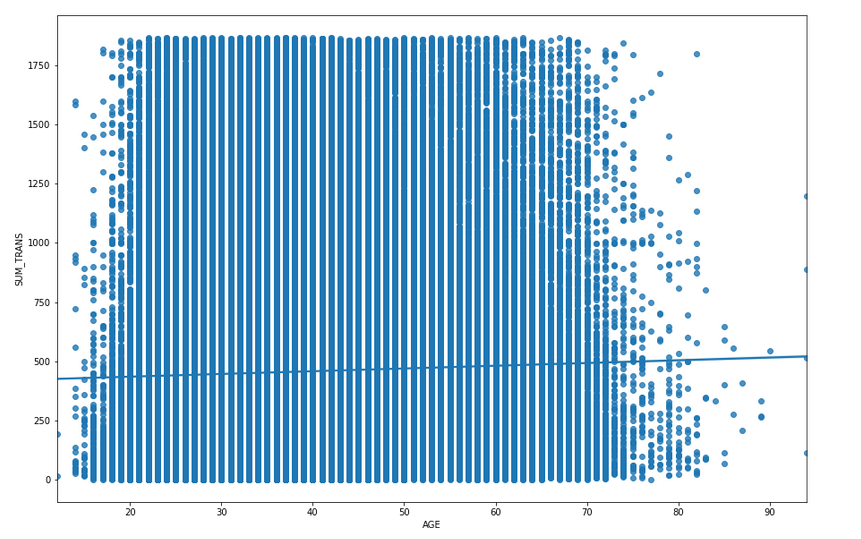
Выше видно, что на предпочтения пользователей по покупкам оказывает сумму транзакции, что означает, то что люди всегда обращают внимание на ценник прежде чем, что то преобрести, а так же не мало важным атрибутом является код региона, а это значит, что люди имеют разный доход в зависимости от региона

#### Catplot



Выше можно заметить, что какого-то сильного отрыва в предпочтениях и стоимости покупки в зависимости от гендера нет, но все же мужчины покупают больше, особенно в дорогом сегменте

#### Regplot



На графике выше видно, что самые дорогие транзакции совершаются в возрасте от 30 до 70 лет

Далее следует классификация, тут по результатам метрик (я рассматривал метрики - f1 и accuracy sore, f так же f1 по macro avg), лучшим был KNN обе из выбранных метрик показывают 0.99

Изображение выглядит как стол

Автоматически созданное описание

**В итоге** было сохранено 2 модели, это модель скалирования данных и модель классификации KNN

**Сессия 3:**

**Основная задача** в 4 сессии — это разработка API и разработка приложения (рекомендательной системы).

**Решения:**

Для решения этой задачи было разработано API в виде класса, такое решение выбрано, так как это простая реализация и легкий способ дальнейшего внедрения в любые программные продукты в виде подключения простого модуля.

Изображение выглядит как текст

Автоматически созданное описание

### RecomendationSystemV1

Был разработан программный продукт на основе API, реализация была исполнена в виде консольного приложения, где консольный ввод служил интерфейсом для взаимодействия с пользователем. Так же и бот и API получили программную документацию и руководство пользователя.

Изображение выглядит как текст

Автоматически созданное описание

**Заключение:**

В заключение могу сказать, что реализация программного продукта прошла успешно, и приложение получилось вполне дееспособным,

API можно в дальнейшем доработать и построить более масштабную рекомендательную систему.