**(Настя) Введение в предметную область**. С развитием цифровых технологий конкуренция сместилась от создания лучшего продукта/услуги к созданию лучшего клиентского опыта. На представленной диаграмме красными маркерами отмечены банки-лидеры по индексу потребительской лояльности; серыми маркерами отмечены отстающие по этому показателю. На графике мы можем увидеть, насколько изменился чистый процентный доход в зависимости от страны за три года. Обратим внимание на то, что у лидеров чистый процентный доход увеличился; у отстающих банков же он либо остался на прежнем уровне, либо уменьшился. Исходя из вышесказанного, клиентский опыт очень важен, и поэтому наш проект нацелен на его анализ в разных банках.

**(Артем) Актуальность и ЦА.** Причиной актуальности нашего продукта является отсутствие в открытом доступе программ/сайтов, занимающихся сравнением клиентского опыта в различных банках.

**Постановка задачи.** Нашей задачей является создание модели оценки качества банковских каналов, продуктов и услуг с визуализацией результата, позволяющей сравнить любой банк с рынком в среднем и с лучшими игроками рынка, определить области улучшения и дальнейшего развития.

**(Настя) Категории.** Мы проводим анализ по следующим категориям:

1. Удобство офиса
2. Банкоматы
3. Уровень сервиса
4. Персонал
5. Продукты и услуги
6. Дистанционные каналы обслуживания

**Про аналоги.** На данный момент существуют программы, являющиеся аналогами нашему сайту, однако все они являются частями закрытых систем и недоступны простым пользователям. Именно поэтому мы не знаем, как они выглядят и какими качествами обладают аналоги. В то время как наш проект открыт и общедоступен.

**(Артем) Ход работы**

Нашу работу мы начали со скрепинга адресов банков с сайта 1000 банков.ru.

С помощью этих адресов мы собрали отзывы с яндекс карт, а с гугл карт у нас к сожалению не получилось, так как API не выдает более пяти релевантных отзывов даже с Premium-аккаунтом. Также нам не удалось произвести скрэппинг отзывов с сайта, потому что каждый раз генерируется новый номер сессии, который невозможно предсказать.

Несмотря на проблему с гугл, у нас получилось собрать около 500 тысяч отзывов.

Для автоматического выделения категорий, о которых пишут пользователи, консультант посоветовал нам воспользоваться программой KNIME Analytics Platform. К сожалению, это не увенчалось успехом, потому что она не может выделить достаточное количество различных категорий (например, несколько раз повторялась категория, похожая на «Уровень сервиса»). Стоит отметить, что иногда получалось так, что один выделившийся топик можно разделить даже на несколько категорий.

**(Настя)**

Поэтому мы решили произвести ручную разметку случайной выборки отзывов(1600) на основе категоризации предложенной заказчиком.

Далее мы воспользовались этими данными для обучения моделей классификации с помощью библиотеки scikit learn на сервисе google colab. Используя эти модели мы пролэйблили все оставшиеся отзывы в базе данных. И для визуализации мы воспользовались связкой из трёх библиотек.