Тестовое задание на позицию аналитик-разработчик в домен Supply Chain Management / Assortment

### Контекст по компании Autodoc, департамент Private Brands, отдел Analytics, команда Data Analytics

Компания Автодок - это площадка e-com (маркетплейс) по продаже автозапчастей в Европе. По ссылке на сайте подрбная информация <https://autodoc.group/en/about-us>

Как у любой e-com компании,перед нами стоят задачи генерации ассортимента и доступности товаров для продажи клиентам.

Pirvate Brands занимается созданием собственной торговой марки в каналах продаж Autodoc, поэтому нам требуется проводить сплошной анализ ассортимента рынка автозапчастей, чтобы быть впереди конкурентов.

Перед нами стоят задачи:

1. Какой ассортимент генерировать ?
2. Как сгруппировать ассортимент по взаимозаменяемости и схожести - чтобы правильно себя позиционировать?
3. Как оценить потенциал спроса товара?
4. Как управлять ассортиментом - ввод и вывод товара от чего он зависит?
5. Какой спрос мы покрываем своим предложением ?
6. Как понять где мы находимся по отношению к нашим конкурентам альтернативным брендам и товарам ?
7. Какие существуют сегменты пользовательского спроса в ассортименте (разные характеристики товара, например чугунные диски или хромированные, ценовой сегмент, пользовательский например разделение по авто),
8. Как оптимизировать ассортимент Private Brands для максимизации выручки при высокой оборачиваемости запасов?
9. Как обеспечить спрос запасами на складах в Европе при ограничениям бюджета, складской площади, спроса и т.д.?
10. Какой товар заказать, ассортимент широкий - есть дублирование ассортимента и разные бренды которые друг друга дублируют ?
11. У кого заказать товар - поставщиков много мы заказываем по всей Азии/Средней Азии/Европе/ ?
12. По какой цене заказать товар, дублируется прайсы - есть разницы от альтернатив на рынке , мы не можем закупать дороже цены продажи альтернатив -иначе какой смысл ?
13. Сколько заказать товара для покрытия спроса клиентами ?
14. Как учесть риски времени производства и доставки товара ?
15. Какая частота заказов у поставщиков и какую нам держать частоту, отчего она зависит ?
16. Как оптимизировать заказ moq - минимальное кол-во заказа товара в партии, mov - минимальный объем заказа партии, lot - кратность товара в партии?
17. Как максимизировать оборачиваемость заказа при ограничении бюджета заказа например у вас есть 1 млн долларов что вы закажете для максимизации метрик- ведь деньги не бесконечны ?
18. Как учесть риски в заказе ? Как рассчитать страховой запас ?
19. Мы влезем в складские площади ?
20. Мы сможем доставить весь заказ или нужно на партии разделять ? Как это сделать выгодно ?
21. От чего зависит наш запас на складе ?
22. Какой оптимальный запас на складах нужно держать ? На какое время ? от чего зависит?
23. Как учесть запасы на b2b/b2c клиентов ? b2b покупают много, но непредсказуемо, b2c мало но стабильно ?
24. Как нам мониторить запасы,быстро распродается или товар не продается и лежит камнем, как мы можем контролировать оборачиваемость запасов?

Это лишь некоторые вводные вопросы, которые решает команда Аналитики в Autodoc Private Brands

Также в нашем отделе существуют другие команды, мы работаем вместе над задачи нашего e-com бизнеса:

* DMT - управление данными, основная задача поставка данных и инфраструктура (GitLab|GCP|DWH Clickhouse|Airflow|Metabase|JupyterLab|venv и окружение) - они тоже ищут ребят, попозже скину описание, можете мне написать.
* BI & Reporting - визуализация данных, разработка отчетности и ad-hoc задачи (PowerBI|Metabase|Superset|MSSQL|Slack-оповещения+боты отчетности|даги пайплайнов отчетности)
* Data Analytics - исследование данных для извлечения ценности для бизнеса ($) (EDA|ML|lgbm|pytorch|nlp|aiohttp|fastapi|validation|a/b testing|)

Мы работаем в партнерстве с нашими бизнес-заказчиками и решаем задачи и вызовы вместе, работаем в таск трекер jira, документацию ведем в confluence, код пишем на Python, храним код в gitlab.

Пропишу стек подробно

В нашей Data Analytics команде 12 человек и два доменных направления:

* Управление запасами,цепочками поставок и закупок (SCM): цель - обеспечить склады запасами, повышение уровня сервиса и доступности товара для клиентов, оборачиваемости. Продукты и пайплайны в домене:
* Автозаказ - генерация заказов поставщикам при разных ограничениях условий поставки, бюджета, места на складе.
* Прогноз спроса - ассортимент более 1 млн товаров, сейчас катается lgbm+средние+сезонность+тренды товары с историей и также оценка спроса на новые товары без истории (оцениваем по размеченным альтернативам от команды аналитики ассортимента)
* Прогноз запасов - моделирование поступления запасов на год вперед для оценки потребности площадей, и на краткосрочные периоды день/неделя/месяц, есть базлайн , сейчас пишем lgbm под задачу
* Мониторинг вымывания ассортимента и уровня запасов
* Прогноз времени (LeadTime) от выполнения заказа до поставки товара на склад
* Управление ассортиментом (Assortment): цель - оптимизация ассортиментной матрицы, повышение выручки, покрытие ассортиментом автозапчастей парка автомобилей на рынке Европы. Продукты и пайплайны в домене:
* Рекомендация товарного ассортимента для ввода в матрицу ассортимента товаров на основании анализа внешних и внутренних данных рынка e-com и текущей матрицы.
* Кластеризация и матчинг ассортимента >9 млн.товаров (+иногда парсинг) - группировка ассортимента по характеристикам и применимости запчастей, генерация групп субститутов (альтернативных товаров)
* Ценообразование и сегментация - алгоритмы калькуляции цен,A/B тесты, группировка ассортимента по ценовой сегментации
* Оценка потенциала рынка автозапчастей - оценка ввода нового ассортимента по отношению к карпарку рынка

Многие вопросы решаются без Машинного обучения и нейронных сетей, но существуют хайлоад проекты, где требуется знание МЛ и AI технологий, это проекты описанные выше. Мы поддерживаем эти сервисы для пользователей, отвечаем за используемые данные, весь препроцессинг, метрики валидации и тестирования, и конечно за алгоритмы на продакшн и улучшение метрик сервиса и самого сервиса. Так как многое катается на бэкенде и хранится в базе на клике, на API мы еще все сервисы не перевели. Также есть огромное раздолье на использование алгоритмов машинного обучения, например сейчас нам требуется создать Рекомендательную систему генерации ассортимента, чтобы ускорить работу с ассортиментом у пользователей / Создать сервис прогнозирования запасов, чтобы снизить убытки от вымывания и перезакупа, когда товар стоит без движения, быстро принимать решение о планировании действий на складе.

### Тестовое задание:

* Срок на решение тестового задания - 1 неделя = 7 дней вкл выходные,
* Определите сроки, когда вы отдадите решение. Желательно написать мне после получения тестового в срок 1-2 дня , когда вы его отдадите. Чтобы вам ответить на этот вопрос, стоит:
  + прочитать тестовое задание
  + провести оценку по времени, оценить свою нагрузку
  + дать ответ

У вас два набора данных по доменам SCM и Ассортимент.

* Представьте, что вы уже работаете у нас и решаете задачи по исследованию данных.
* Вы можете выбрать 1 набор данных (любой) или исследовать оба набора - на ваш выбор.
* Если у вас не хватает оперативной памяти для обработки наборов данных, то вы можете сократить набор на свое усмотрение.
* Также на ваш выбор и ваше усмотрение способы и подходы к исследованию данных - методы, способы и даже вопросы, которые вы хотите решить на усмотрение исследователя, возможно вы больше сосредоточитесь на МЛ модели, чем на разведке данных и вопросах заказчика, возможно на исследовании данных и получения знаний из набора данных.
* Возможно вам будет тяжело провести свободную исследовательскую работу, тогда мы предлагаем вам оценить свои навыки по тестовому для младших аналитиков-разработчиков – лежит в **Тестовое\_junior\_task.7z**
* После оформления исследования, скиньте решение на:
  + почту: [d.asalhanov@autodoc.eu](mailto:d.asalhanov@autodoc.eu)
  + telegram: @cosa31 ваше решение в виде Jupyter Notebook
* Решение будет проверяться в течении 2-3 дней
* Как будет оцениваться решение:
  + Команда будет оценивать решение голосованием
  + Постарайтесь оформлять исследование так будто вы отдаете его бизнес-заказчику, желательно
    - комменты,
    - выводы,
    - ваш ход мыслей почему пришли к этому выводу,
    - почему использовали этот метод,
    - визуализация - лучше один раз показать.
* Решение будет оцениваться, как **отчет исследование бизнес заказчику**
  + решили ли вы вопросы заказчика,
  + какие гипотезы проверили и как,
  + как провели eda и что нашли,
  + как решили задачу прогноза или матчинга товаров, или сегментации, возможно некоторые задачи ML не решить.

Наборы данных

**Supply Chain Management** - **SCM\_data.zip**

Описание набора и полей данных:

* test\_fcast\_qty.csv - Динамика продаж товаров (все артикулы из одной товарной категории)
  + article\_id - id артикула
  + Qty - продано штук
  + Date - дата создания заказа
  + Country - страна
  + FinResult - финансовый результат от продажи (выручка)
* test\_fcast\_stock.csv - Остатки на складе по артикулам
  + article\_id - id артикула
  + Stock - остаток в штуках
  + Date - дата записи остатка

**Assortment** - **assortment\_test.zip**

Описание набора и полей данных:

* assortment\_test.csv - ассортимент двух товарных категорий Brake pad set и Brake disc
  + article\_id: id артикула
  + brand\_number: номер бренда
  + category\_name: имя товарной категории
  + price: цена товара в Евро
  + sales\_qty: продажи за год, в штуках
  + sales\_netto: продажи за год, в ценах себестоимости (в закупочных ценах)
  + sales\_revenue: продажи за год, в ценах продажи
  + sales\_count\_country: кол-во стран в которых продавался товар за год
  + all\_OEM\_number: список OEM номеров, для определения взаимозаменяемости запчастей.

Вам предстоит:

1. Провести EDA набора данных. Найдем полезные для бизнеса знания из набора данных, что мы можем сделать, чтобы лучше понять набор данных и сделать выводы полезные для бизнеса?
   1. Визуализация
   2. Описательная статистика
   3. Корреляции и связи
   4. Двумерный анализ
   5. Многомерный анализ
   6. Аномалии и выбросы
   7. Типы данных
2. Существует ряд вопросов и гипотез, которые вам стоит проверить:
   1. У набора данных есть прогностический потенциал, с помощью методов машинного обучения мы можем спрогнозировать продажи по странам и категориям, или категориям, или товарам. Группировка прогнозирования продаж на ваше усмотрение.
   2. Ассортимент в категории взаимозаменяем по ОЕM номерам, мы можем сгруппировать ассортимент по ОЕM номерам в группы альтернатив для дальнейшей оценки спроса группы и анализа.
   3. Динамика запасов логируется, мы можем определить оптимальный уровень запасов по месяцам в зависимости от сезонности, спрогнозировать запасы по товарам или категориям товаров, или определить интервалы уровня запасов по товарам, товарным категориям в штуках.
   4. Мы можем понять сколько запасов нам необходимо для оптимального покрытия спроса каждый месяц.
   5. Мы хотим понять как товары деляться по ценовому сегменту
3. Вы можете предложить решение своих вопросов и гипотез, не обязательно идти по списку выше) Возможно вы найдете много полезного
4. Подобрать модель машинного обучения:
   1. Прогноз спроса. Горизонт и группировку (по дням, неделям, месяцам, кварталам, годам) прогнозирования выбирает исследователь, в зависимости от данных (могу предложить на 3 месяцев вперед по месяцам, не забудьте провести валидацию модели)
      1. Прогнозирования спроса по товарам - выбрать 100 товаров для прогноза
      2. Прогнозирование спроса по категориям - выбрать 10 странам для прогноза
      3. Валидация, Метрики
   2. Кластеризация ассортимента.
      1. Требуется сгруппировать ассортимент по взаимозаменяемости - у 2 товаров есть схожие ОЕ номера значит товары могут быть альтернативами.
      2. Также вы можете использовать любой другой способ кластеризации ассортимента, цель: собрать в одной группе схожие альтернативные товары (субституты)
      3. Провести сегментацию ассортимента по цене.

### Полезные ссылки:

* <https://www.linkedin.com/pulse/power-exploratory-data-analysis-eda-science-basics-best-soni/>
* <https://otexts.com/fpp3/>
* <https://www.kaggle.com/learn/time-series>
* <https://www.kaggle.com/code/tanulsingh077/twitter-sentiment-extaction-analysis-eda-and-model>
* <https://habr.com/ru/companies/aliexpress_russia/articles/686672/>
* <https://habr.com/ru/companies/ozontech/articles/648231/>