Ориентировочный **стек** технологий для DS: Python, Pandas, Matplotlib, Seaborn, Scipy, Numpy, TSFresh, Statsmodels, Facebook Prophet, Scikit-learn, XGBoost, SHAP, Docker, Flask, Git, GitHub, SQL, и базовые навыки Linux.

**22 сентября – 10 октября**

* **DS**
  + **Изучение задачи и датасетов**
    - Загрузить в Pandas
    - Изучить колонки и строки

23 сент. Вебинар по прогнозированию временных рядов, библиотекам и моделям

* + **Сделать разведочный анализ данных**
    - (День — Товар — Магазин — Продажа в штуках — Продажа в рублях — доля продаж по ПРОМО)
  + Построить графики временных рядов
  + Исследовать пропуски, выбросы и корреляции между товарами, заполнить пропуски, нивелировать выбросы
  + Провести предобработку данных
  + **Исследовать признаки**
    - Посчитать цены
      * Стоимость одного товара (за день)
      * Стоимость группы товаров (за день)

25 сент. Консультация с наставником

* + **Фича-инжиниринг: сгенерировать дополнительные признаки**
    - Лаговые признаки (признаки за предыдущие периоды)
    - Скользящие средние
    - Временные признаки
    - Использовать библиотеку TSFresh (генерирует множество признаков)
* Построить прогноз временных рядов с использованием ARIMA и FaceBook Profet
* Сезонные признаки (покажем откуда грузить данные по погоде) \* — доп. источник
* Исследовать другие таблицы и сформировать фичи на свое усмотрение
* Реализовать кросс-валидацию временных рядов
* **Построить ML-модель(и)**
  + Перебрать регрессией
  + Обучить регрессию на деревьях
  + Обучить бустинги
  + Обучить ARIMA
* Проверить качество модели

29 сент. QnA по докеру

* Сделать Фичи-импотанс и проанализировать качество признаков
* Упаковать модель в докер-контейнер

1 окт. - 3 окт. – проверка работы ревьюерами

3 - 10 окт. Внесение правок, финализировать работу, засинхронизироваться с командой и представить командное решение