В данном тестовом задании, вам необходимо построить прогноз продаж на месяц (март 2020).

У вас имеются исторические данные за 2 полных года, а также 1 квартал 2020 года. В наборе данных представлены продажи по двум брендам в 3х магазинах.

sales\_train – продажи за период с 2018.01.01 по 2019.03.31

* Date – Дата продажи
* Month – Месяц
* Year – Год
* Date\_block\_num – Порядковый номер месяца
* Week – Порядковый номер недели
* Shop\_id – код магазина
* Item\_id – код товара
* Brand\_id – код бренда
* Item\_cnt\_day – кол-во продаж в штуках
* Item\_price – прайсовая цена позиции
* Discamount – скидка
* Promo – метка наличия акции в данный период (1 – да, 0 – нет). Действия со скидкой на конкретные позиции в бренде.
* Size\_disc – размер скидки (данное поле только для Promo = 1)
* Promo\_time – время действия акции в днях
* Spec\_promo – специальное предложение. Отдельный вид активности, обладающий некоторыми отличиями от обычных сетевых промо. Действует как правило – месяц.

Items – в данной таблице представлена информация по брендам. Категории позиций и код категории.

* Item\_id – код товара
* Brand\_id – код бренда
* Item\_category\_id – код категории товара
* Item\_category – категория товара

test\_submission – образец того как должен выглядеть итоговый файл.

Ожидаемое содержание тестового задания кандидата:

* Проведите предобработку данных (поиск аномалий, удаление/срез выбросов и др.). Проведите анализ временных рядов (графическое представление, описательная статистика).
* Подведите итоги по результатам, проведенной очистки, предобработки данных. Сформируйте выводы на основании проведенного анализа.
* Feature engineering
* Какие из алгоритмов подойдут лучше. Какими данными на ваш взгляд необходимо обогатить текущий набор данных, для более точного, эффективного прогнозирования продаж.
* Построение прогноза и его оценка. Пожалуйста, дайте комментарии по выбранной вами метрике для оценки прогноза. MAE, RMSE, др? и почему.