

Обработка данных выполнена в Jupyter notebook, расчет прогноза – MS Excel.

1.1 Основная часть

Загружены необходимые модули, загружены данные из исходного файла. Проанализированы для дальнейшей обработки.

1.2 Проведен анализ данных по товарообороту на пустые значения, всплески. Обработаны данные для приведения к стандартному формату дат. После анализа средних показателей и перцентилей отброшены данные с завышенным чеком как некорректные. 2013 год содержит значительное количество недельных пропусков, что может свидетельствовать о нерегулярности сбора данных.

1.3 Выполнен анализ месячных колебаний товарооборота в целом и по каждому году по отдельности. По каждому году также построены линии тренда. Построен график ускорения роста. На основании указанных материалов сделан вывод, о наличии ежемесячных колебаний товарооборота. Также по годам с 2013 по 2017 видно явное ускорение роста, в 2017 ускорение замедляется, а в 2018 идет на спад. Следовательно, для прогноза необходимо использовать кривую 3-го порядка (график ускорения роста показывает два экстремума). Наложение графиков по каждому году иллюстрирует схожие сезонные колебания. Отношение показателей ускорения роста позволяют выделить коэффициенты для прогноза.

1.4 Выполнен анализ среднего чека. 2013 год характерен наличием значительного количества колебаний с пределом, значительно превышающим средний максимум по всем 6 годам. Что в совокупности с нерегулярностью недельного ТО позволяет сделать вывод о некорректном сборе данных (навалены в кучу по месяцам). Проведенный анализ ежемесячных колебаний среднего чека не выявил существенных особенностей за исключением некоторого его падения в конце декабря. В принципе колебания среднего чека в течении анализируемого периода коррелируются с динамикой курса рубля. После роста среднего чека в 14 году к 15 начинается инфляция, продолжающаяся до начала 16-го, а затем постепенно с выравниванием курса в 16-17 средний чек растет. В 18 начинает постепенно падать вместе с курсом рубля

1.5 Повторные продажи. Выполнена группировка данных и подсчитаны количества заказов для каждого уникального пользователя с разбивкой по месяцам. Далее данные сгруппированы по 3-м группам (с 1 покупкой, с 2 и от 3). По результатам разбивки построены графики роста групп в динамике. Наиболее быстрорастущая группа – с 1-й покупкой. Остальные две группы с небольшим дифференциалом повторяют друг друга. График количества пользователей показывает определенную годовую зависимость. Очевидно, что в декабре ежегодный всплеск активности пользователей.

1.6 Выполняется в MS Excel. Проведено прогнозирование для имеющихся данных на линейной модели и на модели, основанной на полиноме 3 порядка с проверкой на данных 2018 года. Расчет выполнялся дважды, с учетом 2013 года и без него. При расчете с учетом 2013 года отклонение прогноза имеет нестабильное значение по месяцам. При выполнении расчета без учета 2013 года отклонение по месяцам имеет линейный тренд. При выполнении расчета посредством кривой 3-го порядка среднее отклонение прогноза составляет примерно 1,26. Что в свою очередь коррелируется с коэффициентом отношения ускорения роста 2017 -2018 гг., который равен 1,27. На основании анализа данных графика Year commodity circulations growth acceleration можно сделать вывод, что снижение скорости роста при неизменных остальных условиях будет оставаться в течении следующего года постоянным. Следовательно, мы можем применить коэффициент 1,27. Поскольку, судя по статистике курса рубля на 2019 год (плавное снижение с 69 до 61 рубля) можно предугадать небольшой рост среднего чека, то и скорость падения роста может быть стабильна. Таким образом принимаем для прогнозирования модель без учета 2013 года и построенную на кривой 3-го порядка.

1.7 Построен график с прогнозируемыми данными. Для проверки необходимо заполнить поля, выделенные желтым цветом фактическими данными в файле MS Excel.