Вам предложены данные сети кофеен. Целевой переменной выступает средний товарооборот за определенный промежуток времени (переменная «target»). Данная переменная известна для обучающей выборки и неизвестна для тестовой. Для каждой кофейни даны:

1. Идентификатор (id).
2. Расстояние до метро.
3. Набор признаков, характеризующих окружающие кофейню объекты различных категорий в указанном радиусе относительно кофейни: количество отзывов, количество объектов, средний рейтинг объектов. В категорию «Кофейни» в данных признаках рассматриваемая кофейня не включается.
4. Набор признаков, характеризующих интенсивность транспортных и пешеходных потоков в указанном радиусе или пешей доступности относительно кофейни. Интенсивность рассчитана отдельно для пешеходов, автомобилей, суммы первых и вторых; для разного времени суток (утро, день, вечер, весь день); для числа сигналов и числа пользователей (юзеров); для резидентов района, в котором находится рассматриваемая кофейня, и для работников этого района.
5. Набор признаков, характеризующих жилые дома в указанном радиусе или пешей доступности относительно кофейни. Данные признаки рассчитываются отдельно для числа квартир и цены квадратного метра.
6. Набор признаков, характеризующих мобильный трафик в указанном радиусе относительно кофейни. Данные признаки рассчитываются отдельно для групп, разделенных по возрасту и доходу, а также для типов трафика для района, где расположена рассматриваемая кофейня (домашний, рабочий, транзитный, длительный, их сумма).
7. Набор признаков, характеризующих тип кофейни. Данная бинарная переменная определяется специфическим расположением кофейни (в бизнес-центре, в крупном транспортном узле (хаб), в любом здании кроме бизнес-центра и торгового центра (улица), в торговом центре).

Примечание: «heaviside» и «gauss» в названиях признаков характеризуют методику расчета расстояния.

Задание: создайте подходящую модель, наиболее точно прогнозирующую товарооборот кофеен. При создании модели рекомендуется опираться на бизнес-смысл исследуемой отрасли. Примите к сведению, что среди посетителей данной кофейни преобладают женщины. Проинтерпретируйте признаки, используемые в модели. В качестве ключевой метрики используйте MAPE (mean average percentage error). Выведите ее значение на кросс-валидации на обучающей выборке. Заполните спрогнозированными на тестовой выборке при помощи полученной модели значениями приложенную форму.

Формат отчета: Моделирование необходимо выполнить на python. На выходе нужно предоставить jupyter\_notebook / rmarkdown / аналог с понятным описанием логики построения модели и интерпретацией результатов или аналог в doc / ppt / pdf / html.