

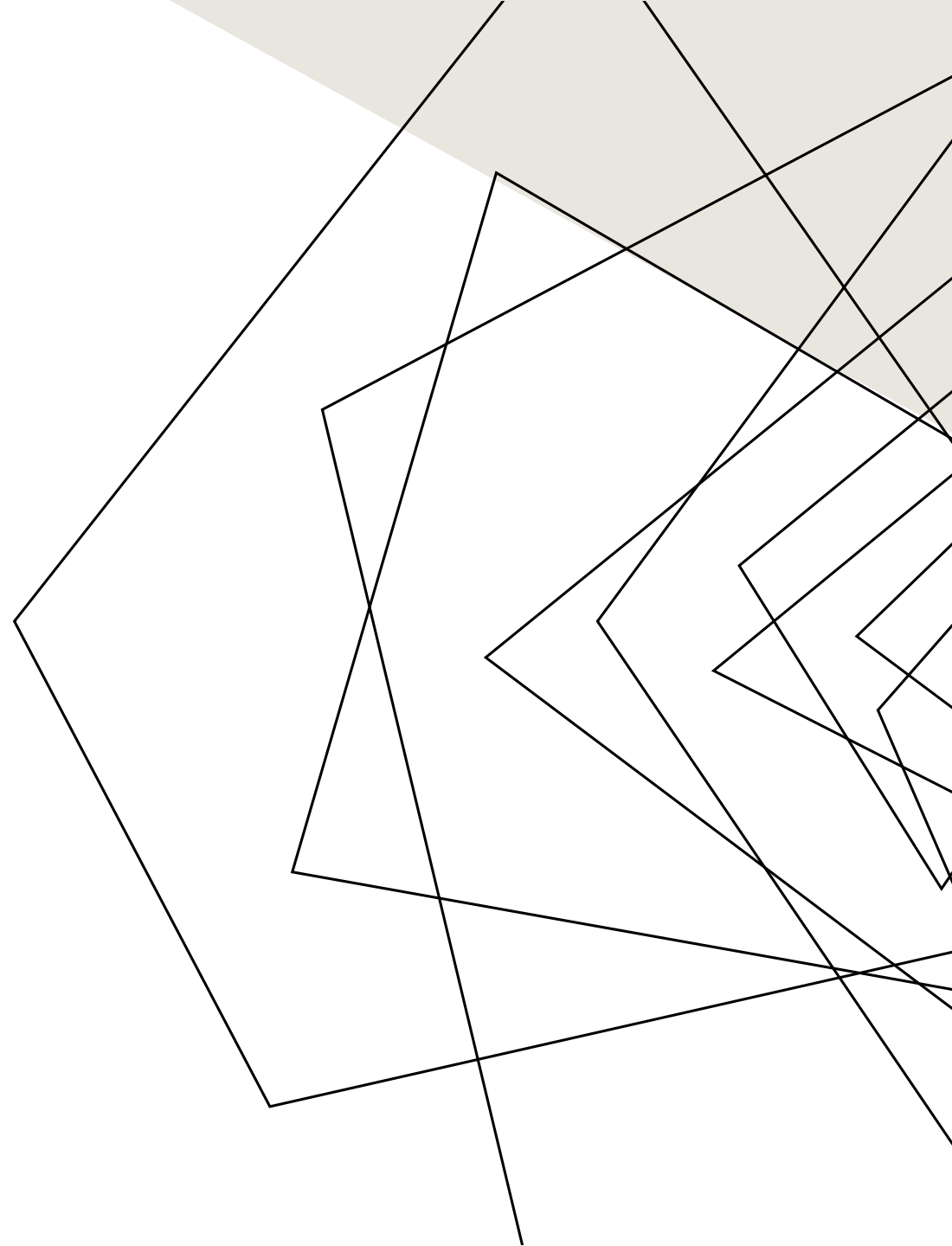
Abstract geometric lines in the top left corner of the page, consisting of several white lines of varying lengths and orientations that intersect to form a series of overlapping triangles and polygons.

Ляра Максим Витальевич

**РЕШЕНИЕ
ТЕСТОВОГО
ЗАДАНИЯ НА
ПОЗИЦИЮ DATA
SCIENTIST**

ОБО МНЕ

Data Scientist, аналитик с опытом работы в крупной FMCG-компании. Умею находить общий язык как со смежными ролями, так и с бизнесом, переводить потребности и задачи компании в задачи в области аналитики, Data Science

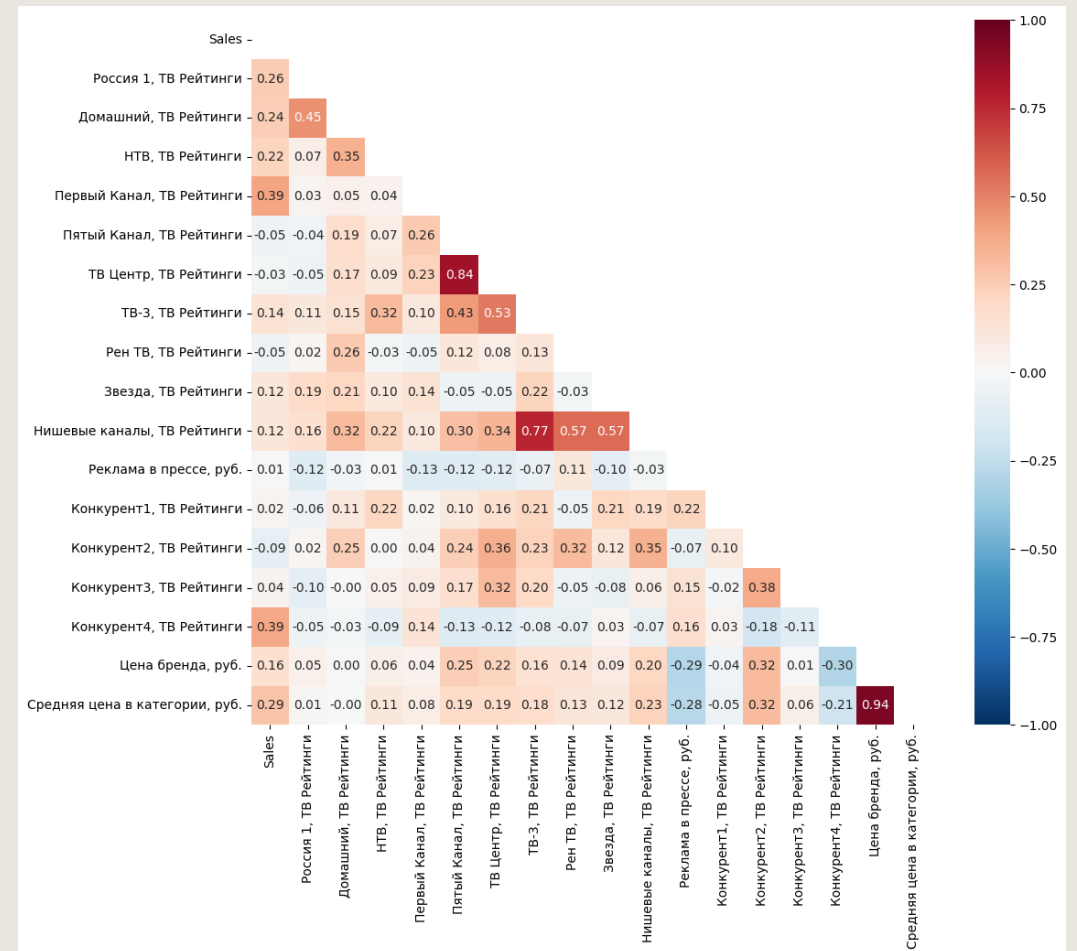


Задачи

- 1. Построить понедельный прогноз продаж на 2013г. (с 31.12.2012 по 29.09.2013), обосновать
- 2. Количественно оценить влияние рекламной активности на ТВ
- 3. Построить график динамики целевого показателя + график оценок
- 4. Интерпретация результатов, рекомендации по улучшению модели

КОРЕЛЛЯЦИОННЫЙ АНАЛИЗ

- 'ТВ Центр, ТВ Рейтинги': высокая корреляция с 'Пятый Канал, ТВ Рейтинги' (0.84) и 'ТВ-3, ТВ Рейтинги' (0.43).
Решение: **удалить** 'ТВ Центр, ТВ Рейтинги' для снижения мультиколлинеарности.
- "Цена бренда, руб.": высокая корреляция с "Средняя цена в категории, руб." (0.94).
Решение: **создать** признак "Отношение цены к цене конкурентов" и **удалить** исходные.
- 'Нишевые каналы, ТВ Рейтинги': корреляция с 'ТВ-3' (0.77), 'ТВ Рейтинги Звезда' (0.57), и 'Звезда' (0.57).
Решение: **удалить** 'Нишевые каналы, ТВ Рейтинги'.



1. ПРОГНОЗ ПРОДАЖ НА 2013 ГОД: МЕТОД И ОБОСНОВАНИЕ

Метод:

- Выбрана **Lasso-регрессия** — подходит для задач с высокой мультиколлинеарностью, что актуально из-за значительных корреляций между признаками по ТВ-каналам.


Параметры модели:

- Тестовый размер:** 30% всей выборки, **без перемешивания** данных, чтобы сохранить временную структуру данных (работа с временным рядом).
- Масштабирование данных** с использованием **StandardScaler** — необходимо для Lasso, но снижает интерпретируемость коэффициентов.

Дополнительно:

- Для понимания значимости отдельных признаков был построен классический линейный регрессионный анализ на тех же данных (без масштабирования), что позволило получить более интерпретируемые коэффициенты.





2. КОЛИЧЕСТВЕННАЯ ОЦЕНКА ВЛИЯНИЯ РЕКЛАМНОЙ АКТИВНОСТИ НА ТВ

Описательная статистика коэффициентов
рекламной активности:

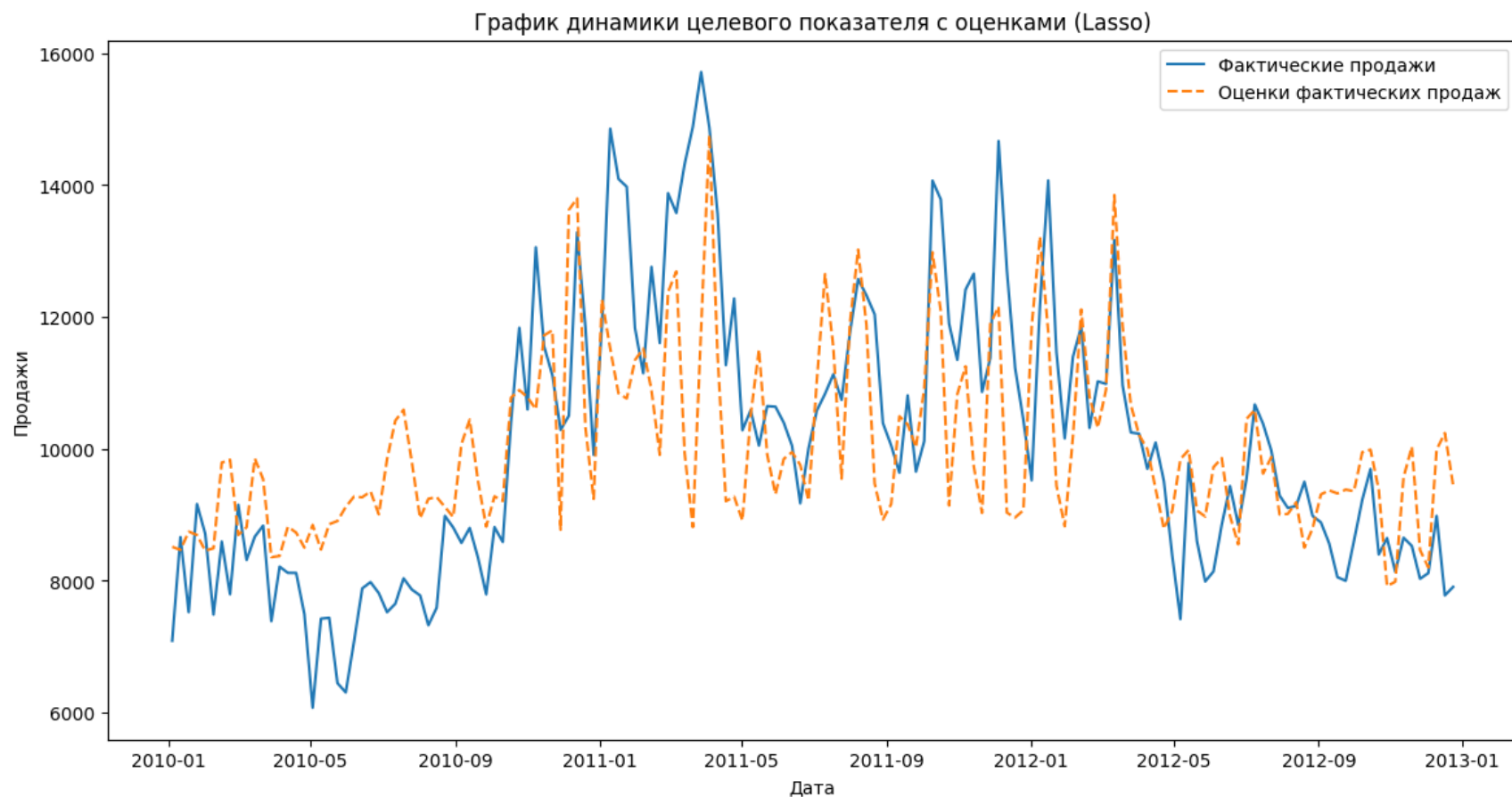
- Среднее значение: 31.8
- Стандартное отклонение: 60.0
- Минимум: -49.0
- Медиана (50%): 14.1
- Максимум: 135.9

Суммарное влияние рекламной активности:
254.2

Коэффициенты влияния по каналам:

- Россия 1: 28.2
- Домашний: -49.0
- НТВ: 100.6
- Первый Канал: 38.5
- Пятый Канал: -0.0
- ТВ-3: 0.0
- Рен ТВ: 0.0
- Звезда: 135.9

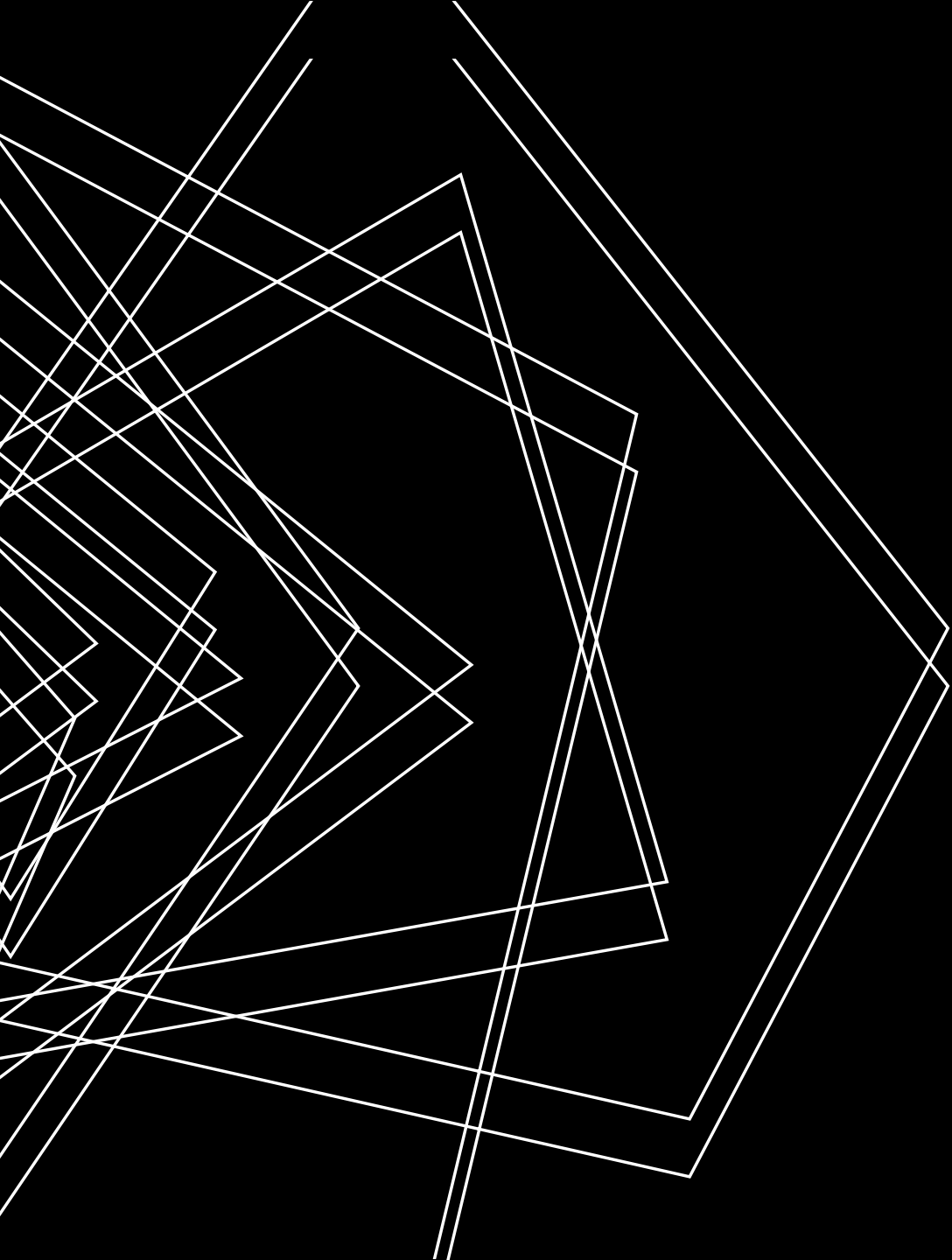
3. ГРАФИК ДИНАМИКИ ЦЕЛЕВОГО ПОКАЗАТЕЛЯ + ГРАФИК ОЦЕНОК



4. ИНТЕРПРЕТАЦИЯ РЕЗУЛЬТАТОВ, РЕКОМЕНДАЦИИ ПО УЛУЧШЕНИЮ МОДЕЛИ

	Коэффициенты	Стандартные ошибки	t-статистика	p-значения
const	-864.67	2735.08	-0.32	0.75
Россия 1, ТВ Рейтинги	28.23	10.25	2.76	0.01
Домашний, ТВ Рейтинги	-49.00	64.60	-0.76	0.45
НТВ, ТВ Рейтинги	100.62	36.88	2.73	0.01
Первый Канал, ТВ Рейтинги	38.52	10.08	3.82	0.00
Пятый Канал, ТВ Рейтинги	-0.00	0.00	-2.80	0.01
ТВ-3, ТВ Рейтинги	0.00	0.00	3.16	0.00
Рен ТВ, ТВ Рейтинги	0.00	0.00	3.07	0.00
Звезда, ТВ Рейтинги	135.90	199.44	0.68	0.50
Реклама в прессе, руб.	7.03	8.88	0.79	0.43
Конкурент1, ТВ Рейтинги	-0.63	13.18	-0.05	0.96
Конкурент2, ТВ Рейтинги	3.04	67.03	0.05	0.96
Конкурент3, ТВ Рейтинги	60.11	33.18	1.81	0.07
Конкурент4, ТВ Рейтинги	129.24	28.36	4.56	0.00
Отношение цены к цене конкурентов	16939.15	4721.64	3.59	0.00

- **Коэффициенты:**
- **Россия 1, НТВ, Первый Канал** — значительное влияние на продажи.
- **Домашний, ТВ-3, Рен ТВ** — нулевой эффект на продажи.
- **Конкуренты:** высокий эффект от конкурента 4, наибольшее снижение продаж при высокой цене по сравнению с конкурентами.
- **Рекомендации по улучшению:**
- **Новые признаки:** учесть **сезонность, цены конкурентов, соцмедиа, промо-акции.**
- **Модели:** рассмотреть **нелинейные методы** (градиентный бустинг, деревья решений) для сложных зависимостей.
- **Временные лаги:** учесть отложенное влияние рекламы.
- **Проверка на мультиколлинеарность:** использовать VIF для коррелирующих признаков.



Ссылка на решение:

https://github.com/maxlyara1/test_task_DS