

Аналитический отчет по результатам исследования работы магазина спортивных товаров

Цель нашего анализа — выявить категории клиентов, для которых с высокой точностью можно предсказать выбор продукта компании.

Задачи:

- проанализировать эффективность проведенной маркетинговой кампании;
- разделить клиентов на кластеры;
- построить модель предсказания выбора продукта клиентом.

Задача 1.

Маркетинговая кампания проводилась в период с 5-го по 16-й день и включала в себя предоставление персональной скидки 5 000 клиентам через email-рассылку.

Для анализа проведенной маркетинговой кампании, мы разделили данные на тех, кому предоставлялась скидка и кому нет.

Сравним данные о сумме проданных товаров в двух группах.

Нулевая гипотеза: доходы в тестовой и контрольной группах не отличаются.

Альтернативная гипотеза: доходы отличаются в группах.

Доход в тестовой группе: 114,343,171 руб.

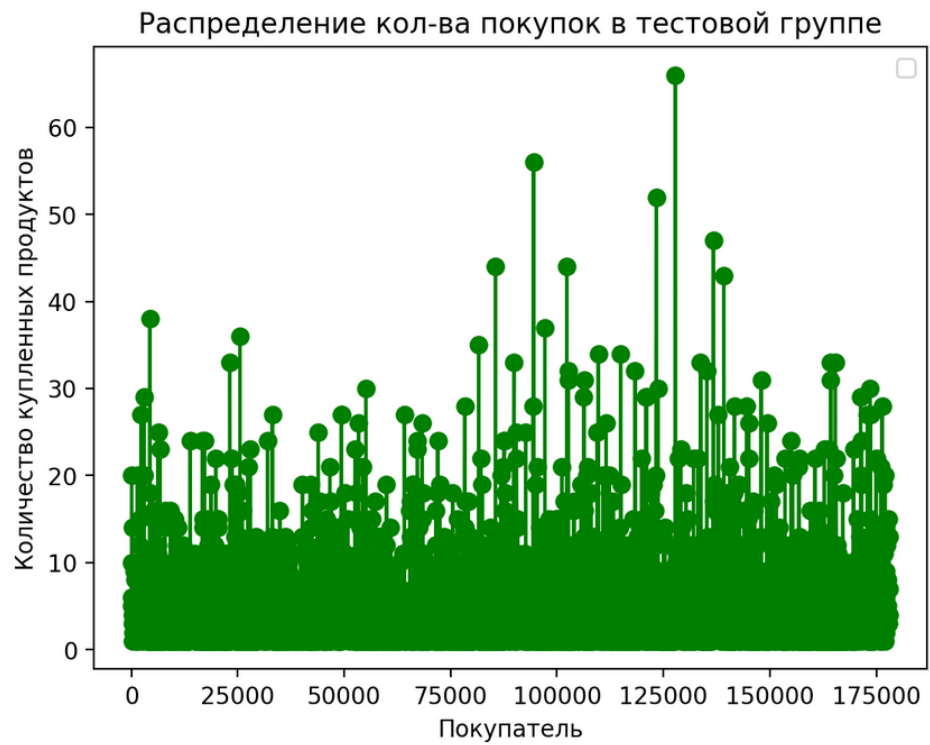
Доход в контрольной группе: 94,508,396 руб.

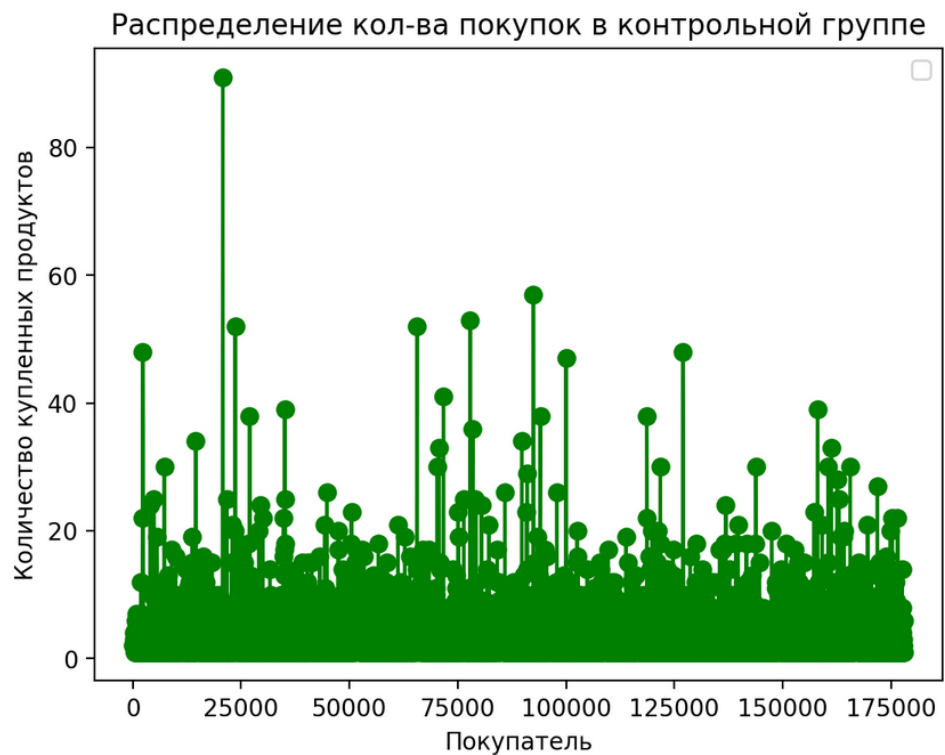
Статистический тест Манна-Уитни показал, что разница доходов является статистически значимой, следовательно мы отвергаем нулевую гипотезу и признаем, что в разрезе метрики полученного дохода, маркетинговая кампания показала успешный результат.

Следующей метрикой возьмем распределение количества купленных товаров каждым покупателем.

Нулевая гипотеза: разницы в количестве купленных товаров каждым покупателем нет.

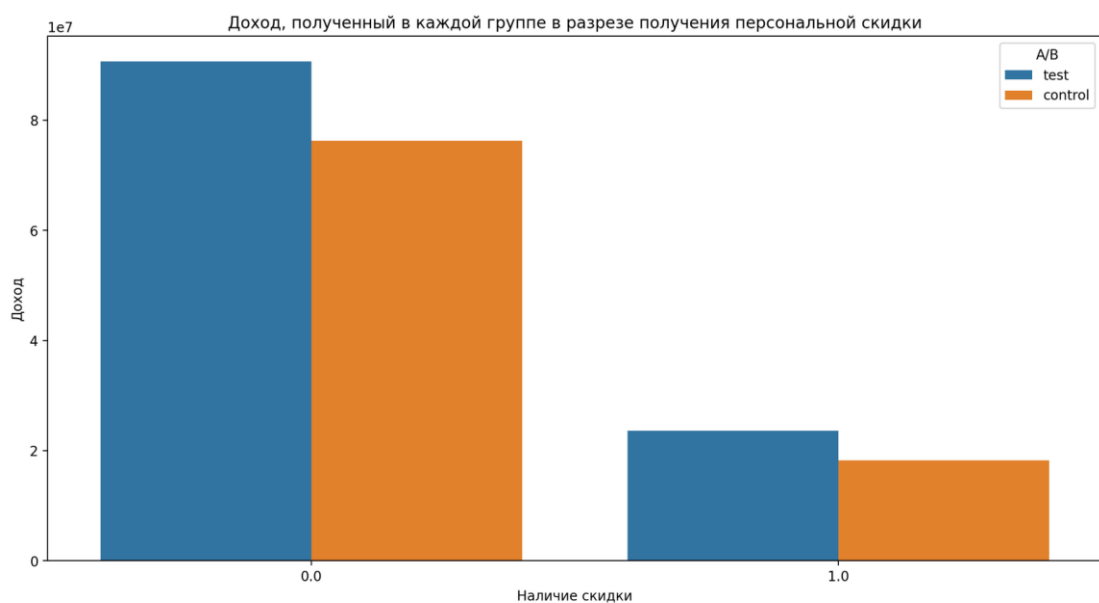
Альтернативная гипотеза: разница существует.





Статистический тест Манна-Уитни показал, что разница распределения количества покупок в группах является статистически значимой, следовательно мы отвергаем нулевую гипотезу и признаем, что в разрезе метрики частоты покупок, маркетинговая кампания показала успешный результат.

Далее, мы сравнили доходы, полученные в каждой группе, учитывая наличие дополнительной персональной скидки.



По представленному графику видно, что доходы в тестовой группе выше и при наличии персональной скидки и без нее.

Итог: маркетинговая кампания показала положительные результаты.

Задача 2.

В первую очередь мы объединили данные, проверили их на пропуски и выявили аномалии.

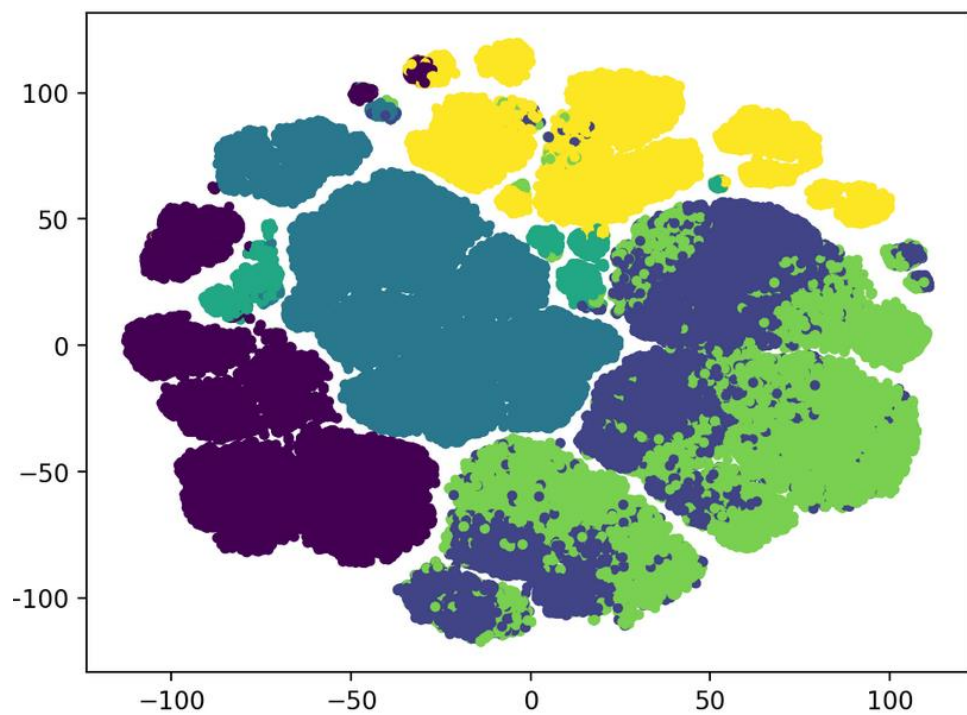
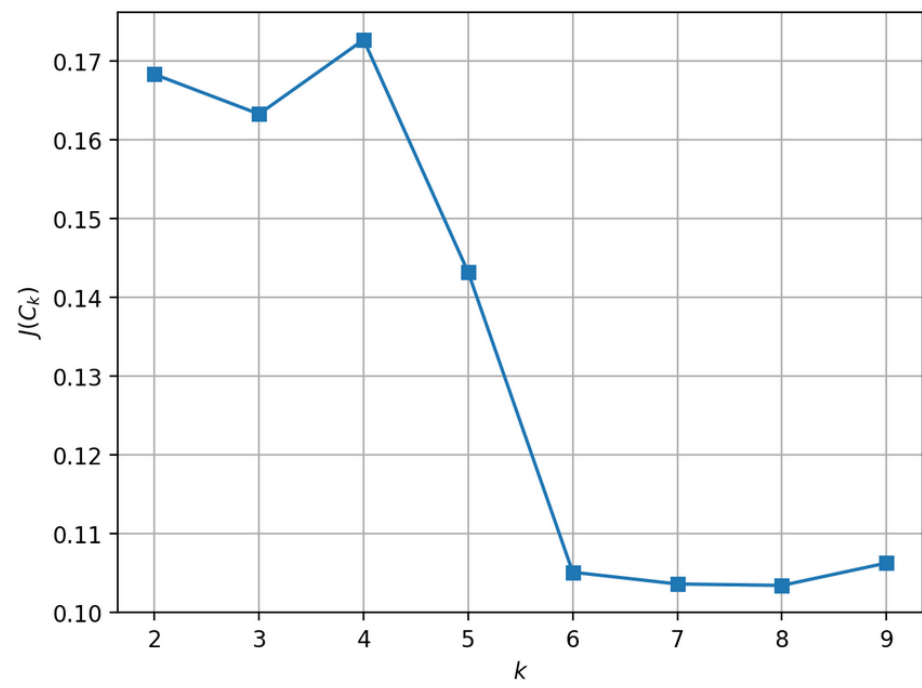
Пропуски в колонке с цветом товара заменили на значение «неизвестно», а в колонке с принадлежностью продукта к полу клиента заполнили исходя из наличия в описании продукта определений «мужская», «женская», «для мальчиков», «для девочек» и т.д., а также значением «неизвестно», в случае, когда ни один вариант не был найден.

Аномальные значения в колонке возраста покупателя (например, возраст некоторых покупателей был равен 5), заменены на среднее значение возраста всех покупателей.

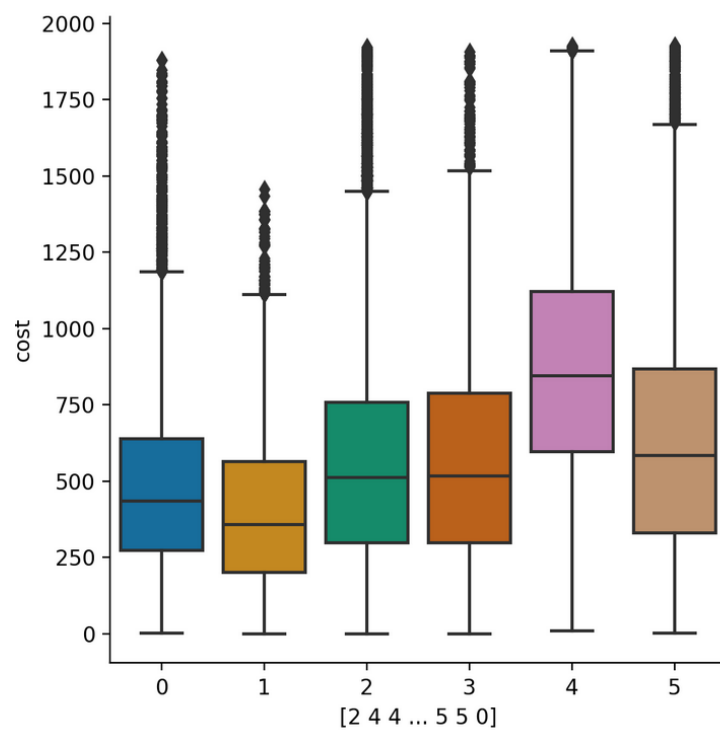
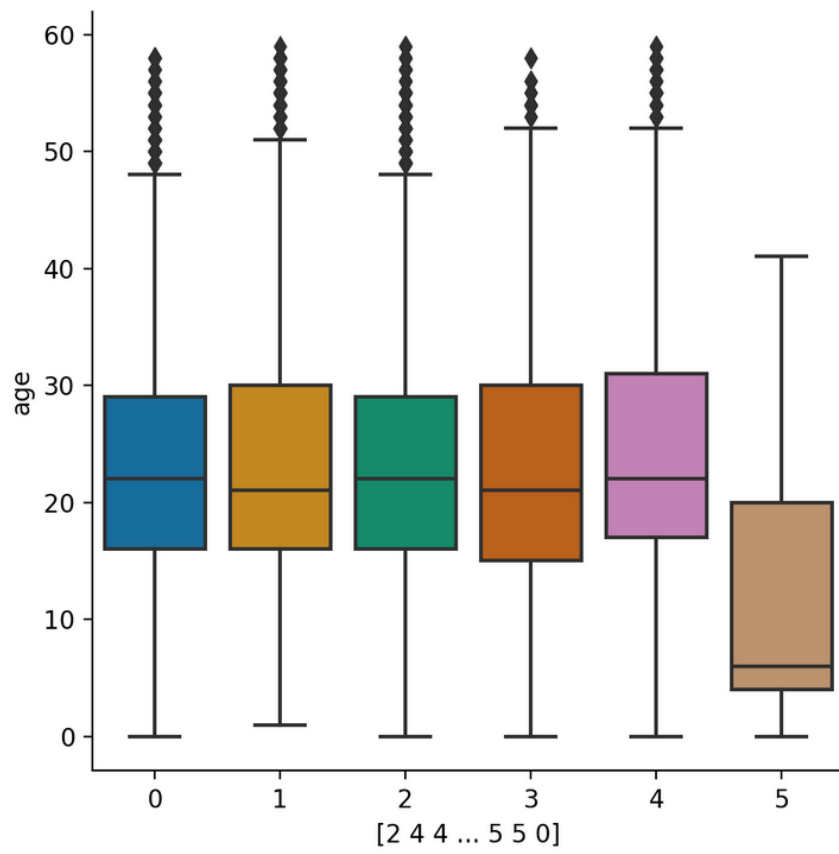
Значения колонки «product» содержат очень много описательных слов, в связи с чем в таблице очень много категорий товаров. Для того, чтобы оставить более обширные категории, мы почистили каждое описание продукта, убрав принадлежность к полу («мужская», «женская», «для мальчиков», «для девочек» и т.д.), а также названия фирм производителя.

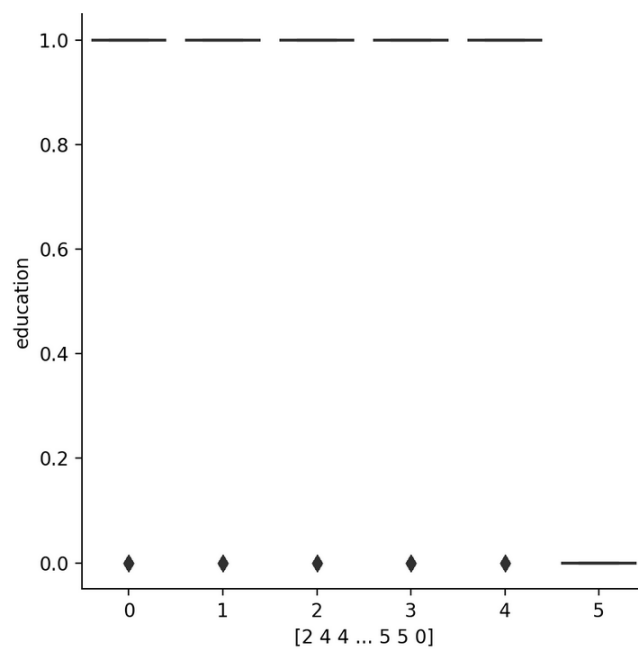
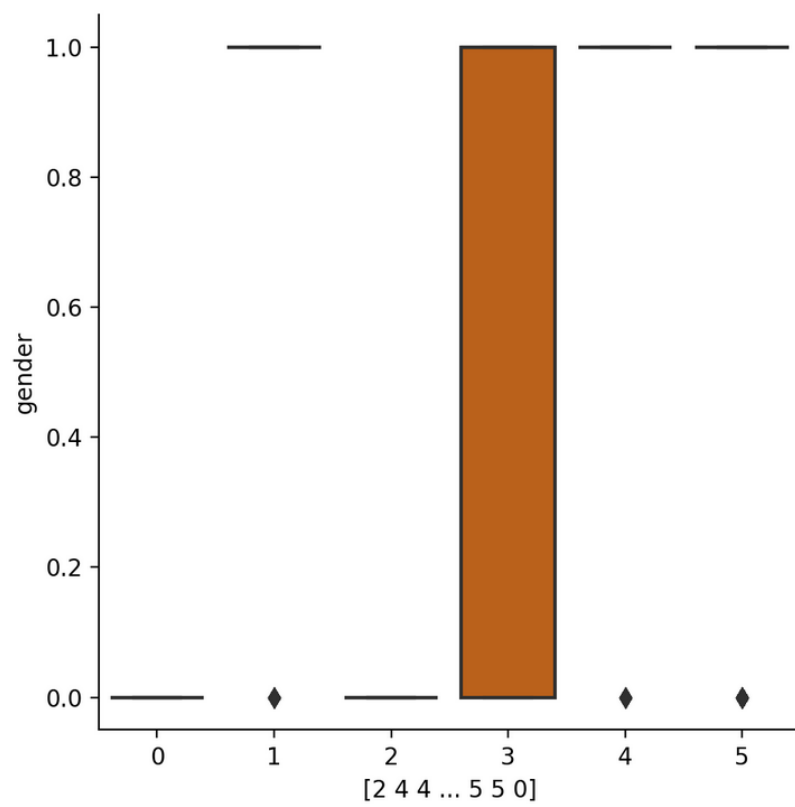
В колонке с цветом товара присутствуют сокращения на английском (заменены на цвет «неизвестно»), а также цвета, перечисленные через «/» (заменены на «и»).

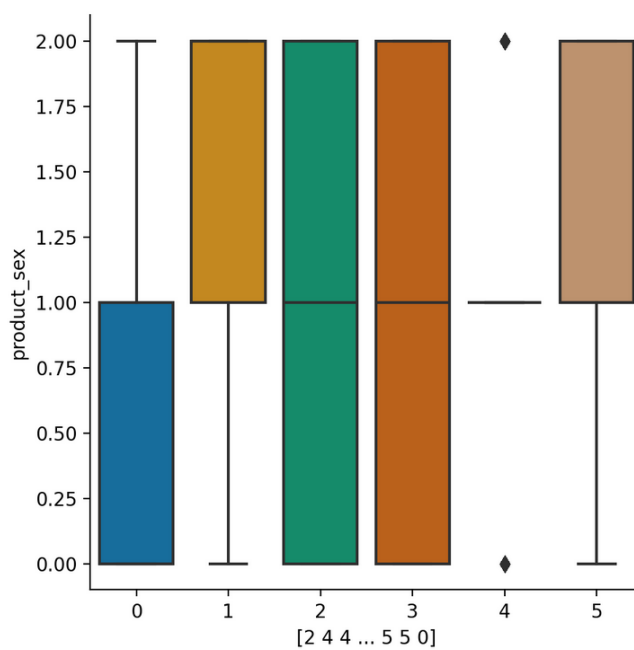
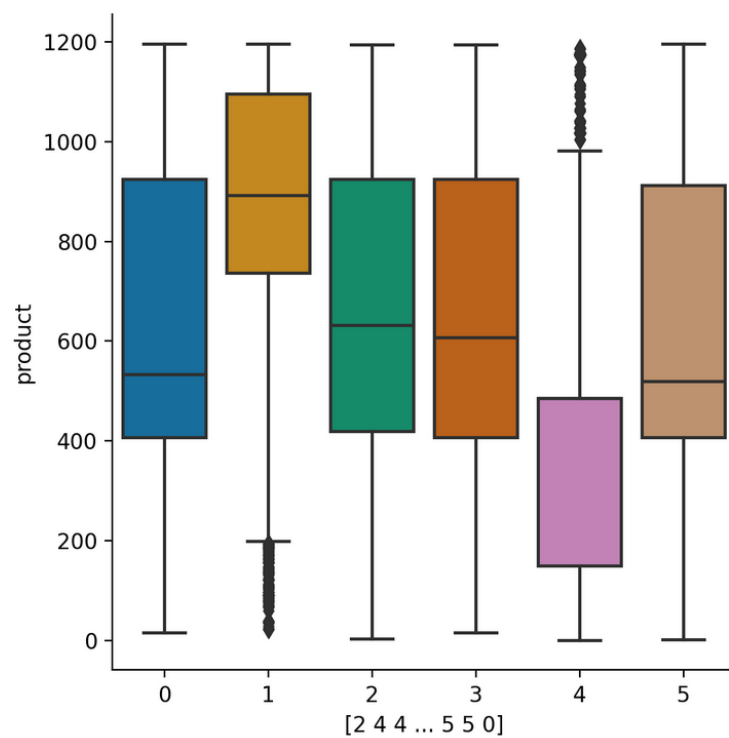
Для определения числа кластеров, использовали метод локтя, по итогам которого выбрано количество кластеров равное 6.



Далее мы построили графики для определения явных признаков каждого кластера.







Проанализировав полученные визуализации, мы описали каждый кластер.

Кластер 0:

В основном покупатели 30-45 лет, женского пола, имеющие среднее образование, покупающие продукты для женского или мужского пола, 2-ой ценовой категории, где 1-ая – продукты минимальной стоимости.

Кластер 1:

В основном покупатели 30-45 лет, мужского пола, имеющие среднее образование, покупающие продукты для мужского пола или не имеющие принадлежности к полу, 1-ой ценовой категории.

Кластер 2:

В основном покупатели 30-45 лет, женского пола, имеющие среднее образование, покупающие продукты всех типов принадлежности к полу, 3-ей ценовой категории.

Кластер 3:

В основном покупатели 25-40 лет, женского и мужского пола, имеющие среднее образование, покупающие продукты всех типов принадлежности к полу, 4-ой ценовой категории.

Кластер 4:

В основном покупатели 35-50 лет, мужского пола, имеющие среднее образование, покупающие продукты для мужского пола, 6-ой ценовой категории.

Кластер 5:

В основном покупатели 20-35 лет, мужского пола, имеющие высшее образование, покупающие продукты для мужского пола или не имеющие принадлежности к полу, 5-ой ценовой категории.

Задача 3.

Лучшие показатели показала модель логистической регрессии:

`model.score(x_test, y_test) = 0.25288333333333335`

Однако, значение равное 25%, само по себе очень низкое. Скорее всего это связано с несовершенной системой категорий продуктов, цветов, а также принадлежности продукта к полу.

Итоги:

В качестве пожеланий руководству магазина для улучшения дальнейшей работы:

- разбить продукты на более точные категории, либо предоставить таблицу сопоставлений;
- предоставить таблицу соответствий цветов, указанных на латинице русским значениям;
- усовершенствовать систему распознавания принадлежности к полу каждого продукта.

Проводить новые маркетинговые кампании, опираясь на итоги проведенной кластеризации (например, предлагать скидки на мужские товары кластерам 1, 4, 5, а промокоды предлагать исходя из ценовой категории продуктов каждого кластера – чем выше стоимость продуктов, тем больше скидка и т.д.).

Проведенная маркетинговая кампания показала хорошие результаты. Ее можно продолжать использовать, учтя проведенный анализ.