

Дипломная работа по программе «Аналитик данных»

# Прогнозирование спроса в онлайн ритейле

Александр Сипко

Группа: AML-42

Руководитель: Даниил Корбут



## 1. Актуальность выбранной темы «Прогнозирование спроса в онлайн ритейле»

---

**Прогнозирование спроса** — это ключевая задача для онлайн-ритейла, оказывающая значительное влияние на управление запасами, логистику и планирование продаж.

*«Точные прогнозы позволяют снизить издержки, дефицита товаров и улучшить обслуживание клиентов, что даёт компаниям конкурентное преимущество, повышая эффективность операций и прибыль.»*

### **Стейкхолдеры:**

**Коммерческий директор** – повышение прибыли.

**Менеджеры по закупкам** – минимизация издержек, планирование поставок.

**Логистический отдел** – сокращение затрат и управление запасами.

**Маркетологи** – планирование активностей и промо-компаний.

**IT-отдел** – поддержка инфраструктуры для сбора и обработки данных.

**Финансовый отдел** – бюджетирование и оценка рентабельности.

### **Источник данных:**

#### **Выгрузка Excel содержащая данные о продажах за 3 года**

- InvoiceNo — номер заказа
- StockCode — номер товара
- Description — название товара
- Quantity — количество
- InvoiceDate — дата и время транзакции
- UnitPrice — стоимость
- CustomerID — id клиента
- Country — страна (нет данных о значении поля)

### **Пред обработка данных:**

- были заменены пропущенные значения;
- устранены некорректные записи;
- обработаны описания товаров для дальнейшего анализа

### **3. План работ и решаемые задачи «Прогнозирование спроса в онлайн ритейле»**

---

**1. Постановка задачи**

**2. Подготовка данных для работы**

**3. Кластеризация товаров и присвоение каждому кластеру наименования для дальнейшей работы**

**5. Анализ продаж по категориям**

**6. Прогнозирование продаж**

**7. Выводы о проделанной работе**

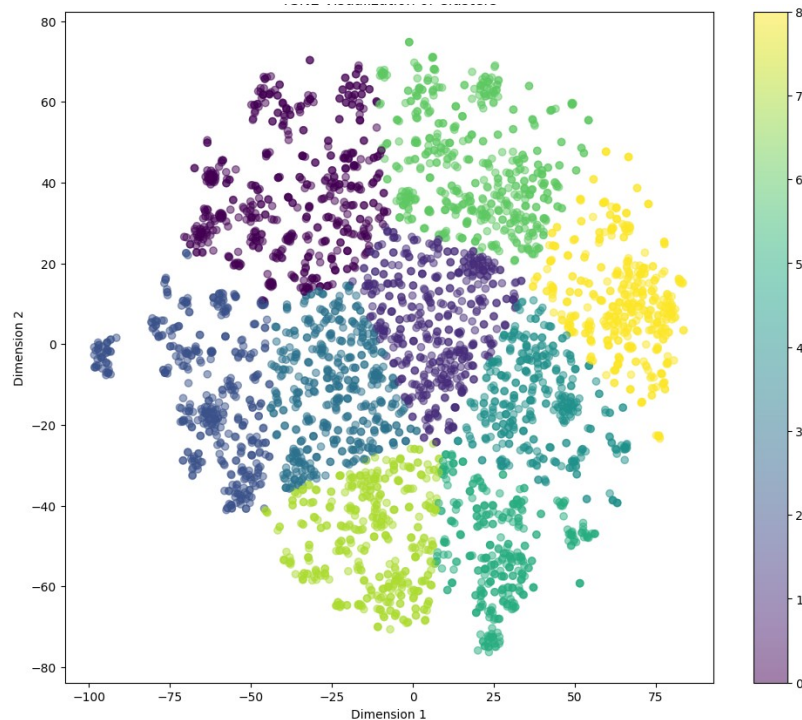
Использование текстовых моделей позволило преобразовать наименование товаров в эмбединги, а затем, используя **K-Means**, разделить товары на кластеры. Для оценки качества кластеризации использовалась оценка **Silhouette Score** (значение от -1 до 1, где 1 указывает на хорошую кластеризацию).

В работе применялись предобученные модели, такие как **paraphrase-MiniLM-L6-v2**, **all-MiniLM-L6-v2** и другие, для представления товаров в векторном пространстве.

**Best model:** paraphrase-MiniLM-L6-v2

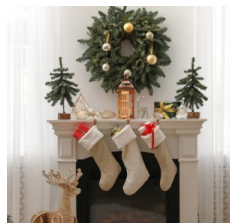
**Best number of clusters (K):** 9

**Best Silhouette Score:** 0.08163392543792725



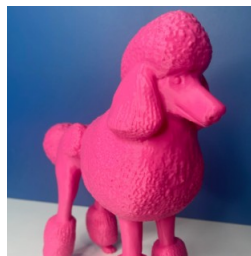
## 5. Осмысление кластеров

## «Прогнозирование спроса в онлайн ритейле»

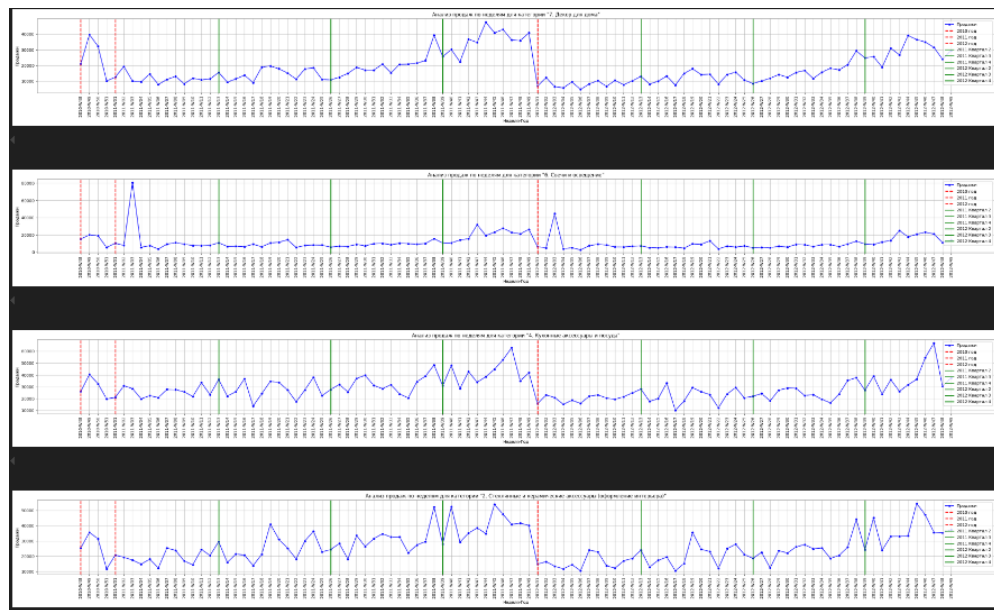


*\* нумерация фото слева на право, начиная с верхнего ряда*

1. Стилизованные вещи интерьера с уникальным дизайном
2. Стекланные и керамические аксессуары (оформление интерьера)
3. Часы и рождественские украшения
4. Кухонные аксессуары и посуда
5. Игрушки или детские товары
6. Свечи и освещение
7. Декор для дома
8. Праздничные украшения
9. Браслеты бижутерия



**Временные ряды кластеров товаров были разбиты с шагом в 1 неделю для визуальной оценки и определения наличия сезонности. Ниже представлены примеры построения графиков для визуального анализа.**



Временные ряды отражают сезонность для различных категорий.

Также отмечается короткий диапазон временного ряда, необходимый для формирования устойчивого сезонного паттерна.

\* полный перечень графиков представлен в приложенном к проекту \*.ipynb файле

**Для прогнозирования временного ряда использовались модели Prophet, ARIMA. Была сформирована функция для определения наилучшей модели и оптимальных гиперпараметров. Для этого применялись следующие метрики:**

**Для оценки моделей временных рядов использовались метрики:**

- Среднеквадратическая ошибка (MSE)
- Средняя абсолютная ошибка (MAE)
- Коэффициент детерминации ( $R^2$ )
- Средняя абсолютная процентная ошибка (MAPE)
- Сравнение с базовой линией, построенной на основании скользящего среднего.

Для прогнозирования продаж была применена модель Prophet, которая показала высокую точность на тестовых данных. Построенные прогнозы охватывали период в 12 недель, с использованием данных последних 4 недель в качестве исторического контекста.



**Примеры категории товаров демонстрирующие улучшение метрик, по сравнению с бейзлайн**

### 8. Праздничные украшения



### 3. Часы и рождественские украшения



\* полный перечень графиков представлен в приложенном к проекту \*.ipynb файле

**Оценка качества прогнозирования продаж для каждой категории**

	Category	Baseline MSE	Baseline MAE	Baseline $R^2$	Prophet MSE	Prophet MAE	Prophet $R^2$
0	7. Декор для дома	1.890789e+07	3342.723684	0.684877	1.848973e+07	3315.071250	0.729523
1	6. Свечи и освещение	1.127422e+07	2185.328947	0.643953	2.253267e+07	3962.301665	0.320746
2	4. Кухонные аксессуары и посуда	8.003183e+07	7222.868421	0.426662	6.589700e+07	6478.019519	0.495356
3	2. Стекланные и керамические аксессуары (оформ...	5.283409e+07	5454.565789	0.422399	5.362271e+07	6075.268235	0.477214
4	8. Праздничные украшения	4.353351e+07	4145.197368	0.654384	1.095873e+08	7927.023640	0.076631
5	1. Стилизованные вещи интерьера с уникальным д...	1.364643e+07	3223.578947	0.448022	2.097828e+07	3840.286861	0.121624
6	5. Игрушки или детские товары	5.240364e+06	1786.486842	0.547259	9.362651e+06	2493.913050	0.185415
7	9. Браслеты бижутерия	4.651652e+05	509.197368	0.586350	1.076343e+06	893.679541	-0.051781
8	3. Часы и рождественские украшения	7.531862e+07	5792.407895	0.555005	3.908192e+07	4488.381398	0.761454
9	11. Прочие	2.732996e+05	426.539474	0.147828	1.358731e+06	1014.865336	-3.644213

Показатель с высоким  $R^2$  и низким MSE и MAE, а так же показавшим улучшение показателей по сравнению с байзлайн, могут быть использованы для принятия управленческих решений в бизнесе, другие категорию требуют более аккуратного использования или увеличения диапазона данных для прогнозирования и повышения качества модели

***Проект создал основу для дальнейшего совершенствования инструментов прогнозирования в онлайн-ритейле, обеспечивая бизнес необходимыми аналитическими данными для принятия обоснованных решений. Также в данном проекте использовались мощности видеокарт для расчета данных, все данные взяты из открытых источников и не содержат коммерческой тайны.***

**Спасибо за внимание!**

