



Федеральное государственное образовательное бюджетное
учреждение высшего образования

**«ФИНАНСОВЫЙ УНИВЕРСИТЕТ ПРИ ПРАВИТЕЛЬСТВЕ
РОССИЙСКОЙ ФЕДЕРАЦИИ»
(Финансовый университет)**

**Институт развития профессиональных
компетенций и квалификаций**

ИТОГОВАЯ РАБОТА

Группа обучения	«AD_potok_2_11.05.23»
Срок обучения	«11.05.23»
Козеровский Максим Витальевич	
Номер Кейса	«10»
Название Датасета	«Продажа авиамоделей в США за период 2017–2019 гг.»



Ссылка на Excel файл кейса № 10

<https://docs.google.com/spreadsheets/d/1zTyJZXKnN38o431OAIX958pTTSSZR9PD/edit?usp=sharing&ouid=112933034334416794803&rtpof=true&sd=true>

Ссылка на очищенные и подготовленные данные:

https://drive.google.com/file/d/1Autj_hYQ8YXhmakiW5_B11do4O5E5aJa/view?usp=sharing

Ссылка на файл KNIME:

<https://drive.google.com/file/d/18wCkYPY822GLXV71J6Ye26lHa2SAohX1/view?usp=sharing>

Ссылка на Colaboratory файл:

https://colab.research.google.com/drive/1Ey5xhJ2XfJOM9ehGDDDF2fwx8OdcEU2Y?usp=drive_link

Ссылка на Tableau:

https://public.tableau.com/views/ADcourse_FinalTask_case10/Analyzingtheworkofthecompanysellingaircraftmodels?:language=en-US&publish=yes&:display_count=n&:origin=viz_share_link

Кейс

Данные выбранного кейса включают в себя информацию о продажах авиамоделей в US за период с 2017 по 2019 годы. Каждая строка датасета это уникальный заказ.

Столбцы в нём следующие –

RangeIndex: 377741 entries, 0 to 377740

Data columns (total 26 columns):

#	Column	Non-Null Count	Dtype
0	OrderNumber	377741 non-null	object
1	OrderDate	377741 non-null	object
2	ShipDate	377741 non-null	object
3	Quantity	377741 non-null	int64
4	UnitPrice	377741 non-null	float64
5	DiscountAmount	377741 non-null	float64
6	PromotionCode	377741 non-null	object
7	ProductName	377741 non-null	object
8	ProductCategory	377741 non-null	object
9	ItemGroup	377741 non-null	object
10	KitType	377741 non-null	object
11	Channels	377741 non-null	int64
12	Demographic	377741 non-null	object
13	RetailPrice	377741 non-null	float64
14	StateName	377741 non-null	object
15	RegionName	377741 non-null	object
16	UsedPromoCode	377741 non-null	int64
17	Year	377741 non-null	int64
18	SumOrder	377741 non-null	float64
19	Profit	377741 non-null	float64
20	Marginality	377741 non-null	float64
21	ABC_Class_2017	377741 non-null	object
22	ABC_Class_2018	377741 non-null	object
23	ABC_Class_2019	377741 non-null	object
24	XYZ	377741 non-null	object

25 RFM 377741 non-null int64
dtypes: float64(6), int64(5), object(15)

По порядку:

Изначально полученные данные. Номер заказа (уникальное значение), дата заказа(год/месяц/день), дата отгрузки товара (год/месяц/день), количество в заказе, закупочная цена, размер скидки, промо-код, название номенклатуры, категория товара, общая группа товара, готовая модель (rtf) или собирать (kit), количество радиоканалов у модели, уровень сложности авиамодели, цена продажи, штат покупки, регион покупки (укрупнённые зоны проживания покупателей).

Данные полученные в процессе обогащения. Столбец параметра показывающий использовал ли покупатель промо-код. Год совершения заказа. Общая сумма заказа. Посчитанная прибыль. Маржинальность. ABC классификация товара по годам. XYZ классификация товара за весь период. RFM классификация штатов где живут покупатели.

KNIME.

Для очистки и обработки данных использовалась швейцарская программа KNIME.

При обработке проведенного анализа (ABC, XYZ и RFM) можно сделать вывод что несмотря на то, что общая выручка ритейлера растёт, из года в год сокращается количество номенклатурных позиций приносящие 80%-ю долю выручки, большинство ассортимента имеет устойчивую динамику к понижению объёма продаж. При этом немногочисленные товары категории 'А' относятся к группе 'Z', они имеют коэффициент вариативности выше 25%, спрос спрогнозировать сложно. Всего два товара со стабильным спросом имеют низкие продажи. По мимо этого покупатели (они разбиты по штатам) зачастую имеют низкую активность и низкий объём покупок.

На основе этих исследований данных можно сделать следующие выводы:

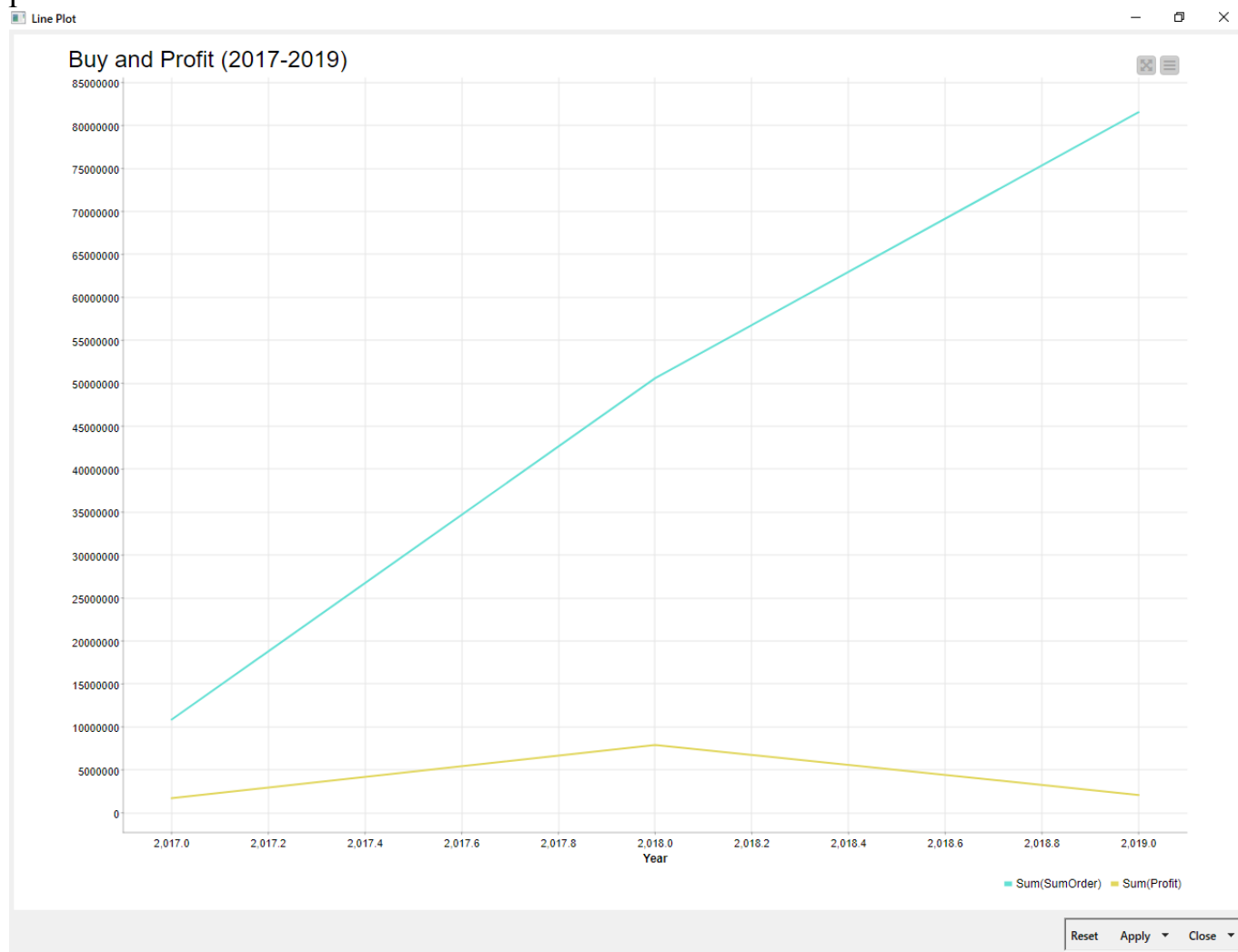
1. Т.к. подавляющее большинство заказов имеют одну позицию, а количество заказов растёт, то отдел маркетинга компании успешно привлекает клиентов.
2. При этом частота покупок со стороны клиентов низкая и так же низкий средний чек. К этому стоит добавить, что большинство клиентов не использует промо-коды. Значит с новыми клиентами плохо ведётся последующая работа. Привлечь нового клиента всегда дороже, чем работать с постоянным.
3. 90% номенклатурных позиций имеют нестабильный спрос. Нестабильность всегда несет риски для бизнеса. За три года сбора данных ассортимент практически не менялся. Имеется только одна авиамодель, которая была введена в продажи в 2018 году и поднялась в категорию "А".



EDA в KNIME.

Общая динамика за отчетный годы по Выручке и прибыли

рис. 1



Видно, что при общем росте Выручки, доходность падает начиная с 2019

Цены покупки и продажи товаров (2017, 2018, 2019 годы)

рис.2

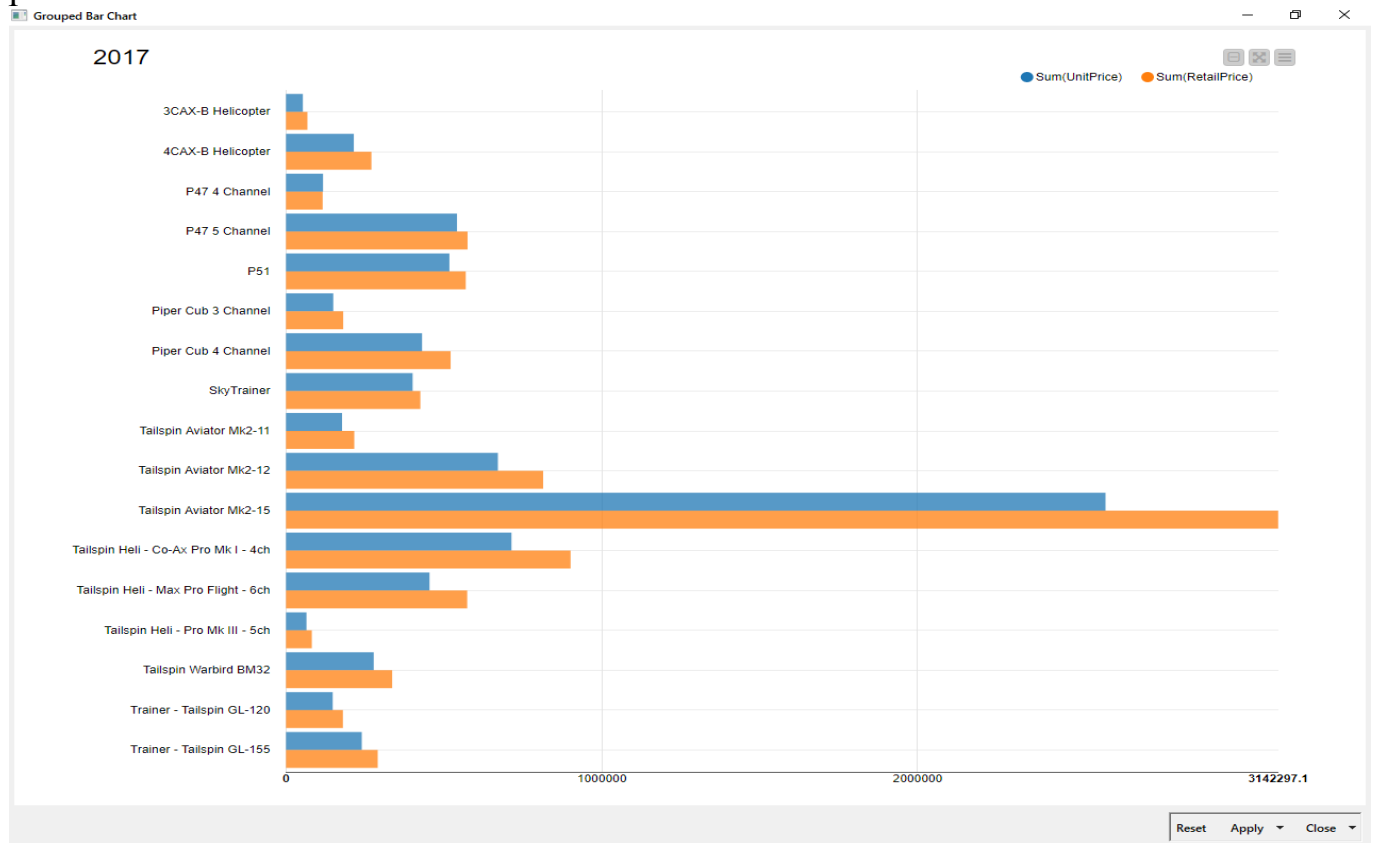


рис.3

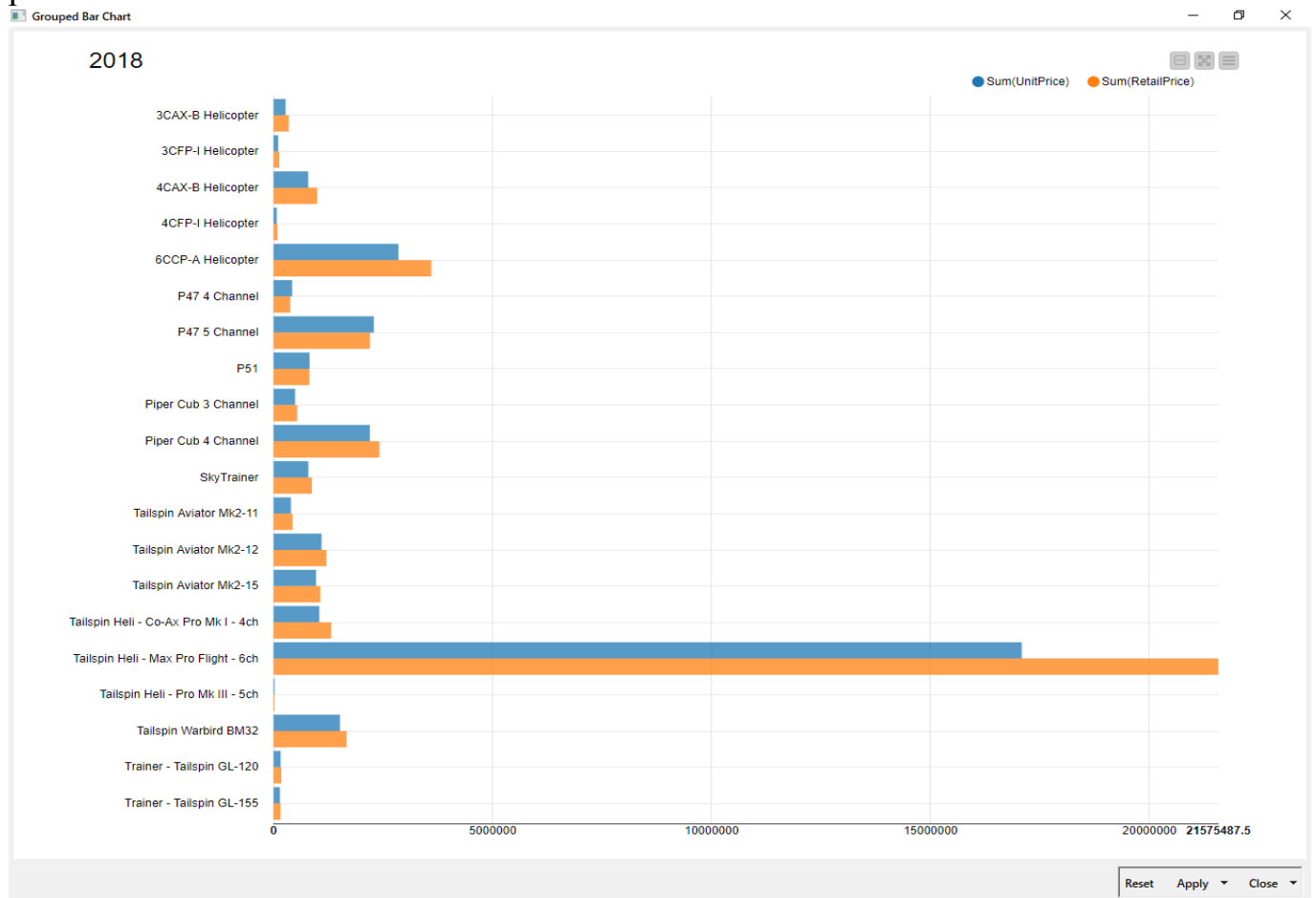
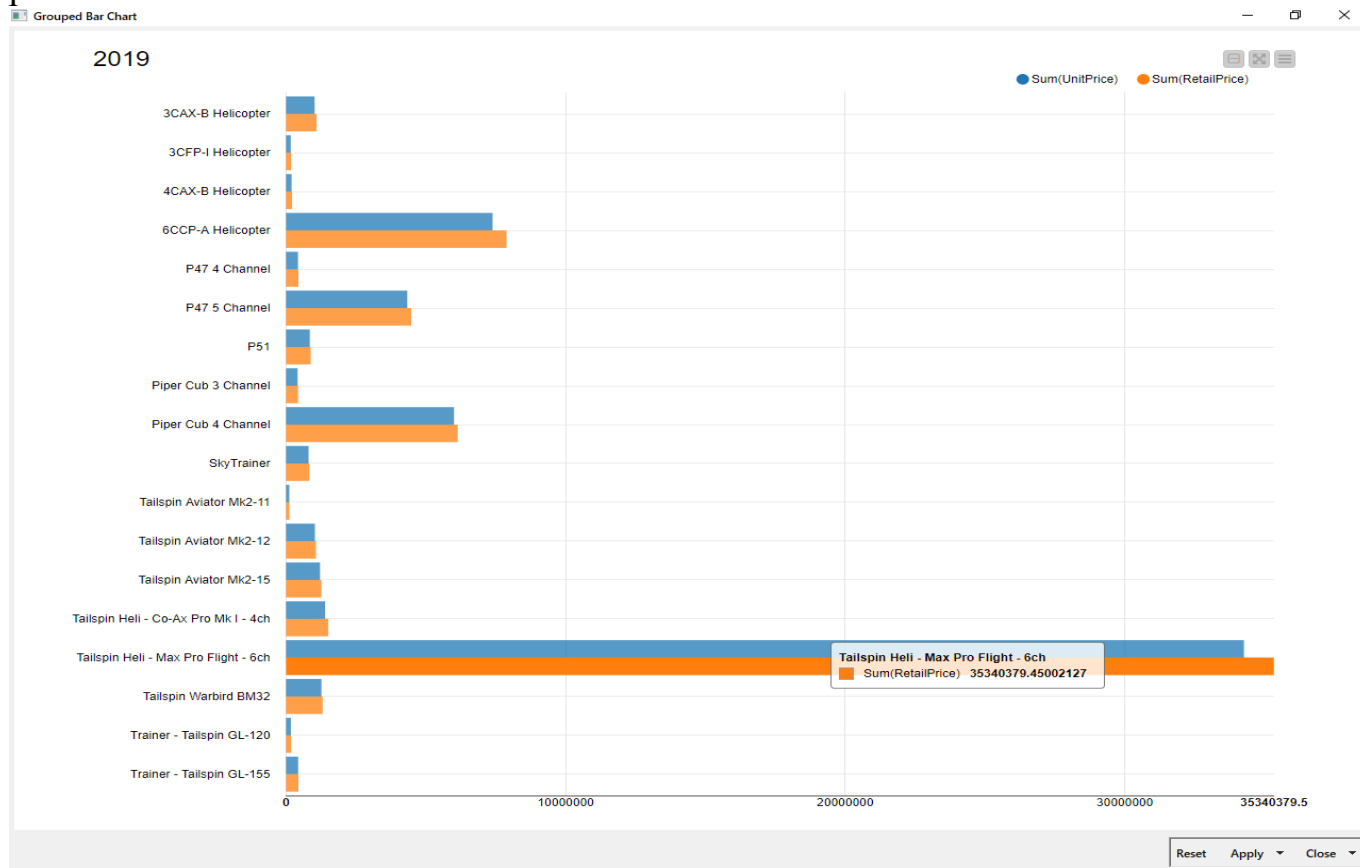


рис.4



При анализе данной визуализации становится ясно, что доходность падает по причине сильного снижения маржинальности. Часть заказов закрывалась с отрицательным результатом.

Необходимы новые поставщики для уменьшения закупочных цен и возможно для предоставления клиентам нового товара. Как видно из данных компания продаёт одни и те же авиамодели, но это тот тип вещей, что не подразумевают повторные покупки того же самого. Нужно либо стимулировать создание коллекций, либо предлагать дополнительные опции к существующим моделям.

ML в KNIME.

Обучались и тестировались три модели машинного обучения. Это Linear regression, XGBoost Linear regression, Time series analysis в двух видах Auto-SARIMA и ARIMA. Решалась задача предсказания получения прибыли.

Программный продукт позволяет логически расставлять необходимые узлы и обучаться на заранее разбитых на два датафрейма (тренировочный и тестовый) данных. После обучения тестовую часть данных прогоняем через созданную систему предсказания.

Результаты по одному из ключевых показателей R^2 следующие:

1. Linear regression - 0.87
2. XGBoost Linear regression - 0.65
3. Auto-SARIMA - 0.39
4. ARIMA – (-0.3)

Для EDA использовались следующие библиотеки Python pandas-profiling и Python dataprep. Обе библиотеки предоставляют интерактивные визуализации и таблицы с отчетами. Основным инструментом, который используют эти библиотеки это группировка, подсчет суммы, среднего, минимум, максимум и прочие агрегирующие функции.

Например:

- Топ 5 дат заказов это три дня в сентябре и два в ноябре. Возможно, это рядом с какими-то национальными праздниками или прошла удачно рекламная кампания.
- Если посмотреть на даты отправления заказов, то видно в какие периоды происходили сбои и это не всегда периоды с большим количеством покупок.
- Количество штук номенклатуры в заказах от 1 до 3. Подавляющее большинство — это покупка одной авиамоделей.
- В закупке лидируют дорогие товары.
- 84% покупок совершаются без скидки.
- Эти 3 наименования товара составляют более 50% оборота компании. Tailspin Heli - Max Pro Flight - 6ch - 26.3%, 6ССР-А Helicopter - 12.7%, Piper Cub 4 Channel - 12.4%
- С большим отрывом лидирую две категории товаров Collective pitch - 147083 шт. и Trainer - 115627 шт.
- Чуть больше покупают вертолеты чем самолеты
- Готовые модели (RTF) покупают 56.0% людей, собирать (KIT) любят 44.0%
- В основном предпочитают профессиональные (Professional - 147083 шт.) и продвинутые (Intermediate - 127671 шт.) модели с большим количеством каналов связи и подключения 6 и 4.
- Цены продаж сильно коррелируются с товарами из сегмента с дорогой закупкой.
- California, Florida и New York штаты лидеры покупок.
- По годам и количеству заказов 2019–203620 шт., 2018–129682 шт., 2017–44439 шт.
- Доход за эти годы. 117728 заказов, а это 31.2% закончились с прибылью 0.
- Моржинальность имеет низкий процент. 117492 продажи (31.1%) совершены с нулевой морженальностью.
- В 2017 году по классу «А» 233501 продажи, в 2018–220381, а в 2019 уже 194006.
- 91.0% заказов пришёлся на товары категории «Z»
- RFM классификация показывает, что на клиентов кластера 311 приходится 77.9% покупок, т. е. редкие и дешёвые, из группы 333 совершилось 11.6%, 322–10.1%. Есть более 20% клиентов, с которыми можно поработать и улучшить положение дел фирмы.

Как и в случае с предыдущим анализом данных (в KNIME), проблема видна очень хорошо, это низкая рентабельность совершаемых сделок. Необходимо искать новых партнеров и снижать закупку или менять ассортимент. Ещё вариант уйти в



сегмент элитных продаж и торговать только дорогими моделями. Меньше, но с более высокой доходностью.

AutoGluon в Colaboratory.

Для работы с моделями машинного обучения был выбран фреймворк AutoML - AutoGluon.

Система была развернута в блокноте Colaboratory, разбит датафрейм на тренировочный и тестовый. При запуске обучения так же была выбрана целевая колонка Прибыль.

AutoGluon подбирает модели под предоставленные данные и производит вычисления и подбор гиперпараметров.

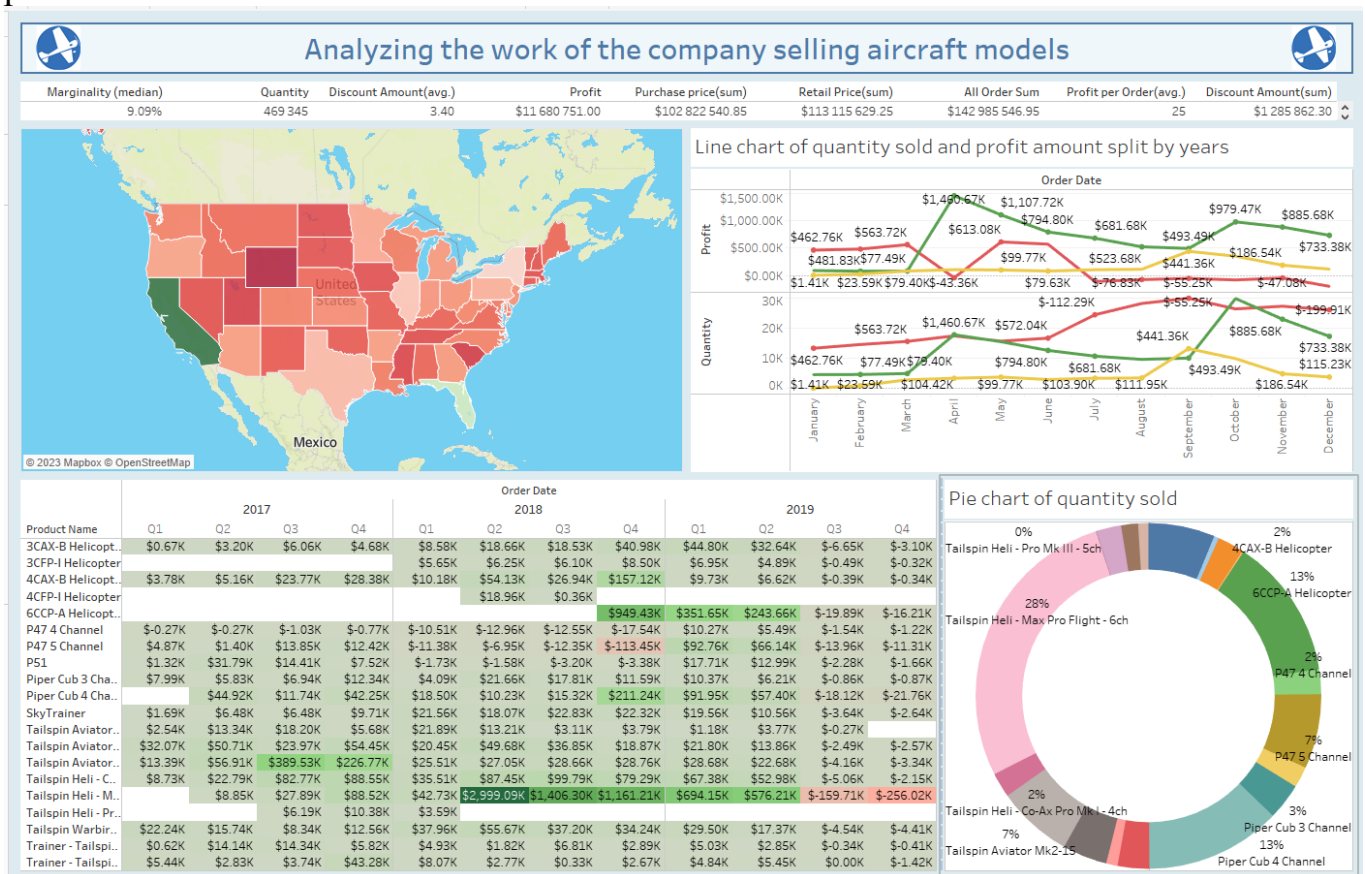
Как и в первом случае авто ML выбрал модели связанные с regression, это модели KNNModel, RFModel, LGBModel, CatBoostModel, WeightedEnsembleModel. В итоге лучшей моделью оказался CatBoost. Один из ключевых показателей R^2 у данного метода составил – 0.999.

Показатель высокий, но есть большая вероятность, что модель переобучилась.

Из плюсов это простота запуска процессов, буквально несколько строк кода. Но расчёт идёт очень продолжительное время. Модель в KNIME имеет хорошие результаты, представляет из себя low-code подход и тренирует модель гораздо быстрее.

Созданный дашборд показывает основные бизнес показатели (прибыль, сумму заказов, среднюю маржинальность, общую сумму скидки, средняя прибыль на заказ и пр.).

рис.5



Визуализация подсвечивает ту же проблему, что и предыдущие анализы. Если 2017 и 2018 годы имеют четкую корреляцию с объёмом продаж и прибылью, то с середины 2019 виден рост продаж и падение рентабельности. Так же первые два года показывают сезонные пики продаж (начало лета и начало осени). Бизнес вёл себя предсказуемо и стабильно.

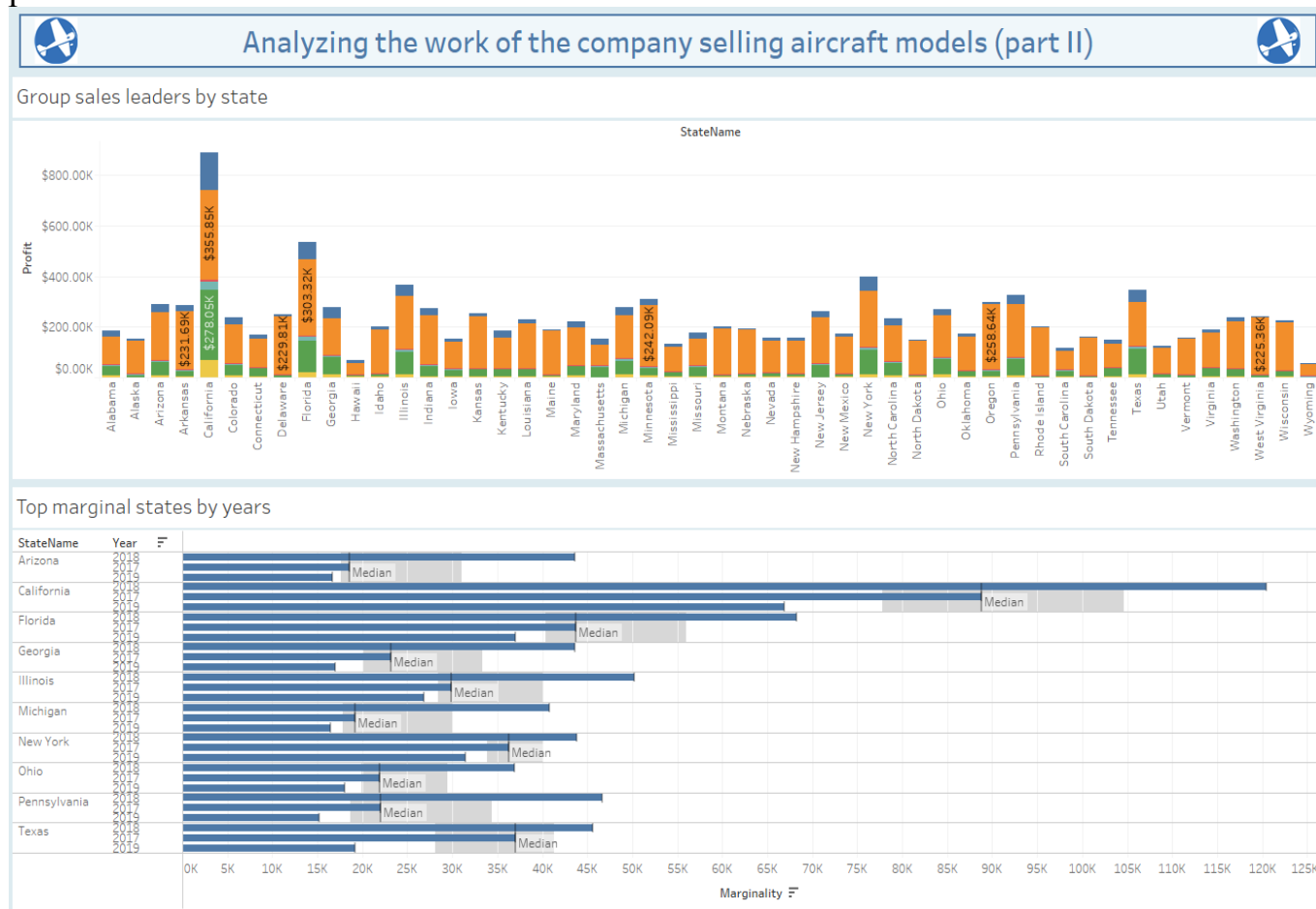
Так же 95% штатов имеют низкую покупательскую активность данной продукции (нужны дополнительные данные, что бы понимать это общее отсутствие интереса или прямые конкуренты более успешны).

Высокую прибыльность в данном периоде показывала только одна авиамодель, и то со второго квартала 2018 года видна динамика падения и последнее полугодие 2019 этот товар продавался в убыток.

Следующий дашборд показывает территориальных лидеров как по количеству заказов, так и по сумме, и по маржинальности сделок. Явных лидера в этих показателях два, это Калифорния и Флорида. Штаты с большим количеством теплых солнечных дней. Сумма заказов и маржинальность там очень высокие, но проблемы

бизнеса там точно такие же как и штатах с худшими показателями. Проблема системная.

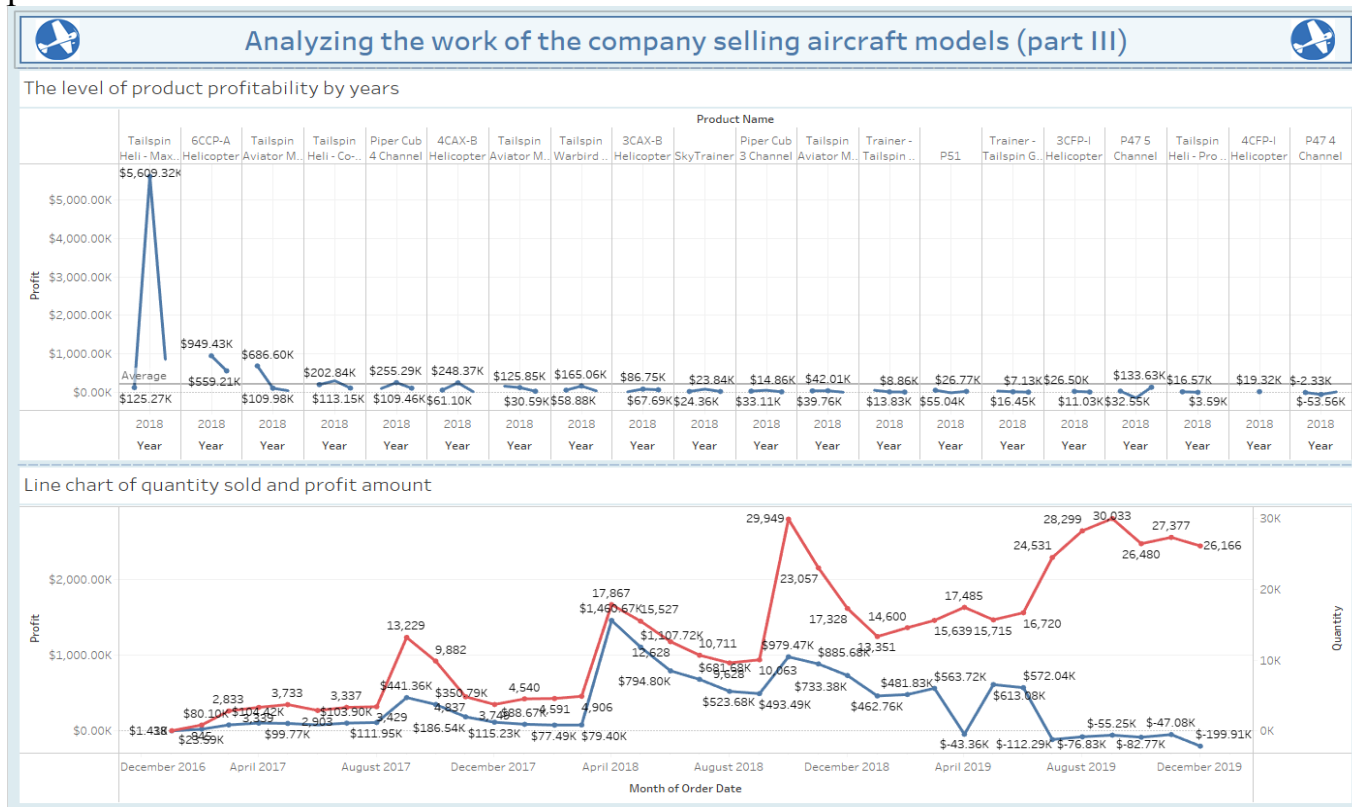
рис.6



Третий дашборд как раз и демонстрирует причину плачевных финансовых результатов. На нём видно, как сильно падает доходность номенклатурных позиций и что это происходит стабильно из периода в период. За редким исключением, и при этом данные позиции имеют ничтожные обороты.

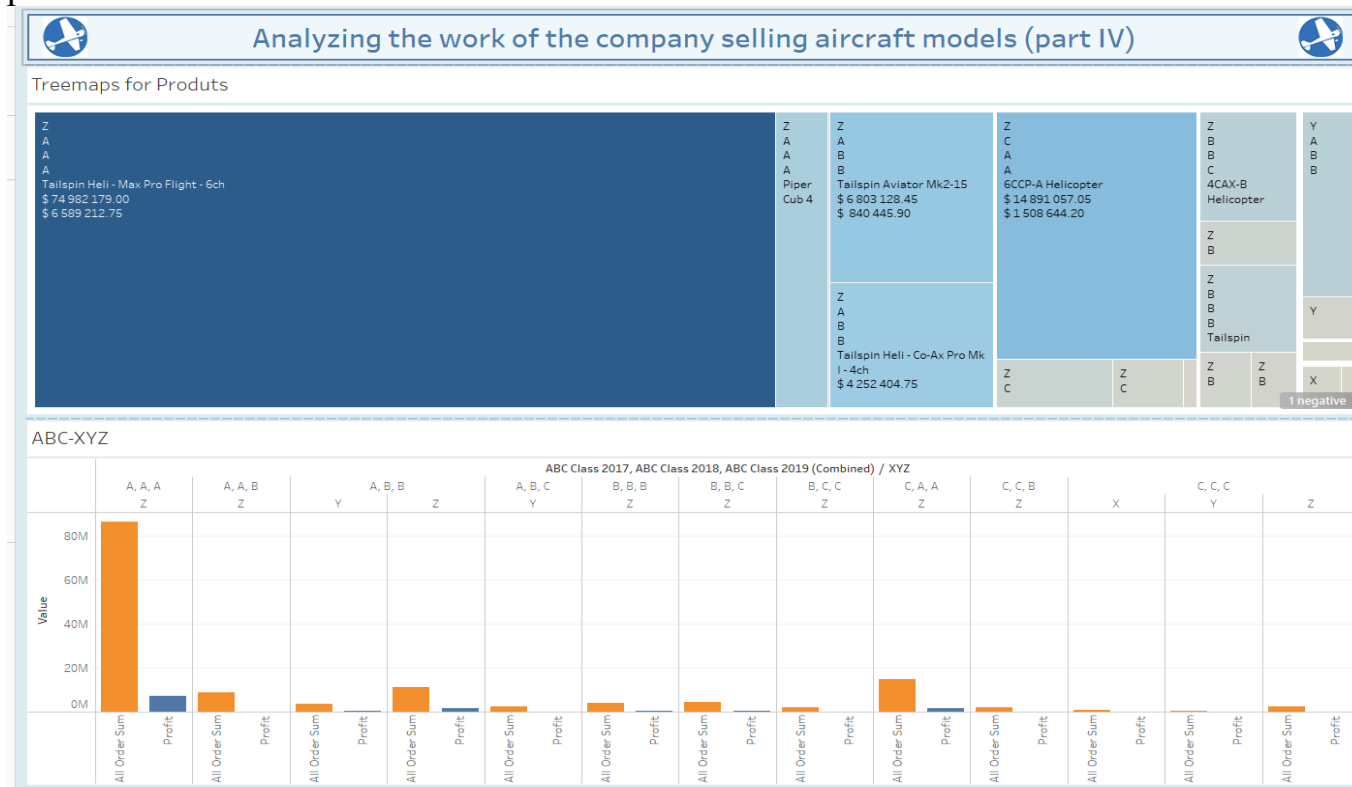
На втором графике видно, что в первоначальные периоды количество проданного товара напрямую влияло на прибыльность. Но с мая 2019 бизнес-процессы дали сбой и прибыльность стала сильно падать на фоне увеличения количества сделок.

рис.7



Заключительный дашборд показывает финансовые показатели в разрезе ABC и XYZ анализа. Опять это показывает скудность продаваемого ассортимента и большую разницу между получаемыми за заказы деньгами и прибылью от этих заказов.

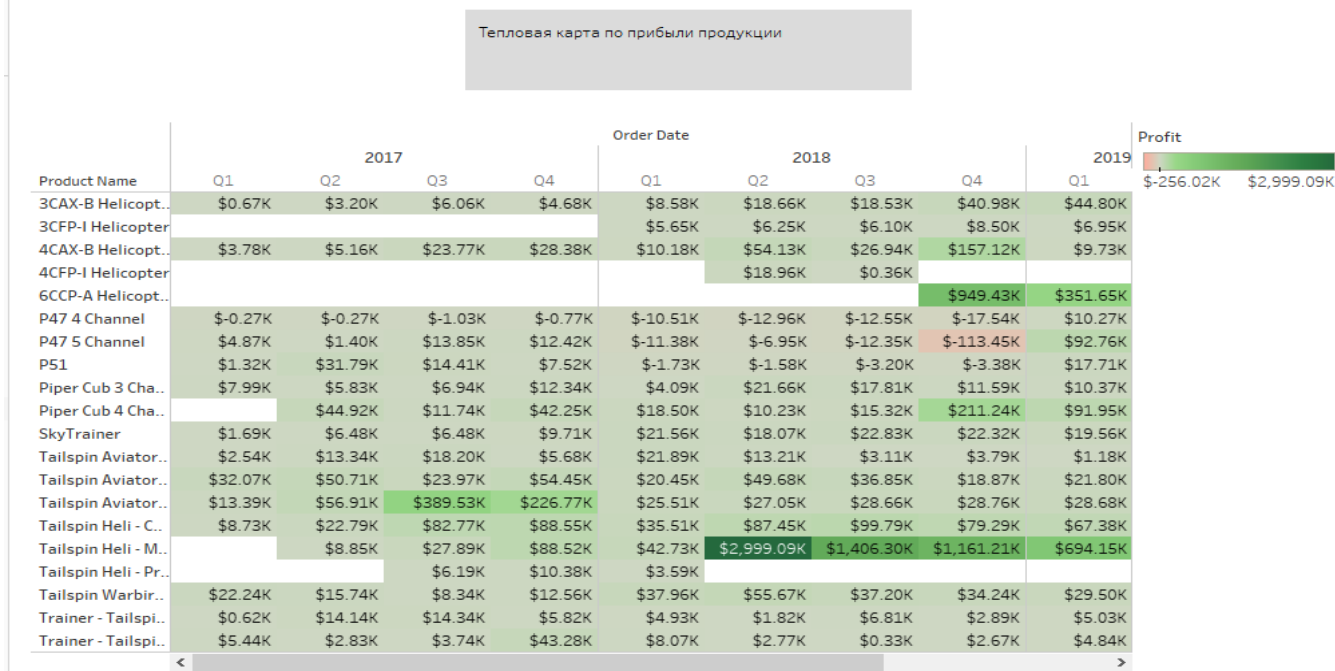
рис.8





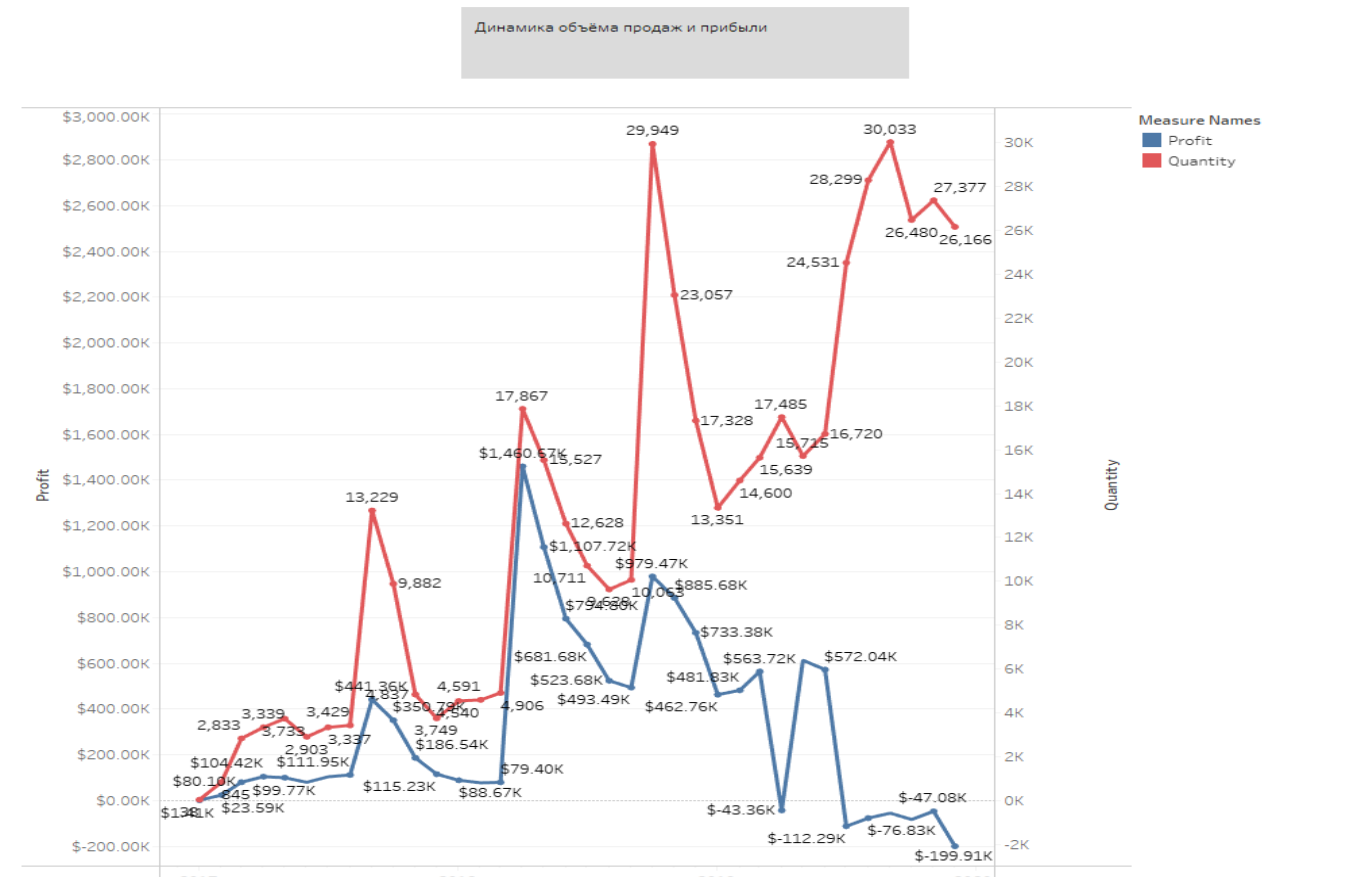
В конце было создано несколько Историй.
Тепловая карта по прибыли продукции.
рис.9

Heatmap on the profit of products

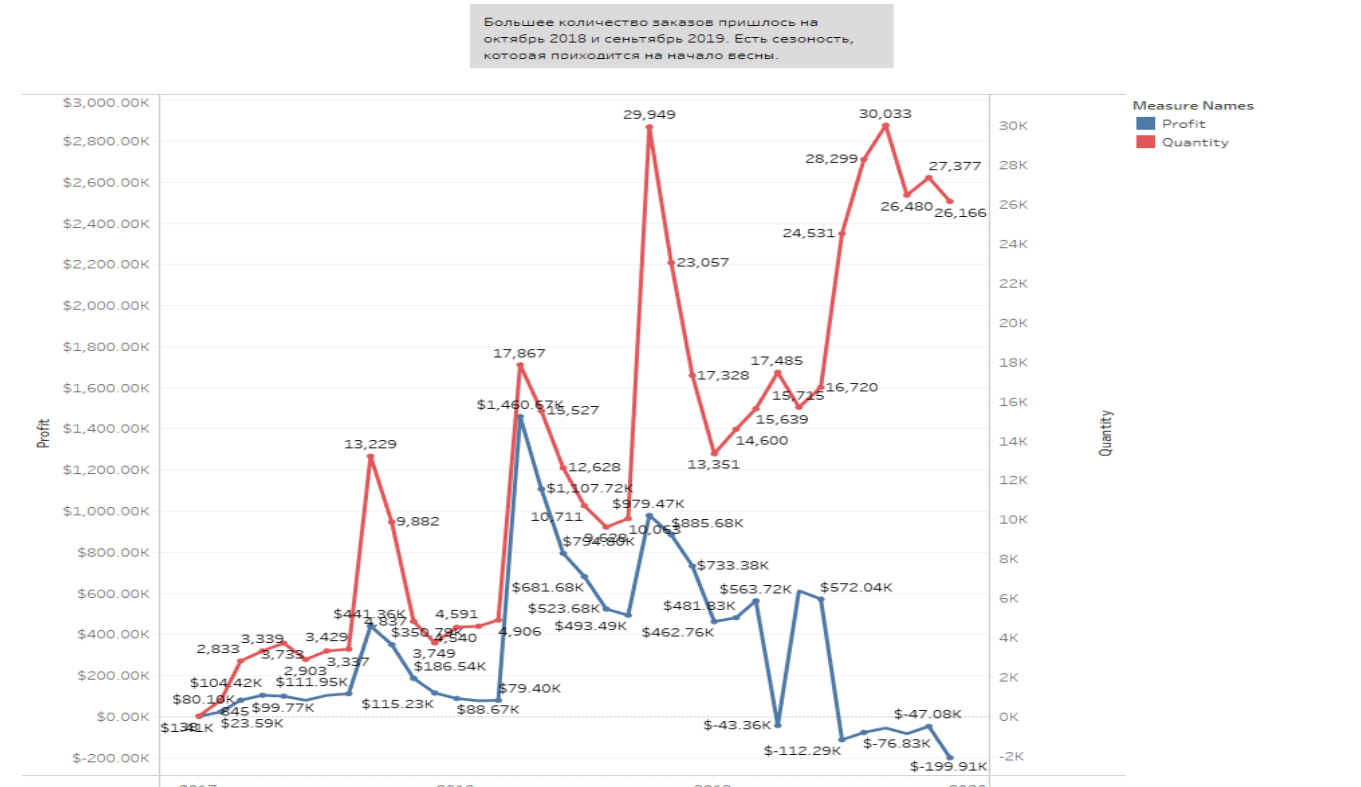


Динамика объёма продаж и прибыли.
рис.10

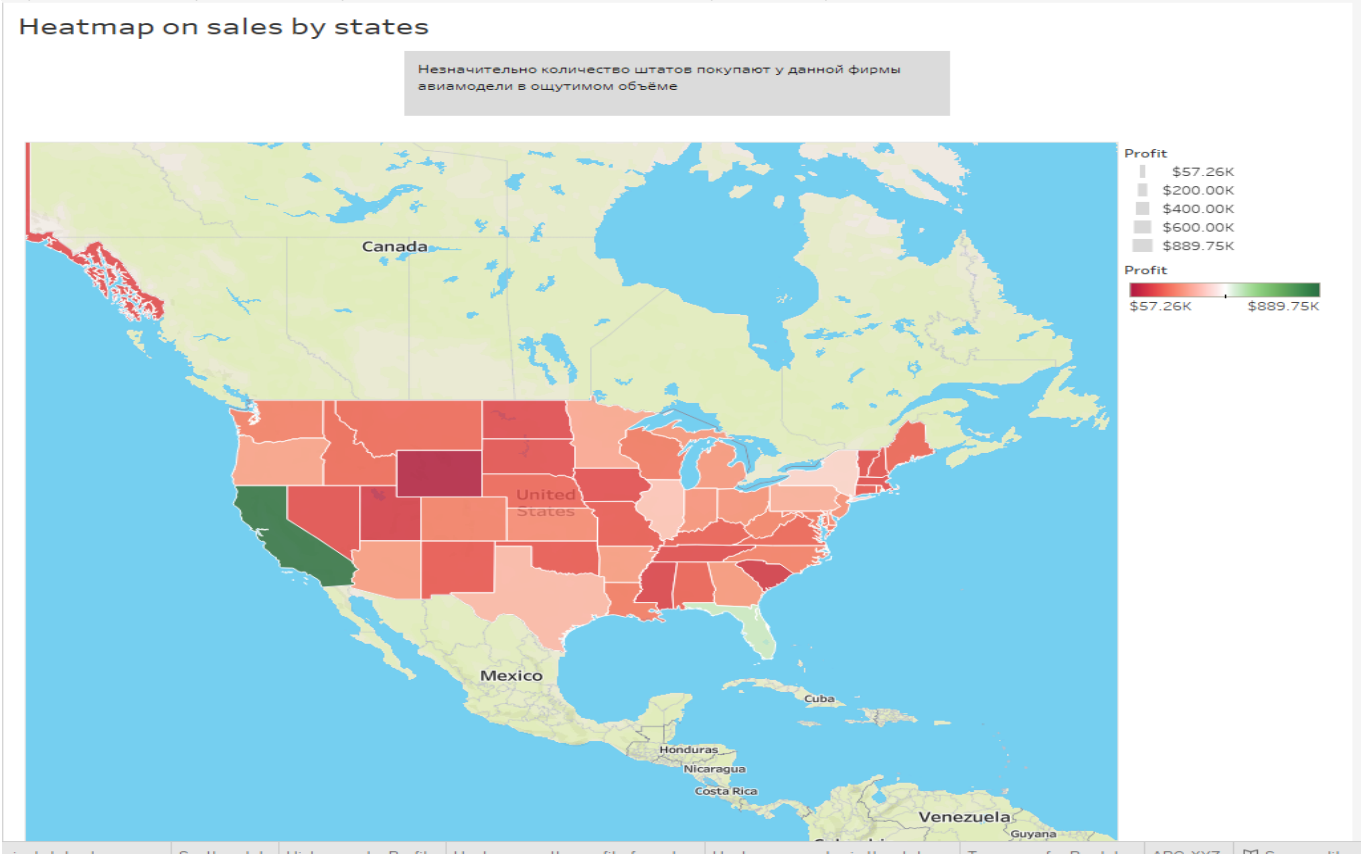
Dynamics of sales volume and profit



Сезонность.
рис.11
Seasonality



Продажи по штатам.
рис.12





Все визуализации созданные в Tableau интегрированы в **Colaboratory**.

Спасибо за внимание.