M-Fashion Data Analysis

By Nur Inna Alfianinda

Table of contents

Background

Methodology

Design

- Determining the purpose
- Determining hypothesis
- Data scope: Jan-Oct 2021

Conduct

- Data source: **Kaggle**
- Data cleaning and processing:
  - Spreadsheet
  - Python
  - Tableau

Synthesize

- Data modelling
  1. Correlation Analysis
  2. Customer Segmentation
  3. Forecasting – Nov-Dec 2021
- Summarize and interpret the results
- Recommendation

Company Overview

Jan - Oct 2021

M-Fashion is a fashion e-commerce start-up based in Australia established in 2021.

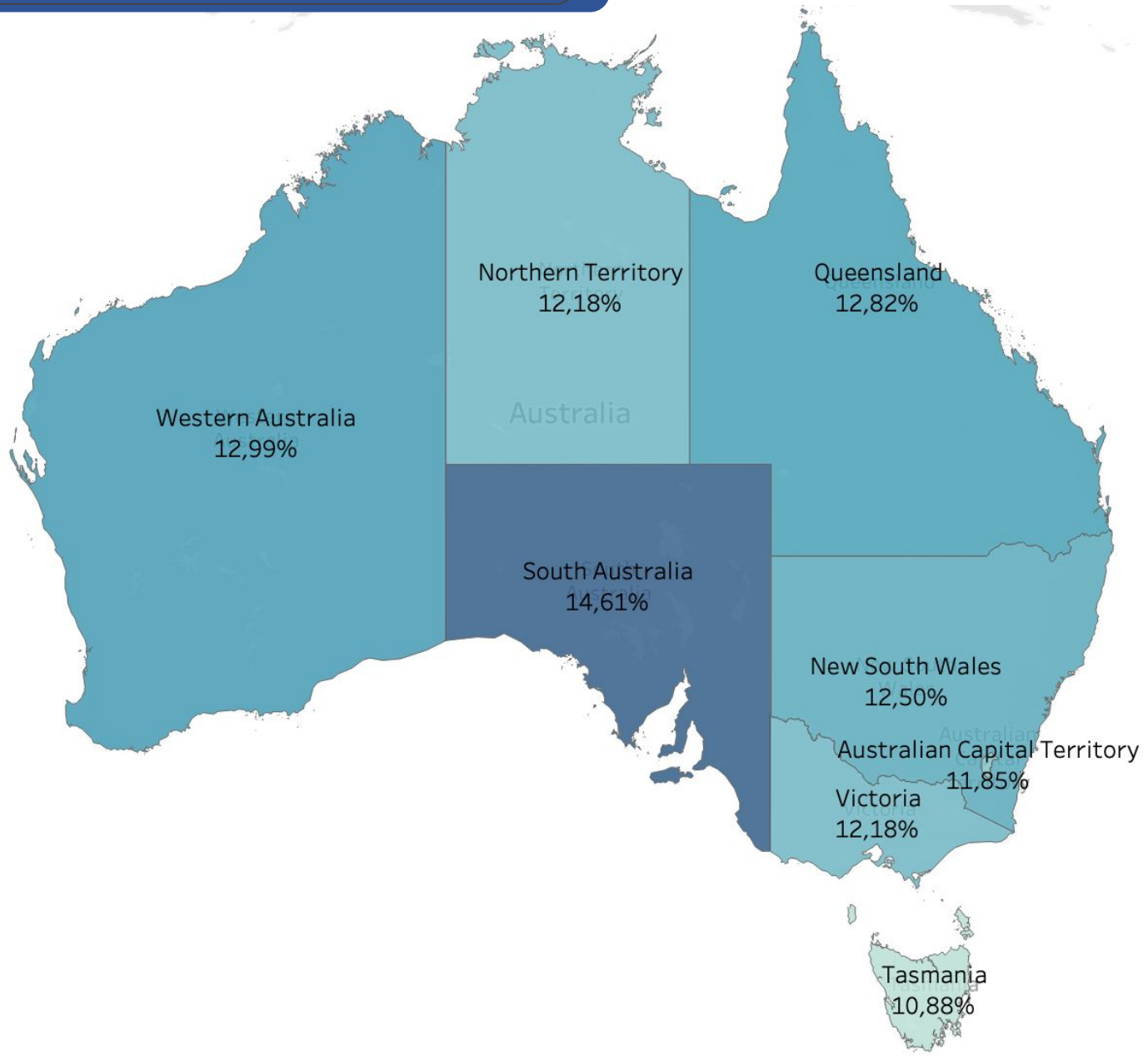
M-Fashion has three product categories: jackets, shirts, and trousers

$24,913
Total Profit

9,962
Products Sold

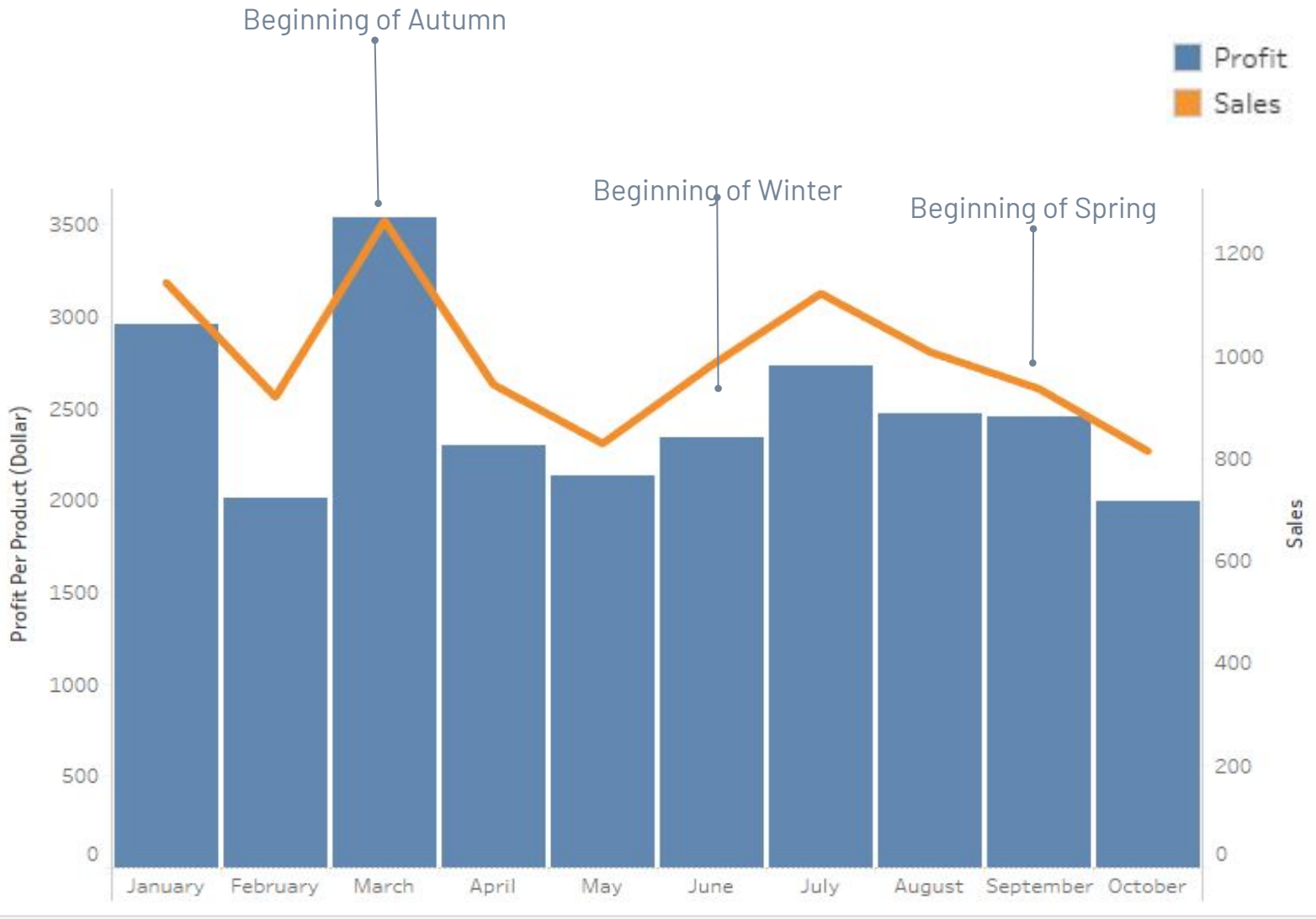
616
Total Customers

Customer Distribution Maps



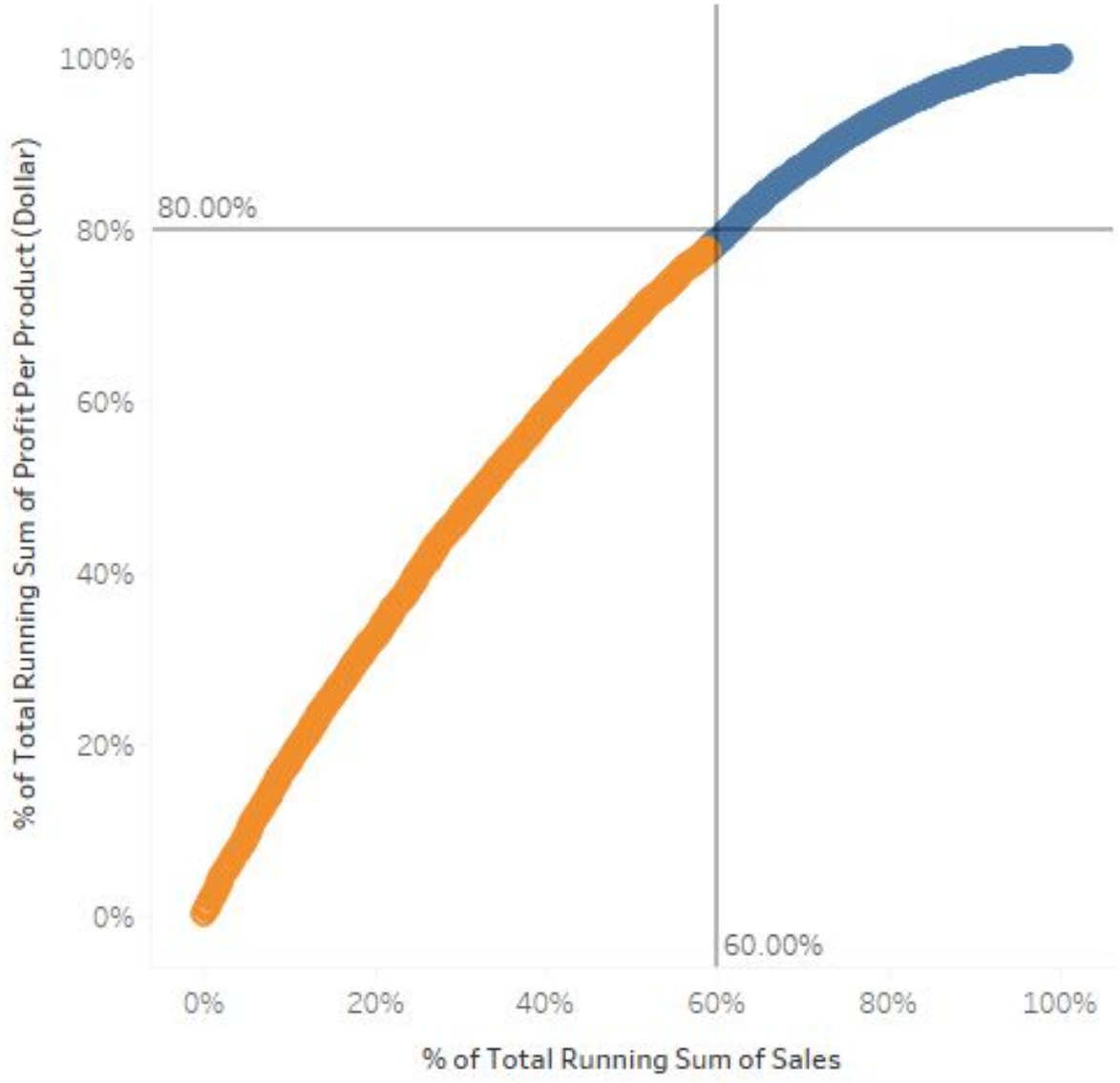
Highest proportion for customer distribution is **South Australia at 14,61%**

Sales and Profit Overtime



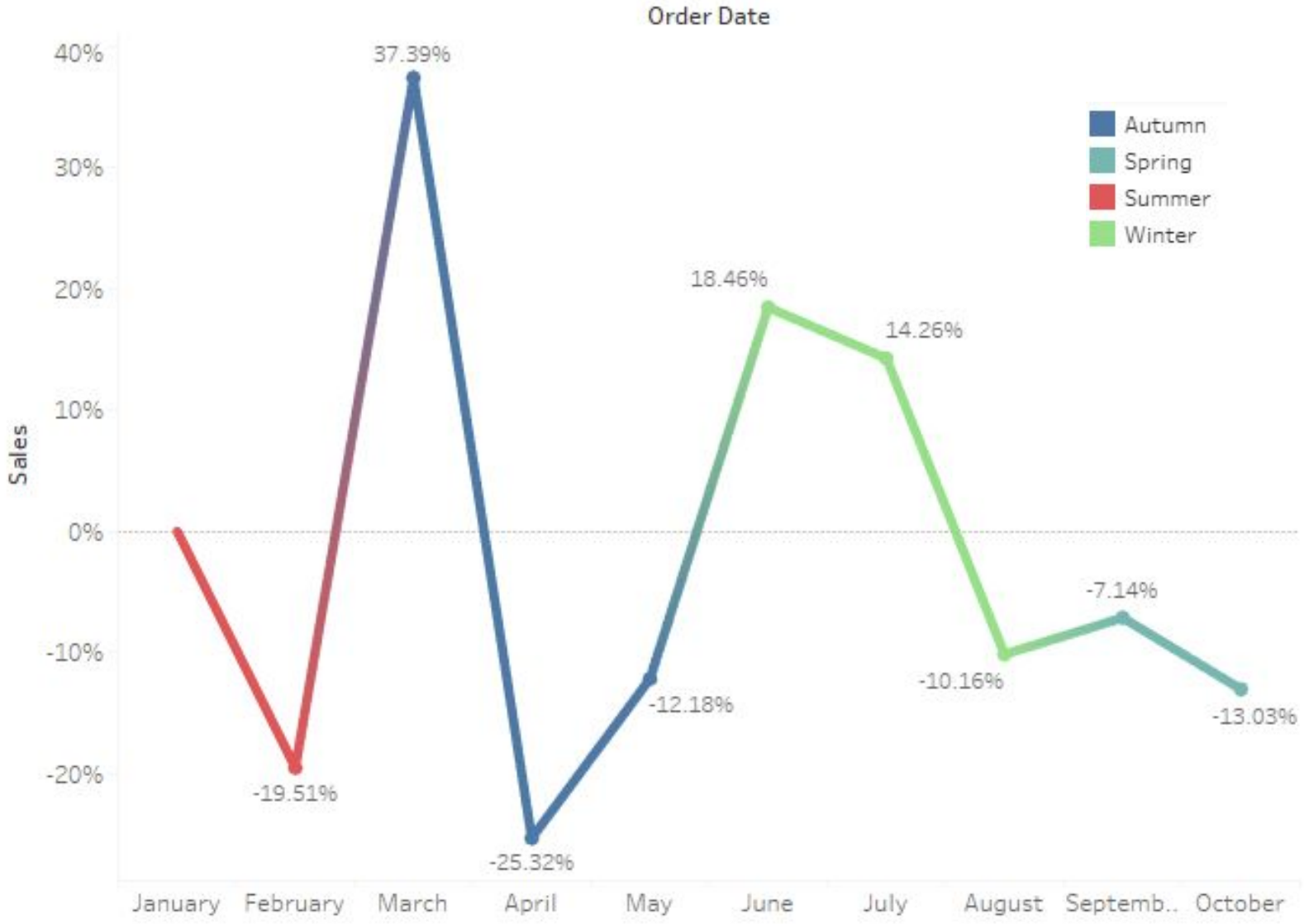
- Sales and Profit experienced a **significant increase in March and June** which is the **beginning of the Autumn and Winter season**, then decreased in the following month
- Overall sales from January to October 2021 did not experience significant movement

80 percent of profit comes from 60 percent of order



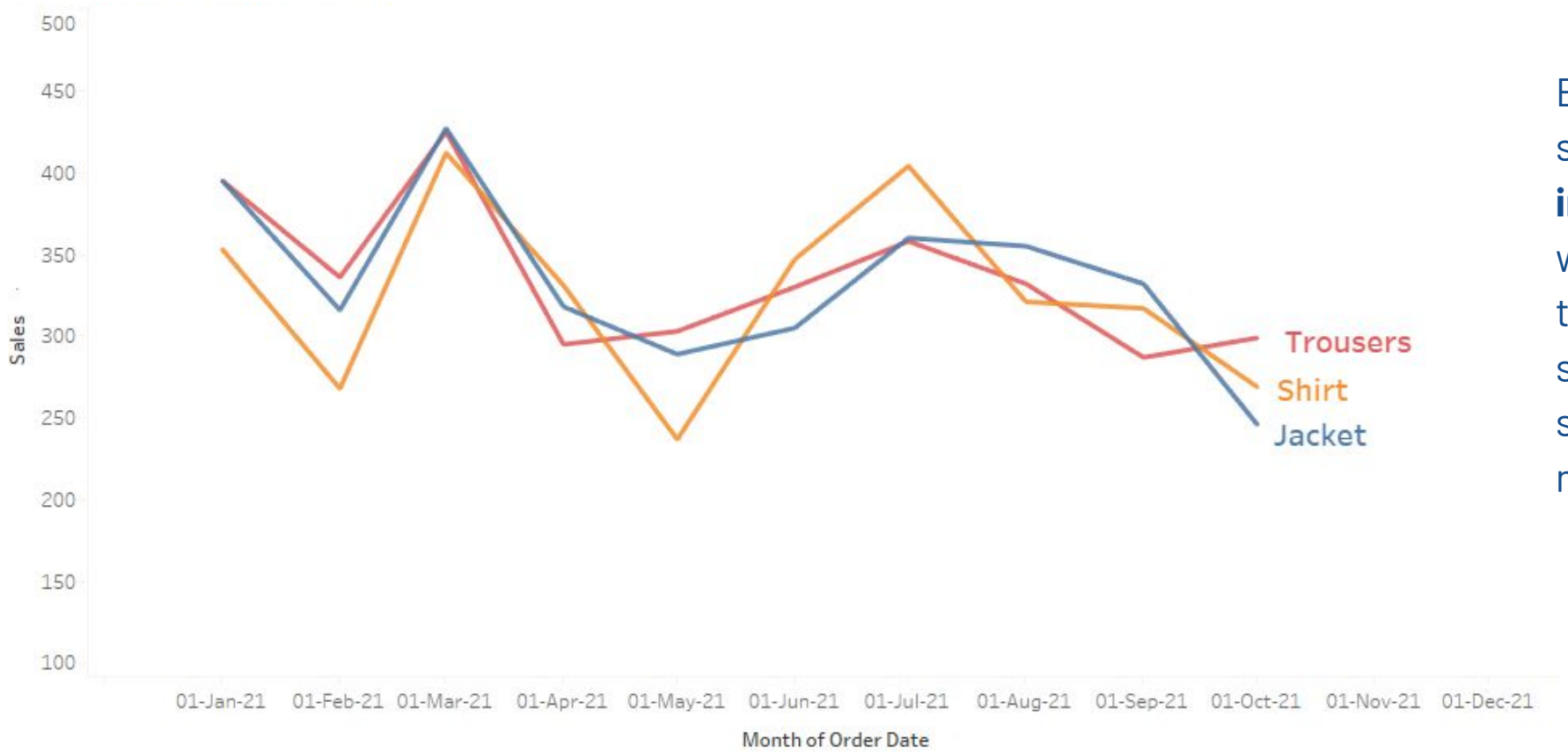
- **Majority of profit** comes from a **60% of orders**
- It requires the large number of sales contribution to earn profit

Percentage Change in Sales Overtime



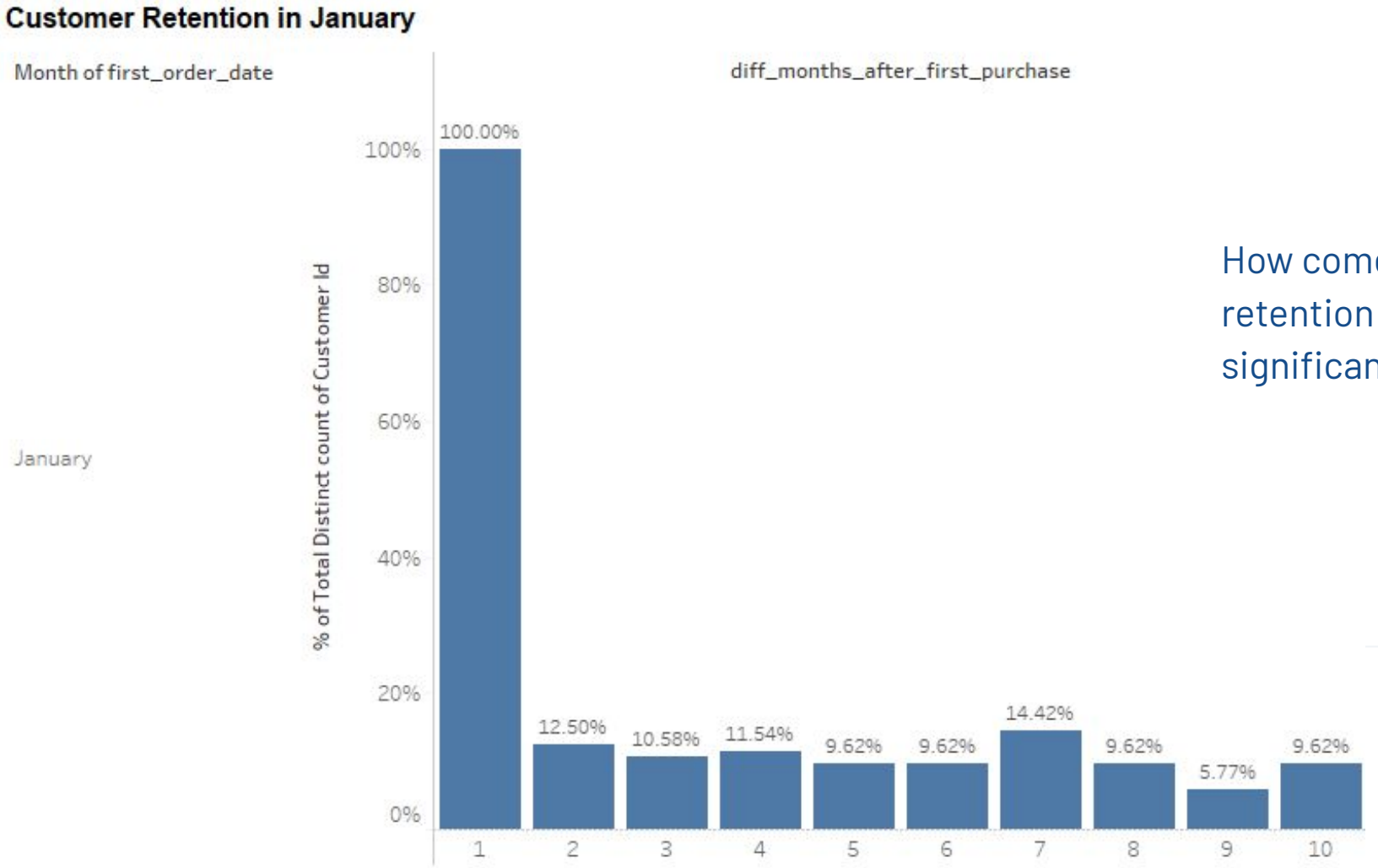
- A significant **increase 37.39%** was sales at the **beginning of the Autumn season** and then **decreased drastically -25.32%** in the **following month**
- Overall, the **sales trend decreased by 28.78%**

Sales of Each Product per Month



Each product experienced a significant **increase in sales in March, June, and July** which is the beginning of the **Autumn and Winter** season, then decreased significantly in the following month.

The percentage of customer retention after the first month of placing an order is quite low (January)



How come customer retention drop significantly?

Project Goals

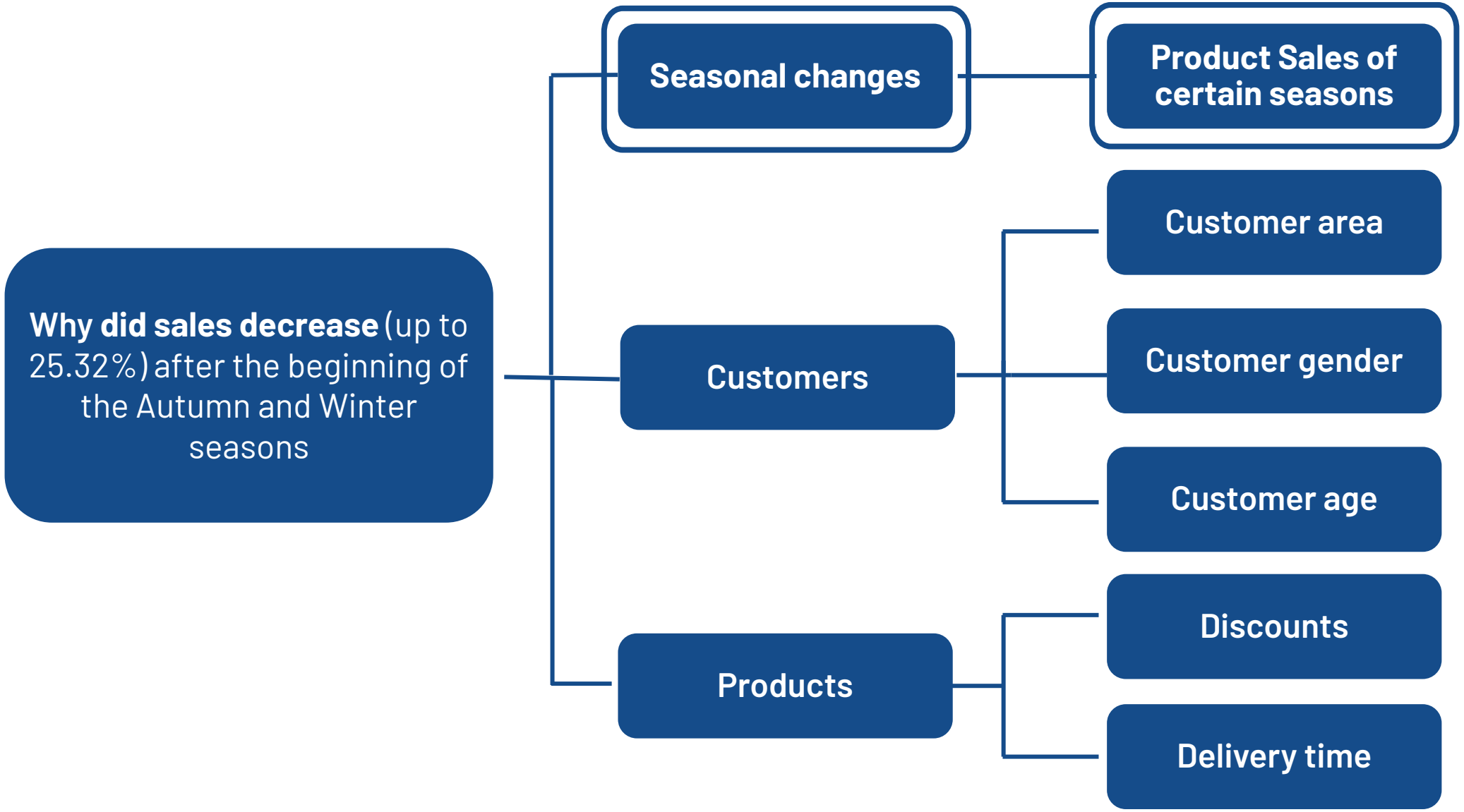
Problem Statement

1. There is a fluctuation in sales with a decrease (up to 25.32%) after the beginning of the Autumn and Winter seasons
2. Overall sales trend January-October 2021 decreased by 28.78%

Objectives

1. How to prevent the decline in product sales (up to 25.32%) after the beginning of the Autumn and Winter seasons?
2. How to increase product sales by at least 28.78% in 2022?

Root Cause Analysis

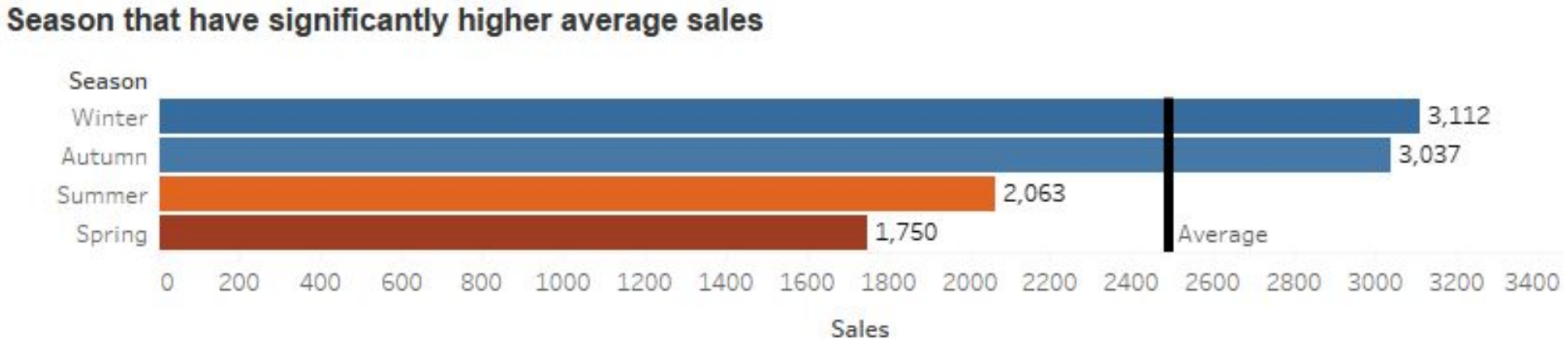


Modelling Results

- Correlation Analysis
- Customer Segmentation
- Forecasting
- Recommendation

1. **There is a fluctuation in sales with a decrease (up to 25.32%) after the beginning of the Autumn and Winter seasons**

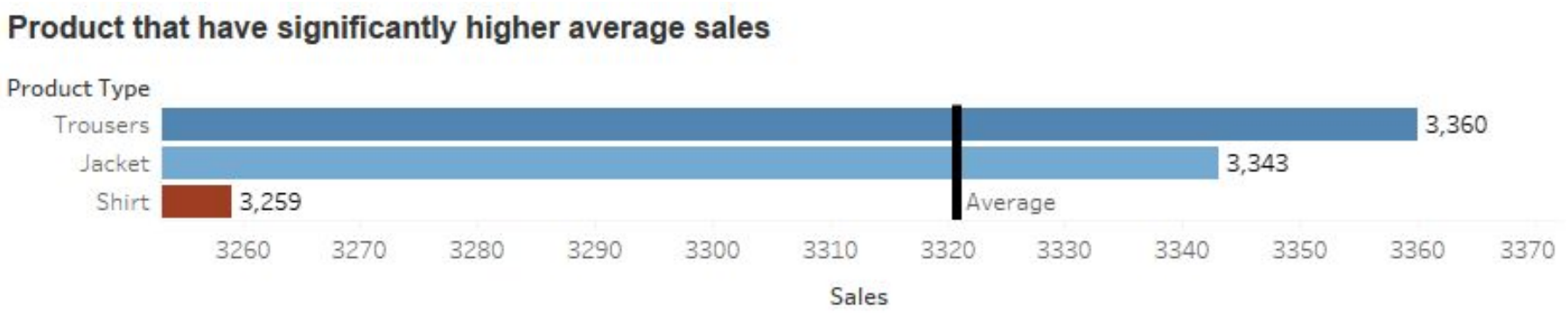
There is correlation between changing seasons and total sales



p-value < 0.05
ANOVA test

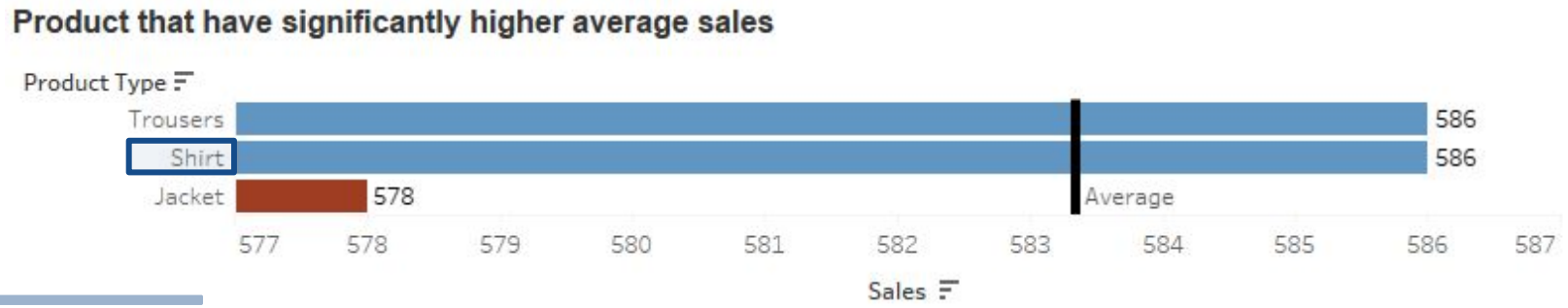
- More people make transactions above the average sales in the winter and autumn seasons
- Transactions in spring and summer have below average sales
- **Need a strategy and review of sales in the spring and summer to improve customer retention**

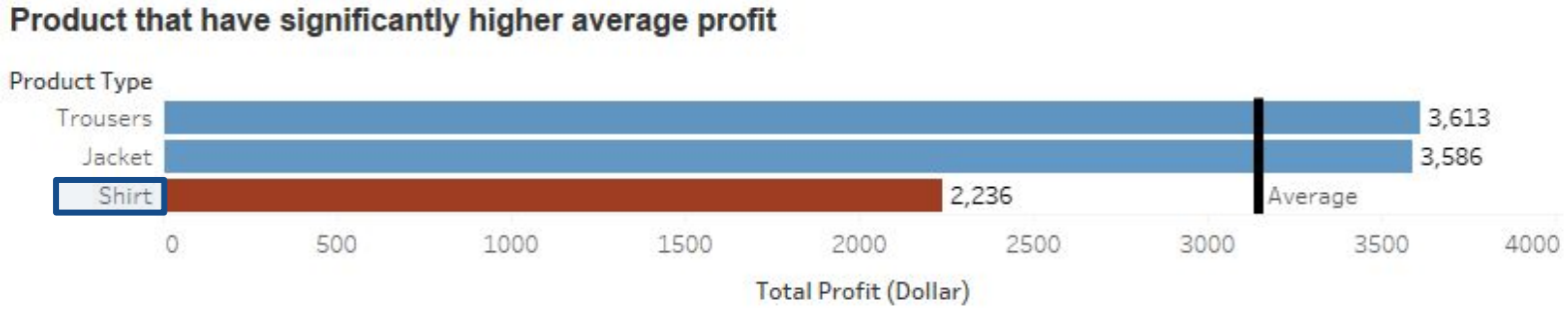
Trousers have the highest average sales and shirts are below average sales in 2021



- More people make transactions above the average sales for trousers and jackets
- While shirt sales are below average
- **Need a strategy and review of shirt products to increase sales, whether it's from the quality or the model**

Sales anomaly in spring: shirt has the highest average sales, but the profit is quite small

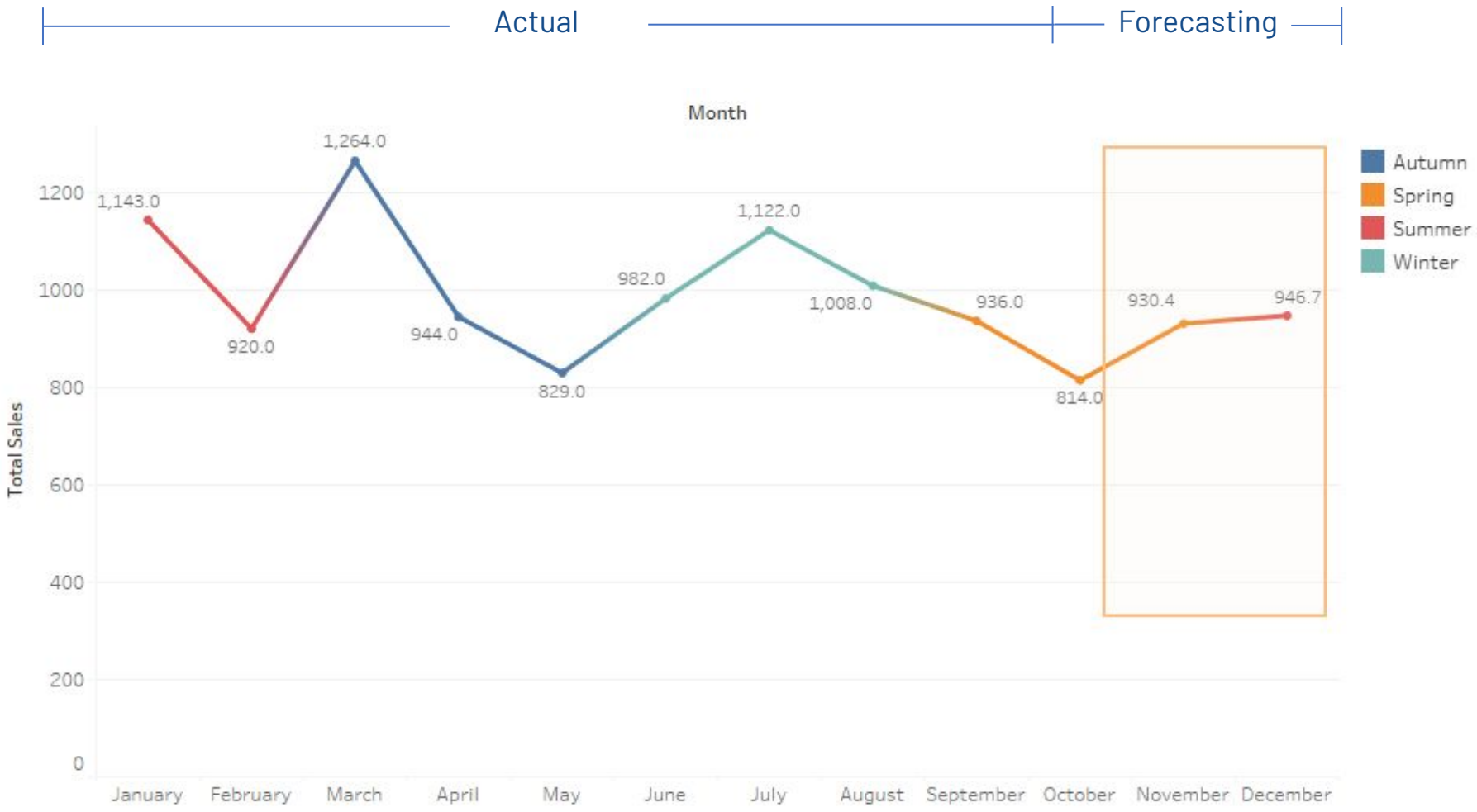




- The company held promos in spring, such as buy two get one free, so that sales increased but the profit was quite small
- **Need a review of the promo management**
- **Need a strategy and review of shirt**

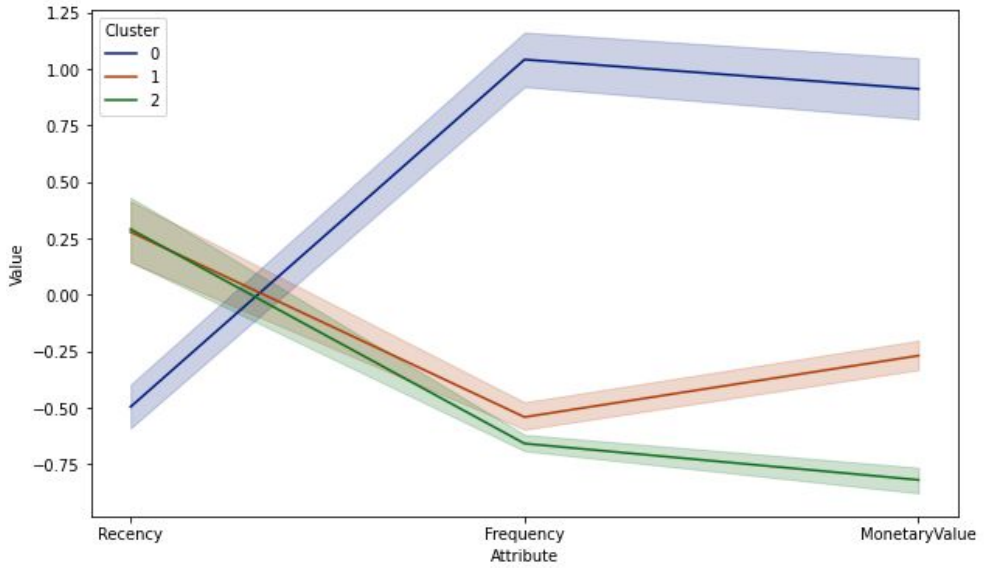
2. Overall sales trend January-October 2021 decreased by 28.78%

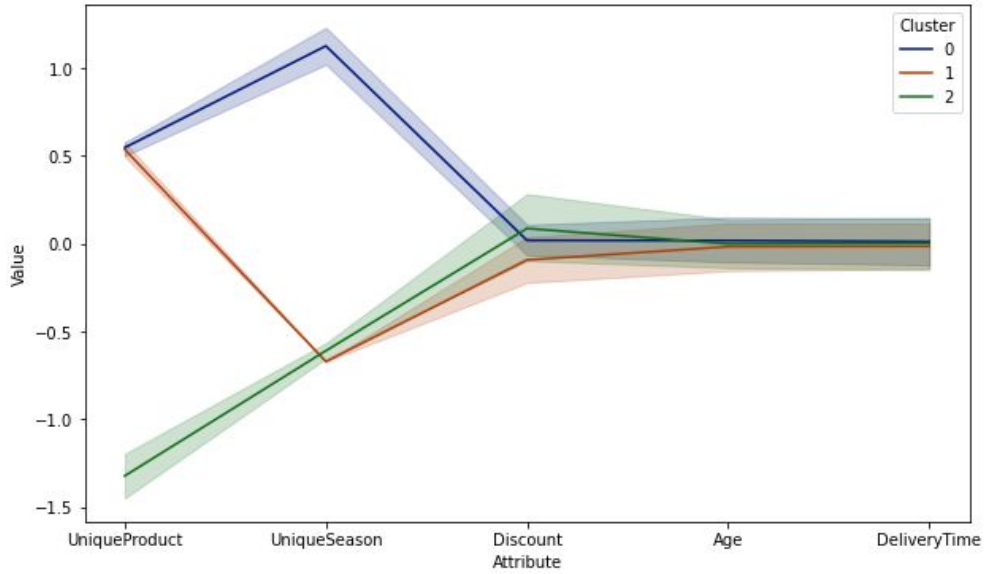
Forecasting



Sales increase from October until December 2021

Correlation & Customer Segmentation





Product discounts, customer age, and delivery time have no effect on sales

Cluster 0 – Loyal Customers (43,02%)
Customers buy products recently, frequently, vary, highest spending, and do not depend on the season.

Cluster 1 –Need-based Customers (35,39%)
Customers buy the product at the old time, quite often, vary, moderate spending, and depend on the season.

Cluster 2 – Churn Customers (21,59%)
Customers buy the product at the old time, rarely, do not vary, lowest spending, and depend on the season.

Recommendations

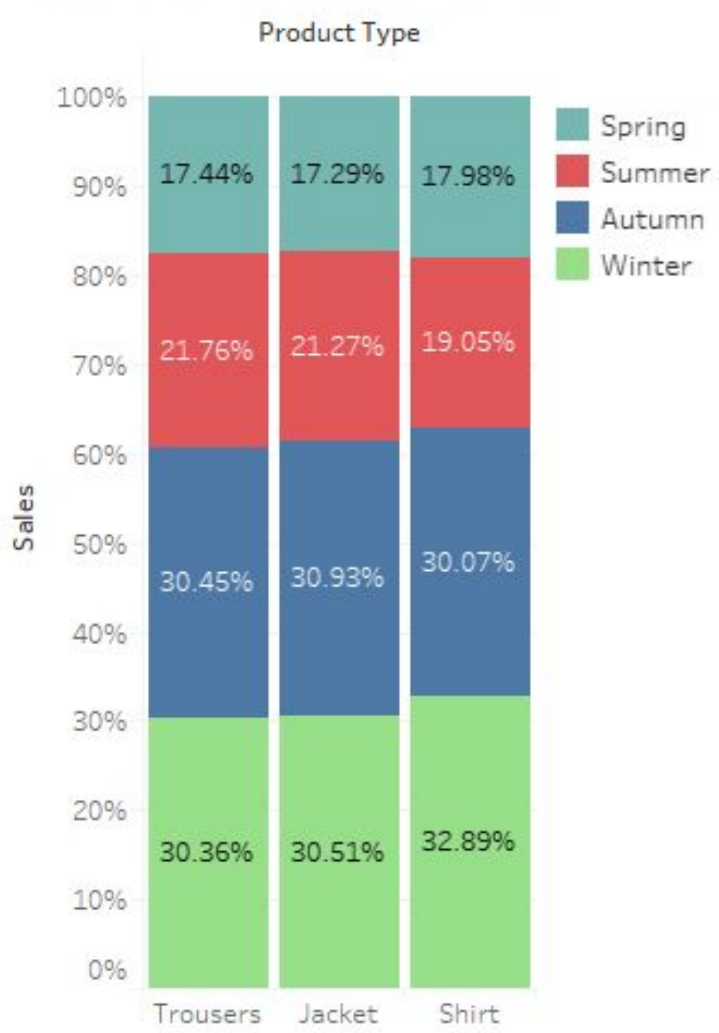
Loyal Customers	• System algorithms: showing more various complementary product that don't depend on season
Need-based Customers	• Product recommendation ads on social media based on searching history • Displays the top ten similar products/brands in the application
Churn Customers	• Give special vouchers with a specific validity period • New product notification via email or app

	Urgent	Not Urgent
Important	**Product review (model and quality)**	Gain customer (advertising, add product variations)
Not Important	Promo management	Feedback improvement service

Thank You

Appendix

Sales of Each Product per Season



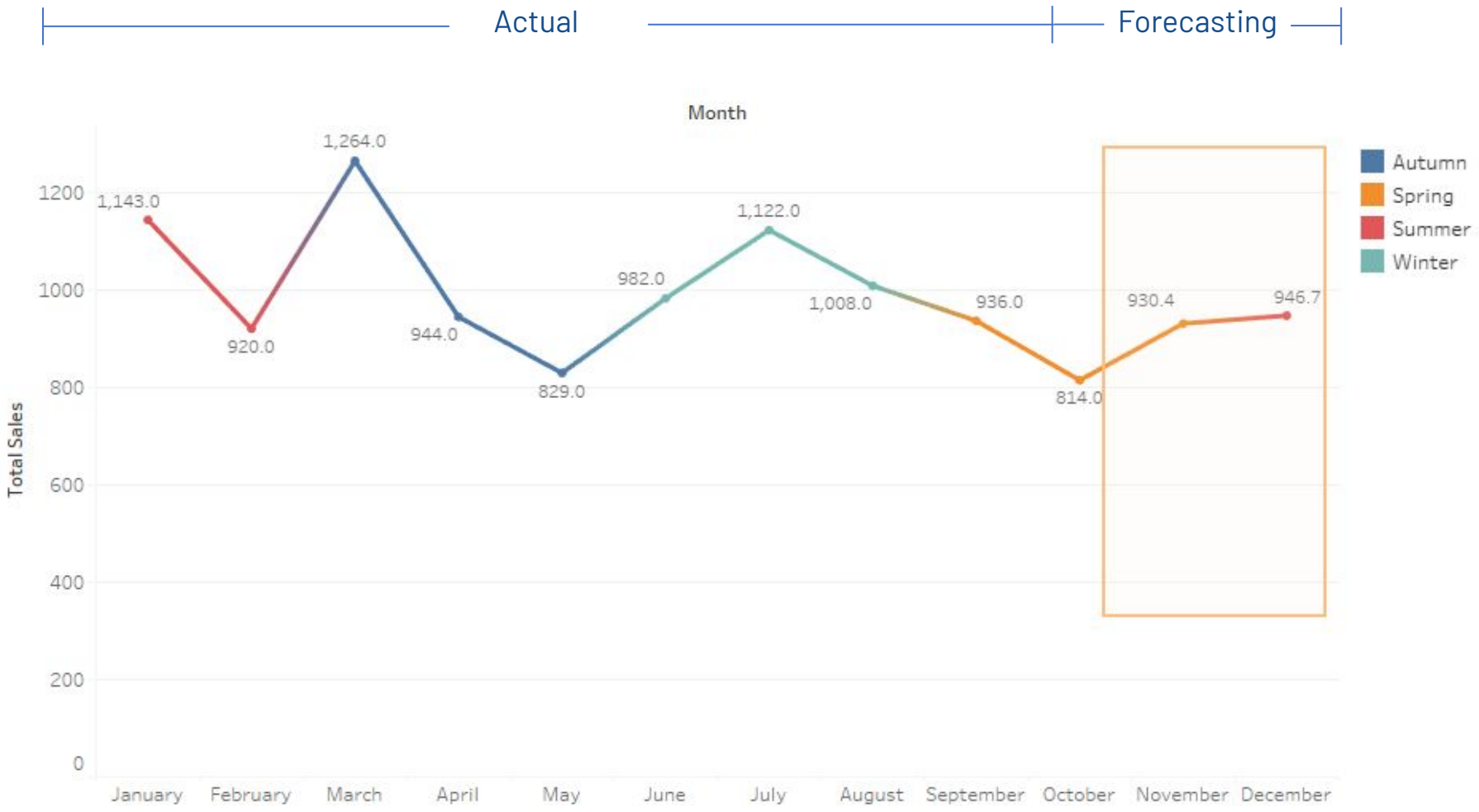
- There is no significant difference in the number of sales ratio based on seasons per product
- All products have a high number of sales in Winter and Autumn
- All products also have low sales in Summer and Spring

Cohort Analysis

Customer Retention

	diff_months_after_first_purchase									
Month of first_order_date	1	2	3	4	5	6	7	8	9	10
January	100.00%	12.50%	10.58%	11.54%	9.62%	9.62%	14.42%	9.62%	5.77%	9.62%
February	100.00%	16.88%	11.69%	14.29%	6.49%	12.99%	7.79%	7.79%	10.39%	
March	100.00%	5.56%	5.56%	12.22%	13.33%	13.33%	12.22%	12.22%		
April	100.00%	5.88%	11.76%	11.76%	8.82%	7.35%	8.82%			
May	100.00%	14.58%	2.08%	2.08%	12.50%	6.25%				
June	100.00%	14.04%	1.75%	10.53%	7.02%					
July	100.00%	15.22%	8.70%	15.22%						
August	100.00%	6.90%	3.45%							
September	100.00%	6.82%								
October	100.00%									

Forecasting



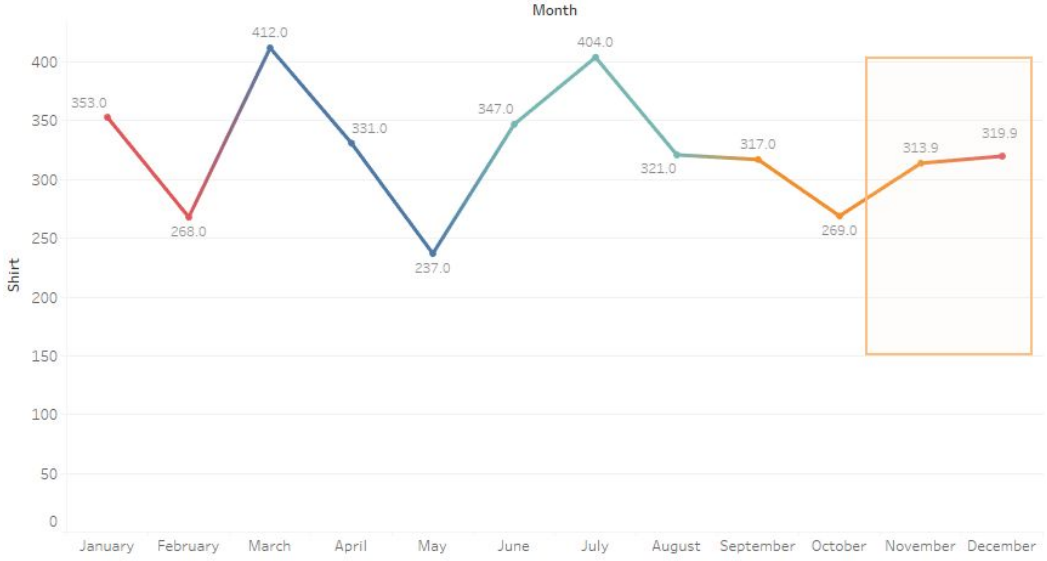
r = - 0.44
p-value = 0.15
pearson test

- The correlation of sales of all products every month has decreased, as can be seen from the **negative r** value.
- However, the p-value > 0.05 indicates that although there was a decrease in sales, the decline was not significant or sales in 2021 were **stagnant**

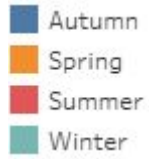
Forecasting Per Product Type

Jackets

Shirts



Trousers



Data Modelling

Google Colab