



# A Retail Analysis

Ikram Mohamad



# Introduction

## Background :

- Customer transactions dataset
- Online unique all-occasion gift ware retailer

## Data :

- 525461 observations
- Invoice dates ranged from 2009-12-01 to 2010-12-09

## Goal :

- Identify types of customers
- Predict future orders

# Agenda

Data Analysis

---

Customer Types

---

Forecasting

---

Q & A

---

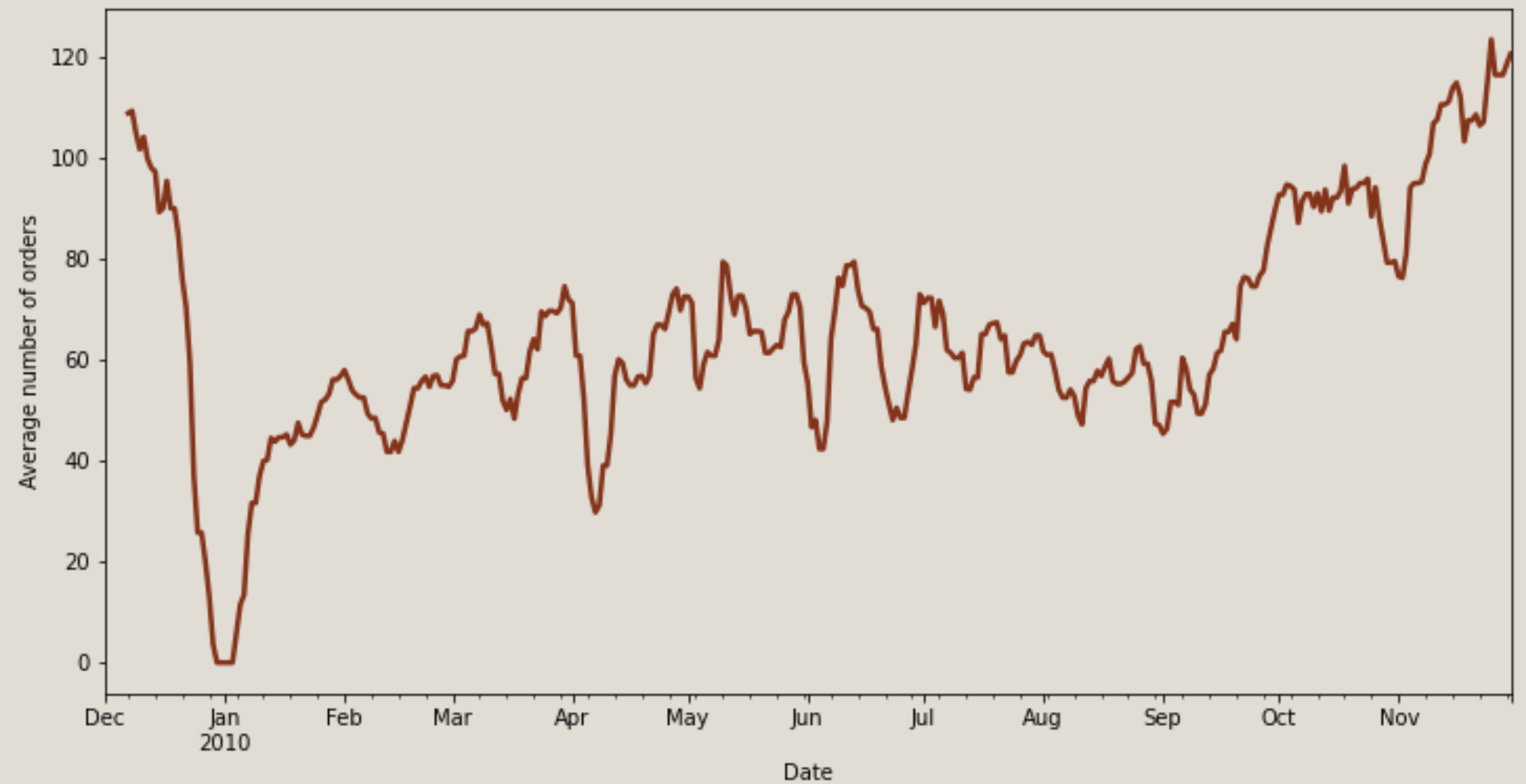
# Data Analysis

---

# The data

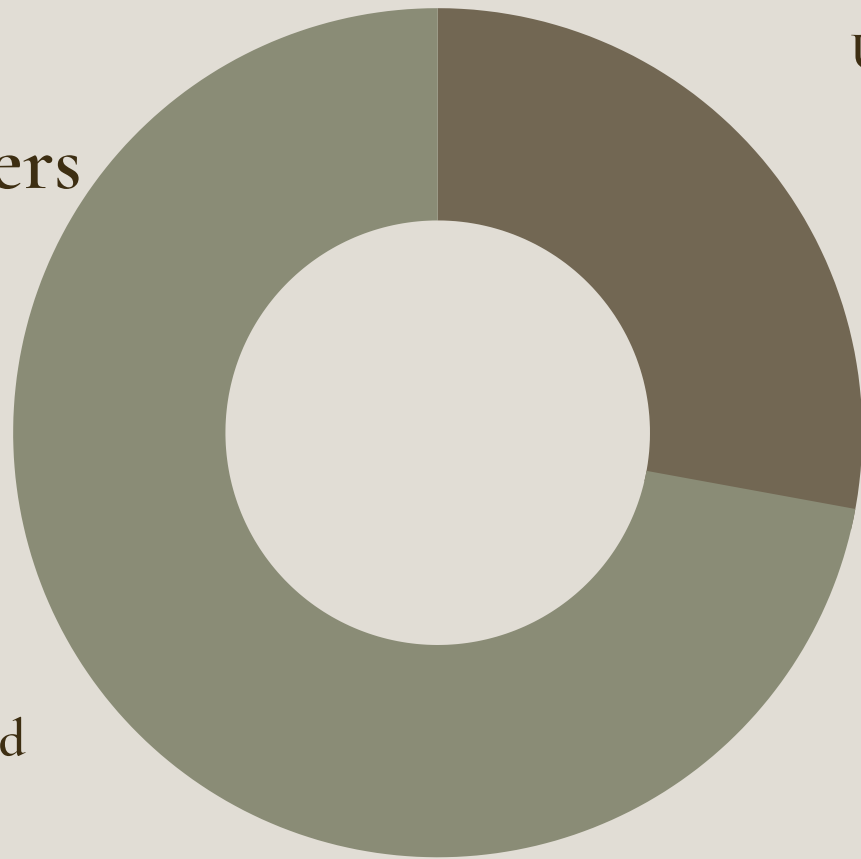
## Columns :

- Invoice
- StockCode
- Description
- Quantity
- InvoiceDate
- Price
- CustomerID
- Country



- The retailer is closed on Saturdays, except for the first Saturday of December
- Included returns, postage costs and admin updates
- Customers from 39 countries

Total  
Customers



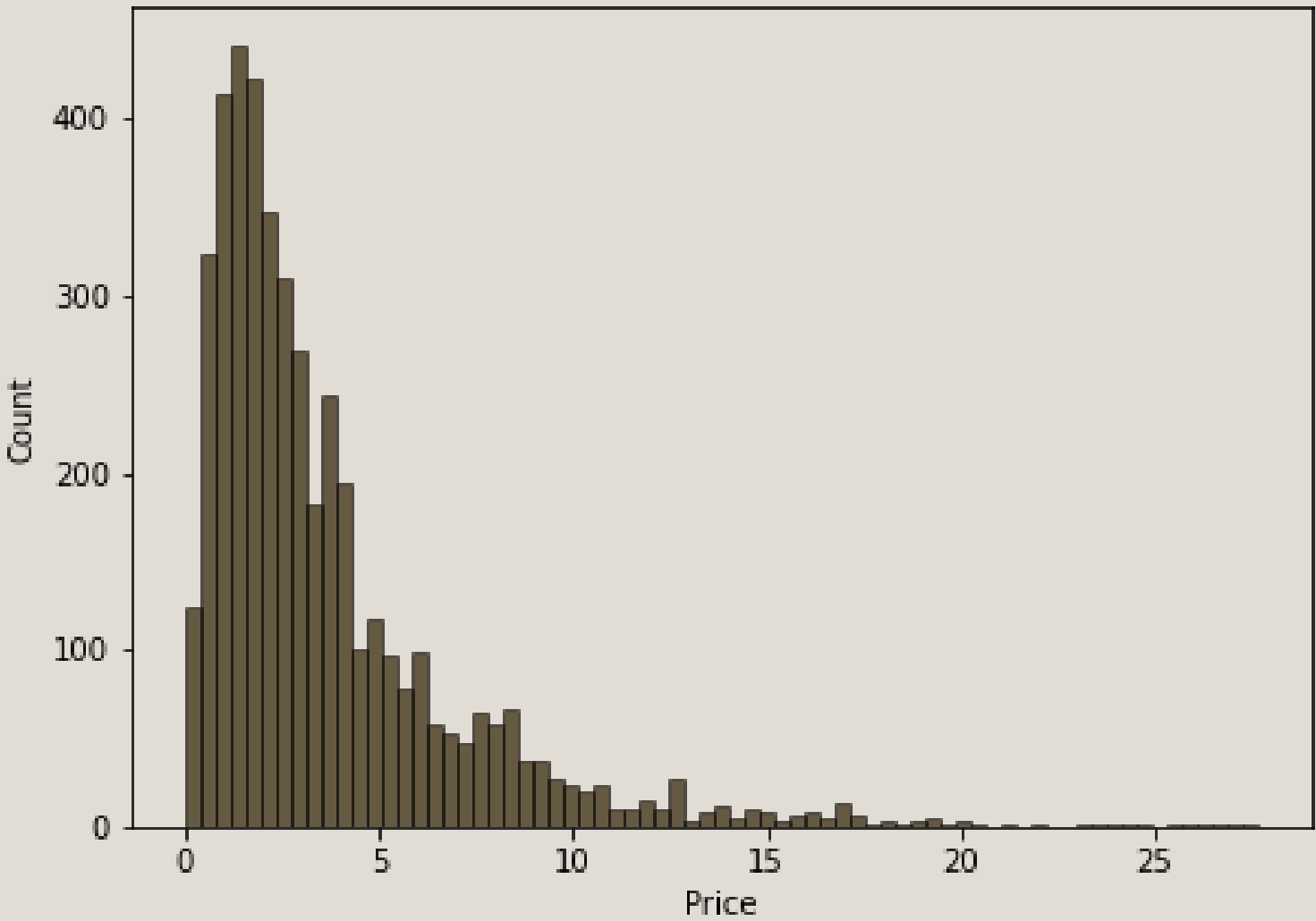
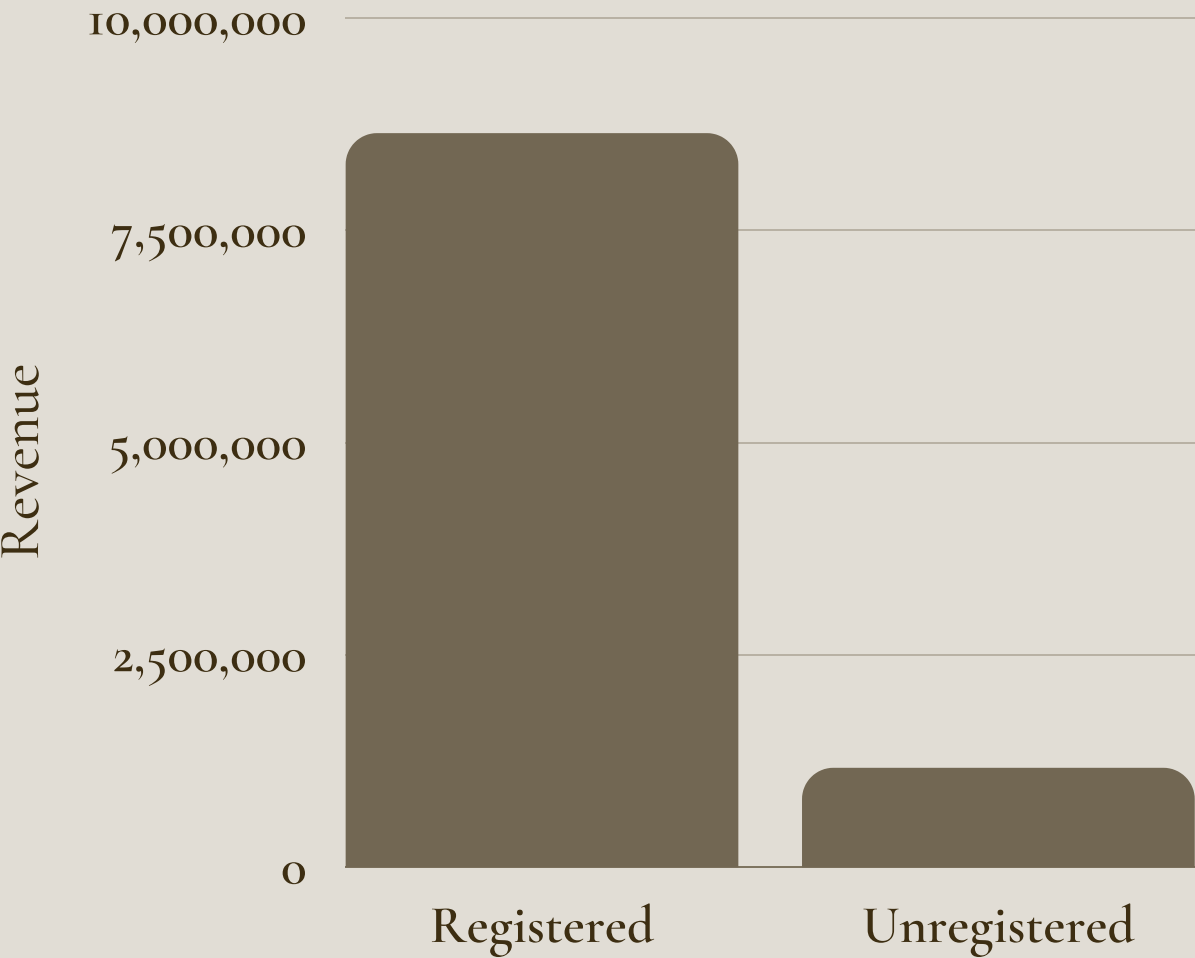
Unregistered  
27.9%

Registered  
72.1%

# Customers & Products

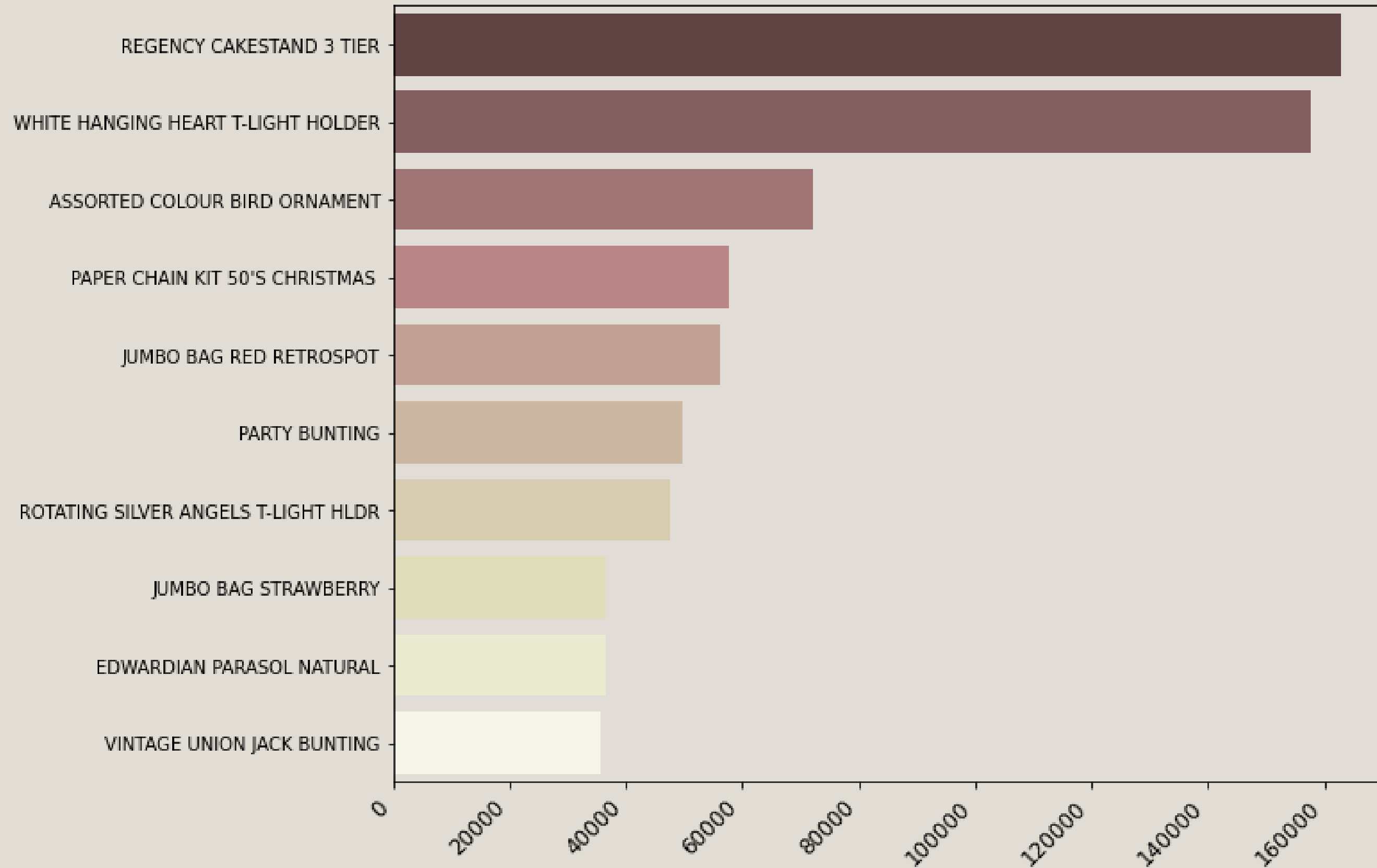
- The largest orders each consisted of 87167 items across 45 products, purchased by customers 14277 and 13687, at a cost of £11,880.84
- There are 4533 unique products
- 99% of products priced less than £27.67

Revenue by Registration



# Best Selling Products

---

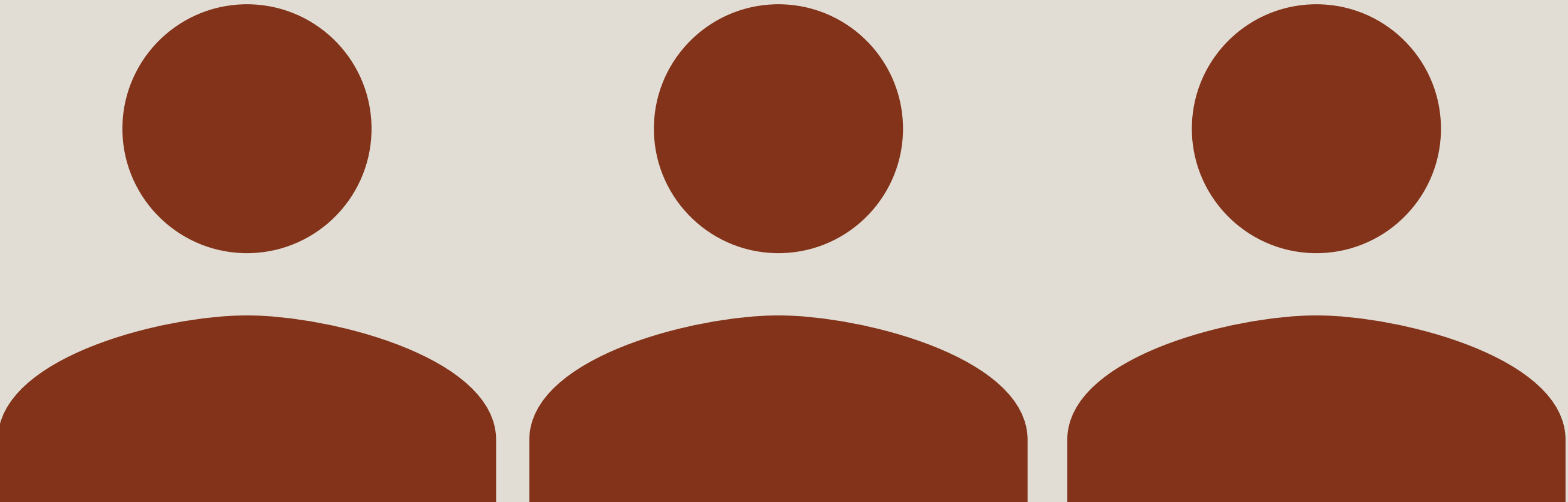


---

# Customer Types

---





# Customer Profiles

The Regular Customer

2741 customers  
6.3 orders  
£2938  
82.12%

The Unloyal Customer

1544 customers  
1.1 orders  
£381  
6%

The One-off Customer

1656 customers  
1 order  
£703  
11.88%

Cluster

Number of customers  
Number of orders  
Yearly spend  
Revenue(%)

Preparation

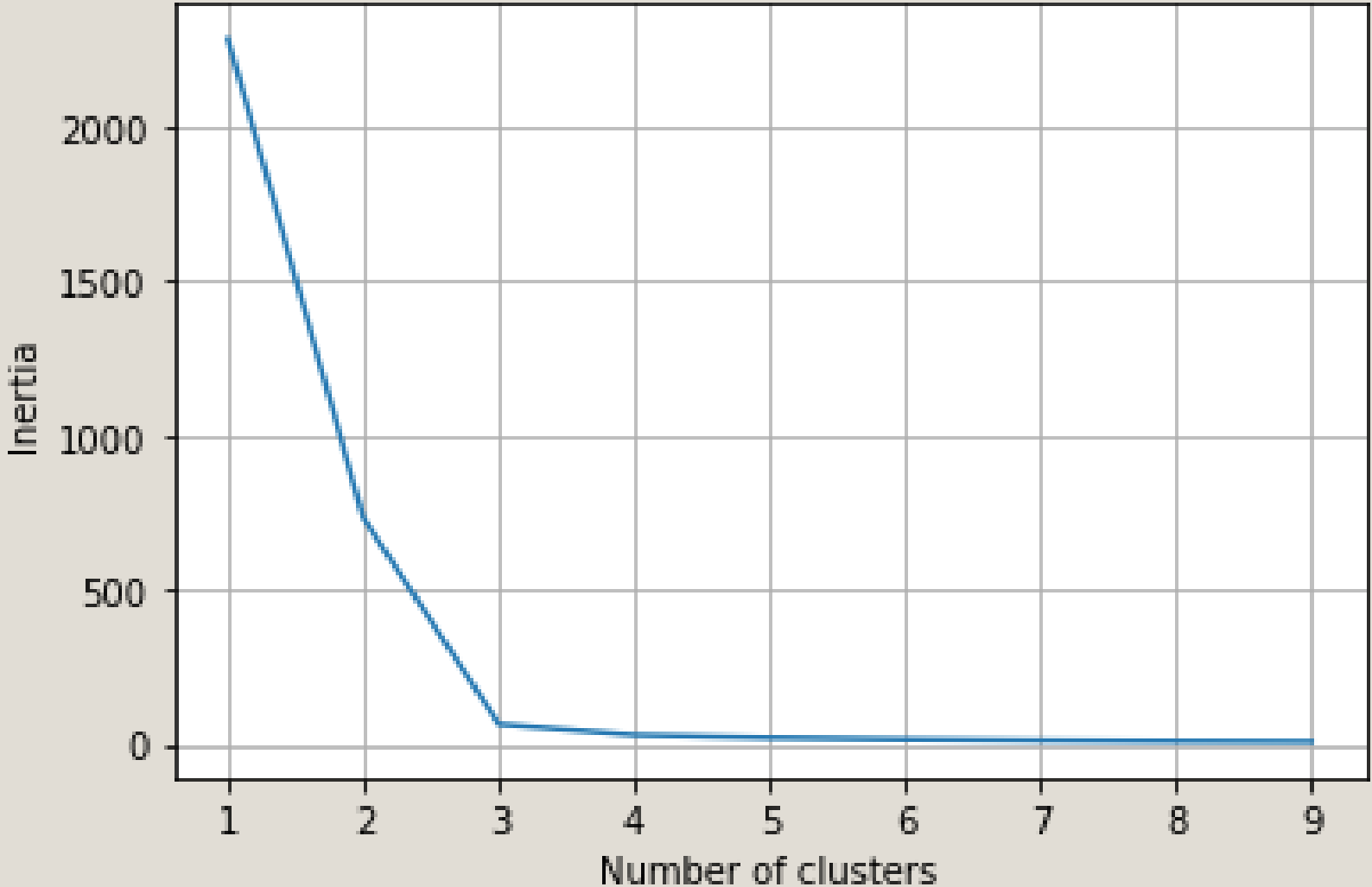
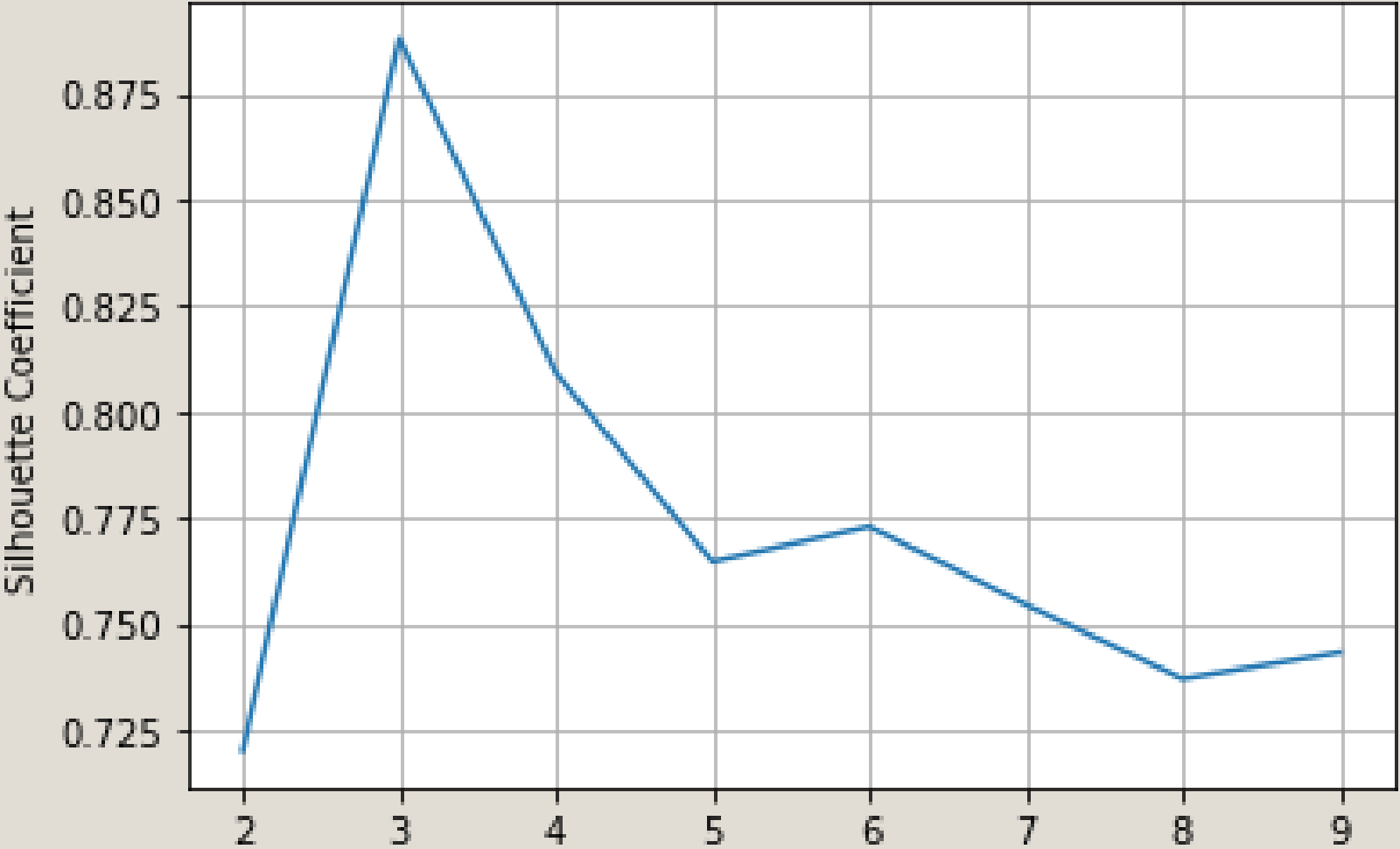
Dropped cancelled orders  
Single Customer View  
Used MinMax Scaler

Features

Revenue  
InvoiceCount  
AverageDuration  
Registered

Model

Kmeans  
3 clusters  
Silhouette score = 0.893



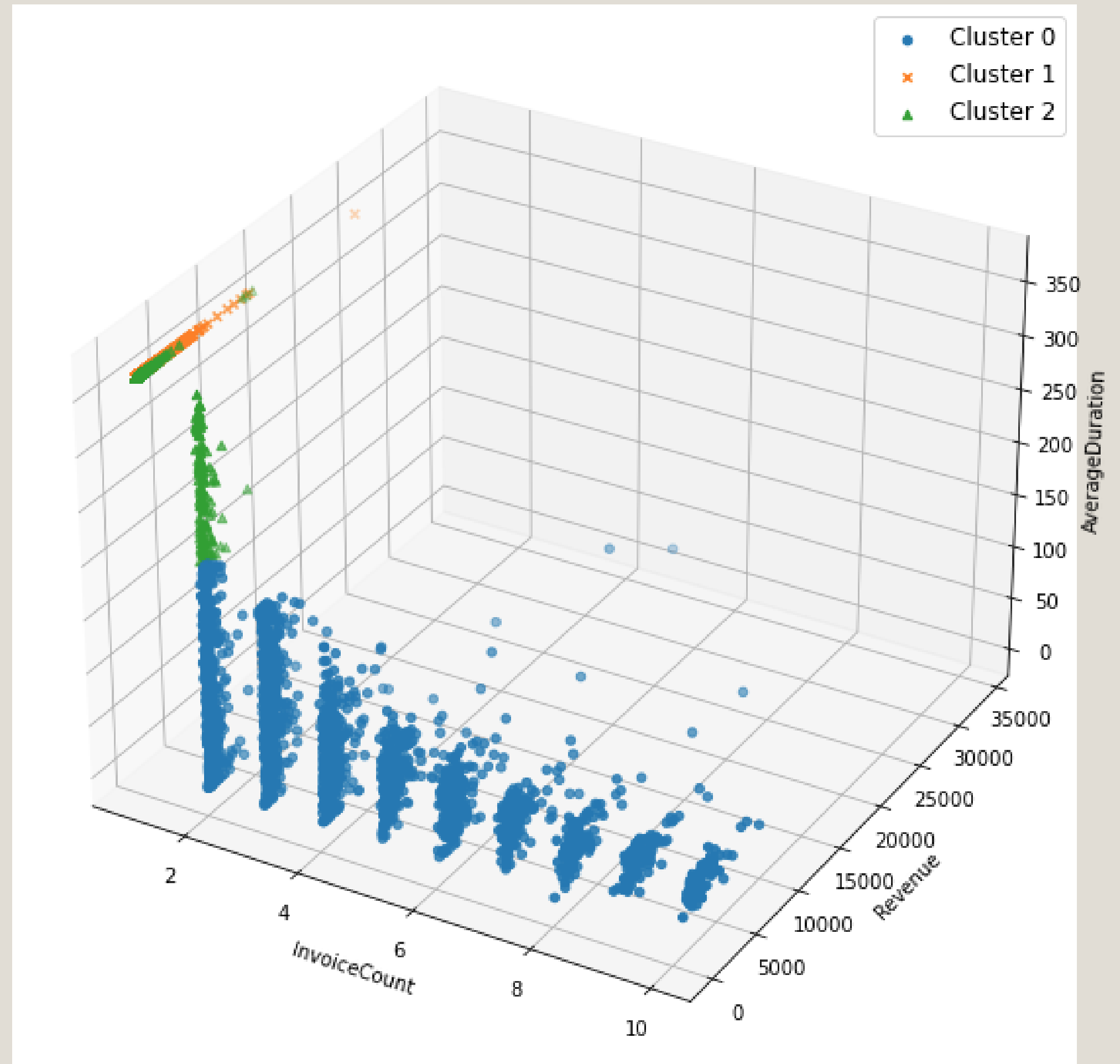
# Results

Displaying 5609 customers

Cluster 0 – Regular Customers

Cluster 1 – One-off Customers

Cluster 2 – Unloyal Customers

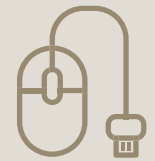


---

# Forecasting

---

# Timeseries



Features

InvoiceCount



Time

12 month period from  
December 2009



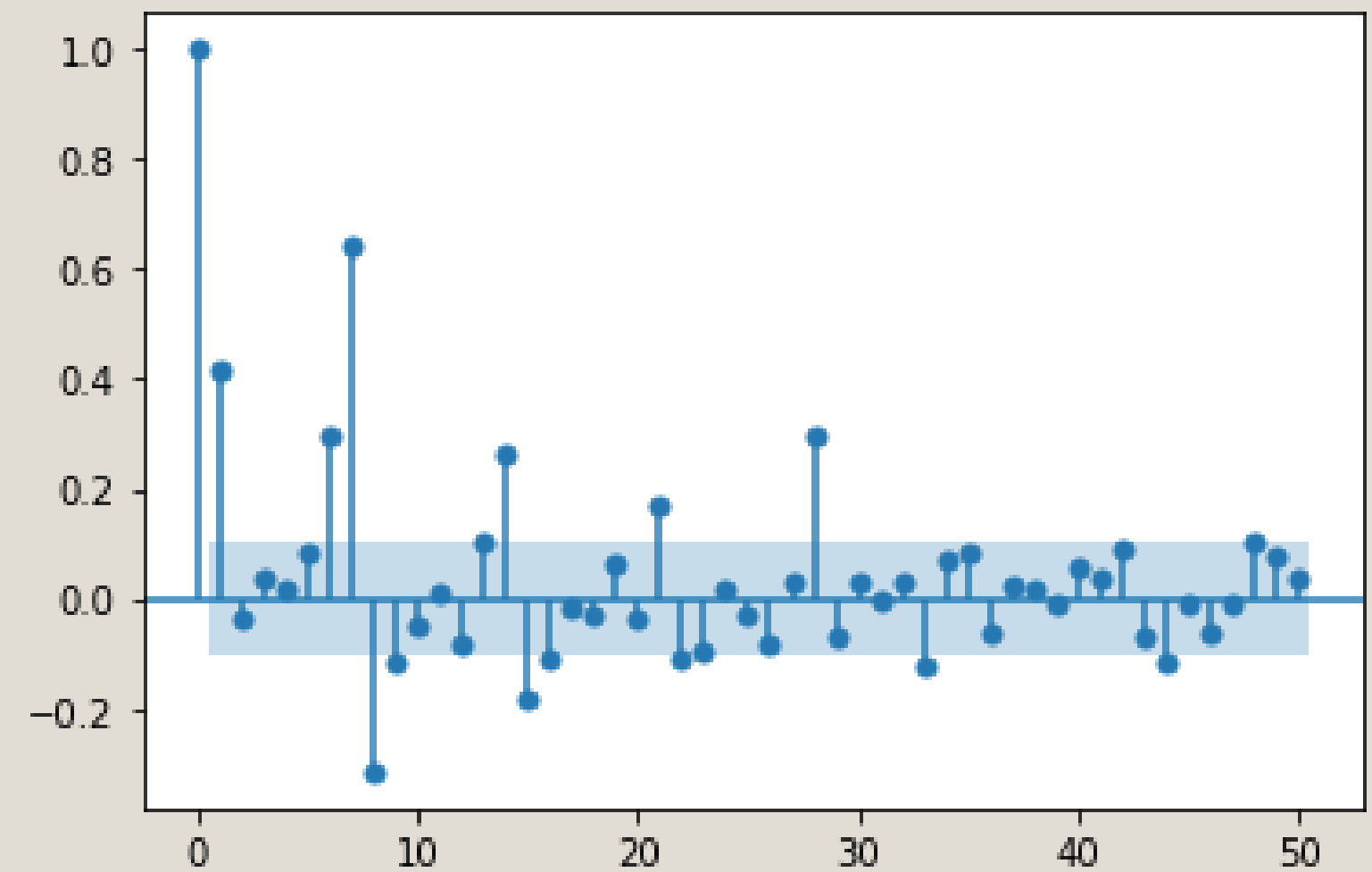
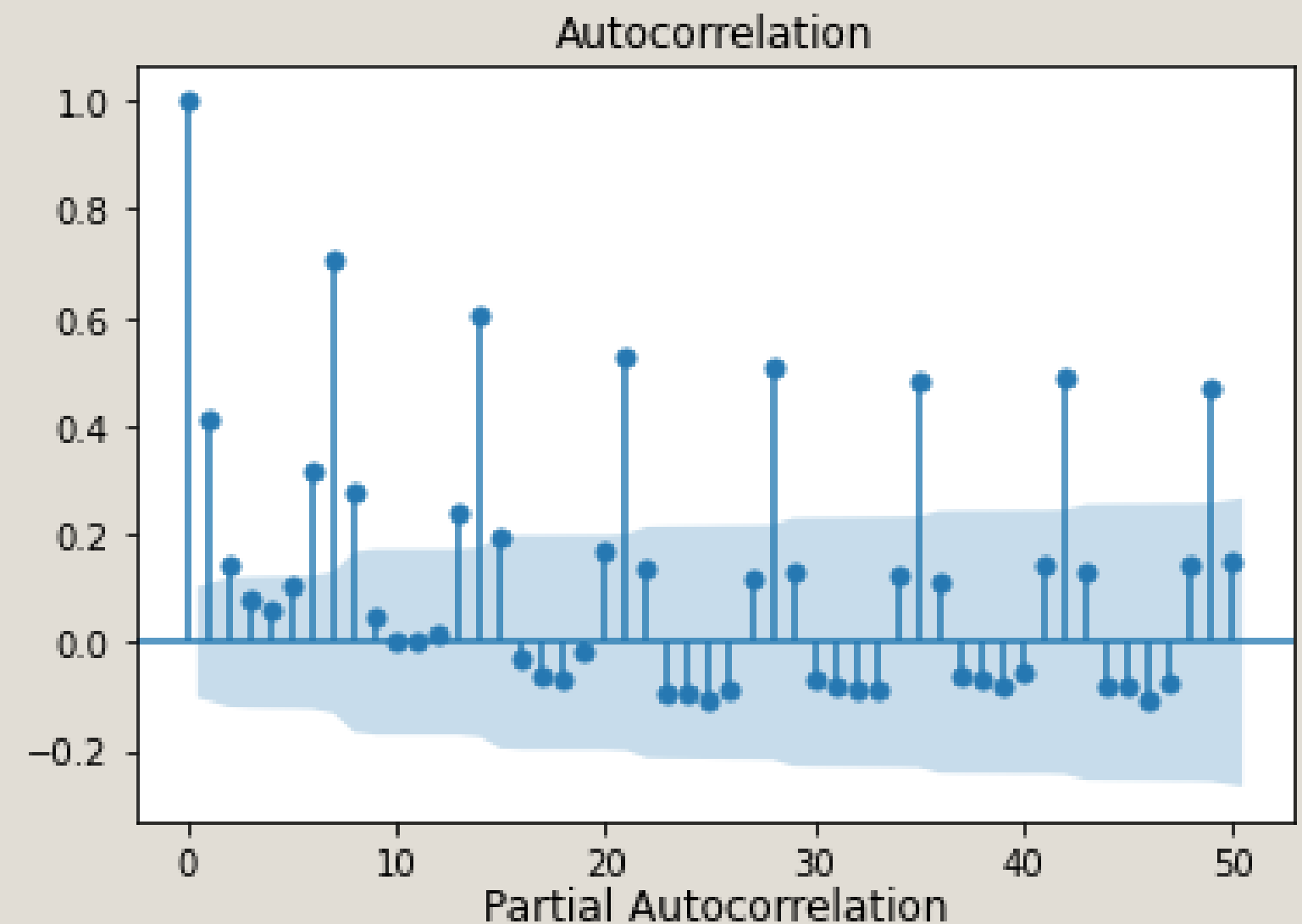
Model

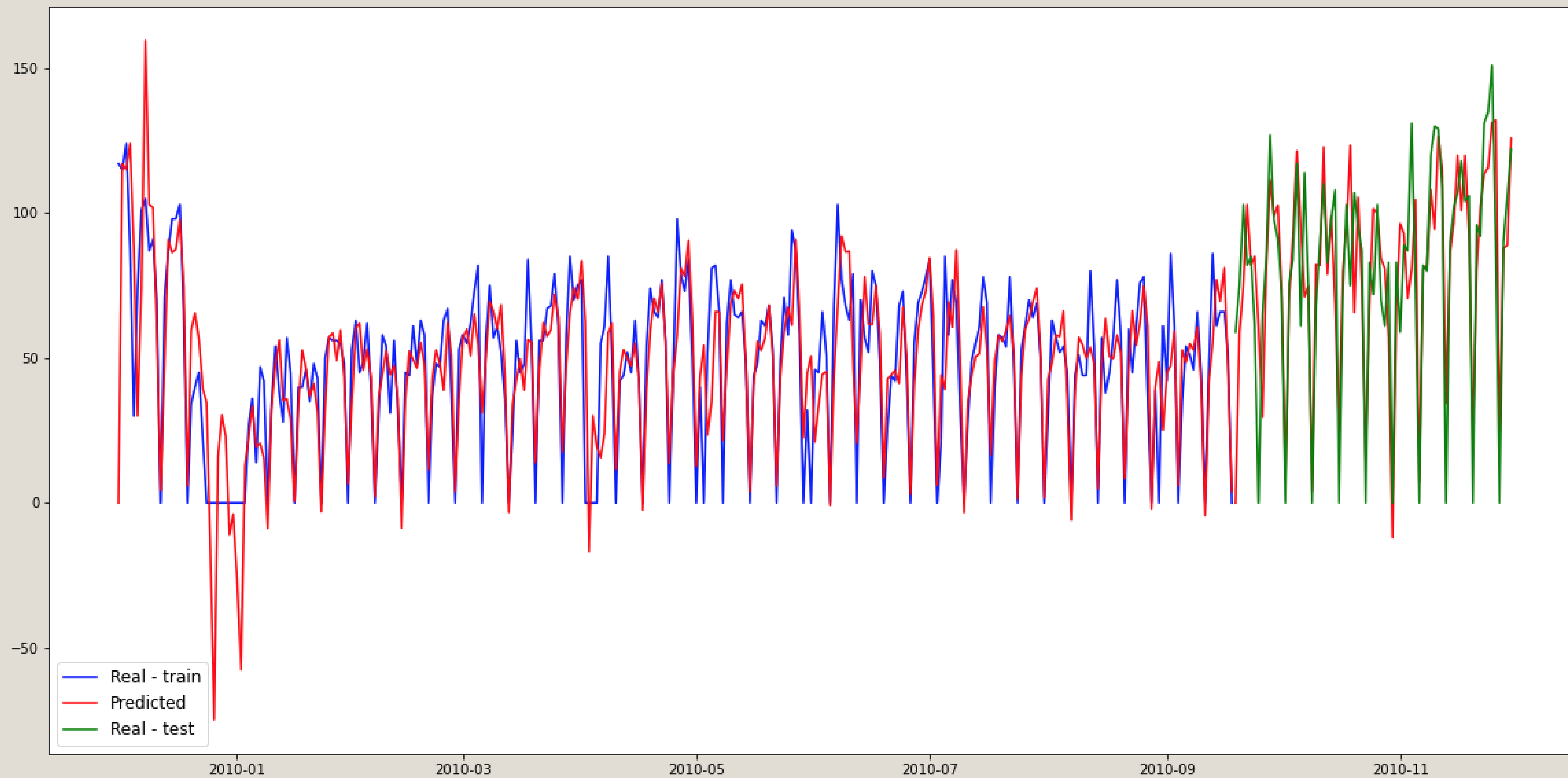
$\text{SARIMAX}(2, 1, 1), (0, 1, 1, 7)$



Evaluation

Rsquared 0.6825





Real vs Predicted



Thank you for listening

---