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 :

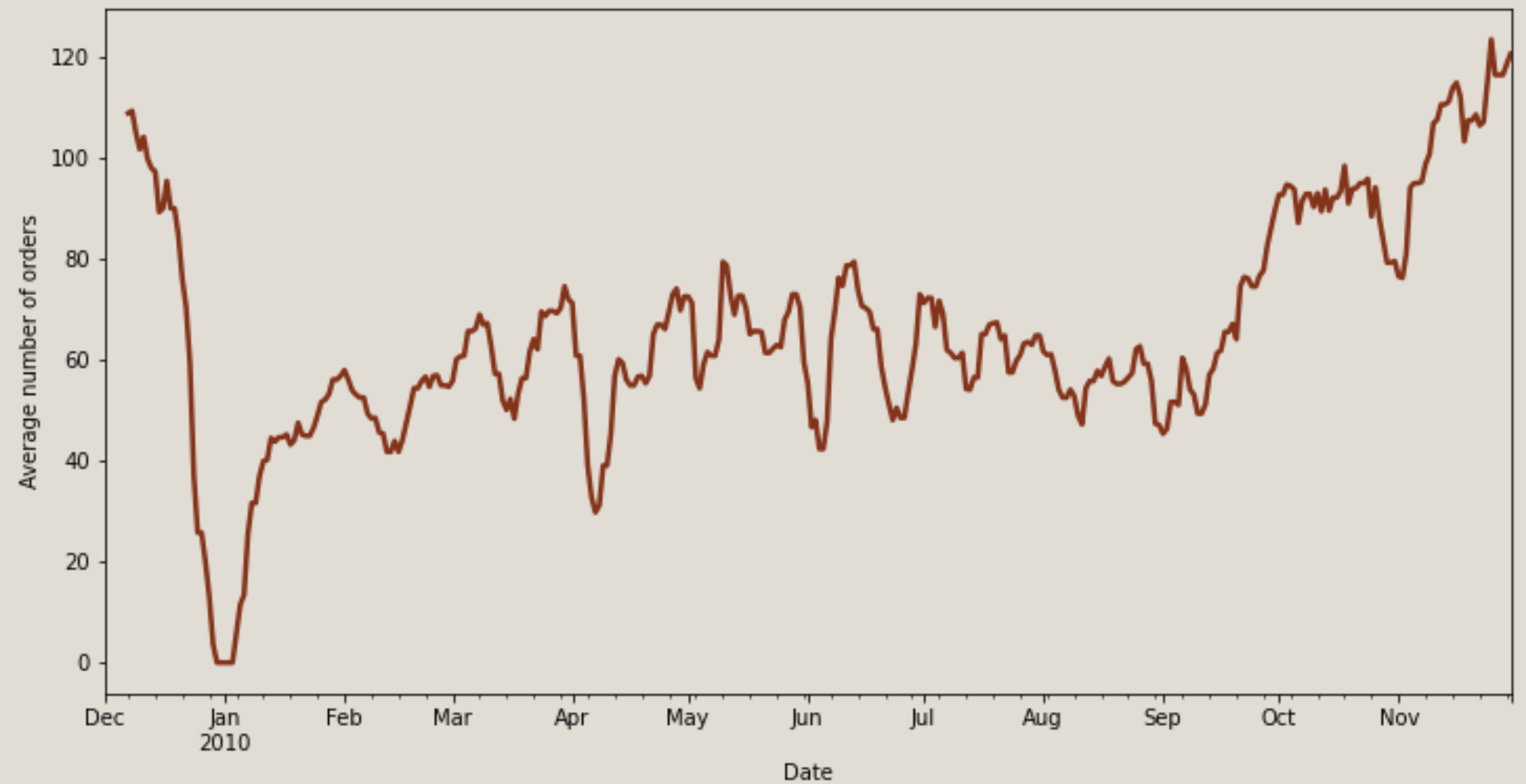- Identify types of customers
- Predict future orders

Agenda

Data Analysis

The data

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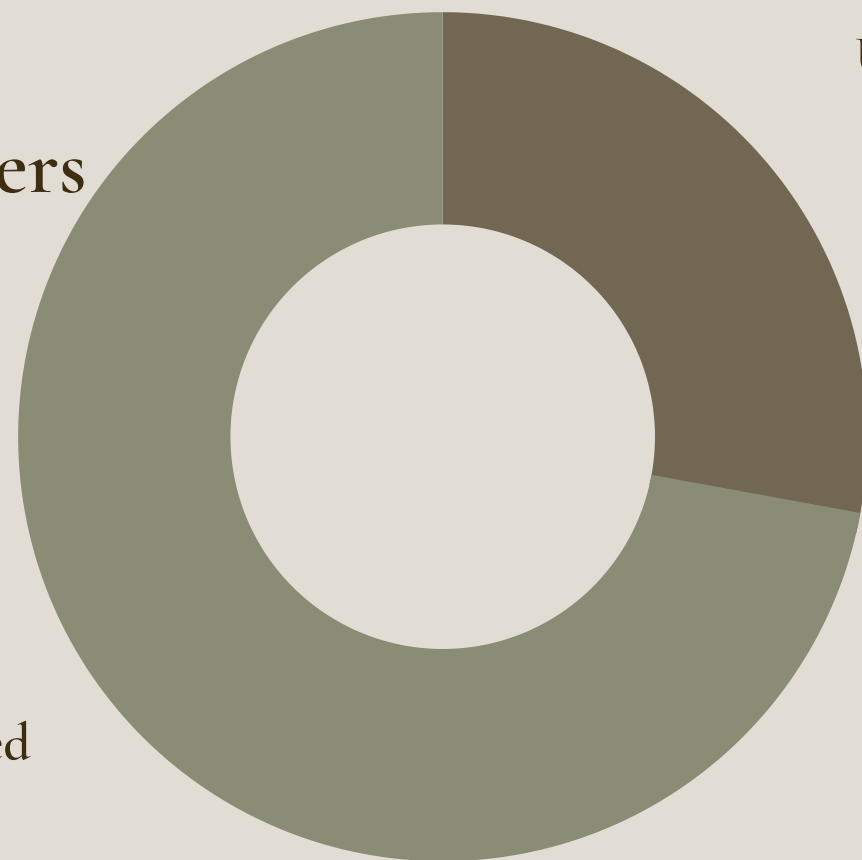- InvoiceDate
- Price
- CustomerID
- Country

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

Customers & Products

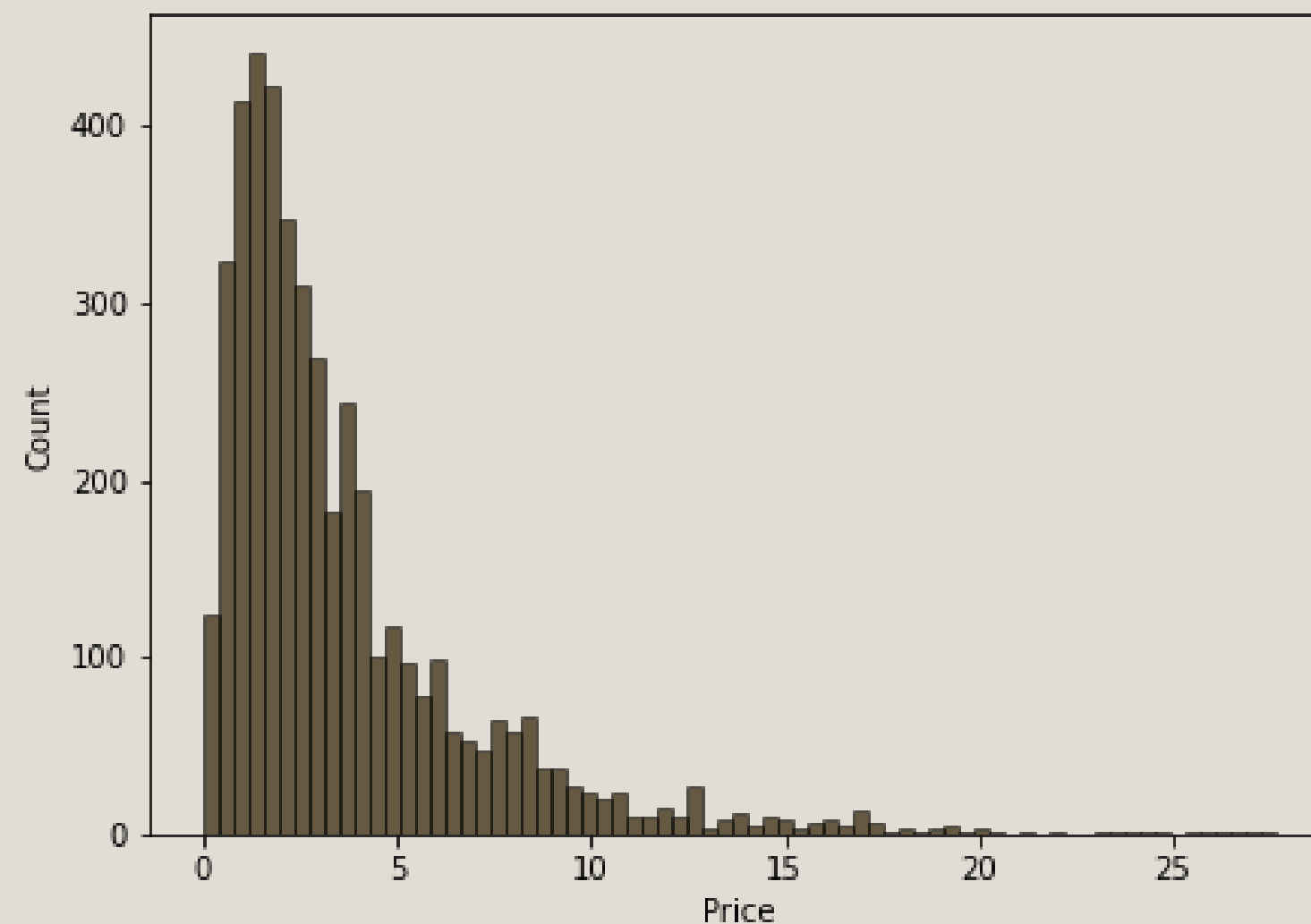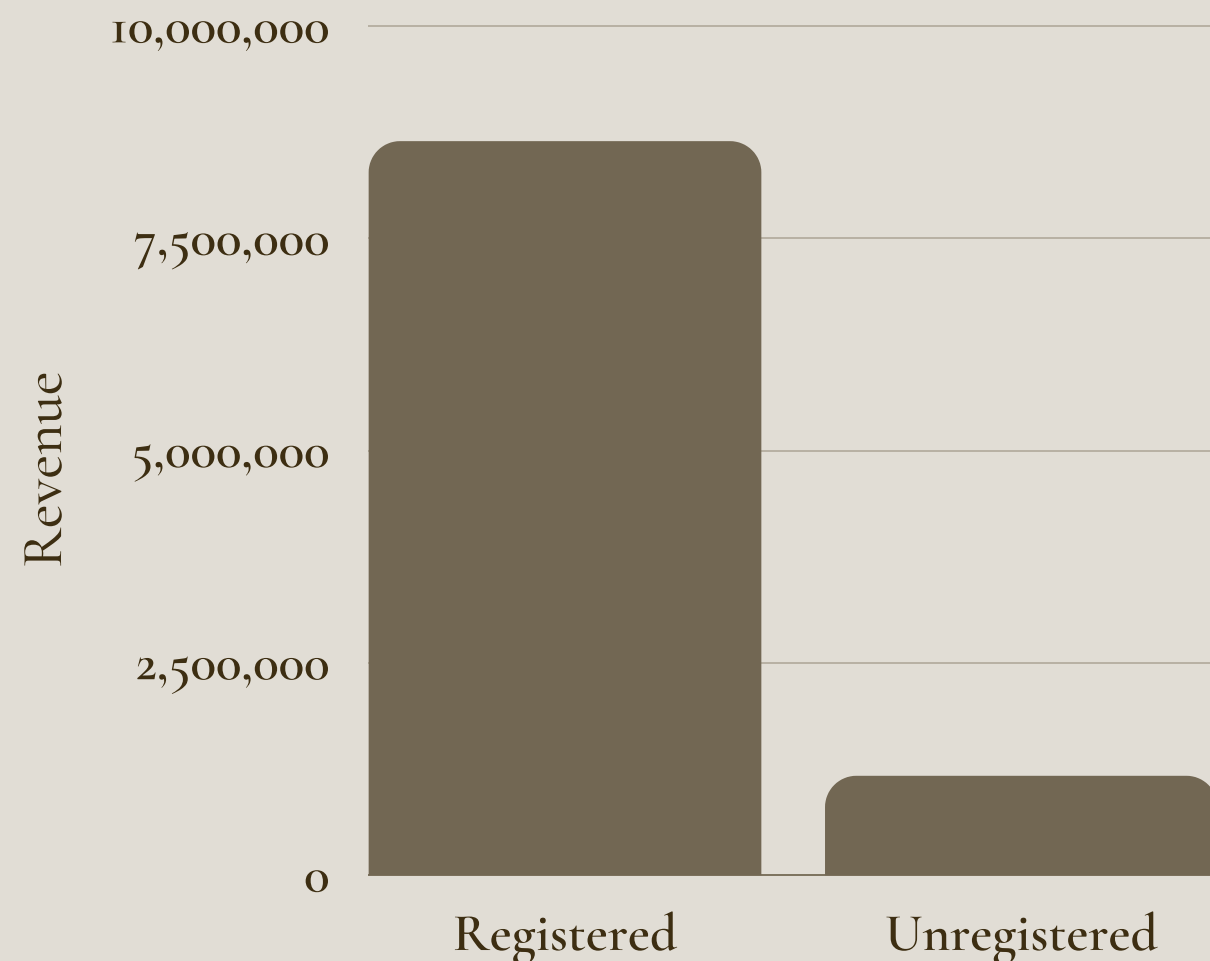**Total Customers**

Unregistered 27.9%

Registered 72.1%
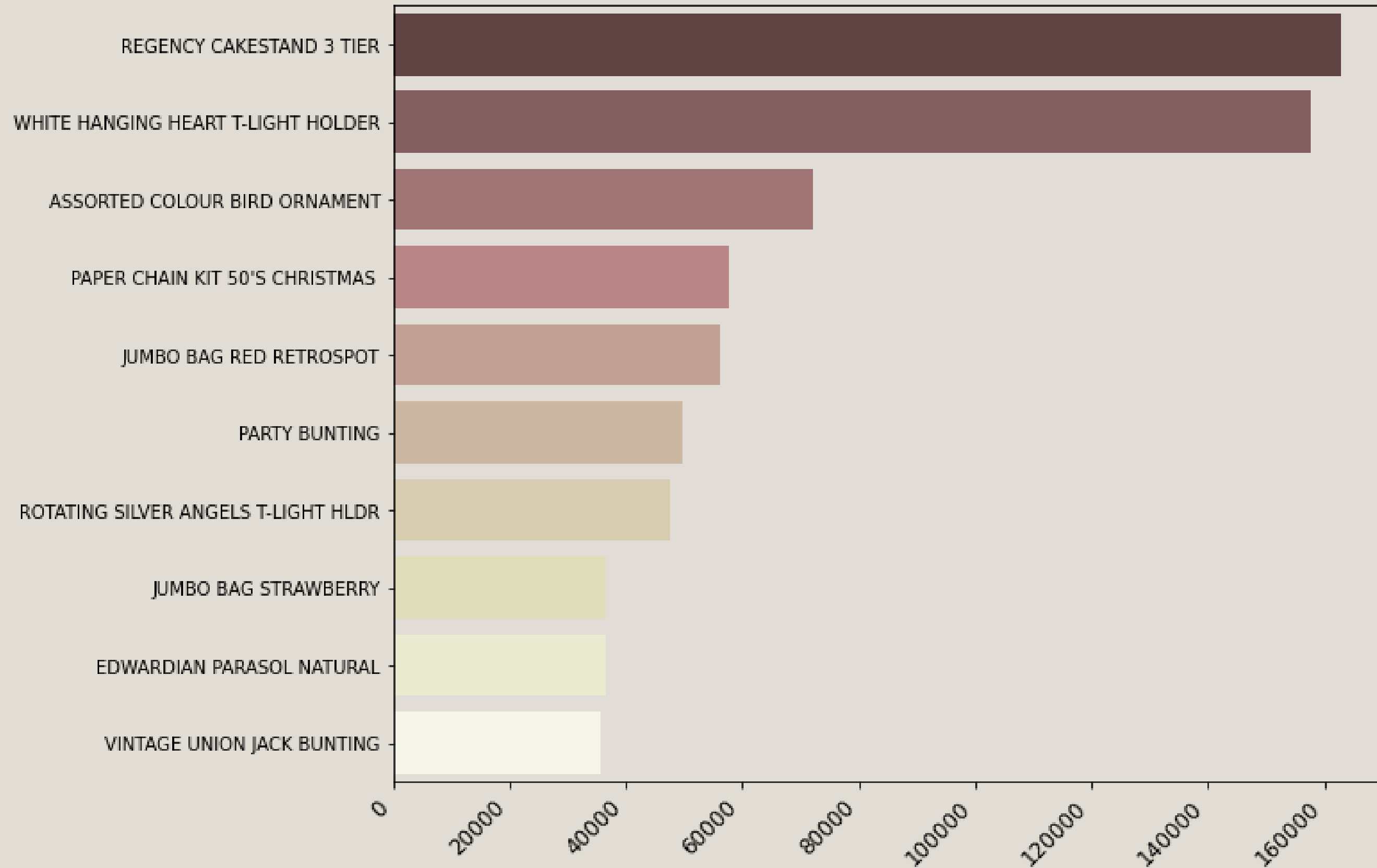
- The largest orders each consisted of 87167 items across 45 products, purchased by customers 14277 and 13687, at a cost of £11,880.84
- There are 4533 unique products
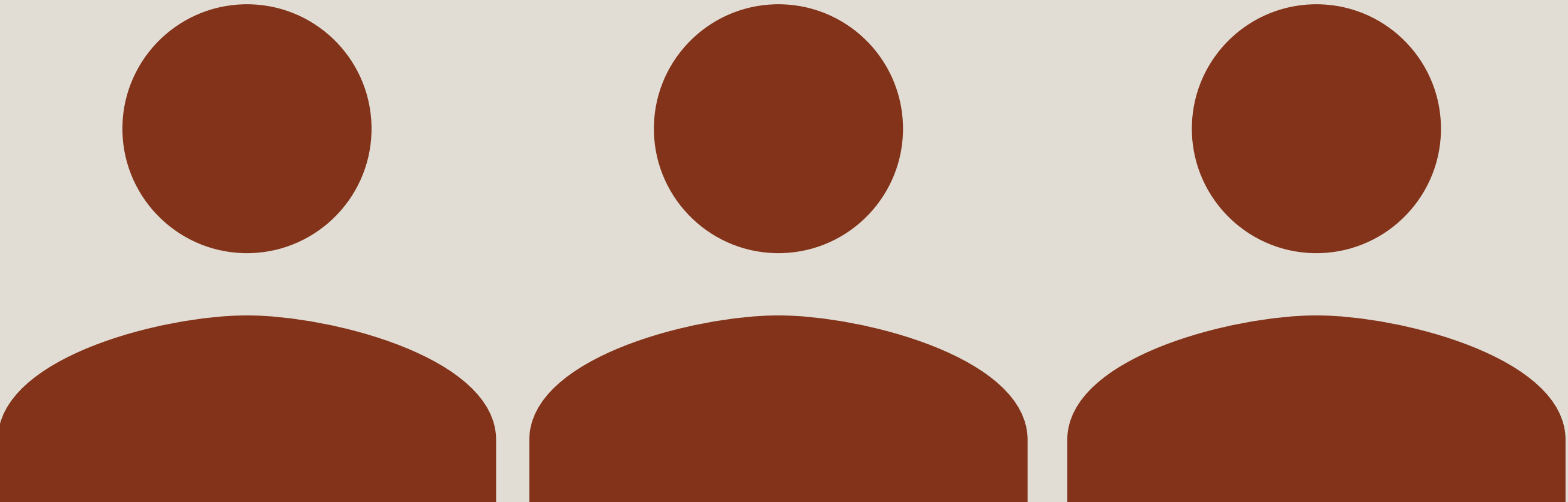- 99% of products priced less than £27.67

Revenue by Registration

Revenue (y-axis): 0, 2,500,000, 5,000,000, 7,500,000, 10,000,000

Categories: Registered, Unregistered

Best Selling Products

Customer Types

Customer Profiles

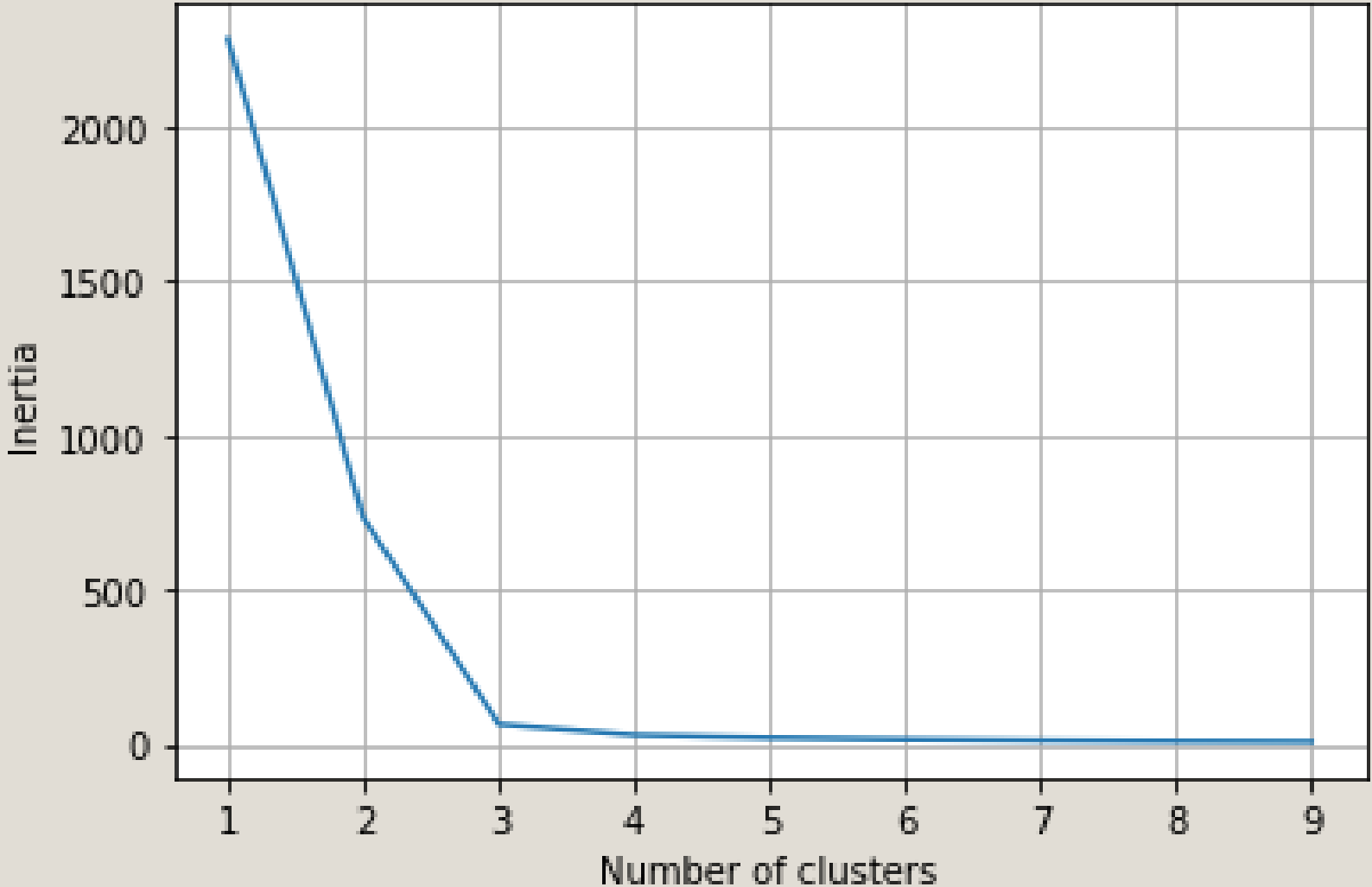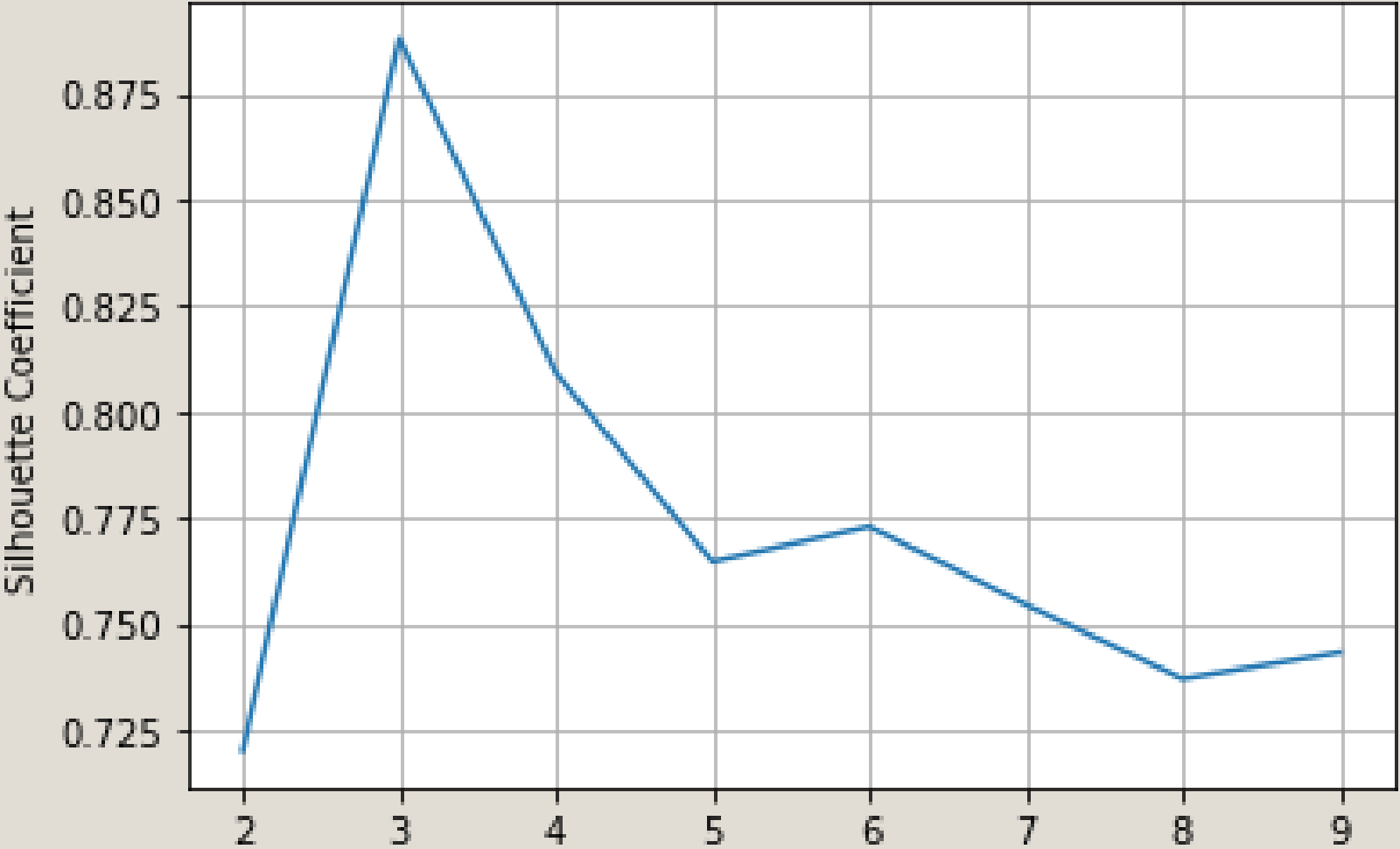| The Regular Customer	The Unloyal Customer	The One-off Customer	Cluster
2741 customers	1544 customers	1656 customers	Number of customers
6.3 orders	1.1 orders	1 order	Number of orders
£2938	£381	£703	Yearly spend
82.12%	6%	11.88%	Revenue(%)

Preparation

Dropped cancelled orders

Single Customer View

Used MinMax Scaler

Features

Revenue

InvoiceCount

AverageDuration

Registered

Model

Kmeans

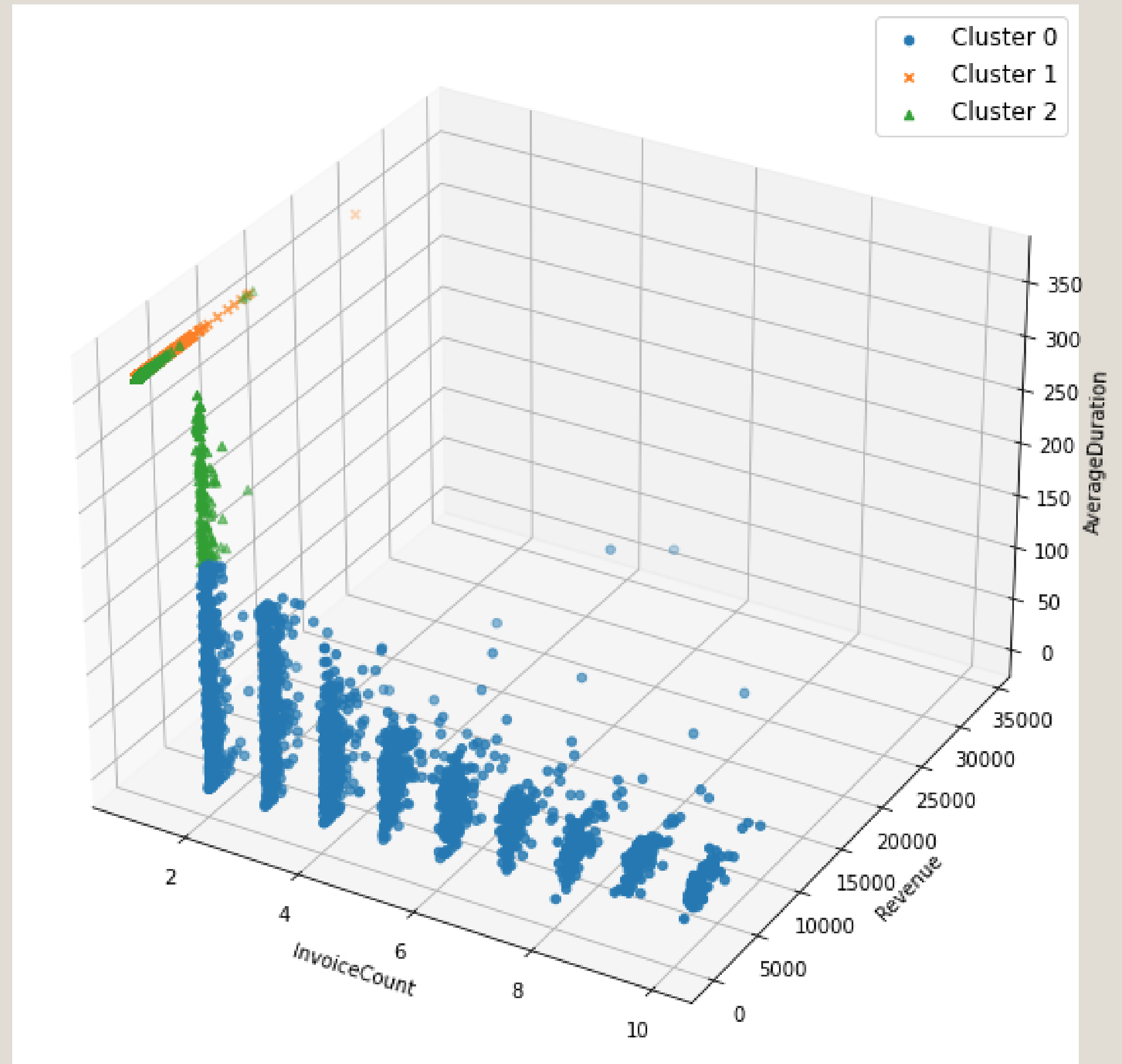3 clusters

Silhouette score = 0.893

Results

Displaying 5609 customers
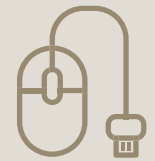
Cluster 0 – The Regular Customers
Cluster 1 – The One-off Customers
Cluster 2 – The Unloyal Customers

Forecasting

Timeseries

Features

InvoiceCount

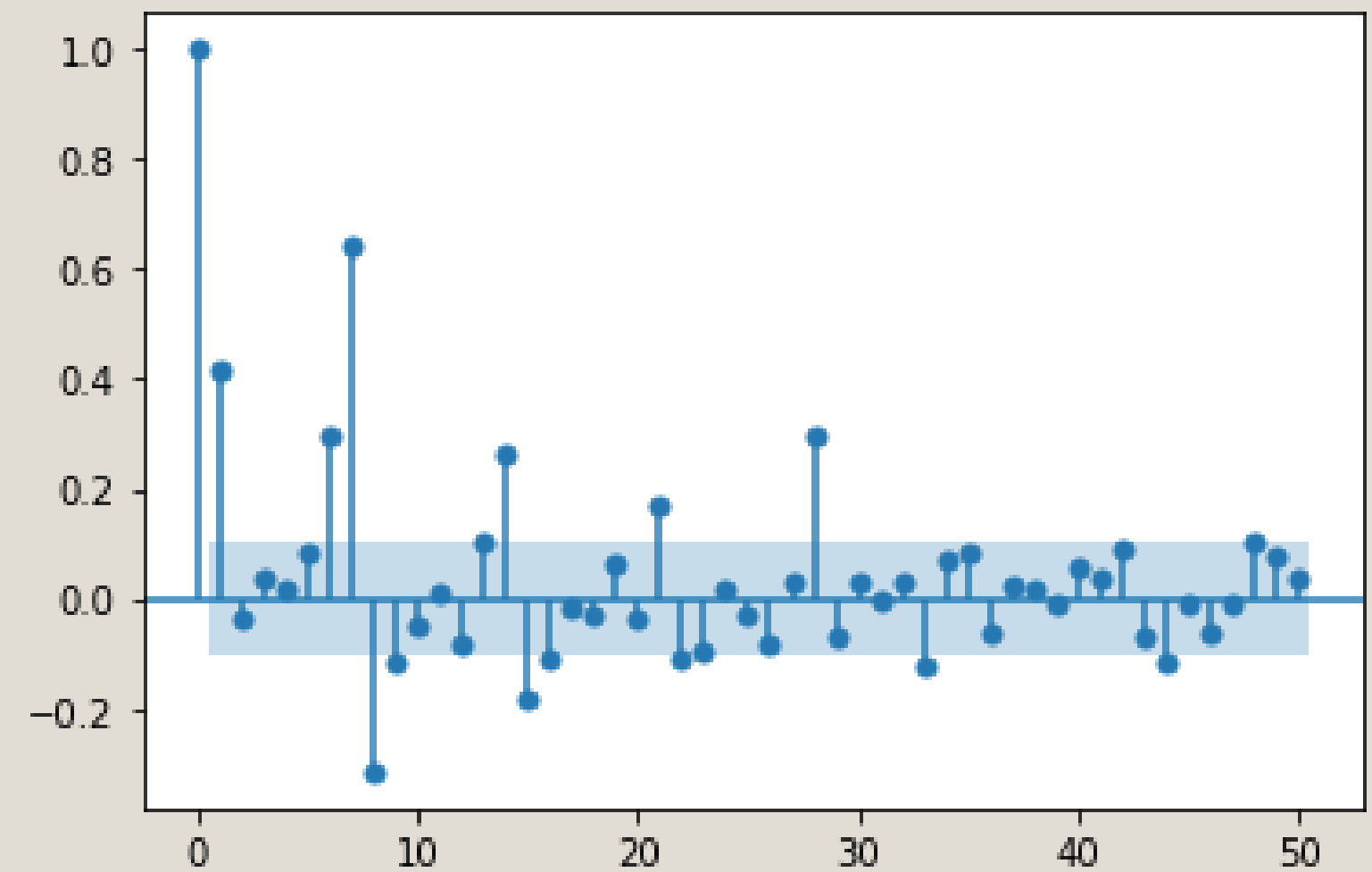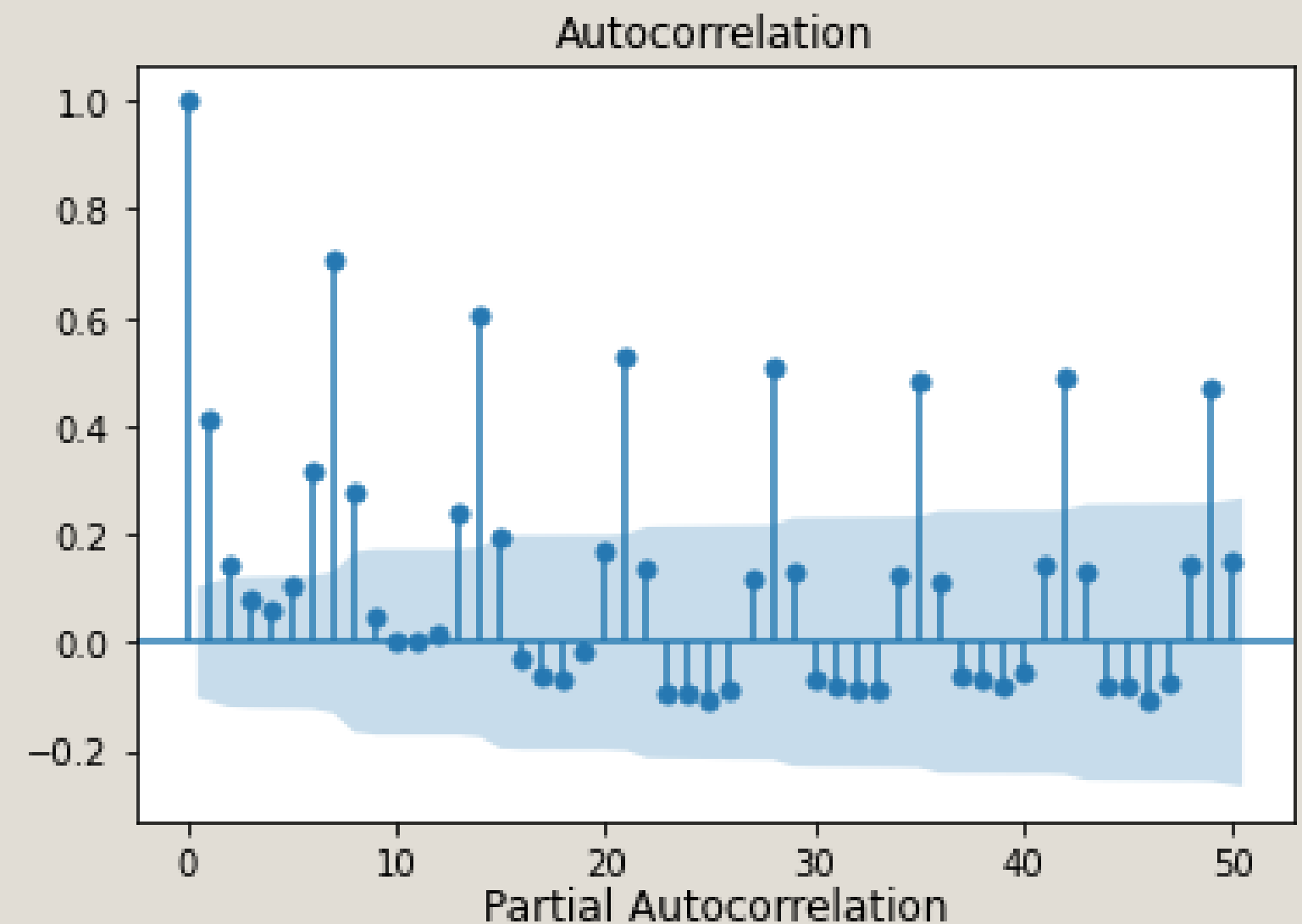**Time**
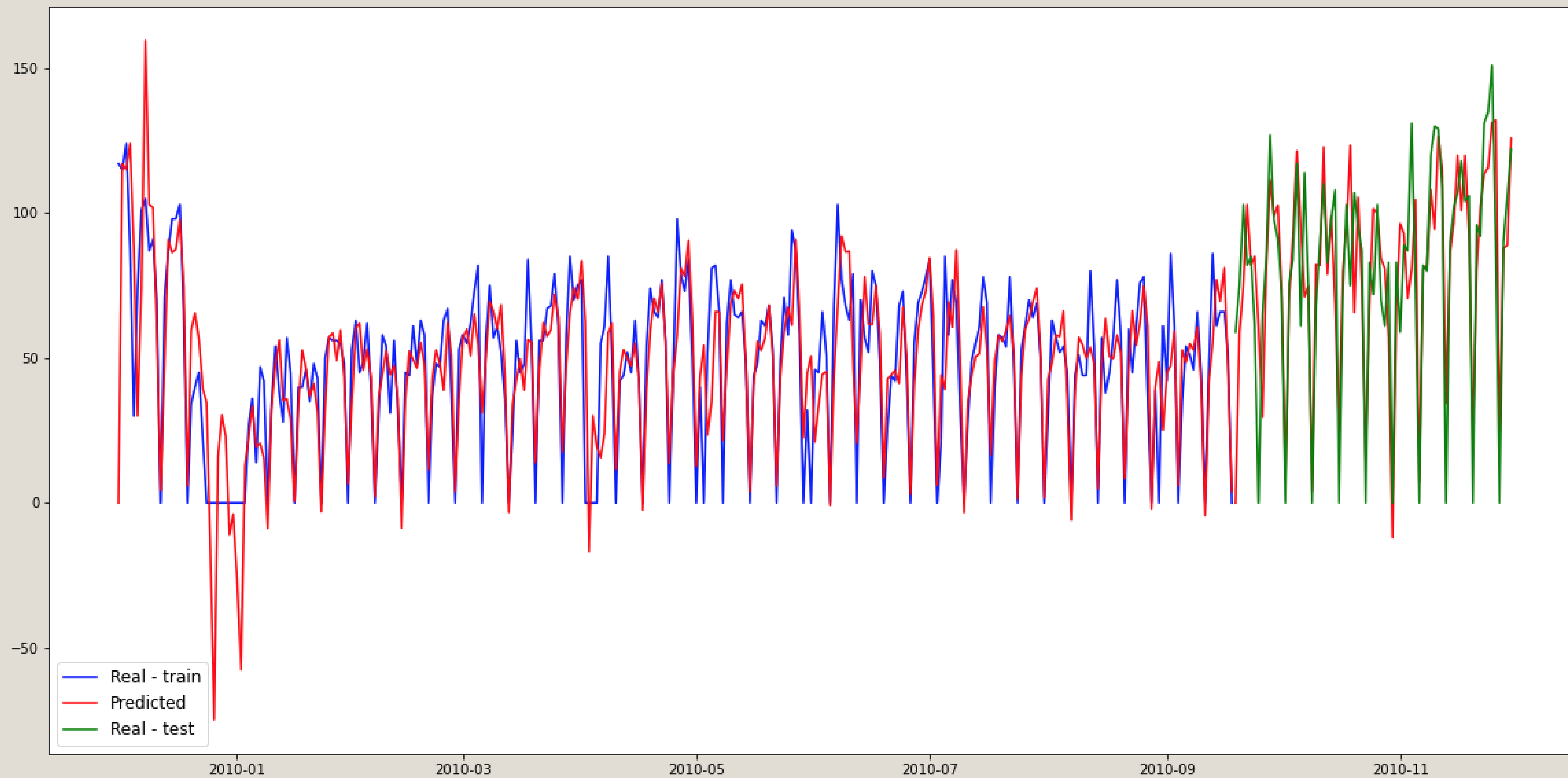
12 month period from December 2009

Model

SARIMAX(2, 1, 1),(0, 1, 1, 7)

Evaluation

Rsquared 0.6825

Real vs Predicted

Thank you for listening