# THE FATE OF ONLINE RETAIL POST-PANDEMIC

## INTRODUCTION

### OVERVIEW

COVID-19 lockdowns had forced consumers to utilise online shopping, causing significant increases in online retail rates many refer to as an "e-commerce turning point".

The question yet to be answered is : will the elevated online retail rates be **sustainable** or is it likely to **revert** to pre-pandemic levels once the crisis subsides?

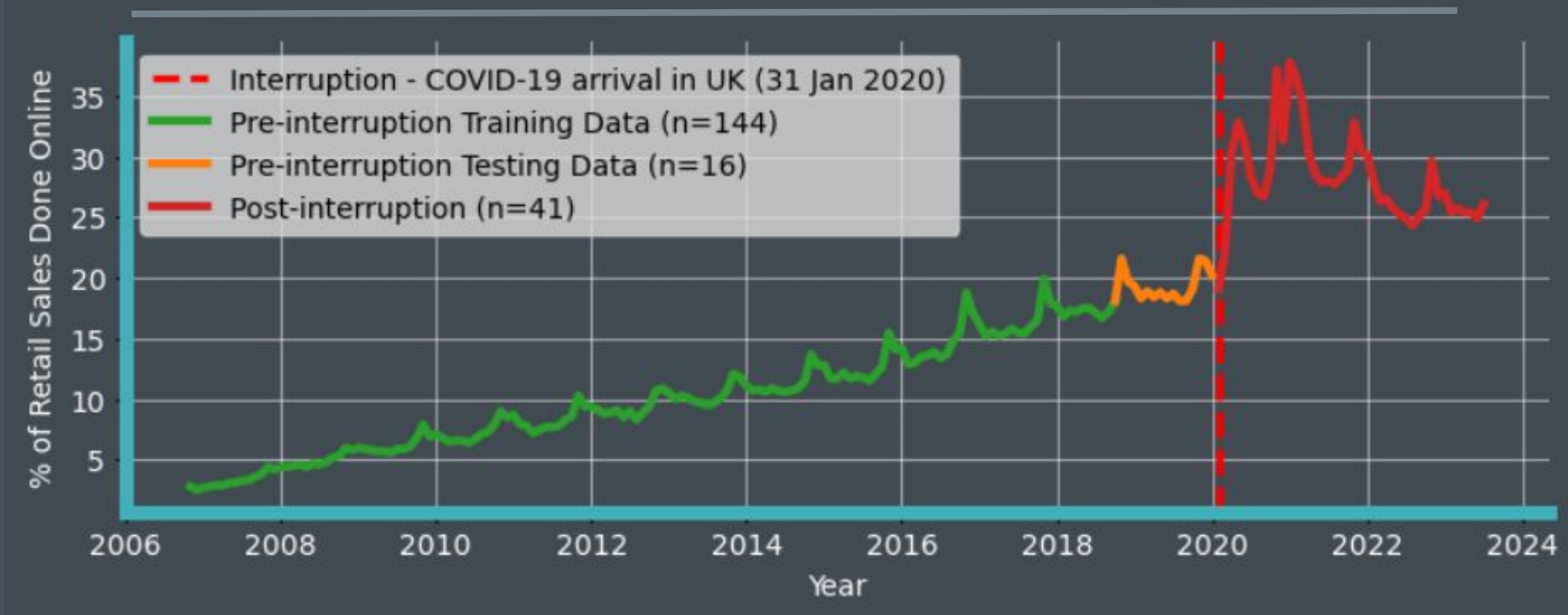
This analysis seeks to give valuable information for businesses, policymakers, and stakeholders in making informed decisions and strategic planning in e-commerce.

### METHODOLOGY

By exploring a **Counterfactual SARIMA** (**S**easonal **A**uto**R**egressive **I**ntegrated **M**oving **A**verage) **Model**, we aim to compare existing data, to the hypothetical trajectory should the pandemic not have happened.

Monthly data from November 2006 to July 2023 from the United Kingdom was chosen as it is the world's **3rd largest digital sector** and would be a good model for analysis

### INITIAL TIME SERIES PLOT




A **large increase** in online retail sales can be observed upon the arrival of COVID-19.

However, there is also a **visible decline** from 2022 to 2023 : Post-Pandemic

- **+0.9%** Jan 2019 → Jan 2020
- **+17.6%** Jan 2020 → Jan 2021
- **-3.1%** Jan 2022 → Jan 2023

For a better assessment accounting for non-stationariness, seasonality and trends, the data was split into **90%: Training Data** and **10%: Testing Data** for SARIMA Forecasting.

## ANALYTICS

### COUNTERFACTUAL MODEL CREATION

#### PART 1: Seasonal-Trend Decomposition through LOESS (STL)


**Observed Data:** Decomposed using STL into the three other components

**Seasonal:** Events such as Christmas in December cause annual (m=12) fluctuations in online retail sales

**Trend:** General trend, here it indicates a mostly linear increasing trend. This violates the stationary assumption for SARIMA Forecasting.

**Residual:** The remainder after accounting for the trend and seasonal components.

#### PART 2: Dickey-Fuller Test for Non-stationariness

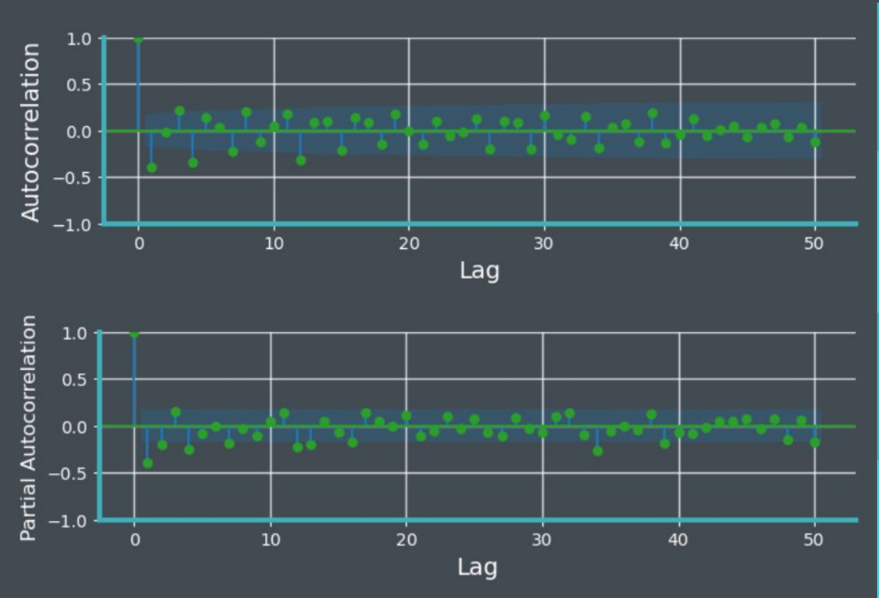
**Null ($H_0$) :** Time Series is Non-Stationary

**Alternate ($H_1$):** Time Series is Stationary

After differencing, the null hypothesis is rejected ($p<0.05$) and the data is now stationary.

**0.995** P-value (3sf) → Differencing the data according to the trend and period (d=1, D=1, m=12) → **0.000** P-value (3sf)

#### PART 3: Autocorrelation and Partial Autocorrelation Analysis




**AutoRegressive (AR) and Moving Average (MA):** Rely on autocorrelation (the correlation between a value and it's past values) in the data points and residuals respectively.

**Autocorrelation Plot**: Suggests how many terms is needed for the MA (q, Q) component in order to account for the autocorrelation in the model, it shows a significant autocorrelation at lag 1, 4 and 12.

**Partial Autocorrelation Plot**: Suggests how many terms is needed for the AR (p, P) component, shows a significant autocorrelation at lag 1,2 and 4, and an exponential decay at 12, 24, 36.

**Optimal combination of orders:** Possible values for AR orders is (p = 1,2,4, P = 1), and for MA orders (q = 1,4, Q = 0).

#### PART 4: Model Selection

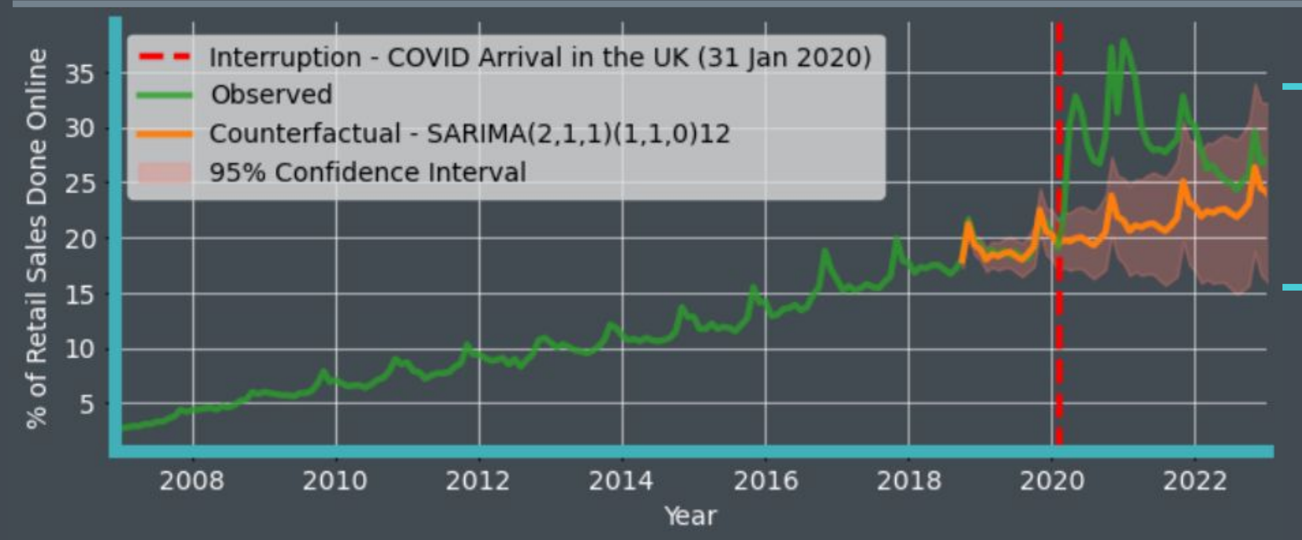
The performance of a SARIMA Model can be measured by the following metrics: **Akaike Information Criterion (AIC)** and **Bayesian Information Criterion (BIC)** on the training data, and **Mean Absolute Percentage Error (MAPE)** on the testing data.

They measure model accuracy and overfit/complexity (The lower the better!), and were used to select the optimal model.

| 100 | 115 | 1.7062% |
|---|---|---|
| AIC (3sf) | BIC (3sf) | MAPE (5sf) |

Metrics of the selected model: SARIMA(2,1,1)(1,1,0)12

### INTERRUPTED TIME SERIES ANALYSIS




Statistically significant difference between the observed values and the counterfactual values ($p<0.05$) from 2020/2 to 2022/1 given by confidence intervals

From the **Counterfactual SARIMA Model**, the following can be observed:

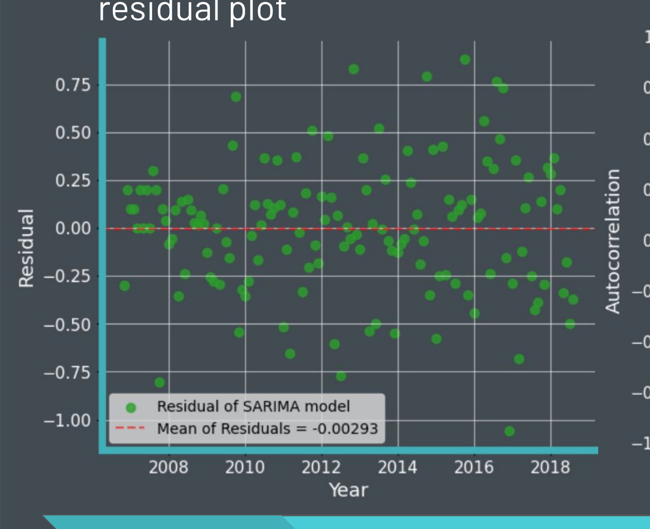
1. The spikes and seasonality observed in the observed data is **not necessarily a result of the covid-pandemic** but just natural fluctuations due to holiday seasons etc.
2. As of 2023 July, observed data is slightly higher than counterfactual by about **2.6%**, however, the difference had **stopped being statistically significant since January 2022.**

### CONCLUSION AND EVALUATION

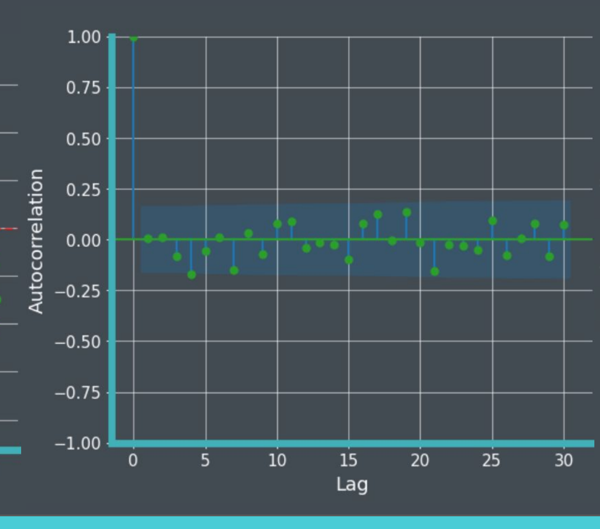
Through in-depth analysis, we conclude that although online retail rates are still expected to rise at a generally linear trend with seasonal periods, businesses and stakeholders should not be mistaken with the notion that COVID-19 has forever revolutionised online retail or has made it into something much larger than it is. This is as there is no statistically significant difference (confidence level of 95%) between the existing data points and the hypothetical counterfactual model as of 2023/7, and we can confidently conclude that the rise of online retail during COVID-19 has not triggered a permanent shift in consumer behaviour.

#### Evaluating Assumptions of the Analysis

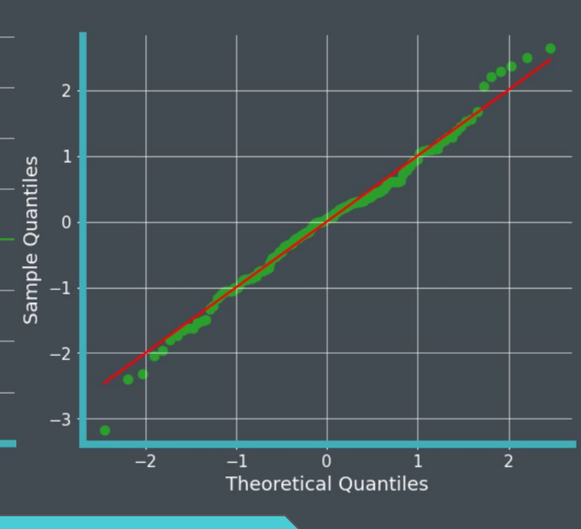
**1. Mean** of residuals close to 0 (Mean = -0.00293)(5dp)
**2. Constant variance** shown in residual plot




**2. No Autocorrelation** in residuals indicating all relevant information was used in model




**3. Normality** of residuals shown by sample quantiles following identity line in Quantile-Quantile Plot




#### Bibliography


Office for National Statistics (2023) Retail Sales Index Time Series (DRSI): https://www.ons.gov.uk/businessindustryandtrade/retailindustry/timeseries/j4mc/drsi

International Trade Administration: Impact of COVID-19 on eCommerce: https://www.trade.gov/impact-covid-pandemic-ecommerce#:~:text=This%20chart%20shows%20us%20clearly,forecast%20sales%20growth%20rates%2C%20respectively

United Nations Conference on Trade and Development (15 March 2021): How COVID-19 triggered the digital and e-commerce turning point https://unctad.org/news/how-covid-19-triggered-digital-and-e-commerce-turning-point#:~:text=As%20lockdowns%20became%20the%20new,to%20about%2017%25%20in%202020