Accessible Inventory Forecasting and Analytics

Degree: Data Science BSc (Hons)

Module: CIS3140 Research and Development Project

By: Christopher Diaz Montoya

ID: 24707686

Module Lecturer: Ella Pereira

Module Supervisor: Marcello Trovati

Date due: 10th of May 2023

Date of Submission: 15th of May 2023 (Due to health)

‘This Report is submitted in partial fulfilment of the requirements for the BSc Honours Computing (or Web Systems Development, Computer Science) Degree at Edge Hill University.’

i: Table of Contents

ii Table of Tables

iii Table of Figures

iv: Abstract

In 2023 there are many e-commerce companies who use data for forecasting. Using tools to forecast inventory and sales is something companies are happy to invest money in as it would improve profits (Abeysekara and Rupasinghe, 2019). Forecasting does this by reducing overstocking as there is a cost to store goods, paired with reducing the amount of stock that would pass its sell by date, it also reduces understocking inventory which leads to unhappy customers and loss of sales. (Puneet et al, 2021).

The methodology involves data integration, cleaning, analysing, and applying machine learning algorithms to the data, both regression and market basket analysis algorithms were used. This was done using the python programming language on Google Colab., a guide to embed Chat GPT was added to allow users to learn python.

The result is a cleaned data set, informative visualisation, machine learning algorithms, and a chat bot. All on an evolvable interface which small businesses could use to their needs. Linear regression proved to be a good model and the forecasting was accurate. Upon completing the prototype, it was clear the data industry requires expertise, cash, and time to develop functional products such as Tableau and Looker.

Chapter 1: Introduction

This thesis focuses on the creation of a prototype designed to assist small businesses in decision making. It will provide an analytical aspect to show the data in a visual format along with some machine learning algorithms to provide deeper insight into the data. The aim is to give small businesses a tool that is free, reliable, a learning environment, evolvable, connects to the weather to see if it affects sales, and visualises the data well.

This report will show the creation of a sales forecasting and analytical tool that was originally intended for just inventory forecasting, this change was done to provide more valuable insights for small businesses. This includes but is not limited to:

- Sales Visualisation
- Inventory Management
- Financial Planning
- Decision Making

This would accept a spreadsheet of data in a specific format that can be updated and then produce a forecast of inventory along with having a basic analytical dashboard for viewing historical data. This would be done with underlying machine learning algorithms which are designed to use historic data to predict future trends.

The tool used for the project was Google Colab. This is a cloud-based coding environment which allows for the python programming language to be used. It uses Google's hardware and makes it easily accessible for the creator and user to view.

After creating a free analytical and forecasting prototype on Google Colab, it was then assessed at the end of the project to see how effective it is, reliable it is and how small and family run e-commerce businesses can implement this to make strategic choices for the business at a low cost and how they can use the tool continually.

Chapter 1.2: Aims and Objectives

The aim of the project is to create accessible analytical and forecasting tool. The aim is to provide a free, accessible, trustworthy, and visualised sales inventory prototype for e-commerce businesses. When breaking each down into objectives and how they were to be approached and how the approach of each objective will be evaluated in the conclusion. To do this some key objectives were established to measure success in the project which can be viewed in Table 1 – Prototype Objectives.

Table 1 - Prototype Objectives

Objectives	How	Approach
Trustworthy	The forecasting algorithm must have a high degree of confidence.	The machine learning algorithm created will be evaluated using ensemble learning and then evaluating the algorithm for accuracy in a number of ways such as through accuracy, f1-score or Mean Squared Error (MSE). Along with a forecasting buffer for improved ML algorithm accuracy.
Cheap and Accessible	Trying to not include a charge, making it easy to access online.	This was done be keeping everything on Google Collab, making a public notebook so it is easily accessible with help on using the notebook.
Visually Appropriate	Making it easy to understand the visualised data and navigate through the page.	This is done by not only ensuring that the graphs are colourful in way that makes them easy to interpret, but ensuring they show what is

		needed, along with having the graphs appropriately labelled using libraries in python.
Check if weather effects sales forecasting	Adding historical weather data to the date of sales to see if there is a correlation.	An API will be used to connect historic data of the weather and will be added into the data frame, this allows for the weather to be a analysed as a variable in a heat map as weather is overlooked (Abeysekara and Rupasinghe, 2019).
Making it a learning tool and easily editable	Using Google Colab as a public notebook and adding Chat GPT.	The Google Colab notebook was publicly published to and a guide on how to download the Chat GPT extension was incorporated.
Market Basket Analysis	Choosing a specific algorithm to find which products to recommend together.	This was done using a frequency and association algorithm, this checks common pairs. The algorithm used was the Apriori algorithm. Finally, the pairs will be displayed to show the products.

Chapter 1.3: Scope

The scope of this project includes creating the prototype from scratch, this tool focuses on, sales and product analysis, teaching businesses about data and everything in section 1.2. This is aimed for small e-commerce stores to aid in processing and visualising their data.

Chapter 1.4: Methods

The methodology used for the project was to gather data, prepare it, visualise the data and use specialised algorithms on the data for deeper insights. This was all implemented within Google Colab, taking advantage of the cloud-based functionality to code in the python programming language.

Chapter 1.5: Outline of Thesis

This report is split into multiple sections. It will go over the creation of the analytical and forecasting system. The project will be split into following; the project background; the literature review; the projects aims and objectives; methodology and build; legal, social and ethical considerations taken; and the conclusion.

Chapter 2: Project Background/ Literature Review

Chapter 2.1: Overview

One aim is to make this tool as accessible as possible for family run businesses, small, and medium enterprises (SMEs) as data tools are not cheap (Benhamida et al, 2020). One important project aims and cause for the project is to create a cheap, trustworthy, accessible Artificial Intelligence and analytical tool for small e-commerce business owners as data tools are not cheap (Benhamida et al, 2020). The goal is to create a place where the past years sales can be tracked and analysed along with being able to forecast sales to allow for better business decisions.

Chapter 2.2: Existing Tools and Techniques

Currently analytical tools exist such as Stock & Buy and Tableau for e-commerce companies, but they cost over £50 a month which is not viable for small businesses. Businesses track their sales and track inventory as closely as possible to not make any decisions which could cause an effect to the business. This leaves Excel but it is not great for pre-processing data for analysis, while it can be done a coded pipeline where the data can just be dropped would be ideal and was what gave the idea for this prototype.

When looking at machine learning techniques which could be used for sales or inventory forecasting which are coded manually and not paid for on python, a regression-based algorithm was found to be of the best fit when there is historic data for a product, this algorithm is used in industry (Ranjitha and Spandana, 2021). When there is little to no historical data it can be difficult to predict the future, here a clustering-based algorithm was found to be best, this is as it allows it to be grouped to similar products (Benhamida et al., 2020). No algorithm would be able to be 100% accurate. This meant that a buffer should be added. When thinking of worst-case scenarios an algorithm could over or under predict sales. As these algorithms are not vastly off, a buffer could be included to ensure this does not happen which could affect business decisions. The buffer has been proven to work well with regression algorithms (Puneet, 2021). The algorithm is meant to be a predictive guide and not a concrete view into the future, although that is the aim.

When looking into how to visualise the data a free to create interactive dashboard, without having to learn HTML or Java script, was found, Dash. This allows to integrate python libraries and their analytical tools onto the web, the loading time was quick as well (Clement et al, 2020). The downside of this tool was that additional coding is needed and learning how to use a new tool would take additional time. Other tools found were Tableau and Looker, these are easy to use analytical platforms that allow users with data knowledge to make dashboards quickly, this is free for businesses and quick to use and link to data. It is roughly over £50 a month which becomes pricey for small businesses over a year and cannot be used as the point is to make this tool accessible to anyone. The final tool found was Google Colab, this is a learning environment that is used a lot in the Data Science field. A lot of cutting-edge tools such as Facebook's Detectron is on there as a learning notebook available for the public. This is a perfect tool as it can display data quickly from Google Drive using python libraries with a lot of code hidden quickly by hiding a section. If the user wants to learn about how, it was made or if they have python coding experience in the data field, the user can open the hidden code and read and make personal changes this allows the prototype to be both a learning and an evolving environment. As such Google Colab was the best fit for the project and was used in building the prototype.

When looking at incorporating a chat bot, Open AI have created a form of complex Machine Learning algorithm that can be embedded into applications by installing an extension, instead of hard coding and learning how to create a Chat Bot, this saved the need for research as Chat GPT embedded with a couple clicks (Developer Tools, 2023).

When looking into connecting weather data, an API would be best to connect historic data of the weather and will be added into the data frame, this allows for the weather to be a analysed as a variable in a heat map as weather is overlooked in the found papers (Abeysekara and Rupasinghe, 2019). A number of API's were found, the problem found with these API's is that they charge after a certain number of requests and did not allow the project to run smoothly so this was not incorporated into the prototype (Open weather, 2023).

Chapter 3: Methodology and Development

Chapter 3.1: Data Collection and Preparation

Finding an available data set that was free to use and had the necessary columns was difficult to find, eBay and amazon had some available datasets at a cost. As such different websites such as Google, GitHub and Kaggle were investigated to find a free dataset suitable for inventory forecasting or sales data. A data set was found that had columns which were deemed a good fit on Kaggle (Basri, 2022), after consideration on if the columns had the date, price, stock information, location of sale and total transaction value. As the data set took longer to find than anticipated, it set the project back a week. The data set found was one that contained all the transactions for a business which is in the UK.

When looking at storing the data, companies can store data on the cloud in Google Drive, this was deemed a better choice over storing the data locally as it can be linked directly, and everything is stored on the Google cloud ecosystem. The data set was then uploaded with the pandas library after mounting Google Drive into Google Colab. This would not work, and the data set would not open, this had been encountered last year on another project and the Chardent library was used to help detect the encoded value to open the data.

The data was now uploaded. The data types were investigated, and object data types had to be delt with. Many of the columns were objects when they had to be something else, for example the Invoice Date column had to be converted to a date time data type to a numeric and the Description column a string which would then become numerical.

Each row was either its own invoice or it could have been part of one, there was a Quantity column and UnitPrice column but not a total so one was created by multiplying the two columns Quantity and UnitPrice. Invoices are not always sent out on the day of sale; this could impact forecasting. From here the data was checked to see if the Quantity column had any negative values, this was assumed to be cancelled

or returned orders due to having the letter C at the beginning of the InvoiveNo for negative quantity in a row. This was used to create a new column that stated if it was a sale or cancelled. Next the data set was explored for duplicate values, this was important as there were over ten thousand duplicate rows which were then dropped using if-else statement to helps keep track of duplicate values.

When doing the worst product graph, there were some strange highly negative values, this was then taken out manually in the data prep. As when most were taken out the accuracy of the sales forecast dropped. As such the outliers were not removed at first due to this. The outliers dropping was added in the end and it helped improve the algorithm accuracy as it was paired with specific rows that had already been removed such as “Amazon Fees”. The box plot was learnt in a previous project and creates an inter quartile range in which outliers would sit outside of the box.

In the end multiple processed data frames were in the algorithm, this was to have some fully and some partially encoded for different analysis.

Chapter 3.2: Linking the Weather API

When linking the weather API learning needed to be done. To begin with how to call an API key was learnt. After getting and setting up the API Key the data was ready to be connect and be downloaded in JSON format (Open Weather, n/a). Once this was done the data was explored about to understand what data was available and how this could be structured into a pandas data frame. The weather was a very high number and at first it was assumed to be Fahrenheit, but it was too high, as such it was noted the weather was in kelvin, a function was made to turn temperature to Celsius. This was then converted to a data frame using pd.DataFrame. After the data was turned into a data frame it needed to then be re-run over the dates in the data and add the weather description and temperature to the main data frame. This would not work no matter what. It was then discovered that this cannot be done as the data set is too large and the number of requests exceeded the number of free requests allowed per day from the API (Open Weather, n/a).

Chapter 3.3: Data Analysis

Plotly, seaborn, and matplotlib are all libraries in python which aid in data visualisation. Seaborn is very quick to write in python and plotly makes the graphs seem more professional. At first seaborn was used and then as there was a bit of time at the end they were changed to plotly. The colour schemes are one colour to not over complicate the graphs, but they are different shades. The graphs are also appropriately named. Each graph has its own colour, so they are easily identifiable, best products is green and worst is red. They were done using cheat sheets created online to aid in remembering all the use cases for graphs depending on the library (Upadhyay, 2018)(Cotton, 2022).

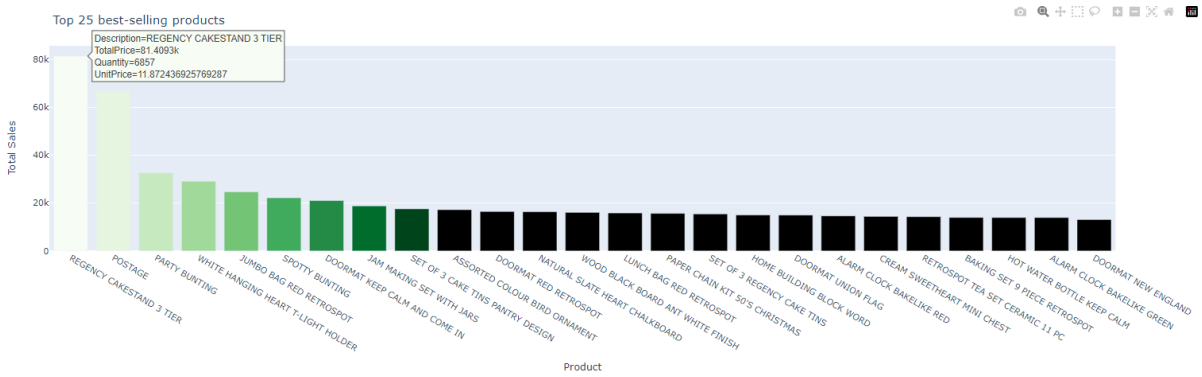


Figure 1 - Plotly interactive graph (25 Best products)

The InvoiceDate column was split into multiple columns, Date, Week, Month, Quarter, and Year. The column helps the number date or month in the row. The were all changed to date data type with pandat.datetime(). This was then used to identify which day of the week it is and be able to add and IsWeekend column, a IsHoliday column could have also been added for more insight given more time.

The next code is on encoding the data, the data is encoded using two ways, label and one hot encoding. Label encoding turns all the values to numbers so if there is a giraffe, zebra, and a lion in the data, giraffes would be 0, zebras 1, and lion 2. One hot encoding on the other hand makes all values 0 and 1, what this does is makes sure they are all of numbers going up in value but will be assigned 1's or a 0's in a specific way. In this case one hot encoding is better as the higher the number in label encoding the more weight it has to skewing the machine learning algorithm.

After this, code was written to check all the dates in which there was not a sale. The findings were that December surprising as it was the month with the most dates with an invoice being processed missing. This is surprising as it could be assumed December is the busiest time of year for e-commerce stores. This could be due to the bank holiday and invoices being sent out when employees were back from work.

A seaborn heatmap was created. This assisted in linear correlation analysis. This is a tool used in previous projects and effectively shows the relationship between columns. To do this the seaborn library was used and the data had to be encoded first as it only works with numerical values. As the weather could not be included in the data frame it could not be included in the heatmap as desired to see if there was a correlation. The heatmap was done twice once on the fully encoded data and once on a grouped data set. The heatmap did not show any notable correlations.

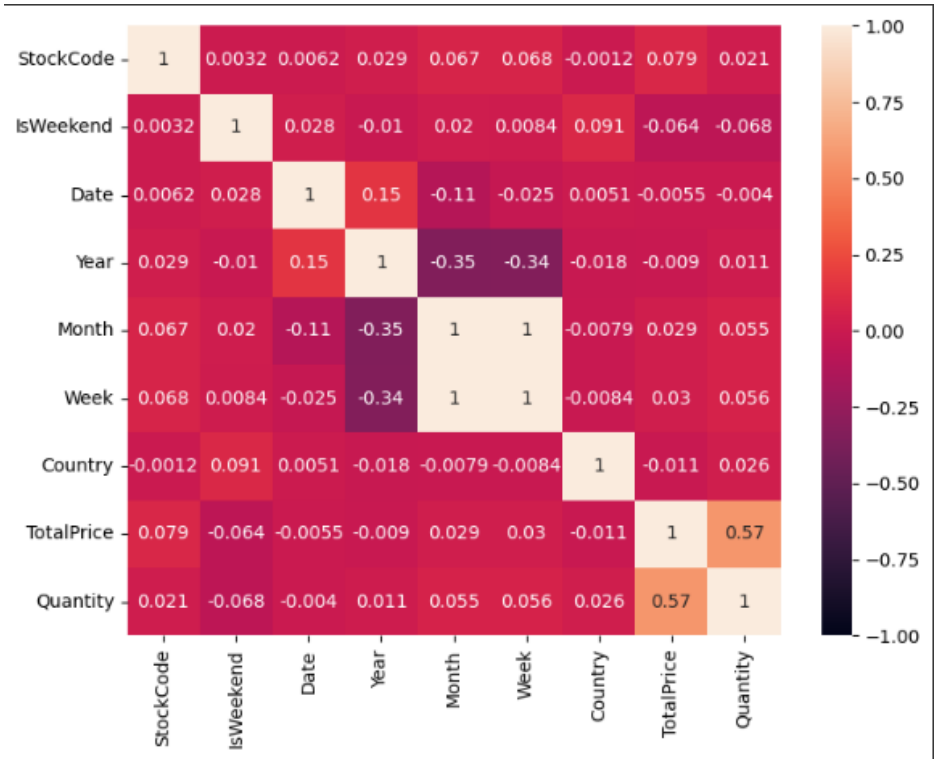


Figure 2 - Grouped Heatmap

The piece of that visualisation if monthly and weekly sales. This was done using seaborn and has the data for 2010 and 2011. There is not much data in 2010 which is why it is missing from the figure. The figure uses the year column for the line, the month or week for the axis and TotalSales on the y-axis. This is correct as the weekly sales add up to the monthly sales.

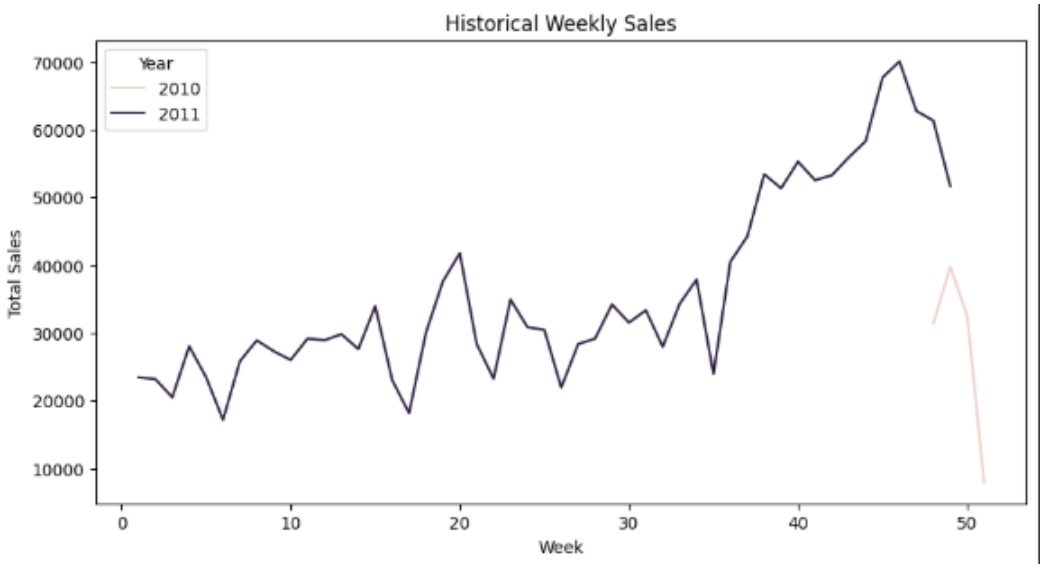


Figure 3 - Weekly Sales

Chapter 3.4: Machine Learning Algorithm 1 (Sales Forecasting)

Machine Learning algorithms used in this case are algorithms that can predict future trend from historic data or find hidden patterns. They are a form of artificial intelligence. For sales forecasting a supervised machine learning model was used. As regression-based algorithms were found to work best on sales forecasting (Ranjitha and Spandana, 2021), a few regression models were created to find the best one. To do this the data needed to be split into train and test. This was done using the hold out method with 75% training and 25% test. K-folds was originally going to be used but the model training time took too long, and the tool needs to be accessible and not take a long time to load without having to pay for use of upgraded hardware in Google's servers for Colab.

Three regression algorithms were used, Linear Regression, Decision Tree Regressor, and XGBRegressor. They were each trained and their RMSE and R^2 was calculated:

- Linear Regression RMSE was 9.726 and R^2 = 1.0
- Decision Tree Regression RMSE was 4.154 and R^2 = 0.948
- XGB Regression RMSE was 4.147 and R^2 = 0.948

As simple is best and the scores were quite good, the liner regression was chosen for forecasting.

When trying to visualise what the model predicted in terms of sales over a month, it was found to be very difficult due to using trial and error for the unknown rows in Appendix C. To add to this outlier removal using boxplots was not used at the beginning and when recapping an old project, it was found and was deemed an essential part of data processing.

As time had been taken up from finding the dataset, trying to connect the weather API, and learning about Market Basket Analysis, it did not leave a lot of time to visualise the forecast in different ways such as monthly, quarterly and weekly. In image the result can be shown, and the True and Predicted values merged perfectly. Upon review this could have been investigated with the other models as well to see how they all compare.

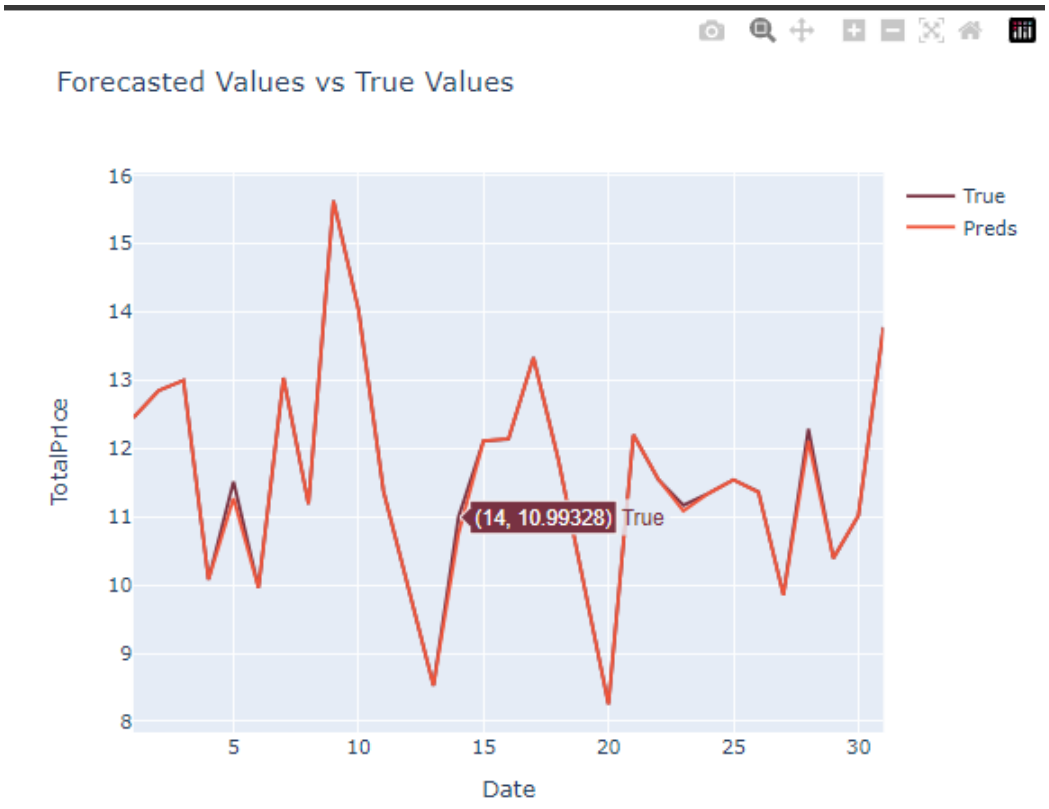


Figure 4 - Predicted vs True Sales (Sales Forecasting)

Chapter 3.5: Market basket Analysis (Market Basket Analysis)

Market basket analysis is used to help recommend products for consumers at the time of checkout or underneath an individual product. This can aid in boosting sales

for a company. This was done using an unsupervised machine learning algorithm, these are designed to help find patterns in data. It does this be checking the frequency of items paired together and the subsets frequency with other products. (Gupta, 2023).

This was something new added as opposed to adding in k-means clustering for products with little historic data to help predict future trends based on other products in similar clusters. This was changed to another type of unsupervised machine learning. This was to help improve employment after the project as clustering had been done multiple times for other projects.

Before starting the market basket analysis, the data was checked. Invoice rows only hold data for one product regardless of the description. The data was checked to then see how many rows have the same invoice number but different descriptions. The result showed there were duplicates so the Apriori algorithm was then investigated.

To perform the market basket analysis a second data frame was created to ensure the Description column had not been encoded, this was to allow ease of product identification when the results came back. In the results for the products commonly sold together, they seem to be products that are used for an afternoon tea, they are not vastly different from each other, so the recommendation algorithm performed as expected.

```
Top items frequently paired together:

PINK REGENCY TEACUP AND SAUCER, GREEN REGENCY TEACUP AND SAUCER, REGENCY CAKESTAND 3 TIER
ROSES REGENCY TEACUP AND SAUCER , PINK REGENCY TEACUP AND SAUCER
ROSES REGENCY TEACUP AND SAUCER , PINK REGENCY TEACUP AND SAUCER, REGENCY CAKESTAND 3 TIER
PINK REGENCY TEACUP AND SAUCER, GREEN REGENCY TEACUP AND SAUCER
PINK REGENCY TEACUP AND SAUCER, REGENCY CAKESTAND 3 TIER
GREEN REGENCY TEACUP AND SAUCER, REGENCY CAKESTAND 3 TIER
```

Figure 5 - Items frequently bought together (Market Basket Analysis)

Chapter 3.6: Chat GPT Extension

When the project was near the finishing stages, it was realised the OpenWeather API had a paywall as it was not working for over half the project duration. As such another objective was added at the end of the project. The project had to be able to evolve to a business's individual needs. Adding a Chat GPT extension allows for it to be

integrated into the code, read it, and explain it so new users can both learn and code quicker.

Chapter 3.7: Usability

The Google Colab interface is arguable one of the best integrated development environments (IDE), when comparing to other coding interfaces such as Visual Studio and Pycharm. They run all the code at once and do not allow for individual lines of code to run or for text boxes to placed inside like Google Colab. This means error handling code would have to incorporated such as the try catch method to ensure the whole code runs without breaking which would make it harder to understand the code for non-technical users in small businesses wanting to learn. Furthermore, these IDEs are not cloud based and the user would have to download the IDE and set it up which would make the code unusable for some small businesses as they may not have the needed expertise. A table of contents was also added in Google Colab in the left side with the 3 bullet points, this allows for easy navigation through the notebook. This can be viewed in Appendix C.

Chapter 3.8: Summary of Mythology and Build

The build was successful for the data set with some improvement that could be made on the forecasting. In regard to the build being successful for all businesses to just upload their dataset the project was not. This was because a survey needs to be done with multiple small e-commerce stores to see how their data is stored and what data is stored inside. The data set used had Amazon Fees in the rows. These fees could have been subtracted from the total revenue to gauge the total profit, but this could not be done as the cost for each product is not there, and for delivery it is not clear. When removed the accuracy of the algorithm improved, but others made the accuracy of the drop significantly, a list of the values that were not sales in the dataset can be found in Appendix C.

Chapter 4: Legal, Social and Ethical issues

When discussing legal, social and ethical requirements this is refers to considerations that must have been made for the prototype, these include things to consider which are against the law and things to consider which even if they are not illegal, they do not go against a user's or society's morals.

Chapter 4.1: Legal issues

Legal issues refer to anything that was done in the project that could cause the creator to be fined or acted upon as the project lead allowed it to directly disobey UK and EU laws (Oxford Learner's dictionary, N/A), anything encompassing the prototype must be in line with UK and EU data laws. GDPR stands for General Data Protection Regulations, and these are the European Union's laws on how to protect data (Gov, 2018). Furthermore, a company can be ISO certified, these are worldwide data protection standards to help companies have better protection on their data and prove their commitment as a business to data protection. Regarding breaking laws with the prototype, it does not as Google has an agreement for data transfers between the EU and USA. The data is deleted at runtime which further protects the data for businesses.

When looking at what legal requirements and adhering to them better, ensure GDPR is fully followed by ensuring data is stored securely in ideally UK servers. Google Collab adheres to this by allowing users to choose the location of where the data is stored (Google, n/a). Another legal requirement that was considered was liability, as the algorithm is a prototype that was mentioned to ensure any businesses using this algorithm are aware.

Chapter 4.2: Social Issues

When discussing social issues, these are the issues which affect a society and a wide range of people depending on the society. This can encompass anything that conflicts the society's morals and culture and ensures equal access to all (Oxford Learner's Dictionary, N/A). Social requirements that were present in this prototype involve ensuring no one can access other businesses data. As Google Collab is a collaborative environment the notebook could be changed by multiple users and make permanent edits to the master notebook. This could cause a mischievous user to ruin the algorithm or use it to download other business' data. To combat this the master notebook was locked so no edits could be made unless a local version was stored. Equal access to all was also another social issue, this was completed by finding a cloud-based environment which anyone with internet has access to.

Chapter 4.3: Ethical Issues

When looking at potential ethical issues, this refers to something the project can legally do but probably should not. Ethics is something a person believes in or follows, this can be influenced by a person's religion, culture, and moral choices (Oxford Learner's dictionary, N/A). When looking at ethical problems, the creator of the notebook must ensure the data visualised is accurate and the model algorithm is accurate, while this is a prototype data integrity is a must. Another ethical problem was ensuring the users know how the whole algorithm works and what their data would be put through, the explanation of how the algorithm works was able to explain how the code works to non-technical users. Another ethical problem would be selling company data, this can cause backlash and unhappiness from users as this is company data, they may want to keep private.

Summary

Looking back at the aim and the objectives set at the beginning of the project each will be evaluated.

1. Confidence in algorithm accuracy
    - The algorithm has a near perfect score and is simple.
2. Free and accessible
    - The code was done on Google Colab which is online and free to use.
3. Visually appropriate analysis
    - The graphs show the appropriate insight to help companies make strategic decisions.
4. Link weather data
    - The weather data could not be added due to needing to pay.
5. Learning and evolving environment
    - The code was all done and visualised on Google Colab so it can be edited in an easy-to-use IDE.
6. Market Basket Analysis
    - The Apriori algorithm works and displays the products that are recommended together.
7. Adding Chat GPT
    - The Chat GPT extension was added successfully and works.

Overall, there are 7 objectives, 6 if linking the weather data was excluded. This means only one objective was not fully completed due to a paywall. There was not enough time to spend a lot of time figuring out why the weather API would suddenly stop working. Alongside this learning new algorithms for market basket analysis also took up more time then needed, as data processing and cleaning is the most time-consuming process of the project, this was checked over after it was completed, and changes had to be made to improve the accuracy of the algorithm.

Contributions and limitations

The project makes some notable contributions to the sector of sales analysis and forecasting for small e-commerce businesses. It addresses the need of data tools for businesses as they are costly and time consuming to make, it did this by creating an accessible and free too which incorporates AI to help small business owners or employees to use the notebook. The tools help bridge the gap between expensive data tools and small e-commerce businesses. The prototype also shows sales and trend in the data which are beneficial in making business decisions.

However, the project does have some limitations. It could not include the weather data which did not allow for the correlation analysis (heatmap) against the rest of the data. This could prevent potential patterns from being discovered for both creating graphs and forecasting. The tool may be too much for non-technical users as it can be intimidating seeing a lot of code along with the need to dedicate time and brain power to evolving the prototype.

Overall, the projects contributions provide an accessible and affordable prototype to experiment with or use as a business tool that visualises and forecasts data and provides usable insight. Future evolutions of this could fix the limitations.

Improvements to prototype

In terms of what could be done better given more time there are many. To begin with a survey with multiple e-commerce stores could be done to see how they capture data to format the processing accordingly in the notebook to make it usable for any data set. The weather data could be explored to find other ways of getting that data. The code could be organised better so that all the analysis is in one easy to find location. More graphs could be made for company specific metrics, to do this the code needs to be able to change in a tailored manner for each business, which it is able to do. This could pose a problem to non-technical users who could accidentally break the code.

If they do, they can redownload the notebook but thoughts on how to handle this would be done as well. And the final mention on improvements would be adding different forms of market basket analysis in different countries so companies can trial each and see which was the most effective.

References

ABEYSEKARA, K, T. and RUPASINGHE, S., 2019. Analysis of Influential Factors for Inventory Forecasting Systems. 2019 IEEE 5th International Conference for Convergence in Technology (I2CT). 29-31 March 2019. Bombay, India. Piscataway, N.J.: Institute of Electrical and Electronics Engineers. pp. 1-4. Available from: https://ieeexplore.ieee.org/document/9033725 [Accessed 8 November 2022].

BASRI, A, H., 2022., *E-commerce $ Forecasting Fbprophet + Optuna* [online]. Available from: 📈 E-Commerce 💲 Forecasting Fbprophet + Optuna 💰 | Kaggle [Accessed 15 February 2023].

BENHAMIDA, F, Z. KADDOURI, O. OUHROUCHE, T. BENAICHOUCHE, M. CASADO-MANSILLA, D. and LOPEZ-DE-LPINA, D., 2020. Stock&Buy: A New Demand Forecasting Tool for Inventory Control. 2020 5th International Conference on Smart and Sustainable Technologies (SpliTech), 23-26 September 2020, Split, Croatia. Piscataway, N.J.: Institute of Electrical and Electronics Engineers. Pp. 1-6. Available from: https://ieeexplore.ieee.org/document/9243824 [Accessed 07 November 2022].

CLEMEMT, F., KAUR, A., SEDGHI, M., KRISHNASWAMY, D. and PUNITHAKUMAR, K. 2020. Interactive Data Driven Visualization for Covid-19 with Trends, Analysis and Forecasting. 2020 24th International Conference Information Visualization (IV). 07– 11 September 2020. Melbourne, Australia. Piscataway, N.J.: Institute of Electrical and Electronics Engineers. pp. 593-598. Available from: https://ieeexplore.ieee.org/stamp/stamp.jsp?tp=&arnumber=9373291 [Accessed 11 of November 2022].

COTTON, R., 2022., *Plotly Express Cheat Sheet* [online]. Available from: https://www.datacamp.com/cheat-sheet/plotly-express-cheat-sheet [Accessed 12 April 2023].

DEVELOPERTOOLS., 2023. *ChatGPT for Google Colab* [online]. Available from: https://chrome.google.com/webstore/detail/chatgpt-for-google-colab/dfhfeifekpgapdlhfakecbbinnnfoohh [Accessed 05 May 2023].

OPENWEATHER., N/A. *Detailed Pricing* [online]. Available from: https://openweathermap.org/full-price [Accessed 20 April 2023].

OPENWEATHER., N/A. *History API* [online]. Available from: https://openweathermap.org/history [Accessed 25 February 2023].

OXFORD LEARNERS DICTIONARY., N/A. ethics [online]. Available from: https://www.oxfordlearnersdictionaries.com/definition/american_english/ethic [Accessed 01 April 2023].

OXFORD LEARNERS DICTIONARY., N/A. illegal [online]. Available from: https://www.oxfordlearnersdictionaries.com/definition/english/illegal_1 [Accessed 01 April 2023].

OXFORD LEARNERS DICTIONARY., N/A. legal [online]. Available from: https://www.oxfordlearnersdictionaries.com/definition/english/legal?q=legal [Accessed 01 April 2023].

OXFORD LEARNERS DICTIONARY., N/A. society [online]. Available from: society noun - Definition, pictures, pronunciation and usage notes | Oxford Advanced Learner's Dictionary at OxfordLearnersDictionaries.com [Accessed 01 April 2023].

GOOGLE., n/a. *Data regions: Choose a geographic location for your data* [online]. Available from: https://support.google.com/a/answer/7630496?hl=en [Accessed 01 March 2023].

GOV, 2018., *Data Protection* [online]. Available from: https://www.gov.uk/data-protection#:~:text=The%20Data%20Protection%20Act%202018%20is%20the%20U

K's%20implementation%20of,used%20fairly%2C%20lawfully%20and%20transparently [Accessed 01 April 2023].

GUPTA, A., 2023., I*mplementing Apriori algorithm in Python* [online]. Available from: https://www.geeksforgeeks.org/implementing-apriori-algorithm-in-python/#article-meta-div [Accessed 01 May 2023].

HOSSAIN, M., SATTAR, S, M, H, A., PAUL, K, M., 2019. Market Basket Analysis Using Apriori and FP Growth Algorithm, In: *2019 22nd International Conference on Computer and Information Technology (ICCIT)*, 18-20 December 2019. Dhaka, Bangladesh, N.J.: Institute of Electrical and Electronics Engineers 2019, pp. 1-6.

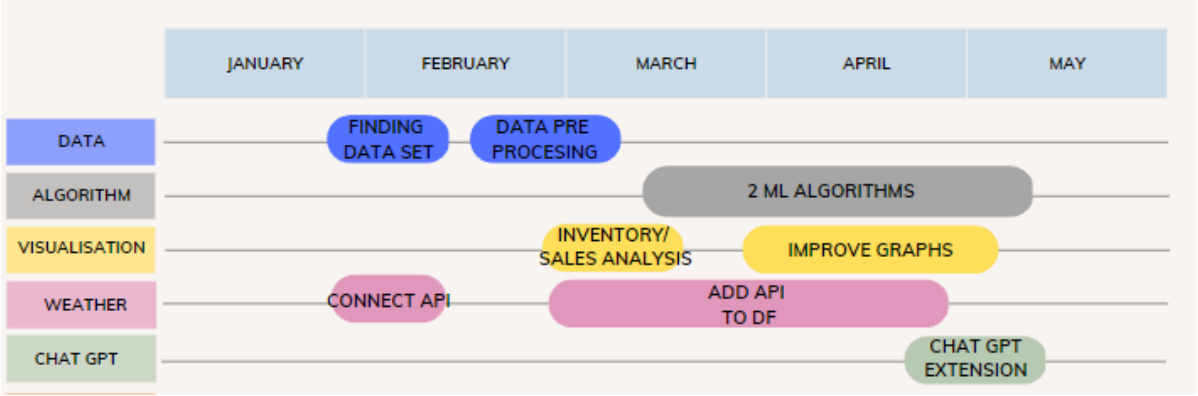
PUNEET,. SHARMA, S. DEEPIKA, D. and SINGH, G., 2021. Intelligent Warehouse Stocking Using Machine Learning. 2021 International Conference of Mobile Networks and Wireless Communications (ICMNWC), 03-04 December 2021, Tumkur, Karnataka, India. Piscataway, N.J.: Institute of Electrical and Electronics Engineers. pp. 1-6. Available from: https://ieeexplore.ieee.org/document/9688530 [Accessed 06 November 2022].

RANJITHA, P., and SPANDANA, M., 2021. Predictive Analysis for Big Mart Sales Using Machine Learning Algorithms. 5th International Conference on Intelligent Computing and Control Systems (ICICCS). 06-08 May 2021. Madurai, India. Piscataway, N.J.: Institute of Electrical and Electronics Engineers. pp. 1416-1421. Available from: https://ieeexplore.ieee.org/abstract/document/9432109/metrics#metrics [Accessed 11 November 2022].

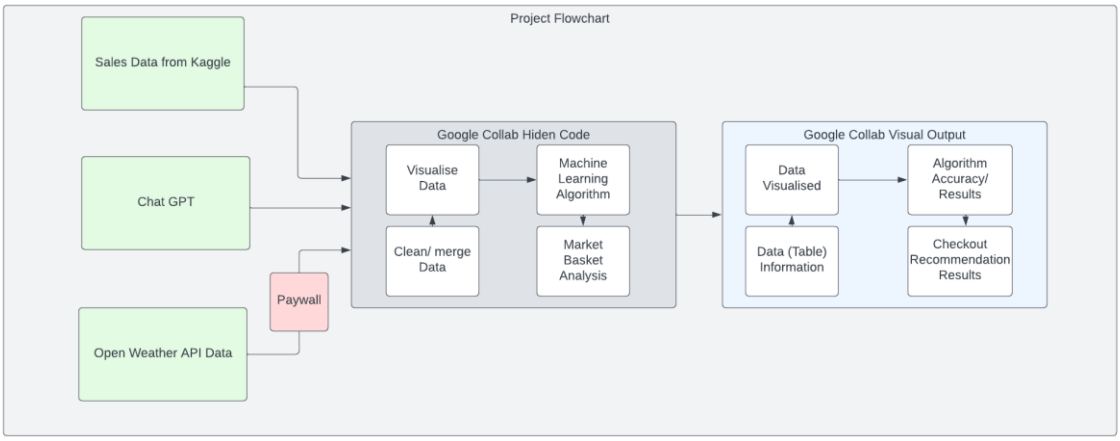
STOCK&BUY,. N/A. *Pricing tailored for businesses of all sizes.* Available from: [14 days Free Trial] Inventory Management Software | Stock&Buy (stockandbuy.com) [Accessed 11 November 2022].

UPADHYAY, N., 2018. *Python seaborn cheat_sheet* [online]. Available from: https://www.slideshare.net/nishantupadhyaysbi/python-seaborn-cheatsheet [Accessed 18 March 2023].

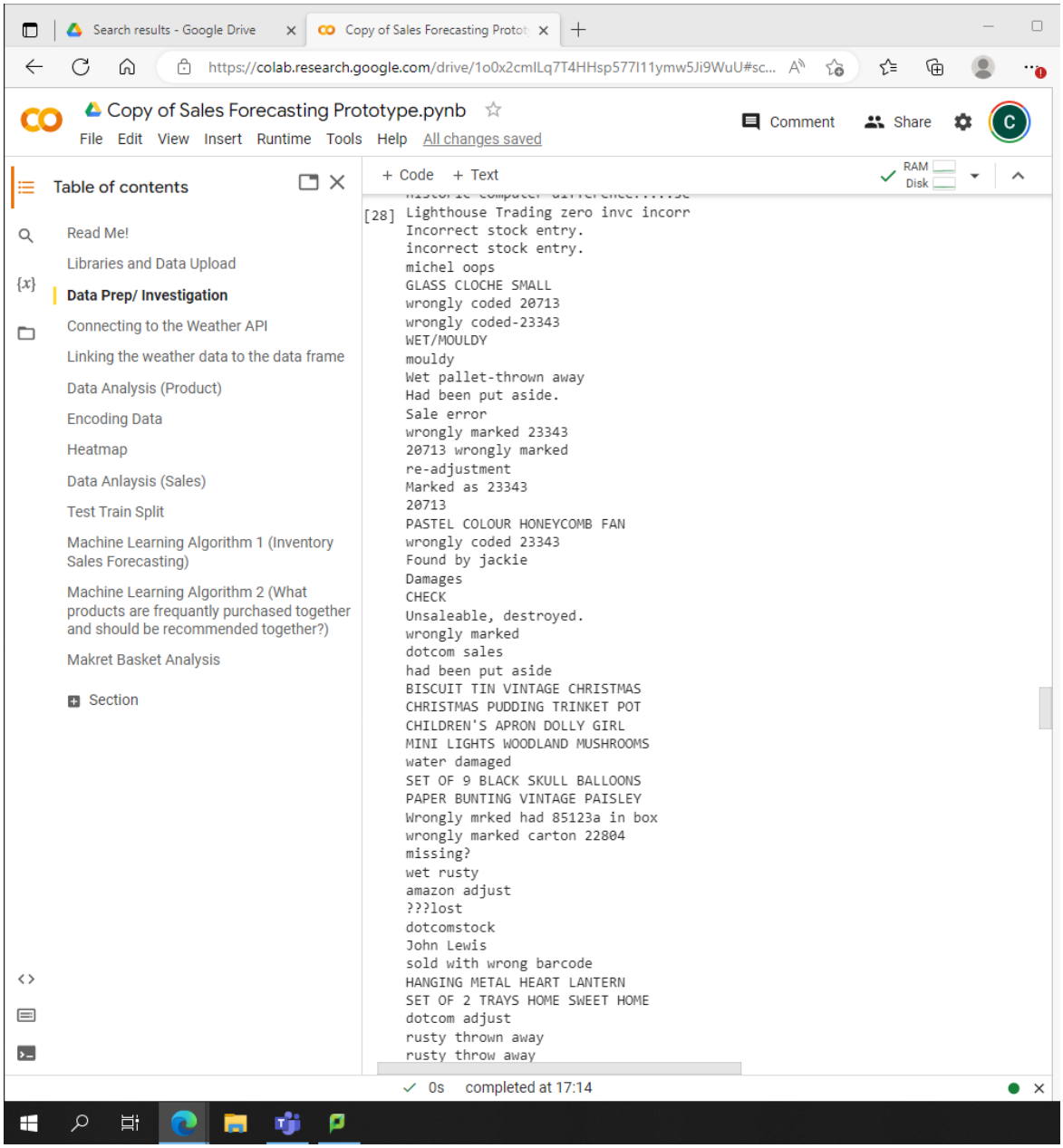
Appendix A: Prototype plan



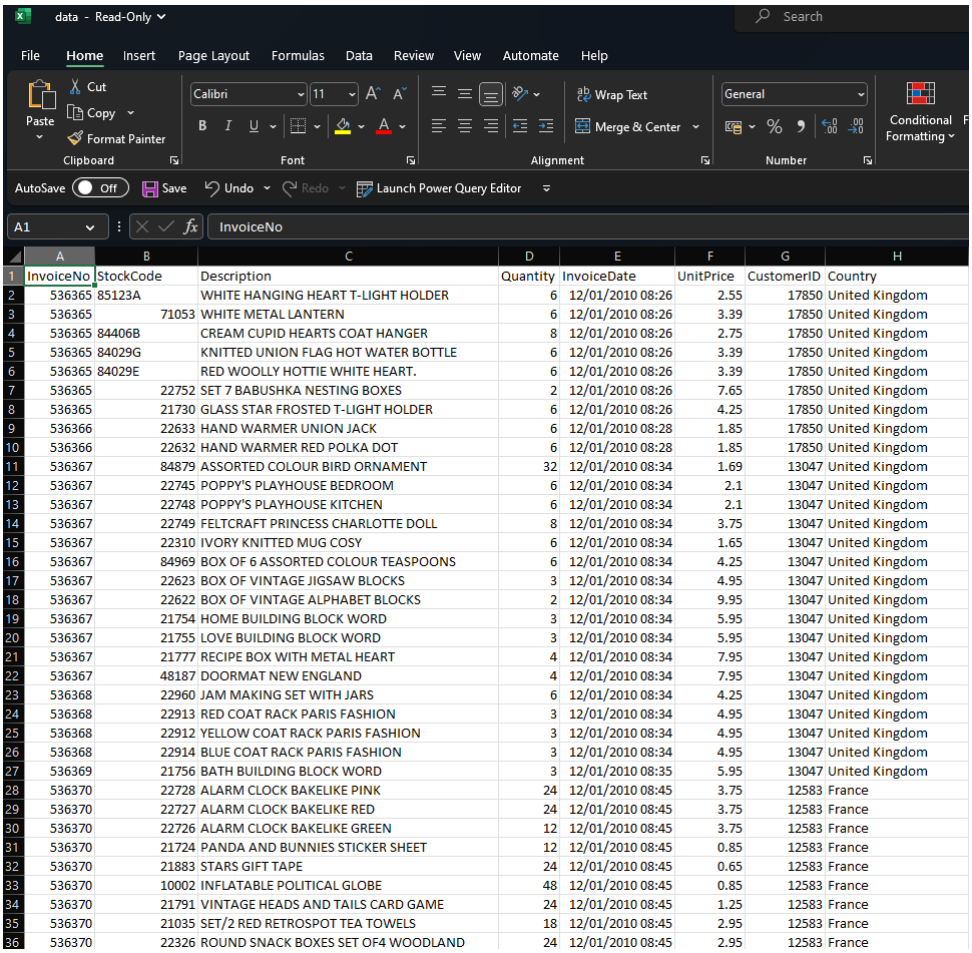
Appendix B: Project Flow



Appendix C: Rows in Data



Appendix D: Raw Data



data - Read-Only

	A	B	C	D	E	F	G	H
1	InvoiceNo	StockCode	Description	Quantity	InvoiceDate	UnitPrice	CustomerID	Country
2	536365	85123A	WHITE HANGING HEART T-LIGHT HOLDER	6	12/01/2010 08:26	2.55	17850	United Kingdom
3	536365	71053	WHITE METAL LANTERN	6	12/01/2010 08:26	3.39	17850	United Kingdom
4	536365	84406B	CREAM CUPID HEARTS COAT HANGER	8	12/01/2010 08:26	2.75	17850	United Kingdom
5	536365	84029G	KNITTED UNION FLAG HOT WATER BOTTLE	6	12/01/2010 08:26	3.39	17850	United Kingdom
6	536365	84029E	RED WOOLLY HOTTIE WHITE HEART.	6	12/01/2010 08:26	3.39	17850	United Kingdom
7	536365	22752	SET 7 BABUSHKA NESTING BOXES	2	12/01/2010 08:26	7.65	17850	United Kingdom
8	536365	21730	GLASS STAR FROSTED T-LIGHT HOLDER	6	12/01/2010 08:26	4.25	17850	United Kingdom
9	536366	22633	HAND WARMER UNION JACK	6	12/01/2010 08:28	1.85	17850	United Kingdom
10	536366	22632	HAND WARMER RED POLKA DOT	6	12/01/2010 08:28	1.85	17850	United Kingdom
11	536367	84879	ASSORTED COLOUR BIRD ORNAMENT	32	12/01/2010 08:34	1.69	13047	United Kingdom
12	536367	22745	POPPY'S PLAYHOUSE BEDROOM	6	12/01/2010 08:34	2.1	13047	United Kingdom
13	536367	22748	POPPY'S PLAYHOUSE KITCHEN	6	12/01/2010 08:34	2.1	13047	United Kingdom
14	536367	22749	FELTCRAFT PRINCESS CHARLOTTE DOLL	8	12/01/2010 08:34	3.75	13047	United Kingdom
15	536367	22310	IVORY KNITTED MUG COSY	6	12/01/2010 08:34	1.65	13047	United Kingdom
16	536367	84969	BOX OF 6 ASSORTED COLOUR TEASPOONS	6	12/01/2010 08:34	4.25	13047	United Kingdom
17	536367	22623	BOX OF VINTAGE JIGSAW BLOCKS	3	12/01/2010 08:34	4.95	13047	United Kingdom
18	536367	22622	BOX OF VINTAGE ALPHABET BLOCKS	2	12/01/2010 08:34	9.95	13047	United Kingdom
19	536367	21754	HOME BUILDING BLOCK WORD	3	12/01/2010 08:34	5.95	13047	United Kingdom
20	536367	21755	LOVE BUILDING BLOCK WORD	3	12/01/2010 08:34	5.95	13047	United Kingdom
21	536367	21777	RECIPE BOX WITH METAL HEART	4	12/01/2010 08:34	7.95	13047	United Kingdom
22	536367	48187	DOORMAT NEW ENGLAND	4	12/01/2010 08:34	7.95	13047	United Kingdom
23	536368	22960	JAM MAKING SET WITH JARS	6	12/01/2010 08:34	4.25	13047	United Kingdom
24	536368	22913	RED COAT RACK PARIS FASHION	3	12/01/2010 08:34	4.95	13047	United Kingdom
25	536368	22912	YELLOW COAT RACK PARIS FASHION	3	12/01/2010 08:34	4.95	13047	United Kingdom
26	536368	22914	BLUE COAT RACK PARIS FASHION	3	12/01/2010 08:34	4.95	13047	United Kingdom
27	536369	21756	BATH BUILDING BLOCK WORD	3	12/01/2010 08:35	5.95	13047	United Kingdom
28	536370	22728	ALARM CLOCK BAKELIKE PINK	24	12/01/2010 08:45	3.75	12583	France
29	536370	22727	ALARM CLOCK BAKELIKE RED	24	12/01/2010 08:45	3.75	12583	France
30	536370	22726	ALARM CLOCK BAKELIKE GREEN	12	12/01/2010 08:45	3.75	12583	France
31	536370	21724	PANDA AND BUNNIES STICKER SHEET	12	12/01/2010 08:45	0.85	12583	France
32	536370	21883	STARS GIFT TAPE	24	12/01/2010 08:45	0.65	12583	France
33	536370	10002	INFLATABLE POLITICAL GLOBE	48	12/01/2010 08:45	0.85	12583	France
34	536370	21791	VINTAGE HEADS AND TAILS CARD GAME	24	12/01/2010 08:45	1.25	12583	France
35	536370	21035	SET/2 RED RETROSPOT TEA TOWELS	18	12/01/2010 08:45	2.95	12583	France
36	536370	22326	ROUND SNACK BOXES SET OF4 WOODLAND	24	12/01/2010 08:45	2.95	12583	France

Appendix E: Project Artefact

https://colab.research.google.com/drive/1o0x2cmlLq7T4HHsp577I11ymw5Ji9WuU?usp=sharing