

Prediction and analysis of online sales in retail industry

Domain: Retail Industry – Online Sales

Introduction

Studying the behavior of the online retail sales is one of the key factor for the business to expand. Over a decade of time, the demand for online sales has seen a tremendous rise in terms of customers and business. When compared to traditional methods of retail, online retail stands out in various ways. Understanding the customer purchase history to help effectively promote future product offerings to customer-base is one of the key solution. It helps establish successful product types and improve future offerings.

Background

A large variety of choices to choose from, firsthand information on price differences and discreetness of the activity are few of the main reasons that tempts people to buy things online. Eight in 10 Americans are now shopping online, according to a new study from Pew Research. More users these days prefer paying through their mobile phones than cash. Online sales is one such safe and convenient gateway that allows them to depend on this means of payment.

Dataset Description

The dataset focusses data for a period of two years. The data extracted maps to the online sales, predicting the change in sales. It analyzes which of the items are frequently purchased together amongst many other conclusions that can draw from the dataset. The preprocessed data consists of nearly 0.3 million rows without any null values.

Variable Name	Description	Data Type
InvoiceNo	Invoice Number	Nominal
StockCode	Product code	Nominal
Description	Description of the item	Nominal
Quantity	The number of products per transaction	Numeric
InvoiceDate	Invoice Data and Time	Numeric
UnitPrice	Unit price of each item	Numeric
CustomerID	Customer Number	Nominal
Country	Name of the country	Nominal
Total Price	Total Price of the order	Numeric

	A	B	C	D	E	F	G	H	I
1	InvoiceNo	StockCode	Description	Quantity	InvoiceDate	UnitPrice	CustomerID	Country	Total Price
287677	563425	23308	SET OF 60 VINTAGE LEAF CAKE CASES	4	_2011August	0.55	16752	United Kingdom	2.2
287678	563425	23296	SET OF 6 TEA TIME BAKING CASES	4	_2011August	1.25	16752	United Kingdom	5
287679	563425	23293	SET OF 12 FAIRY CAKE BAKING CASES	4	_2011August	0.83	16752	United Kingdom	3.32
287680	563425	23189	SET OF 12 FORK CANDLES	4	_2011August	2.89	16752	United Kingdom	11.56
287681	563425	23301	GARDENERS KNEELING PAD KEEP CALM	4	_2011August	1.45	16752	United Kingdom	5.8
287682	563425	23300	GARDENERS KNEELING PAD CUP OF TEA	4	_2011August	1.45	16752	United Kingdom	5.8
287683	563426	22080	RIBBON REEL POLKADOTS	4	_2011August	1.65	16752	United Kingdom	6.6
287684	563426	23306	SET OF 36 DOILIES PANTRY DESIGN	4	_2011August	1.45	16753	United Kingdom	5.8
287685	563427	17003	BROCADE RING PURSE	4	_2011August	0.29	16753	United Kingdom	1.16
287686	563427	20777	CHRYSANTHEMUM NOTEBOOK	4	_2011August	1.65	16753	United Kingdom	6.6
287687	563427	22560	TRADITIONAL MODELLING CLAY	4	_2011August	1.25	16753	United Kingdom	5
287688	563427	22491	PACK OF 12 COLOURED PENCILS	4	_2011August	0.85	16753	United Kingdom	3.4
287689	563427	22544	MINI JIGSAW SPACEBOY	4	_2011August	0.42	16753	United Kingdom	1.68
287690	563427	22559	SEASIDE FLYING DISC	4	_2011August	1.25	16753	United Kingdom	5
287691	563427	22545	MINI JIGSAW BUNNIES	4	_2011August	0.42	16753	United Kingdom	1.68
287692	563427	20658	RED RETROSPOT LUGGAGE TAG	4	_2011August	1.25	16753	United Kingdom	5
287693	563427	22278	OVERNIGHT BAG VINTAGE ROSE PAISLEY	4	_2011August	4.95	16753	United Kingdom	19.8

	A	B	C	D	E	F	G	H	I
1	InvoiceNo	StockCode	Description	Quantity	InvoiceDate	UnitPrice	CustomerID	Country	Total Price
2	537240	21980	PACK OF 12 RED RETROSPOT TISSUES	1	_2010December	0.85	12477	United Kingdom	0.85
3	537240	21982	PACK OF 12 SUKI TISSUES	1	_2010December	0.85	12477	United Kingdom	0.85
4	537240	21985	PACK OF 12 HEARTS DESIGN TISSUES	1	_2010December	0.85	12477	United Kingdom	0.85
5	537240	21990	MODERN FLORAL STATIONERY SET	1	_2010December	2.51	12477	United Kingdom	2.51
6	537240	21991	BOHEMIAN COLLAGE STATIONERY SET	1	_2010December	2.51	12478	United Kingdom	2.51
7	537240	21992	VINTAGE PAISLEY STATIONERY SET	1	_2010December	2.51	12478	United Kingdom	2.51
8	537240	21993	FLORAL FOLK STATIONERY SET	1	_2010December	2.51	12478	United Kingdom	2.51
9	537240	22027	TEA PARTY BIRTHDAY CARD	1	_2010December	0.85	12478	United Kingdom	0.85
10	537240	22042	CHRISTMAS CARD SINGING ANGEL	1	_2010December	0.43	12478	United Kingdom	0.43
11	537240	22059	CERAMIC STRAWBERRY DESIGN MUG	1	_2010December	3.36	12478	United Kingdom	3.36
12	537240	22064	PINK DOUGHNUT TRINKET POT	1	_2010December	3.36	12478	United Kingdom	3.36
13	537240	22065	CHRISTMAS PUDDING TRINKET POT	1	_2010December	3.36	12478	United Kingdom	3.36

Outcomes of the Project

- 1) Items customers have purchased together frequently
- 2) Who are the most loyal customers?
- 3) Percentage change in total sales over a period of time
- 4) Number of sales and customers on a particular day.
- 5) Which item was widely sold in which country?
- 6) Predicting the sale of the most widely sold item in a particular location for next two months.

Resources Needed:

1. Weka

Approaches in the Project:

- Prediction Algorithm
- Frequent pattern tree Algorithm

Timeline

We would like to submit the entire work by end of March

References:

<https://techcrunch.com/2016/12/19/79-percent-of-americans-now-shop-online-but-its-cost-more-than-convenience-that-sways-them/>

<http://archive.ics.uci.edu/ml/datasets/Online+Retail>

<https://www.yellowfinbi.com/blog/2010/10/yfcommunitynews-data-analysis-for-the-retail-industry-part-1-100068>