*Analisis Association Rule pada Online Market Dataset Menggunakan Apriori Algorithm*

Nadya Yuniar Desi Prameswari1), Santi Wulan P2), Kartika Fithriasari3) dan Irhamah4)

Departemen Statistika, Fakultas Sains dan Analitika Data

Institut Teknologi Sepuluh Nopember (ITS)

Jl. Arief Rahman Hakim, Surabaya 60111 Indonesia

*e-mail*: [yuniarnadya4@gmail.com](mailto:yuniarnadya4@gmail.com)1), [santi\_wp@statistika.its.ac.id](mailto:santi_wp@statistika.its.ac.id)2), [kartika\_f@statistika.its.ac.id](mailto:kartika_f@statistika.its.ac.id)3), dan irhamah@statistika.its.ac.id4)

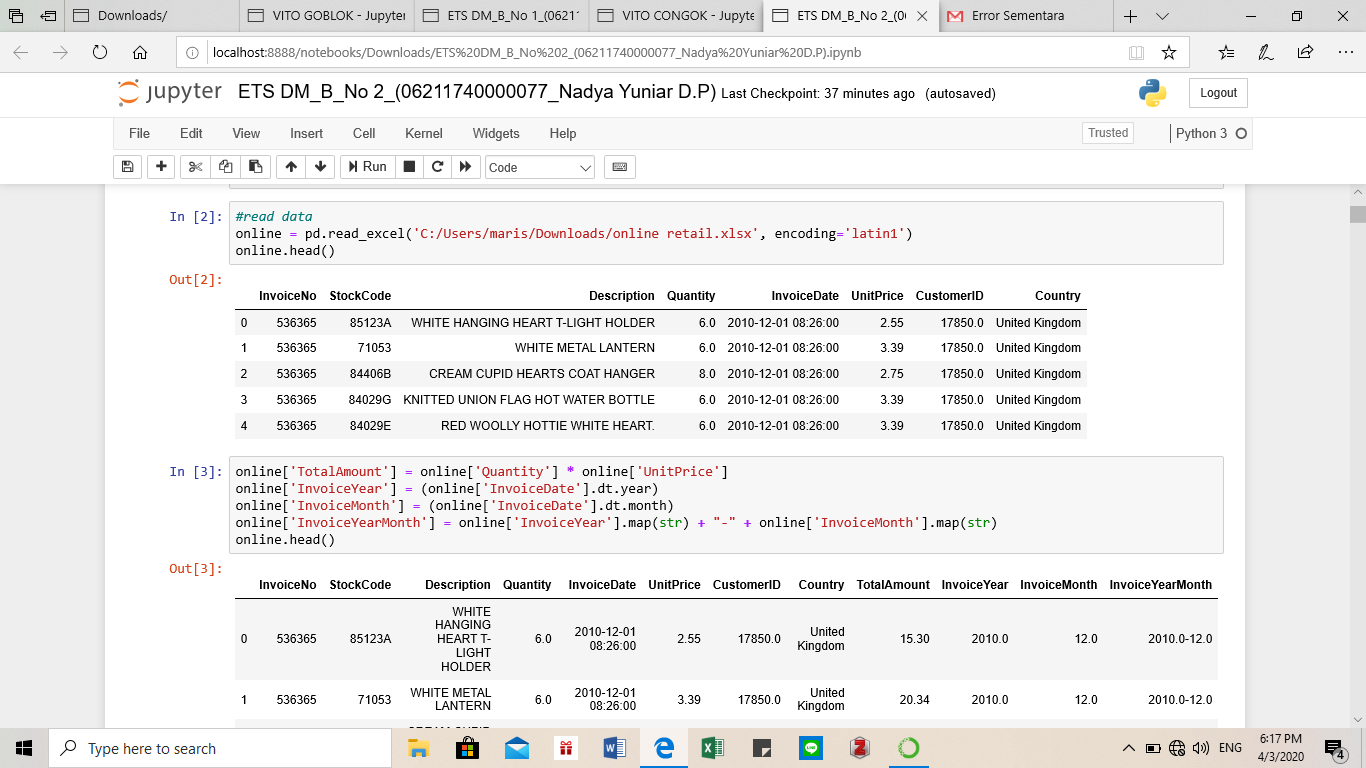
Algoritma Apriori merupakan algoritma pengambilan data dengan aturan asosiatif (Association Rule) untuk menentukan hubungan asosiatif suatu kombinasi item. Analisis pola frekuensi tinggi dengan algoritma apriori mencari kombinasi item yang memenuhi syarat minimum dari nilai support dalam basis data. Dalam menentukan association rule, terdapat suatu interestingness measure (ukuran kepercayaan) yang didapatkan dari hasil pengolahan data dengan perhitungan tertentu umunya ada 3 ukuran, yaitu:

1. Support, yoitu suatu ukuran yang menunjukkan seberapa besar tingkat dominasi suatu item/itemset dari keseluruhan transaksi.
2. Confidence, yaitu suatu ukuran yang menunjukkan hubungan antar dua item secara conditional (missal, seberapa sering item B dibeli jika orang membeli item A)
3. Lift, yaitu suatu ukuran yang menunjukkan besarnya kemungkinan dua item dapat dibeli secara bersamaan.

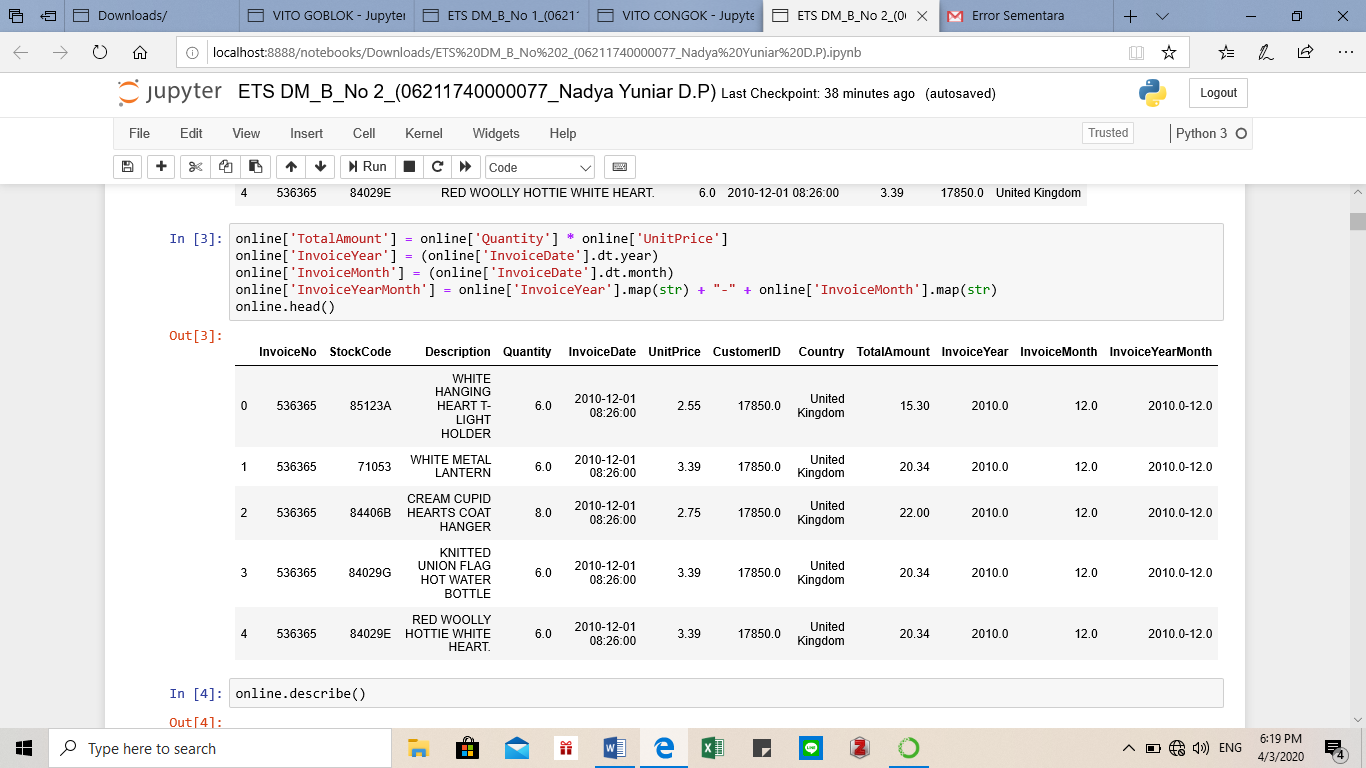
Pembentukan aturan asosiasi setelah semmua pola frekuensi tinggi ditemukan, barulah dicari aturan asosiasi yang memenuhi syarat minimum untuk confidence.

Pada penelitian kali ini akan dilakukan analisis *association rule* pada online market dataset di negara United Kingdom dan didapati langkah-langkah dan hasil analisis sebagai berikut:

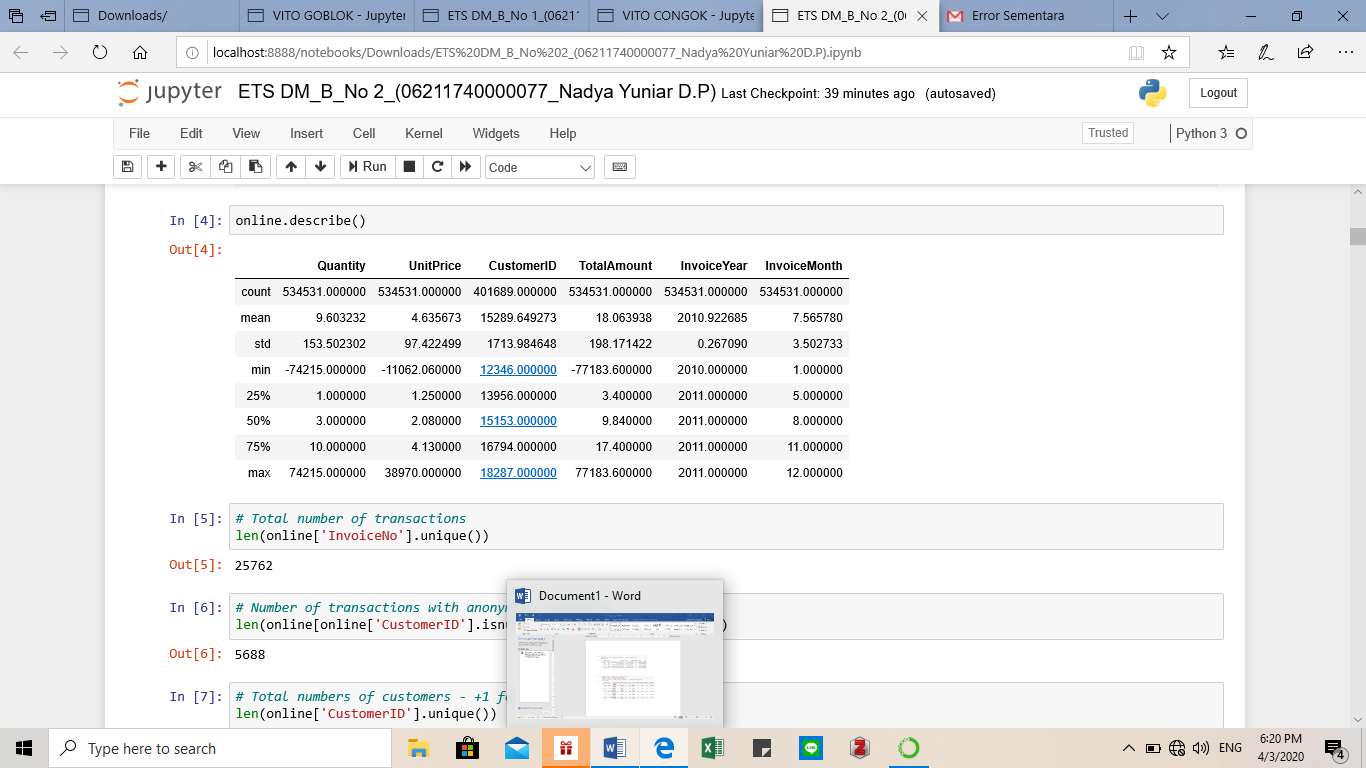
1. Membaca data pada Python



1. Menambahkan *field* atau variabel TotalAmount, InvoiceYear, InvoiceMonth dan InvoiceYearMonth

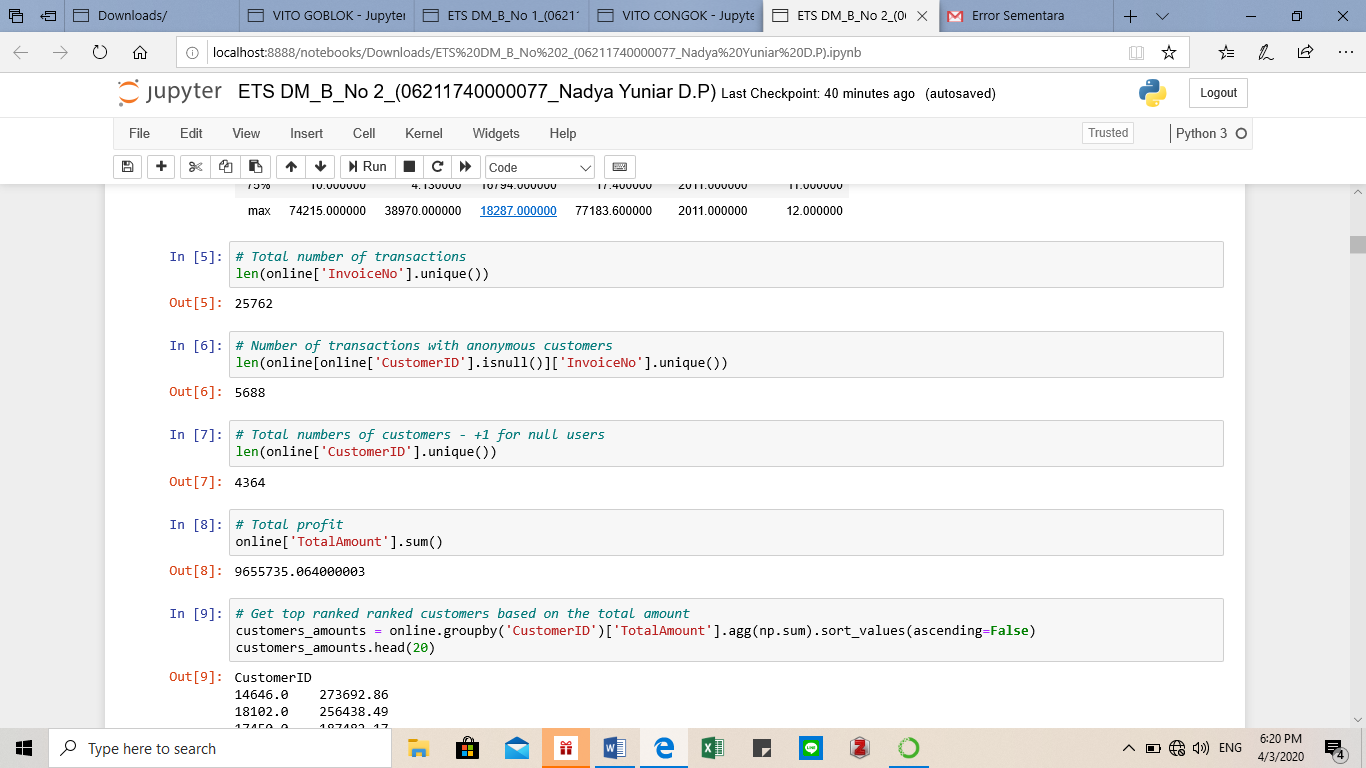


1. Menghitung statistika deskriptif dari setiap variabel



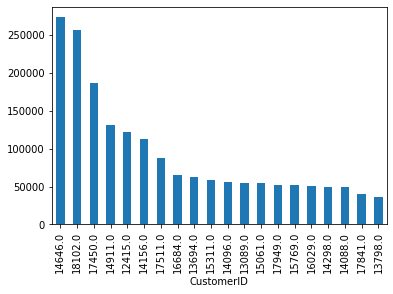
Dari gambar di atas, dapat diketahui nilai dari statistika deskriptif dari tiap variabel mulai dari mean, median hingga nilai minimum dan maksimum.

1. Menghitung Total transaction



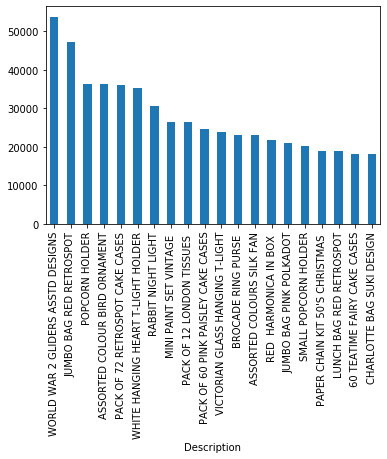
Dari gambar di atas dapat diketahui bahwa jumlah transaksi ada sebanyak 25.762 transaksi. Jumlah transaksi dari pelanggan tak dikenal terdapat 5.688 transaksi. Jumlah pelanggan ada sebanyak 4.364 transaksi dan total profit yang didapat sebanyak 9655735,064000003

1. Histogram dari ranking pelanggan berdasarkan Total Amount



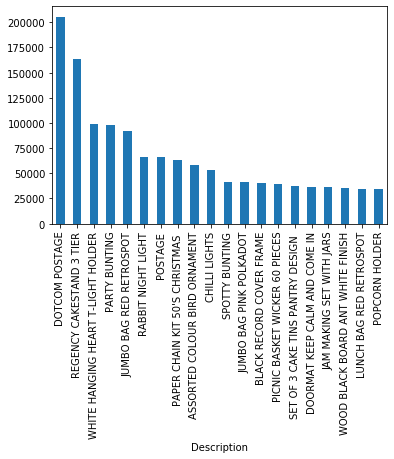
Dari histogram di atas dapat disimpulkan bahwa pelanggan yang menghabiskan banyak uang saat berbelanja ialah pelanggan dengan CustomerID 14646.0 dan yang paling sedikit ialah pelanggan dengan CustomerID 13798.0

1. Histogram dari banyaknya barang yang lebih cepat habis berdasarkan kuantitas



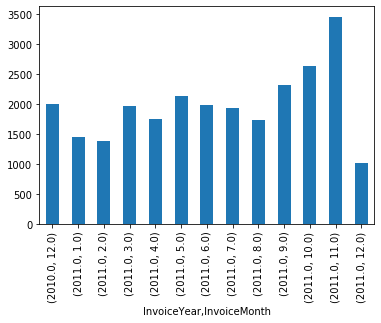
Barang yang cepat habis ialah World War 2 Gliders Asstd Design dengan kuantitas sebanyak 53751. Sedangkan dengan kuantitas paling sedikit yaitu Charlotte Bag Suki Design dengan sebanyak 17977.

1. Barang yang cepat habis berdasarkan total jumlah



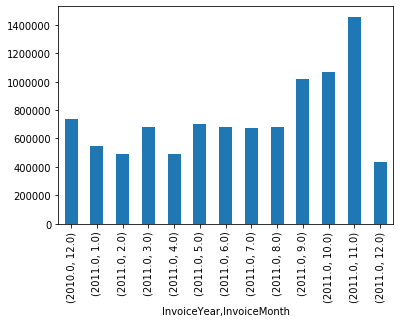
Barang yang lebih cepat habis berdasarkan total pembelian ialah Dotcom Postage dengan jumlah nominal pembelian 205553,47 dan yang paling sedikit ialah Popcorn Holder dengan nominal pembelian sebesar 33969,49.

1. Eksplorasi nomor invoice berdasarkan bulan



Pada gambar di atas, jumlah nomor invoice yang masuk pada bulan 12 tahun 2010 ada sebanyak 1996, pada 2011 bulan Januari ada sebanyak 1448 dan pada bulan Desember tahun 2011 ada sebanyak 1011 nomor invoice yang masuk.

1. Eksplorasi Total Amount berdasarkan bulan



Berdasarkan histogram di atas, Total Amount terbanyak terjadi pada bulan 11 tahun 2011 dan yang paling sedikit pada bulan Desember 2011 dengan Total Amount sebesar 431427,970.

1. Eksplorasi Total Amount berdasarkan negara

Country

United Kingdom 8.114621e+06

Netherlands 2.786464e+05

EIRE 2.610493e+05

Germany 2.185575e+05

France 1.959105e+05

Australia 1.353302e+05

Switzerland 5.589583e+04

Spain 5.411026e+04

Belgium 4.041901e+04

Sweden 3.637415e+04

Norway 3.490813e+04

Japan 3.461606e+04

Portugal 2.927234e+04

Finland 2.222669e+04

Channel Islands 1.995054e+04

Denmark 1.866518e+04

Italy 1.650603e+04

Cyprus 1.279131e+04

Austria 1.014928e+04

Hong Kong 1.003784e+04

Singapore 9.054690e+03

Israel 7.867420e+03

Poland 7.193340e+03

Unspecified 4.746650e+03

Greece 4.644820e+03

Iceland 4.299800e+03

Canada 3.599680e+03

Malta 2.503190e+03

United Arab Emirates 1.877080e+03

USA 1.730920e+03

Lebanon 1.693880e+03

Lithuania 1.661060e+03

European Community 1.291750e+03

Brazil 1.143600e+03

RSA 1.002310e+03

Czech Republic 7.077200e+02

Bahrain 5.484000e+02

Saudi Arabia 1.311700e+02

Name: TotalAmount, dtype: float64

Berdasarkan banyaknya total belanja di tiap negara, total belanja tertinggi diperingkati oleh United Kingdom dan yang terendah negara Saudi Arabia.

1. Eksplorasi Total Amount berdasarkan negara

Country

United Kingdom 23332

Germany 599

France 458

EIRE 358

Belgium 118

Spain 104

Netherlands 100

Switzerland 73

Portugal 71

Australia 69

Italy 55

Finland 47

Sweden 46

Norway 40

Channel Islands 32

Japan 28

Poland 24

Denmark 21

Cyprus 20

Austria 19

Hong Kong 15

Unspecified 13

Malta 10

Singapore 10

Israel 9

Iceland 7

USA 7

Greece 6

Canada 6

European Community 5

Czech Republic 5

Lithuania 4

Bahrain 4

United Arab Emirates 3

Saudi Arabia 2

Lebanon 1

RSA 1

Brazil 1

Name: InvoiceNo, dtype: int64

Berdasarkan data di atas, negara dengan jumlah nomor invoice terbanyak diperingkati oleh United Kingdom dengan sebanyak 23332 nomor invoice yang masuk dan yang paling sedikit negara Brazil dengan jumlah nomor invoice masuk 1.

1. Eksplorasi customers id berdasarkan negara

Country

United Kingdom 3944

Germany 94

France 87

Spain 31

Belgium 25

Switzerland 22

Portugal 20

Italy 15

Finland 12

Austria 11

Norway 10

Netherlands 9

Australia 9

Channel Islands 9

Denmark 9

Cyprus 8

Sweden 8

Japan 8

Poland 6

Unspecified 5

Israel 5

Greece 4

USA 4

EIRE 4

Canada 4

Bahrain 3

United Arab Emirates 2

Malta 2

Singapore 1

Brazil 1

Iceland 1

Saudi Arabia 1

Hong Kong 1

RSA 1

European Community 1

Lithuania 1

Lebanon 1

Czech Republic 1

Name: CustomerID, dtype: int64

Berdasarkan data di atas, negara dengan jumlah customer id terbanyak diperingkati oleh United Kingdom dengan sebanyak 3944 customer id dan yang paling sedikit negara Czech dengan jumlah customer id sebanyak 1.

1. Eksplorasi Total Amount berdasarkan negara untuk pelanggan tak terdaftar

Country

United Kingdom 1405968.84

EIRE 12961.00

Hong Kong 10037.84

Unspecified 2082.72

Israel 913.57

France 691.06

Switzerland 645.95

Portugal 307.21

Bahrain 0.00

Name: TotalAmount, dtype: float64

Berdasarkan hasil output di atas, jumlah pelanggan tak terdaftar berdasarkan total belanja paling banyak berada di negara United Kingdom yaitu sebesar 1405968.84 dan yang paling sedikit di negara Bahrain dengan Total Belanja sebesar 0.

1. Eksplorasi Nomor invoice berdasarkan negara untuk pelanggan tak terdaftar

Country

United Kingdom 3601

EIRE 41

Hong Kong 15

Unspecified 5

Switzerland 3

Israel 3

France 3

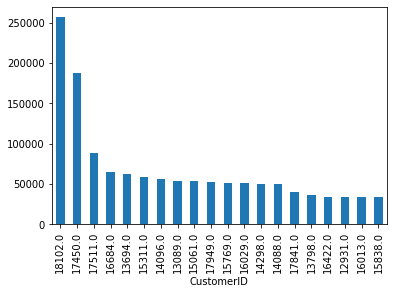
Bahrain 2

Portugal 1

Name: InvoiceNo, dtype: int64

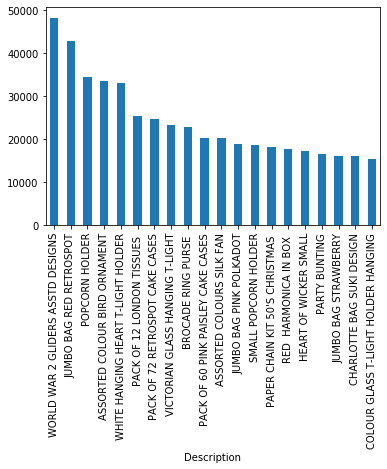
Berdasarkan hasil output di atas, jumlah pelanggan tak terdaftar berdasarkan nomor invoice yang masuk berada di negara United Kingdom yaitu sebanyak 3601 dan yang paling sedikit di negara Portugal dengan nomor invoice yang masuk sebanyak 1.

1. Get United Kingdom top ranked customers based on the total amount



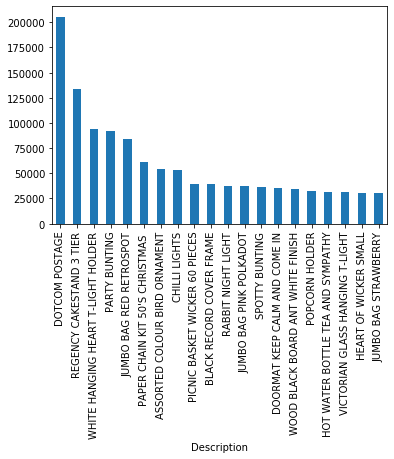
Berdasarkan output di atas, pelanggan dengan customerid 18102 memiliki total amount terbanyak dengan sejumlah 256438.49 dan yang paling sedikit dengan jumlah belanja 33350,76 dengan customerid 15838.

1. United Kingdom frequently sold items by quantitiy



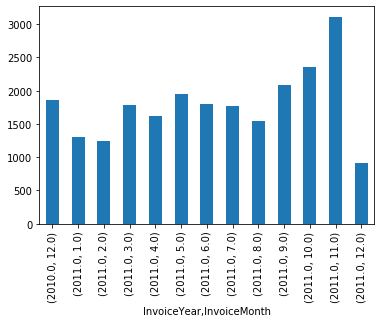
Barang yang lebih cepat habis berdasarkan kuantitas di United Kingdom ialah World War 2 Gliders Assid Designs dengan sebanyak 48230 buah dan yang paling sedikit ialah Colour Glass T-Light Holder Hanging yaitu sebanyak 15400.

1. Frequently sold items by total amount



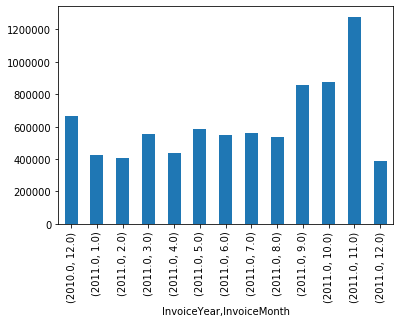
Berdasarkan histogram di atas, di United Kingdom barang yang lebih cepat habis berdasarkan Total belanja ialan Dotcom Postage yaitu sebanyak 205553.47 buah dan yang paling sedikit Jumbo Bag Strawberry sebanyak 30340.90 buah.

1. United Kingdom By Month number of invoices



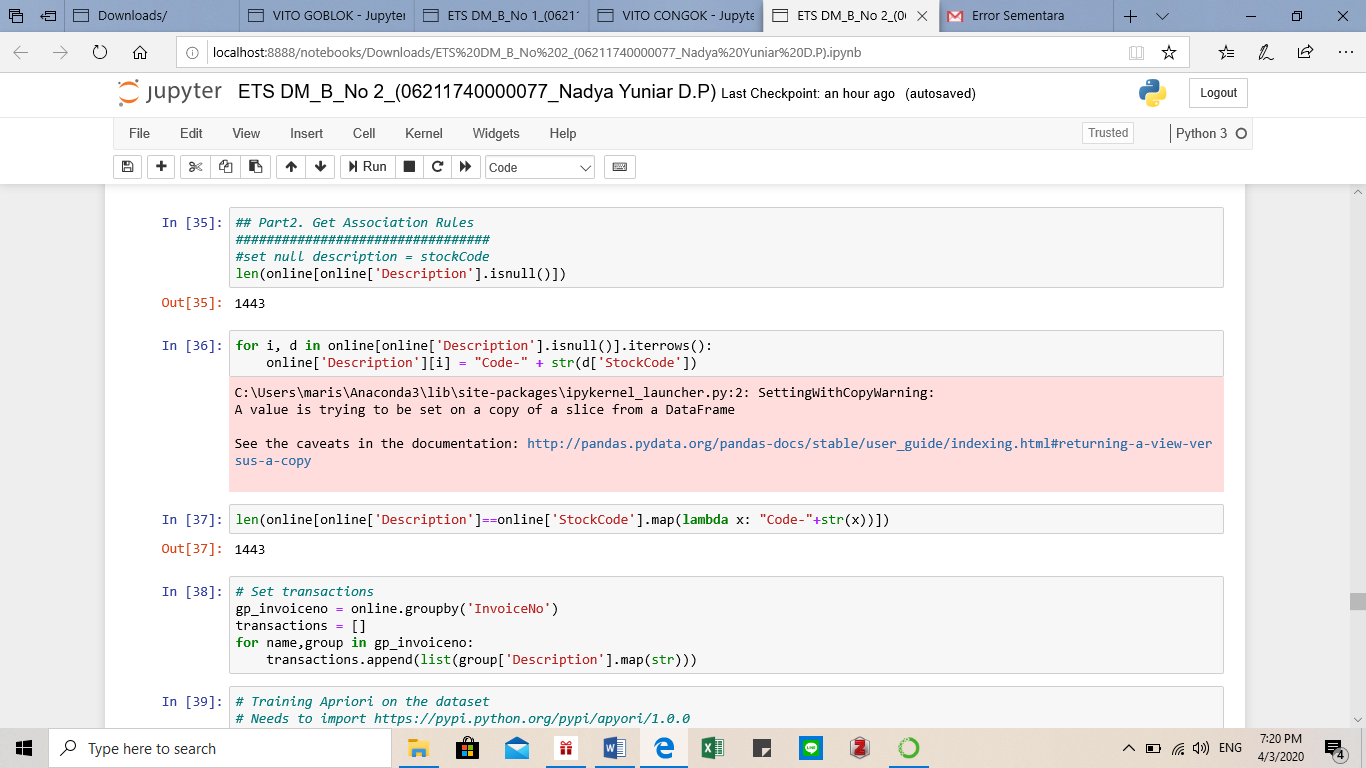
Jumlah nomor invoice yang masuk di United Kingdom paling banyak pada bulan 11 tahun 2011 disusul dengan bulan 10 tahun 2011 dan yang paling sedikit pada bulan Desember tahun 2011.

1. United Kingdom By Month total amounts

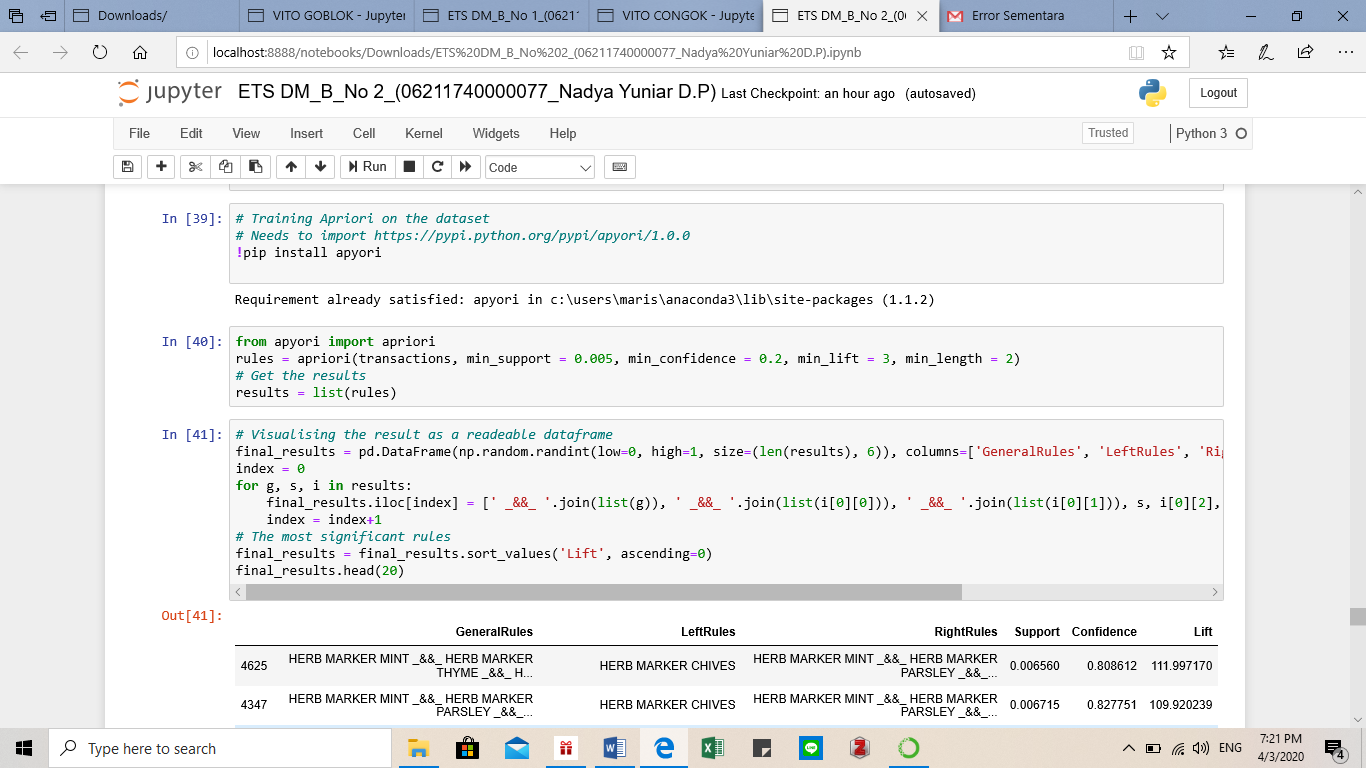


Berdasarkan histogram di atas, total belanja di United Kingdom paling banyak terjadi pada bulan 11 tahun 2011 yaitu sebesar 1279235.91 dan yang paling sedikit pada bulan 12 tahun 2011 dengan total belanja sebesar 386735.240.

1. Mendapatkan Association Rule



1. Melakukan Apriori Algorithm

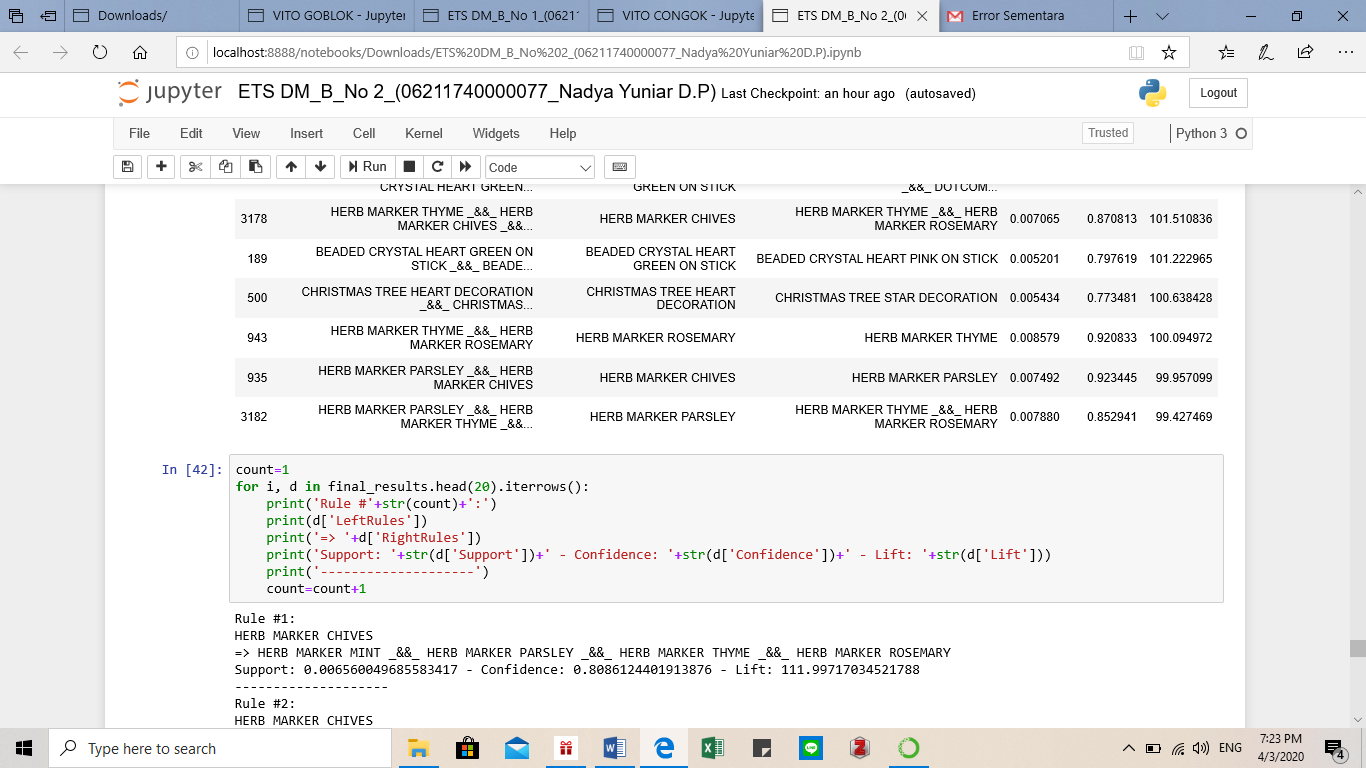


1. Visualising the result as a readeable dataframe



Mendapatkan rule yang paling signifikan yang akan digunakan seperti pada gambar

1. Mengeluarkan Association Rules



1. Otput Association Rule

Rule #1:

HERB MARKER CHIVES

=> HERB MARKER MINT \_&&\_ HERB MARKER PARSLEY \_&&\_ HERB MARKER THYME \_&&\_ HERB MARKER ROSEMARY

Support: 0.006560049685583417 - Confidence: 0.8086124401913876 - Lift: 111.99717034521788

--------------------

Rule #2:

HERB MARKER CHIVES

=> HERB MARKER MINT \_&&\_ HERB MARKER PARSLEY \_&&\_ HERB MARKER THYME

Support: 0.006715317133762906 - Confidence: 0.8277511961722488 - Lift: 109.92023874118286

--------------------

Rule #3:

HERB MARKER CHIVES

=> HERB MARKER MINT \_&&\_ HERB MARKER PARSLEY \_&&\_ HERB MARKER ROSEMARY

Support: 0.006754133995807779 - Confidence: 0.8325358851674641 - Lift: 109.42749731471535

--------------------

Rule #4:

HERB MARKER CHIVES

=> HERB MARKER PARSLEY \_&&\_ HERB MARKER THYME \_&&\_ HERB MARKER ROSEMARY

Support: 0.006909401443987268 - Confidence: 0.8516746411483254 - Lift: 108.08296603577911

--------------------

Rule #5:

HERB MARKER CHIVES

=> HERB MARKER MINT \_&&\_ HERB MARKER THYME \_&&\_ HERB MARKER ROSEMARY

Support: 0.006637683409673162 - Confidence: 0.8181818181818181 - Lift: 106.99492385786802

--------------------

Rule #6:

HERB MARKER CHIVES

=> HERB MARKER MINT \_&&\_ HERB MARKER PARSLEY

Support: 0.0071034857542116294 - Confidence: 0.8755980861244019 - Lift: 106.40168818272095

--------------------

Rule #7:

SET/6 COLLAGE PAPER CUPS

=> SET/6 COLLAGE PAPER PLATES

Support: 0.00609424734104495 - Confidence: 0.887005649717514 - Lift: 106.2839048745237

--------------------

Rule #8:

HERB MARKER CHIVES

=> HERB MARKER PARSLEY \_&&\_ HERB MARKER THYME

Support: 0.0071034857542116294 - Confidence: 0.8755980861244019 - Lift: 105.90214974054857

--------------------

Rule #9:

HERB MARKER CHIVES

=> HERB MARKER MINT \_&&\_ HERB MARKER THYME

Support: 0.006831767719897523 - Confidence: 0.8421052631578947 - Lift: 105.3122125702606

--------------------

Rule #10:

HERB MARKER CHIVES

=> HERB MARKER PARSLEY \_&&\_ HERB MARKER ROSEMARY

Support: 0.0071034857542116294 - Confidence: 0.8755980861244019 - Lift: 104.91701346389229

--------------------

Rule #11:

PACK OF 6 SKULL PAPER CUPS

=> PACK OF 6 SKULL PAPER PLATES

Support: 0.005201459514012887 - Confidence: 0.7790697674418604 - Lift: 104.53330910852712

--------------------

Rule #12:

HERB MARKER CHIVES

=> HERB MARKER MINT \_&&\_ HERB MARKER ROSEMARY

Support: 0.006870584581942396 - Confidence: 0.8468899521531101 - Lift: 102.91310824230388

--------------------

Rule #13:

DOLLY GIRL CHILDRENS BOWL

=> DOLLY GIRL CHILDRENS CUP

Support: 0.005977796754910333 - Confidence: 0.7897435897435897 - Lift: 102.75441595441595

--------------------

Rule #14:

BEADED CRYSTAL HEART GREEN ON STICK

=> BEADED CRYSTAL HEART PINK ON STICK \_&&\_ DOTCOM POSTAGE

Support: 0.005123825789923143 - Confidence: 0.7857142857142858 - Lift: 102.23015873015873

--------------------

Rule #15:

HERB MARKER CHIVES

=> HERB MARKER THYME \_&&\_ HERB MARKER ROSEMARY

Support: 0.007064668892166758 - Confidence: 0.8708133971291867 - Lift: 101.51083591331272

--------------------

Rule #16:

BEADED CRYSTAL HEART GREEN ON STICK

=> BEADED CRYSTAL HEART PINK ON STICK

Support: 0.005201459514012887 - Confidence: 0.7976190476190477 - Lift: 101.22296504808821

--------------------

Rule #17:

CHRISTMAS TREE HEART DECORATION

=> CHRISTMAS TREE STAR DECORATION

Support: 0.005434360686282121 - Confidence: 0.7734806629834253 - Lift: 100.63842848373234

--------------------

Rule #18:

HERB MARKER ROSEMARY

=> HERB MARKER THYME

Support: 0.008578526511916776 - Confidence: 0.9208333333333333 - Lift: 100.09497187060478

--------------------

Rule #19:

HERB MARKER CHIVES

=> HERB MARKER PARSLEY

Support: 0.007491654374660353 - Confidence: 0.9234449760765551 - Lift: 99.95709862892527

--------------------

Rule #20:

HERB MARKER PARSLEY

=> HERB MARKER THYME \_&&\_ HERB MARKER ROSEMARY

Support: 0.007879822995109075 - Confidence: 0.8529411764705882 - Lift: 99.42746872504658

--------------------