Feature extraction - transaction data

December 6, 2017

1 Extrakcia atributov z transakcnych dat

rozne sposoby ako z transakcnych dat vyyazit zaujimave aributy

```
In [1]: %matplotlib inline
        import pandas as pd
        import numpy as np
        import matplotlib.pyplot as plt
        import seaborn

        plt.rcParams['figure.figsize'] = 9, 6
        from IPython.display import Image
```

1.1 Dataset nakupov

```
In [2]: data = pd.read_excel('data/Online Retail.xlsx')

        data['Description'] = data['Description'].str.strip()
        data.dropna(axis=0, subset=['InvoiceNo'], inplace=True)
        data['InvoiceNo'] = data['InvoiceNo'].astype('str')
        data['StockCode'] = data['StockCode'].astype('str')
        data['CustomerID'] = data['CustomerID'].astype('str')
        data = data[~data['InvoiceNo'].str.contains('C')]

        data.head()
```

```
Out[2]:   InvoiceNo StockCode                          Description  Quantity  \
        0    536365    85123A   WHITE HANGING HEART T-LIGHT HOLDER         6
        1    536365     71053                  WHITE METAL LANTERN         6
        2    536365    84406B       CREAM CUPID HEARTS COAT HANGER         8
        3    536365    84029G  KNITTED UNION FLAG HOT WATER BOTTLE         6
        4    536365    84029E       RED WOOLLY HOTTIE WHITE HEART.         6

                  InvoiceDate  UnitPrice CustomerID         Country
        0 2010-12-01 08:26:00       2.55    17850.0  United Kingdom
        1 2010-12-01 08:26:00       3.39    17850.0  United Kingdom
        2 2010-12-01 08:26:00       2.75    17850.0  United Kingdom
        3 2010-12-01 08:26:00       3.39    17850.0  United Kingdom
        4 2010-12-01 08:26:00       3.39    17850.0  United Kingdom
```

```
In [3]: data.describe()
```

```
Out[3]:             Quantity      UnitPrice
        count  532621.000000  532621.000000
        mean       10.239972       3.847621
        std       159.593551      41.758023
        min     -9600.000000  -11062.060000
        25%         1.000000       1.250000
        50%         3.000000       2.080000
        75%        10.000000       4.130000
        max     80995.000000   13541.330000
```

```
In [4]: data[['StockCode', 'Description', 'InvoiceDate', 'Country', 'CustomerID', 'InvoiceNo']].
```

```
Out[4]:        StockCode                         Description          InvoiceDate  \
        count     532621                              531166               532621
        unique      4059                                4194                19990
        top       85123A  WHITE HANGING HEART T-LIGHT HOLDER  2011-10-31 14:41:00
        freq        2271                                2327                 1114
        first        NaN                                 NaN  2010-12-01 08:26:00
        last         NaN                                 NaN  2011-12-09 12:50:00

                       Country CustomerID InvoiceNo
        count           532621     532621    532621
        unique              38       4340     22064
        top     United Kingdom        nan    573585
        freq            487622     134697      1114
        first              NaN        NaN       NaN
        last               NaN        NaN       NaN
```

1.2 Co su transakcne data

Tranakcne data su data opisujuce rozlicne udalosti, transakcie, akcie. Su definovane casovou peciatkou, hodnotou a odkazom na jeden alebo viac objektov (typicky navstivena stranka, zakupeny prodkt, identifikacia zakaznika).

1.3 Typicke ulohy nad takymito datami

- Segmentacia zakaznikov - zhlukovanie
- Predikcia straty zakaznika (churn) - klasifikacia
- Predikcia opakovaneho nakupu (ci a co) - klasifikacia
- Identifikacia spolocne nakupovanych poloziek - hladanie asociacnych pravidiel
- Odporucanie dalsieho obsahu - personalizacia / odporucanie
- Odhad valuacie zakaznika - suma buducich nakupov za nejake obdobie - regresia

2 Extrakcia atributov z tranakcnych dat

Primarny rozdiel oproti datam, ktore bezne pouzivame na trenovanie modelov je v tom, ze v transakcnych datach mame pre jednou entitu viacero pozorovani, ktore za sebou nasleduju v case.

Co teda potrebujeme spravit je previest teito udaje tak, aby sme mali pre jednu entitu jeden riadok a v nom sadu crt.

Najskor treba previest transakcne data do podoby tabulkovych dat (tabular data) = atributov priradenych k pouzivatelom / sedeniam / produktom alebo k hociakym prvkom, pre ktore chceme predikovat

Cize zoskupenie a pocitanie agregacii

Podme vyrabat nejake atributy pre zakaznikov

```
In [5]: customer_data = pd.DataFrame(data.groupby('CustomerID').size()) # celkovy pocet poloziek
        customer_data.columns = ['TransactionCount']
        customer_data.head()
```

```
Out[5]:             TransactionCount
        CustomerID
        12346.0                    1
        12347.0                  182
        12348.0                   31
        12349.0                   73
        12350.0                   17
```

```
In [6]: customer_data['TotalItemCount'] = data.groupby('CustomerID').Quantity.sum() # celkovy po
        customer_data.head()
```

```
Out[6]:             TransactionCount  TotalItemCount
        CustomerID
        12346.0                    1           74215
        12347.0                  182            2458
        12348.0                   31            2341
        12349.0                   73             631
        12350.0                   17             197
```

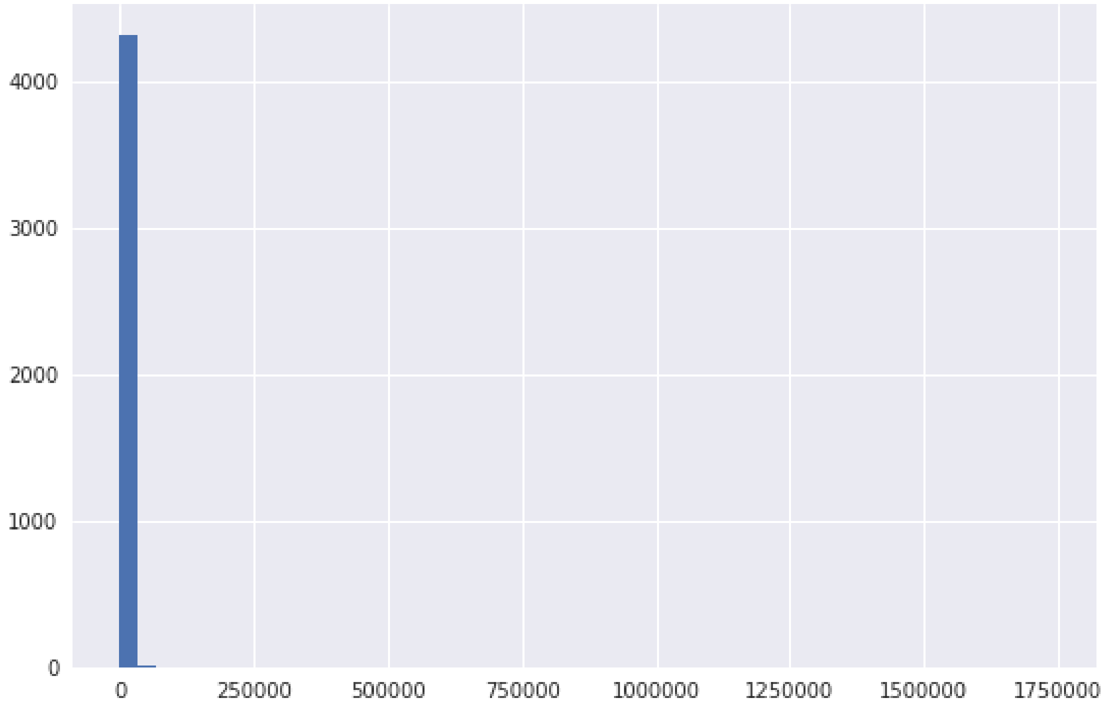
```
In [8]: # valuacia zakaznika
        data['TotalPrice'] = data.Quantity * data.UnitPrice
        customer_data['Valuation'] = data.groupby('CustomerID').TotalPrice.sum()
        customer_data.head()
```

```
Out[8]:             TransactionCount  TotalItemCount  Valuation
        CustomerID
        12346.0                    1           74215   77183.60
        12347.0                  182            2458    4310.00
        12348.0                   31            2341    1797.24
        12349.0                   73             631    1757.55
        12350.0                   17             197     334.40
```

```
In [9]: customer_data['Valuation'].hist(bins=50)
```

```
Out[9]: <matplotlib.axes._subplots.AxesSubplot at 0x7fb31d36e898>
```

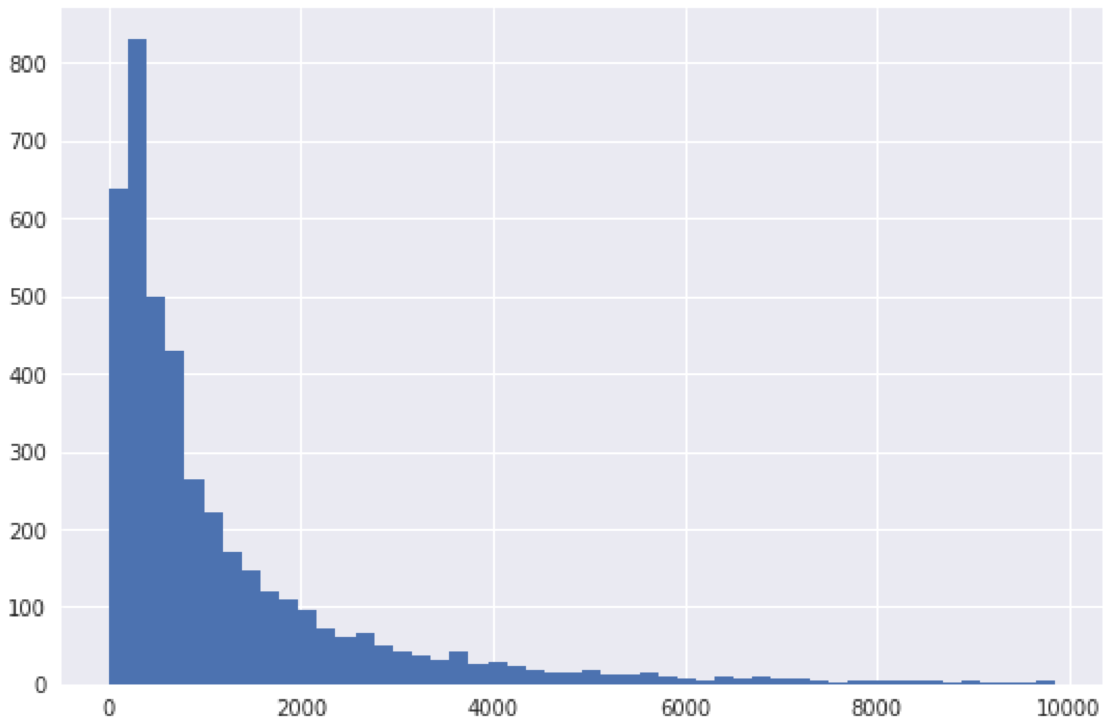
```
/usr/local/lib/python3.5/dist-packages/matplotlib/font_manager.py:1297: UserWarning: findfont: F
  (prop.get_family(), self.defaultFamily[fontext]))
```



```
In [10]: subset = customer_data[customer_data.Valuation < 10000]
         subset.Valuation.hist(bins=50)
```

```
Out[10]: <matplotlib.axes._subplots.AxesSubplot at 0x7fb31d177e10>
```

```
/usr/local/lib/python3.5/dist-packages/matplotlib/font_manager.py:1297: UserWarning: findfont: F
  (prop.get_family(), self.defaultFamily[fontext]))
```



2.0.1 Samozrejme, ze mzoeme agregovat aj kategoricke udaje

```
In [11]: # pouzijeme krajinu, v ktorej si kupil najviac poloziek ako jeho krajinu
         customer_data['Country'] = data.groupby('CustomerID').Country.agg(lambda x: x.value_cou
         customer_data.head()
```

```
Out[11]:             TransactionCount  TotalItemCount  Valuation         Country
         CustomerID
         12346.0                    1           74215   77183.60  United Kingdom
         12347.0                  182            2458    4310.00         Iceland
         12348.0                   31            2341    1797.24         Finland
         12349.0                   73             631    1757.55           Italy
         12350.0                   17             197     334.40          Norway
```

Samozrejme, ze sa da z takychto dat vybrat ovela viac atributov priamo agregaciou.
Daju sa tiez dotiahnut dalsie atributy napriklad z CRM.
Dalsie atributy sa daju vyrabat z tychto.

3 Sedenie (session)

V tomto datasete uz nieco ako sedenie je pripravene (InvoiceNo), kde su zoskupene zaznamy per zakaznik v tom istom case (produkty kupene naraz).

V inych datasetoch toto ale castokrat treba vytvorit: * operacie jedneho pouzivatela za nejake definove casove obdobie * operacie jedneho pouzivatela oddelene casovou medzerou * -| |- oddelene nejakou udalostou

v tomto pripade je to ta prva moznost, a casove okno je 1 minuta

Daju sa potom vytvarat atributy ako: * priemerna dlzka sedenia * pocet akcii per sedenie * priemerna hodnota nejakeho atributu per session

```
In [12]: data.head()
```

```
Out[12]:   InvoiceNo StockCode                          Description  Quantity  \
         0    536365    85123A   WHITE HANGING HEART T-LIGHT HOLDER         6
         1    536365     71053                  WHITE METAL LANTERN         6
         2    536365    84406B       CREAM CUPID HEARTS COAT HANGER         8
         3    536365    84029G  KNITTED UNION FLAG HOT WATER BOTTLE         6
         4    536365    84029E       RED WOOLLY HOTTIE WHITE HEART.         6

                   InvoiceDate  UnitPrice  CustomerID         Country  TotalPrice
         0 2010-12-01 08:26:00       2.55     17850.0  United Kingdom       15.30
         1 2010-12-01 08:26:00       3.39     17850.0  United Kingdom       20.34
         2 2010-12-01 08:26:00       2.75     17850.0  United Kingdom       22.00
         3 2010-12-01 08:26:00       3.39     17850.0  United Kingdom       20.34
         4 2010-12-01 08:26:00       3.39     17850.0  United Kingdom       20.34
```

4 Vyrabanie novych z existujucich atributov

4.1 Pomerove atributy

Pomer voci agregovanej hodnote cez viacero pozorovani.

Napr. Pomer voci priemernemu poctu transakcii per pouzivatel.

Viacero sposobov vyrabania tychto pomerov.

** Pozor ** agregovanu hodnotu pocitat len na trenovacich datach a nie na testovacich. Zaviedla by sa tak informacia z buducnosti

```
In [13]: # pomer voci priemeru / medianu/ kvartilom ...
         mean = customer_data.TransactionCount.mean() # median() # quantile(0.25) # quantile(0.7
         # priemer pocitam vzdy len na trenovacich datach. Na testovacich pouzivame len ten, kto
         customer_data['RatioToMeanTransactionCount'] = customer_data.TransactionCount / mean
         customer_data.head()
```

```
Out[13]:             TransactionCount  TotalItemCount  Valuation         Country  \
         CustomerID
         12346.0                    1           74215   77183.60  United Kingdom
         12347.0                  182            2458    4310.00         Iceland
         12348.0                   31            2341    1797.24         Finland
         12349.0                   73             631    1757.55           Italy
         12350.0                   17             197     334.40          Norway

                     RatioToMeanTransactionCount
```

```
CustomerID
12346.0                        0.008148
12347.0                        1.483006
12348.0                        0.252600
12349.0                        0.594832
12350.0                        0.138523
```

```
In [14]: # rozdiel voci priemeru / ...
         mean = customer_data.TransactionCount.mean() # median() # quantile(0.25) # quantile(0.7
         customer_data['DifFromMeanTransactionCount'] = customer_data.TransactionCount - mean
         customer_data.head()
```

```
Out[14]:             TransactionCount  TotalItemCount  Valuation         Country  \
         CustomerID
         12346.0                    1           74215   77183.60  United Kingdom
         12347.0                  182            2458    4310.00         Iceland
         12348.0                   31            2341    1797.24         Finland
         12349.0                   73             631    1757.55           Italy
         12350.0                   17             197     334.40          Norway

                     RatioToMeanTransactionCount  DifFromMeanTransactionCount
         CustomerID
         12346.0                        0.008148                  -121.723733
         12347.0                        1.483006                    59.276267
         12348.0                        0.252600                   -91.723733
         12349.0                        0.594832                   -49.723733
         12350.0                        0.138523                  -105.723733
```

```
In [15]: # binarna hodnota nad/pod priemerom
         mean = customer_data.TransactionCount.mean() # median() # quantile(0.25) # quantile(0.7
         customer_data['HigherThanMeanTransactionCount'] = customer_data.TransactionCount > mean
         customer_data.head()
```

```
Out[15]:             TransactionCount  TotalItemCount  Valuation         Country  \
         CustomerID
         12346.0                    1           74215   77183.60  United Kingdom
         12347.0                  182            2458    4310.00         Iceland
         12348.0                   31            2341    1797.24         Finland
         12349.0                   73             631    1757.55           Italy
         12350.0                   17             197     334.40          Norway

                     RatioToMeanTransactionCount  DifFromMeanTransactionCount  \
         CustomerID
         12346.0                        0.008148                  -121.723733
         12347.0                        1.483006                    59.276267
         12348.0                        0.252600                   -91.723733
         12349.0                        0.594832                   -49.723733
         12350.0                        0.138523                  -105.723733
```

```
           HigherThanMeanTransactionCount
CustomerID
12346.0                             False
12347.0                              True
12348.0                             False
12349.0                             False
12350.0                             False
```

```
In [16]: # percentil
         from scipy import stats
         customer_data['TransactionCountPercentile'] = stats.rankdata(customer_data.TransactionC
         customer_data.head()
```

```
Out[16]:             TransactionCount  TotalItemCount  Valuation         Country  \
         CustomerID
         12346.0                    1           74215   77183.60  United Kingdom
         12347.0                  182            2458    4310.00         Iceland
         12348.0                   31            2341    1797.24         Finland
         12349.0                   73             631    1757.55           Italy
         12350.0                   17             197     334.40          Norway

                     RatioToMeanTransactionCount  DifFromMeanTransactionCount  \
         CustomerID
         12346.0                        0.008148                  -121.723733
         12347.0                        1.483006                    59.276267
         12348.0                        0.252600                   -91.723733
         12349.0                        0.594832                   -49.723733
         12350.0                        0.138523                  -105.723733

                    HigherThanMeanTransactionCount  TransactionCountPercentile
         CustomerID
         12346.0                             False                    0.841014
         12347.0                              True                   88.087558
         12348.0                             False                   42.108295
         12349.0                             False                   66.555300
         12350.0                             False                   24.919355
```

```
In [17]: # na testovacich datach by sa percentil musel pocitat nejak takto
         train = customer_data.TransactionCount
         test = [10, 20, 30, 40]
         [stats.percentileofscore(train, a, 'rank') for a in test]
```

```
Out[17]: [14.343317972350231,
          28.963133640552996,
          41.036866359447004,
          49.170506912442399]
```

4.2 Pomer voci agregovanej hodnote segmentu

Nepocitat globalnu agregovanu hodnotu, ale pre kazdy segment je spocitat zvlast a potom pouzit tuto specificku hodnotu pre vypocet pomeroveho atributu

```
In [18]: # pomer voci priemru segmentu
         means = customer_data.groupby('Country').TransactionCount.mean()
         means
```

```
Out[18]: Country
         Australia                139.666667
         Austria                   41.000000
         Bahrain                    8.500000
         Belgium                   83.583333
         Brazil                    32.000000
         Canada                    37.750000
         Channel Islands           83.111111
         Cyprus                    96.333333
         Czech Republic            25.000000
         Denmark                   49.375000
         EIRE                    2412.666667
         European Community        60.000000
         Finland                   57.083333
         France                    95.885057
         Germany                   96.191489
         Greece                    36.250000
         Iceland                  182.000000
         Israel                    82.666667
         Italy                     54.142857
         Japan                     40.125000
         Lebanon                   45.000000
         Lithuania                 35.000000
         Malta                     56.000000
         Netherlands              262.555556
         Norway                   107.200000
         Poland                    55.000000
         Portugal                  76.947368
         RSA                       58.000000
         Saudi Arabia               9.000000
         Singapore                222.000000
         Spain                     86.344828
         Sweden                    56.375000
         Switzerland               91.300000
         USA                       44.750000
         United Arab Emirates      34.000000
         United Kingdom           124.691994
         Unspecified               61.000000
         Name: TransactionCount, dtype: float64
```

```
In [19]: customer_data['RatioToMeanTransactionCountPerCountry'] = customer_data.apply(lambda x:
         customer_data.head()
```

```
Out[19]:             TransactionCount  TotalItemCount  Valuation         Country  \
         CustomerID
         12346.0                    1           74215   77183.60  United Kingdom
         12347.0                  182            2458    4310.00         Iceland
         12348.0                   31            2341    1797.24         Finland
         12349.0                   73             631    1757.55           Italy
         12350.0                   17             197     334.40          Norway

                     RatioToMeanTransactionCount  DifFromMeanTransactionCount  \
         CustomerID
         12346.0                        0.008148                  -121.723733
         12347.0                        1.483006                    59.276267
         12348.0                        0.252600                   -91.723733
         12349.0                        0.594832                   -49.723733
         12350.0                        0.138523                  -105.723733

                    HigherThanMeanTransactionCount  TransactionCountPercentile  \
         CustomerID
         12346.0                             False                    0.841014
         12347.0                              True                   88.087558
         12348.0                             False                   42.108295
         12349.0                             False                   66.555300
         12350.0                             False                   24.919355

                     RatioToMeanTransactionCountPerCountry
         CustomerID
         12346.0                                  0.008020
         12347.0                                  1.000000
         12348.0                                  0.543066
         12349.0                                  1.348285
         12350.0                                  0.158582
```

5 Segmentacia

- kategoricka premenna
- Segmenty zlozene na manualne definovanych pravidlach
- Naucenie sa prirodzenych segemntov pomocou zhlukovania

```
In [20]: subset = customer_data.replace([np.inf, -np.inf], np.nan)
         subset = subset.dropna()
```

```
In [21]: from sklearn.cluster import KMeans

         cluster = KMeans(n_clusters=5)
         subset['cluster'] = cluster.fit_predict(subset._get_numeric_data())
         subset.head()
```

```
Out[21]:             TransactionCount  TotalItemCount  Valuation         Country  \
         CustomerID
         12346.0                    1           74215   77183.60  United Kingdom
         12347.0                  182            2458    4310.00         Iceland
         12348.0                   31            2341    1797.24         Finland
         12349.0                   73             631    1757.55           Italy
         12350.0                   17             197     334.40          Norway

                     RatioToMeanTransactionCount  DifFromMeanTransactionCount  \
         CustomerID
         12346.0                        0.008148                  -121.723733
         12347.0                        1.483006                    59.276267
         12348.0                        0.252600                   -91.723733
         12349.0                        0.594832                   -49.723733
         12350.0                        0.138523                  -105.723733

                    HigherThanMeanTransactionCount  TransactionCountPercentile  \
         CustomerID
         12346.0                             False                    0.841014
         12347.0                              True                   88.087558
         12348.0                             False                   42.108295
         12349.0                             False                   66.555300
         12350.0                             False                   24.919355

                     RatioToMeanTransactionCountPerCountry  cluster
         CustomerID
         12346.0                                  0.008020        3
         12347.0                                  1.000000        0
         12348.0                                  0.543066        0
         12349.0                                  1.348285        0
         12350.0                                  0.158582        0
```

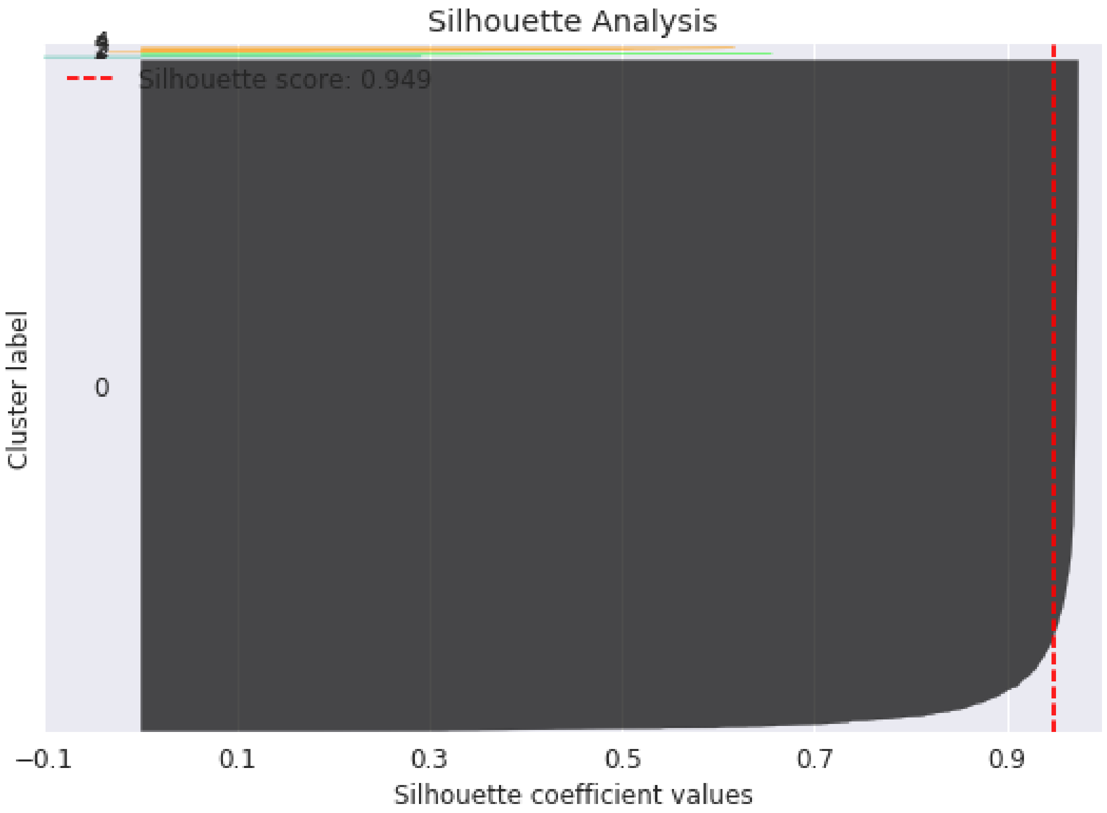
```
In [22]: subset.cluster.value_counts()
```

```
Out[22]: 0    4299
         4      31
         3       6
         2       3
         1       1
         Name: cluster, dtype: int64
```

```
In [23]: import scikitplot as skplt
         skplt.metrics.plot_silhouette(subset._get_numeric_data(), subset.cluster, cmap='nipy_sp
```

```
Out[23]: <matplotlib.axes._subplots.AxesSubplot at 0x7fb3115cdf28>
```

```
/usr/local/lib/python3.5/dist-packages/matplotlib/font_manager.py:1297: UserWarning: findfont: F
  (prop.get_family(), self.defaultFamily[fontext]))
```



Tieto zhluky niesu az take pekne. Chcelo by sa to pohrat s atributmi, ktore pouzivame na zhlukovanie. Mozno nejake transformacie do inych rozdeleni, mozno povyberat len nejaku podmnozinu atributov, mozno popridatavat dalsie aby to davalo zmysel.

Vieme potom ale pouzivat identifikator zhluku ako kategoricku premennu alebo nejaky segment, pre ktore mozeme pocitat agregovane hodnoty a vyrabat pomerove atributy na prirodzenych zhlukoch / segmentoch, ktore su v datach.

6 Spracovanie textovych opisov

Vo vsetobecnosti nema velmi zmysel pocitat kolko krat si pouzivatel kupil jednotlive produkty. Tych produktov je vacsinou tak vela, ze ak by sme chceli kazdy pouzit pocet nakupov kazdeho z nich tak vytvorime velmi riedku maticu a tazko sa z nej bude nieco dat pocitat (samozrejme, existuju aj metody, ktore pracuju prave s takymito maticami)

Castokrat mame k produktom priradene nejake kategorie, ktore nam zoskupuju viacere produkty. Pocitat pocet nakupov per kategoria uz zmysel moze mat. Kde ale zobrat kategorie ak ich nemame priprvene vopred.

Jedna moznost je zhlukovanie na zaklade atributov produktov, tak ako v predchadzajucom priklade.

Dalsia moznost: ak mame url produktu, tak casto je v nej zakodovana kategoria. Da sa vyparsovat

Dalsia moznost: z textovych opisov produktov vytvorit spolocne temy.

```
In [24]: data.head()
```

```
Out[24]:   InvoiceNo StockCode                          Description  Quantity  \
         0    536365    85123A   WHITE HANGING HEART T-LIGHT HOLDER         6
         1    536365     71053                  WHITE METAL LANTERN         6
         2    536365    84406B       CREAM CUPID HEARTS COAT HANGER         8
         3    536365    84029G  KNITTED UNION FLAG HOT WATER BOTTLE         6
         4    536365    84029E       RED WOOLLY HOTTIE WHITE HEART.         6

                   InvoiceDate  UnitPrice CustomerID         Country  TotalPrice
         0 2010-12-01 08:26:00       2.55    17850.0  United Kingdom       15.30
         1 2010-12-01 08:26:00       3.39    17850.0  United Kingdom       20.34
         2 2010-12-01 08:26:00       2.75    17850.0  United Kingdom       22.00
         3 2010-12-01 08:26:00       3.39    17850.0  United Kingdom       20.34
         4 2010-12-01 08:26:00       3.39    17850.0  United Kingdom       20.34
```

```
In [25]: products = data[['StockCode', 'Description']]
         products = products.drop_duplicates(subset='StockCode')
         print(len(products))
         products.head()
```

```
4059
```

```
Out[25]:   StockCode                          Description
         0    85123A   WHITE HANGING HEART T-LIGHT HOLDER
         1     71053                  WHITE METAL LANTERN
         2    84406B       CREAM CUPID HEARTS COAT HANGER
         3    84029G  KNITTED UNION FLAG HOT WATER BOTTLE
         4    84029E       RED WOOLLY HOTTIE WHITE HEART.
```

```
In [26]: import gensim
         from nltk.tokenize import RegexpTokenizer
         from nltk.stem.porter import PorterStemmer

         tokenizer = RegexpTokenizer(r'\w+')
         p_stemmer = PorterStemmer()

         texts = [[p_stemmer.stem(i) for i in tokenizer.tokenize(str(raw).lower())] for raw in p
```

```
In [27]: texts
```

```
Out[27]: [['white', 'hang', 'heart', 't', 'light', 'holder'],
          ['white', 'metal', 'lantern'],
          ['cream', 'cupid', 'heart', 'coat', 'hanger'],
          ['knit', 'union', 'flag', 'hot', 'water', 'bottl'],
          ['red', 'woolli', 'hotti', 'white', 'heart'],
          ['set', '7', 'babushka', 'nest', 'box'],
          ['glass', 'star', 'frost', 't', 'light', 'holder'],
['hand', 'warmer', 'union', 'jack'],
['hand', 'warmer', 'red', 'polka', 'dot'],
['assort', 'colour', 'bird', 'ornament'],
['poppi', 's', 'playhous', 'bedroom'],
['poppi', 's', 'playhous', 'kitchen'],
['feltcraft', 'princess', 'charlott', 'doll'],
['ivori', 'knit', 'mug', 'cosi'],
['box', 'of', '6', 'assort', 'colour', 'teaspoon'],
['box', 'of', 'vintag', 'jigsaw', 'block'],
['box', 'of', 'vintag', 'alphabet', 'block'],
['home', 'build', 'block', 'word'],
['love', 'build', 'block', 'word'],
['recip', 'box', 'with', 'metal', 'heart'],
['doormat', 'new', 'england'],
['jam', 'make', 'set', 'with', 'jar'],
['red', 'coat', 'rack', 'pari', 'fashion'],
['yellow', 'coat', 'rack', 'pari', 'fashion'],
['blue', 'coat', 'rack', 'pari', 'fashion'],
['bath', 'build', 'block', 'word'],
['alarm', 'clock', 'bakelik', 'pink'],
['alarm', 'clock', 'bakelik', 'red'],
['alarm', 'clock', 'bakelik', 'green'],
['panda', 'and', 'bunni', 'sticker', 'sheet'],
['star', 'gift', 'tape'],
['inflat', 'polit', 'globe'],
['vintag', 'head', 'and', 'tail', 'card', 'game'],
['set', '2', 'red', 'retrospot', 'tea', 'towel'],
['round', 'snack', 'box', 'set', 'of4', 'woodland'],
['spaceboy', 'lunch', 'box'],
['lunch', 'box', 'i', 'love', 'london'],
['circu', 'parad', 'lunch', 'box'],
['charlott', 'bag', 'dolli', 'girl', 'design'],
['red', 'toadstool', 'led', 'night', 'light'],
['set', '2', 'tea', 'towel', 'i', 'love', 'london'],
['vintag', 'seasid', 'jigsaw', 'puzzl'],
['mini', 'jigsaw', 'circu', 'parad'],
['mini', 'jigsaw', 'spaceboy'],
['mini', 'paint', 'set', 'vintag'],
['postag'],
['paper', 'chain', 'kit', '50', 's', 'christma'],
['edwardian', 'parasol', 'red'],
['retro', 'coffe', 'mug', 'assort'],
['save', 'the', 'planet', 'mug'],
['vintag', 'billboard', 'drink', 'me', 'mug'],
['vintag', 'billboard', 'love', 'hate', 'mug'],
['wood', '2', 'drawer', 'cabinet', 'white', 'finish'],
['wood', 's', '3', 'cabinet', 'ant', 'white', 'finish'],
['wooden', 'pictur', 'frame', 'white', 'finish'],
['wooden', 'frame', 'antiqu', 'white'],
['victorian', 'sew', 'box', 'larg'],
['hot', 'water', 'bottl', 'tea', 'and', 'sympathi'],
['red', 'hang', 'heart', 't', 'light', 'holder'],
['jumbo', 'bag', 'pink', 'polkadot'],
['jumbo', 'bag', 'baroqu', 'black', 'white'],
['jumbo', 'bag', 'charli', 'and', 'lola', 'toy'],
['strawberri', 'charlott', 'bag'],
['red', '3', 'piec', 'retrospot', 'cutleri', 'set'],
['blue', '3', 'piec', 'polkadot', 'cutleri', 'set'],
['set', '6', 'red', 'spotti', 'paper', 'plate'],
['lunch', 'bag', 'red', 'retrospot'],
['strawberri', 'lunch', 'box', 'with', 'cutleri'],
['lunch', 'box', 'with', 'cutleri', 'retrospot'],
['pack', 'of', '72', 'retrospot', 'cake', 'case'],
['pack', 'of', '60', 'dinosaur', 'cake', 'case'],
['pack', 'of', '60', 'pink', 'paisley', 'cake', 'case'],
['60', 'teatim', 'fairi', 'cake', 'case'],
['tomato', 'charli', 'lola', 'coaster', 'set'],
['charli', 'lola', 'wastepap', 'bin', 'flora'],
['red', 'charli', 'lola', 'person', 'doorsign'],
['jumbo', 'storag', 'bag', 'suki'],
['jumbo', 'bag', 'pink', 'vintag', 'paisley'],
['jam', 'make', 'set', 'print'],
['retrospot', 'tea', 'set', 'ceram', '11', 'pc'],
['girli', 'pink', 'tool', 'set'],
['jumbo', 'shopper', 'vintag', 'red', 'paisley'],
['airlin', 'loung', 'metal', 'sign'],
['white', 'spot', 'red', 'ceram', 'drawer', 'knob'],
['red', 'drawer', 'knob', 'acryl', 'edwardian'],
['clear', 'drawer', 'knob', 'acryl', 'edwardian'],
['photo', 'clip', 'line'],
['felt', 'egg', 'cosi', 'chicken'],
['piggi', 'bank', 'retrospot'],
['skull', 'shoulder', 'bag'],
['you', 're', 'confus', 'me', 'metal', 'sign'],
['cook', 'with', 'wine', 'metal', 'sign'],
['gin', 'tonic', 'diet', 'metal', 'sign'],
['yellow', 'breakfast', 'cup', 'and', 'saucer'],
['pink', 'breakfast', 'cup', 'and', 'saucer'],
['paper', 'chain', 'kit', 'retrospot'],
['small', 'heart', 'flower', 'hook'],
['tea', 'time', 'de', 'tea', 'cosi'],
['felt', 'egg', 'cosi', 'white', 'rabbit'],
['zinc', 'willi', 'winki', 'candl', 'stick'],
['ceram', 'cherri', 'cake', 'money', 'bank'],
['retrospot', 'larg', 'milk', 'jug'],
['set', 'of', '6', 'funki', 'beaker'],
['edwardian', 'parasol', 'black'],
['edwardian', 'parasol', 'natur'],
['ceram', 'strawberri', 'cake', 'money', 'bank'],
['blue', 'owl', 'soft', 'toy'],
['balloon', 'art', 'make', 'your', 'own', 'flower'],
['glass', 'cloch', 'small'],
['gumbal', 'monochrom', 'coat', 'rack'],
['doormat', 'fanci', 'font', 'home', 'sweet', 'home'],
['vintag', 'snake', 'ladder'],
['chocol', 'calcul'],
['recycl', 'bag', 'retrospot'],
['toy', 'tidi', 'pink', 'polkadot'],
['antiqu', 'glass', 'dress', 'tabl', 'pot'],
['ivori', 'giant', 'garden', 'thermomet'],
['3', 'tier', 'cake', 'tin', 'green', 'and', 'cream'],
['3', 'tier', 'cake', 'tin', 'red', 'and', 'cream'],
['set', '3', 'wicker', 'oval', 'basket', 'w', 'lid'],
['wood', 'black', 'board', 'ant', 'white', 'finish'],
['colour', 'glass', 't', 'light', 'holder', 'hang'],
['hang', 'metal', 'heart', 'lantern'],
['hang', 'medina', 'lantern', 'small'],
['natur', 'slate', 'heart', 'chalkboard'],
['heart', 'of', 'wicker', 'small'],
['heart', 'of', 'wicker', 'larg'],
['white', 'lovebird', 'lantern'],
['classic', 'metal', 'birdcag', 'plant', 'holder'],
['cream', 'heart', 'card', 'holder'],
['enamel', 'flower', 'jug', 'cream'],
['enamel', 'fire', 'bucket', 'cream'],
['enamel', 'bread', 'bin', 'cream'],
['jumbo', 'bag', 'dolli', 'girl', 'design'],
['tradit', 'christma', 'ribbon'],
['organis', 'wood', 'antiqu', 'white'],
['lunch', 'bag', 'dolli', 'girl', 'design'],
['white', 'wire', 'egg', 'holder'],
['jumbo', 'bag', 'red', 'retrospot'],
['chilli', 'light'],
['light', 'garland', 'butterfil', 'pink'],
['wooden', 'owl', 'light', 'garland'],
['fairi', 'tale', 'cottag', 'nightlight'],
['home', 'small', 'wood', 'letter'],
['gingham', 'heart', 'doorstop', 'red'],
['five', 'heart', 'hang', 'decor'],
['assort', 'bottl', 'top', 'magnet'],
['fridg', 'magnet', 'us', 'diner', 'assort'],
['homemad', 'jam', 'scent', 'candl'],
['fridg', 'magnet', 'le', 'enfant', 'assort'],
['rose', 'caravan', 'doorstop'],
['5', 'hook', 'hanger', 'magic', 'toadstool'],
['christma', 'light', '10', 'reindeer'],
['vintag', 'union', 'jack', 'cushion', 'cover'],
['set', 'of', '3', 'colour', 'fli', 'duck'],
['set', 'of', '3', 'gold', 'fli', 'duck'],
['red', 'retrospot', 'umbrella'],
['black', 'blue', 'polkadot', 'umbrella'],
['red', 'diner', 'wall', 'clock'],
['blue', 'diner', 'wall', 'clock'],
['ivori', 'diner', 'wall', 'clock'],
['larg', 'heart', 'measur', 'spoon'],
['small', 'heart', 'measur', 'spoon'],
['jam', 'jar', 'with', 'pink', 'lid'],
['jam', 'jar', 'with', 'green', 'lid'],
['rose', 'cottag', 'keepsak', 'box'],
['hang', 'heart', 'zinc', 't', 'light', 'holder'],
['paper', 'chain', 'kit', 'vintag', 'christma'],
['disco', 'ball', 'christma', 'decor'],
['small', 'popcorn', 'holder'],
['larg', 'popcorn', 'holder'],
['set', '20', 'red', 'retrospot', 'paper', 'napkin'],
['set', '6', 'red', 'spotti', 'paper', 'cup'],
['polkadot', 'rain', 'hat'],
['delux', 'sew', 'kit'],
['retrospot', 'heart', 'hot', 'water', 'bottl'],
['english', 'rose', 'hot', 'water', 'bottl'],
['photo', 'cube'],
['3', 'stripey', 'mice', 'feltcraft'],
['set', 'of', '6', 'soldier', 'skittl'],
['tradit', 'wooden', 'skip', 'rope'],
['wooden', 'box', 'of', 'domino'],
['rustic', 'seventeen', 'drawer', 'sideboard'],
['parti', 'cone', 'carniv', 'assort'],
['parti', 'cone', 'candi', 'assort'],
['picnic', 'basket', 'wicker', 'small'],
['star', 'decor', 'paint', 'zinc'],
['retrospot', 'lamp'],
['fanci', 'font', 'birthday', 'card'],
['hand', 'warmer', 'scotti', 'dog', 'design'],
['hand', 'warmer', 'owl', 'design'],
['dog', 'bowl', 'chase', 'ball', 'design'],
['cloth', 'peg', 'retrospot', 'pack', '24'],
['hand', 'over', 'the', 'chocol', 'sign'],
['travel', 'sew', 'kit'],
['black', 'heart', 'card', 'holder'],
['chick', 'grey', 'hot', 'water', 'bottl'],
['small', 'glass', 'heart', 'trinket', 'pot'],
['alarm', 'clock', 'bakelik', 'ivori'],
['alarm', 'clock', 'bakelik', 'orang'],
['hand', 'warmer', 'bird', 'design'],
['ivori', 'embroid', 'quilt'],
['set', 'of', '3', 'black', 'fli', 'duck'],
['pack', 'of', '12', 'red', 'retrospot', 'tissu'],
['red', 'retrospot', 'mug'],
['babushka', 'light', 'string', 'of', '10'],
['doormat', 'fairi', 'cake'],
['strawberri', 'ceram', 'trinket', 'box'],
['pink', 'doughnut', 'trinket', 'pot'],
['silk', 'purs', 'babushka', 'pink'],
['hot', 'water', 'bottl', 'i', 'am', 'so', 'poorli'],
['chocol', 'hot', 'water', 'bottl'],
['white', 'skull', 'hot', 'water', 'bottl'],
['scotti', 'dog', 'hot', 'water', 'bottl'],
['bird', 'hous', 'hot', 'water', 'bottl'],
['boudoir', 'squar', 'tissu', 'box'],
['skull', 'squar', 'tissu', 'box'],
['photo', 'frame', 'cornic'],
['silk', 'purs', 'babushka', 'red'],
['pictur', 'domino'],
['s', '6', 'sew', 'on', 'crochet', 'flower'],
['scandinavian', 'red', 'ribbon'],
['balloon', 'write', 'set'],
['lavend', 'incens', 'in', 'tin'],
['tv', 'dinner', 'tray', 'vintag', 'paisley'],
['set', 'of', '4', 'english', 'rose', 'placemat'],
['set', 'of', '4', 'english', 'rose', 'coaster'],
['tripl', 'photo', 'frame', 'cornic'],
['famili', 'photo', 'frame', 'cornic'],
['mirror', 'disco', 'ball'],
['disco', 'ball', 'rotat', 'batteri', 'oper'],
['silver', 'look', 'mirror'],
['ladi', 'gentlemen', 'metal', 'sign'],
['metal', 'sign', 'her', 'dinner', 'is', 'serv'],
['doormat', 'topiari'],
['bathroom', 'metal', 'sign'],
['kitchen', 'metal', 'sign'],
['toilet', 'metal', 'sign'],
['metal', 'sign', 'take', 'it', 'or', 'leav', 'it'],
['i', 'm', 'on', 'holiday', 'metal', 'sign'],
['grow', 'your', 'own', 'basil', 'in', 'enamel', 'mug'],
['set', '10', 'pink', 'polkadot', 'parti', 'candl'],
['set', '20', 'napkin', 'fairi', 'cake', 'design'],
['set', 'of', '6', 't', 'light', 'snowmen'],
['set', 'of', '6', 't', 'light', 'santa'],
['set', 'of', '9', 'heart', 'shape', 'balloon'],
['sandwich', 'bath', 'spong'],
['appl', 'bath', 'spong'],
['strawberri', 'bath', 'spong'],
['black', 'pirat', 'treasur', 'chest'],
['star', 'portabl', 'tabl', 'light'],
['snowflak', 'portabl', 'tabl', 'light'],
['pink', 'oval', 'jewel', 'mirror'],
['retrospot', 'cigar', 'box', 'match'],
['cosi', 'hour', 'giant', 'tube', 'match'],
['jazz', 'heart', 'purs', 'notebook'],
['candlehold', 'pink', 'hang', 'heart'],
['assort', 'colour', 'mini', 'case'],
['lunch', 'bag', 'spaceboy', 'design'],
['lunch', 'bag', 'woodland'],
['lunch', 'bag', 'pink', 'polkadot'],
['gumbal', 'coat', 'rack'],
['blue', 'new', 'baroqu', 'candlestick', 'candl'],
['pink', 'new', 'baroquecandlestick', 'candl'],
['fairi', 'cake', 'flannel', 'assort', 'colour'],
['bread', 'bin', 'diner', 'style', 'pink'],
['tea', 'time', 'tabl', 'cloth'],
['hot', 'water', 'bottl', 'babushka'],
['heart', 'ivori', 'trelli', 'small'],
['pink', 'drawer', 'knob', 'acryl', 'edwardian'],
['green', 'drawer', 'knob', 'acryl', 'edwardian'],
['blue', 'drawer', 'knob', 'acryl', 'edwardian'],
['set', '6', 'footbal', 'celebr', 'candl'],
['set', 'of', '6', 'girl', 'celebr', 'candl'],
['romant', 'pink', 'ribbon'],
['bright', 'blue', 'ribbon'],
['chocol', 'box', 'ribbon'],
['parti', 'invit', 'footbal'],
['parti', 'invit', 'jazz', 'heart'],
['parti', 'invit', 'spaceman'],
['set', 'of', '3', 'butterfli', 'cooki', 'cutter'],
['set', 'of', '3', 'heart', 'cooki', 'cutter'],
['3', 'piec', 'spaceboy', 'cooki', 'cutter', 'set'],
['pack', 'of', '72', 'skull', 'cake', 'case'],
['pack', 'of', '60', 'spaceboy', 'cake', 'case'],
['72', 'sweetheart', 'fairi', 'cake', 'case'],
['lunch', 'bag', 'suki', 'design'],
['lunch', 'bag', 'car', 'blue'],
['lunch', 'bag', 'black', 'skull'],
['heart', 'ivori', 'trelli', 'larg'],
['set', '5', 'red', 'retrospot', 'lid', 'glass', 'bowl'],
['magic', 'draw', 'slate', 'dinosaur'],
['magic', 'draw', 'slate', 'bake', 'a', 'cake'],
['12', 'pencil', 'tall', 'tube', 'skull'],
['red', 'harmonica', 'in', 'box'],
['blue', 'harmonica', 'in', 'box'],
['skull', 'water', 'transfer', 'tattoo'],
['pack', '3', 'box', 'bird', 'panneton'],
['set', 'of', '20', 'kid', 'cooki', 'cutter'],
['10', 'colour', 'spaceboy', 'pen'],
['victorian', 'glass', 'hang', 't', 'light'],
['singl', 'heart', 'zinc', 't', 'light', 'holder'],
['hang', 'metal', 'star', 'lantern'],
['silver', 'hang', 't', 'light', 'holder'],
['doormat', 'red', 'retrospot'],
['doormat', 'heart'],
['natur', 'slate', 'rectangl', 'chalkboard'],
['lovebird', 'hang', 'decor', 'white'],
['pen', 'assort', 'funni', 'face'],
['card', 'circu', 'parad'],
['wrap', 'cowboy'],
['airlin', 'bag', 'vintag', 'tokyo', '78'],
['set', 'of', '72', 'retrospot', 'paper', 'doili'],
['plaster', 'in', 'tin', 'skull'],
['sleep', 'cat', 'eras'],
['card', 'birthday', 'cowboy'],
['pack', '3', 'box', 'christma', 'panneton'],
['rotat', 'silver', 'angel', 't', 'light', 'hldr'],
['ribbon', 'reel', 'make', 'snowmen'],
['4', 'tradit', 'spin', 'top'],
['bag', '500g', 'swirli', 'marbl'],
['5', 'strand', 'glass', 'necklac', 'crystal'],
['squarecushion', 'cover', 'pink', 'union', 'flag'],
['spaceboy', 'children', 'egg', 'cup'],
['children', 's', 'spaceboy', 'mug'],
['feltcraft', 'cushion', 'owl'],
['black', 'candelabra', 't', 'light', 'holder'],
['toy', 'tidi', 'dolli', 'girl', 'design'],
['set', '12', 'lavend', 'botan', 't', 'light'],
['union', 'jack', 'flag', 'luggag', 'tag'],
['red', 'heart', 'luggag', 'tag'],
['red', 'glass', 'tassl', 'bag', 'charm'],
['clear', 'acryl', 'facet', 'bangl'],
['doormat', 'union', 'jack', 'gun', 'and', 'rose'],
['vanilla', 'scent', 'candl', 'jewel', 'box'],
['full', 'english', 'breakfast', 'plate'],
['magic', 'draw', 'slate', 'circu', 'parad'],
['christma', 'hang', 'heart', 'with', 'bell'],
['cake', 'stand', 'victorian', 'filigre', 'med'],
['paisley', 'pattern', 'sticker'],
['flower', 'sticker'],
['black', 'love', 'bird', 'candl'],
['parti', 'time', 'pencil', 'eras'],
['pink', 'b', 'fli', 'c', 'cover', 'w', 'bobbl'],
['pink', 'union', 'jack', 'luggag', 'tag'],
['6', 'ribbon', 'eleg', 'christma'],
['age', 'glass', 'silver', 't', 'light', 'holder'],
['boom', 'box', 'speaker', 'girl'],
['6', 'ribbon', 'shimmer', 'pink'],
['jumbo', 'bag', 'owl'],
['owl', 'doorstop'],
['paint', 'metal', 'star', 'with', 'holli', 'bell'],
['paint', 'metal', 'heart', 'with', 'holli', 'bell'],
['angel', 'decor', 'star', 'on', 'dress'],
['jumbo', 'bag', 'strawberri'],
['strawberri', 'shopper', 'bag'],
['round', 'snack', 'box', 'set', 'of', '4', 'fruit'],
['round', 'snack', 'box', 'set', 'of', '4', 'skull'],
['dolli', 'girl', 'lunch', 'box'],
['green', 'polkadot', 'plate'],
['blue', 'polkadot', 'plate'],
['red', 'retrospot', 'plate'],
['pink', 'polkadot', 'plate'],
['feltcraft', 'doll', 'molli'],
['feltcraft', 'christma', 'fairi'],
['set', 'of', '3', 'notebook', 'in', 'parcel'],
['red', 'retrospot', 'tape'],
['cosi', 'slipper', 'shoe', 'small', 'red'],
['6', 'ribbon', 'rustic', 'charm'],
['12', 'daisi', 'peg', 'in', 'wood', 'box'],
['5', 'hook', 'hanger', 'red', 'magic', 'toadstool'],
['christma', 'craft', 'tree', 'top', 'angel'],
['christma', 'craft', 'littl', 'friend'],
['cosi', 'slipper', 'shoe', 'small', 'green'],
['feltcraft', 'doll', 'rosi'],
['feltcraft', 'doll', 'emili'],
['feltcraft', 'princess', 'olivia', 'doll'],
['rex', 'cash', 'carri', 'jumbo', 'shopper'],
['feltcraft', 'cushion', 'rabbit'],
['feltcraft', 'cushion', 'butterfli'],
['tote', 'bag', 'i', 'love', 'london'],
['fold', 'umbrella', 'pinkwhit', 'polkadot'],
['fold', 'umbrella', 'cream', 'polkadot'],
['fold', 'umbrella', 'white', 'red', 'polkadot'],
['set', '10', 'light', 'night', 'owl'],
['fold', 'umbrella', 'red', 'white', 'polkadot'],
['s', '4', 'valentin', 'decoupag', 'heart', 'box'],
['soldier', 'egg', 'cup'],
['red', 'retrospot', 'mini', 'case'],
['60', 'cake', 'case', 'vintag', 'christma'],
['ribbon', 'reel', 'christma', 'sock', 'baubl'],
['ribbon', 'reel', 'snowi', 'villag'],
['set', 'of', '20', 'vintag', 'christma', 'napkin'],
['turquois', 'christma', 'tree'],
['red', 'star', 'card', 'holder'],
['wicker', 'wreath', 'small'],
['advent', 'calendar', 'gingham', 'sack'],
['feltcraft', 'butterfli', 'heart'],
['parti', 'cone', 'christma', 'decor'],
['wicker', 'star'],
['vintag', 'snap', 'card'],
['feltcraft', '6', 'flower', 'friend'],
['bird', 'decor', 'red', 'retrospot'],
['christma', 'gingham', 'tree'],
['christma', 'gingham', 'star'],
['christma', 'gingham', 'heart'],
['pack', 'of', '12', 'pink', 'polkadot', 'tissu'],
['feltcraft', 'princess', 'lola', 'doll'],
['pack', 'of', '12', 'london', 'tissu'],
['purpl', 'drawerknob', 'acryl', 'edwardian'],
['children', 's', 'apron', 'dolli', 'girl'],
['children', 'apron', 'spaceboy', 'design'],
['christma', 'light', '10', 'santa'],
['pack', 'of', '60', 'mushroom', 'cake', 'case'],
['60', 'cake', 'case', 'dolli', 'girl', 'design'],
['hand', 'warmer', 'babushka', 'design'],
['set', '12', 'retro', 'white', 'chalk', 'stick'],
['feltcraft', 'hairband', 'red', 'and', 'blue'],
['feltcraft', 'hairband', 'pink', 'and', 'purpl'],
['feltcraft', 'hairband', 'pink', 'and', 'white'],
['tv', 'dinner', 'tray', 'dolli', 'girl'],
['plaster', 'in', 'tin', 'vintag', 'paisley'],
['plaster', 'in', 'tin', 'spaceboy'],
['plaster', 'in', 'tin', 'woodland', 'anim'],
['magic', 'draw', 'slate', 'spaceboy'],
['magic', 'draw', 'slate', 'go', 'to', 'the', 'fair'],
['magic', 'draw', 'slate', 'dolli', 'girl'],
['charli', 'lola', 'red', 'hot', 'water', 'bottl'],
['urban', 'black', 'ribbon'],
['set', 'of', '6', 't', 'light', 'toadstool'],
['rotat', 'leav', 't', 'light', 'holder'],
['cupcak', 'lace', 'paper', 'set', '6'],
['tradit', 'wooden', 'catch', 'cup', 'game'],
['s', '6', 'wooden', 'skittl', 'in', 'cotton', 'bag'],
['red', 'metal', 'beach', 'spade'],
['boy', 'vintag', 'tin', 'seasid', 'bucket'],
['girl', 'vintag', 'tin', 'seasid', 'bucket'],
['grey', 'heart', 'hot', 'water', 'bottl'],
['king', 'choic', 'tea', 'caddi'],
['king', 'choic', 'biscuit', 'tin'],
['metal', 'sign', 'empir', 'tea'],
['fawn', 'blue', 'hot', 'water', 'bottl'],
['pack', 'of', '6', 'birdi', 'gift', 'tag'],
['ceram', 'strawberri', 'money', 'box'],
['ceram', 'heart', 'fairi', 'cake', 'money', 'bank'],
['ceram', 'pirat', 'chest', 'money', 'bank'],
['doorstop', 'retrospot', 'heart'],
['retrospot', 'babushka', 'doorstop'],
['sweetheart', 'wire', 'wall', 'tidi'],
['red', 'retrospot', 'oven', 'glove'],
['namast', 'swagat', 'incens'],
['ceram', 'strawberri', 'design', 'mug'],
['classic', 'rose', 'small', 'vase'],
['hyacinth', 'bulb', 't', 'light', 'candl'],
['christma', 'craft', 'white', 'fairi'],
['cosi', 'hour', 'cigar', 'box', 'match'],
['red', 'tea', 'towel', 'classic', 'design'],
['heart', 'filigre', 'dove', 'small'],
['retrospot', 'children', 'apron'],
['colour', 'pencil', 'brown', 'tube'],
['4', 'ivori', 'dinner', 'candl', 'silver', 'flock'],
['zinc', 'metal', 'heart', 'decor'],
['green', 'fern', 'notebook'],
['blue', 'paisley', 'notebook'],
['chrysanthemum', 'notebook'],
['larg', 'chines', 'style', 'scissor'],
['small', 'chines', 'style', 'scissor'],
['hang', 'glass', 'etch', 'tealight'],
['cook', 'set', 'retrospot'],
['king', 'choic', 'giant', 'tube', 'match'],
['place', 'set', 'white', 'star'],
['s', '15', 'silver', 'glass', 'baubl', 'in', 'bag'],
['french', 'wc', 'sign', 'blue', 'metal'],
['fairi', 'soap', 'soap', 'holder'],
['green', 'giant', 'garden', 'thermomet'],
['blue', 'giant', 'garden', 'thermomet'],
['victorian', 'sew', 'box', 'small'],
['victorian', 'sew', 'box', 'medium'],
['set', '10', 'red', 'polkadot', 'parti', 'candl'],
['washroom', 'metal', 'sign'],
['black', 'sweetheart', 'bracelet'],
['diamant', 'hair', 'grip', 'pack', '2', 'black', 'dia'],
['black', 'diamant', 'expand', 'ring'],
['diamant', 'hair', 'grip', 'pack', '2', 'rubi'],
['diamant', 'hair', 'grip', 'pack', '2', 'montana'],
['blue', 'sweetheart', 'bracelet'],
['pink', 'sweetheart', 'bracelet'],
['purpl', 'sweetheart', 'bracelet'],
['ribbon', 'reel', 'christma', 'present'],
['first', 'aid', 'tin'],
['make', 'your', 'own', 'monsoon', 'card', 'kit'],
['pack', 'of', '12', 'heart', 'design', 'tissu'],
['heart', 'filigre', 'dove', 'larg'],
['swirli', 'circular', 'rubber', 'in', 'bag'],
['vintag', 'paisley', 'stationeri', 'set'],
['set', '10', 'blue', 'polkadot', 'parti', 'candl'],
['ribbon', 'reel', 'sock', 'and', 'mitten'],
['christma', 'metal', 'tag', 'assort'],
['christma', 'tree', 't', 'light', 'holder'],
['star', 't', 'light', 'holder'],
['heart', 't', 'light', 'holder'],
['mini', 'funki', 'design', 'tape'],
['star', 'wooden', 'christma', 'decor'],
['wooden', 'heart', 'christma', 'scandinavian'],
['christma', 'retrospot', 'tree', 'wood'],
['set', '9', 'christma', 't', 'light', 'scent'],
['blue', 'rose', 'fabric', 'mirror'],
['three', 'canva', 'luggag', 'tag'],
['retrospot', 'parti', 'bag', 'sticker', 'set'],
['retrospot', 'small', 'tube', 'match'],
['3d', 'christma', 'stamp', 'sticker'],
['36', 'foil', 'heart', 'cake', 'case'],
['pack', 'of', '6', 'sweeti', 'gift', 'box'],
['cinammon', 'set', 'of', '9', 't', 'light'],
['orang', 'scent', 'set', '9', 't', 'light'],
['ribbon', 'reel', 'stripe', 'design'],
['set', 'of', '2', 'tin', 'jardin', 'de', 'provenc'],
['make', 'your', 'own', 'playtim', 'card', 'kit'],
['dinosaur', 'write', 'set'],
['skull', 'write', 'set'],
['doormat', 'black', 'flock'],
['christma', 'decoupag', 'candl'],
['offic', 'mug', 'warmer', 'polkadot'],
['black', 'record', 'cover', 'frame'],
['record', 'frame', '7', 'singl', 'size'],
['36', 'pencil', 'tube', 'red', 'retrospot'],
['36', 'pencil', 'tube', 'woodland'],
['rock', 'hors', 'red', 'christma'],
['regenc', 'cakestand', '3', 'tier'],
['vintag', 'union', 'jack', 'bunt'],
['retro', 'longboard', 'iron', 'board', 'cover'],
['3d', 'dog', 'pictur', 'play', 'card'],
['save', 'the', 'planet', 'cotton', 'tote', 'bag'],
['paper', 'bunt', 'white', 'lace'],
['gentleman', 'shirt', 'repair', 'kit'],
['card', 'i', 'love', 'london'],
['mirror', 'wall', 'art', 'foxi'],
['mirror', 'wall', 'art', 'gent'],
['airlin', 'bag', 'vintag', 'jet', 'set', 'white'],
['pot', 'shed', 'sow', 'n', 'grow', 'set'],
['airlin', 'bag', 'vintag', 'world', 'champion'],
['pink', 'heart', 'paper', 'garland'],
['pot', 'shed', 'seed', 'envelop'],
['hi', 'tec', 'alpin', 'hand', 'warmer'],
['polyest', 'filler', 'pad', '45x45cm'],
['card', 'dolli', 'girl'],
['card', 'billboard', 'font'],
['raini', 'ladi', 'birthday', 'card'],
['red', 'white', 'scarf', 'hot', 'water', 'bottl'],
['set', 'of', '2', 'tea', 'towel', 'appl', 'and', 'pear'],
['lantern', 'cream', 'gazebo'],
['blue', 'birdhous', 'decor'],
['antiqu', 'silver', 'tea', 'glass', 'engrav'],
['set', 'of', '2', 'wooden', 'market', 'crate'],
['classic', 'french', 'style', 'basket', 'natur'],
['crazi', 'daisi', 'heart', 'decor'],
['card', 'holder', 'gingham', 'heart'],
['small', 'stripe', 'chocol', 'gift', 'bag'],
['tray', 'breakfast', 'in', 'bed'],
['pizza', 'plate', 'in', 'box'],
['pink', 'heart', 'dot', 'hot', 'water', 'bottl'],
['place', 'set', 'white', 'heart'],
['set', 'of', '4', 'napkin', 'charm', 'leav'],
['set', 'of', '4', 'napkin', 'charm', 'star'],
['set', 'of', '4', 'napkin', 'charm', 'cutleri'],
['caravan', 'squar', 'tissu', 'box'],
['button', 'box'],
['12', 'messag', 'card', 'with', 'envelop'],
['small', 'red', 'babushka', 'notebook'],
['herb', 'marker', 'mint'],
['herb', 'marker', 'rosemari'],
['herb', 'marker', 'basil'],
['herb', 'marker', 'chive'],
['herb', 'marker', 'parsley'],
['herb', 'marker', 'thyme'],
['metal', 'merri', 'christma', 'wreath'],
['white', 'christma', 'garland', 'star', 'tree'],
['retrospot', 'giant', 'tube', 'match'],
['36', 'foil', 'star', 'cake', 'case'],
['pack', 'of', '6', 'handbag', 'gift', 'box'],
['garland', 'wooden', 'happi', 'easter'],
['calendar', 'paper', 'cut', 'design'],
['calendar', 'in', 'season', 'design'],
['blue', 'paisley', 'tissu', 'box'],
['pink', 'paisley', 'squar', 'tissu', 'box'],
['toast', 'it', 'happi', 'birthday'],
['set', '10', 'ivori', 'polkadot', 'parti', 'candl'],
['skull', 'storag', 'box', 'small'],
['jumbo', 'storag', 'bag', 'skull'],
['garden', 'path', 'journal'],
['black', 'orang', 'squeezer'],
['set', 'of', '72', 'green', 'paper', 'doili'],
['retrospot', 'red', 'wash', 'up', 'glove'],
['small', 'licoric', 'de', 'pink', 'bowl'],
['swiss', 'roll', 'towel', 'chocol', 'spot'],
['picnic', 'basket', 'wicker', 'larg'],
['cream', 'sweetheart', 'egg', 'holder'],
['pink', '3', 'piec', 'polkadot', 'cutleri', 'set'],
['pink', 'parti', 'bag'],
['blue', 'parti', 'bag'],
['pink', 'cream', 'felt', 'craft', 'trinket', 'box'],
['red', 'retrospot', 'oven', 'glove', 'doubl'],
['ridg', 'glass', 'storag', 'jar', 'cream', 'lid'],
['pack', 'of', '20', 'napkin', 'pantri', 'design'],
['sweetheart', 'cakestand', '3', 'tier'],
['christma', 'hang', 'star', 'with', 'bell'],
['multicolour', 'confetti', 'in', 'tube'],
['chocol', '3', 'wick', 'morri', 'box', 'candl'],
['set', 'of', '16', 'vintag', 'pistachio', 'cutleri'],
['christma', 'toilet', 'roll'],
['white', 'bell', 'honeycomb', 'paper'],
['4', 'purpl', 'flock', 'dinner', 'candl'],
['pink', 'white', 'breakfast', 'tray'],
['english', 'rose', 'spirit', 'level'],
['larg', 'round', 'wicker', 'platter'],
['black', 'tea', 'towel', 'classic', 'design'],
['christma', 'light', '10', 'vintag', 'baubl'],
['rose', 'regenc', 'teacup', 'and', 'saucer'],
['t', 'light', 'holder', 'hang', 'lace'],
['hang', 'heart', 'mirror', 'decor'],
['cream', 'slice', 'flannel', 'chocol', 'spot'],
['cream', 'slice', 'flannel', 'pink', 'spot'],
['green', 'regenc', 'teacup', 'and', 'saucer'],
['eight', 'piec', 'dinosaur', 'set'],
['head', 'and', 'tail', 'sport', 'fun'],
['ass', 'col', 'small', 'sand', 'gecko', 'p', 'weight'],
['enamel', 'measur', 'jug', 'cream'],
['12', 'pencil', 'small', 'tube', 'skull'],
['12', 'pencil', 'small', 'tube', 'woodland'],
['blue', 'paper', 'parasol'],
['victorian', 'metal', 'postcard', 'spring'],
['metal', '4', 'hook', 'hanger', 'french', 'chateau'],
['black', 'kitchen', 'scale'],
['red', 'kitchen', 'scale'],
['ivori', 'kitchen', 'scale'],
['multi', 'colour', 'silver', 't', 'light', 'holder'],
['3', 'hook', 'hanger', 'magic', 'garden'],
['assort', 'colour', 'lizard', 'suction', 'hook'],
['jumbo', 'bag', 'woodland', 'anim'],
['children', 's', 'circu', 'parad', 'mug'],
['pack', '3', 'fire', 'engin', 'car', 'patch'],
['card', 'psychedel', 'appl'],
['ivori', 'enchant', 'forest', 'placemat'],
['pack', 'of', '6', 'panneton', 'gift', 'box'],
['measur', 'tape', 'babushka', 'red'],
['red', 'enchant', 'forest', 'placemat'],
['children', 'dolli', 'girl', 'mug'],
['toadstool', 'money', 'box'],
['make', 'your', 'own', 'flowerpow', 'card', 'kit'],
['paper', 'chain', 'kit', 'london'],
['measur', 'tape', 'babushka', 'blue'],
['pencil', 'case', 'life', 'is', 'beauti'],
['wooden', 'rounder', 'garden', 'set'],
['child', 'breakfast', 'set', 'dolli', 'girl'],
['child', 'breakfast', 'set', 'spaceboy'],
['cake', 'plate', 'lovebird', 'white'],
['wooden', 'advent', 'calendar', 'red'],
['green', 'christma', 'tree', 'card', 'holder'],
['string', 'of', 'star', 'card', 'holder'],
['set', '12', 'taper', 'candl'],
['mint', 'kitchen', 'scale'],
['white', 'wood', 'garden', 'plant', 'ladder'],
['glass', 'jar', 'english', 'confectioneri'],
['glass', 'jar', 'digest', 'biscuit'],
['milk', 'bottl', 'with', 'glass', 'stopper'],
['welcom', 'wooden', 'block', 'letter'],
['set', '4', 'modern', 'vintag', 'cotton', 'napkin'],
['vintag', 'union', 'jack', 'apron'],
['swallow', 'squar', 'tissu', 'box'],
['no', 'junk', 'mail', 'metal', 'sign'],
['slate', 'tile', 'natur', 'hang'],
['vintag', 'glass', 'coffe', 'caddi'],
['silver', 'glass', 't', 'light', 'set'],
['ridg', 'glass', 'finger', 'bowl'],
['glass', 'beurr', 'dish'],
['gin', 'tonic', 'diet', 'greet', 'card'],
['birthday', 'card', 'retro', 'spot'],
['wrap', 'red', 'appl'],
['red', 'retrospot', 'wrap'],
['wrap', 'christma', 'screen', 'print'],
['fanci', 'font', 'birthday', 'wrap'],
['classic', 'white', 'frame'],
['black', 'diner', 'wall', 'clock'],
['doormat', 'peac', 'on', 'earth', 'blue'],
['box', 'of', '24', 'cocktail', 'parasol'],
['grow', 'your', 'own', 'plant', 'in', 'a', 'can'],
['plaster', 'in', 'tin', 'circu', 'parad'],
['plaster', 'in', 'tin', 'strongman'],
['mini', 'jigsaw', 'dinosaur'],
['mini', 'jigsaw', 'bake', 'a', 'cake'],
['mini', 'jigsaw', 'dolli', 'girl'],
['wooden', 'school', 'colour', 'set'],
['pack', 'of', '12', 'tradit', 'crayon'],
['card', 'parti', 'game'],
['booz', 'women', 'greet', 'card'],
['pack', 'of', '12', 'suki', 'tissu'],
['pack', 'of', '12', 'woodland', 'tissu'],
['pack', 'of', '12', 'skull', 'tissu'],
['pack', 'of', '12', 'blue', 'paisley', 'tissu'],
['pack', 'of', '12', 'circu', 'parad', 'tissu'],
['pack', 'of', '12', 'spaceboy', 'tissu'],
['lad', 'onli', 'tissu', 'box'],
['red', 'retrospot', 'tissu', 'box'],
['pack', 'of', '12', 'pink', 'paisley', 'tissu'],
['gingerbread', 'man', 'cooki', 'cutter'],
['gin', 'and', 'tonic', 'mug'],
['if', 'you', 'can', 't', 'stand', 'the', 'heat', 'mug'],
['i', 'can', 'onli', 'pleas', 'one', 'person', 'mug'],
['pink', 'heart', 'shape', 'egg', 'fri', 'pan'],
['200', 'red', 'white', 'bendi', 'straw'],
['recip', 'box', 'retrospot'],
['recip', 'box', 'pantri', 'yellow', 'design'],
['recip', 'box', 'blue', 'sketchbook', 'design'],
['20', 'dolli', 'peg', 'retrospot'],
['no', 'sing', 'metal', 'sign'],
['biscuit', 'tin', 'vintag', 'red'],
['emerg', 'first', 'aid', 'tin'],
['polkadot', 'pen'],
['decor', 'plant', 'pot', 'with', 'friez'],
['french', 'blue', 'metal', 'door', 'sign', '5'],
['chest', 'of', 'drawer', 'gingham', 'heart'],
['red', 'retrospot', 'cake', 'stand'],
['joy', 'wooden', 'block', 'letter'],
['noel', 'wooden', 'block', 'letter'],
['peac', 'wooden', 'block', 'letter'],
['offic', 'mug', 'warmer', 'black', 'silver'],
['set', 'of', '2', 'christma', 'decoupag', 'candl'],
['cosmet', 'bag', 'vintag', 'rose', 'paisley'],
['pink', 'rose', 'fabric', 'mirror'],
['silver', 'plate', 'candl', 'bowl', 'small'],
['small', 'regal', 'silver', 'candlepot'],
['antiqu', 'tall', 'swirlglass', 'trinket', 'pot'],
['antiqu', 'glass', 'pedest', 'bowl'],
['gumbal', 'magazin', 'rack'],
['red', 'flower', 'crochet', 'food', 'cover'],
['cherri', 'crochet', 'food', 'cover'],
['strawberri', 'raffia', 'food', 'cover'],
['polka', 'dot', 'raffia', 'food', 'cover'],
['doormat', 'spotti', 'home', 'sweet', 'home'],
['t', 'light', 'glass', 'flute', 'antiqu'],
['silver', 'candlepot', 'jardin'],
['polkadot', 'mug', 'pink'],
['paper', 'bunt', 'colour', 'lace'],
['antiqu', 'glass', 'heart', 'decor'],
['set', 'of', '6', 'kashmir', 'folkart', 'baubl'],
['christma', 'retrospot', 'angel', 'wood'],
['christma', 'retrospot', 'star', 'wood'],
['heart', 'wooden', 'christma', 'decor'],
['larg', 'hang', 'ivori', 'red', 'wood', 'bird'],
['small', 'hang', 'ivori', 'red', 'wood', 'bird'],
['grand', 'chocolatecandl'],
['fairi', 'cake', 'birthday', 'candl', 'set'],
['3', 'hook', 'photo', 'shelf', 'antiqu', 'white'],
['cream', 'sweetheart', 'wall', 'cabinet'],
['doormat', 'english', 'rose'],
['doormat', 'union', 'flag'],
['romant', 'imag', 'notebook', 'set'],
['curiou', 'imag', 'notebook', 'set'],
['vintag', 'keepsak', 'box', 'pari', 'day'],
['red', 'gingham', 'rose', 'jewelleri', 'box'],
['blue', 'charli', 'lola', 'person', 'doorsign'],
['charli', 'lola', 'extrem', 'busi', 'sign'],
['charlott', 'bag', 'suki', 'design'],
['lolita', 'design', 'cotton', 'tote', 'bag'],
['let', 'go', 'shop', 'cotton', 'tote', 'bag'],
['paint', 'your', 'own', 'canva', 'set'],
['camouflag', 'led', 'torch'],
['blue', 'dragonfli', 'helicopt'],
['yellow', 'shark', 'helicopt'],
['red', 'shark', 'helicopt'],
['tool', 'box', 'soft', 'toy'],
['doctor', 's', 'bag', 'soft', 'toy'],
['carriag'],
['jumbo', 'bag', 'scandinavian', 'paisley'],
['jumbo', 'bag', 'toy'],
['decor', 'rose', 'bathroom', 'bottl'],
['decor', 'cat', 'bathroom', 'bottl'],
['rain', 'poncho', 'retrospot'],
['christma', 'tree', 'decor', 'with', 'bell'],
['christma', 'tree', 'heart', 'decor'],
['christma', 'tree', 'star', 'decor'],
['christma', 'hang', 'tree', 'with', 'bell'],
['glitter', 'star', 'garland', 'with', 'bell'],
['glitter', 'christma', 'tree', 'with', 'bell'],
['lilac', 'diamant', 'pen', 'in', 'gift', 'box'],
['heart', 'gift', 'tape'],
['cake', 'and', 'bow', 'gift', 'tape'],
['bingo', 'set'],
['garden', 'metal', 'sign'],
['union', 'stripe', 'with', 'fring', 'hammock'],
['suki', 'shoulder', 'bag'],
['skull', 'design', 'flannel'],
['kitten', 'design', 'flannel'],
['set', 'of', '6', 'strawberri', 'chopstick'],
['modern', 'floral', 'stationeri', 'set'],
['bohemian', 'collag', 'stationeri', 'set'],
['floral', 'folk', 'stationeri', 'set'],
['robot', 'birthday', 'card'],
['christma', 'card', 'screen', 'print'],
['christma', 'pud', 'trinket', 'pot'],
['choc', 'truffl', 'gold', 'trinket', 'pot'],
['brown', 'pirat', 'treasur', 'chest'],
['small', 'white', 'retrospot', 'mug', 'in', 'box'],
['6', 'ribbon', 'empir'],
['ribbon', 'reel', 'polkadot'],
['ribbon', 'reel', 'flora', 'fauna'],
['paper', 'bunt', 'retrospot'],
['ladl', 'love', 'heart', 'red'],
['ladl', 'love', 'heart', 'pink'],
['angel', 'decor', '3', 'button'],
['star', 'decor', 'rustic'],
['heart', 'decor', 'with', 'pearl'],
['heart', 'decor', 'rustic', 'hang'],
['heart', 'garland', 'rustic', 'pad'],
['famili', 'album', 'white', 'pictur', 'frame'],
['cake', 'stand', 'victorian', 'filigre', 'small'],
['local', 'cafe', 'mug'],
['milk', 'pan', 'red', 'retrospot'],
['fri', 'pan', 'union', 'flag'],
['easter', 'decor', 'natur', 'chick'],
['wash', 'bag', 'vintag', 'rose', 'paisley'],
['pig', 'keyr', 'with', 'light', 'sound'],
['coffe', 'mug', 'dog', 'ball', 'design'],
['coffe', 'mug', 'cat', 'bird', 'design'],
['tea', 'cosi', 'red', 'stripe'],
['offic', 'mug', 'warmer', 'choc', 'blue'],
['heart', 'decor', 'paint', 'zinc'],
['dove', 'decor', 'paint', 'zinc'],
['parti', 'pizza', 'dish', 'blue', 'polkadot'],
['tea', 'bag', 'plate', 'red', 'retrospot'],
['charlott', 'bag', 'pink', 'polkadot'],
['glass', 'jar', 'king', 'choic'],
['glass', 'jar', 'daisi', 'fresh', 'cotton', 'wool'],
['airlin', 'bag', 'vintag', 'jet', 'set', 'brown'],
['wall', 'tidi', 'retrospot'],
['toy', 'tidi', 'spaceboy'],
['paperweight', 'king', 'choic'],
['magnet', 'pack', 'of', '4', 'retro', 'photo'],
['lipstick', 'pen', 'red'],
['toothpast', 'tube', 'pen'],
['enamel', 'water', 'can', 'cream'],
['water', 'can', 'pink', 'bunni'],
['set', 'of', '9', 'black', 'skull', 'balloon'],
['tv', 'dinner', 'tray', 'air', 'hostess'],
['empir', 'union', 'jack', 'tv', 'dinner', 'tray'],
['water', 'can', 'garden', 'marker'],
['birdhous', 'garden', 'marker'],
['daisi', 'garden', 'marker'],
['set', 'of', '2', 'tin', 'vintag', 'bathroom'],
['child', 'garden', 'spade', 'pink'],
['child', 'garden', 'rake', 'pink'],
['child', 'garden', 'brush', 'pink'],
['mini', 'jigsaw', 'bunni'],
['holiday', 'fun', 'ludo'],
['tradit', 'model', 'clay'],
['cardhold', 'gingham', 'star'],
['wooden', 'croquet', 'garden', 'set'],
['set', 'of', '4', 'napkin', 'charm', 'heart'],
['red', 'babi', 'bunt'],
['french', 'kitchen', 'sign', 'blue', 'metal'],
['set', 'of', '6', 'ribbon', 'vintag', 'christma'],
['funki', 'diva', 'pen'],
['small', 'purpl', 'babushka', 'notebook'],
['larg', 'red', 'babushka', 'notebook'],
['larg', 'purpl', 'babushka', 'notebook'],
['chalkboard', 'kitchen', 'organis'],
['flute', 'antiqu', 'candl', 'holder'],
['card', 'motorbik', 'santa'],
['card', 'christma', 'villag'],
['cream', 'wall', 'planter', 'heart', 'shape'],
['vintag', 'cream', 'dog', 'food', 'contain'],
['bread', 'bin', 'diner', 'style', 'ivori'],
['love', 'heart', 'napkin', 'box'],
['number', 'tile', 'cottag', 'garden', '8'],
['yellow', 'giant', 'garden', 'thermomet'],
['metal', 'decor', 'naughti', 'children'],
['36', 'doili', 'dolli', 'girl'],
['circu', 'parad', 'children', 'egg', 'cup'],
['blue', 'victorian', 'fabric', 'oval', 'box'],
['red', 'victorian', 'fabric', 'oval', 'box'],
['set', 'of', '3', 'bird', 'light', 'pink', 'feather'],
['pink', 'white', 'christma', 'tree', '60cm'],
['pink', 'and', 'white', 'christma', 'tree', '120cm'],
['pink', 'christma', 'flock', 'droplet'],
['acryl', 'jewel', 'icicl', 'pink'],
['smallfolkart', 'baubl', 'christma', 'dec'],
['folkart', 'zinc', 'heart', 'christma', 'dec'],
['rose', 'folkart', 'heart', 'decor'],
['ceram', 'cake', 'stand', 'hang', 'cake'],
['condiment', 'tray', '4', 'bowl', 'and', '4', 'spoon'],
['tea', 'time', 'oven', 'glove'],
['scotti', 'children', 'apron'],
['pink', 'fairi', 'cake', 'children', 'apron'],
['scotti', 'dog', 'babi', 'bib'],
['doormat', 'welcom', 'puppi'],
['hang', 'jam', 'jar', 't', 'light', 'holder'],
['set', 'of', '6', 'halloween', 'ghost', 't', 'light'],
['ivori', 'pillar', 'candl', 'silver', 'flock'],
['ivori', 'pillar', 'candl', 'gold', 'flock'],
['set', '3', 'rose', 'candl', 'in', 'jewel', 'box'],
['set', '3', 'ocean', 'scent', 'candl', 'jewel', 'box'],
['set', '3', 'vanilla', 'scent', 'candl', 'in', 'box'],
['set', '3', 'christma', 'decoupag', 'candl'],
['tumbler', 'baroqu'],
['tumbler', 'new', 'england'],
['laundri', '15c', 'metal', 'sign'],
['hot', 'bath', 'metal', 'sign'],
['metal', 'sign', 'cupcak', 'singl', 'hook'],
['metal', 'sign', 'cupcak', 'singl', 'hook'],
['metal', 'sign', 'cupcak', 'singl', 'hook'],
['charli', 'lola', 'red', 'hot', 'water', 'bottl'],
['charli', 'lola', 'blue', 'hot', 'water', 'bottl'],
['charli', 'lola', 'pink', 'hot', 'water', 'bottl'],
['pink', 'purpl', 'retro', 'radio'],
['cherri', 'blossom', 'decor', 'flask'],
['set', 'of', '12', 'vintag', 'postcard', 'set'],
['set', 'of', '6', 'vintag', 'notelet', 'kit'],
['envelop', '50', 'romant', 'imag'],
['envelop', '50', 'blossom', 'imag'],
['yuletid', 'imag', 'gift', 'wrap', 'set'],
['set', '4', 'red', 'mini', 'rose', 'candl', 'in', 'bowl'],
['15cm', 'christma', 'glass', 'ball', '20', 'light'],
['cream', 'sweetheart', 'letter', 'rack'],
['small', 'squar', 'cut', 'glass', 'candlestick'],
['bead', 'crystal', 'heart', 'pink', 'small'],
['bead', 'crystal', 'heart', 'green', 'on', 'stick'],
['charli', 'lola', 'biscuit', 'tin'],
['red', 'dragonfli', 'helicopt'],
['sew', 'susan', '21', 'needl', 'set'],
['basket', 'of', 'flower', 'sew', 'kit'],
['victorian', 'sew', 'kit'],
['easter', 'bunni', 'garland', 'of', 'flower'],
['set', 'of', '3', 'babushka', 'stack', 'tin'],
['midnight', 'blue', 'pair', 'heart', 'hair', 'slide'],
['silver', 'm', 'o', 'p', 'orbit', 'drop', 'ear'],
['edwardian', 'drop', 'ear', 'jet', 'black'],
['rubi', 'glass', 'cluster', 'ear'],
['necklac', 'bracelet', 'set', 'blue', 'hibiscu'],
['fruit', 'salad', 'bag', 'charm'],
['green', 'murano', 'twist', 'bracelet'],
['pink', 'glass', 'tassl', 'bag', 'charm'],
['turquois', 'glass', 'tassl', 'bag', 'charm'],
['ant', 'copper', 'red', 'boudicca', 'bracelet'],
['silver', 'amethyst', 'drop', 'ear', 'leaf'],
['silver', 'lariat', 'black', 'stone', 'ear'],
['green', 'enamel', 'glass', 'hair', 'comb'],
['letter', 'd', 'bling', 'key', 'ring'],
['letter', 'g', 'bling', 'key', 'ring'],
['letter', 'h', 'bling', 'key', 'ring'],
['letter', 'j', 'bling', 'key', 'ring'],
['letter', 'r', 'bling', 'key', 'ring'],
['dotcom', 'postag'],
['fairi', 'cake', 'notebook', 'a5', 'size'],
['fairi', 'cake', 'notebook', 'a6', 'size'],
['english', 'rose', 'notebook', 'a7', 'size'],
['fairi', 'cake', 'notebook', 'a7', 'size'],
['mous', 'toy', 'with', 'pink', 't', 'shirt'],
['dog', 'toy', 'with', 'pink', 'crochet', 'skirt'],
['boy', 'alphabet', 'iron', 'on', 'patch'],
['green', 'rose', 'washbag'],
['soft', 'pink', 'rose', 'towel'],
['mint', 'green', 'rose', 'towel'],
['pink', 'butterfli', 'handbag', 'w', 'bobbl'],
['antiqu', 'silver', 'tea', 'glass', 'etch'],
['set', 'of', '72', 'pink', 'heart', 'paper', 'doili'],
['denim', 'patch', 'purs', 'pink', 'butterfli'],
['blue', 'patch', 'purs', 'pink', 'heart'],
          ['red', 'spotti', 'biscuit', 'tin'],
          ['vintag', 'cream', 'cat', 'food', 'contain'],
          ['asstd', 'design', 'race', 'car', 'pen'],
          ['kitti', 'pencil', 'eras'],
          ['brocad', 'ring', 'purs'],
          ['origami', 'sandlewood', 'incens', 'flower'],
          ['origami', 'vanilla', 'incens', 'candl', 'set'],
          ['origami', 'jasmin', 'incens', 'candl', 'set'],
          ['origami', 'lavend', 'incens', 'candl', 'set'],
          ['origami', 'rose', 'incens', 'candl', 'set'],
          ['origami', 'opium', 'incens', 'candl', 'set'],
          ['origami', 'sandlewood', 'incens', 'cand', 'set'],
          ['blue', 'glass', 'gem', 'in', 'bag'],
          ['porcelain', 'butterfli', 'oil', 'burner'],
          ['vippassport', 'cover'],
          ['red', 'retrospot', 'luggag', 'tag'],
          ['first', 'class', 'holiday', 'purs'],
          ['red', 'retrospot', 'bowl'],
          ['strawberri', 'dream', 'child', 'umbrella'],
          ['woodland', 'charlott', 'bag'],
          ['red', 'retrospot', 'charlott', 'bag'],
          ['gold', 'mini', 'tape', 'measur'],
          ['black', 'mini', 'tape', 'measur'],
          ['abstract', 'circl', 'journal'],
          ['french', 'lattic', 'cushion', 'cover'],
          ['french', 'paisley', 'cushion', 'cover'],
          ['french', 'floral', 'cushion', 'cover'],
          ['porcelain', 't', 'light', 'holder', 'assort'],
          ['red', 'floral', 'feltcraft', 'shoulder', 'bag'],
          ['pink', 'blue', 'felt', 'craft', 'trinket', 'box'],
          ['12', 'pencil', 'small', 'tube', 'red', 'retrospot'],
          ['12', 'pencil', 'tall', 'tube', 'woodland'],
          ['jazz', 'heart', 'address', 'book'],
          ...]
```

```
In [28]: from gensim import corpora, models

         dictionary = corpora.Dictionary(texts)
```

```
In [29]: corpus = [dictionary.doc2bow(text) for text in texts]
```

```
In [30]: %%time
         ldamodel = gensim.models.ldamodel.LdaModel(corpus, num_topics=10, id2word = dictionary,
```

```
CPU times: user 1min 24s, sys: 8 ms, total: 1min 24s
Wall time: 1min 24s
```

```
In [31]: ldamodel[corpus[0]]
```

```
Out[31]: [(0, 0.014290038097251899),
          (1, 0.67984400020886682),
          (2, 0.014285714290632264),
          (3, 0.014286441191028693),
          (4, 0.20585210712574467),
          (5, 0.014285714293329692),
          (6, 0.01428601340584398),
          (7, 0.014291743146230055),
          (8, 0.014289446767260549),
          (9, 0.014288781473811477)]
```

```
In [32]: sorted(ldamodel[corpus[0]], key=lambda x: -x[1])
         # sorted(ldamodel[corpus[0]], key=lambda x: -x[1])[0][0]
```

```
Out[32]: [(1, 0.6797963853761807),
          (4, 0.20589973110575185),
          (7, 0.014291741929744449),
          (0, 0.014290033885212446),
          (8, 0.014289446662695621),
          (9, 0.014288778493976542),
          (3, 0.014286440488835688),
          (6, 0.014286013474124979),
          (5, 0.014285714293326453),
          (2, 0.014285714290630176)]
```

```
In [33]: product_topics = [sorted(ldamodel[product], key=lambda x: -x[1])[0][0] for product in c
```

```
In [34]: product_topics[:10]
```

```
Out[34]: [1, 4, 3, 1, 4, 0, 1, 9, 7, 1]
```

```
In [35]: products['topic'] = product_topics

         products.index = products.StockCode
         products = products.drop('StockCode' ,1)
         products.head()
```

```
Out[35]:                                 Description  topic
         StockCode
         85123A      WHITE HANGING HEART T-LIGHT HOLDER      1
         71053                      WHITE METAL LANTERN      4
         84406B          CREAM CUPID HEARTS COAT HANGER      3
         84029G     KNITTED UNION FLAG HOT WATER BOTTLE      1
         84029E          RED WOOLLY HOTTIE WHITE HEART.      4
```

```
In [36]: products['topic'].value_counts()
```

```
Out[36]: 0    556
         7    501
```

```
3    495
9    468
1    420
4    419
5    373
8    342
6    300
2    185
Name: topic, dtype: int64
```

```
In [37]: data.head()
```

```
Out[37]:   InvoiceNo StockCode                          Description  Quantity  \
         0    536365    85123A   WHITE HANGING HEART T-LIGHT HOLDER         6
         1    536365     71053                  WHITE METAL LANTERN         6
         2    536365    84406B       CREAM CUPID HEARTS COAT HANGER         8
         3    536365    84029G  KNITTED UNION FLAG HOT WATER BOTTLE         6
         4    536365    84029E       RED WOOLLY HOTTIE WHITE HEART.         6

                   InvoiceDate  UnitPrice CustomerID         Country  TotalPrice
         0 2010-12-01 08:26:00       2.55    17850.0  United Kingdom       15.30
         1 2010-12-01 08:26:00       3.39    17850.0  United Kingdom       20.34
         2 2010-12-01 08:26:00       2.75    17850.0  United Kingdom       22.00
         3 2010-12-01 08:26:00       3.39    17850.0  United Kingdom       20.34
         4 2010-12-01 08:26:00       3.39    17850.0  United Kingdom       20.34
```

```
In [38]: data_with_topic = data.join(products[['topic']], on='StockCode', rsuffix='_')
         data_with_topic.head(10)
```

```
Out[38]:   InvoiceNo StockCode                          Description  Quantity  \
         0    536365    85123A   WHITE HANGING HEART T-LIGHT HOLDER         6
         1    536365     71053                  WHITE METAL LANTERN         6
         2    536365    84406B       CREAM CUPID HEARTS COAT HANGER         8
         3    536365    84029G  KNITTED UNION FLAG HOT WATER BOTTLE         6
         4    536365    84029E       RED WOOLLY HOTTIE WHITE HEART.         6
         5    536365     22752         SET 7 BABUSHKA NESTING BOXES         2
         6    536365     21730    GLASS STAR FROSTED T-LIGHT HOLDER         6
         7    536366     22633               HAND WARMER UNION JACK         6
         8    536366     22632            HAND WARMER RED POLKA DOT         6
         9    536367     84879        ASSORTED COLOUR BIRD ORNAMENT        32

                   InvoiceDate  UnitPrice CustomerID         Country  TotalPrice  topic
         0 2010-12-01 08:26:00       2.55    17850.0  United Kingdom       15.30      1
         1 2010-12-01 08:26:00       3.39    17850.0  United Kingdom       20.34      4
         2 2010-12-01 08:26:00       2.75    17850.0  United Kingdom       22.00      3
         3 2010-12-01 08:26:00       3.39    17850.0  United Kingdom       20.34      1
         4 2010-12-01 08:26:00       3.39    17850.0  United Kingdom       20.34      4
         5 2010-12-01 08:26:00       7.65    17850.0  United Kingdom       15.30      0
         6 2010-12-01 08:26:00       4.25    17850.0  United Kingdom       25.50      1
```

```
7 2010-12-01 08:28:00       1.85    17850.0  United Kingdom       11.10      9
8 2010-12-01 08:28:00       1.85    17850.0  United Kingdom       11.10      7
9 2010-12-01 08:34:00       1.69    13047.0  United Kingdom       54.08      1
```

```
In [39]: ["topic_{}".format(i) for i in range(10)]
```

```
Out[39]: ['topic_0',
          'topic_1',
          'topic_2',
          'topic_3',
          'topic_4',
          'topic_5',
          'topic_6',
          'topic_7',
          'topic_8',
          'topic_9']
```

```
In [40]: topics_per_customer = (data_with_topic.groupby(['CustomerID', 'topic'])
                   .size().unstack().reset_index().fillna(0)
                   .set_index('CustomerID'))

         topics_per_customer.columns = ["topic_{}".format(i) for i in range(10)]
         topics_per_customer.head(10)
```

```
Out[40]:             topic_0  topic_1  topic_2  topic_3  topic_4  topic_5  topic_6  \
         CustomerID
         12346.0         0.0      0.0      0.0      0.0      0.0      1.0      0.0
         12347.0        30.0     19.0     14.0     13.0     26.0     24.0     12.0
         12348.0        19.0      0.0      1.0      0.0      0.0      5.0      0.0
         12349.0        15.0     10.0      1.0      6.0      5.0     14.0      9.0
         12350.0         0.0      0.0      0.0      1.0      0.0      4.0      1.0
         12352.0         9.0     11.0      2.0      7.0      3.0     17.0      5.0
         12353.0         3.0      0.0      0.0      0.0      0.0      1.0      0.0
         12354.0         6.0      5.0      3.0      2.0      4.0     19.0      2.0
         12355.0         4.0      0.0      0.0      1.0      0.0      3.0      0.0
         12356.0        22.0      1.0      5.0      3.0      4.0     15.0      0.0

                     topic_7  topic_8  topic_9
         CustomerID
         12346.0         0.0      0.0      0.0
         12347.0        17.0     17.0     10.0
         12348.0         4.0      2.0      0.0
         12349.0         4.0      5.0      4.0
         12350.0         3.0      5.0      3.0
         12352.0        12.0      6.0     13.0
         12353.0         0.0      0.0      0.0
         12354.0         4.0      4.0      9.0
         12355.0         3.0      0.0      2.0
         12356.0         5.0      4.0      0.0
```

```
In [41]: customer_data.head()
```

```
Out[41]:             TransactionCount  TotalItemCount  Valuation         Country  \
         CustomerID
         12346.0                    1           74215   77183.60  United Kingdom
         12347.0                  182            2458    4310.00         Iceland
         12348.0                   31            2341    1797.24         Finland
         12349.0                   73             631    1757.55           Italy
         12350.0                   17             197     334.40          Norway

                     RatioToMeanTransactionCount  DifFromMeanTransactionCount  \
         CustomerID
         12346.0                        0.008148                  -121.723733
         12347.0                        1.483006                    59.276267
         12348.0                        0.252600                   -91.723733
         12349.0                        0.594832                   -49.723733
         12350.0                        0.138523                  -105.723733

                    HigherThanMeanTransactionCount  TransactionCountPercentile  \
         CustomerID
         12346.0                             False                    0.841014
         12347.0                              True                   88.087558
         12348.0                             False                   42.108295
         12349.0                             False                   66.555300
         12350.0                             False                   24.919355

                     RatioToMeanTransactionCountPerCountry
         CustomerID
         12346.0                                  0.008020
         12347.0                                  1.000000
         12348.0                                  0.543066
         12349.0                                  1.348285
         12350.0                                  0.158582
```

```
In [42]: customer_data.join(topics_per_customer)
```

```
Out[42]:             TransactionCount  TotalItemCount     Valuation         Country  \
         CustomerID
         12346.0                    1           74215  7.718360e+04  United Kingdom
         12347.0                  182            2458  4.310000e+03         Iceland
         12348.0                   31            2341  1.797240e+03         Finland
         12349.0                   73             631  1.757550e+03           Italy
         12350.0                   17             197  3.344000e+02          Norway
         12352.0                   85             536  2.506040e+03          Norway
         12353.0                    4              20  8.900000e+01         Bahrain
         12354.0                   58             530  1.079400e+03           Spain
         12355.0                   13             240  4.594000e+02         Bahrain
         12356.0                   59            1591  2.811430e+03        Portugal
```

```
12357.0               131         2708  6.207670e+03     Switzerland
12358.0                19          248  1.168060e+03         Austria
12359.0               248         1622  6.372580e+03          Cyprus
12360.0               129         1165  2.662060e+03         Austria
12361.0                10           91  1.899000e+02         Belgium
12362.0               266         2229  5.226230e+03         Belgium
12363.0                23          408  5.520000e+02     Unspecified
12364.0                85         1506  1.313100e+03         Belgium
12365.0                22          174  6.413800e+02          Cyprus
12367.0                11          173  1.689000e+02         Denmark
12370.0               167         2353  3.545690e+03          Cyprus
12371.0                63          591  1.887960e+03     Switzerland
12372.0                52          794  1.298040e+03         Denmark
12373.0                14          197  3.646000e+02         Austria
12374.0                33          342  7.429300e+02         Austria
12375.0                17          178  4.575000e+02         Finland
12377.0                77          944  1.628120e+03     Switzerland
12378.0               219         2529  4.008620e+03     Switzerland
12379.0                40          407  8.522400e+02         Belgium
12380.0               104         1127  2.724810e+03         Belgium
...                   ...          ...           ...             ...
18245.0               175         1788  2.567060e+03  United Kingdom
18246.0                 4          166  5.961000e+02  United Kingdom
18248.0                47          511  8.040200e+02  United Kingdom
18249.0                 8          128  9.534000e+01  United Kingdom
18250.0                21          199  3.492700e+02  United Kingdom
18251.0                16         7824  4.314720e+03  United Kingdom
18252.0                98          239  5.266700e+02  United Kingdom
18255.0                 6           74  1.033000e+02  United Kingdom
18257.0               118         1041  2.337630e+03  United Kingdom
18259.0                42          714  2.338600e+03  United Kingdom
18260.0               134         1478  2.643200e+03  United Kingdom
18261.0                21          146  3.242400e+02  United Kingdom
18262.0                13          182  1.494800e+02  United Kingdom
18263.0                61         1468  1.213160e+03  United Kingdom
18265.0                46          311  8.015100e+02  United Kingdom
18268.0                 1            2  2.550000e+01  United Kingdom
18269.0                 7           76  1.686000e+02  United Kingdom
18270.0                11          101  2.831500e+02  United Kingdom
18272.0               166         2050  3.078580e+03  United Kingdom
18273.0                 3           80  2.040000e+02  United Kingdom
18274.0                11           88  1.759200e+02  United Kingdom
18276.0                14          186  3.358600e+02  United Kingdom
18277.0                 8           68  1.103800e+02  United Kingdom
18278.0                 9           66  1.739000e+02  United Kingdom
18280.0                10           45  1.806000e+02  United Kingdom
18281.0                 7           54  8.082000e+01  United Kingdom
18282.0                12          103  1.780500e+02  United Kingdom
```

```
18283.0                    756          1397  2.094880e+03  United Kingdom
18287.0                     70          1586  1.837280e+03  United Kingdom
nan                     134697        272328  1.733153e+06  United Kingdom

            RatioToMeanTransactionCount  DifFromMeanTransactionCount  \
CustomerID
12346.0                        0.008148                  -121.723733
12347.0                        1.483006                    59.276267
12348.0                        0.252600                   -91.723733
12349.0                        0.594832                   -49.723733
12350.0                        0.138523                  -105.723733
12352.0                        0.692613                   -37.723733
12353.0                        0.032594                  -118.723733
12354.0                        0.472606                   -64.723733
12355.0                        0.105929                  -109.723733
12356.0                        0.480755                   -63.723733
12357.0                        1.067438                     8.276267
12358.0                        0.154819                  -103.723733
12359.0                        2.020799                   125.276267
12360.0                        1.051141                     6.276267
12361.0                        0.081484                  -112.723733
12362.0                        2.167470                   143.276267
12363.0                        0.187413                   -99.723733
12364.0                        0.692613                   -37.723733
12365.0                        0.179264                  -100.723733
12367.0                        0.089632                  -111.723733
12370.0                        1.360780                    44.276267
12371.0                        0.513348                   -59.723733
12372.0                        0.423716                   -70.723733
12373.0                        0.114077                  -108.723733
12374.0                        0.268897                   -89.723733
12375.0                        0.138523                  -105.723733
12377.0                        0.627426                   -45.723733
12378.0                        1.784496                    96.276267
12379.0                        0.325935                   -82.723733
12380.0                        0.847432                   -18.723733
...                                 ...                          ...
18245.0                        1.425967                    52.276267
18246.0                        0.032594                  -118.723733
18248.0                        0.382974                   -75.723733
18249.0                        0.065187                  -114.723733
18250.0                        0.171116                  -101.723733
18251.0                        0.130374                  -106.723733
18252.0                        0.798542                   -24.723733
18255.0                        0.048890                  -116.723733
18257.0                        0.961509                    -4.723733
18259.0                        0.342232                   -80.723733
18260.0                        1.091883                    11.276267
```

```
18261.0                        0.171116                   -101.723733
18262.0                        0.105929                   -109.723733
18263.0                        0.497051                    -61.723733
18265.0                        0.374826                    -76.723733
18268.0                        0.008148                   -121.723733
18269.0                        0.057039                   -115.723733
18270.0                        0.089632                   -111.723733
18272.0                        1.352632                     43.276267
18273.0                        0.024445                   -119.723733
18274.0                        0.089632                   -111.723733
18276.0                        0.114077                   -108.723733
18277.0                        0.065187                   -114.723733
18278.0                        0.073335                   -113.723733
18280.0                        0.081484                   -112.723733
18281.0                        0.057039                   -115.723733
18282.0                        0.097781                   -110.723733
18283.0                        6.160178                    633.276267
18287.0                        0.570387                    -52.723733
nan                         1097.562770                 134574.276267

           HigherThanMeanTransactionCount  TransactionCountPercentile  \
CustomerID
12346.0                             False                    0.841014
12347.0                              True                   88.087558
12348.0                             False                   42.108295
12349.0                             False                   66.555300
12350.0                             False                   24.919355
12352.0                             False                   70.529954
12353.0                             False                    4.711982
12354.0                             False                   59.965438
12355.0                             False                   19.193548
12356.0                             False                   60.460829
12357.0                              True                   81.831797
12358.0                             False                   27.511521
12359.0                              True                   92.373272
12360.0                              True                   81.486175
12361.0                             False                   14.343318
12362.0                              True                   93.122120
12363.0                             False                   32.718894
12364.0                             False                   70.529954
12365.0                             False                   31.578341
12367.0                             False                   16.094470
12370.0                              True                   86.739631
12371.0                             False                   62.534562
12372.0                             False                   56.808756
12373.0                             False                   20.645161
12374.0                             False                   43.778802
12375.0                             False                   24.919355
```

```
12377.0                             False                   67.937788
12378.0                              True                   90.990783
12379.0                             False                   49.170507
12380.0                             False                   76.163594
...                                   ...                         ...
18245.0                              True                   87.453917
18246.0                             False                    4.711982
18248.0                             False                   54.124424
18249.0                             False                   10.956221
18250.0                             False                   30.345622
18251.0                             False                   23.594470
18252.0                             False                   74.366359
18255.0                             False                    7.764977
18257.0                             False                   79.447005
18259.0                             False                   50.656682
18260.0                              True                   82.246544
18261.0                             False                   30.345622
18262.0                             False                   19.193548
18263.0                             False                   61.497696
18265.0                             False                   53.375576
18268.0                             False                    0.841014
18269.0                             False                    9.377880
18270.0                             False                   16.094470
18272.0                              True                   86.566820
18273.0                             False                    3.525346
18274.0                             False                   16.094470
18276.0                             False                   20.645161
18277.0                             False                   10.956221
18278.0                             False                   12.603687
18280.0                             False                   14.343318
18281.0                             False                    9.377880
18282.0                             False                   17.718894
18283.0                              True                   99.308756
18287.0                             False                   65.483871
nan                                  True                  100.000000

            RatioToMeanTransactionCountPerCountry  topic_0  topic_1  topic_2  \
CustomerID
12346.0                                  0.008020      0.0      0.0      0.0
12347.0                                  1.000000     30.0     19.0     14.0
12348.0                                  0.543066     19.0      0.0      1.0
12349.0                                  1.348285     15.0     10.0      1.0
12350.0                                  0.158582      0.0      0.0      0.0
12352.0                                  0.792910      9.0     11.0      2.0
12353.0                                  0.470588      3.0      0.0      0.0
12354.0                                  0.671725      6.0      5.0      3.0
12355.0                                  1.529412      4.0      0.0      0.0
12356.0                                  0.766758     22.0      1.0      5.0
```

```
12357.0                                   1.434830     17.0     15.0      4.0
12358.0                                   0.463415      3.0      0.0      0.0
12359.0                                   2.574394     24.0     36.0      4.0
12360.0                                   3.146341     26.0      9.0      5.0
12361.0                                   0.119641      3.0      2.0      0.0
12362.0                                   3.182453     44.0     29.0     15.0
12363.0                                   0.377049     11.0      1.0      6.0
12364.0                                   1.016949     22.0      3.0     17.0
12365.0                                   0.228374      4.0      2.0      4.0
12367.0                                   0.222785      2.0      2.0      0.0
12370.0                                   1.733564     20.0     22.0      9.0
12371.0                                   0.690033     20.0     18.0      0.0
12372.0                                   1.053165     10.0      4.0      1.0
12373.0                                   0.341463      2.0      5.0      0.0
12374.0                                   0.804878      9.0      6.0      1.0
12375.0                                   0.297810      1.0      3.0      0.0
12377.0                                   0.843373     10.0      7.0      2.0
12378.0                                   2.398686     41.0     20.0     13.0
12379.0                                   0.478564      8.0      7.0      0.0
12380.0                                   1.244267     18.0     21.0      5.0
...                                            ...      ...      ...      ...
18245.0                                   1.403458     27.0     22.0      9.0
18246.0                                   0.032079      0.0      0.0      2.0
18248.0                                   0.376929      8.0      8.0      2.0
18249.0                                   0.064158      0.0      2.0      0.0
18250.0                                   0.168415      5.0      4.0      0.0
18251.0                                   0.128316     12.0      0.0      0.0
18252.0                                   0.785937     20.0     13.0      4.0
18255.0                                   0.048119      0.0      4.0      0.0
18257.0                                   0.946332     18.0     22.0      5.0
18259.0                                   0.336830      2.0     21.0      1.0
18260.0                                   1.074648     30.0     22.0     13.0
18261.0                                   0.168415      5.0      1.0      1.0
18262.0                                   0.104257      6.0      3.0      0.0
18263.0                                   0.489205     11.0     11.0      3.0
18265.0                                   0.368909      9.0      4.0      4.0
18268.0                                   0.008020      1.0      0.0      0.0
18269.0                                   0.056138      0.0      4.0      0.0
18270.0                                   0.088217      1.0      3.0      0.0
18272.0                                   1.331280     50.0     32.0      5.0
18273.0                                   0.024059      0.0      0.0      0.0
18274.0                                   0.088217      5.0      0.0      2.0
18276.0                                   0.112277      2.0      2.0      0.0
18277.0                                   0.064158      0.0      1.0      1.0
18278.0                                   0.072178      0.0      3.0      0.0
18280.0                                   0.080198      1.0      1.0      0.0
18281.0                                   0.056138      2.0      0.0      0.0
18282.0                                   0.096237      1.0      1.0      1.0
```

```
18283.0                                   6.062939    153.0    162.0     23.0
18287.0                                   0.561383     18.0     16.0      0.0
nan                                    1080.237759  21125.0  17875.0   6235.0

            topic_3  topic_4  topic_5  topic_6  topic_7  topic_8  topic_9
CustomerID
12346.0         0.0      0.0      1.0      0.0      0.0      0.0      0.0
12347.0        13.0     26.0     24.0     12.0     17.0     17.0     10.0
12348.0         0.0      0.0      5.0      0.0      4.0      2.0      0.0
12349.0         6.0      5.0     14.0      9.0      4.0      5.0      4.0
12350.0         1.0      0.0      4.0      1.0      3.0      5.0      3.0
12352.0         7.0      3.0     17.0      5.0     12.0      6.0     13.0
12353.0         0.0      0.0      1.0      0.0      0.0      0.0      0.0
12354.0         2.0      4.0     19.0      2.0      4.0      4.0      9.0
12355.0         1.0      0.0      3.0      0.0      3.0      0.0      2.0
12356.0         3.0      4.0     15.0      0.0      5.0      4.0      0.0
12357.0        15.0     17.0     23.0      9.0      4.0     10.0     17.0
12358.0         0.0      0.0      6.0      0.0      2.0      0.0      8.0
12359.0        24.0     27.0     42.0     12.0     28.0     24.0     27.0
12360.0        25.0      3.0     20.0      9.0      7.0      7.0     18.0
12361.0         0.0      1.0      1.0      1.0      2.0      0.0      0.0
12362.0        19.0     24.0     46.0     20.0     36.0     12.0     21.0
12363.0         0.0      1.0      2.0      1.0      0.0      0.0      1.0
12364.0         7.0      2.0      8.0      1.0      8.0      5.0     12.0
12365.0         1.0      1.0      2.0      1.0      4.0      2.0      1.0
12367.0         1.0      1.0      0.0      0.0      3.0      1.0      1.0
12370.0        13.0     33.0     16.0      9.0     15.0     20.0     10.0
12371.0         4.0      4.0      6.0      3.0      7.0      0.0      1.0
12372.0         1.0      9.0      8.0      7.0      7.0      0.0      5.0
12373.0         0.0      0.0      3.0      1.0      2.0      0.0      1.0
12374.0         2.0      1.0      4.0      7.0      1.0      0.0      2.0
12375.0         1.0      1.0      3.0      0.0      3.0      0.0      5.0
12377.0         6.0     14.0      8.0      9.0      7.0      5.0      9.0
12378.0        18.0     20.0     43.0      8.0      8.0     17.0     31.0
12379.0         2.0      3.0      9.0      5.0      3.0      0.0      3.0
12380.0         8.0      4.0     27.0      4.0      8.0      0.0      9.0
...             ...      ...      ...      ...      ...      ...      ...
18245.0        30.0     13.0     28.0     12.0     16.0      4.0     14.0
18246.0         0.0      1.0      0.0      1.0      0.0      0.0      0.0
18248.0         4.0      1.0     12.0      4.0      1.0      2.0      5.0
18249.0         0.0      0.0      4.0      0.0      2.0      0.0      0.0
18250.0         1.0      2.0      8.0      0.0      1.0      0.0      0.0
18251.0         0.0      0.0      4.0      0.0      0.0      0.0      0.0
18252.0         3.0      6.0     12.0      7.0     14.0     10.0      9.0
18255.0         1.0      0.0      0.0      0.0      1.0      0.0      0.0
18257.0         5.0      3.0     27.0      5.0      6.0      9.0     18.0
18259.0         0.0      0.0      0.0      4.0     12.0      0.0      2.0
18260.0         3.0     15.0     17.0     11.0     10.0      3.0     10.0
```

```
18261.0         2.0      4.0      0.0      1.0      3.0      3.0      1.0
18262.0         0.0      1.0      2.0      0.0      0.0      1.0      0.0
18263.0         7.0      4.0     11.0      1.0      0.0      7.0      6.0
18265.0         3.0      2.0     10.0      3.0      4.0      1.0      6.0
18268.0         0.0      0.0      0.0      0.0      0.0      0.0      0.0
18269.0         0.0      0.0      1.0      2.0      0.0      0.0      0.0
18270.0         0.0      4.0      0.0      1.0      1.0      0.0      1.0
18272.0         5.0     10.0     36.0     13.0      3.0      2.0     10.0
18273.0         0.0      0.0      3.0      0.0      0.0      0.0      0.0
18274.0         1.0      0.0      3.0      0.0      0.0      0.0      0.0
18276.0         1.0      1.0      3.0      1.0      3.0      1.0      0.0
18277.0         0.0      3.0      1.0      0.0      0.0      0.0      2.0
18278.0         1.0      1.0      2.0      0.0      0.0      0.0      2.0
18280.0         1.0      3.0      2.0      0.0      1.0      0.0      1.0
18281.0         1.0      1.0      0.0      1.0      0.0      0.0      2.0
18282.0         3.0      1.0      3.0      0.0      2.0      0.0      0.0
18283.0        24.0     19.0    177.0     35.0     45.0     34.0     84.0
18287.0        10.0      1.0     11.0      3.0      3.0      3.0      5.0
nan         11963.0  11597.0  20423.0   8731.0  11334.0  11417.0  13997.0

[4340 rows x 19 columns]
```

este by bolo dobre tie pocty normalizovat na podiely z celkoveho poctu zakupenych produktov

6.0.1 V tomto priklade su spojene dve ukazky:

- na vytvorenie segmentov sa da pouzit aj textovy opis a nie len zhlukovanie na zaklade ostatnych atributov
- daju sa pouzit segmenty z naviazanej entity na vytvorenie atributov jednoducho spocitanim vyskytov

6.1 Kodovanie kategorickych dat

6.1.1 Existuje viacero sposobov ako transformovat kategoricku hodnotu na cislo

http://www.willmcginnis.com/2015/11/29/beyond-one-hot-an-exploration-of-categorical-variables/

1. Ordinal - priradzovanie cisel postupne roznym hodnotam
2. One-hot - z kazdej kategorie vznikne stlpec s hodnotou 1 v tych riadkoch, ktore boli nastavene na tuto hodnotu. inde 0
3. Binary - zoberie sa ordinal, zakoduju sa tie cisla ako binarne, kazda cislica binarneho cisla je pouzita ako stlpec a tam kde bola na zodpovedajucom mieste 1, tam bude v stlpci 1 a inak 0
4. Sum - porovnava sa priemer zavyslej premennej na riadokch jednej skupiny oproti priemeru zavyslej premennej na celej datovej sade
5. Helmert - velmi podobne ako Sum, len jedinecnost categorickej hodnoty je dana inou kombinaciou hodnot

6. BackwardDifferenceEncoder - velmi podobne ako Sum, len jedinecnost categorickej hodnoty je dana inou kombinaciou hodnot
7. Polynomial - trenuje koeficienty ploynomialnej regresie rozneho stupna, ktore sa daju pouzit na regresiu zavyslej premennej (neviem aky to ma zmysel pre maly pocet roznych hodnot) (treba ordinalne premenne)
8. Hash - zahashuje string kategorickej premennej a moduluje ho poctom roznych hodnot. Je v sklearn http://scikit-learn.org/stable/modules/generated/sklearn.feature_extraction.FeatureHasher.html

```
In [43]: import category_encoders as ce
```

```
In [44]: encoder = ce.OrdinalEncoder()
         # encoder = ce.OneHotEncoder() # tento je aj v sklearn.preprocessing.OneHotEncoder, tu
         # encoder = ce.BinaryEncoder()
         # encoder = ce.PolynomialEncoder()
         # encoder = ce.HashingEncoder()
         # encoder = ce.HelmertEncoder()
         # encoder = ce.BackwardDifferenceEncoder()
         # encoder = ce.SumEncoder()
```

```
In [45]: X = customer_data[['Country']]
         X_cleaned = encoder.fit_transform(X)
         X_cleaned.head(5)
```

```
Out[45]:             Country
         CustomerID
         12346.0           0
         12347.0           1
         12348.0           2
         12349.0           3
         12350.0           4
```

```
In [46]: # Niekore sposoby kodovania su zavisle na zavislej premennej

         encoder = ce.SumEncoder()
         encoder.fit(X, customer_data.TransactionCount)
         X_cleaned = encoder.transform(X)
         X_cleaned.head(10)
```

```
Out[46]:             col_Country_0  col_Country_1  col_Country_2  col_Country_3  \
         CustomerID
         12346.0               1.0            1.0            0.0            0.0
         12347.0               1.0            0.0            1.0            0.0
         12348.0               1.0            0.0            0.0            1.0
         12349.0               1.0            0.0            0.0            0.0
         12350.0               1.0            0.0            0.0            0.0
         12352.0               1.0            0.0            0.0            0.0
         12353.0               1.0            0.0            0.0            0.0
         12354.0               1.0            0.0            0.0            0.0
```

```
12355.0               1.0            0.0            0.0            0.0
12356.0               1.0            0.0            0.0            0.0

            col_Country_4  col_Country_5  col_Country_6  col_Country_7  \
CustomerID
12346.0               0.0            0.0            0.0            0.0
12347.0               0.0            0.0            0.0            0.0
12348.0               0.0            0.0            0.0            0.0
12349.0               1.0            0.0            0.0            0.0
12350.0               0.0            1.0            0.0            0.0
12352.0               0.0            1.0            0.0            0.0
12353.0               0.0            0.0            1.0            0.0
12354.0               0.0            0.0            0.0            1.0
12355.0               0.0            0.0            1.0            0.0
12356.0               0.0            0.0            0.0            0.0

            col_Country_8  col_Country_9       ...        col_Country_27  \
CustomerID                                     ...
12346.0               0.0            0.0       ...                   0.0
12347.0               0.0            0.0       ...                   0.0
12348.0               0.0            0.0       ...                   0.0
12349.0               0.0            0.0       ...                   0.0
12350.0               0.0            0.0       ...                   0.0
12352.0               0.0            0.0       ...                   0.0
12353.0               0.0            0.0       ...                   0.0
12354.0               0.0            0.0       ...                   0.0
12355.0               0.0            0.0       ...                   0.0
12356.0               1.0            0.0       ...                   0.0

            col_Country_28  col_Country_29  col_Country_30  col_Country_31  \
CustomerID
12346.0                0.0             0.0             0.0             0.0
12347.0                0.0             0.0             0.0             0.0
12348.0                0.0             0.0             0.0             0.0
12349.0                0.0             0.0             0.0             0.0
12350.0                0.0             0.0             0.0             0.0
12352.0                0.0             0.0             0.0             0.0
12353.0                0.0             0.0             0.0             0.0
12354.0                0.0             0.0             0.0             0.0
12355.0                0.0             0.0             0.0             0.0
12356.0                0.0             0.0             0.0             0.0

            col_Country_32  col_Country_33  col_Country_34  col_Country_35  \
CustomerID
12346.0                0.0             0.0             0.0             0.0
12347.0                0.0             0.0             0.0             0.0
12348.0                0.0             0.0             0.0             0.0
12349.0                0.0             0.0             0.0             0.0
```

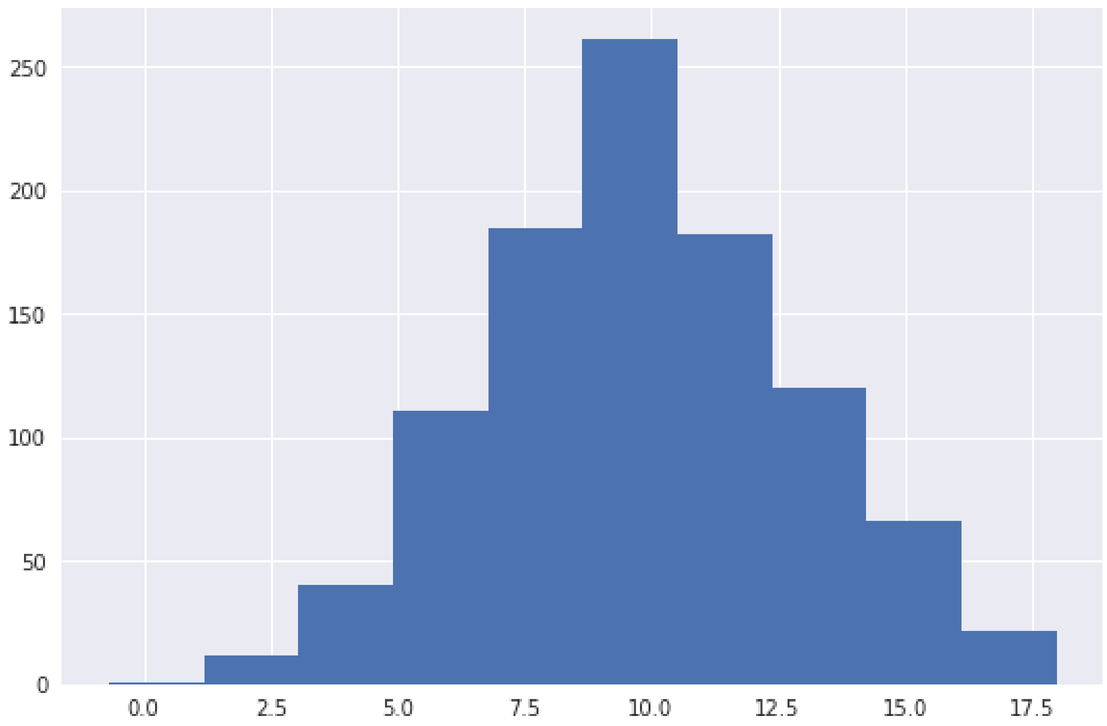
```
12350.0                0.0             0.0             0.0             0.0
12352.0                0.0             0.0             0.0             0.0
12353.0                0.0             0.0             0.0             0.0
12354.0                0.0             0.0             0.0             0.0
12355.0                0.0             0.0             0.0             0.0
12356.0                0.0             0.0             0.0             0.0

            col_Country_36
CustomerID
12346.0                0.0
12347.0                0.0
12348.0                0.0
12349.0                0.0
12350.0                0.0
12352.0                0.0
12353.0                0.0
12354.0                0.0
12355.0                0.0
12356.0                0.0

[10 rows x 37 columns]
```

6.2 Transformacia numerickych atributov na kategoricke - diskretizacia

Pre niektore algoritmy ciselne data niesu uplne vhodne. Napriklad taky Naivny Bayes potrebuje kategoricke hodnoty. Tie sa daju vytvorit zaradenim numerickych hodnot do intervalov.

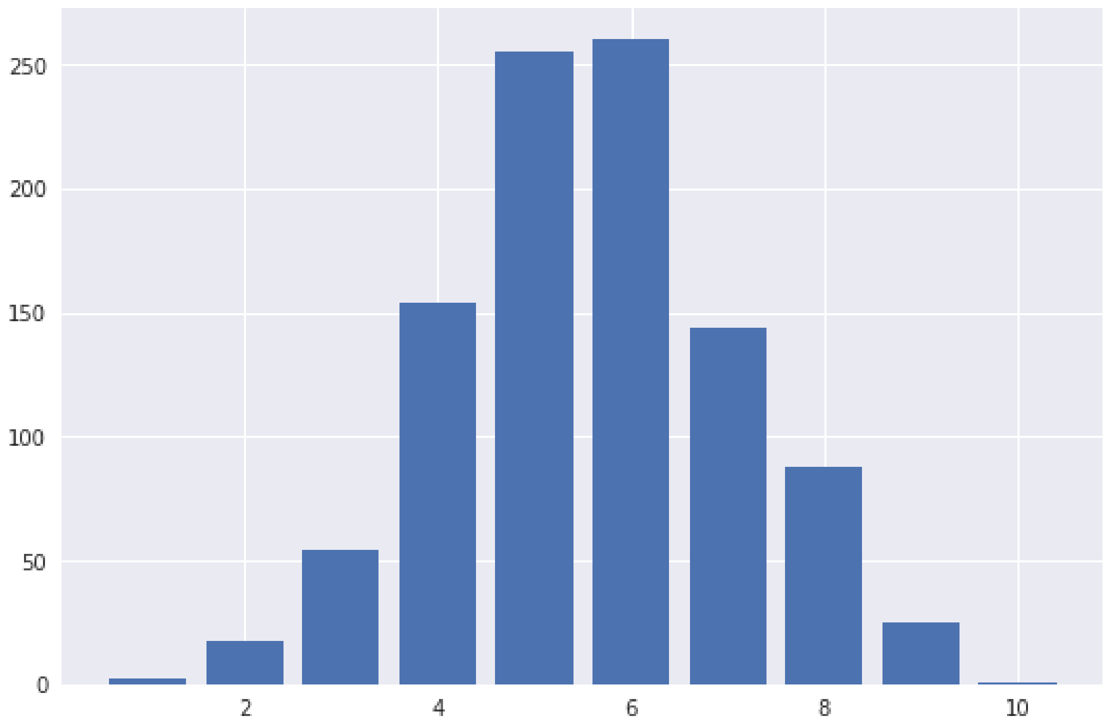
Intervaly su bud manulane nastavene, alebo take, aby ich sirka bola rovnaka (Equal width binning/discretization) alebo pocty prvkov (Equal frequency binning/discretization), ktore do nich spadaju boli priblyzne rovnake.

```
In [47]: sample = stats.norm(10,3).rvs(1000)
         plt.hist(sample)

Out[47]: (array([   1.,   12.,   40.,  111.,  185.,  261.,  182.,  120.,   66.,   22.]),
          array([ -0.68007087,   1.18588389,   3.05183865,   4.91779341,
                   6.78374817,   8.64970293,  10.51565769,  12.38161245,
                  14.24756721,  16.11352197,  17.97947673]),
          <a list of 10 Patch objects>)

/usr/local/lib/python3.5/dist-packages/matplotlib/font_manager.py:1297: UserWarning: findfont: F
  (prop.get_family(), self.defaultFamily[fontext]))
```



6.2.1 Rovnaka sirka intervalov

```
In [48]: bin_count = 10
         bins = np.linspace(sample.min(), sample.max(), bin_count)
         ew_digitized = np.digitize(sample, bins)
         ew_digitized[:10]

Out[48]: array([6, 7, 9, 4, 5, 3, 4, 6, 4, 6])

In [49]: plt.bar(*np.unique(ew_digitized, return_counts=True))

Out[49]: <Container object of 10 artists>

/usr/local/lib/python3.5/dist-packages/matplotlib/font_manager.py:1297: UserWarning: findfont: F
  (prop.get_family(), self.defaultFamily[fontext]))
```



```
In [50]: bins

Out[50]: array([ -0.68007087,   1.3932122 ,   3.46649526,   5.53977833,
                  7.6130614 ,   9.68634446,  11.75962753,  13.83291059,
                 15.90619366,  17.97947673])
```

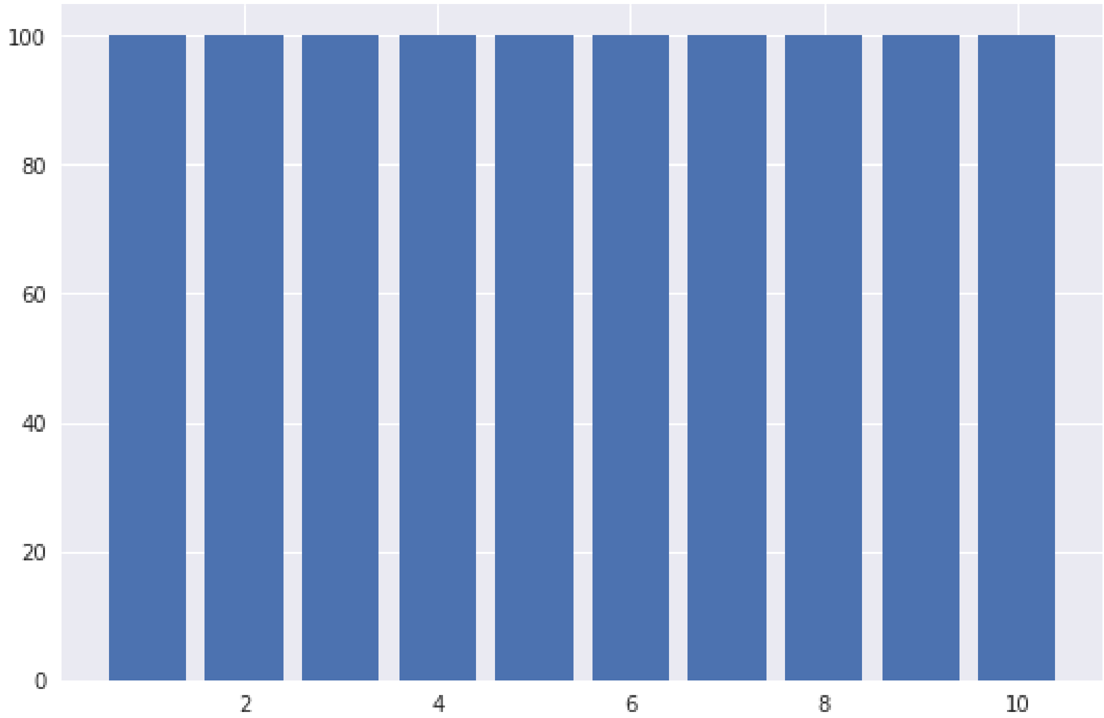
6.2.2 Rovnaka pocetnost intervalov

```
In [51]: bin_count = 10
         percentiles = np.linspace(0, 100, bin_count, endpoint = False)
         bins = list(map(lambda x: np.percentile(sample, x), percentiles))
         ef_digitized = np.digitize(sample, bins)
         ef_digitized[:10]

Out[51]: array([ 8,  9, 10,  2,  4,  1,  2,  7,  2,  7])

In [52]: percentiles

Out[52]: array([  0.,  10.,  20.,  30.,  40.,  50.,  60.,  70.,  80.,  90.])

In [53]: bins

Out[53]: [-0.68007086965670283,
          5.9862092770812128,
          7.2532891806649262,
         8.345836574524963,
         9.0608287153825486,
         9.7801664556088479,
         10.466713219144406,
         11.305542255603619,
         12.451353275387463,
         14.09125231359128]
```

```
In [54]: plt.bar(*np.unique(ef_digitized, return_counts=True))
```

```
Out[54]: <Container object of 10 artists>
```

```
/usr/local/lib/python3.5/dist-packages/matplotlib/font_manager.py:1297: UserWarning: findfont: F
  (prop.get_family(), self.defaultFamily[fontext]))
```



Predchadzajuci obrazok zobrazuje transformovane data pomocou transformacie natrenovanej na tych istych datach

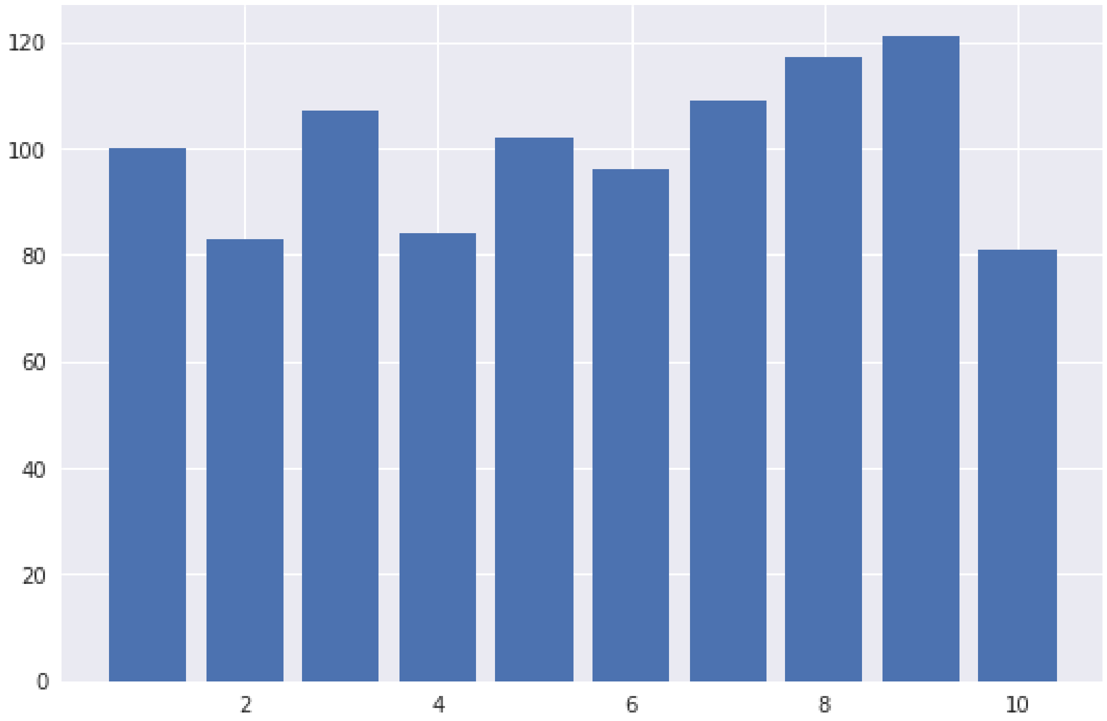
Co ked zoberiem data (z rovnakeho rozdelenia) a skusim ich transformovat pomocou vopred pripravenej transformacie?

```
In [55]: sample2 = stats.norm(10,3).rvs(1000)
```

```
In [56]: # nezmenim biny, cize pouzivam transformaciu, ktoru som natrenoval na trenovacich datac
         ef_digitized_test = np.digitize(sample2, bins)
         plt.bar(*np.unique(ef_digitized_test, return_counts=True))
```

```
Out[56]: <Container object of 10 artists>
```

```
/usr/local/lib/python3.5/dist-packages/matplotlib/font_manager.py:1297: UserWarning: findfont: F
  (prop.get_family(), self.defaultFamily[fontext]))
```



Tie intervaly zrazu niesu take pekne vyrovnane. To sa ale da cakat.

Vsimnite si intervaly na koncoch. Tam je rozdiel velmi velky. Je to kvoli tomu, ze tam bol long tail.

7 Zmena v case

Zmena spravania voci nejakemu inemu casovemu oknu.

- running window (posuvne okno - stale rovnako velke)
- landmark window (od nejakeho momentu dalej)
- dumping window (pomale zabudanie)

```
In [57]: data.head()
```

```
Out[57]:   InvoiceNo StockCode                          Description  Quantity  \
         0    536365    85123A   WHITE HANGING HEART T-LIGHT HOLDER         6
         1    536365     71053                  WHITE METAL LANTERN         6
         2    536365    84406B       CREAM CUPID HEARTS COAT HANGER         8
         3    536365    84029G  KNITTED UNION FLAG HOT WATER BOTTLE         6
         4    536365    84029E       RED WOOLLY HOTTIE WHITE HEART.         6

                   InvoiceDate  UnitPrice CustomerID         Country  TotalPrice
         0 2010-12-01 08:26:00       2.55    17850.0  United Kingdom       15.30
         1 2010-12-01 08:26:00       3.39    17850.0  United Kingdom       20.34
         2 2010-12-01 08:26:00       2.75    17850.0  United Kingdom       22.00
         3 2010-12-01 08:26:00       3.39    17850.0  United Kingdom       20.34
         4 2010-12-01 08:26:00       3.39    17850.0  United Kingdom       20.34
```

```
In [58]: %%time

         # rozdelim si data na posledny den, kedy zakaznik robil obrat a vsetky jeho ostatne dni
         # chcem spravit atributy, ktore by odrazali zmenu medzi poslednym dnom a dnami predtym
         before = []
         after = []

         sorted_data = data.sort_values(['CustomerID', 'InvoiceDate'], ascending=False)

         cid, idate = sorted_data.iloc[0][['CustomerID', 'InvoiceDate']]
         for index, row in sorted_data.iterrows():
             if (cid == row.CustomerID) and (idate == row.InvoiceDate):
                 after.append(row)
             elif cid != row.CustomerID:
                 cid = row.CustomerID
                 idate = row.InvoiceDate
                 after.append(row)
             else:
                 before.append(row)
         print(len(before), len(after))

         before_df = pd.DataFrame(before)
         after_df = pd.DataFrame(after)
```

```
440031 92590
CPU times: user 2min 24s, sys: 1.28 s, total: 2min 25s
Wall time: 2min 25s
```

```
In [59]: before_df.head()
```

```
Out[59]:        InvoiceNo StockCode                    Description  Quantity  \
         541264    581497     20719         WOODLAND CHARLOTTE BAG        33
         541265    581497     20723       STRAWBERRY CHARLOTTE BAG        42
         541266    581497     20724    RED RETROSPOT CHARLOTTE BAG        55
         541267    581497     20727        LUNCH BAG  BLACK SKULL.         8
         541268    581497     21212  PACK OF 72 RETROSPOT CAKE CASES         7

                        InvoiceDate  UnitPrice CustomerID         Country  TotalPrice
         541264 2011-12-09 10:23:00       2.46        nan  United Kingdom       81.18
         541265 2011-12-09 10:23:00       2.46        nan  United Kingdom      103.32
         541266 2011-12-09 10:23:00       2.46        nan  United Kingdom      135.30
         541267 2011-12-09 10:23:00       4.96        nan  United Kingdom       39.68
         541268 2011-12-09 10:23:00       2.08        nan  United Kingdom       14.56

In [60]: # rozdiel oproti predchadzajucemu casovemu oknu

In [61]: customer_data['BeforeItemCount'] = before_df.groupby('CustomerID').Quantity.sum()
         customer_data['AfterItemCount'] = after_df.groupby('CustomerID').Quantity.sum()

In [62]: customer_data.head()

Out[62]:             TransactionCount  TotalItemCount  Valuation         Country  \
         CustomerID
         12346.0                    1           74215   77183.60  United Kingdom
         12347.0                  182            2458    4310.00         Iceland
         12348.0                   31            2341    1797.24         Finland
         12349.0                   73             631    1757.55           Italy
         12350.0                   17             197     334.40          Norway

                     RatioToMeanTransactionCount  DifFromMeanTransactionCount  \
         CustomerID
         12346.0                        0.008148                  -121.723733
         12347.0                        1.483006                    59.276267
         12348.0                        0.252600                   -91.723733
         12349.0                        0.594832                   -49.723733
         12350.0                        0.138523                  -105.723733

                    HigherThanMeanTransactionCount  TransactionCountPercentile  \
         CustomerID
         12346.0                             False                    0.841014
         12347.0                              True                   88.087558
         12348.0                             False                   42.108295
         12349.0                             False                   66.555300
         12350.0                             False                   24.919355

                     RatioToMeanTransactionCountPerCountry  BeforeItemCount  \
         CustomerID
         12346.0                                  0.008020              NaN
         12347.0                                  1.000000           2266.0
         12348.0                                  0.543066           2124.0
         12349.0                                  1.348285              NaN
         12350.0                                  0.158582              NaN

                     AfterItemCount
         CustomerID
         12346.0              74215
         12347.0                192
         12348.0                217
         12349.0                631
         12350.0                197
```

```
In [63]: customer_data['DifBeforeAfterItemCount'] = customer_data['BeforeItemCount'] - customer_
         customer_data['RatioBeforeAfterItemCount'] = customer_data['BeforeItemCount'] / custome
         customer_data.head()
```

```
Out[63]:             TransactionCount  TotalItemCount  Valuation         Country  \
         CustomerID
         12346.0                    1           74215   77183.60  United Kingdom
         12347.0                  182            2458    4310.00         Iceland
         12348.0                   31            2341    1797.24         Finland
         12349.0                   73             631    1757.55           Italy
         12350.0                   17             197     334.40          Norway

                     RatioToMeanTransactionCount  DifFromMeanTransactionCount  \
         CustomerID
         12346.0                        0.008148                  -121.723733
         12347.0                        1.483006                    59.276267
         12348.0                        0.252600                   -91.723733
         12349.0                        0.594832                   -49.723733
         12350.0                        0.138523                  -105.723733

                    HigherThanMeanTransactionCount  TransactionCountPercentile  \
         CustomerID
         12346.0                             False                    0.841014
         12347.0                              True                   88.087558
         12348.0                             False                   42.108295
         12349.0                             False                   66.555300
         12350.0                             False                   24.919355

                     RatioToMeanTransactionCountPerCountry  BeforeItemCount  \
         CustomerID
         12346.0                                  0.008020              NaN
         12347.0                                  1.000000           2266.0
         12348.0                                  0.543066           2124.0
         12349.0                                  1.348285              NaN
         12350.0                                  0.158582              NaN

                     AfterItemCount  DifBeforeAfterItemCount  RatioBeforeAfterItemCount
         CustomerID
         12346.0              74215                      NaN                        NaN
         12347.0                192                   2074.0                  11.802083
         12348.0                217                   1907.0                   9.788018
         12349.0                631                      NaN                        NaN
         12350.0                197                      NaN                        NaN
```

A teraz mozem nad tymi oknami pocitat znovavsetky mozne metriky. Napriklad: * pocas akeho dlheho casoveho useku nazbieral data v tom before okne * aky je priemerny pocet poloziek nakupu v tychto dvoch oknach a aky je ich vztah * ake su dlhe prestavky medzi nakupmi * ...

8 Dnes sme si strucne presli

- vypocet agregovanych hodnot
- sedenia
- pomerove atributy
- pomer voci agregovanej hodnote segmentu
- segmentacia pomocou zhlukovania
- pozuitie textovych opisov na vytvorenie segmentov
- pouzitie naviazanej premennej na vytvorenie atributov spocitanim vyskytov
- kodovanie kategorickych premennych
- diskretizacia dat
- modelovanie zmien v case

9 Co tu este chyba?

- normalizacia atributov
- feature construction / feature explostion
- vyber atributov
- redukcia dimenzionality

```
In [ ]:
```