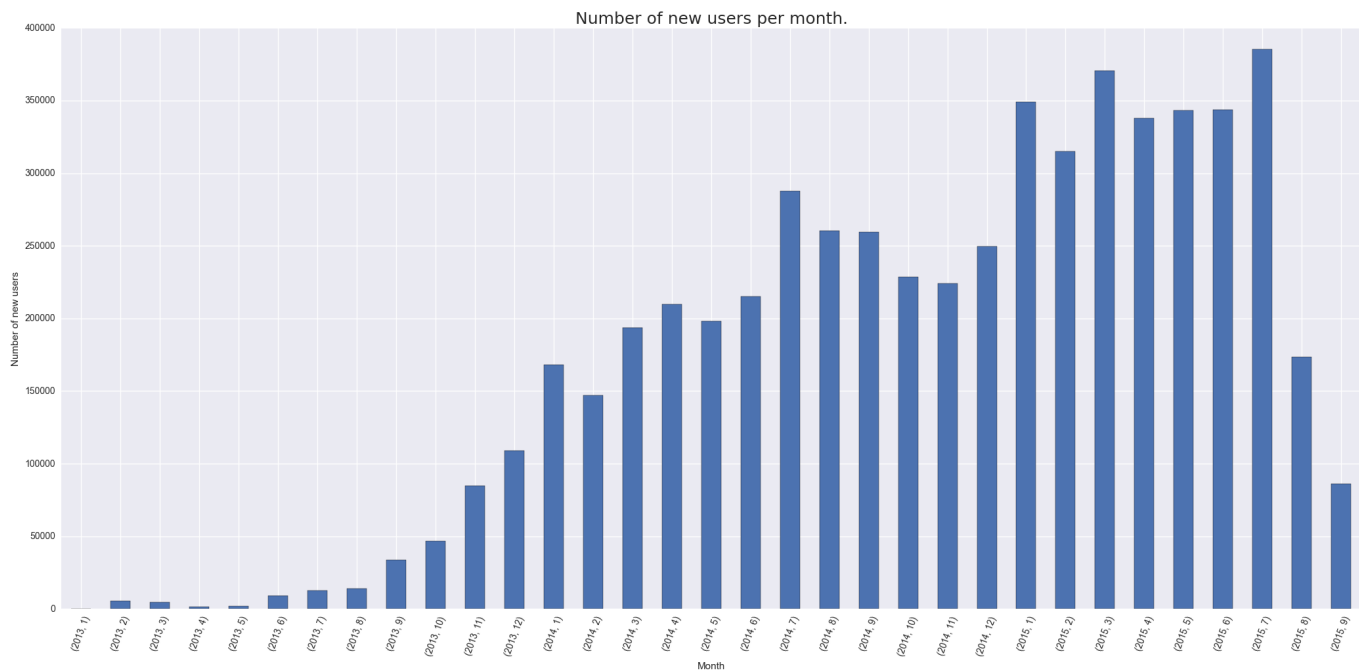
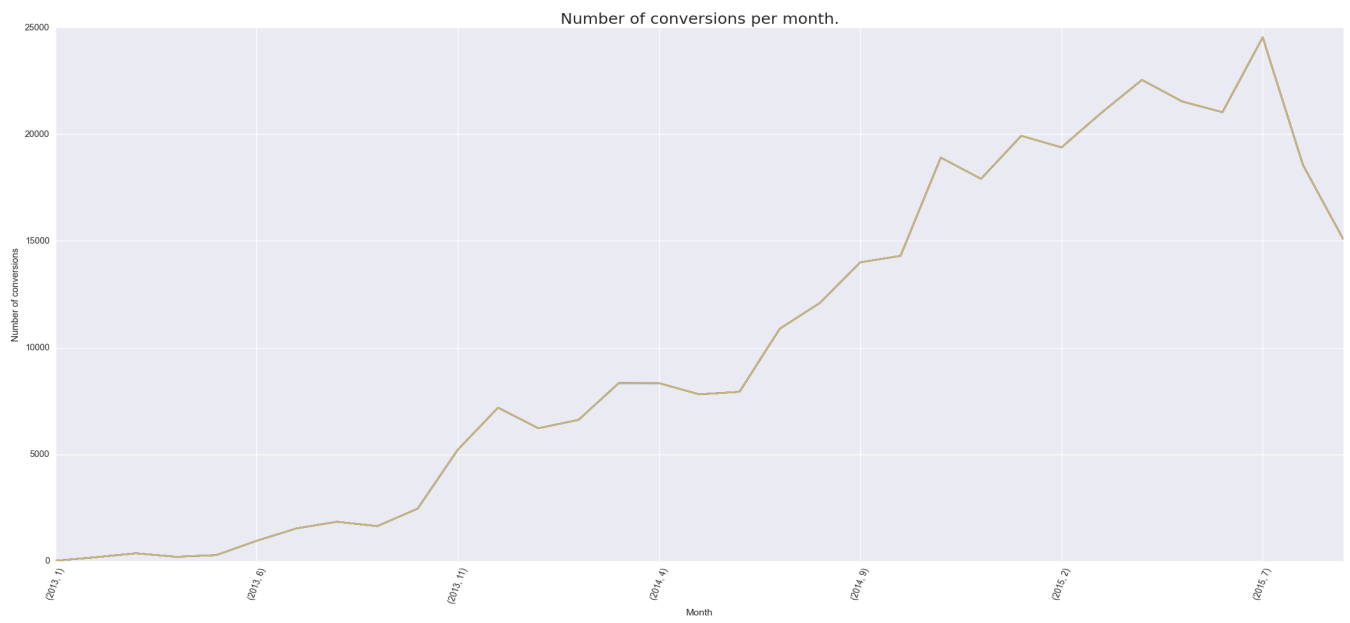


number_of_new_users_per_month()	
Data	
Tables	Columns
Users	'signupTime', 'userId'
Properties	
dropped 'Nan' and 'None' values from 'signupTime' in Users	
Actions	
performing count() operation on 'userId', grouped by year and month in 'signupTime'	
sorting 'signupTime' by year and month	
Axes	
x: year and month	
y: number of registrations	



As can be seen from plot, the number of new users was constantly increasing since beginning of the site till July 2015, then, as plot shows, was a rapid crash in the number of newly registered people.

number_of_conversions_per_month()	
Data	
Tables	Columns
Conversions	'timestamp'
Properties	
dropped 'Nan' and 'None' values from 'timestamp' in Conversions	
Actions	
performing count() operation on rows grouped by year and month	
sorting 'timestamp' by year and month	
Axes	
x: year and month	
y: number of conversions	



As plot indicates the number of conversions per month reached a peak in July 2015, then began to decrease.

## number\_of\_items\_purchased\_per\_month()

### Data

#### Tables

Conversions

#### Columns

'timestamp', 'quantity'

### Properties

dropped 'Nan' and 'None' values from 'timestamp' and 'quantity' in Conversions

### Actions

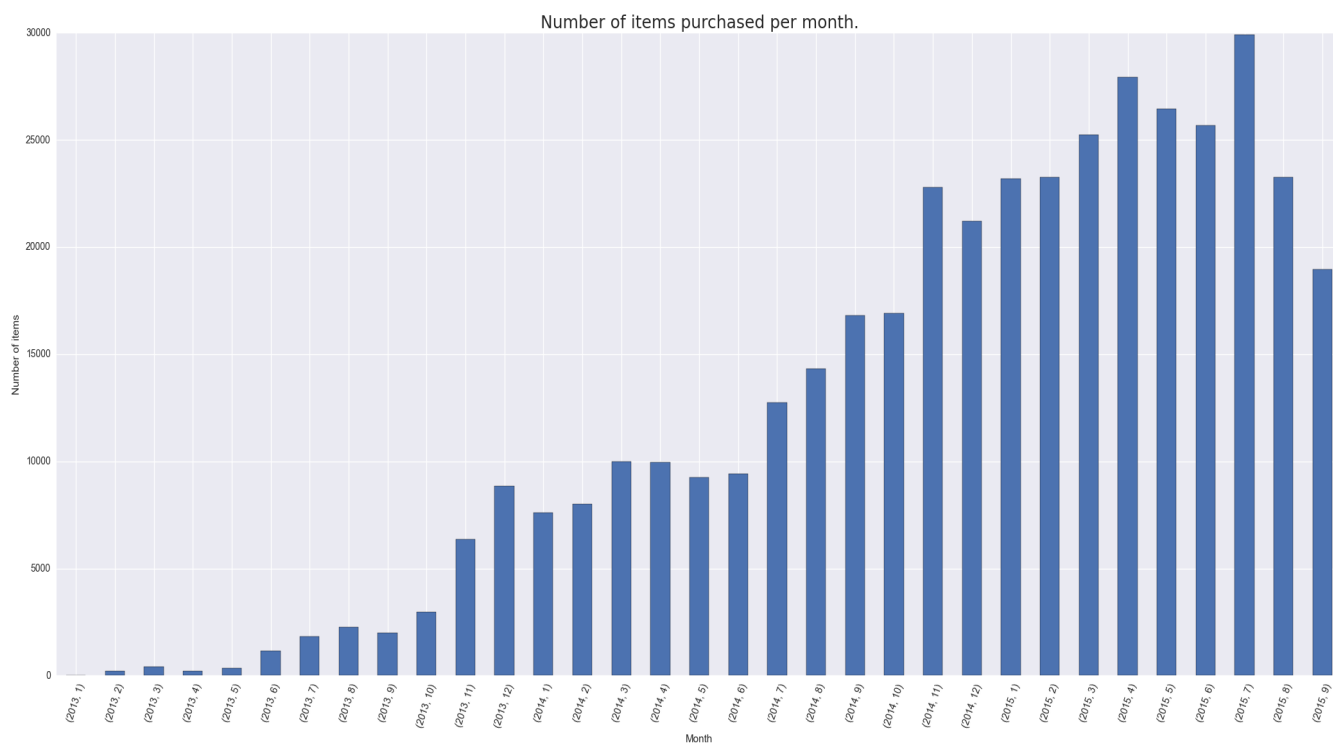
performing sum() operation on 'quantity', rows grouped by year and month of conversion

sorting 'timestamp' by year and month

### Axes

x: year and month

y: number of purchased items



As can be seen, again, the peak is reached in July 2015, then decrease appeared.

income_per_month()	
Data	
Tables	Columns
Conversions	'timestamp', 'price'
Properties	
dropped 'Nan' and 'None' values from 'timestamp' and 'price' in Conversions	
Actions	
performing sum() operation on 'price', rows grouped by year and month of conversion	
sorting 'timestamp' by year and month	
Axes	
x: year and month	
y: income	



As plots shows among 2013, 2014 and 2015 income was constantly growing. Interesting points might be seen in November 2014 and July 2015.

**number\_of\_items\_purchases\_per\_user\_in\_the\_first\_week\_after\_sign\_in()**

**Data**

Tables	Columns
Conversions	'timestamp', 'userId', 'quantity'
Users	'signupTime', 'userId'

**Properties**

dropped 'Nan' and 'None' values from 'timestamp', 'userId', 'quantity' in Conversions

dropped 'Nan' and 'None' values from 'signupTime', 'userId' in Users

**Actions**

joining Conversions and Users on 'userId'

adding additional column to joined structure: 'week\_after' - date week after signing in

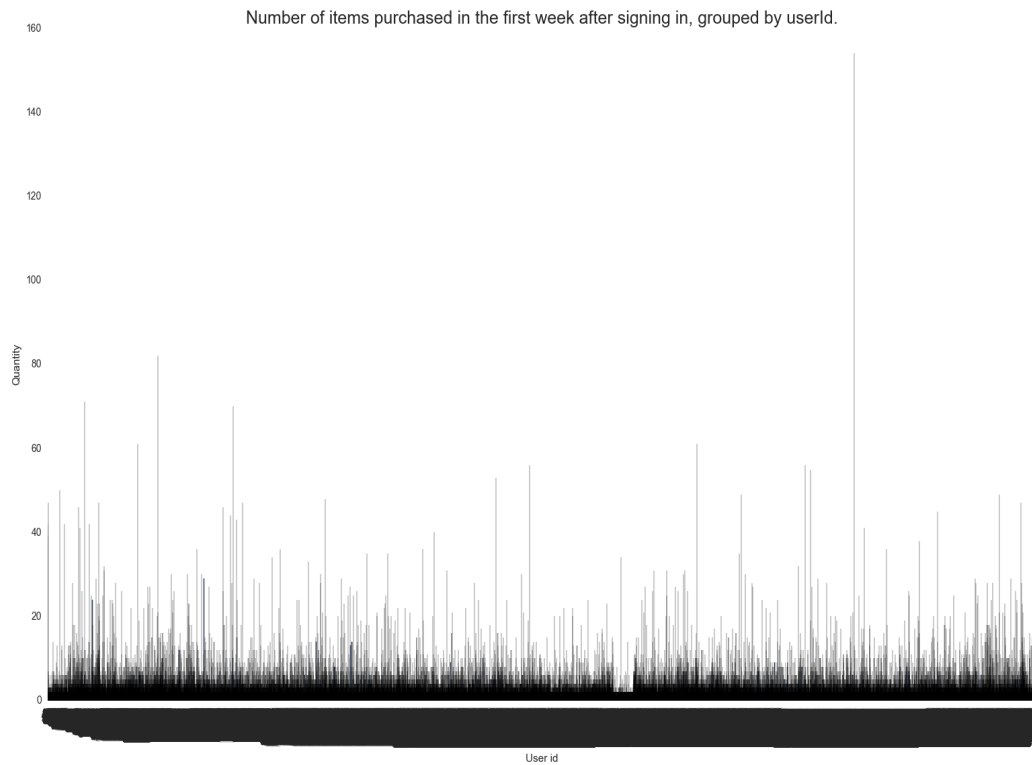
filtering 'timestamp' - rows only with 'timestamp' <= 'week\_after' preserved

performing sum() operation on 'quantity', rows grouped by 'userId'

**Axes**

x: user id

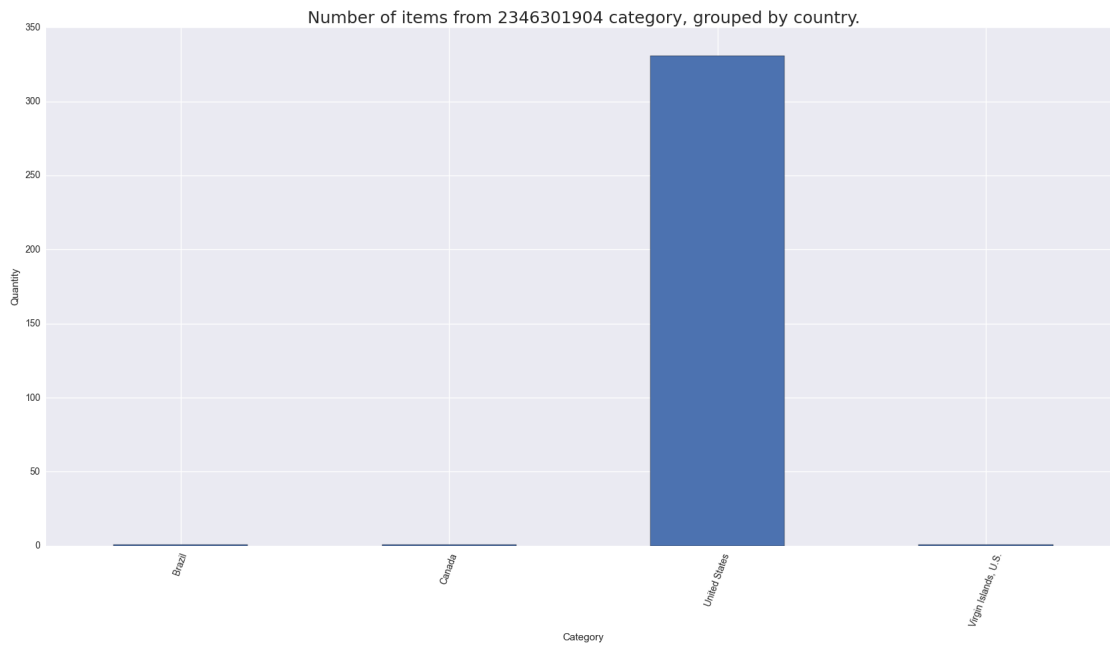
y: quantity of purchased products



The plot indicates that beside single examples, majority of users purchased no more than 20 items during firs week after signing in.

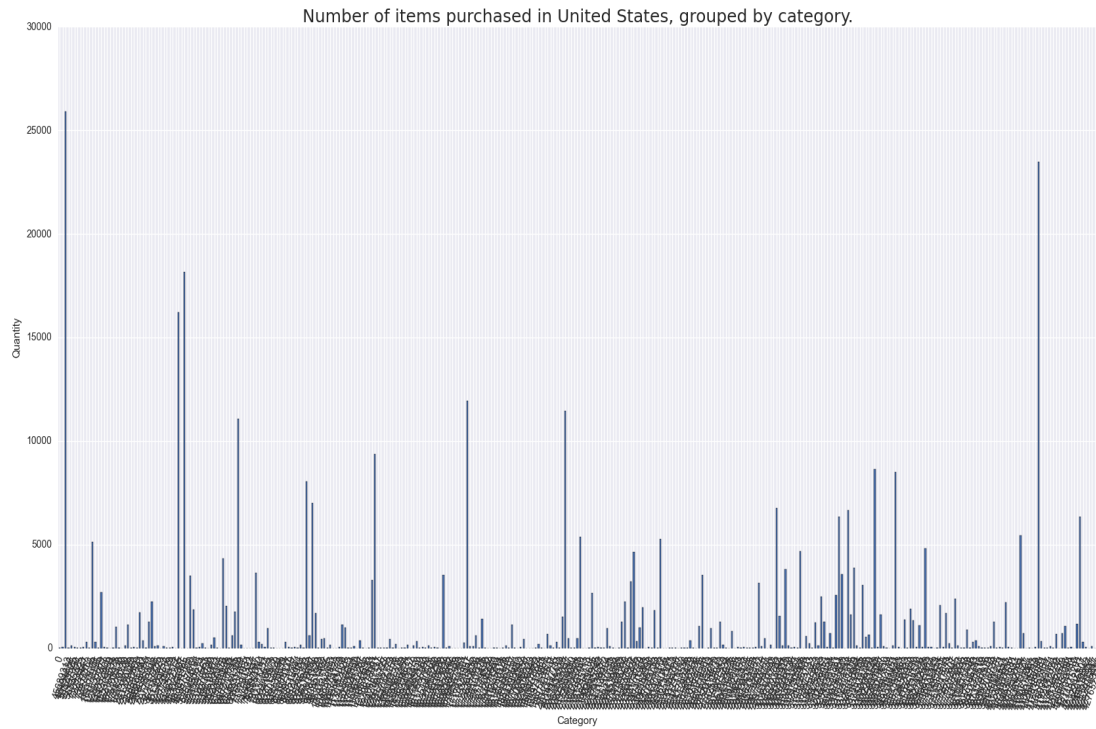


number_of_items_purchased_from_particular_category_grouped_by_count ry(category)	
Data	
Tables	Columns
Conversions	'itemId', 'userId', 'quantity'
Items	'itemId', 'category'
Users	'userId', 'registerCountry'
Properties	
dropped 'Nan' and 'None' values from 'category' in Items	
dropped 'Nan' and 'None' values from 'quantity' in Conversions	
dropped 'Nan' and 'None' values from 'registerCountry' in Users	
Actions	
joining Items and Conversions on 'itemId' and futher with Users on 'userId'	
filtering joined data on 'category' property	
performing sum on 'quantity' in rows grouped by 'registerCountry'	
Axes	
x: country	
y: quantity	



Above plot is generated with filter category == 2346301904. However one plot is not representative enough, what might be seen among figures is that number of purchases in United States is the biggest.

<b>number_of_items_purchased_in_particular_country_grouped_by_category(country)</b>	
<b>Data</b>	
<b>Tables</b>	<b>Columns</b>
Conversions	'itemId', 'userId', 'quantity'
Items	'itemId', 'category'
Users	'userId', 'registerCountry'
<b>Properties</b>	
dropped 'Nan' and 'None' values from 'category' in Items	
dropped 'Nan' and 'None' values from 'quantity' in Conversions	
dropped 'Nan' and 'None' values from 'registerCountry' in Users	
<b>Actions</b>	
joining Items and Conversions on 'itemId' and futher with Users on 'userId'	
filtering joined data on 'country' property	
performing sum on 'quantity' in rows grouped by 'category'	
<b>Axes</b>	
x: category	
y: quantity	



Above plot is an example generated for United States. It indicates that some categories are extremely popular, whereas purchases in the others are on similar level.

## number\_of\_purchased\_items\_grouped\_by\_categories\_in\_all\_countries()

### Data

Tables	Columns
Conversions	'itemId', 'userId', 'quantity'
Items	'itemId', 'category'
Users	'userId', 'registerCountry'

### Properties

dropped 'Nan' and 'None' values from 'category' in Items

dropped 'Nan' and 'None' values from 'quantity' in Conversions

dropped 'Nan' and 'None' values from 'registerCountry' in Users

### Actions

joining Items and Conversions on 'itemId' and further with Users on 'userId'

generating y-axis' ticks on unique 'registerCountry' values

generating x-axis' ticks on unique 'category' values

filtering data on 'registercountry' and 'category' property

performing sum() operation on 'quantity' in filtered rows

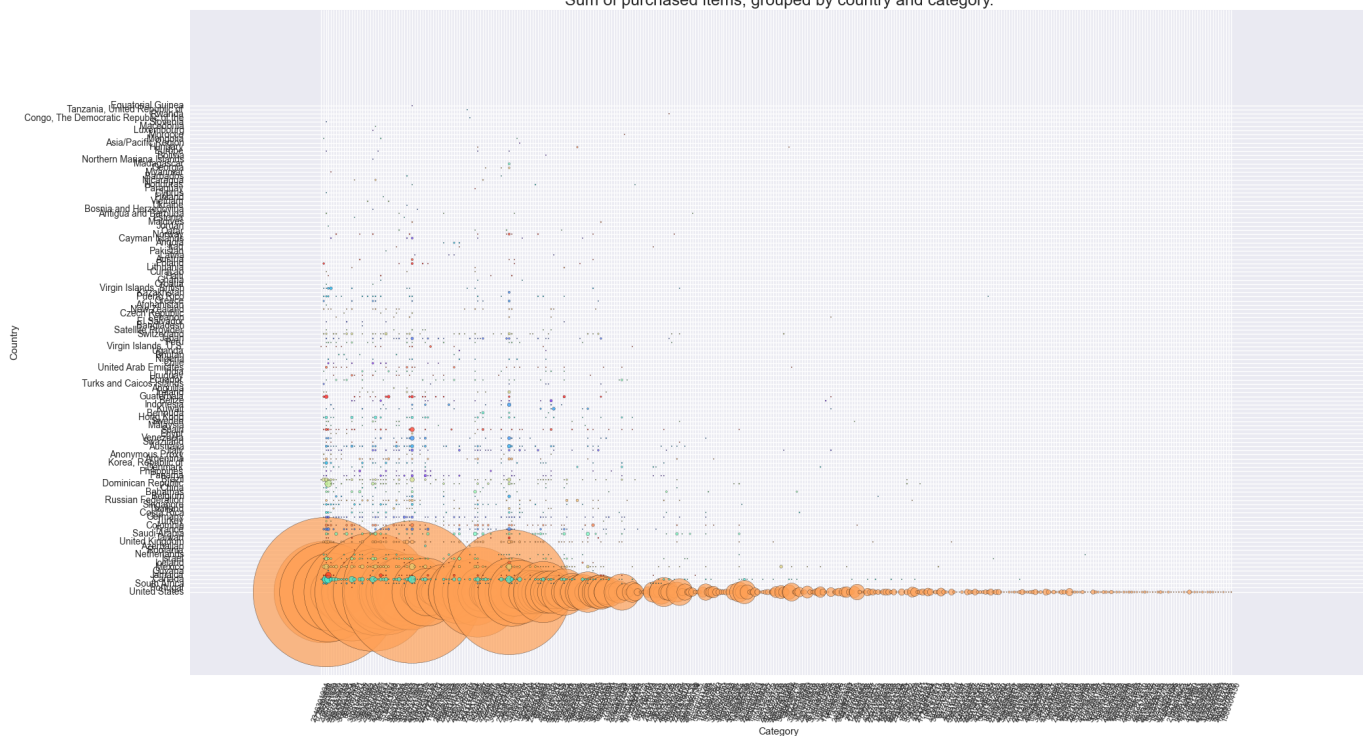
setting ticks and labels on the plot

### Axes

x: category

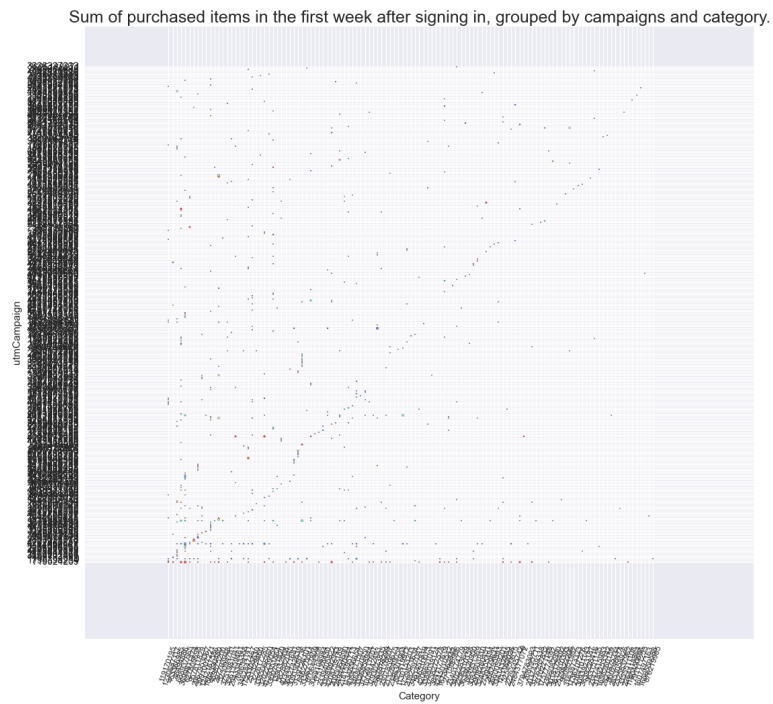
y: country

Sum of purchased items, grouped by country and category.



As can be seen United States are extremely important client for the service. The others countries whose impact in total amount of purchased products is significant are: Dominican Republic, Guatemala, Spain, Venezuela Canada and Mexico.

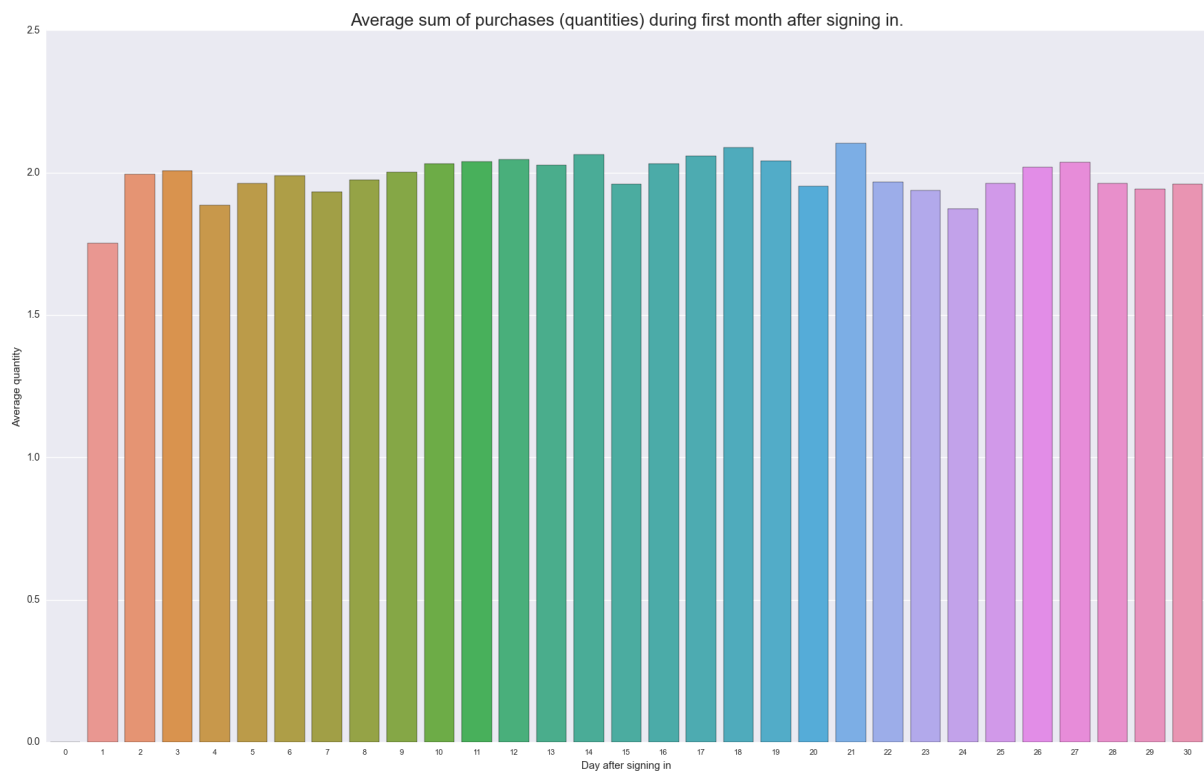
<b>number_of_purchased_items_after_seeing_campaigns_grouped_by_categories()</b>	
<b>Data</b>	
<b>Tables</b>	<b>Columns</b>
Conversions	'itemId', 'userId', 'quantity'
Users_Ads	'itemId', 'category'
Users	'userId', 'registerCountry'
Items	
<b>Properties</b>	
dropped 'Nan' and 'None' values from 'timestamp' in Conversions	
dropped 'Nan' and 'None' values from 'signupTime' in Users	
dropped 'Nan' and 'None' values from 'utmCampaign' in Users_Ads	
dropped 'Nan' and 'None' values from 'category' in Items	
<b>Actions</b>	
adding additional column to Users: 'week_after' - date week after signing in	
joining Users, Users_ads and Conversions on 'userId' and further with Items on 'itemId'	
filtering joined structure: joined['timestamp'] <= joined['week_after']	
generating y-axis' ticks on unique 'utmCampaign' values	
generating x-axis' ticks on unique 'category' values	
filtering data on 'utmCampaign' property	
performing sum() operation on 'quantity' in filtered rows	
setting ticks and labels on the plot	
<b>Axes</b>	
x: category	
y: utmCampaign	



Above plot is generated on a random sample (0.3) of the data.



<b>average_number_of_purchased_items_during_the_first_month_after_signing_in()</b>	
<b>Data</b>	
<b>Tables</b>	<b>Columns</b>
Conversions	'userId', 'timestamp', 'quantity'
Users	'userId', 'signupTime'
<b>Properties</b>	
dropped 'Nan' and 'None' values from 'timestamp' and 'quantity' in Conversions	
dropped 'Nan' and 'None' values from 'signupTime' in Users	
<b>Actions</b>	
joining Users and Conversions on 'userId'	
adding additional column: 'purchase_day' – number of days after 'signupTime', when conversion was completed	
filtering joined structure: $0 \leq \text{'purchase\_day'} < 30$	
grouping rows on 'purchase_day' and 'userId' (one user can have many conversions during one day)	
performing sum() operation on grouped structure, a result is number of purchased items in particular day after signing in	
counting average: for each day in the range (0, 30]	
<b>Axes</b>	
x: day after signing in	
y: average quantity	



As it can be seen, the plot is monotonic, the average number of purchased items per day during first month after signing in, oscillates around 2.