Spain Translation A/B Test

- Confirm that the test is actually negative. That is, it appears that the old version of the site with just one translation across Spain and LatAm performs better
- Higher conversion dare using one translation cross spain (test=0) and LatAM compare with localized translation (test=1) confirms the test is actually negative

conversion_Sum	test	Record_Count	Conversion_Rate
8949	0	185311	0.048
9367	1	215774	0.043

Z-score = 7.35
P-value = 1
With 100% confidence level, we can conclude that the old version outperforms the localized translation version in conversion rate

- The steps involved:
  - Merge test_table.csv with user_table.csv (use user_id as key and inner join operation)
  - Select non-spain countries records
  - Group conversion count by test value (0 and 1)
  - Compute conversion rate using conversion_sum / read count

Spain Translation A/B Test

- Explain why that might be happening. Are the localized translations really worse?

Large discrepancy in the sample size between the control & test sets is not observed in the other fields and values

conversion_Sum	test	country	Record_Count	Conversion_Rate
141	0	Argentina	9356	0.015
513	1	Argentina	37377	0.014
274	0	Bolivia	5550	0.049
267	1	Bolivia	5574	0.048
474	0	Chile	9853	0.048
507	1	Chile	9884	0.051
1411	0	Colombia	27088	0.052
1364	1	Colombia	26972	0.051
139	0	Costa Rica	2660	0.052
145	1	Costa Rica	2649	0.055
395	0	Ecuador	8036	0.049
385	1	Ecuador	7859	0.049
220	0	El Salvador	4108	0.054
195	1	El Salvador	4067	0.048
386	0	Guatemala	7622	0.051
365	1	Guatemala	7503	0.049
222	0	Honduras	4361	0.051
200	1	Honduras	4207	0.048
3178	0	Mexico	64209	0.049
3290	1	Mexico	64275	0.051
180	0	Nicaragua	3419	0.053
179	1	Nicaragua	3304	0.054
92	0	Panama	1966	0.047
98	1	Panama	1985	0.049
177	0	Paraguay	3650	0.048
182	1	Paraguay	3697	0.049
842	0	Peru	16869	0.050
850	1	Peru	16797	0.051
5	0	Uruguay	415	0.012
48	1	Uruguay	3719	0.013
813	0	Venezuela	16149	0.050
779	1	Venezuela	15905	0.049

There is a large discrepancy in the sample size between the control & test sets of Argentina and Uruguay

source	conversion_Sum	test	Record_Count	Conversion_Rate
Ads	3620	0	74352	0.049
Ads	3763	1	86448	0.044
Direct	1833	0	37238	0.049
Direct	1830	1	43047	0.043
SEO	3496	0	73721	0.047
SEO	3774	1	86279	0.044

device	conversion_Sum	test	Record_Count	Conversion_Rate
Mobile	4019	0	82350	0.049
Mobile	4236	1	96256	0.044
Web	4930	0	102961	0.048
Web	5131	1	119518	0.043

browser_language	conversion_Sum	test	Record_Count	Conversion_Rate
EN	1207	0	25777	0.047
EN	1283	1	30106	0.043
ES	7508	0	154416	0.049
ES	7816	1	179598	0.044
Other	234	0	5118	0.046
Other	268	1	6070	0.044

conversion_Sum	test	sex	Record_Count	Conversion_Rate
3724	0	F	77096	0.048
3900	1	F	89909	0.043
5225	0	M	108215	0.048
5467	1	M	125865	0.043

ads_channel	conversion_Sum	test	Record_Count	Conversion_Rate
Bing	254	0	5565	0.046
Bing	274	1	6508	0.042
Facebook	1402	0	27846	0.050
Facebook	1439	1	32607	0.044
Google	1355	0	27923	0.049
Google	1392	1	32378	0.043
NA	5329	0	110959	0.048
NA	5604	1	129326	0.043
Other	71	0	1716	0.041
Other	73	1	1930	0.038
Yahoo	538	0	11302	0.048
Yahoo	585	1	13025	0.045

conversion_Sum	test	age	Record_Count	Conversion_Rate
519	0	18	10556	0.049
534	1	18	12412	0.043
567	0	19	11023	0.051
561	1	19	12701	0.044
524	0	20	11107	0.047
513	1	20	12821	0.040
534	0	21	11301	0.047
586	1	21	13150	0.045
519	0	22	11127	0.047
585	1	22	13012	0.045
525	0	23	10983	0.048
559	1	23	12927	0.043
534	0	24	10885	0.049
593	1	24	12801	0.046
497	0	25	10578	0.047
554	1	25	12549	0.044
481	0	26	10225	0.047
540	1	26	12020	0.045
463	0	27	9763	0.047
468	1	27	11259	0.042
461	0	28	9068	0.051
468	1	28	10558	0.044
405	0	29	8463	0.048
384	1	29	9912	0.039
383	0	30	7896	0.049
379	1	30	9179	0.041
322	0	31	7269	0.044
348	1	31	8382	0.042
299	0	32	6548	0.046
335	1	32	7665	0.044
286	0	33	5846	0.049
301	1	33	6722	0.045
251	0	34	5151	0.049
269	1	34	5969	0.045
216	0	35	4651	0.046
247	1	35	5333	0.046
207	0	36	3923	0.053
200	1	36	4550	0.044
169	0	37	3341	0.051
179	1	37	3916	0.046
165	0	38	3005	0.055
138	1	38	3388	0.041
106	0	39	2369	0.045
129	1	39	2808	0.046
118	0	40	2087	0.057
116	1	40	2340	0.050
94	0	41	1722	0.055
79	1	41	1966	0.040
58	0	42	1339	0.043
60	1	42	1636	0.037
59	0	43	1095	0.054
38	1	43	1272	0.030
49	0	44	908	0.054
41	1	44	1007	0.041
35	0	45	737	0.047

Spain Translation A/B Test

- Explain why that might be happening. Are the localized translations really worse?

conversion_Sum	test	country	Record_Count	Conversion_Rate
141	0	Argentina	9356	0.015
513	1	Argentina	37377	0.014
274	0	Bolivia	5550	0.049
267	1	Bolivia	5574	0.048
474	0	Chile	9853	0.048
507	1	Chile	9884	0.051
1411	0	Colombia	27088	0.052
1364	1	Colombia	26972	0.051
139	0	Costa Rica	2660	0.052
145	1	Costa Rica	2649	0.055
395	0	Ecuador	8036	0.049
385	1	Ecuador	7859	0.049
220	0	El Salvador	4108	0.054
195	1	El Salvador	4067	0.048
386	0	Guatemala	7622	0.051
365	1	Guatemala	7503	0.049
222	0	Honduras	4361	0.051
200	1	Honduras	4207	0.048
3178	0	Mexico	64209	0.049
3290	1	Mexico	64275	0.051
180	0	Nicaragua	3419	0.053
179	1	Nicaragua	3304	0.054
92	0	Panama	1966	0.047
98	1	Panama	1985	0.049
177	0	Paraguay	3650	0.048
182	1	Paraguay	3697	0.049
842	0	Peru	16869	0.050
850	1	Peru	16797	0.051
5	0	Uruguay	415	0.012
48	1	Uruguay	3719	0.013
813	0	Venezuela	16149	0.050
779	1	Venezuela	15905	0.049

There is a large discrepancy in the sample size between the control & variation sets of Argentina and Uruguay

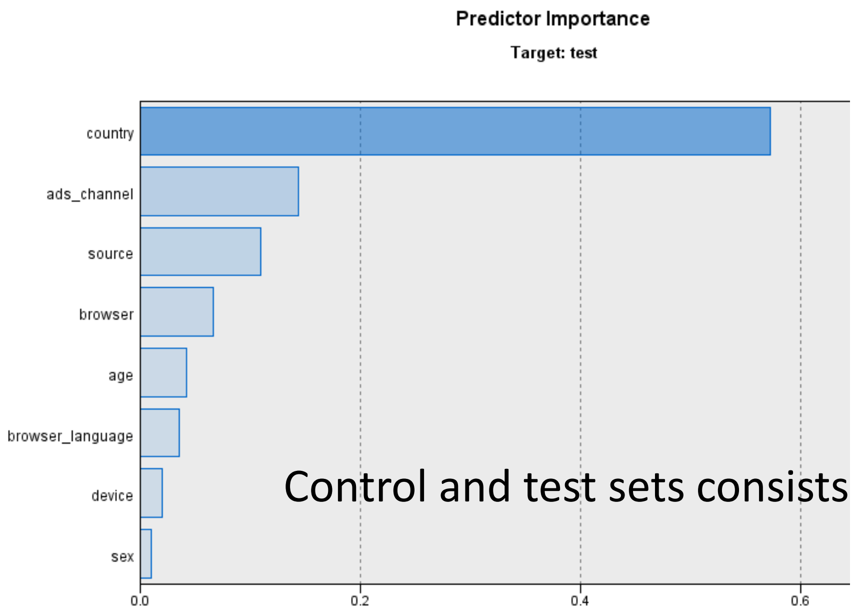
- The negative test is the result of sampling bias represented by the large discrepancy in the sample size between the control & test sets of Argentina and Uruguay

- Records of these two countries are removed from the control and test sets. The conversion rates of both control and test people are found to be equal (5%). In conclusion, the localized translations are not really worse than Spanish translation if the sampling bias is removed.

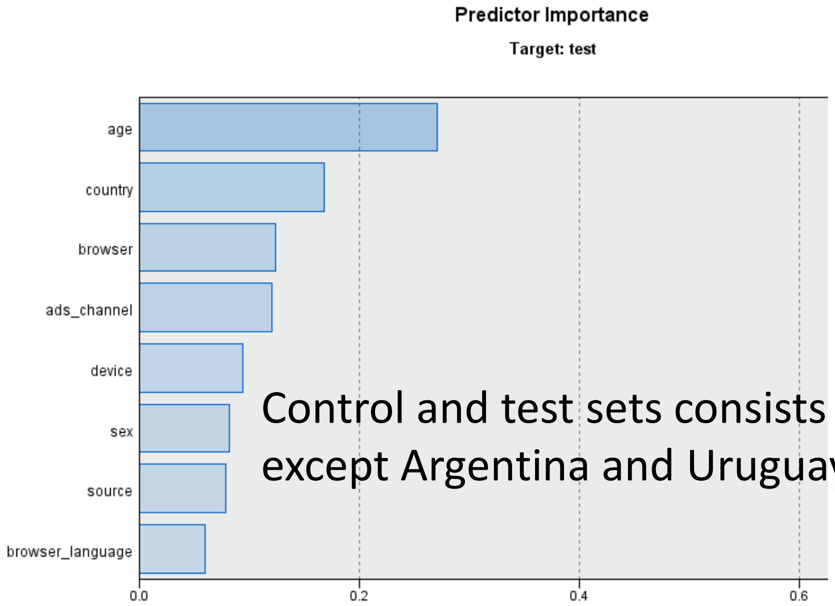
conversion_Sum	test	Record_Count	Conversion_Rate
8803	0	175540	0.050
8806	1	174678	0.050

Spain Translation A/B Test

- If you identified what was wrong, design an algorithm that would return FALSE if the same problem is happening in the future and TRUE if everything is good and the results can be trusted.
  - My proposed solution to help answering the sampling bias problem is by fitting the dataset (using test as the target variable and the rest of variables as independent variables) to a machine learner (i.e. chaid decision tree, random forest, gbm, or neural network etc) and obtain and evaluate the relative predictor importance result
  - The predictor importance chart indicates the relative importance of each variable in determining which records goes into control or test set. Since the values are relative, the sum of the values for all predictors on the display is 1.0
  - As shown on the left figure, country is the prevalent variable (importance ~= 0.6 or 60%) found after fitting a machine learner, which is also indicative that the sampling process according to country is not random
  - So we can apply this technique to return FALSE if there is one ore more independent variables found to be prevalent in the relative predictor importance result, if not, return TRUE and everything is good and results can be trusted



Control and test sets consists of all countries

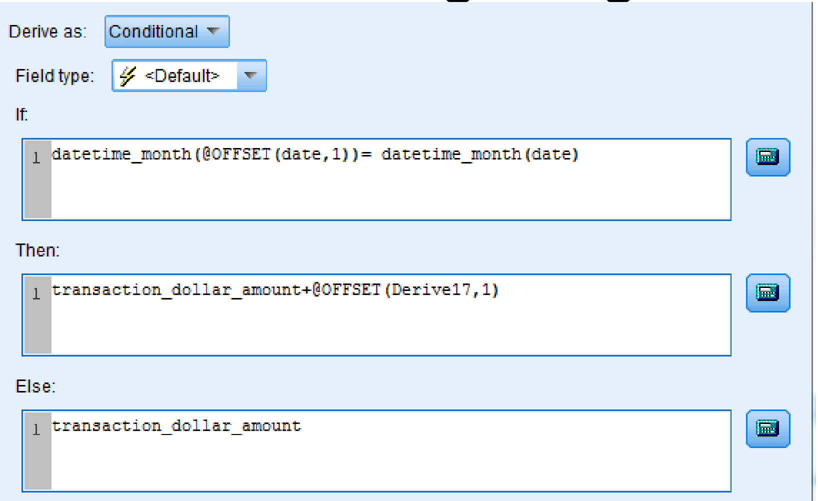


Control and test sets consists of all countries except Argentina and Uruguay

Credit Card Transaction

- Build an algorithm that identifies users who have exceeded their monthly credit limit on any day.
- Using a series of processing steps to formulate this algorithm
  - Join the transactions and cc_info datasets using credit_card field as the key
  - Sort the cred_card field in ascending order, followed by date in ascending order, so that we can have the table formatted according to account holder first then his/her transactions in sequential date format
  - Then accumulate the transaction amounts up until to date

but refresh at the beginning of every month, see the figures below



date	transaction_dollar_amount	Derive17
2015-08-26 18:53:43	13.430	6964.160
2015-08-26 22:44:20	48.870	7013.030
2015-08-26 23:33:58	122.710	7135.740
2015-08-27 20:01:25	87.700	7223.440
2015-08-27 21:34:58	89.010	7312.450
2015-08-27 22:16:39	160.670	7473.120
2015-08-28 16:07:40	109.940	7583.060
2015-08-28 18:30:34	62.090	7645.150
2015-08-29 00:35:11	104.920	7750.070
2015-08-29 16:38:37	44.780	7794.850
2015-08-29 18:34:04	122.650	7917.500
2015-08-29 18:36:10	118.760	8036.260
2015-08-29 21:24:04	26.490	8062.750
2015-08-30 14:29:15	133.360	8196.110
2015-08-30 21:26:52	56.890	8253.000
2015-08-31 02:03:20	98.750	8351.750
2015-08-31 16:54:11	89.760	8441.510
2015-09-01 00:50:44	118.840	118.840
2015-09-01 18:53:30	122.800	241.640
2015-09-01 22:55:09	93.730	335.370
2015-09-02 01:54:20	124.630	460.000
2015-09-02 20:54:00	905.540	1365.540
2015-09-03 16:41:23	141.320	1506.860
2015-09-03 18:36:35	115.630	1622.490
2015-09-03 19:17:22	91.740	1714.230
2015-09-03 20:46:36	68.380	1782.610
2015-09-04 20:11:59	143.650	1926.260
2015-09-04 22:03:33	162.050	2088.310
2015-09-04 22:08:31	80.480	2168.790
2015-09-04 22:55:36	162.610	2331.400
2015-09-05 20:12:57	200.340	2531.740
2015-09-05 21:36:43	115.020	2646.760
2015-09-07 00:29:10	99.450	2746.210
2015-09-07 21:58:15	110.130	2856.340

Credit Card Transaction

- Using a series of processing steps to formulate this algorithm
  - Check the accumulated transaction amount with respect to the monthly credit limit, and set a flag if the limit is exceeded

Yes/No Flag as indicato
of over spending limit

credit_card	date	transaction_dollar_amount	Long	Lat	city	state	zipcode	credit_card_limit	Derive17	ExceedLimit_Flag
4298557099672376	2015-10-21 02:41:24	46.780	-79.871	40.454	Pittsburgh	PA	15201	8000	5556.840	NO
4298557099672376	2015-10-21 20:40:41	103.690	-79.950	40.424	Pittsburgh	PA	15201	8000	5660.530	NO
4298557099672376	2015-10-21 21:06:47	101.670	-79.969	40.433	Pittsburgh	PA	15201	8000	5762.200	NO
4298557099672376	2015-10-22 04:32:57	44.410	-79.899	40.422	Pittsburgh	PA	15201	8000	5806.610	NO
4298557099672376	2015-10-22 18:32:36	122.000	-79.977	40.434	Pittsburgh	PA	15201	8000	5928.610	NO
4298557099672376	2015-10-22 18:48:38	47.330	-79.917	40.549	Pittsburgh	PA	15201	8000	5975.940	NO
4298557099672376	2015-10-22 18:51:17	134.140	-80.019	40.465	Pittsburgh	PA	15201	8000	6110.080	NO
4298557099672376	2015-10-22 21:03:36	155.970	-79.923	40.504	Pittsburgh	PA	15201	8000	6266.050	NO
4298557099672376	2015-10-22 21:16:58	22.860	-79.914	40.413	Pittsburgh	PA	15201	8000	6288.910	NO
4298557099672376	2015-10-22 21:57:13	73.660	-79.995	40.395	Pittsburgh	PA	15201	8000	6362.570	NO
4298557099672376	2015-10-22 22:39:37	117.080	-79.879	40.441	Pittsburgh	PA	15201	8000	6479.650	NO
4298557099672376	2015-10-23 00:22:10	172.400	-79.978	40.422	Pittsburgh	PA	15201	8000	6652.050	NO
4298557099672376	2015-10-23 18:00:40	31.950	-79.914	40.495	Pittsburgh	PA	15201	8000	6684.000	NO
4298557099672376	2015-10-23 20:49:01	70.070	-79.987	40.544	Pittsburgh	PA	15201	8000	6754.070	NO
4298557099672376	2015-10-24 00:48:03	97.830	-79.904	40.527	Pittsburgh	PA	15201	8000	6851.900	NO
4298557099672376	2015-10-24 19:21:26	6.660	-79.968	40.465	Pittsburgh	PA	15201	8000	6858.560	NO
4298557099672376	2015-10-24 20:04:00	136.950	-79.983	40.402	Pittsburgh	PA	15201	8000	6995.510	NO
4298557099672376	2015-10-24 20:57:04	87.340	-79.934	40.406	Pittsburgh	PA	15201	8000	7082.850	NO
4298557099672376	2015-10-24 22:46:02	68.470	-80.020	40.454	Pittsburgh	PA	15201	8000	7151.320	NO
4298557099672376	2015-10-24 23:04:45	6.320	-79.984	40.406	Pittsburgh	PA	15201	8000	7157.640	NO
4298557099672376	2015-10-24 23:25:52	53.550	-79.976	40.485	Pittsburgh	PA	15201	8000	7211.190	NO
4298557099672376	2015-10-25 00:14:24	24.180	-80.027	40.474	Pittsburgh	PA	15201	8000	7235.370	NO
4298557099672376	2015-10-25 21:16:57	70.790	-79.884	40.509	Pittsburgh	PA	15201	8000	7306.160	NO
4298557099672376	2015-10-26 00:27:54	84.230	-79.904	40.476	Pittsburgh	PA	15201	8000	7390.390	NO
4298557099672376	2015-10-26 00:47:04	81.180	-80.035	40.488	Pittsburgh	PA	15201	8000	7471.570	NO
4298557099672376	2015-10-26 01:26:05	23.450	-80.008	40.423	Pittsburgh	PA	15201	8000	7495.020	NO
4298557099672376	2015-10-26 17:03:24	46.880	-79.966	40.553	Pittsburgh	PA	15201	8000	7541.900	NO
4298557099672376	2015-10-26 19:55:11	19.030	-80.038	40.507	Pittsburgh	PA	15201	8000	7560.930	NO
4298557099672376	2015-10-26 21:03:25	43.060	-79.903	40.453	Pittsburgh	PA	15201	8000	7603.990	NO
4298557099672376	2015-10-26 21:03:26	13.070	-79.891	40.477	Pittsburgh	PA	15201	8000	7617.060	NO
4298557099672376	2015-10-26 21:03:27	87.890	-80.001	40.491	Pittsburgh	PA	15201	8000	7704.950	NO
4298557099672376	2015-10-27 17:42:56	112.260	-79.989	40.477	Pittsburgh	PA	15201	8000	7817.210	NO
4298557099672376	2015-10-27 22:45:57	96.760	-79.885	40.452	Pittsburgh	PA	15201	8000	7913.970	NO
4298557099672376	2015-10-27 23:16:09	108.770	-79.970	40.414	Pittsburgh	PA	15201	8000	8022.740	YES
4298557099672376	2015-10-28 01:54:11	53.330	-79.958	40.498	Pittsburgh	PA	15201	8000	8076.070	YES
4298557099672376	2015-10-28 17:02:05	97.480	-80.002	40.426	Pittsburgh	PA	15201	8000	8173.550	YES
4298557099672376	2015-10-28 19:30:44	56.280	-79.982	40.408	Pittsburgh	PA	15201	8000	8229.830	YES
4298557099672376	2015-10-28 20:23:44	39.160	-79.922	40.488	Pittsburgh	PA	15201	8000	8268.990	YES
4298557099672376	2015-10-28 20:44:21	51.460	-80.023	40.513	Pittsburgh	PA	15201	8000	8320.450	YES
4298557099672376	2015-10-28 21:08:56	122.750	-79.898	40.527	Pittsburgh	PA	15201	8000	8443.200	YES
4298557099672376	2015-10-28 21:16:46	62.500	-79.958	40.483	Pittsburgh	PA	15201	8000	8505.700	YES
4298557099672376	2015-10-29 01:54:49	71.770	-79.954	40.415	Pittsburgh	PA	15201	8000	8577.470	YES
4298557099672376	2015-10-29 18:10:04	44.730	-79.912	40.479	Pittsburgh	PA	15201	8000	8622.200	YES
4298557099672376	2015-10-29 20:05:28	53.840	-79.971	40.465	Pittsburgh	PA	15201	8000	8676.040	YES
4298557099672376	2015-10-29 23:27:43	68.150	-79.997	40.400	Pittsburgh	PA	15201	8000	8744.190	YES
4302633772248169	2015-07-31 14:42:21	93.780	-74.593	40.396	Dayton	NJ	8810	16000	93.780	NO
4302633772248169	2015-07-31 15:38:00	22.970	-74.598	40.355	Dayton	NJ	8810	16000	116.750	NO
4302633772248169	2015-07-31 17:47:22	46.620	-74.523	40.337	Dayton	NJ	8810	16000	163.370	NO
4302633772248169	2015-07-31 18:56:50	21.840	-74.516	40.351	Dayton	NJ	8810	16000	185.210	NO
4302633772248169	2015-07-31 19:13:01	19.040	-74.479	40.396	Dayton	NJ	8810	16000	204.250	NO

The credit card holder exceeded his/her card spending limit on 2015-10-27

Credit Card Transaction

- Implement an unsupervised classifier which returns suspicious / anomalous transactions
  - There are two scenarios to this question
    - Overall anomalous transactions based on all transactions of individual card holder
    - Anomalous transactions from the card holder’s usual spending patterns (meaning he/she used to spend on two groups of items, expensive items (i.e. gadgets) and cheap items (i.e grocery items), so it is incorrect to classify a transaction as anomalous because he/she occasionally shop for expensive items and hence its frequency is considerably less the frequency of grocery shopping
    - Both scenarios can be answered using anomaly detection algorithm which has clustering technique built into it.
    - My suggestion is to build an anomaly detection model for every card holder, and since clustering algorithm takes in multi dimensional variables, I suggest to include more relevant variables (types of item, purchase location, online/offline purchase etc). For this example, I use transaction_dollar_amount as the only variable in the clustering algorithm.

Credit Card Transaction

- Implement an unsupervised classifier which returns suspicious / anomalous transactions
  - Answer (using credit_card=1003715054175576 as an example)

Preset the number of cluster (peer groups) = 3, The mean transaction_amount of these groups are: 137.756, 63.829 and 898.75 respectively. We can then classify peer group-3 (6 records) transactions are anomalous because they are farthest away from the mean of all transaction amounts

- Peer group-1: 97 records
  - Anomalies: found 4 records from an estimated total of 97 records

Contribution	Count	Average index
transaction_dollar_amount	4	1.0

    - Residual of the unreported reasons: 0%
  - Peer group profile
    - transaction_dollar_amount (137.756)
      - Standard deviation: 23.741
- Peer group-2: 158 records
  - Anomalies: found 9 records from an estimated total of 158 records

Contribution	Count	Average index
transaction_dollar_amount	9	1.0

    - Residual of the unreported reasons: 0%
  - Peer group profile
    - transaction_dollar_amount (63.829)
      - Standard deviation: 27.441
- Peer group-3: 6 records
  - Peer group profile
    - transaction_dollar_amount (898.748)
      - Standard deviation: 60.371

credit_card	date	transaction_dollar_amount	Long	Lat	city	state	zipcode	credit_card_limit	$O-Anomaly	$O-AnomalyIndex	$O-PeerGroup	$O-Field-1
1003715054175576	2015-08-12 15:58:53	88.870	-80.169	40.207	Houston	PA	15342	20000	F	0.865	2	transaction_dollar_amount
1003715054175576	2015-08-10 20:24:04	92.210	-80.152	40.257	Houston	PA	15342	20000	F	1.079	2	transaction_dollar_amount
1003715054175576	2015-08-10 17:37:51	68.860	-80.190	40.234	Houston	PA	15342	20000	F	0.142	2	transaction_dollar_amount
1003715054175576	2015-08-09 23:45:30	6.820	-80.245	40.221	Houston	PA	15342	20000	T	4.008	2	transaction_dollar_amount
1003715054175576	2015-08-31 02:03:20	98.750	-80.215	40.325	Houston	PA	15342	20000	F	1.576	2	transaction_dollar_amount
1003715054175576	2015-08-08 20:41:08	73.080	-80.183	40.166	Houston	PA	15342	20000	F	0.215	2	transaction_dollar_amount
1003715054175576	2015-08-30 21:26:52	56.890	-80.213	40.291	Houston	PA	15342	20000	F	0.170	2	transaction_dollar_amount
1003715054175576	2015-09-08 18:59:41	92.490	-80.188	40.265	Houston	PA	15342	20000	F	1.099	2	transaction_dollar_amount
1003715054175576	2015-08-07 22:06:15	87.040	-80.220	40.194	Houston	PA	15342	20000	F	0.759	2	transaction_dollar_amount
1003715054175576	2015-08-29 21:24:04	26.490	-80.211	40.265	Houston	PA	15342	20000	F	1.786	2	transaction_dollar_amount
1003715054175576	2015-09-08 16:43:42	99.270	34.429	24.660	Houston	PA	15342	20000	F	1.620	2	transaction_dollar_amount
1003715054175576	2015-08-07 17:42:28	53.800	-80.250	40.218	Houston	PA	15342	20000	F	0.233	2	transaction_dollar_amount
1003715054175576	2015-08-07 02:28:31	71.020	-80.283	40.189	Houston	PA	15342	20000	F	0.174	2	transaction_dollar_amount
1003715054175576	2015-08-06 23:52:05	57.560	-80.202	40.268	Houston	PA	15342	20000	F	0.159	2	transaction_dollar_amount
1003715054175576	2015-08-06 21:38:18	84.880	-80.244	40.277	Houston	PA	15342	20000	F	0.644	2	transaction_dollar_amount
1003715054175576	2015-08-06 19:20:58	61.270	-80.252	40.259	Houston	PA	15342	20000	F	0.120	2	transaction_dollar_amount
1003715054175576	2015-08-06 17:43:13	86.910	-80.228	40.240	Houston	PA	15342	20000	F	0.752	2	transaction_dollar_amount
1003715054175576	2015-08-06 17:08:38	38.620	-80.258	40.238	Houston	PA	15342	20000	F	0.875	2	transaction_dollar_amount
1003715054175576	2015-08-04 16:29:10	75.360	-80.200	40.163	Houston	PA	15342	20000	F	0.272	2	transaction_dollar_amount
1003715054175576	2015-08-04 00:50:23	91.790	-80.289	40.289	Houston	PA	15342	20000	F	1.051	2	transaction_dollar_amount
1003715054175576	2015-08-29 16:38:37	44.780	-80.233	40.315	Houston	PA	15342	20000	F	0.548	2	transaction_dollar_amount
1003715054175576	2015-08-03 18:51:08	54.870	-80.185	40.200	Houston	PA	15342	20000	F	0.208	2	transaction_dollar_amount
1003715054175576	2015-08-03 00:27:43	80.670	-80.238	40.225	Houston	PA	15342	20000	F	0.453	2	transaction_dollar_amount
1003715054175576	2015-08-02 22:40:59	85.740	-80.245	40.255	Houston	PA	15342	20000	F	0.689	2	transaction_dollar_amount
1003715054175576	2015-09-07 21:58:16	61.030	-80.250	40.162	Houston	PA	15342	20000	F	0.121	2	transaction_dollar_amount
1003715054175576	2015-08-01 19:10:09	97.350	-80.164	40.203	Houston	PA	15342	20000	F	1.461	2	transaction_dollar_amount
1003715054175576	2015-08-01 17:44:19	96.970	-80.145	40.286	Houston	PA	15342	20000	F	1.431	2	transaction_dollar_amount
1003715054175576	2015-08-28 18:30:34	62.090	-80.180	40.209	Houston	PA	15342	20000	F	0.115	2	transaction_dollar_amount
1003715054175576	2015-08-01 10:48:03	51.270	-80.177	40.313	Houston	PA	15342	20000	F	0.301	2	transaction_dollar_amount
1003715054175576	2015-10-18 01:01:41	888.040	-80.232	40.297	Houston	PA	15342	20000	F	0.333	3	transaction_dollar_amount
1003715054175576	2015-10-01 20:44:09	818.470	-80.280	40.201	Houston	PA	15342	20000	F	1.790	3	transaction_dollar_amount
1003715054175576	2015-09-24 22:10:07	925.780	-80.164	40.238	Houston	PA	15342	20000	F	0.484	3	transaction_dollar_amount
1003715054175576	2015-08-24 15:14:52	859.310	-80.202	40.222	Houston	PA	15342	20000	F	0.682	3	transaction_dollar_amount
1003715054175576	2015-09-11 19:50:02	995.350	-80.127	40.226	Houston	PA	15342	20000	F	2.395	3	transaction_dollar_amount
1003715054175576	2015-09-02 20:54:00	905.540	-80.237	40.291	Houston	PA	15342	20000	F	0.316	3	transaction_dollar_amount

The field which indicates the peer group-3 in the transaction table

Credit Card Transaction

- Implement an unsupervised classifier which returns suspicious / anomalous transactions
  - Answer (using credit_card=1003715054175576 as an example)

To answer the second scenario, simply detect transactions which are located far away from the mean/centroid of each cluster and flag them as anomalous transactions.

- Peer group-1: 97 records
  - Anomalies: found 4 records from an estimated total of 97 records

Contribution	Count	Average index
transaction_dollar_amount	4	1.0

    - Residual of the unreported reasons: 0%
  - Peer group profile
    - transaction_dollar_amount (137.756)
      - Standard deviation: 23.741
- Peer group-2: 158 records
  - Anomalies: found 9 records from an estimated total of 158 records

Contribution	Count	Average index
transaction_dollar_amount	9	1.0

    - Residual of the unreported reasons: 0%
  - Peer group profile
    - transaction_dollar_amount (63.829)
      - Standard deviation: 27.441
- Peer group-3: 6 records
  - Peer group profile
    - transaction_dollar_amount (898.748)
      - Standard deviation: 60.371

This is an anomalous transaction which belongs to peer group-2 as its transaction amount of 7.27 is far below the mean of group=2 mean

1003715054175576	2015-08-06 17:08:38	38.620	-80.258	40.238	Houston	PA	15342	20000	F	0.875	2
1003715054175576	2015-08-04 16:29:10	75.360	-80.200	40.163	Houston	PA	15342	20000	F	0.272	2
1003715054175576	2015-08-04 00:50:23	91.790	-80.289	40.289	Houston	PA	15342	20000	F	1.051	2
1003715054175576	2015-08-03 19:00:49	106.180	-80.164	40.173	Houston	PA	15342	20000	F	1.678	1
1003715054175576	2015-08-03 18:51:08	54.870	-80.185	40.200	Houston	PA	15342	20000	F	0.208	2
1003715054175576	2015-08-03 00:27:43	80.670	-80.238	40.225	Houston	PA	15342	20000	F	0.453	2
1003715054175576	2015-08-02 22:40:59	85.740	-80.245	40.255	Houston	PA	15342	20000	F	0.689	2
1003715054175576	2015-10-23 16:22:08	7.270	-80.194	40.300	Houston	PA	15342	20000	T	3.947	2
1003715054175576	2015-10-17 23:02:40	201.810	-80.215	40.281	Houston	PA	15342	20000	T	6.610	1
1003715054175576	2015-10-16 21:07:15	11.520	-80.163	40.262	Houston	PA	15342	20000	T	3.393	2
1003715054175576	2015-10-16 16:29:54	0.770	-80.178	40.278	Houston	PA	15342	20000	T	4.875	2
1003715054175576	2015-10-09 19:59:16	6.930	-80.135	40.214	Houston	PA	15342	20000	T	3.993	2
1003715054175576	2015-10-07 01:16:12	0.970	-80.197	40.226	Houston	PA	15342	20000	T	4.845	2
1003715054175576	2015-09-19 21:22:52	190.270	-80.193	40.303	Houston	PA	15342	20000	T	4.480	1
1003715054175576	2015-09-18 18:04:16	0.880	-80.133	40.204	Houston	PA	15342	20000	T	4.859	2
1003715054175576	2015-08-09 23:45:30	6.820	-80.245	40.221	Houston	PA	15342	20000	T	4.008	2
1003715054175576	2015-08-26 18:53:43	13.430	-80.238	40.249	Houston	PA	15342	20000	T	3.159	2
1003715054175576	2015-08-14 21:15:45	9.340	-80.281	40.204	Houston	PA	15342	20000	T	3.672	2
1003715054175576	2015-09-05 20:12:57	200.340	-80.231	40.244	Houston	PA	15342	20000	T	6.316	1
1003715054175576	2015-09-07 21:58:17	199.570	-80.146	40.188	Houston	PA	15342	20000	T	6.165	1

This is an anomalous transaction which belongs to peer group-1 as its transaction amount of 190 is far away from the mean of group=1 mean