

Survival and Cohort Analysis Using Excel

By Noviandi

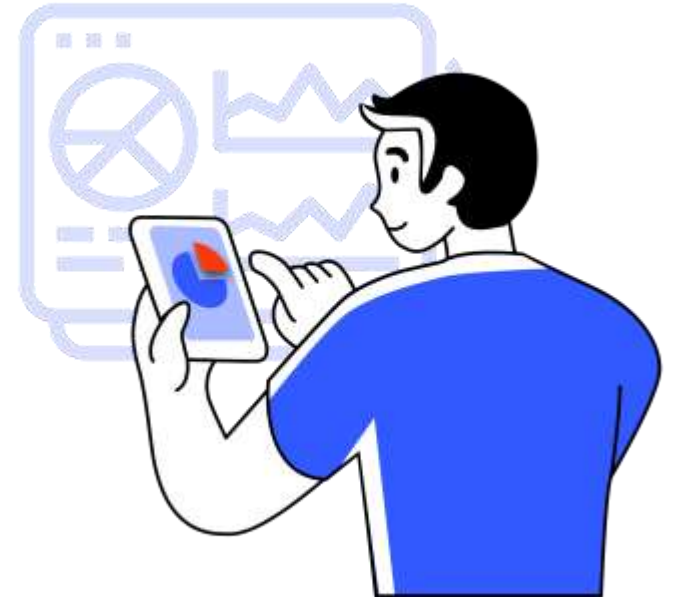


Table of Contents

1. Dataset & Tools
2. Objectives
3. Process
4. Summary



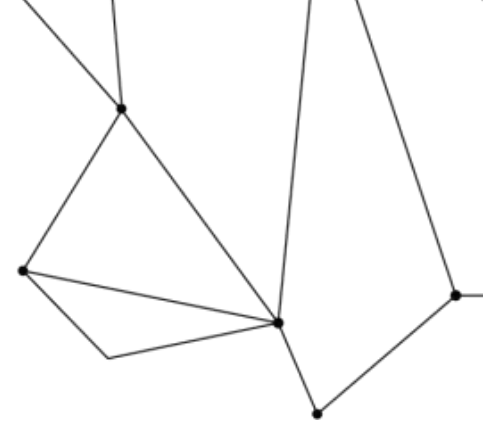
Dataset and Tools

Dataset: <https://www.kaggle.com/datasets/mashlyn/online-retail-ii-uci>

The dataset contains 15,433 rows and 8 columns

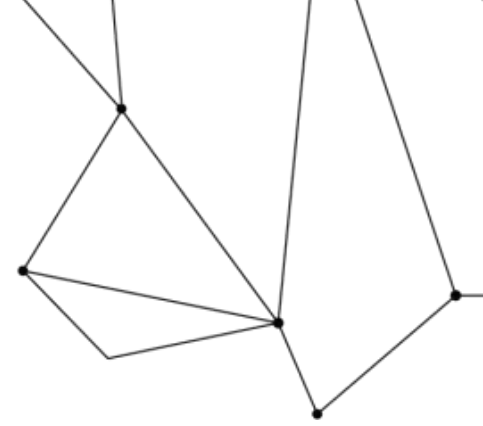
Attribute Information:

1. Country: Germany & Ireland
2. Customer ID: Customer number
3. Description: Product name
4. Invoice Date: Invoice date and time
5. Stock Code: Product (item) code
6. Invoice: ID of order
7. Price: Product price per unit
8. Quantity: The quantity of each product per transaction



Objectives

Analyze data using survival analysis and cohort analysis



What is Survival Analysis?

Survival analysis is a collection of statistical procedures for data analysis where the outcome variable of interest is time until an event occurs.

What is Cohort Analysis?

Cohort analysis is a kind of behavioral analytics that breaks the data in a dataset into related groups before analysis. These groups, or cohorts, usually share common characteristics or experiences within a defined time-span



PROCESS





Survival Analysis

1. Create new sheet to find unique customer ID and first invoice date values.

To get the first date of the invoice, look for the unique value of each customer ID

`=UNIQUE(online_retail_II_germany_and_ir!B:B)`

By using the MINIFS function we can find the first date of the transaction (first_invoice_date)

`=MINIFS(online_retail_II_germany_and_ir!D:D;online_retail_II_germany_and_ir!B:B;'first_invoice lookup'!A3)`

Change the data type of the first_invoice_date column to the date data type (column C)

A	B	C
Customer ID	first_invoice date	first_invoice date
14156	14/01/2011	14/01/2011
14911	09/01/2011	09/01/2011
12673	17/01/2011	17/01/2011
12474	14/01/2011	14/01/2011
12708	14/01/2011	14/01/2011
12501	07/01/2011	07/01/2011
12500	10/01/2011	10/01/2011
12527	18/01/2011	18/01/2011
12601	11/01/2011	11/01/2011
12665	19/01/2011	19/01/2011
12480	07/01/2011	07/01/2011
12481	07/01/2011	07/01/2011
12626	17/01/2011	17/01/2011
12530	07/01/2011	07/01/2011
12712	12/01/2011	12/01/2011
12471	13/01/2011	13/01/2011
12524	10/01/2011	10/01/2011

Survival Analysis

2. Added new columns

first_invoice_date

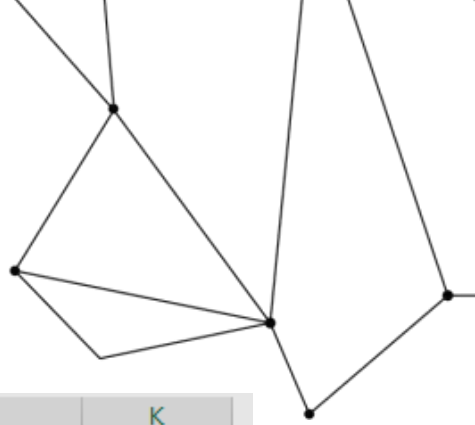
month_group

month_age

to fill in the first_invoice_date column you can
use the VLOOKUP or XLOOKUP function

```
=XLOOKUP(B968;'first_invoice lookup'!A:A;'first_invoice lookup'!C:C)
```

first_invoice date and month_group are created
to find out the first month and last month of the
invoice so we can know the month_age / range
of transactions that occur.



F	G	H	I	J	K
Invoice	Price	Quantit	first_invoice_date	month_grou	month_age
573748	4.95	1	09/01/2011	01/11/2011	10
572459	12.75	1	09/01/2011	01/10/2011	10
575728	16.95	1	09/01/2011	01/11/2011	10
574841	50	1	14/01/2011	01/11/2011	10
575735	50	1	14/01/2011	01/11/2011	10
572459	50	1	09/01/2011	01/10/2011	10
573391	50	1	09/01/2011	01/10/2011	10
573581	50	1	09/01/2011	01/10/2011	10
574745	50	1	09/01/2011	01/11/2011	10
574874	50	1	09/01/2011	01/11/2011	10
575728	4.95	2	09/01/2011	01/11/2011	10
575728	5.75	2	09/01/2011	01/11/2011	10
573892	7.5	2	09/01/2011	01/11/2011	10
573391	7.95	2	09/01/2011	01/10/2011	10
573581	7.95	2	09/01/2011	01/10/2011	10

Survival Analysis

3. Pivot Table

Create a summary group using pivot tables

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P
1																
2																
3	Count of Customer ID by Year and Month															
4	Years	month_group	0	1	2	3	4	5	6	7	8	9	10	11	Grand Total	
5	2011	Jan	878	73												951
6	2011	Feb	214	428												642
7	2011	Mar	216	51	752	89										1108
8	2011	Apr	176	13	136	416	5									746
9	2011	May	168	68	100	192	733	28								1289
10	2011	Jun	182		12	6	138	729	125							1192
11	2011	Jul	263	4		76	90	75	648	215						1371
12	2011	Aug	193	37	50	30	19	38	57	867	97					1388
13	2011	Sep	144	96	93	42	32	45	111	66	696	523				1848
14	2011	Oct	200	48	44	80		99	21	133	138	1320	74			2157
15	2011	Nov	123	52		63	29	81		19	93	101	1175	369		2105
16	2011	Dec	38	24	13	13	18	28			37		89	376		636
17	Grand Total		2795	894	1200	1007	1064	1123	962	1300	1061	1944	1338	745	15433	
18																
19																
20																
21																
22																
23																
24																
25																
26																
27																
28																
29																
30																
31																
32																
33																
34																

PivotTable Fields

Choose fields to add to report:

Search

- ☐ Country
- ☒ Customer ID
- ☐ Description
- ☐ Invoice Date
- ☐ Stock Code
- ☐ Invoice

Drag fields between areas below:

Filters	Columns
	month_age
Rows	Values
Years	Count of Customer ID
month_group	



Copy and paste the value of pivot table to a new sheet to make it easier to analyze

Survival Analysis

4. Edit Table

edit the table to simplify the analysis process such as creating a new column obs_group: Concat year and month

obs_group	Grand Total	0	1	2	3	4	5	6	7	8	9	10	11
2011 : Jan	951	73	0	0	0	0	0	0	0	0	0	0	0
2011 : Feb	642	428	0	0	0	0	0	0	0	0	0	0	0
2011 : Mar	1108	892	841	89	0	0	0	0	0	0	0	0	0
2011 : Apr	746	570	557	421	5	0	0	0	0	0	0	0	0
2011 : May	1289	1121	1053	953	761	28	0	0	0	0	0	0	0
2011 : Jun	1192	1010	1010	998	992	854	125	0	0	0	0	0	0
2011 : Jul	1371	1108	1104	1104	1028	938	863	215	0	0	0	0	0
2011 : Aug	1388	1195	1158	1108	1078	1059	1021	964	97	0	0	0	0
2011 : Sep	1848	1704	1608	1515	1473	1441	1396	1285	1219	523	0	0	0
2011 : Oct	2157	1957	1909	1865	1785	1785	1686	1665	1532	1394	74	0	0
2011 : Nov	2105	1982	1930	1930	1867	1838	1757	1757	1738	1645	1544	369	0
2011 : Dec	636	598	574	561	548	530	502	502	502	465	465	376	0
Grand Total :	15433	12638	11744	10544	9537	8473	7350	6388	5088	4027	2083	745	0

5. User Profile and Num of User Loss

Copy the pivot table that has been edited into a user profile table and a num of user loss table

To get the number of user profiles,
subtract the grand total with the
number of user loss

C40		=B40-C56													
	A	B	C	D	E	F	G	H	I	J	K	L	M	N	
38		User profile:													
39	obs_group	Grand Total	0	1	2	3	4	5	6	7	8	9	10	11	
40	2011 : Jan	951	73	0	0	0	0	0	0	0	0	0	0	0	
41	2011 : Feb	642	428	0	0	0	0	0	0	0	0	0	0	0	
42	2011 : Mar	1108	892	841	89	0	0	0	0	0	0	0	0	0	
43	2011 : Apr	746	570	557	421	5	0	0	0	0	0	0	0	0	
44	2011 : May	1289	1121	1053	953	761	28	0	0	0	0	0	0	0	
45	2011 : Jun	1192	1010	1010	998	992	854	125	0	0	0	0	0	0	
46	2011 : Jul	1371	1108	1104	1104	1028	938	863	215	0	0	0	0	0	
47	2011 : Aug	1388	1195	1158	1108	1078	1059	1021	964	97	0	0	0	0	
48	2011 : Sep	1848	1704	1608	1515	1473	1441	1396	1285	1219	523	0	0	0	
49	2011 : Oct	2157	1957	1909	1865	1785	1785	1686	1665	1532	1394	74	0	0	
50	2011 : Nov	2105	1982	1930	1930	1867	1838	1757	1757	1738	1645	1544	369	0	
51	2011 : Dec	636	598	574	561	548	530	502	502	502	465	465	376	0	
52	Grand Total :	15433	12638	11744	10544	9537	8473	7350	6388	5088	4027	2083	745	0	
53															
54		Nbr of user loss:													
55	obs_group	Grand Total	0	1	2	3	4	5	6	7	8	9	10	11	
56	2011 : Jan	951	878	73											
57	2011 : Feb	642	214	428											
58	2011 : Mar	1108	216	51	752	89									
59	2011 : Apr	746	176	13	136	416	5								

Retention rate per Age can be calculated by dividing the user profile by the grand total

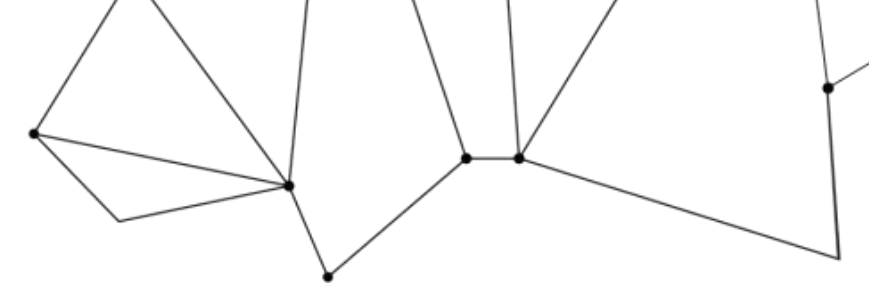
C24														
=C40/\$B40														
	A	B	C	D	E	F	G	H	I	J	K	L	M	N
22		Retention % per age:												
23	obs_group	Grand Total	0	1	2	3	4	5	6	7	8	9	10	11
24	2011 : Jan	100.00%	7.68%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%
25	2011 : Feb	100.00%	66.67%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%
26	2011 : Mar	100.00%	80.51%	75.90%	8.03%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%
27	2011 : Apr	100.00%	76.41%	74.66%	56.43%	0.67%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%
28	2011 : May	100.00%	86.97%	81.69%	73.93%	59.04%	2.17%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%
29	2011 : Jun	100.00%	84.73%	84.73%	83.72%	83.22%	71.64%	10.49%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%
30	2011 : Jul	100.00%	80.82%	80.53%	80.53%	74.98%	68.42%	62.95%	15.68%	0.00%	0.00%	0.00%	0.00%	0.00%
31	2011 : Aug	100.00%	86.10%	83.43%	79.83%	77.67%	76.30%	73.56%	69.45%	6.99%	0.00%	0.00%	0.00%	0.00%
32	2011 : Sep	100.00%	92.21%	87.01%	81.98%	79.71%	77.98%	75.54%	69.53%	65.96%	28.30%	0.00%	0.00%	0.00%
33	2011 : Oct	100.00%	90.73%	88.50%	86.46%	82.75%	82.75%	78.16%	77.19%	71.02%	64.63%	3.43%	0.00%	0.00%
34	2011 : Nov	100.00%	94.16%	91.69%	91.69%	88.69%	87.32%	83.47%	83.47%	82.57%	78.15%	73.35%	17.53%	0.00%
35	2011 : Dec	100.00%	94.03%	90.25%	88.21%	86.16%	83.33%	78.93%	78.93%	78.93%	73.11%	73.11%	59.12%	0.00%
36	Average % :	100.00%	81.89%	76.10%	68.32%	61.80%	54.90%	47.63%	41.39%	32.97%	26.09%	13.50%	4.83%	0.00%
37														
38		User profile:												
39	obs_group	Grand Total	0	1	2	3	4	5	6	7	8	9	10	11
40	2011 : Jan	951	73	0	0	0	0	0	0	0	0	0	0	0
41	2011 : Feb	642	428	0	0	0	0	0	0	0	0	0	0	0
42	2011 : Mar	1108	892	841	89	0	0	0	0	0	0	0	0	0
43	2011 : Apr	746	570	557	421	5	0	0	0	0	0	0	0	0
44	2011 : May	1289	1121	1053	953	761	28	0	0	0	0	0	0	0
45	2011 : Jun	1192	1010	1010	998	992	854	125	0	0	0	0	0	0

7. Retention % through period:

To get the Retention Rate through period value, the value in the retention data per age is multiplied by the previous month

[illegible]

Survival Analysis



8. Color Scales

To highlight the date, use conditional formatting color scale

Retention % through period:														
obs_group	First_transaction		0	1	2	3	4	5	6	7	8	9	10	11
2011 : Jan	100.00%	7.68%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%
2011 : Feb	100.00%	66.67%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%
2011 : Mar	100.00%	80.51%	61.11%	6.10%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%
2011 : Apr	100.00%	76.41%	57.05%	42.14%	0.38%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%
2011 : May	100.00%	86.97%	71.04%	60.40%	43.65%	1.28%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%
2011 : Jun	100.00%	84.73%	71.79%	70.94%	69.68%	59.62%	7.51%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%
2011 : Jul	100.00%	80.82%	65.08%	64.84%	60.38%	51.30%	43.07%	9.87%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%
2011 : Aug	100.00%	86.10%	71.83%	66.60%	62.00%	59.26%	56.12%	51.09%	4.85%	0.00%	0.00%	0.00%	0.00%	0.00%
2011 : Sep	100.00%	92.21%	80.23%	71.33%	65.34%	62.15%	58.90%	52.53%	45.87%	18.67%	0.00%	0.00%	0.00%	0.00%
2011 : Oct	100.00%	90.73%	80.30%	76.52%	71.55%	68.48%	64.68%	60.34%	54.82%	45.90%	2.22%	0.00%	0.00%	0.00%
2011 : Nov	100.00%	94.16%	86.33%	84.06%	81.32%	77.44%	72.88%	69.67%	68.92%	64.52%	57.32%	12.86%	0.00%	0.00%
2011 : Dec	100.00%	94.03%	84.86%	79.61%	76.00%	71.80%	65.78%	62.30%	62.30%	57.71%	53.46%	43.22%	0.00%	0.00%
Average % :	100.00%	81.89%	62.32%	51.99%	42.22%	33.93%	26.15%	19.71%	13.65%	8.60%	3.52%	0.65%	0.00%	0.00%

Cohort Analysis

Cohort analysis focused on survival analysis per group. for this case the group to be selected is based on country, Germany and Ireland

Add a filter to the pivot table to filter countries into Germany and Ireland

Country	Germany									
Count of Customer ID		month_age								
Years	month_group	0	1	2	3	4	5	6	7	
2011	Jan	778								
2011	Feb	214	82							
2011	Mar	171	51	314	41					
2011	Apr	176	13	136	115					
2011	May	168	68	68	192	268				
2011	Jun	182		12	6	138	294	21		
2011	Jul	263	4		76	45	75	292		
2011	Aug	193	37	50	30	19	38	57	345	
2011	Sep	144	96	93	42	32	45	111	66	
2011	Oct	200	48	44	80		99	21	133	
2011	Nov	123	52		63	29	81		19	
2011	Dec	38	24	13	13	18	28			
Grand Total		2650	475	730	658	549	660	502	563	

PivotTable Fields

Choose fields to add to report:

Search

- ☒ Country
- ☒ Customer ID
- ☐ Description
- ☐ Invoice Date
- ☐ Stock Code
- ☐ Invoice

Drag fields between areas below:

Filters	Columns
Country	month_age
Rows	Values
Years	Count of Customer ID
month_group	

Cohort Analysis

Copy and paste the analysis table that was created earlier.

Change the num of user loss data into country-by-country data (Germany/Ireland)

	Nbr of user loss:												
obs_group	Grand Total	0	1	2	3	4	5	6	7	8	9	10	11
2011 : Jan	778	778											
2011 : Feb	296	214	82										
2011 : Mar	577	171	51	314	41								
2011 : Apr	440	176	13	136	115								
2011 : May	764	168	68	68	192	268							
2011 : Jun	653	182		12	6	138	294	21					
2011 : Jul	755	263	4		76	45	75	292					
2011 : Aug	795	193	37	50	30	19	38	57	345	26			
2011 : Sep	773	144	96	93	42	32	45	111	66	124	20		
2011 : Oct	1286	200	48	44	80		99	21	133	138	523		
2011 : Nov	1095	123	52		63	29	81		19	93	101	512	22
2011 : Dec	317	38	24	13	13	18	28			37		89	57
Grand Total :	8529	2650	475	730	658	549	660	502	563	418	644	601	79

Then the user profile table data, retention % per Age and retention % through period will change.

Cohort Analysis

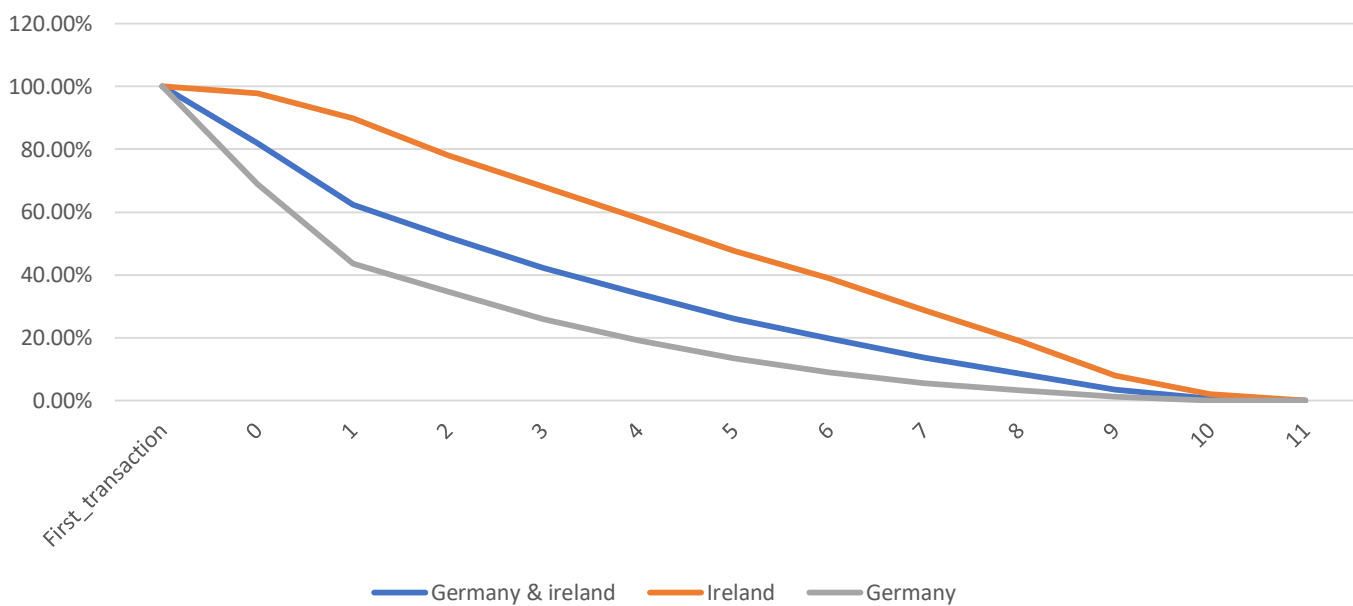
Move the analysis results to a new sheet so that the differences can be compared.

Ireland														
obs_group	First_transaction	0	1	2	3	4	5	6	7	8	9	10	11	
2011: Jan	100.00%	42.20%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	
2011: Feb	100.00%	100.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	
2011: Mar	100.00%	91.53%	83.77%	8.27%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	
2011: Apr	100.00%	100.00%	100.00%	100.00%	1.63%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	
2011: May	100.00%	100.00%	100.00%	93.90%	88.18%	5.01%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	
2011: Jun	100.00%	100.00%	100.00%	100.00%	100.00%	100.00%	19.29%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	
2011: Jul	100.00%	100.00%	100.00%	100.00%	100.00%	92.63%	85.92%	32.35%	0.00%	0.00%	0.00%	0.00%	0.00%	
2011: Aug	100.00%	100.00%	100.00%	100.00%	100.00%	100.00%	100.00%	100.00%	11.97%	0.00%	0.00%	0.00%	0.00%	
2011: Sep	100.00%	100.00%	100.00%	100.00%	100.00%	100.00%	100.00%	100.00%	100.00%	46.73%	0.00%	0.00%	0.00%	
2011: Oct	100.00%	100.00%	100.00%	100.00%	100.00%	100.00%	100.00%	100.00%	100.00%	100.00%	8.50%	0.00%	0.00%	
2011: Nov	100.00%	100.00%	100.00%	100.00%	100.00%	100.00%	100.00%	100.00%	100.00%	100.00%	#####	34.36%	0.00%	
2011: Dec	100.00%	100.00%	100.00%	100.00%	100.00%	100.00%	100.00%	100.00%	100.00%	100.00%	#####	#####	0.00%	
Average % :	100.00%	97.90%	89.90%	78.08%	67.99%	57.98%	47.71%	38.92%	28.66%	18.97%	7.96%	1.96%	0.00%	
Germany														
obs_group	First_transaction	0	1	2	3	4	5	6	7	8	9	10	11	
2011: Jan	100.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	
2011: Feb	100.00%	27.70%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	
2011: Mar	100.00%	70.36%	43.29%	4.37%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	
2011: Apr	100.00%	60.00%	34.23%	14.91%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	
2011: May	100.00%	78.01%	53.91%	41.61%	21.12%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	
2011: Jun	100.00%	72.13%	52.03%	50.70%	48.76%	33.46%	1.55%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	
2011: Jul	100.00%	65.17%	42.12%	41.78%	35.27%	26.53%	18.80%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	
2011: Aug	100.00%	75.72%	53.82%	46.04%	39.52%	35.76%	31.56%	25.12%	1.53%	0.00%	0.00%	0.00%	0.00%	
2011: Sep	100.00%	81.37%	56.11%	39.25%	29.31%	24.38%	19.66%	11.28%	5.06%	0.48%	0.00%	0.00%	0.00%	
2011: Oct	100.00%	84.45%	68.16%	62.39%	54.94%	50.51%	45.04%	39.13%	31.74%	20.90%	0.00%	0.00%	0.00%	
2011: Nov	100.00%	88.77%	74.58%	70.59%	65.76%	59.18%	51.58%	46.54%	45.35%	38.55%	28.28%	0.98%	0.00%	
2011: Dec	100.00%	88.01%	70.80%	61.41%	55.15%	48.08%	38.43%	33.33%	33.33%	26.59%	21.21%	8.28%	0.00%	
Average % :	100.00%	68.93%	43.67%	34.72%	25.80%	19.14%	13.38%	8.89%	5.52%	3.17%	1.24%	0.07%	0.00%	

Cohort Analysis

Create visualizations to make data easier to understand.

Cohort Analysis - Online Transaction Germany vs Ireland



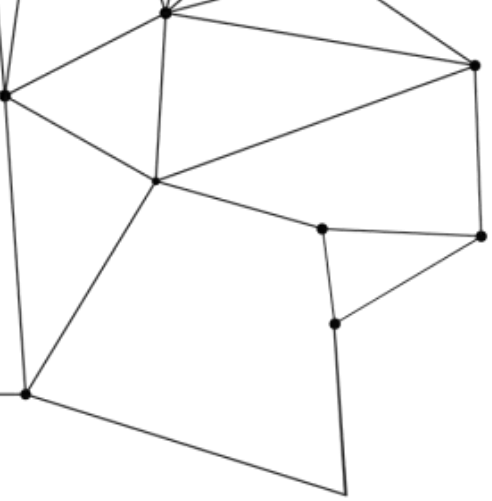
obs_group	First_transaction	0	1	2	3	4	5	6	7	8	9	10	11
Germany & ireland	100.00%	81.89%	62.32%	51.99%	42.22%	33.93%	26.15%	19.71%	13.65%	8.60%	3.52%	0.65%	0.00%
Ireland	100.00%	97.90%	89.90%	78.08%	67.99%	57.98%	47.71%	38.92%	28.66%	18.97%	7.96%	1.96%	0.00%
Germany	100.00%	68.93%	43.67%	34.72%	25.80%	19.14%	13.38%	8.89%	5.52%	3.17%	1.24%	0.07%	0.00%

Summary

From the graph, it can be concluded that Ireland has a better retention rate than Germany. This shows that Irish customers are more loyal than German customers.

German retailers must analyze more deeply to prevent a lack of consumer loyalty





THANK YOU



linkedin.com/noviandi



github.com/noviandi99

